

**Validity and Power in Hemodynamic Response Modeling:**

**A comparison study and a new approach**

**Martin Lindquist\* and Tor D. Wager**

Department of Statistics, Columbia University, New York, NY, 10027

Department of Psychology, Columbia University, New York, NY, 10027

**RUNNING HEAD:** HEMODYNAMIC RESPONSE MODELING

**ADDRESS:**

Martin Lindquist

Department of Statistics

1255 Amsterdam Ave, 10th Floor, MC 4409

New York, NY 10027

Phone: (212) 851-2148

Fax: (212) 851-2164

E-Mail: [martin@stat.columbia.edu](mailto:martin@stat.columbia.edu)

**KeyWords:** fMRI, Hemodynamic response, magnitude, delay, latency, brain imaging, timing, analysis

**ABSTRACT**

One of the advantages of event-related fMRI is that it permits estimation of the shape of the hemodynamic response (HRF) elicited by cognitive events. Although studies to date have focused almost exclusively on the magnitude of evoked HRFs across different tasks, there is growing interest in testing other statistics, such as the time-to-peak and duration of activation as well. Although there are many ways to estimate such parameters, we suggest three criteria for optimal estimation: 1) the relationship between parameter estimates and neural activity must be as transparent as possible, 2) parameter estimates should be independent of one another, so that true differences in one parameter (e.g. delay) are not confused for apparent differences in other parameters (e.g. magnitude), and 3) statistical power should be maximized. In this work, we introduce a new modeling technique, based on the superposition of three inverse logit functions (IL), designed to achieve these criteria. In simulations based on real fMRI data, we compare the IL model with several other popular methods, including smooth finite impulse response (FIR) models, the canonical HRF with derivatives, nonlinear fits using a canonical HRF, and a standard canonical model. The IL model achieves the best overall balance between parameter interpretability and power. The FIR model was the next best choice, with gains in power at some cost to parameter independence. We provide software implementing the IL model.

## INTRODUCTION

Linear and nonlinear statistical models of fMRI data simultaneously incorporate information about the shape, timing, and magnitude of task-evoked hemodynamic responses. Most brain research to date has focused on the magnitude of evoked activation, although magnitude cannot be measured without assuming or measuring timing and shape information as well. Currently, however, there is increasing interest in measuring onset, peak latency and duration of evoked fMRI responses (Bellgowan, Saad, & Bandettini, 2003; Henson, Price, Rugg, Turner, & Friston, 2002; Hernandez, Badre, Noll, & Jonides, 2002; Menon, Luknowsky, & Gati, 1998; Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000; Rajapakse, Kruggel, Maisog, & von Cramon, 1998; Saad, DeYoe, & Ropella, 2003). Measuring timing and duration of brain activity has obvious parallels to the measurement of reaction time widely used in psychological and neuroscientific research, and thus may be a powerful tool for studying brain correlates of human performance. Recent studies, for instance, have found that although event-related BOLD responses evolve slowly in time, meaningful latency differences between averaged responses on the order of 100-200 ms can be detected (Aguirre, Singh, & D'Esposito, 1999; Bellgowan, Saad, & Bandettini, 2003; Formisano & Goebel, 2003; Formisano et al., 2002; Henson, Price, Rugg, Turner, & Friston, 2002; Hernandez, Badre, Noll, & Jonides, 2002; Liao et al., 2002; Richter et al., 2000). In addition, accurate modeling of hemodynamic response function (HRF) shape may prevent both false positive and negative results from arising due to ill-fitting constrained canonical models (Calhoun, Stevens, Pearlson, & Kiehl, 2004; Handwerker, Ollinger, & D'Esposito, 2004).

A number of fitting procedures exist that potentially allow for characterization of the latency and duration of fMRI responses. It requires only a model that extracts the shape of the hemodynamic response function (HRF) to different types of cognitive events. In analyzing the shape, summary measures of psychological interest (e.g., magnitude, delay, and duration) can be extracted. In this paper, we focus on the estimation of response height (H), time-to-peak (T), and full-width at half-max (W) as potential measures of response magnitude, latency, and duration (Fig. 1). These are not the only measures that are of interest—time-to-onset is also important, though it appears to be related to T but less reliable (Miezin et al., 2000)—but they capture some important aspects of the response that may be of interest to psychologists, as they relate to the latency and duration of brain responses to cognitive events. As we show here, not all modeling strategies work equally well for this purpose—i.e., they differ in the validity and the statistical precision of the estimates they provide.

Ideally, estimated parameters of the HRF (e.g., H, T, and W) should be interpretable in terms of changes in neuronal activity, and they should be estimated such that statistical power is maximized. The issue of interpretability is complex, as the evoked HRF is a complex, nonlinear function of the results of neuronal and vascular changes (Buxton, Wong, & Frank, 1998; Logothetis, 2003; Mechelli, Price, & Friston, 2001; Vazquez & Noll, 1998; Wager, Vazquez, Hernandez, & Noll, 2005). Essentially, the problem can be divided into two parts, shown in Figure 2.

The first issue is the question of whether changes in physiological, neuronal-level parameters (such as the magnitude, delay, and duration of evoked changes in neuronal activity) translate into changes in corresponding parameters of the HRF. Potential

relationships are schematically depicted on the left side of Figure 2. Ideally, changes in neuronal parameters would each produce unique changes in one parameter of the HR shape, shown as solid arrows. However, neuronal changes may produce true changes in multiple aspects of the HR shape, as shown by the dashed arrows on the left side of Figure 2. The second issue is whether changes in the evoked HR are uniquely captured by parameter estimates of H, T, and W. That is, whatever combination of neuro-vascular effects leads to the evoked BOLD response, does the statistical model of the HRF recover the true magnitude, time to peak, and width of the response? This issue concerns the accuracy of the statistical model of the evoked response and the independence of H, T, and W parameter estimates, irrespective of whether the true HR changes were produced by uniquely interpretable physiological changes.

In this paper, we start from the assumption that meaningful changes can be captured in a linear or nonlinear time-invariant system, and chiefly address the second issue of whether commonly used HR models can accurately estimate true changes in the height, time to peak, and width of HR responses. That is, we assess the interpretability of H, T, and W estimates (right boxes in Fig. 2) given true changes in the shape of the evoked signal response (center boxes in Fig. 2).

Importantly, however, the complex relationship between neuronal activity and evoked signal response also places important constraints on the ultimate neuronal interpretation of evoked fMRI signal. While a full analysis of BOLD physiology is beyond the scope of the current work, we provide a brief analysis of some important constraints in the discussion, and refer the reader to more detailed descriptions of BOLD physiology (Buxton, Wong, & Frank, 1998; Logothetis, 2003; Mechelli, Price, & Friston,

2001; Vazquez & Noll, 1998; Wager, Vazquez, Hernandez, & Noll, 2005). In spite of this limitation, the estimation of the magnitude, latency, and width of empirical BOLD responses to psychological tasks is of great interest, because these responses may provide meaningful brain-based correlates of cognitive activity (e.g., Bellgowan, Saad, & Bandettini, 2003; Henson et al., 2002).

