NeuroImage **12,** 466–477 (2000)
doi:10.1006/nimg.2000.0630, available online at http://www.idealibrary.com on IDEAL®

# Nonlinear Responses in fMRI: The Balloon Model, Volterra Kernels, and Other Hemodynamics

K. J. Friston, A. Mechelli, R. Turner, and C. J. Price

*The Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1N 3BG, United Kingdom*

Received March 30, 2000

**There is a growing appreciation of the importance of nonlinearities in evoked responses in fMRI, particularly with the advent of event-related fMRI. These nonlinearities are commonly expressed as interactions among stimuli that can lead to the suppression and increased latency of responses to a stimulus that are incurred by a preceding stimulus. We have presented previously a model-free characterization of these effects using generic techniques from nonlinear system identification, namely a Volterra series formulation. At the same time Buxton *et al.* (1998) described a plausible and compelling dynamical model of hemodynamic signal transduction in fMRI. Subsequent work by Mandeville *et al.* (1999) provided important theoretical and empirical constraints on the form of the dynamic relationship between blood flow and volume that underpins the evolution of the fMRI signal. In this paper we combine these system identification and model-based approaches and ask whether the Balloon model is sufficient to account for the nonlinear behaviors observed in real time series. We conclude that it can, and furthermore the model parameters that ensue are biologically plausible. This conclusion is based on the observation that the Balloon model can produce Volterra kernels that emulate empirical kernels. To enable this evaluation we had to embed the Balloon model in a hemodynamic input-state-output model that included the dynamics of perfusion changes that are contingent on underlying synaptic activation. This paper presents (i) the full hemodynamic model (ii), how its associated Volterra kernels can be derived, and (iii) addresses the model's validity in relation to empirical nonlinear characterisations of evoked responses in fMRI and other neurophysiological constraints.** 

***Key Words:*** **nonlinear system identification; functional neuroimaging; fMRI; hemodynamic response function; Volterra series; Balloon Model; Windkessel.**

## INTRODUCTION

This paper is about the nonlinear aspects of evoked responses in functional neuroimaging and represents a dynamical approach to modeling and characterizing event-related signals in fMRI. It aims to: (i) show that the Balloon/Windkessel model (Buxton and Frank, 1997; Buxton *et al.*, 1998; Mandeville *et al.*, 1999) is sufficient to account for nonlinearities in event-related responses that are seen empirically and (ii) describe a nonlinear dynamical model that couples changes in synaptic activity to fMRI signals. This hemodynamic model obtains by combining the Balloon/Windkessel model (henceforth Balloon model) with a model of how synaptic activity causes changes in regional flow. Subsequent communications will use this model to explore the dependence of fMRI signals on various parameters pertaining to experimental design and the dynamics of the underlying neuronal response.

In Friston *et al.* (1994) we presented a linear model of hemodynamic responses in fMRI time-series, wherein underlying neuronal activity (inferred on the basis of changing stimulus or task conditions) is convolved, or smoothed with a *hemodynamic response function*. In Friston *et al.* (1998) we extended this model to cover nonlinear responses using a Volterra series expansion. At the same time Buxton and colleagues developed a mechanistically compelling model of how evoked changes in blood flow were transformed into a blood oxygenation level dependent (BOLD) signal (Buxton *et al.*, 1998). A component of the Balloon model, namely the relationship between blood flow and volume, was then elaborated in the context of standard windkessel theory by Mandeville *et al.* (1999). The Volterra approach, in contradistinction to other nonlinear characterizations of hemodynamic responses (c.f. Vazquez and Noll, 1996), is model-independent, in the sense that Volterra series can model the behavior of any nonlinear time-invariant dynamical system.[1] The

[1] In principle Volterra series can represent any dynamical input-state-output system and in this sense a characterisation in terms of Volterra kernels is model independent. However, by using basis functions to constrain the solution space, constraints are imposed on the form of the kernels and, implicitly, the underlying dynamical system (i.e., state-space representation). The characterization is therefore only assumption free to the extent the basis set is sufficiently comprehensive.



1053-8119/00 $35.00

principal aim of this work was to see if the theoretically motivated Balloon model would be sufficient to explain the nonlinearities embodied in a purely empirical Volterra characterisation.

### *Volterra Series*

Volterra series express the output of a system, in this case the BOLD signal from a particular voxel, as a function of some input, here the assumed synaptic activity that is changed experimentally. This series is a function of the input over its recent history and is expressed in terms of generalised convolution kernels. Volterra series are often referred to as nonlinear convolutions or polynomial expansions with memory. They are simply Taylor expansions extended to cover dynamical input-state-output systems by considering the effect of the input now and at all times in the recent past. The zero-order kernel is simply a constant about which the response varies. The first order kernel represents the weighting applied to a sum of inputs over the recent past (c.f. the hemodynamic response function) and can be thought of as the change in output for a change in the input at each time point. Similarly, the second order coefficients represent interactions that are simply the effect of the input at one point in time on its contribution at another. The second order kernel comprises coefficients that are applied to interactions among (i.e., products of) inputs, at different times in the past, to predict the response.

In short the output can be considered a nonlinear convolution of the input where nonlinear behaviours are captured by high order kernels. For example, the presence of a stimulus can be shown to attenuate the magnitude of, and induce a longer latency in, the response to a second stimulus that occurs within a second or so. The example shown in Fig. 1 comes from our previous analysis (Friston *et al.*, 1998) and shows how a preceding stimulus can modify the response to a subsequent stimulus. This sort of effect led to the notion of *hemodynamic refractoriness* and is an important example of nonlinearity in fMRI time-series.

The important thing about Volterra series is that they do not refer to all the hidden state variables that mediate between the input and output (e.g., blood flow, venous volume, oxygenation, the dynamics of endothelium derived relaxing factor, kinetics of cerebral metabolism etc.). This renders them very powerful because they provide for a complete specification of the dynamical behaviour of a system without ever having to measure the state variables or making any assumptions about how these variables interact to produce a response. On the other hand the Volterra formulation is impoverished because it yields no mechanistic insight into how the response is mediated. The alternative is to posit some model of interacting state variables and establish the validity of that model in relation to observed input-output behaviors and the dynamics of the state variables themselves. This involves specifying a series of differential equations that express the change in one state variable as a function of the others and the input. Once these equations are specified the equivalent Volterra representation can be derived analytically (see the Appendix for details). The Balloon model is a comprehensive example of such a model.

