Parameter estimation efficiency using nonlinear models in fMRI

Thomas Deneux, Olivier Faugeras

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Parameter estimation efficiency using nonlinear models in fMRI

Thomas Deneux — Olivier Faugeras

N° 5758

Novembre 2005

Thème BIO

ISSN 0249-6399 ISRN INRIA/RR--5758--FR+ENG

INRIA
SOPHIA ANTIPOLIS

Parameter estimation efficiency using nonlinear models in fMRI

Thomas Deneux * , Olivier Faugeras †

Thème BIO — Systèmes biologiques
Projets Odyssée

Rapport de recherche n° 5758 — Novembre 2005 — 35 pages

Abstract: There is an increasing interest in using physiologically plausible models in fMRI analysis. These models do raise new mathematical problems in terms of parameters estimation and interpretation of the measured data.

We present some theoretical contributions in this area, using different variations of the Balloon Model (Buxton et al., 1998; Friston et al., 2000; Buxton et al., 2004) as example models. We propose 1) a method to analyze the models dynamics and their stability around equilibrium, 2) a new way to derive least square energy gradient for parameter estimation, 3) a quantitative measurement of parameter estimation efficiency, and 4) a statistical test for detecting voxel activations.

We use these methods in a visual perception checker-board experiment. It appears that the different hemodynamic models considered better capture some features of the response than linear models. In particular, they account for small nonlinearities observed for stimulation durations between 1 and 8 seconds. Nonlinearities for stimulation shorter than one second can also be explained by a neural habituation model (Buxton et al., 2004), but further investigations should assess whether they are rather not due to nonlinear effects of the flow response.

Moreover, the tools we have developed prove that statistical methods that work well for the GLM can be nicely adapted to nonlinear models. The activation maps obtained in both frameworks are comparable.

Key-words: Nonlinear hemodynamic, Balloon Model, system identification

* thomas.deneux@ens.fr
† olivier.faugeras@sophia.inria.fr

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 — Télécopie : +33 4 92 38 77 65

Utilisation de modèles non-linéaires en IRMf: estimation de paramètres et applications

Résumé : L'utilisation de modèles physiologiquement plausible dans l'analyse des données d'IRM fonctionnelle connaît un intérêt grandissant. Ces modèles soulèvent de nouveaux problèmes mathématiques en ce qui concerne l'estimation de leur paramètres et l'interprétation des données.

Nous présentons des contributions théroriques dans ce domaine, en utilisant plusieurs variations du "Balloon Model" (Buxton et al., 1998; Friston et al., 2000; Buxton et al., 2004). Nous développons 1) une méthode pour analyser les dynamisques des modèles, et leur stabilité autour de l'équilibre, 2) une nouvelle manière de calculer le gradient de l'énergie des moindres carrés utilisée dans l'estimation des paramètres, 3) une mesure quantitative de la précision de cette estimation des paramètres, et 4) un test statistique pour détecter l'activation voxel par voxel.

Nous utilisons ces méthodes pour l'analyse d'une expérience de perception. Les modèles hémodynamiques considérés sont apparus capables de rendre compte de certains aspects de la réponse que les modèles linéaires ignoraient. En particulier, les non-linéarités observées pour des stimulations de une à huit secondes. Les non-linéarités observées pour des stimulations plus courtes ont également pu être expliquées par un modèle d'habituation neuronale (Buxton et al., 2004), mais nous nous demandons si elles n'ont pas lieu en réalité dans la réponse du flux sanguin.

Les outils que nous avons développés prouvent que les méthodes statistiques couramment utilisée dans le cadre du Modèle Linéaire Général (GLM) peuvent ête adaptées aux modèles non-linéaires. Les cartes d'activation obtenues avec les deux approches sont en réalité très similaires.

Mots-clés : modèle hémodynamique, identification de système dynamique non linéaire

1 Introduction

Most fMRI analyses rely on the hypothesis that the BOLD response is an affine function of the neural activity. This hypothesis allows the use of such powerful tools as linear regressions and statistical tests (Friston et al., 1995).

Many studies have considered the question of the range of validity of this linear assumption. They all agree on the fact that it holds for stimulation duration or interstimulus intervals (ISI) larger than a threshold. The value of this threshold varies among studies : 2-3 seconds (Boynton et al., 1996; Dale and M., 1997) to 4-6 seconds (Birn et al., 2001; Glover, 1999; Miller et al., 2001; Vazquez and Noll, 1998). Studies involving other measurement modalities established that some nonlinearities in the BOLD were not present at the neural level, and hence were due to hemodynamic effects: blood flow measurement in humans via Arterial Spin Labelling (Miller et al., 2001; Obata et al., 2004), or electrical activity in animals (Janz et al., 2001). Other objections to linearity were raised also, like the apparition of a drift in the BOLD during long stimulations (Krüger et al., 1999).

Besides these experimental observations, there has been a sustained effort to model the long chain that extends between neural activity and the BOLD response: which part of neural activity best correlates with fMRI? (Logothetis and Pfeuffer, 2004), energy consumption and metabolic demand (Aubert and Costalat, 2002), blood flow increase signal (Friston et al., 2000; Glover, 1999), vascular mechanic and oxygen extraction (Buxton et al., 1998; Hoge et al., 2005; Zheng et al., 2002), paramagnetic effect of the deoxyhemoglobin (Ogawa et al., 1993; Davis et al., 1998). A very detailed model can be found in (Aubert and Costalat, 2002), while (Buxton et al., 2004) presents a simplified synthesis.

There have been several attempts to handle nonlinearities in fMRI studies. Some consist in characterizing them as empirical functions of the stimulation patterns, via Volterra kernels (Friston et al., 1998) or specific basis functions (Wager et al., 2005) that could be integrated to the GLM. Others rather bring physiological models in the analysis: they replace linear regression by fitting model output to measured data via parameter estimation.

Thus, Friston and colleagues (Friston et al., 2000) worked with Buxton's Balloon Model (Buxton et al., 1998), to which they added a damped oscillator in order to model blood flow. They estimated the model parameters in activated voxels using a Volterra kernel expansion to characterize the model dynamics. Neural signal time courses were approximated by the stimulus up to a scaling factor called neural efficiency, which was estimated as well. Later (2002) they introduced a Bayesian estimation framework that allowed the use of priors on parameters values, didn't need the Volterra kernels any more, and eventually produced a posteriori probability distributions of the parameters.

Riera et al. (2004) used the same physiological model in a more general framework, allowing noise in the evolution equations in addition to measurement noise. They used a local linearization filter in the spirit of Kalman filter methodology. This filter allowed them to compute an estimation of the system input. This estimation had to lie in a specific vector space, otherwise the problem was underdetermined. This method can be compared to deconvolution tentatives in the linear framework (Glover, 1999).

We develop our estimation algorithm in the same framework as Friston et al. (2000), i.e. with a stimulus-locked input to the system, no priors on the parameters, and a Gaussian measurement noise model. These hypotheses allow an easier interpretation and quantification of the estimation results, since there are only a few indeterminate variables. Estimating the model parameters can be formulated as a simple least square minimization problem. In order to achieve the energy minimization, we propose to compute the system output gradient with respect to the parameters through the integration of a new differential system. It requires no particular form for the input, makes no linear approximation, and the estimation is robust to low frequency drifts in the data.

Once the parameters have been estimated, it is highly important to evaluate the estimation accuracy. In effect, it turns out that some parameters are poorly identifiable, because they do not influence much the model output, or because their effect on the ouput interferes with that of other parameters. We propose a sensitivity analysis method, relying on the system output derivative with respect to the parameters, to quantify the identifiability of each parameter. A parallel can be made with the Bayesian framework in (Friston, 2002).

Last, with the interpretation of estimation results the question arises of detecting activations. As proposed by Friston et al. , the efficiency parameter estimated in each voxel is a good candidate to measure activation. But due to identification problems its values do interfere with those of other parameters. So we prefer an activation detection based on statistical significance of the fit between predicted output and measured data. We thus propose a F-test to answer this question.

Before we dive into the details of our contribution, we start with a short analysis of the Balloon Model dynamics.

2 Methods

Physiological models for the BOLD response can be formulated as input-state-output systems (Friston et al., 2000), the input u being the stimulus function, the state x being a set of non-measurable variables, and the output y being the BOLD signal at the same voxel. This system is driven by the following evolution and measurement equations:

$$\left\{ \begin{array}{rcl} \dot{x}(t,u,\theta) & = & F(x(t,u,\theta),u(t),\theta) \\ y(t,u,\theta) & = & G(x(t,u,\theta),\theta), \end{array} \right. \tag{1}$$

where F and G are nonlinear functions, and θ represents the set of model parameters.

The Balloon Model proposed by Buxton and al. (1997; 1998) and completed by Friston (2002) (flow dynamic) describes the dynamics of the blood flow f, the blood venous volume

v, the veins deoxyhemoglobin content q (these values are normalized and thus equal 1 at rest), and the BOLD signal y:

$$\left\{\begin{array}{rcl} \ddot{f} &=& \epsilon\, u - \kappa_s \dot{f} - \kappa_f (f-1) \\ \dot{v} &=& \frac{1}{\tau}(f - v^{1/\alpha}) \\ \dot{q} &=& \frac{1}{\tau}(f\frac{1-(1-E_0)^{1/f}}{E_0} - v^{1/\alpha - 1}q) \\ y &=& V_0(k_1 q + k_2 q/v + k_3 v) \approx V_0(a_1(1-q) - a_2(1-v)). \end{array}\right. \tag{2}$$

Note that to match the general formulation in (1), it is necessary to add $\dot{f}$ as a state variable (Friston et al. considered it as a physiological "flow inducing signal"). The system evolution parameters are the neural efficiency ϵ, the flow decay κ_s, the flow time constant κ_f, the venous transit time τ, Grub's parameter α, the oxygen extraction at rest E_0 and the blood volume fraction at rest V_0; they may vary across brain regions and across subjects. The measurement parameters a_1 and a_2 are scanner-dependent. Their values have been evaluated to $a_1 = 7E_0 + 2$ and $a_2 = -2E_0 + 2.2$ for a 1.5 T scanner (Buxton et al., 1998). But our experimental data was acquired on a 3T scanner, and less is known at this field strength, except that the volume effect a_2 is smaller. We used $a_2 = a_1/9$ and considered the product $b = V_0(a_1 + a_2)$ as an unknown quantity, leading to measurement equation $y = b(0.9 * (1-q) - 0.1 * (1-v))$. There is too much indetermination in the parameters estimation indeed, to allow us to estimate both a_1 and a_2.

There have been several enhancements of the Balloon Model since then, and Buxton et al. (2004) put them together nicely recently. Three more variables are considered: the metabolism (CMRO2) m becomes an independent variable instead of the flow-locked expression $f\frac{1-(1-E_0)^{1/f}}{E_0}$; the neural activity N is the output of a simple habituation model (with a neural inhibition I) instead of the stimulus-locked expression ϵu. Flow and metabolism are not described by an evolution equations any more, but as convolutions of neural activity with gamma-variate functions:

$$\left\{\begin{array}{rcl} N &=& \epsilon\, u - I \\ \dot{I} &=& \frac{1}{\tau_I}(\kappa_n N - I) \\ f &=& 1 + (f_1 - 1)\, h_f(t - \delta_t) * N \\ m &=& 1 + (m_1 - 1)\, h_m(t) * N \\ \dot{v} &=& \frac{1}{\tau}(f - (v^{1/\alpha} + \tau_{visc}\dot{v})) \\ \dot{q} &=& \frac{1}{\tau}(m - \frac{q}{v}(v^{1/\alpha} + \tau_{visc}\dot{v})) \\ y &=& V_0(a_1(1-q) - a_2(1-v)), \end{array}\right. \tag{3}$$

with

$$\left\{\begin{array}{rcl} h_f(t) &=& \frac{1}{6\tau_f}(\frac{t}{\tau_f})^3 e^{-\frac{t}{\tau_f}} \\ h_m(t) &=& \frac{1}{6\tau_m}(\frac{t}{\tau_m})^3 e^{-\frac{t}{\tau_m}}. \end{array}\right.$$

Additional parameters are the inhibitory time constant τ_I, the inhibitory gain factor κ_n, the normalized CBF and CMRO2 responses to sustained activity f_1 and m_1, the delay

δ_t between CMRO2 and CBF responses, the widths τ_f and τ_m of the CBF and CMRO2 impulse responses, and the volume viscoelastic time constants τ^+_{visc} and τ^-_{visc} (an hysteresis rule is authorized for the volume dynamics: the viscosity parameter τ_{visc} can take 2 different values whether $\frac{\partial v}{\partial t} > 0$ ($\tau_{visc} = \tau^+_{visc}$) or $\frac{\partial v}{\partial t} < 0$ ($\tau_{visc} = \tau^-_{visc}$)).