***To assume or not to assume?***

Typically used linear and nonlinear models for the HRF vary greatly in the degree to which they make *a priori* assumptions about the shape of the response. In the most extreme case, the shape of the HRF is completely fixed; a canonical HRF is assumed, and the height (i.e., amplitude) of the response alone is allowed to vary (Worsley & Friston, 1995). The magnitude of the height parameter is taken to be an estimate of the strength of activation. By contrast, one of the most flexible models, a finite impulse response (FIR) basis set, contains one free parameter for every time-point following stimulation in every cognitive event type modeled (Glover, 1999; Goutte, Nielsen, & Hansen, 2000; Ollinger, Shulman, & Corbetta, 2001). Thus, the model is able to estimate an HRF of arbitrary shape for each event type in each voxel of the brain. A popular related technique is the *selective averaging* of responses following onsets of each trial type ((Dale & Buckner, 1997; Maccotta, Zacks, & Buckner, 2001); a time x condition ANOVA model is often used to test for differences between event types).

Many basis sets fall somewhere midway between these two extremes and have an intermediate number of free parameters, providing the ability to model a family of plausible HRFs throughout the brain. For example, a popular choice is to use a canonical

HRF and its derivatives with respect to time and dispersion (we use TD to denote this hereafter (Friston, Josephs, Rees, & Turner, 1998; Henson, Price, Rugg, Turner, & Friston, 2002)). Such approaches also include the use of basis sets composed of principal components (Aguirre, Zarahn, & D'Esposito, 1998; Woolrich, Behrens, & Smith, 2004), cosine functions (Zarahn, 2002), radial basis functions (Riera et al., 2004), spline basis sets, and a Gaussian model (Rajapakse, Kruggel, Maisog, & von Cramon, 1998). Recently a method was introduced (Woolrich, Behrens, & Smith, 2004), which allows the specification of a set of optimal basis functions. In this method a large number of sensibly shaped HRFs are randomly generated, and singular value decomposition is used on the set of functions to find a small number of basis sets that optimally span the space of the generated functions. Another promising approach uses spectral basis functions to provide independent estimates of magnitude and delay in a linear modeling framework (Liao et al., 2002).

Because linear regression is limited in its ability to provide independent estimates of multiple parameters of the HRF, a number of researchers have used nonlinear fitting of a canonical function with free parameters for magnitude and onset/peak delay (Kruggel & von Cramon, 1999; Kruggel, Wiggins, Herrmann, & von Cramon, 2000; Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000). The most common criticisms of such approaches are their computational costs and potential convergence problems, although increases in computational power make nonlinear estimation over the whole brain feasible.

In general, the more basis functions used in a linear model or the more free parameters in a nonlinear one, the more flexible the model is in measuring the magnitude

and other parameters of interest. However, flexibility comes at a cost: More free parameters means more error in estimating them, fewer degrees of freedom, and decreased power and validity if the model regressors are collinear. In addition, even if the basis functions themselves are orthogonal, as with a principal components basis set, this does not guarantee that the regressors, which model multiple overlapping events throughout an experiment, are orthogonal. Finally, it is easier and statistically more powerful to interpret differences between task conditions (e.g., A – B) on a single parameter such as height than it is to test for differences in multiple parameters (A1A2A3 – B1B2B3)—conditional, of course, on the interpretability of those parameter estimates. The temporal derivative of the canonical SPM HRF, for example, is not uniquely interpretable in terms of activation delay; both magnitude and delay are functions of the two parameters (Calhoun, Stevens, Pearlson, & Kiehl, 2004; Liao et al., 2002).

All these problems suggest that using a single, canonical HRF is a good choice. Indeed, it offers optimal power if the shape is specified exactly correctly. However, the shape of the HRF varies as a function of both task and brain region, and any fixed model is bound to be wrong in much of the brain (Birn, Saad, & Bandettini, 2001; Handwerker, Ollinger, & D'Esposito, 2004; Marrelec, Benali, Ciuciu, Pelegriani-Issac, & Poline, 2003; Wager, Vazquez, Hernandez, & Noll, 2005). If the model is incorrectly specified, then statistical power will decrease, and the model may also produce invalid and biased results. In addition, using a canonical HRF provides no way to assess latency and duration—in fact, differences between conditions in response latency will be confused for differences in amplitude (Calhoun, Stevens, Pearlson, & Kiehl, 2004).



Thus, neither the fixed-response nor the completely flexible response appear to be optimal solutions, and using a restricted set of basis functions is an alternative that may preserve validity and power within a plausible range of true HRFs (Woolrich, Behrens, & Smith, 2004). However, an advantage of the more flexible models is that *height*, *latency*, and *response width* (duration) can potentially be assessed. This paper is dedicated to consideration of the validity and power of such estimates using several common basis sets. In this work, we also introduce a new technique for modeling the HRF, based on the superposition of three inverse logit functions (IL), which balances the need for both interpretability and flexibility of the model. In simulations based on actual HRFs measured in a group of 10 participants, we compare the performance of this model to four other popular choices of basis functions. These include an enhanced smooth FIR filter (Goutte, Nielsen, & Hansen, 2000), a canonical HRF with time and dispersion derivatives (TD; (Calhoun, Stevens, Pearlson, & Kiehl, 2004; Friston, Josephs, Rees, & Turner, 1998)), the nonlinear fit of a gamma function used by Miezin et al. (NL, (Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000)) and the canonical SPM HRF (Friston, Josephs, Rees, & Turner, 1998). We show that the IL model can capture magnitude, delay, and duration of activation with less error than the other methods tested, and provides a promising way to flexibly but powerfully test the magnitude and timing of activation across experimental conditions.

### ***What makes a good model?***

Ideally, differences in estimates of H, T, and W across conditions would reflect differences in the height, time-to-peak, and width of the true BOLD response (and, ideally, unique changes in corresponding neuronal effects as well, though this is unlikely

under most conditions due to the complex physiology underlying the BOLD effect). These relationships are shown as solid lines connecting true signal responses and estimated responses in the right side of Fig. 2. A 1:1 mapping between true and estimated parameters would render estimated parameters uniquely interpretable in terms of the underlying shape of the BOLD response. As the example above illustrates, however, there is not always a clean 1:1 mapping, indicated by the dashed lines in Fig 2. True differences in delay may appear as estimated differences in H (for example), if the model cannot accurately account for differences in delay. This potential for cross-talk exists among all the estimated parameters. We refer to this potential as *confusability*, defined as the bias in a parameter estimate that is induced by true changes in another nominally unrelated parameter. In our simulations, based on empirical HRFs, we independently varied true height, time to peak, and response width (so that the true values are known). We show that there is substantial confusability between true differences and estimates, and that this confusability is dependent on the HRF model used. Thus, the chosen modeling system places practical constraints on the interpretability of H, T, and W estimates.