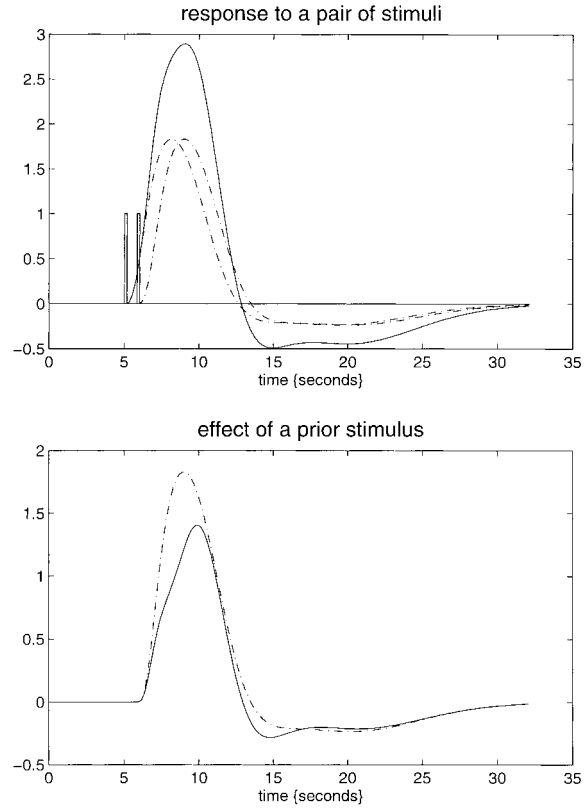


**FIG. 1.** Top panel: Simulated responses to a pair of words (bars) one second apart, presented together (solid line) and separately (broken lines) based on the kernels shown in Fig. 4. Lower panel: The response to the second word when presented alone (broken line as above) and when preceded by the first (solid line). The latter obtains by subtracting the response to the first word from the response to both. The difference reflects the effect of the first word on the response to the second.

### *The Balloon Model*

The Balloon model (Buxton and Frank, 1997; Buxton *et al.*, 1998) is an input-state-output model with two-state variables volume ($v$) and deoxyhemoglobin content ($q$). The input to the system is blood flow ($f_{in}$) and the output is the BOLD signal ($y$). The BOLD signal is partitioned into an extra and intravascular component, weighted by their respective volumes. These signal components depend on the deoxyhemoglobin content and render the signal a nonlinear function of $v$ and $q$. The effect of flow on $v$ and $q$ (see below) determines the

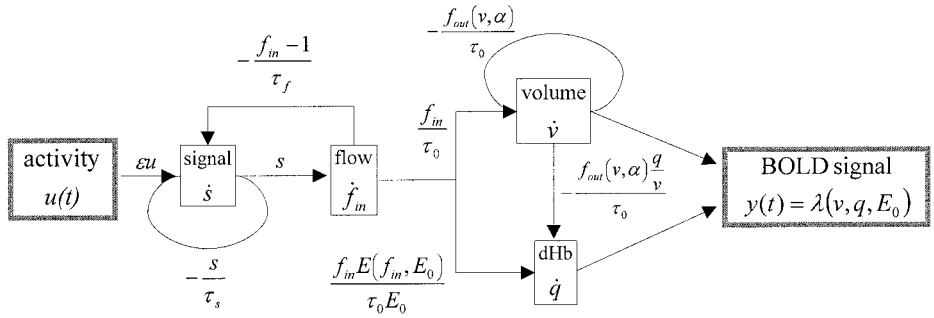


**FIG. 2.** Schematic illustrating the organization of the hemodynamic model. This is a fully nonlinear single input $u(t)$, single output $y(t)$ state model with four state variables $s$, $f_{in}$, $v$, and $q$. The form and motivation for the changes in each state variable, as functions of the others, are described in the main text.

output and it is these effects that are the essence of the Balloon model: Increases in flow effectively inflate a venous "balloon" such that deoxygenated blood is diluted and expelled at a greater rate. The clearance of deoxyhemoglobin reduces intravoxel dephasing and engenders an increase in signal. Before the balloon has inflated sufficiently the expulsion and dilution may be insufficient to counteract the increased delivery of deoxygenated blood to the venous compartment and an "early dip" in signal may be expressed. After the flow has peaked, and the balloon has relaxed again, reduced clearance and dilution contribute to the poststimulus undershoot commonly observed. This is a simple and plausible model that is predicated on a minimal set of assumptions and relates closely to the windkessel formulation of Mandeville *et al.* (1999). Furthermore the predictions of the Balloon model concur with the steady-state models of Hoge and colleagues, and their elegant studies of the relationship between blood flow and oxygen consumption in human visual cortex (e.g., Hoge *et al.*, 1999).

The Balloon model is inherently nonlinear and may account for the sorts of nonlinear interactions revealed by the Volterra formulation. One simple test of this hypothesis is to see if the Volterra kernels associated with the Balloon model compare with those derived empirically. The Volterra kernels estimated in Friston *et al.* (1998) clearly did not use flow as input because flow is not measurable with BOLD fMRI. The input comprised a stimulus function as an index of synaptic activity. In order to evaluate the Balloon model in terms of these Volterra kernels it has to be extended to accommodate the dynamics of how flow is coupled to synaptic activity encoded in the stimulus function. This paper presents one such extension.

In summary the Balloon model deals with the link between flow and BOLD signal. By extending the model to cover the dynamic coupling of synaptic activity and flow a complete model, relating experimentally induced changes in neuronal activity to BOLD signal, obtains. The input–output behavior of this model can be compared to the real brain in terms of their respective Volterra kernels.

### *Overview*

The remainder of this paper is divided into three sections. In the next section we present a hemodynamic model of the coupling between synaptic activity and BOLD response that builds upon the Balloon model. The second section presents an empirical evaluation of this model by comparing its Volterra kernels with those obtained using real fMRI data. This is not a trivial exercise because: (i) there is no guarantee that the Balloon model could produce the complicated forms of the kernels seen empirically, and (ii) even if it could, the parameters needed to do so may be biologically implausible. This section provides estimates of these parameters, which allow some comment on the face validity of the model, in relation to known physiology. The final section presents a discussion of the results in relation to known biophysics and neurophysiology.

This paper is concerned with the validation and evaluation of the Balloon model, in relation to the Volterra characterisations, and the hemodynamic model presented below in relation to real hemodynamics. Subsequent papers will use the model to address some important issues related to experimental design and the sorts of neuronal dynamics that BOLD signals are most sensitive to.

## THE HEMODYNAMIC MODEL

In this section we describe a hemodynamic model that mediates between synaptic activity and measured BOLD responses. This model essentially combines the Balloon model and a simple linear dynamical model of changes in regional cerebral blood flow (rCBF) caused by neuronal activity. The model architecture is summarized in Fig. 2. To motivate the model components more clearly we will start at the output and work toward the input.

*The Balloon Component*

This component links rCBF and the BOLD signal as described in Buxton *et al.* (1998). All variables are expressed in normalized form, relative to resting values. The BOLD signal $y(t) = \lambda(v, q, E_0)$ is taken to be a static nonlinear function of normalized venous volume ($v$), normalized total deoxyhemoglobin voxel content ($q$) and resting net oxygen extraction fraction by the capillary bed ($E_0$)

$$\begin{aligned} y(t) &= \lambda(v, q, E_0) = V_0(k_1(1 - q) \\ &\quad + k_2(1 - q/v) + k_3(1 - v)) \\ k_1 &= 7E_0 \\ k_2 &= 2 \\ k_3 &= 2E_0 - 0.2, \end{aligned} \tag{1}$$

where $V_0$ is resting blood volume fraction. This signal comprises a volume-weighted sum of extra- and intravascular signals that are functions of volume and deoxyhemoglobin content. The latter are the state variables whose dynamics need specifying. The rate of change of volume is simply

$$\tau_0 \dot{v} = f_{in} - f_{out}(v). \tag{2}$$

See Mandeville *et al.* (1999) for an excellent discussion of this equation in relation to windkessel theory. Equation (2) says that volume changes reflect the difference between inflow $f_{in}$ and outflow $f_{out}$ from the venous compartment with a time constant $\tau_0$. This constant represents the mean transit time (i.e., the average time it takes to traverse the venous compartment or for that compartment to be replenished) and is $V_0/F_0$, where $F_0$ is resting flow. The physiology of the relationship between flow and volume is determined by the evolution of the transit time. Mandeville *et al.* (1999) reformulated the temporal evolution of transit time into a description of the dynamics of resistance and capacitance of the balloon using windkessel theory ("windkessel" means leather bag). This enabled them to posit a form for the temporal evolution of a downstream elastic response to arteriolar vasomotor changes and estimate mean transit times using measurements of volume and flow, in rats, using fMRI and laser-Doppler flowmetry. We will compare these estimates to our empirical estimates in the next section.