We note that the neural habituation model results in

$$N = \epsilon\, u - \epsilon\, h_I * u,$$

where

$$h_I(t) = \frac{\kappa_n}{\tau_I} e^{-\frac{\kappa_n+1}{\tau_I}t}.$$

Thus we can write

$$\left\{ \begin{array}{lcl} f = 1 + \xi\,(h_f(t-\delta_t) - h_f(t-\delta_t) * h_I) * u & = & 1 + \xi\, H_f * u \\ m = 1 + \frac{\xi}{n}\,(h_m - h_m * h_I) * u & = & 1 + \frac{\xi}{n}\, H_m * u, \end{array} \right.$$

where $\xi = \epsilon(f_1 - 1)$, and $n = (f_1 - 1)/(m_1 - 1)$ is the steady-state flow-metabolism ratio. In the following, we will estimate ξ and n instead of ϵ, f_1 and m_1.

We also note that volume evolution equation can be transformed in

$$\dot v = \frac{1}{\tau + \tau_{visc}}(f - v^{\frac{1}{\alpha}}) = \left\{ \begin{array}{ll} \frac{1}{\tau+\tau^+_{visc}}(f - v^{\frac{1}{\alpha}}) & \text{if } f^\alpha > v \\ \frac{1}{\tau+\tau^-_{visc}}(f - v^{\frac{1}{\alpha}}) & \text{if } f^\alpha < v. \end{array} \right.$$

Hence, the new Balloon Model formulation becomes

$$\left\{ \begin{array}{lcl} \dot v & = & \frac{1}{\tau+\tau_{visc}}(1 + \xi\, H_f * u - v^{1/\alpha}) \\ \dot q & = & \frac{1}{\tau}(1 + \frac{\xi}{n}\, H_m * u - \frac{q}{v}(v^{1/\alpha} + \tau_{visc}\dot v)) \\ y & = & V_0(a_1(1-q) - a_2(1-v)). \end{array} \right. \tag{4}$$

The tools we develop in the following section will be used with the two models ((2) and (4)).

2.1 System dynamic and stability

Before analyzing in detail these two models it is interesting to build some intuition for their dynamics. We refer to previous studies for several simulated time courses of state variables and BOLD output (Buxton et al., 1998; Friston et al., 2000; Riera et al., 2004). Figure 1 demonstrates nonlinear effects of the initial Balloon Model in the response peak amplitudes by comparing responses to given stimulation lengths to their prediction from responses to shorter stimulations.

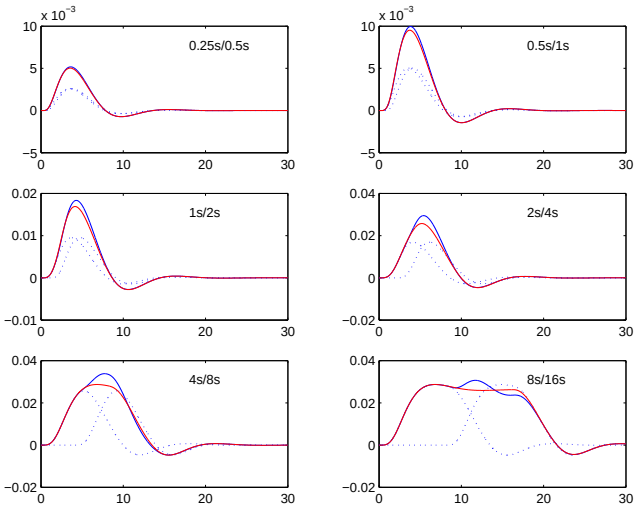

Figure 1: Balloon model simulation for increasing stimulation lengths and visualization of nonlinearities. The response to stimulation length $2T$ is compared to the sum of 2 shifted responses to stimulation length T. Nonlinearities are stronger for 2s/4s and 4s/8s comparisons ($\epsilon = 0.4$, $\kappa_s = 0.65$, $\kappa_f = 0.4$, $\tau = 1$, $\alpha = 0.4$, $E_0 = 0.4$, $V_0(a_1 + a_2) = 0.1$).

The hemodynamic main effect seems to be roughly a smoothing of the input, and it looks unlikely that any special dynamics like bifurcations, limit cycles... can occur. Indeed we prove that for a constant input u_0, there is only one stable equilibrium point.

Let us consider here the first Balloon Model formulation (2). The flow dynamic equation is a pure linear damped oscillator. It can then be computed exactly by a convolution

$$f(t) = 1 + k * u(t).$$

If we assume $\Delta = \kappa_s^2 - 4\kappa_f > 0$, we have:

$$k(t) = \epsilon e^{-\frac{\kappa_s}{2}t} cos(\frac{\sqrt{4\kappa_f - \kappa_s^2}}{2}t)$$

(if we had $\Delta < 0$, k would have been of a different form, with exponentials only).

Since it is a linear convolution, the flow cannot have any special dynamic (if the input is constant the flow converges necessarily to the equilibrium point $1 + \epsilon u_0/\kappa_f$).

One remark is that since $k(t) < 0$ for some t, f can theoretically have negative values, even if $u(t) > 0$, $\forall t$. Figure 2 shows that we can obtain flow time courses with non-realistic

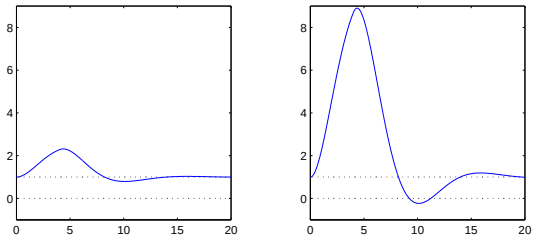

Figure 2: (A) flow time courses for a 4s input ($\epsilon = 0.4$, $\kappa_s = 0.65$, $\kappa_f = 0.4$, $\tau = 1$, $\alpha = 0.4$, $E_0 = 0.4$, $V_0(a_1 + a_2) = 0.1$). (B) flow time course becomes unrealistic for $\epsilon = 3$!

values, e.g., negative, if the product ϵu is too large (it does not happen in practice since it would require non-realistic values for ϵ).

Volume only depends on flow. If we note $v(f) = f^\alpha$, $\dot{v}$ in (2) has the same sign as $v(f) - v$. The equation looks like an exponential decay to steady state, though it is nonlinear. If the input is constant, the flow and the volume necessarily converge to their equilibrium points $(1 + \epsilon u_0/\kappa_f)$ and $(1 + \epsilon u_0/\kappa_f)^\alpha$.

In a similar way, if we note $q(v, f) = f\frac{1-(1-E_0)^{1/f}}{E_0}\, v^{1-1/\alpha}$, $\dot{q}$ is the same sign as $q(v, f) - q$. If the input is constant the deoxyhemoglobin content eventually converges to an equilibrium point.

We just gave an intuitive proof of the system stability for a constant input. From a more mathematical viewpoint, it is also possible to show it by examining the eigenvalues of the Jacobian of the evolution function F at the equilibrium point x_0.

x_0 is determined by equalling the time derivative $F(x_0, u_0, \theta)$ to zero:

$$x_0 = \begin{pmatrix} 0 \\ 1 + \frac{\epsilon u_0}{\kappa_f} \\ (1 + \frac{\epsilon u_0}{\kappa_f})^\alpha \\ \frac{1-(1-E_0)^{1/(1+\frac{\epsilon u_0}{\kappa_f})}}{E_0}(1 + \frac{\epsilon u_0}{\kappa_f})^\alpha \end{pmatrix}.$$

The jacobian of F at x is:

$$\frac{\partial F}{\partial x} = \begin{pmatrix} -\kappa_s & -\kappa_f & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\tau} & -\frac{x_3^{1/\alpha-1}}{\alpha\tau} & 0 \\ 0 & \frac{\partial F_4}{\partial x_2} & \frac{1}{\tau}(1 - \frac{1}{\alpha})x_3^{1/\alpha-2}x_4 & -\frac{x_3^{1/\alpha-1}}{\tau} \end{pmatrix},$$

with

$$\frac{\partial F_4}{\partial x_2} = \frac{1}{\tau}\left(\frac{1 - (1 - E_0)^{1/x_2}}{E_0} - \frac{\log(1 - E_0)(1 - E_0)^{1/x_2}}{E_0 x_2}\right),$$

and its eigenvalues evaluated at x_0 are

$$\{-\frac{\kappa_s+\sqrt{\kappa_s^2-4\kappa_f}}{2}, -\frac{\kappa_s-\sqrt{\kappa_s^2-4\kappa_f}}{2}, -\frac{(1+\frac{\epsilon u_0}{\kappa_f})^{1-\alpha}}{\alpha\tau}, -\frac{(1+\frac{\epsilon u_0}{\kappa_f})^{1-\alpha}}{\tau}\}$$

(they can be obtained as follows: note that the matrix $\frac{\partial F}{\partial x}$ is trigonal by blocks with block sizes equal to 2, 1, and 1; the four eigenvalues are respectively the 2 eigenvalues of the first 2x2 block, and the third and fourth diagonal terms).

Since the physiological parameters are always positive, these eigenvalues are either real and negative or have negative real parts: the system is always stable around equilibrium.

Similar considerations (intuitive interpretation of the equations as well as differentiation of the evolution function) do lead to the conclusion of the uniqueness and stability of the equilibrium point in the second Balloon Model formulation.

2.2 Parameter estimation

We model the measured data as the sum:

$$y = f(u,\theta) + e, \qquad e \sim \mathcal{N}(0,\Sigma), \tag{5}$$

where $f(u,\theta)$ is the output of the dynamical system with input u (stimulus-locked activity) and parameters θ, and e is a Gaussian noise with variance Σ. This does not necessarily ignore physiological noise: if the nonlinear effects of the model are small enough, and if we note e_n the cortical noise (ongoing activity) and e_m the measurement noise, we can make the following approximation:

$$\begin{aligned} y &= f(u+e_n,\theta) + e_m \\ &\approx f(u,\theta) + f(e_n,\theta) + e_m = f(u,\theta) + e. \end{aligned} \tag{6}$$

If e_n and e_m are supposed Gaussian, then resulting e is also a Gaussian colored noise. In fact, we assume in our study a white noise $\Sigma = \sigma^2 I$. The methods we present can be extended to a colored noise $\Sigma = \sigma^2\Sigma_0$, but it would require to estimate autocorrelations to define Σ_0 ; this will be discussed later.

Parameter estimation is obtained by maximizing the likelihood of the measured data y with respect to the parameters θ:

$$\begin{aligned} \hat{\theta} &= \text{argmax}_\theta\ p(y|\theta) \\ &= \text{argmin}_\theta\ -\log p(y|\theta) \\ &= \text{argmin}_\theta\ \frac{(f(u,\theta)-y)^T\Sigma^{-1}(f(u,\theta)-y)}{2}. \end{aligned}$$

Under the white noise assumption, it leads us to minimize the energy

$$\mathcal{E}(\theta) = (f(u,\theta)-y)^T(f(u,\theta)-y).$$

To minimize $\mathcal{E}(\theta)$, we use a Levenberg Marquardt algorithm (Press et al., 1992; Marquardt, 1963), implemented in the Matlab function 'lsqcurvefit'. The algorithm needs at

each iteration step the Jacobian of the system output with respect to the parameters $\frac{\partial f}{\partial \theta}$.

As predicted state and output $x(t,u,\theta)$ and $y(t,u,\theta)$ are defined by a differential system, it is possible to define $\frac{\partial x}{\partial \theta}(t,u,\theta)$ and $\frac{\partial y}{\partial \theta}(t,u,\theta)$ with a new differential system .