Of course, the interpretability of H, T, and W estimates also depends on the relationship between underlying changes in neural activity and changes in the magnitude and shape of the true fMRI signal (Buckner, 2003; Buxton, Wong, & Frank, 1998; Logothetis, 2003; Riera et al., 2004), shown by solid arrows (expected relationships) and dashed arrows (problematic relationships) on the left side of Fig 2. Underlying BOLD physiology limits the ultimate interpretability of the parameter estimates in terms of physiological parameters—e.g., prolonged changes in postsynaptic activity. Because of

the complexity of making such interpretations, we do not attempt to relate BOLD signal to underlying neuronal activity, but rather treat the evoked HRF as a signal of interest. Future work may provide the basis for more accurate models of BOLD responses with physiological parameters that can be practically applied to cognitive studies (e.g., Buxton, Wong, & Frank, 1998). For the present, we feel it is important to acknowledge some of the theoretical limitations imposed by BOLD physiology on the interpretation of evoked BOLD magnitude, latency, and response width, and thus we return to this point in the following sections.

## METHODS

In this section we introduce a method for modeling the hemodynamic response function, based on the superposition of 3 inverse logit (IL) functions, and describe how it compares to four other popular techniques — a non-linear fit on two gamma functions (NL), the canonical HRF + temporal derivative (TD), a finite impulse response basis set (FIR), and the canonical SPM HRF (Gam) — in simulations based on empirical fMRI data.

### *Overview of the Models*

We begin with an overview of the models included in our simulation study.

#### *(i) The Inverse Logit Model*

The logit function is defined as  $x = \log(p(1-p)^{-1})$ , where  $p$  takes values between 0 and 1. Conversely, we can express  $p$  in terms of  $x$  as

$$p = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}}. \quad (1)$$

This function is typically referred to as the inverse logit function and an example is shown in Fig. 3A. In the continuation we will denote this function as  $L(x)$ , i.e.  $L(x) = p$ .

It is important to note a number of important properties of  $L(x)$ . It is an increasing function of  $x$ , which takes the values 0 and 1 in the limits. In addition,  $L(t - T) = 0.5$  when  $t = T$ .

To derive a model for the hemodynamic response function that can efficiently capture the details that are inherent in the function, such as the positive rise and the post-activation undershoot, we will use a superposition of three separate inverse logit functions. The first describing the rise following activation, the second the subsequent decrease and undershoot, while the third describes the stabilization of the HRF, shown in Fig. 3A-C.

Our model of the hemodynamic response function,  $h(t)$ , can therefore be written in the following form:

$$h(t | \theta) = \alpha_1 L((t - T_1)/D_1) + \alpha_2 L((t - T_2)/D_2) + \alpha_3 L((t - T_3)/D_3). \quad (2)$$

In this particular model the function  $h(t)$  will be based on nine variable parameters (seven free parameters after imposing additional constraints), given by  $\theta = (\alpha_1, T_1, D_1, \alpha_2, T_2, D_2, \alpha_3, T_3, D_3)$ . The  $\alpha$  parameters control the direction and

amplitude of the curve. If  $\alpha$  is positive,  $\alpha \cdot L(x)$  will be an increasing function that takes values between 0 and  $\alpha$ . If  $\alpha$  is negative,  $\alpha \cdot L(x)$  will be a decreasing function that takes values between 0 and  $-\alpha$ . The parameter  $T$  is used to shift the center of the function  $T$  time units. In effect it defines the time point,  $x$ , where  $L(x) = 1/2$  and can be used as a measure of the time to half-peak. Finally the parameter  $D$  controls the angle of the slope of the curve, and works as a scaling parameter.

In our implementation of the model we begin by constraining the amplitude of the third inverse logit function, so that the fitted response ends at magnitude 0, by setting  $\alpha_3 = |\alpha_2| - |\alpha_1|$ . In addition we want the function  $h(t; \theta)$  to begin at zero at the time point  $t = 0$ . Therefore we place the constraint  $h(0 | \theta) = 0$  on the model, which implies that

$$\alpha_2 = \alpha_1 \frac{(L(-T_3/D_3) - L(-T_1/D_1))}{(L(-T_3/D_3) + L(-T_2/D_2))}. \quad (3)$$

By applying these two constraints on the amplitude of the basis functions, this leads to a model with 7 variable parameters. Fig. 3A–C shows an example of how varying the parameters can control the shape of the function  $L(x)$ . By superimposing these three curves we obtain the function depicted in Fig. 3D, which shows an example of an IL fit (solid line) to an empirical HRF (dashed line). Note that this function efficiently captures the major details typically present in the HRF and illustrates how effective three inverse logit functions can be in describing its basic shape.

The interpretability of the parameters in the model are increased if the first and second and the second and third IL functions are made as orthogonal as possible to one another. This will be true if the following conditions hold:

$$T_2 - T_1 > (D_1 + D_2)k \quad (4)$$

and

$$T_3 - T_2 > (D_2 + D_3)k, \quad (5)$$

where  $k$  is a constant (see Appendix for more details). To ensure that these constraints hold, restrictions can be placed on the space of possible parameter values allowed in fitting the model.

### *Problem Formulation*

Let us define  $f(t|\theta)$  to be the convolution between the IL model for the hemodynamic response, denoted by  $h(t|\theta)$ , and a known stimulus function,  $s(t)$ . Our non-linear regression model for the fMRI response at time  $t_i$  can be written as

$$y_i = f(t_i|\theta) + \varepsilon_i \quad (6)$$

where  $\varepsilon_i \sim N(0, V\sigma^2)$ . In matrix format we can write this as

$$Y = F(X;\theta) + E \quad (7)$$

where  $Y = (y_1, \dots, y_N)^T$  is the data vector,  $E = (\varepsilon_1, \dots, \varepsilon_N)^T$  is a noise vector and  $F(X; \theta) = (f(t_1; \theta), \dots, f(t_N; \theta))^T$ .

The goal of our analysis is to find the parameters  $\theta^*$  such that the model best fits the data in the least-squares sense, i.e. we seek

$$\theta^* = \arg \min S(\theta) \quad (8)$$

where

$$S(\theta) = (Y - F(X; \theta))^T V^{-1} (Y - F(X; \theta)) . \quad (9)$$

Under the assumption that the noise is independent and identically distributed (*iid*), then  $V = I$  and Eq. (9) can be written on the form

$$S(\theta) = \sum_{i=1}^n (y_i - f(t_i; \theta))^2 . \quad (10)$$

In this situation the value  $\theta^*$  that maximizes  $S(\theta)$  is equivalent to the maximum likelihood estimate (MLE) of  $\theta$ .