Note that outflow is a function of volume. This function models the balloon-like capacity of the venous compartment to expel blood at a greater rate when distended. We model it with a single parameter $\alpha$ based on the windkessel model

$$f_{out}(v) = v^{1/\alpha}, \tag{3}$$

where $1/\alpha = \gamma + \beta$ (c.f. Eq. (6) in Mandeville *et al.*, 1999). $\gamma = 2$ represents laminar flow. $\beta > 1$ models diminished volume reserve at high pressures and can be thought of as the ratio of the balloon's capacitance to its compliance. At steady state empirical results from PET suggest $\alpha \approx 0.38$ (Grubb *et al.*, 1974). However, when flow and volume are changing dynamically, this value is smaller. Mandeville *et al.* (1999) were the first to measure the dynamic flow–volume relationship and estimated $\alpha \approx 0.18$, after 6 s of stimulation, with a projected asymptotic (steady-state) value of 0.36.

The change in deoxyhemoglobin $\dot{q}$ reflects the delivery of deoxyhemoglobin into the venous compartment minus that expelled (outflow times concentration)

$$\tau_0 \dot{q} = f_{in} \frac{E(f_{in}, E_0)}{E_0} - f_{out}(v) q/v, \tag{4}$$

where $E(f_{in}, E_0)$ is the fraction of oxygen extracted from the inflowing blood. This is assumed to depend on oxygen delivery and is consequently flow-dependent. A reasonable approximation for a wide range of transport conditions is (Buxton *et al.*, 1998).

$$E(f_{in}, E_0) = 1 - (1 - E_0)^{1/f_{in}} \tag{5}$$

The second term in Eq. (4) represents an important nonlinearity: The effect of flow on signal is largely determined by the inflation of the balloon, resulting in an increase of $f_{out}(v)$ and clearance of deoxyhemoglobin. This effect depends upon the concentration of deoxyhemoglobin such that the clearance attained by the outflow will be severely attenuated when the concentration is low (e.g., during the peak response to a prior stimulus). The implications of this will be illustrated in the next section.

This concludes the Balloon model component, where there are only three unknown parameters that determine the dynamics $E_0$, $\tau_0$, and $\alpha$, namely resting oxygen extraction fraction ($E_0$), mean transit time ($\tau_0$), and a stiffness exponent ($\alpha$) specifying the flow–volume relationship of the venous balloon. The only thing required, to specify the BOLD response, is inflow.

*rCBF Component*

It is generally accepted that, over normal ranges, blood flow and synaptic activity are linearly related. A recent empirical verification of this assumption can be found in Miller *et al.* (2000), who used MRI perfusion imaging to address this issue in visual and motor cortices. After modeling neuronal adaptation they were able to conclude, "Both rCBF responses are consistent with a linear transformation of a simple nonlinear neural response model." Furthermore our own work using PET and fMRI replications of the same experi-

ments suggests that the observed nonlinearities enter into the translation of rCBF into a BOLD response (as opposed to a nonlinear relationship between synaptic activity and rCBF) in the auditory cortices (see Friston *et al.*, 1998). Under the constraint that the dynamical system linking synaptic activity and rCBF is linear we have chosen the most parsimonious model

$$\dot{f}_{\text{in}} = s, \tag{6}$$

where $s$ is some flow inducing signal defined, operationally, in units corresponding to the rate of change of normalised flow (i.e., $s^{-1}$). Although it may seem more natural to express the effect of this signal directly on vascular resistance ($r$), for example $\dot{r} = -s$, Eq. (6) has the more plausible form. This is because the effect of signal ($s$) is much smaller when $r$ is small (when the arterioles are fully dilated signals such as endothelium-derived relaxing factor or nitric oxide will cause relatively small decrements in resistance). This can be seen by noting Eq. (6) is equivalent to $\dot{r} = -r^2 s$, where $f_{\text{in}} = 1/r$.

The signal is assumed to subsume many neurogenic and diffusive signal subcomponents and is generated by neuronal activity $u(t)$

$$\dot{s} = \epsilon u(t) - s/\tau_s - (f_{\text{in}} - 1)/\tau_f. \tag{7}$$

$\epsilon$, $\tau_s$, and $\tau_f$ are the three unknown parameters that determine the dynamics of this component of the hemodynamic model. They represent the efficacy with which neuronal activity causes an increase in signal, the time-constant for signal decay or elimination, and the time-constant for autoregulatory feedback from blood flow. The existence of this feedback term can be inferred from: (i) poststimulus undershoots in rCBF (e.g., Irikura *et al.*, 1994) and (ii) the well-characterized vasomotor signal in optical imaging (Mayhew *et al.*, 1998). The critical aspect of the latter oscillatory (~0.1 Hz) component of intrinsic signals is that it shows variable phase relationships from region to region, supporting strongly the notion of local closed-loop feedback mechanisms as modelled in Eq. (6) and Eq. (7).

There are three unknown parameters for each of the two components of the hemodynamic model above (see also Fig. 2 for a schematic summary). Figure 3 illustrates the behavior of the hemodynamic model for typical values of the six parameters ($\epsilon = 0.5$, $\tau_s = 0.8$, $\tau_f = 0.4$, $\tau_0 = 1$, $\alpha = 0.2$, $E_0 = 0.8$, and assuming $V_0 = 0.02$ here and throughout). We have used a very high value for oxygen extraction to accentuate the early dip (see Discussion). Following a short-lived neuronal transient a substantial amount of signal is created and starts to decay immediately. This signal induces an increase in flow that itself augments signal decay, to the extent the signal is suppressed below resting levels (see the upper left panel in Fig. 3). This behavior is homologous to a very dampened oscillator. Increases in flow (lower left panel) dilate the venous balloon which responds by ejecting deoxyhemoglobin. In the first few hundred milliseconds the net deoxyhemoglobin ($q$) increases with an accelerating inflow-dependent delivery. It is then cleared by volume-dependent outflow expressing a negative peak a second or so after the positive volume ($v$) peak (the broken and solid lines in the upper right panel correspond to $q$ and $v$, respectively). This results in an early dip in the BOLD signal followed by a pronounced positive peak at about 4 s (lower right panel) that reflects the combined effects of reduced net deoxyhemoglobin, increased venous volume and consequent dilution of deoxyhemoglobin. Note that the rise and peak in volume (solid line in the upper right panel) lags flow by about a second. This is very similar to the predictions of the windkessel formulation and the empirical results presented in Mandeville *et al.* (1999) (see their Fig. 2). After about 8 s the inflow experiences a rebound due to its suppression of the perfusion signal. The reduced venous volume and ensuing outflow permit a reaccumulation of deoxyhemoglobin and a consequent undershoot in the BOLD signal.

