

Causal Time Series Analysis of functional Magnetic Resonance Imaging Data

Alard Roebroeck
Faculty of Psychology & Neuroscience
Maastricht University, the Netherlands

A.ROEBROECK@MAASTRICHTUNIVERSITY.NL

Anil K. Seth
Sackler Centre for Consciousness Science
University of Sussex, UK

A.K.SETH@SUSSEX.AC.UK

Pedro Valdes-Sosa
Cuban Neuroscience Centre, Playa, Cuba

PETER@CNEURO.EDU.CU

Editor: Florin Popescu and Isabelle Guyon

Abstract

This review focuses on dynamic causal analysis of functional magnetic resonance (fMRI) data to infer brain connectivity from a time series analysis and dynamical systems perspective. Causal influence is expressed in the Wiener-Akaike-Granger-Schweder (WAGS) tradition and dynamical systems are treated in a state space modeling framework. The nature of the fMRI signal is reviewed with emphasis on the involved neuronal, physiological and physical processes and their modeling as dynamical systems. In this context, two streams of development in modeling causal brain connectivity using fMRI are discussed: time series approaches to causality in a discrete time tradition and dynamic systems and control theory approaches in a continuous time tradition. This review closes with discussion of ongoing work and future perspectives on the integration of the two approaches.

Keywords: fMRI, hemodynamics, state space model, Granger causality, WAGS influence

1. Introduction

Understanding how interactions between brain structures support the performance of specific cognitive tasks or perceptual and motor processes is a prominent goal in cognitive neuroscience. Neuroimaging methods, such as Electroencephalography (EEG), Magnetoencephalography (MEG) and functional Magnetic Resonance Imaging (fMRI) are employed more and more to address questions of functional connectivity, inter-region coupling and networked computation that go beyond the ‘where’ and ‘when’ of task-related activity (Friston, 2002; Horwitz et al., 2000; McIntosh, 2004; Salmelin and Kujala, 2006; Valdes-Sosa et al., 2005a). A network perspective onto the parallel and distributed processing in the brain - even on the large scale accessible by neuroimaging methods - is a promising approach to enlarge our understanding of perceptual, cognitive and motor functions. Functional Magnetic Resonance Imaging (fMRI) in particular is increasingly used not only to localize structures involved in cognitive and perceptual processes but also to study the connectivity in large-scale brain networks that support these functions.

Generally a distinction is made between three types of brain connectivity. *Anatomical connectivity* refers to the physical presence of an axonal projection from one brain area to another. Identification of large axon bundles connecting remote regions in the brain has recently become possible non-invasively in vivo by diffusion weighted Magnetic resonance imaging (DWMRI) and fiber tractography analysis (Johansen-Berg and Behrens, 2009; Jones, 2010). *Functional connectivity* refers to the correlation structure (or more generally: any order of statistical dependency) in the data such that brain areas can be grouped into interacting networks. Finally, *effective connectivity* modeling moves beyond statistical dependency to measures of directed influence and causality within the networks constrained by further assumptions (Friston, 1994).

Recently, effective connectivity techniques that make use of the temporal dynamics in the fMRI signal and employ time series analysis and systems identification theory have become popular. Within this class of techniques two separate developments have been most used: Granger causality analysis (GCA; Goebel et al., 2003; Roebroeck et al., 2005; Valdes-Sosa, 2004) and Dynamic Causal Modeling (DCM; Friston et al., 2003). Despite the common goal, there seem to be differences between the two methods. Whereas GCA explicitly models temporal precedence and uses the concept of Granger causality (or G-causality) mostly formulated in a discrete time-series analysis framework, DCM employs a biophysically motivated generative model formulated in a continuous time dynamic system framework. In this chapter we will give a general causal time-series analysis perspective onto both developments from what we have called the Wiener-Akaike-Granger-Schweder (WAGS) influence formalism (Valdes-Sosa et al., in press).

Effective connectivity modeling of neuroimaging data entails the estimation of multivariate mathematical models that benefits from a state space formulation, as we will discuss below. Statistical inference on estimated parameters that quantify the directed influence between brain structures, either individually or in groups (model comparison) then provides information on directed connectivity. In such models, brain structures are defined from at least two viewpoints. From a *structural* viewpoint they correspond to a set of “nodes” that comprise a graph, the purpose of causal discovery being the identification of active links in the graph. The structural model contains i) a selection of the structures in the brain that are assumed to be of importance in the cognitive process or task under investigation, ii) the possible interactions between those structures and iii) the possible effects of exogenous inputs onto the network. The exogenous inputs may be under control of the experimenter and often have the form of a simple indicator function that can represent, for instance, the presence or absence of a visual stimulus in the subject’s view. From a *dynamical* viewpoint brain structures are represented by states or variables that describe time varying neural activity within a time-series model of the measured fMRI time-series data. The functional form of the model equations can embed assumptions on signal dynamics, temporal precedence or physiological processes from which signals originate.

We start this review by focusing on the nature of the fMRI signal in some detail in section 2, separating the treatment into neuronal, physiological and physical processes. In section 3 we review two important formal concepts: causal influence in the Wiener-Akaike-Granger-Schweder tradition and the state space modeling framework, with some emphasis on the relations between discrete and continuous time series models. Building on this discussion, section 4 reviews time series modeling of causality in fMRI data. The review

proceeds somewhat chronologically, discussing and comparing the two separate streams of development (GCA and DCM) that have recently begun to be integrated. Finally, section 5 summarizes and discusses the main topics in general dynamic state space models of brain connectivity and provides an outlook on future developments.

2. The fMRI Signal

The fMRI signal reflects the activity within neuronal populations non-invasively with excellent spatial resolution (millimeters down to hundreds of micrometers at high field strength), good temporal resolution (seconds down to hundreds of milliseconds) and whole-brain coverage of the human or animal brain (Logothetis, 2008). Although fMRI is possible with a few different techniques, the Blood Oxygenation Level Dependent (BOLD) contrast mechanism is employed in the great majority of cases. In short, the BOLD fMRI signal is sensitive to changes in blood oxygenation, blood flow and blood volume that result from oxidative glucose metabolism which, in turn, is needed to fuel local neuronal activity (Buxton et al., 2004). This is why fMRI is usually classified as a ‘metabolic’ or ‘hemodynamic’ neuroimaging modality. Its superior spatial resolution, in particular, distinguishes it from other functional brain imaging modalities used in humans, such as EEG, MEG and Positron Emission Tomography (PET). Although its temporal resolution is far superior to PET (another ‘metabolic’ neuroimaging modality) it is still an order of magnitude below that of EEG and MEG, resulting in a relatively sparse sampling of fast neuronal processes, as we will discuss below. The final fMRI signal arises from a complex chain of processes that we can classify into neuronal, physiological and physical processes (Uludag et al., 2005), each of which contain some crucial parameters and variables and have been modeled in various ways as illustrated in Figure 1. We will discuss each of the three classes of processes to explain the intricacies involved in trying to model this causal chain of events with the ultimate goal of estimating neuronal activity and interactions from the measured fMRI signal.

On the neuronal level, it is important to realize that fMRI reflects certain aspects of neuronal functioning more than others. A wealth of processes are continuously in operation at the microscopic level (i.e. in any single neuron), including maintaining a resting potential, post-synaptic conduction and integration (spatial and temporal) of graded excitatory and inhibitory post synaptic potentials (EPSPs and IPSPs) arriving at the dendrites, sub-threshold dynamic (possibly oscillatory) potential changes, spike generation at the axon hillock, propagation of spikes by continuous regeneration of the action potential along the axon, and release of neurotransmitter substances into the synaptic cleft at arrival of an action potential at the synaptic terminal. There are many different types of neurons in the mammalian brain that express these processes in different degrees and ways. In addition, there are other cells, such as glia cells, that perform important processes, some of them possibly directly relevant to computation or signaling. As explained below, the fMRI signal is sensitive to the local oxidative metabolism in the brain. This means that, indirectly, it mainly reflects the most energy consuming of the neuronal processes. In primates, post-synaptic processes account for the great majority (about 75%) of the metabolic costs of neuronal signaling events (Attwell and Iadecola, 2002). Indeed, the greater sensitivity of fMRI to post-synaptic activity, rather than axon generation and propagation (‘spiking’), has been experimentally verified. For instance, in a simultaneous invasive electrophysiology

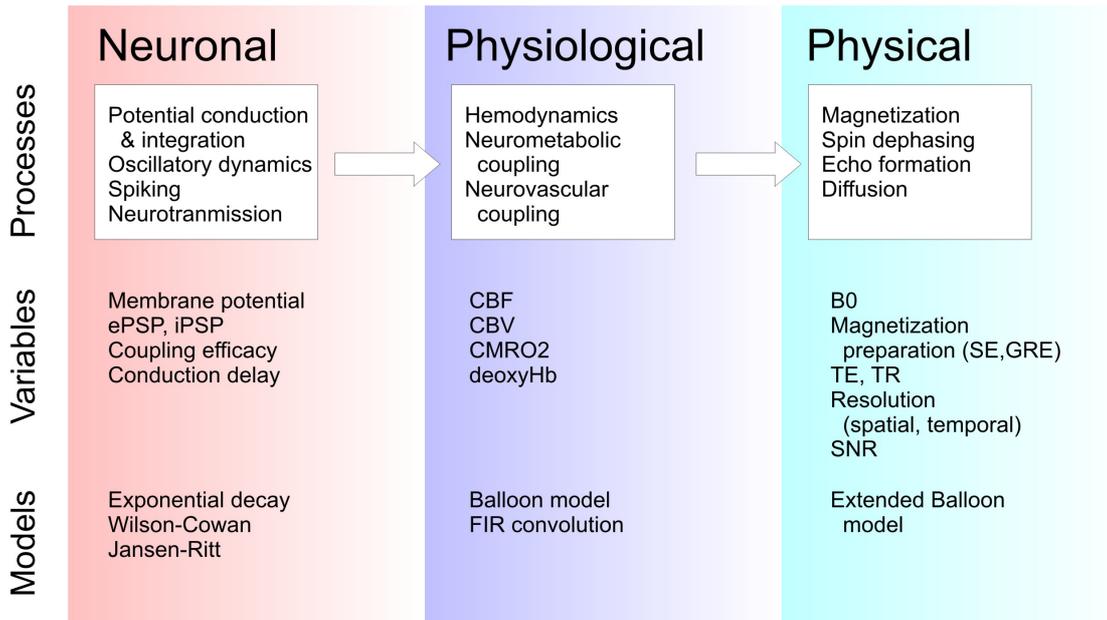


Figure 1: The neuronal, physiological and physical processes (top row) and variables and parameters involved (middle row) in the complex causal chain of events that leads to the formation of the fMRI signal. The bottom row lists some mathematical models of the sub-processes that play a role in the analysis and modeling of fMRI signals. See main text for further explanation.

and fMRI measurement in the primate, Logothetis and colleagues (Logothetis et al., 2001) found the fMRI signal to be more correlated to the mean Local Field Potential (LFP) of the electrophysiological signal, known to reflect post-synaptic graded potentials, than to high-frequency and multi-unit activity, known to reflect spiking. In another study it was shown that, by suppressing action potentials while keeping LFP responses intact by injecting a serotonin agonist, the fMRI response remained intact, again suggesting that LFP is a better predictor for fMRI activity (Rauch et al., 2008). These results confirmed earlier results obtained on the cerebellum of rats (Thomsen et al., 2004).