It is well-known that fMRI noise typically exhibits temporal dependence and it is crucial that this dependence be taken into consideration when fitting the model. In our implementation we assume that the noise term can be modeled using an AR(1) model. As  $F(X; \theta)$  is a non-linear function in  $\theta$ , the process of finding the parameters that

maximize Eq. (9) will almost always involve using an iterative search method. In order to speed-up the computational efficiency of the applied algorithm, we would like to avoid repeatedly inverting the matrix  $V$ . Under the assumption of AR(1) noise we can fortunately express the inverse of  $V$  as,

$$V^{-1} = \begin{bmatrix} 1 & -\phi & 0 & \dots & \dots & 0 \\ -\phi & d & -\phi & 0 & & \\ 0 & -\phi & d & -\phi & 0 & \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & 0 & \ddots & \ddots & -\phi \\ 0 & & & 0 & -\phi & 1 \end{bmatrix} \quad (11)$$

where  $d = 1 + \phi^2$ . Using this expression allows us to circumvent the need for repeated inversion of the correlation matrix and we can rewrite Eq. (9) as

$$S(\theta, \phi) = z_1^2(1 - \phi^2) + \sum_{i=2}^n (z_i - \phi z_{i-1})^2 \quad (12)$$

where

$$z_i = y_i - f(t_i | \theta). \quad (13)$$

Note that for  $\phi = 0$ , the cost functions defined in Eqs. (9) and (12) are equivalent. In the continuation we will include the  $\phi$  term when referring to  $\theta$ , i.e.  $\theta = (\theta, \phi)$ .

The optimization problem stated in Eq. (12) can be solved using a number of different methods. Traditionally deterministic methods for solving the problem have been



used, but recently with increased computational power stochastic approaches have received increased attention.

### *Deterministic Solutions*

The optimization problem stated in Eq. (8) can be solved using numerical algorithms such as the *Gauss-Newton* or *Levenberg-Marquardt* algorithms. Both these methods are iterative and make use of the Jacobian of the objective function at the current solution. In addition, they both have fast rates of convergence. The Gauss-Newton has quadratic convergence, which implies that there exists a constant  $\mu > 0$  such that

$$\lim_k \frac{\|\theta_{k+1} - \theta\|}{\|\theta_k - \theta\|^2} = \mu \quad (14)$$

for each iteration  $k$ , where  $\theta_k$  denotes the estimate of the parameter vector after the  $k^{\text{th}}$  iteration and  $\theta$  the true minimum. The Levenberg-Marquardt algorithm combines the Gauss-Newton algorithm with the method of gradient descent to guarantee convergence with quadratic convergence near the minimum. Though the convergence properties are comparable, the Levenberg-Marquardt algorithm is more robust, in the sense that it is able to find a solution even if it starts out far away from the final minimum. Both the Gauss-Newton and Levenberg-Marquardt algorithms are easily implemented for the IL model, using the fact that the inverse logit function has a straightforward derivative  $L'(x) = L(x)(1 - L(x))$ .

*Stochastic Solutions*

The problem with deterministic methods is that they always converge to the nearest local minimum-error from the initial value, regardless of whether it is a local or global minimum. Hence, the parameter estimate is strongly dependent on the initial values given to the algorithm. As it is common for non-linear functions to have multiple local minima in addition to the global minimum that is being sought, it may be beneficial to use a stochastic approach that samples points across all of parameter space, as they are less likely to converge to a local minimum. Though such methods are computationally slower than deterministic methods, they are more likely to find the global extreme point and will at the very least allow us to investigate whether the fits obtained using the faster deterministic methods are accurate.

The *simulated annealing* algorithm (Metropolis et al. 1953, Kirkpatrick et al. 1983) is one such approach, which involves moving about randomly in parameter space searching for a solution that minimizes the value of the cost function. This method allows for an initially wide exploration of parameter space, which is increasingly narrowed about the global extreme point as the method progresses. This is possible, as the algorithm employs a random search which not only accepts changes that lead to a decrease in the value of the cost function, but also some changes that increase it.

There are four steps to implementing the simulated annealing algorithm:

1. Choose an initial value for the parameter vector  $\theta_0$ . (Unlike the L-M algorithm this choice is not critical).

2. Choose a new candidate solution,  $\theta_{i+1}$ , based on a random perturbation of the current solution of  $\theta_i$ .
3. If the candidate solution decreases the error, as defined by the cost function  $S(\theta)$  (Eq. 12), then automatically accept the new solution. If the error increases, accept the candidate solution with probability  $\min\{\exp((S(\theta_i) - S(\theta_{i+1}))/\tau_i), 1\}$ , where  $\tau_i$  is the so-called *temperature function* at iteration  $i$ . The temperature function decreases for each iteration of the algorithm and as  $\tau_i \rightarrow 0$  the parameters will only be updated if  $\Delta h < 0$ .
4. Update  $\tau_i$  to  $\tau_{i+1}$  and repeat from step 2.

Setting the temperature function is a critical part of the simulated annealing method, as high values of  $\tau$  give wider exploration, and less chance of getting stuck in a local minimum, while lower values reduce the likelihood of moving unless the error is decreased. By starting out with a large value of  $\tau$  and letting it converge to zero, we are allowing for a wide exploration in the beginning of the algorithm, which will narrow as the number of iterations increase. If the temperature function is allowed to decrease at a slow enough pace the global minimum can be reached with probability 1. However, it is typically not practical to use such a slowly decreasing schedule, and therefore it can not be guaranteed that a global optimum will be reached.

The candidate solution is obtained by perturbing the current solution by the outcome of a uniformly distributed random variable, which we will denote  $\Delta\theta$ . In our

implementation we vary the amount each of the components of  $\theta$  are allowed to jump according to the following:

$$\begin{aligned}\Delta T_i &\sim \text{unif}(-r_1, r_1) \\ \Delta D_i &\sim \text{unif}(-r_2, r_2) \\ \Delta \delta_i &\sim \text{unif}(-r_3, r_3) \\ \Delta \phi &\sim \text{unif}(-r_4, r_4).\end{aligned}\tag{15}$$

The objective function, as it is stated in Eq. 12, is not convex. Therefore, whether or not a deterministic solution will converge to its global optimum will strongly depend on the initial values given to the algorithm. To circumvent this issue, we recommend using the simulated annealing approach, and this is the model fitting method we will use in the continuation of this paper. To determine an appropriate temperature function we randomly generated a number of sensibly shaped HRFs, which we used as pilot data to calibrate our schedule. In our implementation we let  $\tau_i = C/\log(1+i)$ , where  $C$  is a large positive number chosen so that the acceptance rate of the algorithm is approximately 80%. For the simulation study performed in this paper we used values on the order of  $r_1 = 5$ ,  $r_2 = 0.1$ ,  $r_3 = 0.1$  and  $\phi = 0.1$ . It should be noted that other distribution functions could have been used instead of the uniform to perturb the solution.