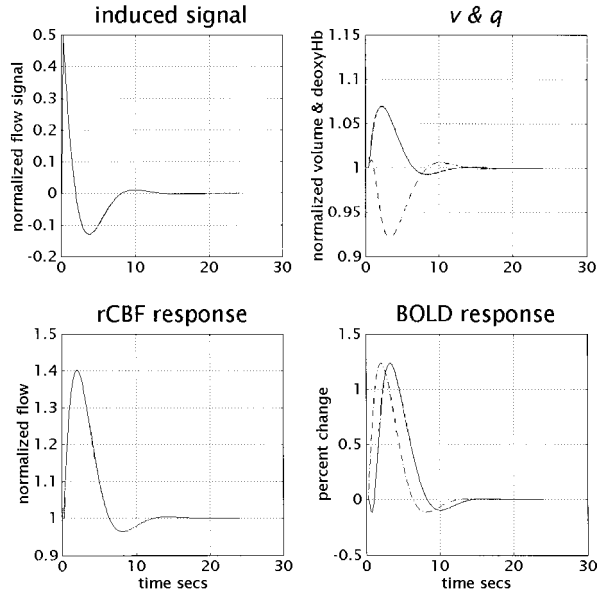


**FIG. 3.** Dynamics of the hemodynamic model. Upper left panel: The time-dependent changes in the neuronally induced perfusion signal that causes an increase in blood flow. Lower left panel: The resulting changes in normalised blood flow ($f$). Upper right panel: The concomitant changes in normalized venous volume ($v$) (solid line) and normalized deoxyhemoglobin content ($q$) (broken line). Lower right panel: The percentage change in BOLD signal that is contingent on $v$ and $q$. The broken line is inflow normalized to the same maximum as the BOLD signal. This highlights the fact that BOLD signal lags the rCBF signal by about a second.

The rCBF component of the hemodynamic model is a linear dynamical system and as such has only zeroth and first order kernels. This means it *cannot* account

for the hemodynamic refractoriness and nonlinearities observed in BOLD responses. Although the rCBF component may facilitate the Balloon component's capacity to model nonlinearities (by providing appropriate input), the rCBF component alone cannot generate second order kernels. The question addressed in this paper is whether the Balloon component can produce second order kernels that are realistic and do so with physiologically plausible parameters.

## MODEL PARAMETER ESTIMATION

In this section we describe the data used to estimate the Volterra kernels. The six unknown parameters of the hemodynamic model that best reproduce these empirical kernels are then identified. By minimizing the difference between the model kernels and the empirical kernels the optimal parameters for any voxel can be determined. The critical questions this section addresses are (i) "can the hemodynamic model account for the form of empirical kernels up to second order?" and (ii) "are the model parameters required to do this physiologically plausible?"

### *Empirical Analyses*

The data and Volterra kernel estimation are described in detail in Friston *et al.* (1998). In brief we obtained fMRI time-series from a single subject at 2 Tesla using a Magnetom VISION (Siemens, Erlangen) whole body MRI system, equipped with a head volume coil. Contiguous multi-slice $T_2^*$-weighted fMRI images were obtained with a gradient echoplanar sequence using an axial slice orientation (TE = 40 ms, TR = 1.7 s, 64 × 64 × 16 voxels). After discarding initial scans (to allow for magnetic saturation effects) each time-series comprised 1200 vol images with 3-mm isotropic voxels. The subject listened to monosyllabic or bisyllabic concrete nouns (i.e., "dog," "radio," "mountain," "gate") presented at five different rates (10, 15, 30, 60, and 90 words per minute) for epochs of 34 s, intercalated with periods of rest. The five presentation rates were successively repeated according to a Latin Square design.

The data were processed within SPM (Wellcome Department of Cognitive Neurology, http://www.fil.ion.ucl.ac.uk/spm). The time-series were realigned, corrected for movement-related effects, and spatially normalized into the standard space of Talairach and Tournoux (1988). The data were smoothed spatially with a 5-mm isotropic Gaussian kernel. Volterra kernels were estimated by expanding the kernels in terms of temporal basis functions and estimating the kernel coefficients up to second order using a generalised linear model (Worsley and Friston, 1995). The basis set comprised three gamma varieties of increasing dispersion and their temporal derivatives (as described in Friston *et al.*, 1998).

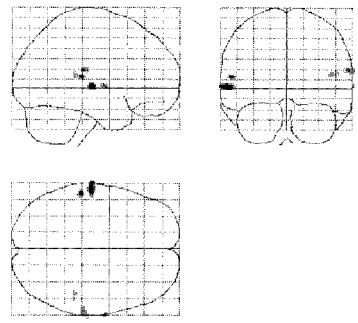

**FIG. 4.** Voxels used to estimate the parameters of the hemodynamic model shown in Fig. 2. This is a SPM{*F*} testing for the significance of the first and second order kernel coefficients in the empirical analysis and represents a maximum intensity projection of a statistical process of the *F* ratio, following a multiple regression analysis at each voxel. This regression analysis estimated the kernel coefficients after expanding them in terms of a small number of temporal basis functions [see Friston *et al.* (1998) for details]. The format is standard and provides three orthogonal projections in the standard space conforming to that described in Talairach and Tournoux (1988). The grey scale is arbitrary and the SPM{*F*} has been thresholded to show the 128 most significant voxels.

The stimulus function $u(t)$, the supposed neuronal activity, was simply the word presentation rate at which the scan was acquired. We selected voxels that showed a robust response to stimulation from two superior temporal regions in both hemispheres (see Fig. 4). These were the 128 voxels showing the most significant response when testing for the null hypothesis that the first and second order kernels were jointly zero. Selecting these voxels ensured that the kernel estimates had minimal variance.

### *Estimating the Model Parameters*

For each voxel we identified the six parameters of the hemodynamic model of the previous section whose kernels corresponded, in a least squares sense, to the empirical kernels for that voxel. To do this we used nonlinear function minimization as implemented in MATLAB5 (MathWorks Inc., MA). The model's kernels were computed, for a given parameter vector, as described in the Appendix and entered, with the corresponding empirical estimates, into the objective function that was minimized.