Neuronal activity, dynamics and computation can be modeled at a different levels of abstraction, including the macroscopic (whole brain areas), mesoscopic (sub-areas to cortical columns) and microscopic level (individual neurons or groups of these). The levels most relevant to modeling fMRI signals are at the macro- and mesoscopic levels. Macroscopic models used to represent considerable expanses of gray matter tissue or sub-cortical structures as Regions Of Interest (ROIs) prominently include single variable deterministic (Friston et al., 2003) or stochastic (autoregressive; Penny et al., 2005; Roebroeck et al., 2005; Valdes-Sosa et al., 2005b) exponential activity decay models. Although the simplicity of such models entail a large degree of abstraction in representing neuronal activity dynamics, their modest complexity is generally well matched to the limited temporal resolution available in fMRI. Nonetheless, more complex multi-state neuronal dynamics models have been investigated in the context of fMRI signal generation. These include the 2 state variable Wilson-Cowan model (Marreiros et al., 2008), with one excitatory and one inhibitory sub-population per ROI and the 3 state variable Jansen-Rit model with a pyramidal excitatory output population and an inhibitory and excitatory interneuron population, particularly in the modeling of simultaneously acquired fMRI and EEG (Valdes-Sosa et al., 2009).

The physiology and physics of the fMRI signal is most easily explained by starting with the physics. We will give a brief overview here and refer to more dedicated overviews (Haacke et al., 1999; Uludag et al., 2005) for extended treatment. The hallmark of Magnetic Resonance (MR) spectroscopy and imaging is the use of the resonance frequency of magnetized nuclei possessing a magnetic moment, mostly protons (hydrogen nuclei, ^1H), called ‘spins’. Radiofrequency antennas (RF coils) can measure signal from ensembles of spins that resonate in phase at the moment of measurement. The first important physical factor in MR is the main magnetic field strength (B_0), which determines both the resonance frequency (directly proportional to field-strength) and the baseline signal-to-noise ratio of the signal, since higher fields make a larger proportion of spins in the tissue available for measurement. The most used field strengths for fMRI research in humans range from 1.5T (Tesla) to 7T. The second important physical factor – containing several crucial parameters – is the MR pulse-sequence that determines the magnetization preparation of the sample and the way the signal is subsequently acquired. The pulse sequence is essentially a series of radiofrequency pulses, linear magnetic gradient pulses and signal acquisition (readout) events (Bernstein et al., 2004; Haacke et al., 1999). An important variable in a BOLD fMRI pulse sequence is whether it is a gradient-echo (GRE) sequence or a spin-echo (SE) sequence, which determines the granularity of the vascular processes that are reflected in the signal, as explained later this section. These effects are further modulated by the echo-time (time to echo; TE) and repetition time (time to repeat; TR) that are usually set by the end-user of the pulse sequence. Finally, an important variable within the pulse sequence is the type of

spatial encoding that is employed. Spatial encoding can primarily be achieved with gradient pulses and it embodies the essence of ‘Imaging’ in MRI. It is only with spatial encoding that signal can be localized to certain ‘voxels’ (volume elements) in the tissue. A strength of fMRI as a neuroimaging technique is that an adjustable trade-off is available to the user between spatial resolution, spatial coverage, temporal resolution and signal-to-noise ratio (SNR) of the acquired data. For instance, although fMRI can achieve excellent spatial resolution at good SNR and reasonable temporal resolution, one can choose to sacrifice some spatial resolution to gain a better temporal resolution for any given study. Note, however, that this concerns the resolution and SNR of the data *acquisition*. As explained below, the physiology of fMRI can put fundamental limitations on the nominal resolution and SNR that is achieved in relation to the neuronal processes of interest.

On the physiological level, the main variables that mediate the BOLD contrast in fMRI are cerebral blood flow (CBF), cerebral blood volume (CBV) and the cerebral metabolic rate of oxygen (CMRO₂) which all change the oxygen saturation of the blood (as usefully quantified by the concentration of deoxygenated hemoglobin). The BOLD contrast is made possible by the fact that oxygenation of the blood changes its magnetic susceptibility, which has an effect on the MR signal as measured in GRE and SE sequences. More precisely, oxygenated and deoxygenated hemoglobin (oxy-Hb and deoxy-Hb) have different magnetic properties, the former being diamagnetic and the latter paramagnetic. As a consequence, deoxygenated blood creates local microscopic magnetic field gradients, such that local spin ensembles dephase, which is reflected in a lower MR signal. Conversely oxygenation of blood above baseline lowers the concentration of deoxy-Hb, which decreases local spin dephasing and results in a higher MR signal. This means that fMRI is directly sensitive to the relative amount of oxy- and deoxy Hb and to the fraction of cerebral tissue that is occupied by blood (the CBV), which are controlled by local neurovascular coupling processes. Neurovascular processes, in turn, are tightly coupled to neurometabolic processes controlling the rate of oxidative glucose metabolism (the CMRO₂) that is needed to fuel neural activity.

Naively one might expect local neuronal activity to quickly increase CMRO₂ and increase the local concentration of deoxy-Hb, leading to a lowering of the MR signal. However, this transient increase in deoxy-Hb or the initial dip in the fMRI signal is not consistently observed and, thus, there is a debate whether this signal is robust, elusive or simply not existent (Buxton, 2001; Ugurbil et al., 2003; Uludag, 2010). Instead, early experiments showed that the dynamics of blood flow and blood volume, the hemodynamics, lead to a robust BOLD signal *increase*. Neuronal activity is quickly followed by a large CBF increase that serves the continued functioning of neurons by clearing metabolic by-products (such as CO₂) and supplying glucose and oxy-Hb. This CBF response is an overcompensating response, supplying much more oxy-Hb to the local blood system than has been metabolized. As a consequence, within 1-2 seconds, the oxygenation of the blood increases and the MR signal increases. The increased flow also induces a ‘ballooning’ of the blood vessels, increasing CBV, the proportion of volume taken up by blood, further increasing the signal.

A mathematical characterization of the hemodynamic processes in BOLD fMRI at 1.5-3T has been given in the biophysical balloon model (Buxton et al., 2004, 1998), schematized in Figure 2A. A simplification of the full balloon model has become important in causal models of brain connectivity (Friston et al., 2000). In this simplified model, the dependence

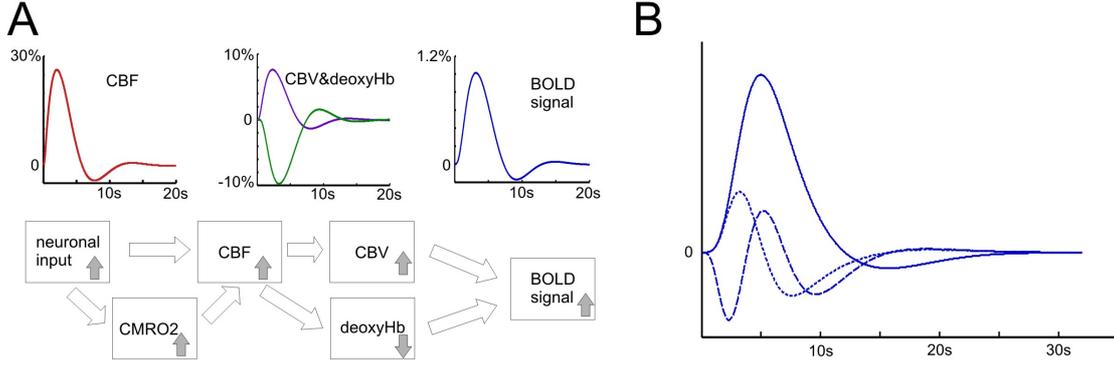


Figure 2: A: Simplified causal chain of hemodynamic events as modeled by the balloon model. Grey arrows show how variable increases (decreases) tend to relate to each other. The dynamic changes after a brief pulse of neuronal activity are plotted for CBF (in red), CBV (in purple), deoxyHb (in green) and BOLD signal (in blue). B: A more abstract representation of the hemodynamic response function as a set of linear basis functions acting as convolution kernels (arbitrary amplitude scaling). Solid line: canonical two-gamma HRF; Dotted line: time derivative; Dashed line: dispersion derivative.

of fractional fMRI signal change $\frac{\Delta S}{S}$, on normalized cerebral blood flow f , normalized cerebral blood volume v and normalized deoxyhemoglobin content q is modeled as:

$$\frac{\Delta S}{S} = V_0 \cdot \left[k_1 \cdot (1 - q) + k_2 \cdot \left(1 - \frac{q}{v}\right) + k_3 \cdot (1 - v) \right] \quad (1)$$

$$\dot{v}_t = \frac{1}{\tau_0} \left(f_t - v_t^{1/\alpha} \right) \quad (2)$$

$$\dot{q}_t = \frac{1}{\tau_0} \left(\frac{f_t \left(1 - (1 - E_0)^{1/f_t}\right)}{E_0} - \frac{q_t}{v_t^{1-1/\alpha}} \right) \quad (3)$$

The term E_0 is the resting oxygen extraction fraction, V_0 is the resting blood volume fraction, τ_0 is the mean transit time of the venous compartment, α is the stiffness component of the model balloon and $\{k_1, k_2, k_3\}$ are calibration parameters. The main simplifications of this model with respect to a more complete balloon model (Buxton et al., 2004) are a one-to-one coupling of flow and volume in (2), thus neglecting the actual balloon effect, and a perfect coupling between flow and metabolism in (3). Friston et al. (2000) augment this model with a putative relation between the a neuronal activity variable z , a flow-inducing signal s , and the normalized cerebral blood flow f . They propose the following relations in which neuronal activity z causes an increase in a vasodilatory signal that is subject to autoregulatory feedback:

$$\dot{s}_t = z_t - \frac{1}{\tau_s} s_t - \frac{1}{\tau_f^2} (f_t - 1) \quad (4)$$

$$\dot{f}_t = s_t \quad (5)$$

Here τ_s is the signal decay time constant, τ_f is the time-constant of the feedback autoregulatory mechanism¹, and f is the flow normalized to baseline flow. The physiological interpretation of the autoregulatory mechanism is unspecified, leaving us with a neuronal activity variable z that is measured in units of s^{-2} . The physiology of the hemodynamics contained in differential equations (2) to (5), on the other hand, is more readily interpretable, and when integrated for a brief neuronal input pulse shows the behavior as described above (Figure 2A, upper panel). This simulation highlights a few crucial features. First, the hemodynamic response to a brief neural activity event is sluggish and delayed, entailing that the fMRI BOLD signal is a delayed and low-pass filtered version of underlying neuronal activity. More than the distorting effects of hemodynamic processes on the temporal structure of fMRI signals per se, it is the *difference* in hemodynamics in different parts of the brain that forms a severe confound for dynamic brain connectivity models. Particularly, the delay imposed upon fMRI signals with respect to the underlying neural activity is known to vary between subjects and between different brain regions of the same subject (Aguirre et al., 1998; Saad et al., 2001). Second, although CBF, CBV and deoxyHb changes range in the tens of percents, the BOLD signal change at 1.5T or 3T is in the range of 0.5-2%. Nevertheless, the SNR of BOLD fMRI in general is very good in comparison to electrophysiological techniques like EEG and MEG.