We tested the convergence properties of the simulated annealing approach at a number of randomly chosen starting points and it converged in a consistent manner to the global minimum. Simulated annealing converged much more reliably than the deterministic methods. In order to better characterize the distribution of the parameter

estimates obtained using simulated annealing, we also performed a series of 1000 simulations on each of 5 plausible signal-to-noise ratio (SNR) levels for fMRI data, ranging from 0.05-0.5. Visual inspection of the distributions suggested that the parameter estimates were normally distributed for each SNR. This conclusion was supported by tests of skewness and kurtosis on each distribution, for which the 95% confidence intervals all contained 0, as expected if parameter estimates follow a normal distribution.

*(ii) Non-Linear fit on two Gamma functions*

The model consists of a linear combination of two Gamma functions with a total of 6 variable parameters, i.e.

$$h(t) = A \left( \frac{t^{\alpha_1-1} \beta_1^{\alpha_1} e^{-\beta_1 t}}{\Gamma(\alpha_1)} - c \frac{t^{\alpha_2-1} \beta_2^{\alpha_2} e^{-\beta_2 t}}{\Gamma(\alpha_2)} \right) \quad (17)$$

where  $A$  controls the amplitude,  $\alpha$  and  $\beta$  control the shape and scale, respectively, and  $c$  determines the ratio of the response to undershoot.  $\Gamma$  represents the gamma function, which acts as a normalizing parameter. This model can fit a wide variety of different HRF shapes within the ranges of commonly observed event-related responses. The six parameters of the model are fit using the Levenberg-Marquardt algorithm.

*(iii) Temporal Derivative*

This model consists of a linear combination of the canonical HRF, which is described in greater detail in (v), and its temporal derivative. Therefore there are two

variable parameters: the amplitudes of the HRF and its derivative. Amplitude estimation was performed using the estimation procedure outlined in Calhoun (Calhoun, Stevens, Pearlson, & Kiehl, 2004).

*(iv) Smooth FIR*

In our implementation we used a semi-parametric smooth FIR model (Goutte, Nielsen, & Hansen, 2000), as it was expected to outperform the standard FIR model. In general, the FIR basis set contains one free parameter for every time point following stimulation in every cognitive event type modeled. Assume that  $x(t)$  is a T-dimensional vector of stimulus inputs, which is equal to 1 at time  $t$  if a stimuli is present at that time point and 0 otherwise. Now we can define the design matrix corresponding to the FIR filter of order  $d$  as,

$$\mathbf{X} = \begin{bmatrix} x(d) & \cdots & x(1) \\ x(d+1) & \cdots & x(2) \\ \vdots & & \vdots \\ x(T) & \cdots & x(T-d+1) \end{bmatrix} \quad (17)$$

In addition, let  $\mathbf{Y}$  be the vector of measurements.

The traditional least-square solution,

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (18)$$

is very sensitive to noise. The individual parameter estimates will also be noisy, which increases the variance of H, T, and W estimates considerably. In particular, FIR HRF estimates contain high-frequency noise that is unlikely to actually be part of the underlying hemodynamic response. To constrain the fit to be smoother (but otherwise of arbitrary shape), Goutte et al. put a Gaussian prior on  $\beta$  and calculated the maximum *a posteriori* estimate:

$$\hat{\beta}_{map} = (\mathbf{X}^T \mathbf{X} + \sigma^2 \Sigma^{-1})^{-1} \mathbf{X}^T \mathbf{Y} \quad (19)$$

where the elements of  $\Sigma$  are given by

$$\Sigma_{ij} = \nu \exp\left(-\frac{h}{2}(i-j)^2\right). \quad (20)$$

This is equivalent to the solution of the least square problem with a penalty function, i.e.,

$\beta_{map}$  is the solution to the problem:

$$\max \left\{ (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \sigma^2 \sum s_{ij} \beta_i \beta_j \right\} \quad (21)$$

where  $s_{ij}$  are the components of the matrix  $\Sigma^{-1}$ . Note that replacing  $\Sigma$  with the identity matrix gives the ridge regression solution (Jain, 1985). As with ridge regression the estimates will be biased with a certain amount of shrinkage.

The parameters of this model are  $h$ ,  $\nu$  and  $\sigma$ . The parameter  $h$  controls the smoothness of the filter and Goutte recommends that this value be set *a priori* to:

$$h = \frac{1}{\sqrt{7/TR}} \quad (22)$$

We used this value in our implementation. In calculating the filter, only the ratio of the parameters  $\nu$  and  $\sigma$  is actually of interest, and we determined empirically, using pilot data, that the ratio:

$$\frac{\sigma^2}{\nu} = 1 \quad (23)$$

gave rise to adequately smooth FIR estimates, without giving rise to significant biases in the estimates due to shrinkage.

#### (v) Gamma

This model again consists of a linear combination of two Gamma functions. However in this implementation all parameters except the amplitude is fixed, giving rise to a model with only one variable parameter. The other parameters were set to be  $\alpha_1 = 6$ ,  $\alpha_2 = 16$ ,  $\beta_1 = \beta_2 = 1$  and  $c = 1/6$ , which are the defaults implemented in SPM99 and SPM2.

#### *Estimating parameters*



After fitting each of the models, the next step is to estimate the height (H), time-to-peak (T) and width (W). Of particular interest is to estimate the difference in H, T and W across different psychological event types. Most of the models used have closed form solutions describing the fits (the Gamma based models & IL), and hence clear estimates of H, T and W can be derived from combinations of parameter estimates. However, a lack of closed form solution (e.g., for FIR models) does not preclude reading off the values from the fits.

When H, T, and W cannot be calculated directly using a closed form solution, we use the following procedure to estimate them from fitted HRF estimates. Height estimates are calculated by taking the derivative of the model function and setting it equal to 0. In order to ensure that this is a maximum, we should check that the second derivative is less than 0. If dual peaks exist, we choose the first one. Hence, our estimate of time-to-peak is  $T = \min\{t \mid h'(t) = 0 \text{ \& } h''(t) < 0\}$ , where  $t$  indicates time and  $h'(t)$  and  $h''(t)$  denote first and second derivatives of the HRF  $h(t)$ . For high-quality HRFs this is sufficient, but in practical application in a wide range of studies, it is also desirable to constrain the peak to be neither the first nor last parameter estimate. To estimate the peak we use  $H = h(T)$ . Finally, to estimate the width we perform the following steps:

- (i) Find the earliest time point  $t_u$  such that  $t_u > T$  and  $h(t_u) < H/2$ , i.e. the last point *before* the peak that lies below half maximum.
- (ii) Find the latest time point  $t_l$  such that  $t_l < T$  and  $h(t_l) < H/2$ , i.e. the last point *after* the peak that lies below half maximum.