Let us go back to the initial system (1) indeed (for clarity, we use here x instead of $x(t,u,\theta)$):

$$\left\{ \begin{array}{rcl} \dot{x} & = & F(x,u(t),\theta) \\ y & = & G(x,\theta). \end{array} \right.$$

Differentiating both sides of these equations with respect to θ, we get the new system

$$\left\{ \begin{array}{rcl} \frac{\partial \dot{x}}{\partial \theta} & = & \frac{\partial F}{\partial x}(x,u(t),\theta)\frac{\partial x}{\partial \theta} + \frac{\partial F}{\partial \theta}(x,u(t),\theta) \\ \frac{\partial y}{\partial \theta} & = & \frac{\partial G}{\partial x}(x,\theta)\frac{\partial x}{\partial \theta} + \frac{\partial G}{\partial \theta}(x,\theta). \end{array} \right.$$

This system can be integrated numerically, using the initial conditions $\frac{\partial x}{\partial \theta}(t=0)=0$ (at time $t=0$, the state variables are at rest and do not depend upon θ). We therefore obtain the numerical values of $\frac{\partial f}{\partial \theta} = (\frac{\partial y}{\partial \theta}(t,u,\theta))_{0\leq t\leq T}$ without using any linearization of the system of equations.

2.3 Handling confounds effects

It is often usefull to ignore a certain set of timecourse components in real datasets, low frequencies for example.

Let us note C the matrix whose columns are the undesirable components. Then $p_C = I - C(C^TC)^{-1}C^T$ is the projector orthogonal to these confounds. Ignoring them consists in fitting $p_C f(u,\theta)$ to $p_C y$ instead of fitting $f(u,\theta)$ to y.

The new energy writes

$$\mathcal{E}(\theta) = (p_C(f(u,\theta)-y))^T\,(p_C(f(u,\theta)-y)) = (f(u,\theta)-y)^T\,p_C\,(f(u,\theta)-y),$$

and the gradient of $p_C f(u,\theta)$ against parameters for using Levenberg-Marquardt algorithm is $p_C\frac{\partial f}{\partial \theta}(u,\theta)$, with $\frac{\partial f}{\partial \theta}(u,\theta)$ computed as previously.

2.4 Sensitivity analysis

A serious obstacle to parameter estimation is the identifiability of the system, i.e. do we have enough information once we know the system input u and output y to determine the parameter values ? Is there a unique solution θ to the equation $y = f(u,\theta)$?

Actually this is hardly the case for the Balloon Model, because the effects of some parameters on the output do interfere with those of others. The extreme case would be for

example if the scale factor on the input (neural efficacy ϵ) and that on the output (V_0) were estimated on data with an input low enough to make the linear approximation of the model hold. Indeed, increasing the first could be compensated by decreasing the second by the same factor to produce exactly the same output. It would not be possible to estimate these 2 parameters, but only their product.

We want to investigate how much the system output is sensitive to changes in one parameter. Let us note θ_i this parameter, and θ_2 the rest of parameters ($\theta = \{\theta_i, \theta_2\}$). Also we note $J = \frac{\partial f}{\partial \theta}(u, \theta)$ the Jacobian of system output, J_i its i^{th} column and J_2 the matrix consisting of the remaining columns. For a small parameter change $d\theta$ we have

$$f(u, \theta + d\theta) = f(u, \theta) + Jd\theta = f(u, \theta) + J_i d\theta_i + J_2 d\theta_2.$$

For a small change $d\theta_i$ of θ_i, f varies by $J_i d\theta_i$; however, if J_i is not orthogonal to the other Jacobian components J_2, part of this variation can be compensated for by a change in the other parameters: $d\theta_2 = -J_2^+ J_i d\theta_i$, where $J_2^+ = (J_2^T J_2)^{-1} J_2^T$ denotes the pseudo-inverse of J_2. We then have:

$$\min_{d\theta_2} \|f(u, \theta + d\theta) - f(u, \theta)\| = \|(I - J_2 J_2^+) J_i d\theta_i\| = \pi_i |d\theta_i|,$$

with

$$\pi_i = \|(I - J_2 J_2^+) J_i\| = \sqrt{J_i^T (I - J_2 J_2^+) J_i}.$$

The bigger π_i, the more identifiable θ_i is. This also means that, for a given percentage x, if θ_i changes by less than $\pi_i^{-1} x \|f(u, \theta)\|$, one can adjust the other parameters θ_2 to make the model output vary by less than $x\%$. Given an input u and an initial parameter set θ_0, our sensitivity analysis consists in considering the sensitivity intervals $[\theta_{0i} - \pi_i^{-1} x \|f(u, \theta_0)\|,\ \theta_{0i} + \pi_i^{-1} x \|f(u, \theta_0)\|\,]$, with $x = 1$ to 5%. They are not confidence intervals for parameter estimation! Rather they indicate that the system output is very little sensitive to changes of θ_i inside these intervals.

Figure 3 shows such a sensitivity analysis with $x = 1\%$ for two different inputs (a single impulse and two successive impulses). The sensitivity intervals are represented in the left column of the figure. For each of the seven parameters (encoded with different colors) we fix it to one of the two bounds of its sensitivity interval compute the values of the other six parameters from $d\theta_2 = -J_2^+ J_i d\theta_i$ and plot the resulting time course. The figure clearly shows that very different parameter sets can result in very similar system outputs (table 2.4 shows the obtained parameter sets and the output variations). It also appears that the sensitivity depends on the input complexity: in the second case the parameters are more identifiable, because the effects of the different parameters can be more diverse and hence less correlated. For that reason, the experimental design we present later uses a large panel of ISI and stimulus duration to increase identifiability.

As a final word of caution, be aware that we have only discussed identifiability at a local scale, i.e. we only considered one minimum of the energy and approximated the shape of

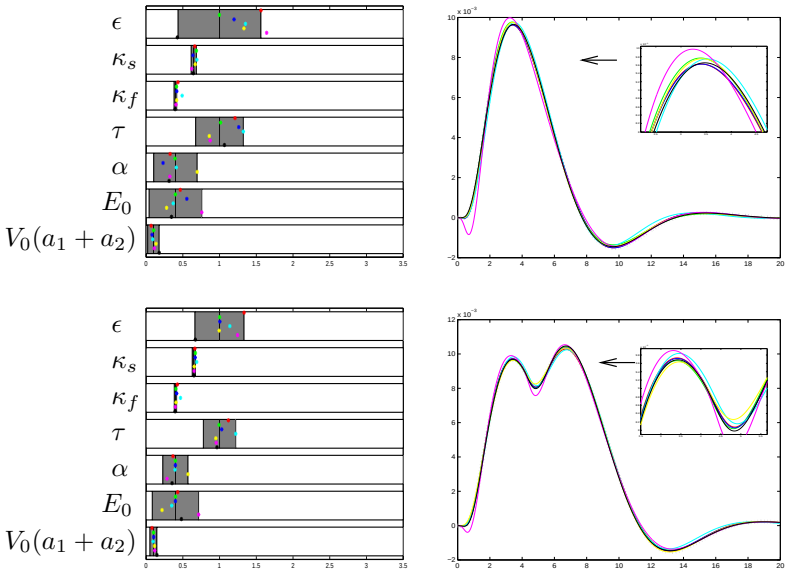

Figure 3: Sensitivity analysis for two different system inputs around a given θ_0. Top: response to an impulse. Bottom: response to 2 consecutive impulses. Left: 1% signal change sensitivity intervals - color stars show different parameter sets with one parameter constrained to be at the edge of its sensitivity interval (e.g., red corresponds to ϵ being fixed and the other parameters computed with $d\theta_2 = -J_2^+ J_i d\theta_i$). Right: output variations for these parameter sets compared to the output for the reference θ_0 (bold dashed line). Values of all parameters and percentages of signal changes are given in table 2.4.

possible model outputs locally with the tangent plane (see figure (4)). But since the shape can be more complicated, indetermination can be even worse than the one resulting from the above discussion.

2.5 A Bayesian formulation of sensitivity

The above results can be related to Bayesian inference principles. Let us note θ_0 the true parameter set and approximate f at θ_0 with its first order Taylor series expansion

$$f(u, \theta) = f(u, \theta_0) + J_{\theta_0}(\theta - \theta_0). \tag{7}$$

We recall the probability model for measured data (5):

$$y = f(u, \theta) + e, \qquad e \sim \mathcal{N}(0, \sigma^2 I).$$

parameter	ϵ	κ_s	κ_f	τ	α	E_0	b	% output change
θ_0	**1**	**.65**	**.4**	**1**	**.4**	**.4**	**.1**	
ϵ fixed	**1.56**	.66	.43	1.2	.33	.46	.07	1%
κ_s fixed	1	**.69**	.41	1.01	.39	.4	.1	2.1%
κ_f fixed	1.2	.64	**.42**	1.26	.23	.55	.08	1%
τ fixed	1.15	.7	.47	**1.32**	.34	.36	.09	3.5%
α fixed	1.25	.68	.42	0.86	**.7**	.28	.15	2%
E_0 fixed	1.52	.56	.37	0.89	.27	**.76**	.09	5.6%
a_1 fixed	0.43	.64	.4	1.07	.31	.35	**.18**	1.3%
parameter	ϵ	κ_s	κ_f	τ	α	E_0	b	% output change
θ_0	**1**	**.65**	**.4**	**1**	**.4**	**.4**	**.1**	
ϵ fixed	**1.33**	.67	.43	1.13	.37	.43	.08	1.29%
κ_s fixed	1	**.67**	.4	0.99	.4	.4	.1	1%
κ_f fixed	1	.67	**.42**	1.03	.4	.4	.1	0.8%
τ fixed	1.12	.69	.47	**1.22**	.39	.35	.1	2.4%
α fixed	0.97	.66	.41	0.95	**.57**	.22	.12	1.1%
E_0 fixed	1.24	.64	.39	0.93	.3	**.72**	.12	3.9%
a_1 fixed	0.67	.65	.39	0.97	.35	.48	**.15**	0.9%

Table 1: Parameter values for the sensitivity analysis in figure 3: quite different parameter sets can lead to very similar system outputs. The output variation is not exactly 1% when one parameter is fixed to the edge of its sensitivity interval, because these intervals were calculated using first order approximations with respect to parameters.

We do not use an a priori Gaussian distribution for the parameters as Friston (2002), but a non-informative degenerate uniform distribution $\theta \sim \mathcal{U}(\mathbb{R})$. See (Kershaw et al., 1999) for using such methods in fMRI data analysis and (Box and Tiao, 1992) for more theoretical details.
Then we can calculate the a posteriori distribution for parameter θ using Bayesian inference

$$p(\theta|y) \propto p(y|\theta)p(\theta),$$

which results in a Gaussian distribution with mean and variance (see Appendix for detail)

$$\left\{ \begin{array}{rcl} E(\theta) & = & \theta_0 + (J^T J)^{-1} J^T (y - f(\theta_0)) \\ V(\theta) & = & (J^T J)^{-1}. \end{array} \right. \tag{8}$$

The variance of parameter θ_i is the i^{th} diagonal term in $V(\theta)$: $(J^T J)^{-1}_{ii}$.

It can also be calculated by forming the marginal a posteriori distribution of parameter θ_i

$$p(\theta_i|y) = \int p(\theta_i, \theta_2|y) d\theta_2.$$

We show by integrating this formula (in appendix) that the a posteriori variance is equal to $\sigma^2\pi_i^{-2}$. This shows that the incertitude in θ_i estimation we established above is proportional to its a posteriori variance when there is a Gaussian white measure noise, and it also gives us the new formula

$$\pi_i^{-1} = \sqrt{(J^tJ)^{-1}_{ii}}.$$

However we have observed in simulation and data (not shown) that actual variance of θ_i is more than $\sigma^2\pi_i^{-2}$. This is probably due to the linearization and white noise assumption. This is the reason why we prefer incertitude intervals as defined previously to statistical confidence intervals.

2.6 Statistical test

We want to establish a statistical test to detect activations voxelwise. We use the presented bayesian framework (but since we do not know the true parameter set θ_0, we use our estimation $\hat{\theta}$ instead). In particular, we still use the linear approximation above (7), since it is too difficult to establish probabilities on parameters in the nonlinear case. This actually means that we approximate the manifold of all possible model outputs by its tangent vectoriel subspace at point $f(\hat{\theta})$ (figure 4A).

Friston (2002) proposed in a similar bayesian framework an estimation detection based on the marginal distribution of neural efficiency parameter ϵ. However, it can happen that this marginal distribution is pretty flat and says '$\epsilon = 0$ is plausible', not because there is no detected activation in the data, but only because ϵ is poorly identifiable, due to interactions with other parameters (see figure 4D-E). For that reason, we would prefer a test based on the whole set of parameters, or equivalently on the model fit to data, by calculating how probable it is that $f(\theta) = 0$ (i.e. $\theta = \hat{\theta} - J^+f(\hat{\theta})$).