Although the balloon model and its variations have played an important role in describing the transient features of the fMRI response and inferring neuronal activity, simplified ways of representing the BOLD signal responses are very often used. Most prominent among these is a linear finite impulse response (FIR) convolution with a suitable kernel. The most used single convolution kernel characterizing the ‘canonical’ hemodynamic response is formed by a superposition of two gamma functions (Glover, 1999), the first characterizing the initial signal increase, the second the later negative undershoot (Figure 2B, solid line):

$$\begin{aligned} h(t) &= m_1 t^{\tau_1} e^{-l_1 t} - c m_2 t^{\tau_2} e^{-l_2 t} \\ m_i &= \max(t^{\tau_i} e^{-l_i t}) \end{aligned} \quad (6)$$

With times-to-peak in seconds $\tau_1 = 6$, $\tau_2 = 16$, scale parameters l_i (typically equal to 1) and a relative amplitude of undershoot to peak of $c = 1/6$.

Often, the canonical two-gamma HRF kernel is augmented with one or two additional orthogonalized convolution kernels: a temporal derivative and a dispersion derivative. Together the convolution kernels form a flexible basis function expansion of possible HRF shapes, with the temporal derivative of the canonical accounting for variation in the response delay and the dispersion derivative accounting for variations in temporal response width (Henson et al., 2002; Liao et al., 2002). Thus, the linear basis function representation is a more abstract characterization of the HRF (i.e. further away from the physiology) that still captures the possible variations in responses.

It is an interesting property of hemodynamic processes that, although they are characterized by a large overcompensating reaction to neuronal activity, their effects are highly

1. Note that we have reparametrized the equation here in terms of τ_f^2 to make τ_f a proper time constant in units of seconds

local. The locality of the hemodynamic response to neuronal activity limits the actual spatial resolution of fMRI. The path blood inflow in the brain is from large arteries through arterioles into capillaries where exchange with neuronal tissue takes place at a microscopic level. Blood outflow takes place via venules into the larger veins. The main regulators of blood flow are the arterioles that are surrounded by smooth muscle, although arteries and capillaries are also thought to be involved in blood flow regulation (Attwell et al., 2010). Different hemodynamic parameters have different spatial resolutions. While CBV and CBF changes in all compartments but mostly venules, oxygenation changes mostly in the venules and veins. Thus, the achievable spatial resolution with fMRI is limited by its specificity to the smaller arterioles and venules and microscopic capillaries supplying the tissue, rather than the larger supplying arteries draining veins. The larger vessels have a larger domain of supply or extraction and, as a consequence, their signal is blurred and mislocalized with respect to active tissue. Here, physiology and physics interact in an important way. It can be shown theoretically – by the effects of thermal motion of spin diffusion over time and the distance of the spins to deoxy-Hb – that the origin of the BOLD signal in SE sequences at high main field strengths (larger than 3T) is much more specific to the microscopic vasculature than to the larger arteries and veins. This does not hold for GRE sequences or SE sequences at lower field strengths. The cost of this greater specificity and higher effective spatial resolution is that SE-BOLD has a lower intrinsic SNR than GRE-BOLD. The balloon model equations above are specific to GRE-BOLD at 1.5T and 3T and have been extended to reflect diffusion effects for higher field strengths (Uludag et al., 2009).

In summary, fMRI is an indirect measure of neuronal and synaptic activity. The physiological quantities directly determining signal contrast in BOLD fMRI are hemodynamic quantities such as cerebral blood flow and volume and oxygen metabolism. fMRI can achieve an excellent spatial resolution (millimeters down to hundreds of micrometers at high field strength) with good temporal resolution (seconds down to hundreds of milliseconds). The potential to resolve neuronal population interactions at a high spatial resolution is what drives attempts at causal time series modeling of fMRI data. However, the significant aspects of fMRI that pose challenges for such attempts are i) the enormous dimensionality of the data that contains hundreds of thousands of channels (voxels) ii) the temporal convolution of neuronal events by sluggish hemodynamics that can differ between remote parts of the brain and iii) the relatively sparse temporal sampling of the signal.

3. Causality and state-space models

The inference of causal influence relations from statistical analysis of observed data has two dominant approaches. The first approach is in the tradition of *Granger causality* or *G-causality*, which has its signature in improved predictability of one time series by another. The second approach is based on graphical models and the notion of intervention (Glymour, 2003), which has been formalized using a Bayesian probabilistic framework termed *causal calculus* or *do-calculus* (Pearl, 2009). Interestingly, recent work has combined the two approaches in a third line of work, termed *Dynamic Structural Systems* (White and Lu, 2010). The focus here will be on the first approach, initially firmly rooted in econometrics and time-series analysis. We will discuss this tradition in a very general form, acknowledging

early contributions from Wiener, Akaike, Granger and Schweder and will follow (Valdes-Sosa et al., in press) in referring to the crucial concept as *WAGS influence*.

3.1. Wiener-Akaike-Granger-Schweder (WAGS) influence

The crucial premise of the WAGS statistical causal modeling tradition is that a cause must precede and increase the predictability of its effect. In other words: a variable X_2 influences another variable X_1 if the prediction of X_1 improves when we use past values of X_2 , given that all other relevant information (importantly: the past of X_1 itself) is taken into account. This type of reasoning can be traced back at least to Hume and is particularly popular in analyzing dynamical data measured as time series. In a formal framework it was originally proposed (in an abstract form) by Wiener (Wiener, 1956), and then introduced into practical data analysis and popularized by Granger (Granger, 1969). A point stressed by Granger is that increased predictability is a necessary but not sufficient condition for a causal relation between time series. In fact, Granger distinguished true causal relations – only to be inferred in the presence of knowledge of the state of the whole universe – from “prima facie” causal relations that we refer to as “influence” in agreement with other authors (Commenges and Gegout-Petit, 2009). Almost simultaneous with Grangers work, Akaike (Akaike, 1968), and Schweder (Schweder, 1970) introduced similar concepts of influence, prompting (Valdes-Sosa et al., in press) to coin the term WAGS influence (for Wiener-Akaike-Granger-Schweder). This is a generalization of a proposal by placeAalen (Aalen, 1987; Aalen and Frigessi, 2007) who was among the first to point out the connections between Granger’s and Schweder’s influence concepts. Within this framework we can define several general types of WAGS influence, which are applicable to both Markovian and non-Markovian processes, in discrete or continuous time.

For three vector time series $X_1(t)$, $X_2(t)$, $X_3(t)$ we wish to know if time series $X_1(t)$ is influenced by time series $X_2(t)$ conditional on $X_3(t)$. Here $X_3(t)$ can be considered any set of relevant time series to be controlled for. Let $X[a, b] = \{X(t), t \in [a, b]\}$ denote the history of a time series in the discrete or continuous time interval $[a, b]$. The first categorical distinction is based on what part of the present or future of $X_1(t)$ can be predicted by the past or present of $X_2(\tau_2)$ $\tau_2 \leq t$. This leads to the following classification (Florens, 2003; Florens and Fougere, 1996):

1. If $X_2(\tau_2) : \tau_2 < t$, can influence any future value of $X_1(t)$ it is a *global* influence.
2. If $X_2(\tau_2) : \tau_2 < t$, can influence $X_1(t)$ at time t it is a *local* influence.
3. If $X_2(\tau_2) : \tau_2 = t$, can influence $X_1(t)$ it is a *contemporaneous* influence.

A second distinction is based on predicting the whole probability distribution (*strong* influence) or only given moments (*weak* influence). Since the most natural formal definition is one of independence, every influence type amounts to the negation of an independence statement. The two classifications give rise to six types of independence and corresponding influence as set out in Table 1.

To illustrate, $X_1(t)$ is **strongly, conditionally, and globally independent** of $X_2(t)$ given $X_3(t)$, if

$$P(X_1(\infty, t] | X_1(t, -\infty], X_2(t, -\infty], X_3(t, -\infty]) = P(X_1(\infty, t] | X_1(t, -\infty], X_3(t, -\infty])$$

Table 1: Types of Influence defined by absence of the corresponding independence relations. See text for acronym definitions.

	Strong (Probability Distribution)	Weak (Expectation)
Global (All horizons)	By absence of strong, conditional, global independence: $X_2(t)$ SCGi $X_1(t) X_3(t)$	By absence of weak, conditional, global independence: $X_2(t)$ WCGi $X_1(t) X_3(t)$
Local (Immediate future)	By absence of strong, conditional, local independence: $X_2(t)$ SCLi $X_1(t) X_3(t)$	By absence of weak, conditional, local in- dependence: $X_2(t)$ WCLi $X_1(t) X_3(t)$
Contemporaneous	By absence of strong, conditional, con- temporaneous indepen- dence: $X_2(t)$ SCCi $X_1(t) X_3(t)$	By absence of weak, conditional, con- temporaneous indepen- dence: $X_2(t)$ WCCi $X_1(t) X_3(t)$

That is: the probability distribution of the future values of X_1 does not depend on the past of X_2 , given that the influence of the past of both X_1 and X_3 has been taken into account. When this condition does not hold we say $X_2(t)$ **strongly, conditionally, and globally influences (SCGi)** $X_1(t)$ given $X_3(t)$. Here we use a convention for intervals $[a,b)$ which indicates that the left endpoint is included but not the right and that b precedes a . Note that the whole future of X_t is included (hence the term “global”). And the whole past of all time series is considered. This means these definitions accommodate non-Markovian processes (for Markovian processes, we only consider the previous time point). Furthermore, these definitions do not depend on an assumption of linearity or any given functional relationship between time series. Note also that this definition is appropriate for point processes, discrete and continuous time series, even for categorical (qualitative valued) time series. The only problem with this formulation is that it calls on the whole probability distribution and therefore its practical assessment requires the use of measures such as mutual information that estimate the probability densities nonparametrically.