### *Results*

The model-based and empirical kernels for the first voxel are shown in Fig. 5. It can be seen that there is a remarkable agreement both in terms of the first and second order kernels. This is important because it suggests that the nonlinearities inherent in the Balloon component of the hemodynamic model are sufficient to account for the nonlinear responses observed in real time-series. The first-order kernel corresponds to the conventional (first-order) hemodynamic response func-

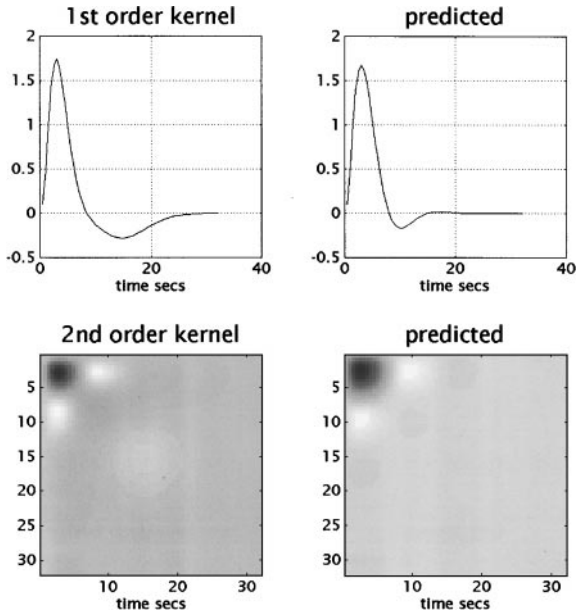


**FIG. 5.** The first and second order Volterra kernels based on parameter estimates from a voxel in the left superior temporal gyrus at −56, −28, 12 mm. These kernels can be thought of as a second order hemodynamic response function. The first order kernels (upper panels) represent the (first order) component usually presented in linear analyses. The second order kernels (lower panels) are presented in image format. The color scale is arbitrary; white is positive and black is negative. The left-hand panels are kernels based on parameter estimates from the analysis described in the legend of Fig. 4. The right hand panels are the kernels associated with the hemodynamic model using parameter estimates that best match the empirical kernels.

tion and shows the characteristic peak at about 4 s and the poststimulus undershoot. The empirical undershoot appears more protracted than the model's prediction, suggesting that the model is not perfect in every respect. The second order kernel has a pronounced negativity on the upper left, flanked by two smaller positivities. This negativity accounts for the refractoriness seen when two stimuli are temporally proximate, where this proximity is defined by the radius of the negative region. From the perspective of the Balloon model the second stimulus is compromised, in terms of elaborating a BOLD signal, because of the venous pooling, and consequent dilution of deoxyhemoglobin, incurred by the first stimulus. This means that less deoxyhemoglobin can be cleared for a given increase in flow. More interesting are the positive regions, which suggest stimuli separated by about 8 s should show super-additive effects. This can be attributed to the fact that, during the flow undershoot following the first stimulus, deoxyhemoglobin concentration is greater than normal (see the upper right panel in Fig. 3), thereby facilitating clearance of deoxyhemoglobin following the second stimulus.

Figure 6 shows the various functions implied by the hemodynamic model parameters averaged over all voxels. These include outflow as a function of venous volume $f_{out}(v, \alpha)$ and oxygen extraction fraction as a function of inflow. The solid line in the upper right panel is extraction per se $E(f_{in}, E_0)$ and the broken line is the net normalized delivery of deoxyhemoglobin to the venous compartment $f_{in}E(f_{in}, E_0)/E_0$. Note that although the fraction of oxygen extracted decreases with flow the net delivery of deoxygenated haemoglobin increases with flow. In other words inflow increases per se actually reduce signal. It is only the secondary effects of inflow on dilution and volume-dependent outflow that cause an increase in BOLD signal. The lower panel depicts the nonlinear function of volume and deoxyhemoglobin that represents BOLD signal $y(t) = \lambda(v, q, E_0)$. Here one observes that positive BOLD signals are expressed only when deoxyhemoglobin is low. The effect of volume is much less marked and tends to affect signal predominantly through dilution. This is consistent with the fact that $k_2 > k_3$ [see Eq. (1)] for the value of $E_0$ estimated for these data.

The distributions of the parameters over voxels are shown in Fig. 7 with their mean in brackets at the top of each panel. It should be noted that the data from which these estimates came were not independent. However, given they came from four different brain

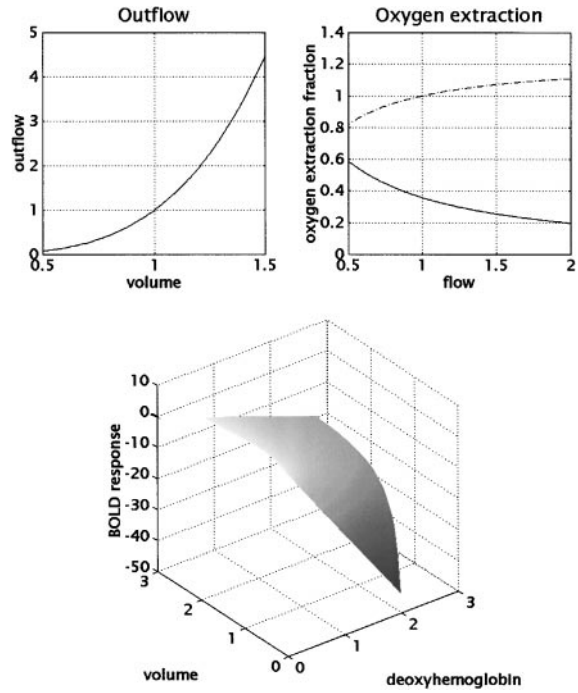


**FIG. 6.** Functions implied by the [mean] hemodynamic model parameters over the voxels shown in Fig. 4. Upper left panel: Outflow as a function of venous volume $f_{out}(v, \alpha)$. Upper right panel: oxygen extraction as a function of inflow. The solid line is extraction per se $E(f_{in}, E_0)$ and the broken line is the net normalised delivery of deoxyhemoglobin to the venous compartment $f_{in}E(f_{in}, E_0)/E_0$. Lower panel: This is a plot of the nonlinear function of volume and deoxyhemoglobin that represents BOLD signal $y(t) = \lambda(v, q, E_0)$.

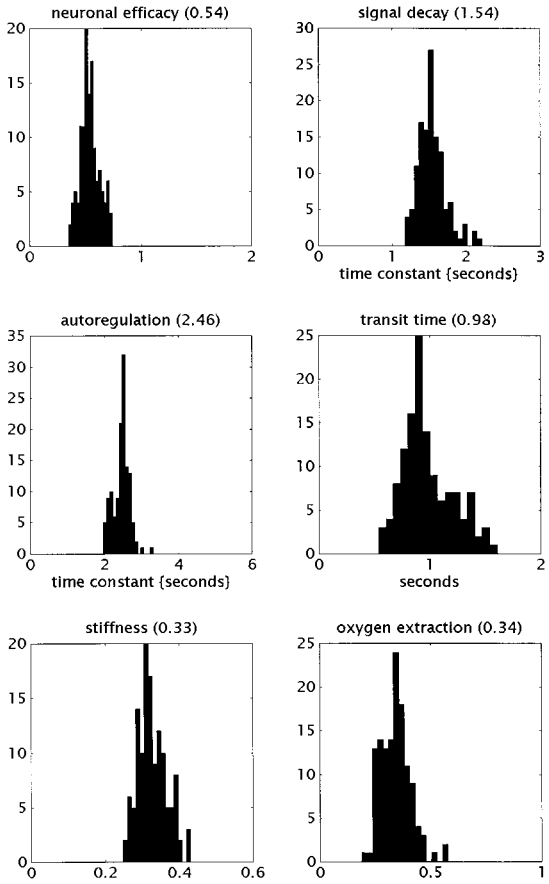


**FIG. 7.** Histograms of the distribution of the six free parameters of the hemodynamic model estimated over the voxels shown in Fig. 3. The number in brackets at the top of each histogram is the mean value for the parameters in question: neuronal efficacy is $\epsilon$, signal decay is $\tau_s$, autoregulation is $\tau_f$, transit time is $\tau_0$, stiffness is $\alpha$, and oxygen extraction is $E_0$.

regions they are remarkably consistent. In the next section we will discuss each of these parameters and the effect it exerts on the BOLD response.