We cannot establish a statistical test directly from θ Gaussian a posteriori distribution calculated above (8), since variance parameter σ^2 is unknown. Again, we use a non-informative degenerate a priori distribution for σ^2, $p(\sigma^2) \sim \frac{1}{\sigma^2}$. Then we can derive (see Appendix) a new a posteriori distribution for θ, that follows a Student law with $\nu = (n-p)$ degrees of freedom (n and p being the numbers of measure point and parameters, respectively), and with mean and variance

$$\left\{ \begin{array}{rcl} E(\theta) & = & \hat{\theta} \\ V(\theta) & = & \hat{\sigma}^2\,(J^TJ)^{-1}, \end{array} \right. \tag{9}$$

where

$$\hat{\sigma}^2 \quad = \quad \tfrac{1}{n-p}\,(y - f(\hat{\theta}))^T(y - f(\hat{\theta})).$$

If we used a colored version of noise $\Sigma = \sigma^2\Sigma_0$, a similar distribution could be obtained, but the number of degrees of freedom ν would depend on the rank of Σ_0 (Friston and Worsley, 1995).

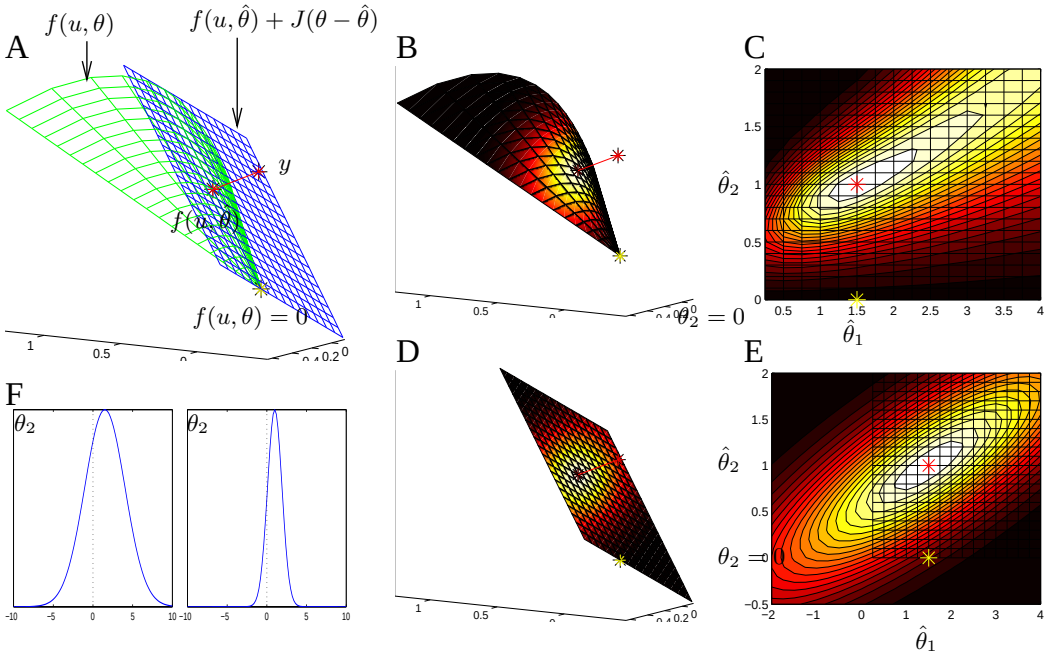

Figure 4: Geometrical interpretation of nonlinear activation detection using the manifold of all possible model outputs inside the space of time courses. Surface was obtained by varying 2 Balloon Model parameters $\theta = \tau, V_0$ and representing the point $f(u,\theta)$ by its values at 3 selected measure points (u=impulse). (A) $f(u,\theta)$ lives in a manifold; maximum likelihood estimate $\hat{\theta}$ is such that point $f(u,\hat{\theta})$ is the point on the manifold the closest to measured data y; the manifold is approximated by its tangent surface at point $f(u,\hat{\theta})$. (B) a posteriori distribution of θ on the manifold; no activation hypothesis $f(u,\theta) = 0$ (yellow point) is unprobable in this example. (C) same distribution represented in the θ space. (D) a posteriori distribution of θ when approximating the manifold by its tangent surface. (E) same distribution represented in the θ space; yellow point (no activation hypothesis) remains unprobable. (F) marginal a posteriori distributions of θ_1 and θ_2 under the approximation: they are pretty flat because of correlations between the 2 parameters; hence, $\theta_1 = 0$ and $\theta_2 = 0$ are probable in these distributions: using the marginal a posteriori distribution could lead to not enough activation detections.

We note that $\hat{\sigma}^2$ is a non-biased estimator of variance σ^2. Since θ has a Student distribution $t_\nu(\hat{\theta}, V(\theta))$, $A(\theta) = \frac{1}{p}(\theta-\hat{\theta})^T V(\theta)^{-1}(\theta-\hat{\theta})$ has a Fisher distribution $\mathcal{F}(p,\nu)$ (Kershaw et al., 1999):

$$\begin{aligned} A(\theta) &= \tfrac{1}{p}(\theta-\hat{\theta})^T \tfrac{J^TJ}{\hat{\sigma}^2}(\theta-\hat{\theta}) \\ &= \tfrac{(J\theta-J\hat{\theta})^T(J\theta-J\hat{\theta})}{p\hat{\sigma}^2} \\ &= \tfrac{\|f(\theta)-f(\hat{\theta})\|^2}{p\hat{\sigma}^2}. \end{aligned}$$

We can test now how plausible it is that $f(\theta)=0$ by calculating the following statistic:

$$F = A(\hat{\theta}-J^+f(\hat{\theta})) = \frac{\|f(\hat{\theta})\|^2}{p\,\hat{\sigma}^2}$$

$$F = \frac{n-p}{p}\frac{f(\hat{\theta})^Tf(\hat{\theta})}{(y-f(\hat{\theta}))^T(y-f(\hat{\theta}))}. \tag{10}$$

F is a sort of signal to noise ratio. If $f_{p,\nu}$ is the Fisher cumulative distribution function ($f_{p,\nu}(z)=P(F<z)$), the test will consist in calculating the p-value $1-f_{p,\nu}(F)$ of this statistic and declare the voxel activated if this p-value is less than some probability.

This test can be adapted if we have ignored some confounds described by the projection matrix p_C during parameter estimation. It necessits to replace f by p_Cf and J by p_CJ in the above formulas. The new statistic is then

$$F = \frac{n-c-p}{p}\frac{f(\hat{\theta})^Tp_Cf(\hat{\theta})}{(y-f(\hat{\theta}))^Tp_C(y-f(\hat{\theta}))},$$

and has a Fisher distribution with c degrees of freedom less (c being the number of confounds), $\mathcal{F}(p,n-c-p)$.

3 Results

We conducted fMRI experiments in order to question the validity of the Balloon model and the estimation and statistical tools described in the previous sections.

3.1 Experimental data

The stimulus consisted of a full screen binocular flashing checkerboard (12 Hz). A red cross fixation point was used throughout the experiment. Resting condition consisted in a grey screen with the fixation cross. Eight volunteers were used for this study (6 males and 2 females, from 19 to 25 years old, with no vision problem). Brain anatomy and fMRI images were acquired in the La Timone Hospital, Marseille, France, on a 3T scanner with surface coil. The functional scans consisted in 11 coronal occipital slices, each voxel being 2 x 2 x 2 mm, with interscan interval TR = 825ms.

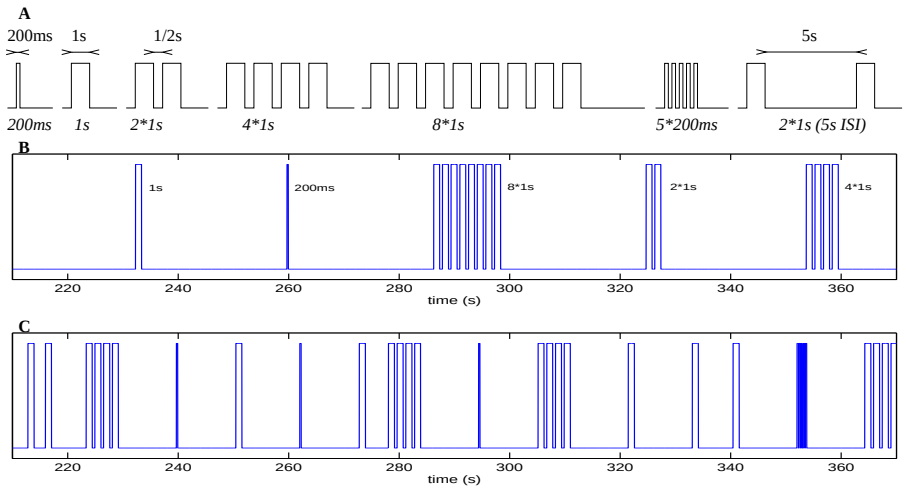

Figure 5: Experimental design. (A) 7 stimulation designs (B) first paradigm (stimulation are separated by 25s rest periods). (C) second paradigm (stimulation are separated by 10s rest periods or less)

In order to test the validity of the Balloon Model, we tried to make the BOLD response as nonlinear as possible with respect to the stimulus. Moreover, we wanted these nonlinearities to be due mostly to vascular effects, and minimize neuronal nonlinearities such as habituation.

For this purpose, we varied the stimulus durations by using repetitions of a 1s checkerboard presentation. We used from 1 up to 8 successive presentations separated by half a second. We prefered such block repetitions to prolonged stimulation to prevent as much as possible neural habituation. Indeed, if there is a strong transient activity at the start of the stimulation, there is more chance that this transient be replicated at each repetition, whereas it would only appear once in the case of a longer stimulation. We also used one 200ms presentation, 5 successive 200ms presentations spaced by 200ms blank, and a sequence of two 1s blocks spaced by 5s (figure 5A). These seven designs provide complementary information that will be discussed below.

We combined the stimuli in two different paradigms. The first one consisted of two 15 minutes runs, each one containing 5 repetitions in random order of 6 different designs (200ms - 1s - 2*1s - 4*1s - 8*1s - 5*200ms (first run) or 2*1s with 5s ISI (second run)), followed by a 25 seconds return to baseline (figure 5B).

The second paradigm consisted of one 10 minutes run containing the 7 designs described above, but separated only by 10s or less. It allows to compare the results when responses are overlapping (figure 5C).

3.2 Data analysis

Five subjects endured the two paradigms explained above (the first and third runs were dedicated to the first paradigm, and the middle one to the second). In a preliminary experiment, one additional subject endured the first paradigm but the data had to be discarded due to the weakness of the response, and two endured the second paradigm.

The functional images were corrected for time delays. One subject needed to be motion-corrected.

For both paradigms, a first SPM study was done, using the stimulus convolved with three basis functions (HRF, HRF time derivative and HRF dispersion).

We extracted a number of mean time courses, focusing on V1 since it is likely the region where neurons respond the most linearly to visual stimulation. The calcarin sulci were located on the anatomical images, and about 20 among the most activated voxels were selected there for each subject and for the left and right cortices. The time course was the first PCA eigenvector of the signals at the selected voxels for all volunteers.

We used these regions time courses to compare and discuss the different models.

Then model parameters estimation was run on each voxel raw time course in a masked brain in order to apply the statistical test.

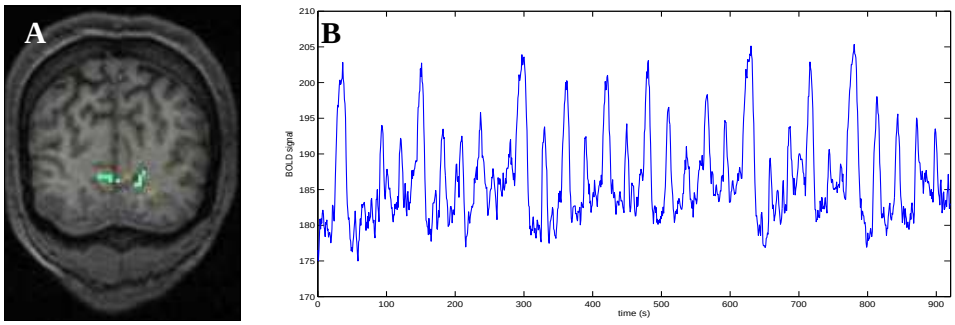

Figure 6: Extracted data on 40 voxels in one subject primary visual cortex. (A) Voxels selected (cyan) according to SPM F-test (light yellow) and anatomical information. (B) First eigenvector of extracted time courses in the left hemisphere.