As an alternative, weak concepts of influence can be defined based on expectations. Consider **weak conditional local independence** *in discrete time*, which is defined:

$$E[X_1[t + \Delta t] | X_1[t, -\infty], X_2[t, -\infty], X_3[t, -\infty]] = E[X_1[t + \Delta t] | X_1[t, -\infty], X_3[t, -\infty]] \quad (7)$$

When this condition does not hold we say X_2 **weakly, conditionally and locally influences (WCLi)** X_1 given X_3 . To make the implementation this definition insightful,

consider a discrete first-order vector auto-regressive (VAR) model for $X = [X_1 X_2 X_3]$:

$$X [t + \Delta t] = AX [t] + e [t + \Delta t] \quad (8)$$

For this case $E [X[t + \Delta t] | X[t, -\infty]] = AX [t]$, and analyzing influence reduces to finding which of the autoregressive coefficients are zero. Thus, many proposed operational tests of WAGS influence, particularly in fMRI analysis, have been formulated as tests of discrete autoregressive coefficients, although not always of order 1. Within the same model one can operationalize **weak conditional instantaneous independence** *in discrete time* as zero off-diagonal entries in the co-variance matrix of the innovations $e[t]$:

$$\Sigma_e = cov [X [t + \Delta t] | X [t, -\infty]] = E [X [t + \Delta t] X' [t + \Delta t] | X [t, -\infty]]$$

In comparison **weak conditional local independence** in *continuous time* is defined:

$$E [Y_1[t] | Y_1(t, -\infty), Y_2(t, -\infty), Y_3(t, -\infty)] = E [Y_1[t] | Y_1(t, -\infty), Y_3(t, -\infty)] \quad (9)$$

Now consider a first-order stochastic differential equation (SDE) model for $Y = [Y_1 Y_2 Y_3]$:

$$dY = BY dt + d\omega \quad (10)$$

Then, since ω is a Wiener process with zero-mean white Gaussian noise as a derivative, $E [Y[t] | Y(t, -\infty)] = BY (t)$ and analysing influence amounts to estimating the parameters B of the SDE. However, if one were to observe a discretely sampled version $X[k] = Y (k\Delta t)$ at sampling interval Δt and model this with the discrete autoregressive model above, this would be inadequate to estimate the SDE parameters for large Δt , since the exact relations between continuous and discrete system matrices are known to be:

$$\begin{aligned} \mathbf{A} &= e^{\mathbf{B}\Delta t} = \mathbf{I} + \sum_{i=1}^{\infty} \frac{\Delta t^i}{i!} \mathbf{B}^i \\ \Sigma_e &= \int_t^{t+\Delta t} e^{\mathbf{B}s} \Sigma_{\omega} e^{\mathbf{B}s} ds \end{aligned} \quad (11)$$

The power series expansion of the matrix exponential in the first line shows A to be a weighted sum of successive matrix powers B^i of the continuous time system matrix. Thus, the A will contain contributions from direct (in B) and indirect (in i steps in B^i) causal links between the modeled areas. The contribution of the more indirect links is progressively down-weighted with the number of causal steps from one area to another and is smaller when the sampling interval Δt is smaller. This makes clear that multivariate *discrete* signal models have some undesirable properties for coarsely sampled signals (i.e. a large Δt with respect to the system dynamics), such as fMRI data. Critically, entirely ruling out *indirect* influences is not actually achieved merely by employing a multivariate discrete model. Furthermore, estimated WAGS influence (particularly the relative contribution of indirect links) is dependent on the employed sampling interval. However, the discrete system matrix still represents the presence and direction of influence, possibly mediated through other regions.

When the goal is to estimate WAGS influence for discrete data starting from a continuous time model, one has to model explicitly the mapping to discrete time. Mapping continuous time predictions to discrete samples is a well known topic in engineering and can be

solved by explicit integration over discrete time steps as performed in (11) above. Although this defines the mapping from continuous to discrete parameters, it does not solve the reverse assignment of estimating continuous model parameters from discrete data. Doing so requires a solution to the aliasing problem (McCrorie, 2003) in continuous stochastic system identification by setting sufficient conditions on the matrix logarithm function to make B above identifiable (uniquely defined) in terms of A . Interesting in this regard is a line of work initiated by Bergstrom (Bergstrom, 1966, 1984) and Phillips (Phillips, 1973, 1974) studying the estimation of continuous time Autoregressive models (McCrorie, 2002), and continuous time Autoregressive Moving Average Models (Chambers and Thornton, 2009) from discrete data. This work rests on the observation that the lag zero covariance matrix Σ_e will show contemporaneous covariance even if the continuous covariance matrix Σ_ω is diagonal. In other words, the discrete noise becomes correlated over the discrete time-series because the random fluctuations are aggregated over time. Rather than considering this a disadvantage, this approach tries to use both lag information (the AR part) and zero-lag covariance information to identify the underlying continuous linear model.

Notwithstanding the desirability of a continuous time model for consistent inference on WAGS influence, there are a few invariances of discrete VAR models, or more generally discrete Vector Autoregressive Moving Average (VARMA) models that allow their carefully qualified usage in estimating causal influence. The VAR formulation of WAGS influence has the property of invariance under invertible linear filtering. More precisely, a general measure of influence remains unchanged if channels are each pre-multiplied with different invertible lag operators (Geweke, 1982). However, in practice the order of the estimated VAR model would need to be sufficient to accommodate these operators. Beyond invertible linear filtering, a VARMA formulation has further invariances. Solo (2006) showed that causality in a VARMA model is preserved under sampling and additive noise. More precisely, if both local and contemporaneous influence is considered (as defined above) the VARMA measure is preserved under sampling and under the addition of independent but colored noise to the different channels. Finally, Amendola et al. (2010) shows the class of VARMA models to be closed under aggregation operations, which include both sampling and time-window averaging.

3.2. State-space models

A general state-space model for a continuous vector time-series $y(t)$ can be formulated with the set of equations:

$$\begin{aligned} \dot{x}(t) &= f(x(t), v(t), \Theta) + \omega(t) \\ y(t) &= g(x(t), v(t), \Theta) + \varepsilon(t) \end{aligned} \tag{12}$$

This expresses the observed time-series $y(t)$ as a function of the state variables $x(t)$, which are possibly hidden (i.e. unobserved) and observed exogenous inputs $v(t)$, which are possibly under control. All parameters in the model are grouped into Θ . Note that some generality is sacrificed from the start since f and g do not depend on t (The model is autonomous and generates stationary processes) or on $\omega(t)$ or $\varepsilon(t)$, that is: noise enters only additively. The first set of equations, the *transition equations* or *state equations*, describe the evolution of the dynamic system over time in terms of stochastic differential equations

(SDEs, though technically only when $\omega(t) = \Sigma \dot{w}(t)$ with $w(t)$ a Wiener process), capturing relations among the hidden state variables $x(t)$ themselves and the influence of exogenous inputs $v(t)$. The second set of equations, the *observation equations* or *measurement equations*, describe how the measurement variables $y(t)$ are obtained from the instantaneous values of the hidden state variables $x(t)$ and the inputs $v(t)$. In fMRI experiments the exogenous inputs $v(t)$ mostly reflect experimental control and often have the form of a simple indicator function that can represent, for instance, the presence or absence of a visual stimulus. The vector-functions f and g can generally be non-linear.

The state-space formalism allows representation of a very large class of stochastic processes. Specifically, it allows representation of both so-called ‘*black-box*’ models, in which parameters are treated as means to adjust the fit to the data without reflecting physically meaningful quantities, and ‘*grey-box*’ models, in which the adjustable parameters *do* have a physical or physiological (in the case of the brain) interpretation. A prominent example of a black-box model in econometric time-series analysis and systems identification is the (discrete) Vector Autoregressive Moving Average model with exogenous inputs (VARMAX model) defined as (Ljung, 1999; Reinsel, 1997):

$$\begin{aligned} F(B)y_t &= G(B)v_t + L(B)e_t \Leftrightarrow \\ \sum_{i=0}^p F_i y_{t-i} &= \sum_{j=0}^s G_j v_{t-j} + \sum_{k=0}^q L_k e_{t-k} \end{aligned} \quad (13)$$

Here, the backshift operator B is defined, for any η_t as $B^i \eta_t = \eta_{t-i}$ and F , G and L are polynomials in the backshift operator, such that e.g. $F(B) = \sum_{i=0}^p F_i B^i$ and p , s and q are the dynamic orders of the VARMAX(p,s,q) model. The minimal constraints on (13) to make it identifiable are $F_0 = L_0 = I$, which yields the *standard* VARMAX representation. The VARMAX model and its various reductions (by use of only one or two of the polynomials, e.g. VAR, VARX or VARMA models) have played a large role in time-series prediction and WAGS influence modeling. Thus, in the context of state space models it is important to consider that the VARMAX model form can be equivalently formulated in a discrete linear state space form:

$$\begin{aligned} x_{k+1} &= \mathbf{A}x_k + \mathbf{B}v_k + \mathbf{K}e_k \\ y_k &= \mathbf{C}x_k + \mathbf{D}v_k + e_k \end{aligned} \quad (14)$$

In turn the *discrete* linear state space form can be explicitly accommodated by the *continuous* general state-space framework in (12) when we define:

$$\begin{aligned} f(x(t), v(t), \Theta) &\simeq Fx(t) + Gv(t) & \omega(t) &= \tilde{K}\varepsilon(t) \\ g(x(t), v(t), \Theta) &\simeq Hx(t) + Dv(t) & \Theta &= \{F, G, H, D, \tilde{K}, \Sigma_e\} \end{aligned} \quad (15)$$

Again, the exact relations between the discrete and continuous state space parameter matrices can be derived analytically by explicit integration over time (Ljung, 1999). And, as discussed above, wherever discrete data is used to model continuous influence relations the problems of temporal aggregation and aliasing have to be taken into account.