- (iii) As both  $t_u$  and  $t_l$  take values below  $0.5H$ , the distance  $d = t_u - t_l$  overestimates the width. Similarly, both  $t_{u-1}$  and  $t_{l+1}$  take values above  $0.5H$ , so the distance  $d = t_{u-1} - t_{l+1}$  underestimates the width. We use linear interpolation to get a better approximation of the time points between  $(t_l, t_{l+1})$  and  $(t_{u-1}, t_u)$  where  $h(t)$  is equal to  $0.5H$ . According to this reasoning, we find that

$$W = (t_{u-1} + \Delta_u) - (t_{l+1} + \Delta_l) \quad (24)$$

where

$$\Delta_l = \frac{h(t_{l+1}) - 0.5H}{h(t_{l+1}) - h(t_l)} \quad (25)$$

and

$$\Delta_u = \frac{h(t_u) - 0.5H}{h(t_u) - h(t_{u-1})}. \quad (26)$$

For high-quality HRFs this procedure suffices, but if the HRF estimates begin substantially above or below 0 (the session mean), then it may be desirable to calculate local HRF deflections by calculating H relative to the average of the first one or two estimates.

For the Gamma based models simple contrasts exist for the magnitude. For TD we use the bias corrected amplitude estimate given by Calhoun et al. (Calhoun, Stevens, Pearlson, & Kiehl, 2004). For the IL model we derive a number of contrasts in the appendix, the results of which are presented here. If the constraints given in [4] and [5] hold, the first and second logit functions are approximately orthogonal and the estimates of H, T and W are given by:

$$H = \alpha_1, \quad (27)$$

$$T = T_1 + D_1 k, \quad (28)$$

and

$$W = T_2 - T_1 - D_2 \log\left(\frac{2\alpha_2}{\alpha_1} - 1\right). \quad (29)$$

Note that the estimates of H and T are independent of one another. The estimate of W depends to a certain degree on both H and T, but the simulation studies we present here show that it is less impacted by changes in H and T than the other models.

Note that although we use model-derived estimates of H, T, and W where possible, the direct approach of estimation from the fitted HRFs is also valid. This is aided in the case of the IL model by the fact that the inverse logit function has a straightforward derivative, as  $L'(x) = L(x)(1 - L(x))$ .

### ***Simulation Study***

The simulations are based on actual HRFs obtained from a visual-motor task in 10 participants (spiral gradient echo imaging at 3T, 0.5 s TR; (Noll, Cohen, Meyer, & Schneider, 1995)). Seven oblique slices were collected through visual and motor cortex at high temporal resolution, 3.12 x 3.12 x 5 mm voxels, TR = 0.5 s, TE = 25 ms, flip angle = 90, FOV = 20 cm. Participants viewed contrast-reversing checkerboards (16 Hz, 250 ms stimulation, full-field to 30 degrees of visual angle) and made manual button-press responses upon detection of each stimulus. ‘Events’ consisted of 1, 2, 5, 6, 10, or 11 such stimuli spaced 1 s apart, followed by 30 s of rest (open-eye fixation). For the simulation

study, we used the 5-stimulus events only; 16 such events were presented to each participant. BOLD activity time-locked to event onset, averaged across a region in the left primary visual cortex defined in a separate localizer scan for each individual, served as the true HRFs in our simulation. Thus, we obtained 10 empirical HRFs, one for each participant. This data has been used previously to describe nonlinearities in BOLD data (Wager, Vazquez, Hernandez, & Noll, 2005).

We began by constructing stimulus functions for 6-minute runs of randomly intermixed event types (A & B), occurring at random intervals of length 2-18 seconds. Assuming a linear time-invariant system, the stimulus functions were convolved with the empirically derived HRFs, and AR(1) noise was added to the resulting time course.

The HRFs for A and B were modified prior to creating the time course in order to create three kinds of “true” effects – an A – B amplitude difference, time-to-peak difference, and duration difference. In total we ran 3 types of simulations:

- S1. (Height mod) The HRF corresponding to event B has half of the amplitude of the HRF corresponding to event A. In this scenario there is a true A – B difference in H of 0.5, but no time-to-peak or duration difference.
  
- S2. (Delay mod) The HRF corresponding to event B has a 3 second onset delay compared to HRF A. In this scenario there is a 3-s difference in T between the HRFs, but no amplitude or duration difference.

- S3. (Duration mod) The width of HRF B is increased by 4 seconds compared to HRF A by extending the time at peak by 8 time points (0.5 s TR). In this scenario there is a 4 s difference in W between the HRFs, but no amplitude or time-to-peak difference.

Each of these three simulations was performed using the HRF for each of the 10 participants without modifications for HRF A, and modified as above for HRF B. For each participant the simulation was repeated 1000 times using different simulated AR(1) noise in each repetition.

We were interested in the efficiency and bias of A – B differences for individuals and in the group analysis treating participant as a random effect. For each participant in each simulation, we estimated A – B differences in H, T, and W. We quantified the relative statistical power of each type of model to recover these “true” effects. We also quantified the confusability of true differences in one effect (e.g., the manipulation of T in S2) with apparent differences (bias) in another (e.g., the estimated W in S2). This was accomplished by examining the relative statistical power across model types for detecting these ‘crossed’ effects, whose magnitude—if H, T, and W estimates are independent—should be 0, as well as calculating how the true change in one parameter induced changes in the bias of the other non-modulated parameters.

### ***Application to voxel-wise time courses***

Using data from the same experiment described in the previous section, we extracted the time courses from individual voxels, contained in the visual cortex, from

each of the 10 subjects. To each voxel-wise time course we applied the five different fitting procedures used in this paper and estimated H, T and W for each.

### *Relationships between neural activity and activation parameters*

Relating neural activity to model parameters is complex, and ultimately places constraints on the interpretation of the parameter estimates. Here, we conduct a preliminary exploration of the conditions under which changes in neuronal activation parameters may lead to specific changes in corresponding HR parameters. We stress that our analysis here is necessarily greatly simplified ; however, it may provide some rules of thumb for the range of conditions under which H, T, and W might roughly correspond to changes in neuronal activity magnitude, onset delay, and duration. For the purposes of this illustration, we assume that changes in neural firing rates (or postsynaptic activity) during brief periods of cognitive activity constitute neural ‘events’ – for example, an ‘event’ may consist of a brief memory refreshing operation that increases neural activity briefly and recurs with some frequency.

The theoretical relationships between neural events (event magnitude, event train onset and event train duration) and fMRI signal (H, T and W) vary depending on the duration of event trains and nonlinear properties of the response. We consider these relationships assuming linear responses and, separately, nonlinear magnitude saturation effects using estimates from previous work (Wager, Vazquez, Hernandez, & Noll, 2005). To construct what HR responses might look like if the response saturates nonlinearly in time, we performed the modified convolution procedure described in Wager et al., 2005, using event trains that varied in event magnitude, onset, and duration. We vary the length of epochs from brief, 1 s events to 18 s stimulation epochs, and consider whether true

differences between two conditions A and B yield estimated differences only in the parameters varied or in others as well.