3.3 Qualitative description of the estimated responses

The responses to each of the seven designs in the first paradigm were time-locked averaged in each region, and a global mean over all subjects calculated as well (figure 7B). We did not apply any high-pass filter, to preserve possible physiological low-frequency components. The baseline signal was estimated by averaging the signal over the 4s before each stimulation.

First we can observe an ascending trend in the estimation of the responses to short duration stimuli, and the signal level before stimulus presentation seems to increase with

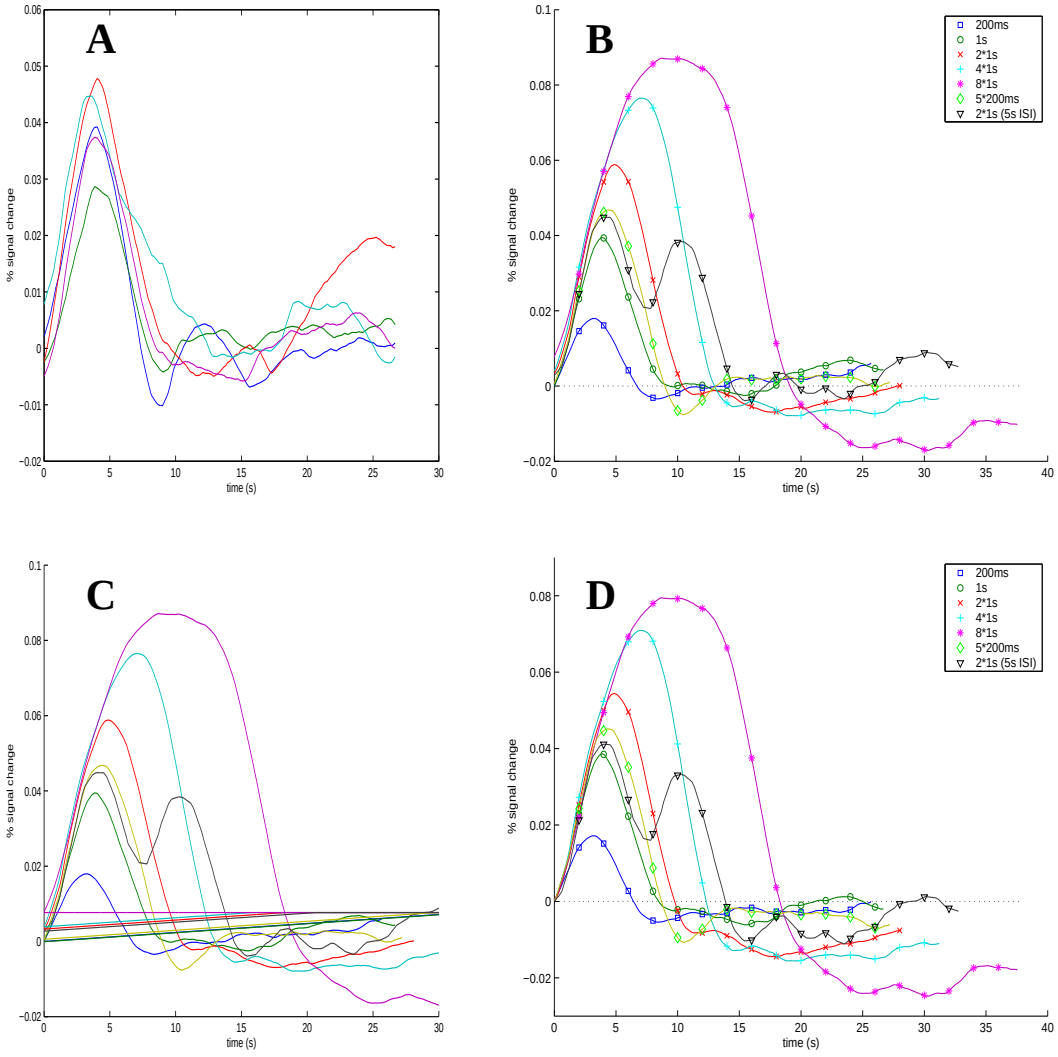

Figure 7: Estimated responses to the 7 designs - mean on 5 subjects. (A) inter-subjects variability (response to the 1s stimulation). (B) mean responses. (C) estimation of an overlapping return to baseline from previous responses. (D) corrected responses.

stimulus duration! This is probably because the responses to long stimulations last more than 25s after stimulation ends, so that the responses to short stimulations are meddled with return to baseline of the previous ones. We tried to correct for this defect by estimating and removing a linear return to baseline (figure 7C,D). Nevertheless, subsequent remarks are robust to this trend removing: the analysis in figure 8 shows the same qualitative behaviors whether the trend has been subtracted or not.

Nonlinearities are clearly present in the short durations range: responses to 1s or 5*200ms stimulations are much smaller than 5 times the response to the 200ms stimulation. We call this a sub-linearity effect below. Moreover, the response to the 1s stimulation is itself smaller than that to the 5*200ms stimulation.

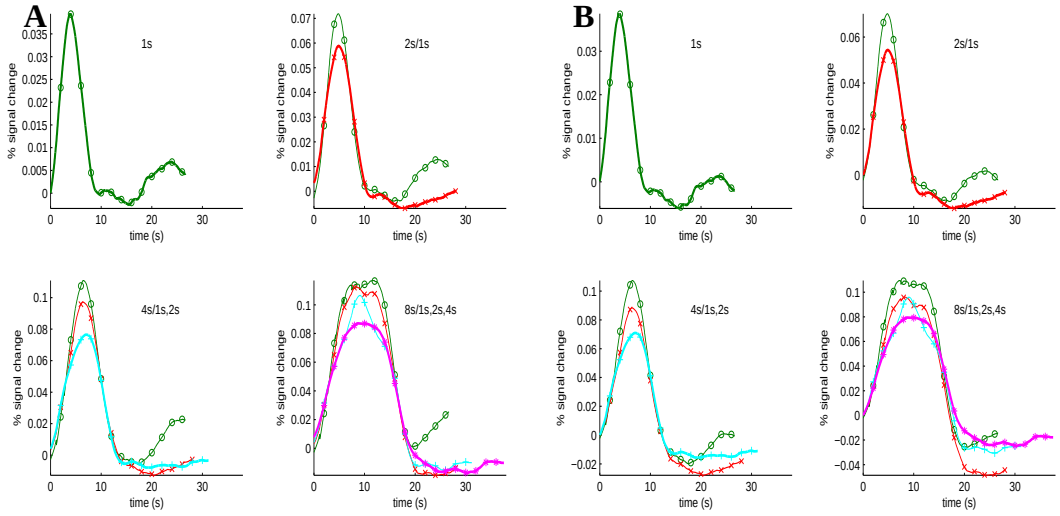

Figure 8: Fit between responses to long durations and their prediction by shorter durations responses. (A) calculations from estimated reponses in figure 7B. (B) from corrected responses in 7D.

For longer stimulations, we study linearity as shown in figure 8: the response to $k * n$ repetitions is predicted by the sum of k shifted responses to n repetitions. We observe that the response to the 1s stimulus overpredicts that to the 2*1s stimulus (see figures 8A and B, upper righthand corners) , which itself overpredicts that to the 4*1s stimulus (see figures 8A and B, lower lefthand corners). It is not clear whether the response to the 4*1s stimulus overpredicts that to the 8*1s stimulus, but anyway, the shapes are significantly different (see figures 8A and B, lower righthand corners).

These results are coherent with other studies (Boynton et al., 1996; Dale and M., 1997; Glover, 1999): when comparing positive responses, the linear assumption for the BOLD

response is acceptable for stimulus durations > 4 seconds, and does not hold for durations < 2 seconds.

Last, there seem to be nonlinear effects in the postimulus undershoots too: the undershoots after longer stimulations appear to last longer than what would be predicted from shorter stimulations (even more than 25s, the time we chose to separate our presentations). This aspect will be tackled in the discussion on hemodynamic models.

3.4 Fitting models to mean responses

We fit different models to the estimated responses by minimizing the least square errors over the seven juxtaposed curves (figure 9).

The results obtained with a linear model consisting of three regressors (the stimulus convolved with a default HRF, the HRF time derivative and the HRF time dispersion, as defined in SPM) are shown in part A of the figure. Sub-linearities are clearly present: the best fit is to the 2s stimulation reponse, but the response peak after the 200ms stimulation is underestimated while those after longer stimulations are overpredicted. Moreover poststimulus undershoots are not fitted well.

The first physiological model we fit to the data (part B of the figure) is the original Balloon Model given by equation (2). It appears to effectively better capture some nonlinearities, thus pointing to a vascular explanation. However, it does not account for short time range nonlinearities (200ms and 5*200ms stimulations). And the poststimulus undershoot is not captured well: on the contrary we can observe oscillations that result from the damped oscillator that models the flow response to neural activity

Adding the two viscosity term ($\tau^+_{visc}/\tau^-_{visc}$) in the volume dynamic results in a more prolonged poststimulus undershoot (part C of the figure). Moreover, we found that there was almost no loss of quality in the fit when values of the Grub parameter α and the extraction at rest E_0 were not estimated, but fixed to some physiologically plausible value (see the next section on sensitivity analysis).

The next estimations use linear convolutions to model the flow and metabolic responses as in equation (3). Since the convolution kernels are positive functions, flow and metabolic responses to a positive neural activity will remain positive, which is probably more realistic than a response oscillating around baseline as previously. We tried an estimation with no viscosity term in the volume dynamics but the poststimulus undershoot was not well predicted again (part D of the figure).

Adding back the two viscosity terms (part E of the figure) results in an estimation comparable to the one shown in part C. Again, we found that we could fix the value of α and impose a coupling between the flow and metabolism responses ($\tau_f = \tau_m$ and $n = (f_1 - 1)/(m_1 - 1) = 2.5$) without significant loss in the quality of the fit. Figure 10 shows a comparison of the flows computed in the two models, Friston's damped oscillator (part A) and Buxton's convolution with a Gamma-variate function (part B).

We tried to estimate the delay between metabolic and flow response (parameter δ_t in equations (3)), but the value came up as zero and resulted in exactly the same plot. A

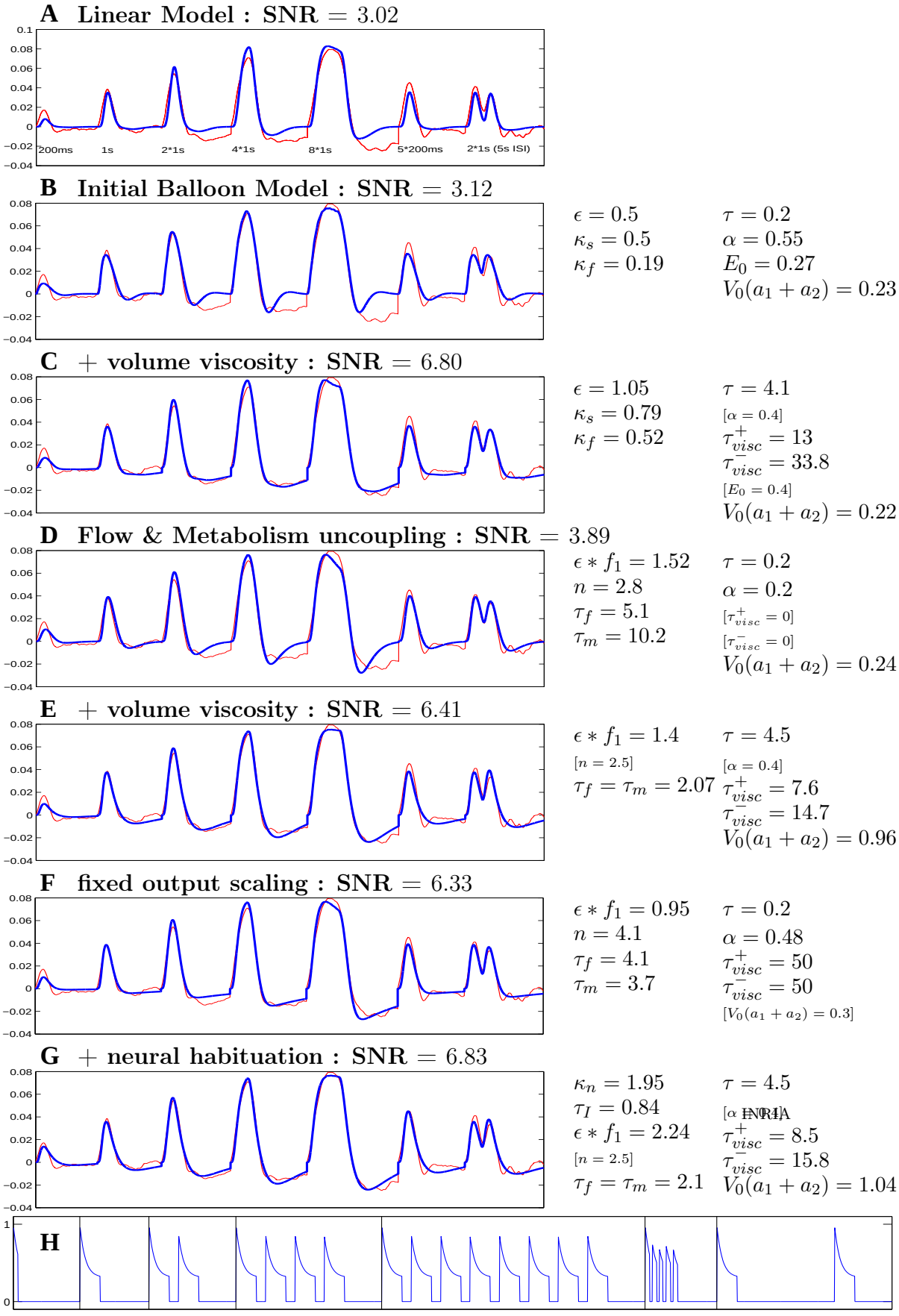

Figure 9: (A)-(G) Fit of different models to the mean responses (we use the corrected