Although analytic solutions for the discretely sampled continuous linear systems exist, the discretization of the *nonlinear* stochastic model (12) does not have a unique global solution. However, physiological models of neuronal population dynamics and hemodynamics are formulated in continuous time and are mostly nonlinear while fMRI data is inherently

discrete with low sampling frequencies. Therefore, it is the discretization of the *nonlinear* dynamical stochastic models that is especially relevant to causal analysis of fMRI data. A local linearization approach was proposed by (Ozaki, 1992) as bridge between discrete time series models and nonlinear continuous dynamical systems model. Considering the nonlinear state equation without exogenous input:

$$\dot{x}(t) = f(x(t)) + \omega(t). \quad (16)$$

The essential assumption in local linearization (LL) of this nonlinear system is to consider the Jacobian matrix $J(l, m) = \frac{\partial f_l(X)}{\partial X_m}$ as constant over the time period $[t + \Delta t, t]$. This Jacobian plays the same role as the autoregressive matrix in the linear systems above. Integration over this interval gives the solution:

$$x_{k+1} = x_k + J^{-1}(e^{J\Delta t} - I) f(x_k) + e_{k+1} \quad (17)$$

where I is the identity matrix. Note integration should not be computed this way since it is numerically unstable, especially when the Jacobian is poorly conditioned. A list of robust and fast procedures is reviewed in (Valdes-Sosa et al., 2009). This solution is locally linear but crucially it changes with the state at the beginning of each integration interval; this is how it accommodates nonlinearity (i.e., a state-dependent autoregression matrix). As above, the discretized noise shows instantaneous correlations due to the aggregation of ongoing dynamics within the span of a sampling period. Once again, this highlights the underlying mechanism for problems with temporal sub-sampling and aggregation for some discrete time models of WAGS influence.

4. Dynamic causality in fMRI connectivity analysis

Two streams of developments have recently emerged that make use of the temporal dynamics in the fMRI signal to analyse directed influence (effective connectivity): Granger causality analysis (GCA; Goebel et al., 2003; Roebroeck et al., 2005; Valdes-Sosa, 2004) in the tradition of time series analysis and WAGS influence on the one hand, and Dynamic Causal Modeling (DCM; Friston et al., 2003) in the tradition of system control on the other hand. As we will discuss in the final section, these approaches have recently started developing into an integrated single direction. However, initially each was focused on separate issues that pose challenges for the estimation of causal influence from fMRI data. Whereas DCM is formulated as an explicit grey box state space model to account for the temporal convolution of neuronal events by sluggish hemodynamics, GCA analysis has been mostly aimed at solving the problem of region selection in the enormous dimensionality of fMRI data.

4.1. Hemodynamic deconvolution in a state space approach

While having a long history in engineering, state space modeling was only introduced recently for the inference of neural states from neuroimaging signals. The earliest attempts targeted estimating hidden neuronal population dynamics from scalp-level EEG data (Hernandez et al., 1996; Valdes-Sosa et al., 1999). This work first advanced the idea that state

space models and appropriate filtering algorithms are an important tool to estimate the trajectories of hidden neuronal processes from observed neuroimaging data if one can formulate an accurate model of the processes leading from neuronal activity to data records. A few years later, this idea was robustly transferred to fMRI data in the form of DCM (Friston et al., 2003). DCM combines three ideas about causal influence analysis in fMRI data (or neuroimaging data in general), which can be understood in terms of the discussion of the fMRI signal and state space models above (Daunizeau et al., 2009a).

First, neuronal interactions are best modeled at the level of unobserved (latent) signals, instead of at the level of observed BOLD signals. This requires a state space model with a dynamic model of neuronal population dynamics and interactions. The original model that was formulated for the dynamics of neuronal states $x = \{x_1, \dots, x_N\}$ is a bilinear ODE model:

$$\dot{x} = \mathbf{A}x + \sum v_j \mathbf{B}^j x + \mathbf{C}v \quad (18)$$

That is, the noiseless neuronal dynamics are characterized by a linear term (with entries in \mathbf{A} representing intrinsic coupling between populations), an exogenous term (with \mathbf{C} representing driving influence of experimental variables) and a bilinear term (with \mathbf{B}^j representing the modulatory influence of experimental variables on coupling between populations). More recent work has extended this model, e.g. by adding a quadratic term (Stephan et al., 2008), stochastic dynamics (Daunizeau et al., 2009b) or multiple state variables per region (Marreiros et al., 2008).

Second, the latent neuronal dynamics are related to observed data by a generative (forward) model that accounts for the temporal convolution of neuronal events by slow and variably delayed hemodynamics. This generative forward model in DCM for fMRI is exactly the (simplified) balloon model set out in section 2. Thus, for every selected region a single state variable represents the neuronal or synaptic activity of a local population of neurons and (in DCM for BOLD fMRI) four or five more represent hemodynamic quantities such as capillary blood volume, blood flow and deoxy-hemoglobin content. All state variables (and the equations governing their dynamics) that serve the mapping of neuronal activity to the fMRI measurements (including the observation equation) can be called the *observation model*. Most of the physiologically motivated generative model in DCM for fMRI is therefore concerned with an observation model encapsulating hemodynamics. The parameters in this model are estimated conjointly with the parameters quantifying neuronal connectivity. Thus, the forward biophysical model of hemodynamics is ‘inverted’ in the estimation procedure to achieve a deconvolution of fMRI time series and obtain estimates of the underlying neuronal states. DCM has also been applied to EEG/MEG, in which case the observation model encapsulates the lead-field matrix from neuronal sources to EEG electrodes or MEG sensors (Kiebel et al., 2009).

Third, the approach to estimating the hidden state trajectories (i.e. filtering and smoothing) and parameter values in DCM is cast in a Bayesian framework. In short, Bayes’ theorem is used to combine priors $p(\Theta|M)$ and likelihood $p(y|\Theta, M)$ into the *marginal likelihood* or *evidence*:

$$p(y|M) = \int p(y|\Theta, M) p(\Theta|M) d\Theta \quad (19)$$

Here, the model M is understood to define the priors on all parameters and the likelihood through the generative models for neuronal dynamics and hemodynamics. A posterior for the parameters $p(\Theta|y, M)$ can be obtained as the distribution over parameters which maximizes the evidence (19). Since this optimization problem has no analytic solution and is intractable with numerical sampling schemes for complex models, such as DCM, approximations must be used. The inference approach for DCM relies on variational Bayes methods (Beal, 2003) that optimize an approximation density $q(\Theta)$ to the posterior. The approximation density is taken to have a Gaussian form, which is often referred to as the ‘‘Laplace approximation’’ (Friston et al., 2007). In addition to the approximate posterior on the parameters, the variational inference will also result into a lower bound on the evidence, sometimes referred to as the ‘‘free energy’’. This lower bound (or other approximations to the evidence, such as the Akaike Information Criterion or the Bayesian Information Criterion) are used for model comparison (Penny et al., 2004). Importantly, these quantities explicitly balance goodness-of-fit against model complexity as a means of avoiding overfitting.

An important limiting aspect of DCM for fMRI is that the models M that are compared also (implicitly) contain an *anatomical model* or *structural model* that contains i) a selection of the ROIs in the brain that are assumed to be of importance in the cognitive process or task under investigation, ii) the possible interactions between those structures and iii) the possible effects of exogenous inputs onto the network. In other words, each model M specifies the nodes and edges in a directed (possibly cyclic) structural graph model. Since the anatomical model also determines the selected part y of the total dataset (all voxels) one cannot use the evidence to compare different anatomical models. This is because the evidence of different anatomical models is defined over different data. Applications of DCM to date invariably use very simple anatomical models (typically employing 3-6 ROIs) in combination with its complex parameter-rich dynamical model discussed above. The clear danger with overly simple anatomical models is that of spurious influence: an erroneous influence found between two selected regions that in reality is due to interactions with additional regions which have been ignored. Prototypical examples of spurious influence, of relevance in brain connectivity, are those between unconnected structures A and B that receive common input from, or are intervened by, an unmodeled region C.

4.2. Exploratory approaches for model selection

Early applications of WAGS influence to fMRI data were aimed at counteracting the problems with overly restrictive anatomical models by employing more permissive anatomical models in combination with a simple dynamical model (Goebel et al., 2003; Roebroeck et al., 2005; Valdes-Sosa, 2004). These applications reflect the observation that estimation of mathematical models from time-series data generally has two important aspects: model selection and model identification (Ljung, 1999). In the *model selection* stage a class of models is chosen by the researcher that is deemed suitable for the problem at hand. In the *model identification* stage the parameters in the chosen model class are estimated from the observed data record. In practice, model selection and identification often occur in a somewhat interactive fashion where, for instance, model selection can be informed by the fit of different models to the data achieved in an identification step. The important point is that model selection involves a mixture of choices and assumptions on the part of the researcher

and the information gained from the data-record itself. These considerations indicate that an important distinction must be made between exploratory and confirmatory approaches, especially in structural model selection procedures for brain connectivity. Exploratory techniques use information in the data to investigate the relative applicability of many models. As such, they have the potential to detect ‘missing’ regions in structural models. Confirmatory approaches, such as DCM, test hypotheses about connectivity within a set of models assumed to be applicable. Sources of common input or intervening causes are taken into account in a multivariate confirmatory model, but only if the employed structural model allows it (i.e. if the common input or intervening node is incorporated in the model).

The technique of Granger Causality Mapping (GCM) was developed to explore all regions in the brain that interact with a single selected reference region using autoregressive modeling of fMRI time-series (Roebroeck et al., 2005). By employing a simple bivariate model containing the reference region and, in turn, every other voxel in the brain, the sources and targets of influence for the reference region can be mapped. It was shown that such an ‘exploratory’ mapping approach can form an important tool in structural model selection. Although a bivariate model does not discern direct from indirect influences, the mapping approach locates potential sources of common input and areas that could act as intervening network nodes. In addition, by settling for a bivariate model one trivially avoids the conflation of direct and indirect influences that can arise in discrete AR model due to temporal aggregation, as discussed above. Other applications of autoregressive modeling to fMRI data have considered full multivariate models on large sets of selected brain regions, illustrating the possibility to estimate high-dimensional dynamical models. For instance, Valdes-Sosa (2004) and Valdes-Sosa et al. (2005b) applied these models to parcellations of the entire cortex in conjunction with sparse regression approaches that enforce an implicit structural model selection within the set of parcels. In another more recent example (Deshpande et al., 2008) a full multivariate model was estimated over 25 ROIs (that were found to be activated in the investigated task) together with an explicit reduction procedure to prune regions from the full model as a structural model selection procedure. Additional variants of VAR model based causal inference that has been applied to fMRI include time varying influence (Havlicek et al., 2010), blockwise (or ‘cluster-wise’) influence from one group of variables to another (Barrett et al., 2010; Sato et al., 2010) and frequency-decomposed influence (Sato et al., 2009).