## DISCUSSION

The main point to be made here is that the Balloon model, suitably extended to incorporate the dynamics of rCBF induction by synaptic activity, is sufficient to reproduce the same form of Volterra kernels that are seen empirically. As such the Balloon model is sufficient to account for the more important nonlinearities observed in evoked fMRI responses. The remainder of this section deals with the validity of the hemodynamic model in terms of the plausibility of the parameter estimates from the previous section. The role of each parameter, in shaping the hemodynamic response, is illustrated in the associated panel in Fig. 8 and is discussed in the following subsections.

### *The Neuronal Efficacy ($\epsilon$)*

This represents the increase in perfusion signal elicited by neuronal activity, expressed in terms of event density (i.e., number of evoked transients per second). From a biophysical perspective it is not exceedingly interesting because it reflects both the potency of the stimulus in eliciting a neuronal response and the efficacy of the ensuing synaptic activity to induce the signal. It is interesting to note, however, that one word per second invokes an increase in normalised rCBF of unity (i.e., in the absence of regulatory effects, a doubling of blood flow over a second). As might be expected changes in this parameter simply modulate the evoked hemodynamic responses (see the first panel in Fig. 8).

### *Signal Decay ($\tau_s$)*

This parameter reflects signal decay or elimination. Transduction of neuronal activity into perfusion

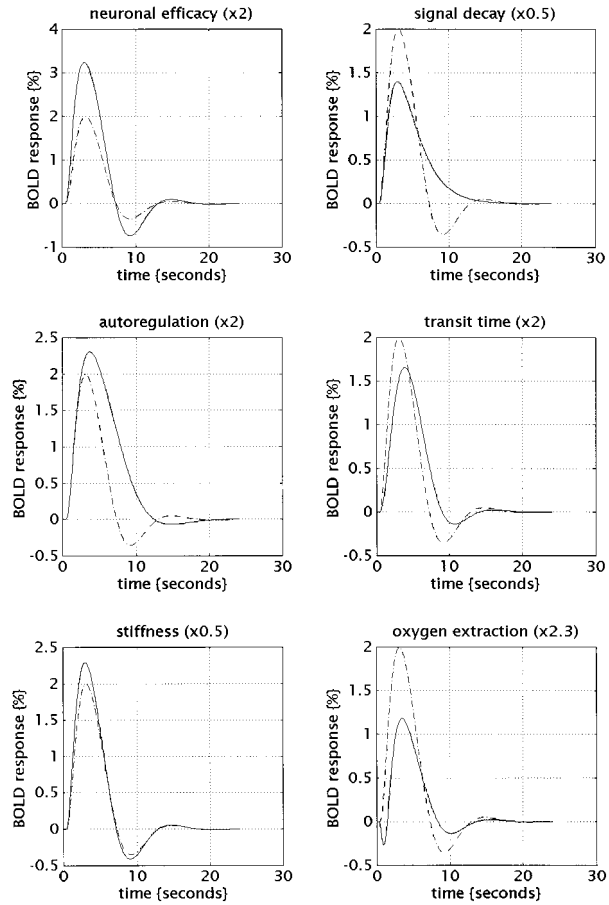


**FIG. 8.** The effects of changing the model parameters on the evoked BOLD response. The number in brackets at the top of each graph is the factor applied to the parameter in question. Solid lines correspond to the response after changing the parameter and the broken line is the response for the original parameter values (the mean values given in Fig. 7): neuronal efficacy is $\epsilon$, signal decay is $\tau_s$, autoregulation is $\tau_f$, transit time is $\tau_0$, stiffness is $\alpha$, and oxygen extraction is $E_0$.

changes, over a few hundred micrometers, has a substantial neurogenic component (that may be augmented by electrical conduction up the vascular endothelium). However, at spatial scales of several mm it is likely that rapidly diffusing spatial signals mediate increases in rCBF through relaxation of arteriolar smooth muscle. There are a number of candidates for this signal, nitric oxide (NO) being the primary one. It has been shown that the rate of elimination is critical in determining the effective time-constants of hemodynamic transduction (Friston, 1995). Our decay parameter had a mean of about 1.54 s giving a half-life $t_{1/2} = \tau_s \ln 2 = 1067$ ms. The half-life of NO is between 100 and 1000 ms (Paulson and Newman, 1987), whereas that of $K^+$ is about 5 s. Our results are therefore consistent with spatial signalling with NO. It should be remembered that the model signal subsumes all the actual signalling mechanisms employed in the real brain. Increases in this parameter dampen the rCBF response to any input and will also suppress the undershoot (see next subsection) because the feedback mechanisms, that are largely responsible for the undershoot, are selectively suppressed (relative to just reducing neuronal efficacy during signal induction).

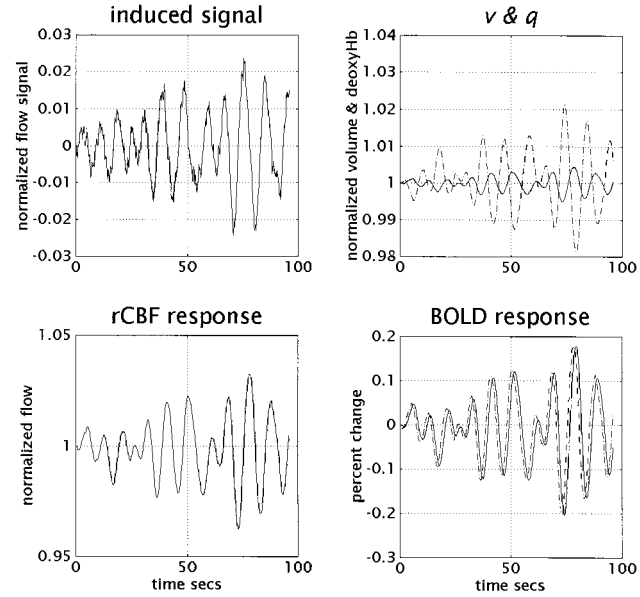


**FIG. 9.** Simulated response to a noisy neuronal input (standard deviation 1/64 and mean of 0) for a model with decreased signal decay (i.e., less dampening). The model parameters were the same as described in the legend of Fig. 3 with the exception of $\tau_s$, which was increased by a factor of 4. The characteristic 0.1 Hz oscillations are very similar to the oscillatory vasomotor signal seen in optical imaging experiments.

### *Autoregulation ($\tau_f$)*

This parameter is the time-constant of the feedback autoregulatory mechanism whose physiological nature remains unspecified (but see Irikura *et al.*, 1994). The coupled differential equations Eq. (6) and Eq. (7) represent a damped oscillator with a resonance frequency of $\bar{\omega} = 1/(2\pi\sqrt{\tau_f}) \approx 0.101$ per second. This is exactly the frequency of the vasomotor signal that typically has a period of about 10 s. This is a pleasing result that emerges spontaneously from the parameter estimation. The nature of these oscillations can be revealed by increasing the signal decay time constant (i.e., reducing the dampening) and presenting the model with low-level random neuronal input (uncorrelated Gaussian noise with a standard deviation of 1/64) as shown in Fig. 9. The characteristic oscillatory dynamics are readily expressed. The effect of increasing the feedback time constant is to decrease the resonance frequency and render the BOLD (and rCBF) response more enduring with a reduction or elimination of the undershoot. The third panel in Fig. 8 shows the effect of doubling $\tau_f$.