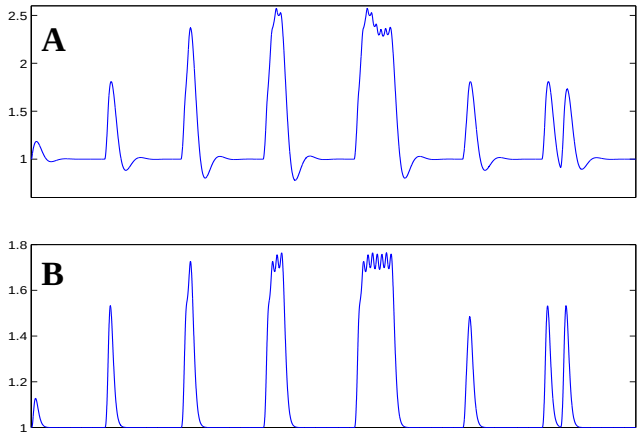

Figure 10: Estimated flow time courses corresponding to the estimations in figure 9C and 9E: they compare Friston's model with a damped oscillator and Buxton's convolution with a Gamma-variate function.

positive value would actually cause an initial dip and a short poststimulus overshoot that we did not observe in our data.

As the estimated values for the output scaling $V_0(a_1+a_2)$ were a bit larger than expected, we tried a new estimation where we imposed the more physiologically plausible values $V_0 = 0.03$ and $a_1 + a_2 = 10$ (part F of the figure), but it resulted in less realistic values for the other parameters.

The last refinement we added to the model was the simple neural habituation proposed by Buxton et al. (2004). Including the parameters κ_n and τ_n in the estimation, i.e. allowing neural habituation, appears to be the only way to correctly predict the 200ms and 5*200ms stimulations responses (part G of the figure). Panel (H) shows the corresponding estimated neural activity.

For the remaining of this paper we focus on the second Balloon Model proposed by Buxton et al. (2004), except that we do not include a delay between metabolic and flow responses (parameter δ_t). The following section discusses the system identifiability and how many parameters we can try to estimate together.

3.5 Sensitivity analysis of the mean responses

Several estimations have been made with the proposed model with different choices of fixed / estimated parameters. For each such choice sensitivity intervals with $x = 2\%$ have been established as described in section 2.4. The results are shown in figure 11.

We recall the definition of the sensitivity intervals: "for every value v of the ith parameter in this interval, under a linear approximation of the system output wrt. the parameters,

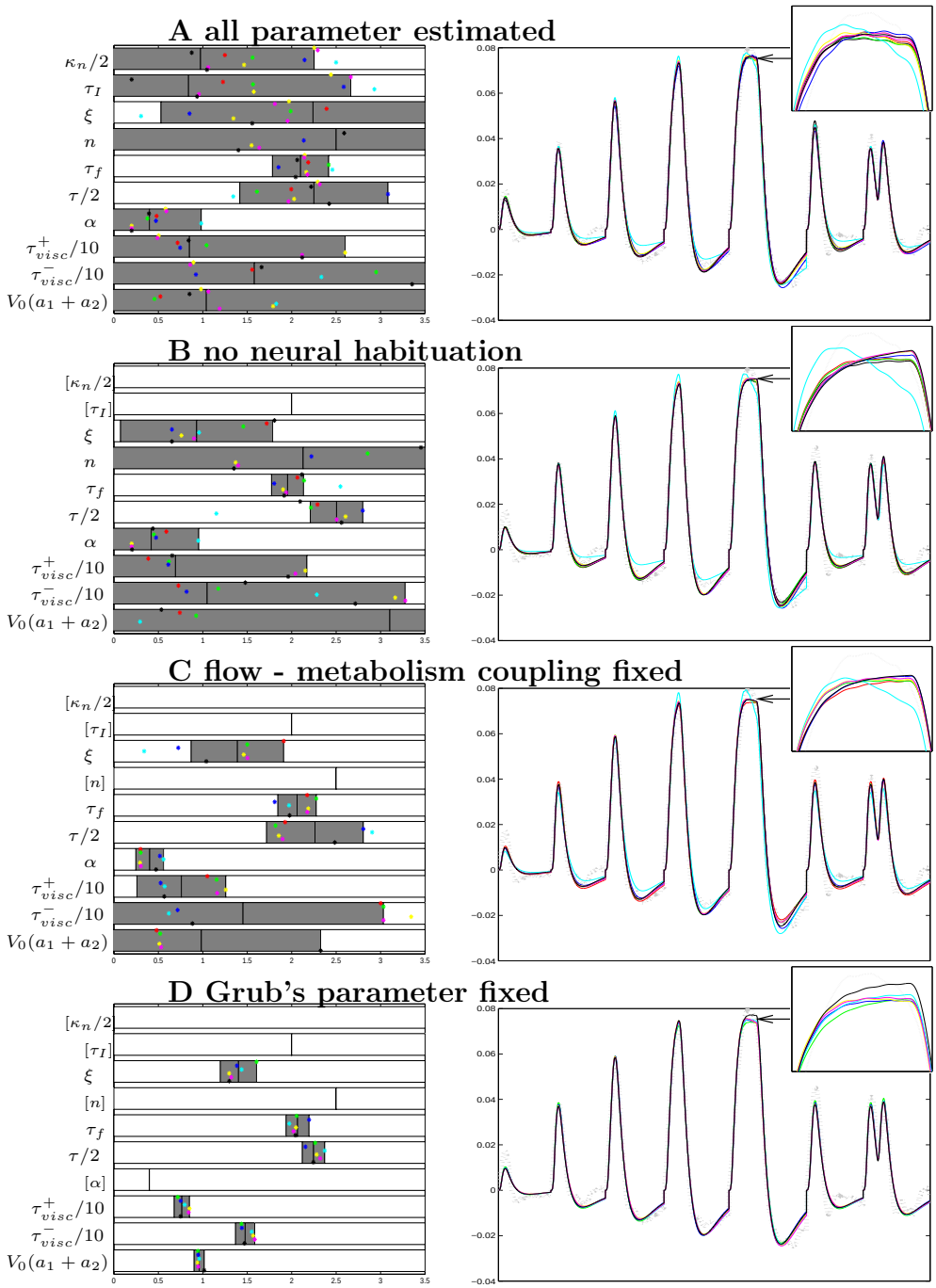

Figure 11: Sensitivity analysis for parameter estimation on the estimated mean responses (similar to figure 3). Left column: parameters sensitivity intervals for 2% output variations; for any value of the corresponding parameter in these intervals it is possible (in the linear approximation case) to find a set of parameters such that the system output is modified by less than 2%; such parameter sets are represented with color stars. Right: measured mean (bold line) and outputs corresponding to these sets. From top to bottom, graphics show different choices for which parameters are estimated or fixed to a canonical physiological value (inside brackets). As expected, the less parameters are estimated, the more identifiable they are.

there exists a parameter set θ' verifying $\theta'_i = v$ such that $\|y(\theta') - y(\theta)\|$ is less that 2% of $\|y(\theta)\|$".

To visualize this assertion, for every estimated parameter, we give an example of a new estimation where that parameter value is fixed to that of the edge of the confidence interval (color stars in the figure) and plot the new time course obtained.

It appears that we cannot estimate all parameters correctly from the fMRI data (top row in the figure) since many sensitivity intervals are very large. To increase the system identifiability (i.e. reduce the sensitivity intervals), we must reduce the number of estimated parameters. Panel 11D (bottom row in the figure) shows the choice we use for the rest of our study: no neural habituation is estimated, the parameters τ_f and τ_m are constrained to be equal, and n and α are fixed, while the six parameters ξ, τ_f, τ, τ^+_{visc}, τ^-_{visc} and $b = V_0(a_1+a_2)$ are estimated from the data.

3.6 Voxel by voxel estimation and activation maps

We estimated the parameters of the model at every voxel for the three runs of each subject. As before with the mean responses, we compared the resulting fit with the one of the linear model with three regressors (canonical HRF + time derivative + time dispersion). Results on the first paradigm are shown for one subject in figure 12: the signal to noise ratio (SNR) is on average 22% stronger for the Balloon Model (upper lefthand corner of the figure). The figure shows details of the predictions of the two models for three voxels (activated, questionable and non-activated). The Balloon Model improvement is mostly in the poststimulus undershoot prediction. The nonlinear effects that the Balloon Model could account for on the estimated means (see figure 9) were not found in this case.

Besides, it appears that parameters unidentifiability is even stronger on voxels time courses than on the means used above: figure 13, to be compared with the last row of figure 11, shows the fits and the sensitivity analysis around estimated parameters, on one activated voxel, for the first and second paradigms.

However, despite these estimation uncertainties due to the correlation between parameters effects, activation detection is still possible with the statistical test based on a signal to noise criterion (see section 2.6). Figure 14 shows the resulting activation detection. The results are comparable to those based on a linear model, which is not surprising since the fits to data are similar (figure 12).

We had to choose a very small p-value. Actually, the computation of the p-values is biased, in the sense that the whiteness hypothesis for the noise does not hold in our experiment with close acquisitions (TR = 825ms), and the number of degrees of freedom should be modified accordingly (Friston and Worsley, 1995).

Despite the fact that the SNR are slightly better for the Balloon Model (figure 12A), the p-values are slightly stronger: this is because the latter has more free parameters (7, versus 3 for the linear model), and this makes the F-statistics smaller (see equation (10)) .