## RESULTS

### *Organization of results*

In three simulations, we varied the true difference in H (S1), T (S2), and W (S3) between two versions of the same empirical HRFs (HRF A and HRF B). In Figs 4-6 the results are shown for each of the three simulations. In the top row, the true effects are shown by horizontal lines, and means and error bars for each of the 10 “participants,” each with a unique empirically-derived HRF, are shown by the vertical lines. In the bottom panels the between-subjects (‘random effects’) means and standard errors are shown. These can be used to assess the significance of the modulated HRF A – HRF B effect in each simulation, as well as biases in estimates of non-modulated parameters. Figure 7A summarizes these results in bias vs. variance plots for the H, T and W effects for each simulation type. Figure 7B (which we denote as confusability plots) shows a scatter plot of the change in bias for the two non-modulated parameters for each simulation type. Tables 1-3 show the average magnitude (M), latency (L) and width (W) over the “participants” and repetitions for each of the five models and event types, and can be used to assess the accuracy of each fit. For comparison purposes, the true values imposed by the manipulations are also shown on the bottom row. Finally, Table 4 provides an overall summary of statistical power for estimating both modulated and non-modulated (cross-talk) effects across all the simulations.

For each simulation type (S1-S3) we will discuss the bias present in the estimates of H, T and W, for both event types (A and B), using each of the five different models. Fig 8 shows typical fits for each event, model and simulation type and gives an indication of the apparent biases present in the estimates. We will also discuss the accuracy of each model in estimating A-B effects, the confusability of modulated effects with those that are not modulated, and the power of each method to detect true effects. Below follows a description of the results for each simulation type.

***Simulation 1: Modulation of height***

The results of simulation 1 are summarized in Table 1 and Figures 4, 7 and 8. Truth was an A – B H difference of 0.5, with no modulation of T or W. Table 1 shows the average estimates of the parameters H, T and W for each event type and each model. The means and error bars for each of the 10 “participants” are shown by the vertical lines in the top panel of Figure 4. In the bottom panel, the between-subjects means and standard errors are shown, as would be most relevant for a group analysis. These results are summarized in the bias vs. variance plot appearing in the first row of Figure 7A. The first column of Figure 7B shows the change in bias in the estimated T and W effect that is induced by the change in height. Finally, the first column of Figure 8 shows a typical fit for each model, selected to be representative of the thousands of model fits performed.

When the height of HRF B is modulated, the IL model gives a good overall fit for each event type, though T is slightly underestimated (Table 1). Figure 4 shows that the IL model produces accurate estimates of the A – B height difference. Further, Table 4 shows that the method is second in statistical power to the smooth FIR model. The IL



model also produces the least bias in both T and W (bias is undesirable) for any of the models (Figures 4 and 7, Table 4). Clearly, there is almost no cross-talk present, as both the A-B latency and width effects are non-significant. This can also be seen in the first column of Figure 7B, as the point corresponding to the IL model lies extremely close to the origin.

The NL model effectively estimates the A-B height difference. However, this model has the least statistical power of all included models. In addition, Table 1 shows that both H and W are underestimated for both HRF A and B. In addition, as Figures 4 and 7 show, amplitude modulation induces bias in estimates of T (HRF B is estimated to peak later).

The TD model gives perhaps the best overall estimates of A-B effects, though it is not the most powerful. Table 1 indicates that in the individual fits for HRF A and B, the estimated parameter values for H and T are consistently close to the true values. However, the estimates of W are underestimated for both event types. Table 4 indicates that the TD model, together with the IL model, has the lowest parameter confusability of all the models—i.e., T and W estimates are relatively unaffected by modulation of H, and are not statistically significant. Each of the other three models has some degree of confusability with T and W.

For the FIR model there is a surprisingly strong bias present in the estimate of both T and W, though the bias in T induced by the amplitude change is a fraction of the power to detect changes in H. The bias arises solely from the estimate of HRF B. The model parameters indicate that this method gives rise to an estimated HRF that is taller and has a shorter width and a later peak than the true curve. The estimate of HRF A on

the other hand is extremely accurate, and this model is the most statistically powerful at detecting the A – B height difference.

Finally, the estimate of height for the Gam model is biased for both HRF A and B, but the estimate of A-B is accurate. The bias arises due to the fact that the true width of the underlying HRFs is shorter than the width of the canonical fitted function, which causes the estimate of H to be too low. Note that the blue bars in Fig. 4 imply that no estimate is available for T and W using the Gam model; i.e. both the width and latency are fixed when using a canonical HRF.

### *Simulation 2: Modulation of hemodynamic delay*

Simulation 2 involved a true 3 s difference in T, and no modulation of H and W. The results are summarized in Table 2 and Figures 5, 7 and 8. Table 2 shows the average estimates over the 1000 repetitions for each event type and model. The results for each of the 10 individual “participants” are shown in Figure 5, while the second row of Figure 7A and the second column of Figure 7B show the bias vs. variance and confusability plots, respectively. Finally, the second column of Figure 8 shows a typical fit for each model, selected to be representative of all the model fits performed.

For true changes in T we obtain a good fit with the IL model, with no significant cross-talk present (Figs. 5 and 7, Table 4). The NL model gives a rather accurate estimate for the difference in time-to-peak, but H and W estimates for HRF B are severely corrupted by the delay. Thus, the delayed HRF B has a substantially smaller estimated magnitude, and modulation of T also induces A–B differences in both the estimates of H and W (Fig. 7 and Table 4).

For the TD model, the estimate of the parameters of HRF B is underestimated for both H and T. The shift is too large for this model to handle, as it can only handle shifts of approximately 1 s. Modulation of T induces A–B differences in both the estimates of H and W (Fig. 7 and Table 4).

The FIR model, on the other hand, gives a good overall fit for both event types with the width being slightly underestimated. The estimates of the A-B differences are extremely accurate, with little to no confusability present. In addition, it is the most statistically powerful at detecting the A – B latency difference.

As expected, the Gam model is unable to handle shifts in T, and a strong bias is induced in H. In addition, since the latency and width are fixed, we have no estimate of these components. These results are not surprising, as this is a highly constrained model that is only effective if the true shape is consistent with the model. It is therefore unable to appropriately model shifts in onset or prolonged duration in the underlying signal.

### ***Simulation 3: Modulation of response width***

Finally, Simulation 3 involved a 4 s extension W for condition B, and no modulation of H or T. The results are summarized in Table 3 and Figures 6, 7 and 8. Table 3 shows the average estimates over the 1000 repetitions for each event type and model, while the results for each of the 10 “participants” are shown in Figure 6. The third row of Figure 7A shows bias vs. variance plots and the third column of Figure 7B shows confusability plots. The last column of Figure 8 shows a typical fit for each model, selected to be representative of the thousands of model fits performed.

When the width of HRF B is extended, the IL model produces differences in estimated W (desirable) and T (undesirable). Figure 6 shows that the IL model provides the most accurate estimates of W, and though the power to detect differences in W is second to the smooth FIR model it is substantially greater than the other models. The IL model also shows the least bias in estimates of H and T. It should be noted from studying Table 3 that in the individual fits for HRF A and B, the estimated parameter values are consistently very close to the true values.