### *Transit Time ($\tau_0$)*

This is an important parameter that determines the dynamics of the signal. It is effectively resting venous volume divided by resting flow, and in our data is estimated at about one second (0.98 s). The transit time through the rat brain is roughly 1.4 seconds at rest and, according to the asymptotic projections for rCBF and volume, falls to 0.73 seconds during stimulation (Mandeville *et al.*, 1999). In other words it takes about a second for a blood cell to traverse the venous compartment. The effect of increasing mean transit time is to slow down the dynamics of the BOLD signal with respect to the flow changes. The shape of the response remains the same but it is expressed more slowly. In the fourth panel of Fig. 8 a doubling of the mean transit time is seen to retard the peak BOLD response by about a second and the undershoot by about 2 s.

### *Stiffness Parameter ($\alpha$)*

Under steady state conditions this would be about 0.38. The mean over voxels considered above was about 0.33. This discrepancy, in relation to steady state levels, is anticipated by the windkessel formulation and is attributable to the fact that volume and flow are in a state of continuous flux during the evoked responses. Recall from Eq. (3) that $1/\alpha = \gamma + \beta = 3.03$, in our data. Under the assumption of laminar flow ($\gamma = 2$), $\beta \approx 1$ which is less than Mandeville *et al.* (1999) found for rats during forepaw stimulation but is certainly in a plausible range. Increasing this parameter increases the degree of nonlinearity in the flow-volume behavior of the venous balloon that underpins the nonlinear behaviors we are trying to account for. However, its direct effect on evoked responses to single stimuli is not very marked. The fifth panel of Fig. 8 shows the effects when $\alpha$ is decreased by 50%.

*Resting Oxygen Extraction ($E_0$)*

This is about 34% and the range observed in our data fit exactly with known values for resting oxygen extraction fraction (between 20 and 55%). Oxygen extraction fraction is a potentially important factor in determining the nature of evoked fMRI responses because it may be sensitive to the nature of the baseline that defines the resting state. Increases in this parameter can have quite profound effects on the shape of the response that bias it toward an early dip. In the example shown (last panel in Fig. 8) the resting extraction has been increased to 78%. This is a potentially important observation that may explain why the initial dip has been difficult to observe in all studies. According to the results presented in Fig. 8 the initial dip is very sensitive to resting oxygen extraction fraction, which should be high before the dip is expressed. Extraction fraction will be high in regions with very low blood flow, or in tissue with endogenously high extraction. It may be that cytochrome oxidase rich cortex, like the visual cortices, may have a higher fraction and be more like to evidence early dips.

In summary the parameters of the hemodynamic model that best reproduce empirically derived Volterra kernels are all biologically plausible and lend the model a construct validity (in relation to the Volterra formulation) and face validity (in relation to other physiological characterisations of the cerebral hemodynamics reviewed in this section). In this extended hemodynamic model nonlinearities, inherent in the Balloon model, have been related directly to nonlinearities in responses. Their role in mediating the poststimulus undershoot is emphasised less here because the rCBF component can model undershoots.

The conclusions above are based only on data from the auditory cortex and from one subject. There is no guarantee that they will generalize. One of our reviewers thought that this paper was more important for its conceptual motivation of modelling than for the specific findings. This is a very valid point. We anticipate that the framework presented here will be refined or changed when applied to other data, or the assumptions upon which it is based are confirmed or refuted.

## CONCLUSION

In conclusion we have developed an input-state-output model of the hemodynamic response to changes in synaptic activity that combines the Balloon model of flow to BOLD signal coupling and a dynamical model of the transduction of neuronal activity into perfusion changes. This model has been characterized in terms of its Volterra kernels and easily reproduces empirical kernels with parameters that are biologically plausible. This means that the nonlinearities inherent in the Balloon model are sufficient to account for hemodynamic refractoriness and other nonlinear aspects of evoked responses in fMRI.

## APPENDIX

Volterra kernels represent a generic and important characterization of the invariant aspects of a nonlinear system (see Bendat, 1990). This appendix describes the nature of these kernels and how they are obtained given the differential equations describing the evolution of the state variables. Consider the single input-single-output (SISO) system

$$\begin{aligned} \dot{X}(t) &= f(X, u(t)) \\ y(t) &= \lambda(X(t)), \end{aligned} \tag{A.1}$$

where, for the hemodynamic model, $X = \{x_1, x_2, x_3, x_4\}^T = \{s, f_{in}, v, q\}^T$ with

$$\begin{aligned} \dot{x}_1 &= f_1(X, u(t)) = \epsilon u(t) - \frac{x_1}{\tau_s} - \frac{x_2 - 1}{\tau_f} \\ \dot{x}_2 &= f_2(X, u(t)) = x_1 \\ \dot{x}_3 &= f_3(X, u(t)) = \frac{1}{\tau_0}(x_2 - f_{out}(x_3, \alpha)) \\ \dot{x}_4 &= f_4(X, u(t)) = \frac{1}{\tau_0}\left(x_2 \frac{E(x_2, E_0)}{E_0} - f_{out}(x_3, \alpha)\frac{x_4}{x_3}\right) \end{aligned}$$

and

$$\begin{aligned} y(t) = \lambda(X(t)) = V_0(k_1(1 - x_4) \\ + k_2(1 - x_4/x_3) + k_3(1 - x_3)). \end{aligned}$$

The Volterra series expresses the output $y(t)$ as a nonlinear convolution of the neuronal inputs $u(t)$, critically without reference to the state variables $X(t)$. This series can be considered a nonlinear convolution that obtains from a functional Taylor expansion of $y(t)$ about $X(0) = X_0 = [0, 1, 1, 1]^T$ and $u(t) = 0$,

$$\begin{aligned} y(t) = \kappa_0(t) + \sum_{i=1}^{\infty} \int_0^t \cdots \int_0^t \kappa_i(t, \sigma_1, \cdots \sigma_i) \\ \times u(\sigma_1) \cdots u(\sigma_i) d\sigma_1 \cdots d\sigma_i \\ \kappa_i(t, \sigma_1, \cdots \sigma_i) = \frac{\partial^i y(t)}{\partial u(\sigma_1) \cdots u(\sigma_i)}, \end{aligned} \tag{A.2}$$

where $\kappa_i$ is the $i$th, generally time-dependent, kernel. The Taylor expansion of $\dot{X}(t)$ about $X_0$ and $u(t) = 0$:

$$\dot{X}(t) \approx f(X_0, 0) + \frac{\partial f(X_0, 0)}{\partial X}(X - X_0) + \frac{\partial^2 f(X_0, 0)}{\partial X \partial u}(X - X_0)u + \frac{\partial f(X_0, 0)}{\partial u}u$$

has a bilinear form following a change of variables (equivalent to adding an extra state variable $x_0(t) = 1$):

$$\dot{X}'(t) \approx AX' + BX'u$$

$$X' = \begin{bmatrix} 1 \\ X \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 0 \\ \left(f(X_0, 0) - \dfrac{\partial f(X_0, 0)}{\partial X} X_0\right) & \dfrac{\partial f(X_0, 0)}{\partial X} \end{bmatrix} \quad \text{(A.3)}$$

$$B = \begin{bmatrix} 0 & 0 \\ \left(\dfrac{\partial f(X_0, 0)}{\partial u} - \dfrac{\partial^2 f(X_0, 0)}{\partial X \partial u} X_0\right) & \dfrac{\partial^2 f(X_0, 0)}{\partial X \partial u} \end{bmatrix}.$$