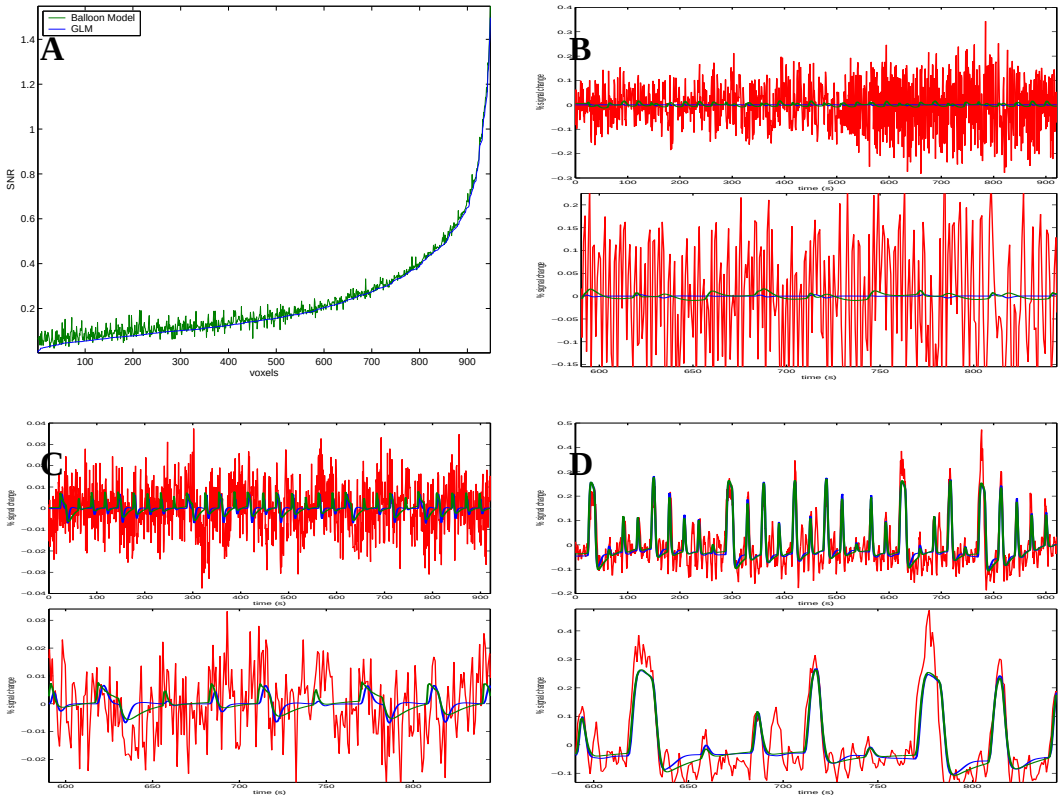

Figure 12: Model to data fits comparison between the GLM and the Balloon Model. (A): plot of the signal to noise ratios for the GLM and the Balloon model at all voxels ($SNR = \|y_{model}\|/\|y - y_{model}\|$ - voxels are sorted by their GLM SNR). (B),(C),(D): details of the fit for three voxels corresponding to the worst, medium and best SNR (red: measured signal, blue: predicted signal by the GLM, green: predicted signal by the Balloon Model).

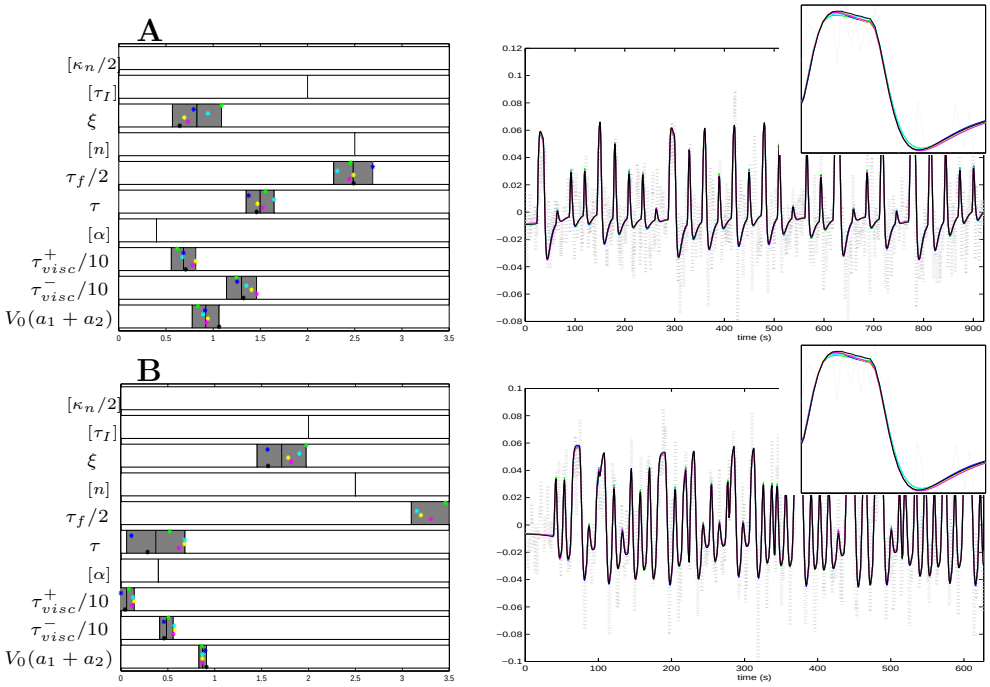

Figure 13: 1% output variations sensitivity analysis around parameters estimated on one voxel time course. (A) First paradigm. (B) Second paradigm.

4 Discussion

4.1 Hemodynamic Models

The results of the tests we performed on the mean responses with different models allow us to discuss a few physiological points about hemodynamics.

We were particularly interested in the nonlinearities of the BOLD response, and to what extent are they due to vascular effects or are already present in the neural response.

As we chose a stimulus pattern which we believe does elicit little neural nonlinearities (repetitions of 1 second stimulation blocks separated by 0.5 s), the nonlinearities between the 1, 2, 4 and 8 repetitions responses should have vascular causes (figure 8). The Balloon Model is able to partly capture these nonlinearities, whereas linear models are not (9B-A).

Nonlinearities can be observed in the response peaks. A first explanation for these nonlinearities can be found in saturation effects: vessels volume is limited and deoxygenation cannot be less than zero. The maximum of these saturations should arise when comparing responses to 2s and 4s stimulations (see figure 1 for Balloon model simulations): for longer

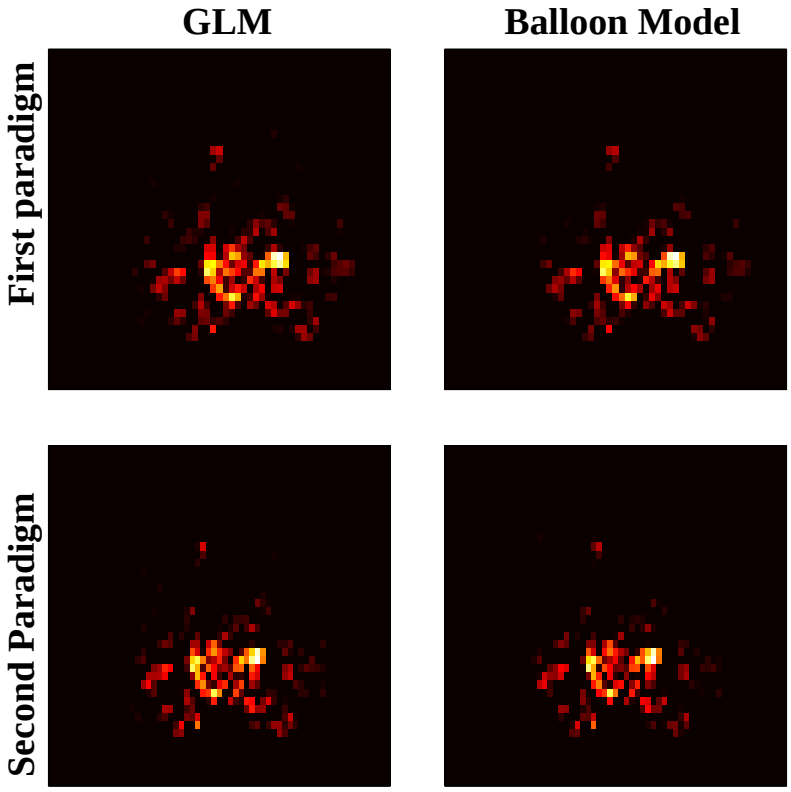

Figure 14: Activation maps with p-value 0.001.

durations, the predictions obtained by adding shifted responses is not stronger than the responses themselves since they do not overlap much; and for shorter durations, saturation is not reached yet. These saturation effects are not very strong however (the errors in peak amplitudes when fitting the data with a linear model (figure 9A, responses to 1s, 4s, 8s stimulations) did not represent more than 10% of the total responses amplitudes in our data).

Meanwhile, the data shows much stronger nonlinearities in the short time range (the response to the 1s stimulation is only twice stronger than that to the 200ms stimulation). These nonlinearities could be explained by habituation effects at the neural scale (figure 9G). However, it is not sure that neurons in V1 exhibit such an habituation. The nonlinearities could also be caused by the flow response process. Glover et al. suggested that the flow response lasts a minimum amount of time independently of the stimulus duration (Glover, 1999). Deciding between the two hypotheses would require measurements with other modal-

ities than fMRI to give direct informations about neural activity and flow response. It would be of great importance to know exactly whether fMRI can provide informations about neural nonlinearities or whether it is the linearity assumption for the flow response that is strongly violated.

In fact, flow modelization still seems to be a wide open field for investigation. We saw in this study that we can obtain comparable fits to the fMRI signal (figure 9C,E) with very different flow time courses (figure 10). We prefered convolutions with gamma-variate functions (Buxton et al., 2004) since the damped oscillator proposal by Friston et al. (2000) can produce non-physiological flow oscillations (figures 2, 10).

Nonlinearities can be found also in the poststimulus undershoot. In the estimated means (figure 7), though this is under standard deviation level, the undershoot magnitude seems to be approximately proportional to stimulus length. However, the return to baseline slope does not increase as fast. Thus, the return to baseline is much longer after stronger undershoots. This causes the drift we observed, and this can be seen even in the raw data (figure 6B: drifts due to gradients artefact have probably cancelled out, and the remaining ascending drifts look much correlated to preceding responses to long stimulations).

The introduction of hysteresis viscosity terms in the volume dynamics allows a better fit to data. However, the explanation for these prolonged undershoot is not necessarily related to volume. Aubert and Costalat (2002) proposed a decline of a tissue oxygen buffer during sustained or repeated activations to explain these long-lasting decreases of the BOLD signal.

4.2 Statistics: noise model and nonlinearities with respect to parameters

We have developed a framework where measured data can be decomposed as a sum of model prediction $f(u, \theta)$ and noise. This framework is somewhat a simplification, since a more complete model should include noise in the input and in the evolution equations, according to the theory of nonlinear stochastic systems. Riera et al. (2004) worked in this more complex framework, estimations being based on Kalman filtering; this approach can have nice applications, e.g. in the fusion with other modalities Riera et al. (2005). However, our framework where we only deal with measurement noise does take into account input and evolution noises (equation (6)). In fact, the approximation would be exact if f was linear with respect to u, and since the effective nonlinearities are not too strong (simulations in figure 1 and data in figure 8), it remains appropriate.

We have worked on parameter estimation, and developed statistical tests to deal with our nonlinear setting. In fact, we assumed a white noise, whereas the actual noise is colored, part of it coming from neural ongoing activity that has been smoothed by the hemodynamic reponse, another part coming from modelling errors. This white noise assumption is not very important for the definition of a least-square estimator of the parameters, it becomes more of a problem for establishing a posteriori probabilities for the parameters: a posteriori variances can be underestimated. Indeed, noise cancels out faster when all time instants are

independent than when they are correlated. If noise is supposed white, statistics will be more confident in the parameter estimation than they should if it were actually colored. Another consequence is the fact that activation detection can be too permissive. It is possible to adapt our methods to colored noise, but it would necessit to estimate the noise autocorrelation and to change the number of degrees of freedom of the statistical test. This could be the subject of a further study.

However, the major problem comes from the nonlinearity of the model with respect to the parameters. Indeed, it is very difficult to assign probabilities to the space of all possible model outputs (figure 4). For estimating parameters, we used a gradient descent algorithm to find, according to an iterative principle, a minimum of least square energy in that space (that can unfortunately be only a local minimum). This algorithm has some nice features: the nonlinearity of f with respect to u is taken into account since we calculate exactly the derivative $\frac{\partial f}{\partial \theta}$ through the integration of a new differential system, and eventual confound effect in the data can be ignored. But setting probabilities is much more difficult, and we had to linearize f with respect to θ at the estimated point. The consequence is that the resulting probabilities are correct only locally. The statistical test we proposed relies on this linearization. In order to take into account the whole space of possible outputs we should somehow scan it completely, which is the idea behind the Markov Chain Monte Carlo algorithms, a much more time consuming way to go (Jacobsen et al., 2005).

4.3 Conclusion: using dynamical systems for fMRI analysis

Our study had two main purposes: testing different models of the hemodynamic response, and develop methods to include them in future fMRI analysis.

First, we confirmed that the Balloon Model was adequate for the fMRI BOLD signal, and took into account some nonlinear effects (saturation) that empirical linear models did not. On one hand, some other effects could necessit further modelling; in particular focusing on the flow response and its possible nonlinearities would be useful if we wanted to tackle neural nonlinearities in fMRI. On the other hand, existing models already offer too much flexibility, and the fMRI signal alone does not allow in general to estimate all parameters, and we would need rather strong physiological priors, or mathematical reformulation of the models, to reduce their number.