The initial developments in autoregressive modeling of fMRI data led to a number of interesting applications studying human mental states and cognitive processes, such as gestural communication (Schippers et al., 2010), top-down control of visual spatial attention (Bressler et al., 2008), switching between executive control and default-mode networks (Sridharan et al., 2008), fatigue (Deshpande et al., 2009) and the resting state (Uddin et al., 2009). Nonetheless, the lack of AR models to account for the varying hemodynamics convolving the signals of interest and aggregation of dynamics between time samples has prompted a set of validation studies evaluating the conditions under which discrete AR models can provide reliable connectivity estimates. In (Roebroeck et al., 2005) simulations were performed to validate the use of bivariate AR models in the face of hemodynamic convolution and sampling. They showed that under these conditions (even without variability in hemodynamics) AR estimates for a unidirectional influence are biased towards inferring bidirectional causality, a well known problem when dealing with aggregated time series

(Wei, 1990). They then went on to show that instead unbiased non-parametric inference for bivariate AR models can be based on a difference of influence terms ($X \rightarrow Y - Y \rightarrow X$). In addition, they posited that inference on such influence estimates should always include experimental modulation of influence, in order to rule out hemodynamic variation as an underlying reason for spurious causality. In Deshpande et al. (2010) the authors simulated fMRI data by manipulating the causal influence and neuronal delays between local field potentials (LFPs) acquired from the macaque cortex and varying the hemodynamic delays of a convolving hemodynamic response function and the signal-to-noise ratio (SNR) and the sampling period of the final simulated fMRI data. They found that in multivariate (4 dimensional) simulations with hemodynamic and neuronal delays drawn from a uniform random distribution correct network detection from fMRI was well above chance and was up to 90% under conditions of fast sampling and low measurement noise. Other studies confirmed the observation that techniques with intermediate temporal resolution, such as fMRI, can yield good estimates of the causal connections based on AR models (Stevenson and Kording, 2010), even in the face of variable hemodynamics (Ryali et al., 2010). However, another recent simulation study, investigating a host of connectivity methods concluded low detection performance of directed influence by AR models under general conditions (Smith et al., 2010).

4.3. Toward integrated models

David et al. (2008) aimed at direct comparison of autoregressive modeling and DCM for fMRI time series and explicitly pointed at deconvolution of variable hemodynamics for causality inferences. The authors created a controlled animal experiment where gold standard validation of neuronal connectivity estimation was provided by intracranial EEG (iEEG) measurements. As discussed extensively in Friston (2009b) and Roebroeck et al. (2009a) such a validation experiment can provide important information on best practices in fMRI based brain connectivity modeling that, however, need to be carefully discussed and weighed. In David et al.'s study, simultaneous fMRI, EEG, and iEEG were measured in 6 rats during epileptic episodes in which spike-and-wave discharges (SWDs) spread through the brain. fMRI was used to map the hemodynamic response throughout the brain to seizure activity, where ictal and interictal states were quantified by the simultaneously recorded EEG. Three structures were selected by the authors as the crucial nodes in the network that generates and sustains seizure activity and further analysed with i) DCM, ii) simple AR modeling of the fMRI signal and iii) AR modeling applied to neuronal state-variable estimates obtained with a hemodynamic deconvolution step. By applying G-causality analysis to deconvolved fMRI time-series, the stochastic dynamics of the linear state-space model are augmented with the complex biophysically motivated observation model in DCM. This step is crucial if the goal is to compare the dynamic connectivity models and draw conclusions on the relative merits of linear stochastic models (explicitly estimating WAGS influence) and bilinear deterministic models. The results showed both AR analysis after deconvolution and DCM analysis to be in accordance with the gold-standard iEEG analyses, identifying the most pertinent influence relations undisturbed by variations in HRF latencies. In contrast, the final result of simple AR modeling of the fMRI signal showed less correspondence with

the gold standard, due to the confounding effects of different hemodynamic latencies which are not accounted for in the model.

Two important lessons can be drawn from David et al.’s study and the ensuing discussions (Bressler and Seth, 2010; Daunizeau et al., 2009a; David, 2009; Friston, 2009b,a; Roebroek et al., 2009a,b). First, it confirms again the distorting effects of hemodynamic processes on the temporal structure of fMRI signals and, more importantly, that the difference in hemodynamics in different parts of the brain can form a confound for dynamic brain connectivity models (Roebroek et al., 2005). Second, state-space models that embody observation models that connect latent neuronal dynamics to observed fMRI signals have a potential to identify causal influence unbiased by this confound. As a consequence, substantial recent methodological work has aimed at combining different models of latent neuronal dynamics with a form of a hemodynamic observation model in order to provide an inversion or filtering algorithm for estimation of parameters and hidden state trajectories. Following the original formulation of DCM that provides a bilinear ODE form for the hidden neuronal dynamics, attempts have been made at explicit integration of hemodynamics convolution with stochastic dynamic models that are interpretable in the framework of WAGS influence.

For instance in (Ryali et al., 2010), following earlier work (Penny et al., 2005; Smith et al., 2009), a discrete state space model with a bi-linear vector autoregressive model to quantify dynamic neuronal state evolution and both intrinsic and modulatory interactions is proposed:

$$\begin{aligned} x_k &= \mathbf{A}x_{k-1} + \sum_{j=1} v_k^j \mathbf{B}^j x_{k-1} + \mathbf{C}v_k^j + \varepsilon_k \\ \mathbf{x}_k^m &= [x_k^m, x_{k-1}^m, \dots, x_{k-L+1}^m] \\ y_k^m &= \beta^m \Phi \mathbf{x}_k^m + e_k^m \end{aligned} \quad (20)$$

Here, we index exogenous inputs with j and ROIs with m in superscripts. The entries in the autoregressive matrix \mathbf{A} , exogenous influence matrix \mathbf{C} and bi-linear matrices \mathbf{B}^j have the same interpretation as in deterministic DCM. The relation between observed BOLD-fMRI data y and latent neuronal sources x is modeled by a temporal embedding of into \mathbf{x}^m for each region or ROI m . This allows convolution with a flexible basis function expansion of possible HRF shapes to be represented by a simple matrix multiplication $\beta^m \Phi \mathbf{x}_k^m$ in the observation equation. Here Φ contains the temporal basis functions in Figure 2B and β^m the basis function parameters to be estimated. By estimating basis function parameters individually per region, variations in the HRF shape between region can be accounted for and the confounding effects of these on WAGS influence estimate can be avoided. Ryali et al. found that robust estimates of parameters $\Theta = \{\mathbf{A}, \mathbf{B}^j, \mathbf{C}, \beta^m, \Sigma_\varepsilon, \Sigma_e\}$ and states x_k can be obtained from a variational Bayesian approach. In their simulations, they show that a state-space model with interactions modeled at the latent level can compensate well for the effects of HRF variability, even when relative HRF delays are opposed to delayed interactions. Note, however, that subsampling of the BOLD signal is not explicitly characterized in their state-space model.

A few interesting variations on this discrete state-space modeling have recently been proposed. For instance in (Smith et al., 2009) a switching linear systems model for latent neuronal state evolution, rather than a bi-linear model was used. This model represents experimental modulation of connections as a random variable, to be learned from the data. This variable switches between different linear system instantiations that each characterize

connectivity in a single experimental condition. Such a scheme has the important advantage that an n -fold cross validation approach can be used to obtain a measure of absolute model-evidence (rather than relative between a selected set of models). Specifically, one could learn parameters for each context-specific linear system with knowledge of the timing of changing experimental conditions in a training data set. Then the classification accuracy of experimental condition periods in a test data set based on connectivity will provide a absolute model-fit measure, controlled for model complexity, which can be used to validate overall usefulness of the fitted model. In particular, this can point to important brain regions missing from the model in case of poor classification accuracy.

Another related line of developments instead has involved generalizing the ODE models in DCM for fMRI to stochastic dynamic models formulated in continuous time (Daunizeau et al., 2009b; Friston et al., 2008). An early exponent of this approach used local linearization in a (generalized) Kalman filter to estimate states and parameters in a non-linear SDE models of hemodynamics (Riera et al., 2004). Interestingly, the inclusion of stochastics in the state equations makes inference on coupling parameters of such models usefully interpretable in the framework of WAGS influence. This hints at the ongoing convergence, in modeling of brain connectivity, of time series approaches to causality in a discrete time tradition and dynamic systems and control theory approaches in a continuous time tradition.

5. Discussion and Outlook

The modeling of an enormously complex biological system such as the brain has many challenges. The abstractions and choices to be made in useful models of brain connectivity are therefore unlikely to be accommodated by one single ‘master’ model that does better than all other models on all counts. Nonetheless, the ongoing development efforts towards improved approaches are continually extending and generalizing the contexts in which dynamic time series models can be applied. It is clear that state space modeling and inference on WAGS influence are fundamental concepts within this endeavor. We end here with some considerations of dynamic brain connectivity models that summarize some important points and anticipate future developments.

We have emphasized that WAGS influence models of brain connectivity have largely been aimed at data driven exploratory analysis, whereas biophysically motivated state space models are mostly used for hypothesis-led confirmatory analysis. This is especially relevant in the interaction between model selection and model identification. Exploratory techniques use information in the data to investigate the relative applicability of many models. As such, they have the potential to detect ‘missing’ regions in anatomical models. Confirmatory approaches test hypotheses about connectivity within a set of models assumed to be applicable.

As mentioned above, the WAGS influence approach to statistical analysis of causal influence that we focused on here is complemented by the interventional approach rooted in the theory of graphical models and causal calculus. Graphical causal models have been recently applied to brain connectivity analysis of fMRI data (Ramsey et al., 2009). Recent work combining the two approaches (White and Lu, 2010) possibly leads the way to a combined causal treatment of brain imaging data incorporating dynamic models and interventions. Such a combination could enable incorporation of direct manipulation of brain activity by

(for example) transcranial magnetic stimulation (Pascual-Leone et al., 2000; Paus, 1999; Walsh and Cowey, 2000) into the current state space modeling framework.

Causal models of brain connectivity are increasingly inspired by biophysical theories. For fMRI this is primarily applicable in modeling the complex chain of events separating neuronal population activity from the BOLD signal. Inversion of such a model (in state space form) by a suitable filtering algorithm amounts to a model-based deconvolution of the fMRI signal resulting in an estimate of latent neuronal population activity. If the biophysical model is appropriately formulated to be identifiable (possibly including priors on relevant parameters), it can take variation in the hemodynamics between brain regions into account that can otherwise confound time series causality analyses of fMRI signals. Although models of hemodynamics for causal fMRI analysis have reached a reasonable level of complexity, the models of neuronal dynamics used to date have remained simple, comprising one or two state variables for an entire cortical region or subcortical structure. Realistic dynamic models of neuronal activity have a long history and have reached a high level of sophistication (Deco et al., 2008; Markram, 2006). It remains an open issue to what degree complex realistic equation systems can be embedded in analysis of fMRI – or in fact: any brain imaging modality – and result in identifiable models of neuronal connectivity and computation.