This formulation is important because the Volterra kernels of bilinear systems have closed-form expressions. The existence of these closed-form expressions is due to the fact that the iterated integrals associated with the system's Generating Series can be expressed in terms of the generalized convolution integrals, of which the Volterra series is comprised (Fliess *et al.*, 1983). Here we take a more heuristic approach and consider the solution to A.2 and its derivatives with respect to the inputs $u(t)$

$$X'(\Delta t) \approx e^{\Delta t(A + Bu(0))} X'(0) \Rightarrow X'(T\Delta t) \approx \prod_{j=T-1}^{0} e^{\Delta t(A + Bu(j\Delta t))} X'(0),\ \Delta t \to 0$$

$$\frac{\partial^i X'(T\Delta t)}{\partial u(\tau_1 \Delta t) \cdots u(\tau_i \Delta t)} = \prod_{j=T-1}^{\tau_i + 1} e^{\Delta t(A + Bu(j\Delta t))} \times B \prod_{j=\tau_i}^{\tau_{i-1}+1} e^{\Delta t(A + Bu(j\Delta t))} \cdots \times B \prod_{j=\tau_1}^{0} e^{\Delta t(A + Bu(j\Delta t))} X'(0).$$

The kernels associated with the state variables $X'$ are these derivatives evaluated at $u(t) = 0$:

$$\chi_i(t, \sigma_1, \cdots \sigma_i) = \frac{\partial^i X'(t)}{\partial u(\sigma_1) \cdots u(\sigma_i)} = e^{(t-\sigma_i)A} B e^{(\sigma_i - \sigma_{i-1})A} \cdots B e^{\sigma_1 A} X'(0);$$

i.e.,

$$\chi_0(t) = e^{tA} X'(0)$$

$$\chi_1(t, \sigma_1) = e^{(t-\sigma_1)A} B e^{\sigma_1 A} X'(0)$$

$$\chi_2(t, \sigma_1, \sigma_2) = e^{(t-\sigma_2)A} B e^{(\sigma_2 - \sigma_1)A} B e^{\sigma_1 A} X'(0)$$

$$\chi_2(t, \sigma_1, \sigma_2, \sigma_3) = \cdots$$

The kernels associated with the output $y(t)$ follow from the chain rule:

$$\kappa_0(t) = \lambda(\chi_0(t))$$

$$\kappa_1(t, \sigma_1) = \frac{\partial \lambda(\chi_0(t))}{\partial X'} \chi_1(t, \sigma_1)$$

$$\kappa_2(t, \sigma_1, \sigma_2) = \frac{\partial \lambda(\chi_0(t))}{\partial X'} \chi_2(t, \sigma_1, \sigma_2) + \chi_1(t, \sigma_1)^{\mathrm{T}} \frac{\partial^2 \lambda(\chi_0(t))}{\partial X'^2} \chi_1(t, \sigma_2)$$

$$\kappa_2(t, \sigma_1, \sigma_2, \sigma_3) = \cdots$$

If the system is fully nonlinear, as in this case, then the kernels can be considered local approximations. In other words the kernels are only valid for inputs (i.e., neuronal activations) of a reasonable magnitude.

## ACKNOWLEDGMENTS

This work was funded by the Wellcome trust. We thank Gary Green for guidance and support in understanding and using Volterra series.

## REFERENCES

Bendat, J. S. 1990. *Nonlinear System Analysis and Identification from Random Data.* Wiley, New York.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* **16:** 4207–4221.

Buxton, R. B., and Frank, L. R. 1997. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J. Cereb. Blood Flow Metab.* **17:** 64–72.

Buxton, R. B., Wong, E. C., and Frank, L. R. 1998. Dynamics of blood flow and oxygenation changes during brain activation: The Balloon model. *MRM* **39:** 855–864.

Fliess, M., Lamnabhi, M., and Lamnabhi-Lagarrigue, F. 1983. An algebraic approach to nonlinear functional expansions. *IEEE Trans. Circuits Syst.* **30:** 554–570.

Friston, K. J., Jezzard, P., and Turner, R. 1994. Analysis of functional MRI time series. *Hum. Brain Mapp.* **1:** 153–171.

Friston, K. J. 1995. Regulation of rCBF by diffusible signals: An analysis of constraints on diffusion and elimination. *Hum. Brain Mapp.* **3:** 56–65.

Friston, K. J., Josephs, O., Rees, G., and Turner, R. 1998. Nonlinear event-related responses in fMRI. *MRM* **39:** 41–52.

Grubb, R. L., Rachael, M. E., Euchring, J. O., and Ter-Pogossian, M. M. 1974. The effects of changes in $PCO_2$ on cerebral blood volume, blood flow and vascular mean transit time. *Stroke* **5:** 630–639.

Hoge, R. D., Atkinson, J., Gill, B., Crelier, G. R., Marrett, S., and Pike, G. B. 1999. Linear coupling between cerebral blood flow and oxygen consumption in activated human cortex. *Proc. Natl. Acad. Sci. USA* **96:** 9403–9408.

Irikura, K., Maynard, K. I., and Moskowitz, M. A. 1994. Importance of nitric oxide synthase inhibition to the attenuated vascular responses induced by topical l-nitro-arginine during vibrissal stimulation. *J. Cereb. Blood Flow Metab.* **14:** 45–48.

Mandeville, J. B., Marota, J. J., Ayata, C., Zararchuk, G., Moskowitz, M. A., Rosen, B., and Weisskoff, R. M. 1999. Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. *J. Cereb. Blood Flow Metab.* **19:** 679–689.

Mayhew, J., Hu, D., Zheng, Y., Askew, S., Hou, Y., Berwick, J., Coffey, P. J., and Brown, N. 1998. An evaluation of linear models analysis techniques for processing images of microcirculation activity. *NeuroImage* **7:** 49–71.

Miller, K. L., Luh, W. M., Liu, T. T., Martinez, A., Obata, T., Wong, E. C., Frank, L. R., and Buxton, R. B. 2000. Characterizing the dynamic perfusion response to stimuli of short duration. *Proc. ISRM* **8:** 580.

Paulson, O. B., and Newman, E. A. 1987. Does the release of potassium from astrocyte endfeet regulate cerebral blood? *Science* **237:** 896–898.

Talairach, J., and Tournoux, P. 1988. *A Co-planar Stereotaxic Atlas of a Human Brain*. Thieme, Stuttgart.

Vazquez, A. L., and Noll, D. C. 1996. Non-linear temporal aspects of the BOLD response in fMRI. *Proc. Int. Soc. Magn. Res. Med.* **3:** S1765.

Worsley, K. J., and Friston, K. J. 1995. Analysis of fMRI time-series revisited—Again. *NeuroImage* **2:** 173–181.