Secondly, we have shown that physiological models expressed as dynamical systems can be used in fMRI analysis to fit predicted responses to the data, instead of linear regression with empirical basis functions. Exact values of the parameters cannot always be obtained, their effects on the signal output being correlated, and it is possible to quantify this undetermination. But activation maps can be established, relying upon an F-test. Activation results are comparable to those obtained with linear models.

Nonlinear models are still expensive in terms of computation time: our algorithm was implemented in C++ and took roughly 10 seconds to perform parameter estimation at one voxel. We think that this is well worth it because of the immense advantage of being well

grounded in Physiology. As they will become more precise they will allow new investigations in fMRI, e.g., neural nonlinearities, and the fusion with other modalities, e.g., EEG.

5 Appendix: a posteriori probabilities in the Bayesian framework

We develop here some computation mentioned in the main part of the paper.

We recall the linearization of the model output with respect to the parameters (7) and the noise model (5)

$$y = f(u, \theta_0) + J(\theta - \theta_0) + e, \qquad e \sim \mathcal{N}(0, \sigma^2 I).$$

We note $\tilde{y} = y - f(u, \theta_0)$. θ a posteriori distribution can be obtained with the Bayesian rule:

$$\begin{aligned} p(\theta|y) &= \frac{p(y|\theta)p(\theta)}{p(y)} \\ &\propto p(y|\theta) \\ &\propto e^{-\frac{1}{2\sigma^2}(\tilde{y}-J(\theta-\theta_0))^T(\tilde{y}-J(\theta-\theta_0))} \\ &\propto e^{-\frac{1}{2\sigma^2}(\theta-\theta_0-(J^TJ)^{-1}J^T\tilde{y})^T(J^TJ)(\theta-\theta_0-(J^TJ)^{-1}J^T\tilde{y})} \\ &\propto e^{-\frac{1}{2\sigma^2}(\theta-E(\theta))^TV(\theta)^{-1}(\theta-E(\theta))}, \end{aligned}$$

where we recognize the mean and variance assessed in (8).

The marginal distribution of the parameter θ_i can be obtained by integrating with respect to the other parameters:

$$\begin{aligned} p(\theta_i|y) &= \int p(\theta_i, \theta_2|y)d\theta_2 \\ &\propto \int e^{-\frac{1}{2\sigma^2}(\theta-E(\theta))^TJ^TJ(\theta-E(\theta))}d\theta_2 \\ &\propto \int e^{-\frac{1}{2\sigma^2}(J_i(\theta_i-E(\theta)_i)+J_2(\theta_2-E(\theta)_2))^T(J_i(\theta_i-E(\theta)_i)+J_2(\theta_2-E(\theta)_2)}d\theta_2 \\ &\propto \int e^{-\frac{1}{2\sigma^2}(\theta_2-E(\theta)_2+(J_2^TJ_2)^{-1}J_2^TJ_i(\theta_i-E(\theta)_i))^TJ_2^TJ_2(\theta_2-E(\theta)_2+(J_2^TJ_2)^{-1}J_2^TJ_i(\theta_i-E(\theta)_i))} \\ &\quad e^{-\frac{1}{2\sigma^2}(\theta_i-E(\theta)_i)^TJ_i^T(I-J_2(J_2^TJ_2)^{-1}J_2^T)J_i(\theta_i-E(\theta)_i)}d\theta_2 \\ &\propto e^{-\frac{1}{2\sigma^2}(\theta_i-E(\theta)_i)^TJ_i^T(I-J_2J_2^+)J_i(\theta_i-E(\theta)_i)}. \end{aligned}$$

We recognize an a posteriori variance for the parameter θ_i, $(J_i^T(I - J_2J_2^+)J_i)^{-1}$.

If we use the non-informative degenerate probability $\sigma^2 \sim \frac{1}{\sigma^2}$ (Kershaw et al., 1999), we get a new a posteriori distribution for θ by integrating with respect to σ^2:

$$\begin{aligned} p(\theta|y) &= \int p(\theta, \sigma^2|y)d\sigma^2 \\ &\propto \int p(y|\theta, \sigma^2)p(\theta)p(\sigma^2)d\sigma^2 \\ &\propto \int \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}+1}}e^{-\frac{1}{2\sigma^2}(\tilde{y}-J(\theta-\theta_0))^T(\tilde{y}-J(\theta-\theta_0))}d\sigma^2. \end{aligned}$$

We recognize an Inverse-Gamma distribution for σ^2 with parameters $a = \frac{n}{2}$ and $b = \frac{1}{2}(\tilde{y} - J(\theta - \theta_0))^T(\tilde{y} - J(\theta - \theta_0))$ (see (Gelman et al., 1998), Annex A, for probability laws and integrations), that integrates in:

$$\begin{aligned} p(\theta|y) &\propto (2\pi)^{-\frac{n}{2}}\Gamma(a)b^{-a} \\ &\propto \Gamma(\tfrac{n}{2})[\tfrac{1}{2}(\tilde{y}-J(\theta-\theta_0))^T(\tilde{y}-J(\theta-\theta_0))]^{-\frac{n}{2}} \\ &\propto [(\theta-\theta_0-(J^TJ)^{-}1J^T\tilde{y})^T\,(J^TJ)\,(\theta-\theta_0-(J^TJ)^{-}1J^T\tilde{y}) \\ &\quad +\tilde{y}^T\,(I-J(J^TJ)^{-}1J^T)\,\tilde{y}]^{-\frac{n}{2}} \\ &\propto [1+\tfrac{1}{\nu}(\theta-E(\theta))^TV(\theta)^{-1}(\theta-E(\theta))]^{-\frac{\nu+p}{2}}. \end{aligned}$$

$p(\theta|y)$ follows a Student law with $\nu = n - p$ degrees of freedom, and with variance $E(\theta)$ and $V(\theta)$. As in practice linearization (7) is done around an estimated parameter set ($\theta_0 = \hat{\theta}$), and since $\hat{\theta}$ minimizes the least-square energy, it follows that $\tilde{y} = (y - f(\hat{\theta}))\perp J$. This leads to some simplifications that prove (9):

$$\begin{aligned} E(\theta) &= \hat{\theta} + (J^TJ)^{-1}J^T\tilde{y} \\ &= \hat{\theta} \end{aligned}$$

and

$$\begin{aligned} \tfrac{1}{\nu}V(\theta)^{-1} &= (JJ^T)/(\tilde{y}^T\,(I-J(JJ^T)J^T)\,\tilde{y}) \\ &= (JJ^T)/(\tilde{y}^T\,\tilde{y}) \\ V(\theta) &= \tfrac{1}{n-p}\,(\tilde{y}^T\,\tilde{y})(JJ^T)^{-1} \\ &= \hat{\sigma}^2(JJ^T)^{-1}. \end{aligned}$$

6 Acknowledgements

We thank Torben Lund and Kristoffer Madsen for their help with preliminary fRMI experiment at the Hvidovre hospital (Copenhaguen), Muriel Roth and Bruno Nazarian for fMRI acquisition at La Timone hospital (Marseille), Jean-Baptiste Poline and Jean Daunizeau for useful discussions.

References

Aubert, A. and Costalat, R. 2002. A model of the coupling between brain electrical activity, metabolism, and hemodynamics : Application to the interpretation of functional neuroimaging. *NeuroImage* 17:1162–1181.

Birn, R. M., Saad, Z. S., and Bandettini, P. A. 2001. Spatial heterogeneity of the nonlinear dynamics in the fmri bold response. *NeuroImage* 14:817–826.

Box, G. E. P. and Tiao, G. C. 1992. Bayesian Inference in Statistical Analysis. New York: Wiley.

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. *The Journal of Neuroscience* 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. of Cerebral Blood Flow and Metabolism* 17:64–72.

Buxton, R. B., Uludağ, K., Dubowitz, D. J., and Liu, T. T. 2004. Modelling the hemodynamic response to brain activation. *NeuroImage* 23:220–233.

Buxton, R. B., Wang, E. C., and Frank, L. R. 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.* 39:855–864.

Dale and M., A. 1997. Selective averaging of rapidly presented individual trials using fmri. 5:329–390.

Davis, T. L., Kwong, K. K., Weisskoff, R. M., and Rosen, B. R. 1998. Calibrated functional mri: mapping the dynamics of oxydative metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 95:1834–1839.

Friston, K., Josephs, O., Rees, G., and Turner, R. 1998. Non-linear event-related responses in fMRI. *Magnetic Resonance in Medicine* 39:41–52.

Friston, K. J. 2002. Bayesian estimation of dynamical systems: an application to fmri. *NeuroImage* 16:513–530.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-B., Frith, C. D., and Frackowiak, R. S. J. 1995. Statistical parametric maps in functional imaging : A general linear approach. *Human Brain Mapping* 2:189–210.

Friston, K. J., Mechelli, A., Turner, R., and Price, C. J. 2000. Nonlinear responses in fmri : the balloon model, volterra kernels, and other hemodynamics. *NeuroImage* 12:466–477.

Friston, K. J. and Worsley, K. J. 1995. Analysis of fmri time-series revisited - again. *NeuroImage* 2:45–53.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 1998. Bayesian Data Analysis. Chapman & Hall, London.

Glover, G. H. 1999. Deconvolution of impulse response in event-related bold fmri. 9:416–29.

Hoge, R. D., Franceschini, M. A., Covolan, R. J. M., Huppert, T., Mandeville, J. B., and Boas, D. A. 2005. Simultaneous recording of task-induced changes in blood oxygenation, volume, and flow during diffuse optical imaging and arterial spin-labelling mri. *NeuroImage* 25:701–707.

Jacobsen, D. J., Hansen, L. K., and Madsen, K. H. 2005. Testing hypotheses on neural activity based on bold hemodynamics. *Neural Information Processing Systems - accepted paper* .

Janz, C., Heinrich, S. P., Kornmayer, J., Bach, M., and Hennig, J. 2001. Coupling of neural activity and bold fmri response: new insights by combination of fmri and vep experiments in transition from single events to continuous stimulation. 46:482–6.

Kershaw, J., Ardekani, B. A., and Kanno, I. 1999. Application of Bayesian inference to fMRI data analysis. *IEEE Trans Med Imaging* 18:1138–1153.

Krüger, G., Kastrup, A., Takahashi, A., and Glover, G. H. 1999. Simultaneous monitoring of dynamic changes in cerebral blood flow and oxygenation during sustained activation of the human visual cortex. *NeuroReport* 10:2939–2943.

Logothetis, N. K. and Pfeuffer, J. 2004. On the nature of the bold fmri contrast mechanism. 22:1517–31.

Marquardt, D. W. 1963. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* 11:431–441.

Miller, K. L., Luh, W.-M., Liu, T. T., Martinez, A., Obata, T., Wong, E. C., Frank, L. R., and Buxton, R. B. 2001. Nonlinear temporal dynamics of the cerebral blood flow response. *Humain Brain Mapping* 13:1–12.

Obata, T., Liu, T. T., Miller, K. L., Luh, W. M., Wong, E. C., and Buxton, R. B. 2004. Discrepancies between bold and flow dynamics in primary and supplementary motor areas : application of the balloon model to the interpretation of bold transients. *NeuroImage* 21:144–153.

Ogawa, S., Menon, R. S., Tank, D. W., Kim, S.-G., Merkle, H., Ellerman, J. M., and Ugurbil, K. 1993. Function brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging: a comparison of signal characteristics with a biophysical model. *Biophys. J.* 64:803–812.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. 1992. Nonlinear models. *Numerical Recipies in C* pp. 681–688.

Riera, J., Aubert, E., Iwata, K., Kawashima, R., Wan, X., and Ozaki, T. 2005. Fusing eeg and fmri based on a bottom-up model: inferring activation and effective connectivity in neural masses. *Phil. Trans. R. Soc.* B 360:1025–1041.

Riera, J., Watanabe, J., Kazuki, I., Naoki, M., Aubert, E., Ozaki, T., and Kawashima, R. 2004. A state-space model of the hemodynamic approach: nonlinear filtering of bold signals. *NeuroImage* 21:547–567.

Vazquez, A. L. and Noll, D. C. 1998. Nonlinear aspects of the bold response in functional mri. *NeuroImage* 7:108–118.

Wager, T. D., Vazquez, A., Hernandez, L., and Noll, D. C. 2005. Accounting for nonlinear bold effects in fmri: parameter estimates and a model for prediction in rapid event-related studies. 25:206–18.

Zheng, Y., Martindale, J., Johnson, D., Jones, M., Berwick, J., and Mayhew, J. 2002. A model of hemodynamic response and oxygen delivery to brain. 16:617–637.

Contents

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399