Two recent developments create opportunities to increase complexity and realism of neuronal dynamics models and move the level of modeling from the macroscopic (whole brain areas) towards the mesoscopic level comprising sub-populations of areas and cortical columns. First, the fusion of multiple imaging modalities, possibly simultaneously recorded, has received a great deal of attention. Particularly, several attempts to model-driven fusion of simultaneously recorded fMRI and EEG data, by inverting a separate observation model for each modality while using the same underlying neuronal model, have been reported (Deneux and Faugeras, 2010; Riera et al., 2007; Valdes-Sosa et al., 2009). This approach holds great potential to fruitfully combine the superior spatial resolution of fMRI with the superior temporal resolution of EEG. In (Valdes-Sosa et al., 2009) anatomical connectivity information obtained from diffusion tensor imaging and fiber tractography is also incorporated. Second, advances in MRI technology, particularly increases of main field strength to 7T (and beyond) and advances in parallel imaging (de Zwart et al., 2006; Heidemann et al., 2006; Pruessmann, 2004; Wiesinger et al., 2006), greatly increase the level spatial detail that are accessible with fMRI. For instance, fMRI at 7T with sufficient spatial resolution to resolve orientation columns in human visual cortex has been reported (Yacoub et al., 2008).

The development of state space models for causal analysis of fMRI data has moved from discrete to continuous and from deterministic to stochastic models. Continuous models with stochastic dynamics have desirable properties, chief among them a robust inference on causal influence interpretable in the WAGS framework, as discussed above. However, dealing with continuous stochastic models leads to technical issues such as the properties and interpretation of Wiener processes and Ito calculus (Friston, 2008). A number of inversion or filtering methods for continuous stochastic models have been recently proposed, particularly for the goal of causal analysis of brain imaging data, including the local linearization and innovations approach (Hernandez et al., 1996; Riera et al., 2004), dynamic expectation maximization (Friston et al., 2008) and generalized filtering (Friston et al., 2010). The

ongoing development of these filtering methods, their validation and their scalability towards large numbers of state variables will be a topic of continuing research.

Acknowledgments

The authors thank Kamil Uludag for comments and discussion.

References

- Odd O. Aalen. Dynamic modeling and causality. *Scandinavian Actuarial journal*, pages 177–190, 1987.
- O.O. Aalen and A. Frigessi. What can statistics contribute to a causal understanding? *Board of the Foundation of the Scandinavian journal of Statistics*, 34:155–168, 2007.
- G. K. Aguirre, E. Zarahn, and M. D’Esposito. The variability of human, bold hemodynamic responses. *Neuroimage*, 8(4):360–9, 1998.
- H Akaike. On the use of a linear model for the identification of feedback systems. *Annals of the Institute of statistical mathematics*, 20(1):425–439, 1968.
- A Amendola, M Niglio, and C Vitale. Temporal aggregation and closure of varma models: Some new results. In F. Palumbo et al., editor, *Data Analysis and Classification: Studies in Classification, Data Analysis, and Knowledge Organization*, pages 435–443. 2010.
- D. Attwell, A. M. Buchan, S. Charpak, M. Lauritzen, B. A. Macvicar, and E. A. Newman. Glial and neuronal control of brain blood flow. *Nature*, 468(7321):232–43, 2010.
- D. Attwell and C. Iadecola. The neural basis of functional brain imaging signals. *Trends Neurosci*, 25(12):621–5, 2002.
- A. B. Barrett, L. Barnett, and A. K. Seth. Multivariate granger causality and generalized variance. *Phys Rev E Stat Nonlin Soft Matter Phys*, 81(4 Pt 1):041907, 2010.
- M.J. Beal. *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, University College London, 2003.
- A R Bergstrom. Nonrecursive models as discrete approximations to systems of stochastic differential equations. *Econometrica*, 34:173–182, 1966.
- A R Bergstrom. Continuous time stochastic models and issues of aggregation. In Z. Griliches and M.D. Intriligator, editors, *Handbook of econometrics*, volume II. Elsevier, 1984.
- M.A. Bernstein, K.F. King, and X.J. Zhou. *Handbook of MRI Pulse Sequences*. Elsevier Academic Press, urlington, 2004.
- S. L. Bressler and A. K. Seth. Wiener-granger causality: A well established methodology. *Neuroimage*, 2010.

- S. L. Bressler, W. Tang, C. M. Sylvester, G. L. Shulman, and M. Corbetta. Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *J Neurosci*, 28(40):10056–61, 2008.
- R. B. Buxton. The elusive initial dip. *Neuroimage*, 13(6 Pt 1):953–8, 2001.
- R. B. Buxton, K. Uludag, D. J. Dubowitz, and T. T. Liu. Modeling the hemodynamic response to brain activation. *Neuroimage*, 23 Suppl 1:S220–33, 2004.
- R. B. Buxton, E. C. Wong, and L. R. Frank. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med*, 39(6):855–64, 1998.
- Marcus J Chambers and Michael A Thornton. Discrete time representation of continuous time arma processes, 2009.
- Daniel Commenges and Anne Gegout-Petit. A general dynamical statistical model with possible causal interpretation. *journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):1–43, 2009.
- J. Daunizeau, O. David, and K. E. Stephan. Dynamic causal modelling: A critical review of the biophysical and statistical foundations. *Neuroimage*, 2009a.
- J. Daunizeau, K. J. Friston, and S. J. Kiebel. Variational bayesian identification and prediction of stochastic nonlinear dynamic causal models. *Physica D*, 238(21):2089–2118, 2009b.
- O. David. fmri connectivity, meaning and empiricism comments on: Roebroek et al. the identification of interacting networks in the brain using fmri: Model selection, causality and deconvolution. *Neuroimage*, 2009.
- O. David, I. Guillemain, S. Saittet, S. Reyt, C. Deransart, C. Segebarth, and A. Depaulis. Identifying neural drivers with functional mri: an electrophysiological validation. *PLoS Biol*, 6(12):2683–97, 2008.
- J. A. de Zwart, P. van Gelderen, X. Golay, V. N. Ikonomidou, and J. H. Duyn. Accelerated parallel imaging for functional imaging of the human brain. *NMR Biomed*, 19(3):342–51, 2006.
- G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. Friston. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Comput Biol*, 4(8):e1000092, 2008.
- T. Deneux and O. Faugeras. Eeg-fmri fusion of paradigm-free activity using kalman filtering. *Neural Comput*, 22(4):906–48, 2010.
- G. Deshpande, X. Hu, R. Stilla, and K. Sathian. Effective connectivity during haptic perception: a study using granger causality analysis of functional magnetic resonance imaging data. *Neuroimage*, 40(4):1807–14, 2008.

- G. Deshpande, S. LaConte, G. A. James, S. Peltier, and X. Hu. Multivariate granger causality analysis of fmri data. *Hum Brain Mapp*, 30(4):1361–73, 2009.
- G. Deshpande, K. Sathian, and X. Hu. Effect of hemodynamic variability on granger causality analysis of fmri. *Neuroimage*, 52(3):884–96, 2010.
- J Florens. Some technical issues in defining causality. *journal of Econometrics*, 112:127–128, 2003.
- J.P. Florens and D. Fougere. Noncausality in continuous time. *Econometrica*, 64(5):1195–1212, 1996.
- K. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Hum Brain Mapp*, 2:56–78, 1994.
- K. Friston. Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annu Rev Neurosci*, 25:221–50, 2002.
- K. Friston. Dynamic causal modeling and granger causality comments on: The identification of interacting networks in the brain using fmri: Model selection, causality and deconvolution. *Neuroimage*, 2009a.
- K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny. Variational free energy and the laplace approximation. *Neuroimage*, 34(1):220–34, 2007.
- K. J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–302, 2003.
- K. J. Friston, A. Mechelli, R. Turner, and C. J. Price. Nonlinear responses in fmri: the balloon model, volterra kernels, and other hemodynamics. *Neuroimage*, 12(4):466–77, 2000.
- K. J. Friston, N. Trujillo-Barreto, and J. Daunizeau. Dem: a variational treatment of dynamic systems. *Neuroimage*, 41(3):849–85, 2008.
- Karl Friston. Hierarchical models in the brain. *PLoS Computational Biology*, 4, 2008.
- Karl Friston. Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS biology*, 7:e33, 2009b.
- Karl Friston, Klaas Stephan, Baojuan Li, and Jean Daunizeau. Generalised filtering. *Mathematical Problems in Engineering*, 2010:1–35, 2010.
- J. F. Geweke. Measurement of linear dependence and feedback between multiple time series. *journal of the American Statistical Association*, 77(378):304–324, 1982.
- G. H. Glover. Deconvolution of impulse response in event-related bold fmri. *Neuroimage*, 9(4):416–29, 1999.
- C. Glymour. Learning, prediction and causal bayes nets. *Trends Cogn Sci*, 7(1):43–48, 2003.

- R. Goebel, A. Roebroeck, D. S. Kim, and E. Formisano. Investigating directed cortical interactions in time-resolved fmri data using vector autoregressive modeling and granger causality mapping. *Magn Reson Imaging*, 21(10):1251–61, 2003.
- C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.
- E.M. Haacke, R.W. Brown, M.R. Thompson, and R. Venkatesan. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. John Wiley and Sons, Inc, New York, 1999.
- M. Havlicek, J. Jan, M. Brazdil, and V. D. Calhoun. Dynamic granger causality based on kalman filter for evaluation of functional network connectivity in fmri data. *Neuroimage*, 53(1):65–77, 2010.
- R. M. Heidemann, N. Seiberlich, M. A. Griswold, K. Wohlfarth, G. Krueger, and P. M. Jakob. Perspectives and limitations of parallel mr imaging at high field strengths. *Neuroimaging Clin N Am*, 16(2):311–20, 2006.
- R. N. Henson, C. J. Price, M. D. Rugg, R. Turner, and K. J. Friston. Detecting latency differences in event-related bold responses: application to words versus nonwords and initial versus repeated face presentations. *Neuroimage*, 15(1):83–97, 2002.
- JL Hernandez, PA Valds, and P Vila. Eeg spike and wave modelled by a stochastic limit cycle. *NeuroReport*, 1996.
- B. Horwitz, K. J. Friston, and J. G. Taylor. Neural modeling and functional brain imaging: an overview. *Neural Netw*, 13(8-9):829–46, 2000.
- H. Johansen-Berg and T.E.J Behrens, editors. *Diffusion MRI: From quantitative measurement to in-vivo neuroanatomy*. Academic Press, London, 2009.
- D.K. Jones, editor. *Diffusion MRI: Theory, Methods, and Applications*. Oxford University Press, Oxford, 2010.
- S. J. Kiebel, M. I. Garrido, R. Moran, C. C. Chen, and K. J. Friston. Dynamic causal modeling for eeg and meg. *Hum Brain Mapp*, 30(6):1866–76, 2009.
- C. H. Liao, K. J. Worsley, J. B. Poline, J. A. Aston, G. H. Duncan, and A. C. Evans. Estimating the delay of the fmri response. *Neuroimage*, 16(3 Pt 1):593–606, 2002.
- L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, New Jersey, 2nd edition, 1999.
- N. K. Logothetis. What we can do and what we cannot do with fmri. *Nature*, 453(7197):869–78, 2008.
- N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fmri signal. *Nature*, 412(6843):150–7, 2001.
- H. Markram. The blue brain project. *Nat Rev Neurosci*, 7(2):153–60, 2006.

- A. C. Marreiros, S. J. Kiebel, and K. J. Friston. Dynamic causal modelling for fmri: a two-state model. *Neuroimage*, 39(1):269–78, 2008.
- J. Roderick McCrorie. The likelihood of the parameters of a continuous time vector autoregressive model. *Statistical Inference for Stochastic Processes*, 5:273–286, 2002.
- J. Roderick Mccrorie. The problem of aliasing in identifying finite parameter continuous time stochastic models. *Acta Applicandae Mathematicae*, 79:9–16, 2003.
- A. R. McIntosh. Contexts and catalysts: a resolution of the localization and integration of function in the brain. *Neuroinformatics*, 2(2):175–82, 2004.
- T Ozaki. A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: A local linearization approach. *Statistica Sinica*, 2:113–135, 1992.
- A. Pascual-Leone, V. Walsh, and J. Rothwell. Transcranial magnetic stimulation in cognitive neuroscience—virtual lesion, chronometry, and functional connectivity. *Curr Opin Neurobiol*, 10(2):232–7, 2000.
- T. Paus. Imaging the brain before, during, and after transcranial magnetic stimulation. *Neuropsychologia*, 37(2):219–24, 1999.
- J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, 2nd edition, 2009.
- W. Penny, Z. Ghahramani, and K. Friston. Bilinear dynamical systems. *Philos Trans R Soc Lond B Biol Sci*, 360(1457):983–93, 2005.
- W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston. Comparing dynamic causal models. *Neuroimage*, 22(3):1157–72, 2004.
- Peter C.B. Phillips. The problem of identification in finite parameter continuous time models. *Journal of Econometrics*, 1:351–362, 1973.
- Peter C.B. Phillips. The estimation of some continuous time models. *Econometrica*, 42:803–823, 1974.
- K. P. Pruessmann. Parallel imaging at high field strength: synergies and joint potential. *Top Magn Reson Imaging*, 15(4):237–44, 2004.
- J. D. Ramsey, S. J. Hanson, C. Hanson, Y. O. Halchenko, R. A. Poldrack, and C. Glymour. Six problems for causal inference from fmri. *Neuroimage*, 49(2):1545–58, 2009.
- A. Rauch, G. Rainer, and N. K. Logothetis. The effect of a serotonin-induced dissociation between spiking and perisynaptic activity on bold functional mri. *Proc Natl Acad Sci U S A*, 105(18):6759–64, 2008.
- G.C. Reinsel. *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York, 2nd edition, 1997.

- J. J. Riera, J. C. Jimenez, X. Wan, R. Kawashima, and T. Ozaki. Nonlinear local electrovascular coupling. ii: From data to neuronal masses. *Hum Brain Mapp*, 28(4):335–54, 2007.
- J. J. Riera, J. Watanabe, I. Kazuki, M. Naoki, E. Aubert, T. Ozaki, and R. Kawashima. A state-space model of the hemodynamic approach: nonlinear filtering of bold signals. *Neuroimage*, 21(2):547–67, 2004.
- A. Roebroeck, E. Formisano, and R. Goebel. Mapping directed influence over the brain using granger causality and fmri. *Neuroimage*, 25(1):230–42, 2005.
- A. Roebroeck, E. Formisano, and R. Goebel. The identification of interacting networks in the brain using fmri: Model selection, causality and deconvolution. *Neuroimage*, 2009a.
- A. Roebroeck, E. Formisano, and R. Goebel. Reply to friston and david after comments on: The identification of interacting networks in the brain using fmri: Model selection, causality and deconvolution. *Neuroimage*, 2009b.
- S. Ryali, K. Supekar, T. Chen, and V. Menon. Multivariate dynamical systems models for estimating causal interactions in fmri. *Neuroimage*, 2010.
- Z. S. Saad, K. M. Ropella, R. W. Cox, and E. A. DeYoe. Analysis and use of fmri response delays. *Hum Brain Mapp*, 13(2):74–93, 2001.
- R. Salmelin and J. Kujala. Neural representation of language: activation versus long-range connectivity. *Trends Cogn Sci*, 10(11):519–25, 2006.
- J. R. Sato, A. Fujita, E. F. Cardoso, C. E. Thomaz, M. J. Brammer, and Jr. E. Amaro. Analyzing the connectivity between regions of interest: an approach based on cluster granger causality for fmri data analysis. *Neuroimage*, 52(4):1444–55, 2010.
- J. R. Sato, D. Y. Takahashi, S. M. Arcuri, K. Sameshima, P. A. Morettin, and L. A. Baccala. Frequency domain connectivity identification: an application of partial directed coherence in fmri. *Hum Brain Mapp*, 30(2):452–61, 2009.
- M. B. Schippers, A. Roebroeck, R. Renken, L. Nanetti, and C. Keysers. Mapping the information flow from one brain to another during gestural communication. *Proc Natl Acad Sci U S A*, 107(20):9388–93, 2010.
- T Schweder. Composable markov processes. *journal of Applied Probability*, 7(2):400–410, 1970.
- J. F. Smith, A. Pillai, K. Chen, and B. Horwitz. Identification and validation of effective connectivity networks in functional magnetic resonance imaging using switching linear dynamic systems. *Neuroimage*, 52(3):1027–40, 2009.
- S. M. Smith, K. L. Miller, G. Salimi-Khorshidi, M. Webster, C. F. Beckmann, T. E. Nichols, J. D. Ramsey, and M. W. Woolrich. Network modelling methods for fmri. *Neuroimage*, 2010.

- V Solo. On causality i: Sampling and noise. *Proceedings of the 46th IEEE Conference on Decision and Control*, pages 3634–3639, 2006.
- D. Sridharan, D. J. Levitin, and V. Menon. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc Natl Acad Sci U S A*, 105(34):12569–74, 2008.
- K. E. Stephan, L. Kasper, L. M. Harrison, J. Daunizeau, H. E. den Ouden, M. Breakspear, and K. J. Friston. Nonlinear dynamic causal models for fmri. *Neuroimage*, 42(2):649–62, 2008.
- I. H. Stevenson and K. P. Kording. On the similarity of functional connectivity between neurons estimated across timescales. *PLoS One*, 5(2):e9206, 2010.
- K. Thomsen, N. Offenhauser, and M. Lauritzen. Principal neuron spiking: neither necessary nor sufficient for cerebral blood flow in rat cerebellum. *J Physiol*, 560(Pt 1):181–9, 2004.
- L. Q. Uddin, A. M. Kelly, B. B. Biswal, F. Xavier Castellanos, and M. P. Milham. Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Hum Brain Mapp*, 30(2):625–37, 2009.
- K. Ugurbil, L. Toth, and D. S. Kim. How accurate is magnetic resonance imaging of brain function? *Trends Neurosci*, 26(2):108–14, 2003.
- K. Uludag. To dip or not to dip: reconciling optical imaging and fmri data. *Proc Natl Acad Sci U S A*, 107(6):E23; author reply E24, 2010.
- K. Uludag, D. J. Dubowitz, and R. B. Buxton. Basic principles of functional mri. In R. Edelman, J. Hesselink, and M. Zlatkin, editors, *Clinical MRI*. Elsevier, San Diego, 2005.
- K. Uludag, B. Muller-Bierl, and K. Ugurbil. An integrative model for neuronal activity-induced signal changes for gradient and spin echo functional imaging. *Neuroimage*, 48(1):150–65, 2009.
- P Valdes-Sosa, J C Jimenez, J Riera, R Biscay, and T Ozaki. Nonlinear eeg analysis based on a neural mass model. *Biological cybernetics*, 81:415–24, 1999.
- P. Valdes-Sosa, A. Roebroeck, J. Daunizeau, and K. Friston. Effective connectivity: Influence, causality and biophysical modeling. *Neuroimage*, in press.
- P. A. Valdes-Sosa. Spatio-temporal autoregressive models defined over brain manifolds. *Neuroinformatics*, 2(2):239–50, 2004.
- P. A. Valdes-Sosa, R. Kotter, and K. J. Friston. Introduction: multimodal neuroimaging of brain connectivity. *Philos Trans R Soc Lond B Biol Sci*, 360(1457):865–7, 2005a.
- P. A. Valdes-Sosa, J. M. Sanchez-Bornot, A. Lage-Castellanos, M. Vega-Hernandez, J. Bosch-Bayard, L. Melie-Garcia, and E. Canales-Rodriguez. Estimating brain functional connectivity with sparse multivariate autoregression. *Philos Trans R Soc Lond B Biol Sci*, 360(1457):969–81, 2005b.

- P. A. Valdes-Sosa, J. M. Sanchez-Bornot, R. C. Sotero, Y. Iturria-Medina, Y. Aleman-Gomez, J. Bosch-Bayard, F. Carbonell, and T. Ozaki. Model driven eeg/fmri fusion of brain oscillations. *Hum Brain Mapp*, 30(9):2701–21, 2009.
- V. Walsh and A. Cowey. Transcranial magnetic stimulation and cognitive neuroscience. *Nat Rev Neurosci*, 1(1):73–9, 2000.
- W. W. S. Wei. *Time Series Analysis: Univariate and Multivariate Methods*. Addison-Wesley, Redwood City, 1990.
- Halbert White and Xun Lu. Granger causality and dynamic structural systems. *Journal of Financial Econometrics*, 8(2):193–243, 2010.
- N. Wiener. The theory of prediction. In E.F. Berkenbach, editor, *Modern Mathematics for Engineers*. McGraw-Hill, New York, 1956.
- F. Wiesinger, P. F. Van de Moortele, G. Adriany, N. De Zanche, K. Ugurbil, and K. P. Pruessmann. Potential and feasibility of parallel mri at high field. *NMR Biomed*, 19(3):368–78, 2006.
- E. Yacoub, N. Harel, and K. Ugurbil. High-field fmri unveils orientation columns in humans. *Proc Natl Acad Sci U S A*, 105(30):10607–12, 2008.