With true differences in W, the amplitude estimate of HRF B using the NL model is consistently underestimated, leading to a bias in H for A – B. Estimated differences in T are also created, and these are actually more reliable than estimates of W (Table 4). Since the shape of the gamma density is fixed in this model, the shape can be scaled but not stretched. Hence, the increased width pulls the function away from its true position during the rise, thus delaying the time-to-peak and shortening the width. Thus, true differences in some measures (H, T, and W) are highly confusable, as they induce estimated differences in multiple measures.

For TD the magnitude estimate of HRF B is consistently overestimated. The estimate of T will be clouded by the estimate of width (T is overestimated, W is underestimated). The added width pulls the function away from its true position during the rise, thus delaying the time-to-peak and thereby shortening the width. The model has difficulty detecting the true A-B effect in W. In fact, estimated differences in both H and T are created which are both more reliable than estimates of W.

The FIR model fits the general shape of both event types well, except for the fact that the FIR model has a difficult time modeling the plateau present in event type B. The

plateau has a length of 4 seconds and the time-to-peak is estimated uniformly over the plateau, giving a mean T estimate that overestimates by approximately 2 s. Estimated differences in T were more reliable than estimates of W.

Lastly, as expected, strong bias exists for the Gam model, as this model is unable to handle prolonged duration in the underlying signal.

### ***Application to voxel-wise time courses***

We applied the 5 fitting methods to time courses obtained from individual voxel contained in the visual cortex. Figure 9 shows the results from one representative subject, whose data consisted of 89 separate voxels. Panels B and C show representative fits from an individual voxel and panel A illustrates the consistency of the estimators over the 89 voxels. Consistency is important, as we expect brain responses in these pre-localized regions of the visual cortex to be relatively homogeneous across voxels (which average over ocular dominance columns and other functional features), and so it is likely that much of the variability across voxels in some of the fits is due to error. The results show that the IL model gives the most consistent estimates across the 89 voxels for each of H, T and W.

### ***Relationships between neural activity and activation parameters***

Figure 10A shows a train of brief stimulus events (vertical lines) occurring every 1 s for 18 s, which are intended to serve as a simplified model of neural activity, and the HR shape that is predicted from the (nonlinear) results in Wager et al. (2005). Different task states may change the *magnitude* of neural activity during events, the *onset latency*

for the event train, and/or the *duration* of the event train. If the ‘true’ HR delay predicted by our model varies as a function of changes in true neural magnitude (and so on for other parameter combinations), then the HRF will be of limited usefulness, because it cannot provide information about the type of neuronal change that occurred.

We first deal with the interpretability of H estimates. For brief events, increases in H were caused by either true increases in magnitude or increases in duration. This is because increases in the duration of brief events (Figs. 10D and 10G) tended to translate into changes in HR height. Changes in H for the three types of simulated neuronal effects (increases in magnitude, onset latency, and duration) are shown by the solid lines in Figs. 10 E, F, and G. Conversely, true increases in magnitude did not evoke changes in T or W (Fig. 10B and 10E).

Figure 10B shows HRFs for conditions A and B (solid and dashed lines, respectively) at short and long epoch durations. Figure 10E shows epoch duration on the x-axis, and parameter differences (A – B) on the y-axis; an ideal, unbiased response would be a flat line at 0.5 for H (solid line), and flat lines at zero for T and W (dashed and dotted lines, respectively). That is, magnitude increases produced expected increases in H for brief events, though observing H cannot tell us about whether the magnitude or duration of neuronal activity was different across conditions. For longer epochs, magnitude increases produced increases only in H; the confusability between true

duration and apparent height fell to zero after about 8 s. Thus, the HRF height for brief events is not uniquely interpretable, but the HRF height for longer epochs is<sup>1</sup>.

We next turn to the interpretability of estimates of T. For brief events, changes in T could be caused by true changes in onset (Figs. 10C and 10F) or by changes in duration (Figs. 10D and 10G). This is because duration increases also increased the peak latency. For longer epochs, T changes could be caused by true changes in onset or changes in height (Figs. 10B and 10E). This is because height increases disproportionately affect the early part of the HR (a nonlinear effect not observed with the linear canonical HRF), shifting T earlier for intense stimuli. Thus, T changes are not uniquely interpretable in terms of neuronal latency.

Changes in W, for short epochs, were not reliably evoked by any method; true changes in duration produced the expected changes in W at much reduced levels (Figs. 10D and 10G). Changes in W for all types of simulated neuronal effects are shown by the dotted lines in Figs. E, F, and G. For long epochs, changes in W were produced only by changes in duration, and these appeared to reach their asymptotic true values with a 10 s stimulation epoch (that is, 10 s for condition A and 13 s for condition B in our simulations). Thus, changes in W may be interpreted as changes in neuronal response duration.

---

<sup>1</sup> Note, however, that these results do not necessarily hold for processes with a different neuronal density (e.g., spike bursts every 500 ms instead of 1 s), and they are presented mainly for illustrative purposes here.

## DISCUSSION

To date most fMRI studies have been primarily focused on estimating the magnitude of evoked HRFs across different tasks. However, there is a growing interest in testing other statistics as well, such as the time-to-peak and duration of activation (Bellgowan, Saad, & Bandettini, 2003; Formisano & Goebel, 2003; Richter et al., 2000). The onset and peak latencies of the HRF can, for instance, provide information about the timing of activation for various brain areas and the width of the HRF provides information about the duration of activation. However, the independence of these parameter estimates has not been properly assessed, as it appears that even if basis functions are independent (or a nonlinear fitting procedure provides nominally independent estimates), the parameter estimates from real data may not be independent.

The present study seeks to both bridge this gap in the literature and present a new estimation method based on the use of inverse logit functions. To assess independence, we determine the amount of confusability between estimates of height (H), time-to-peak (T) and full-width at half-maximum (W) and actual manipulations in the amplitude, time-to-peak and duration of the stimulus. This was investigated using a simulation study that was based on empirical HRFs and illustrated how a variety of popular methods work on actual fMRI data. It is important to note that this is not an exhaustive survey of HRF fitting methods, and some very promising linear methods are not addressed in our simulations (e.g., Liao (Liao et al., 2002); Henson (Henson, Price, Rugg, Turner, & Friston, 2002)). In addition, Ciuciu *et al.* (2003) has introduced an unsupervised FIR model which estimates its parameters using an EM-type algorithm. This promising



approach may potentially improve on the fit of the smoothed (supervised) FIR used in this paper, and decrease the amount of confusability present in that model.

In this work we identified the interpretability of parameter estimates and statistical power to detect true effects as two important criteria for a modeling system. Our results show that with any of the models we tested, there is some degree of confusability between true differences and estimates. With some models, the confusability is profound. For example, delaying the onset of activation by 3 s produced highly reliable changes in estimated response magnitude in most models tested. Even models that attempt to account for delay such as a gamma function with nonlinear fitting (Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000) or temporal and dispersion derivatives (Calhoun, Stevens, Pearlson, & Kiehl, 2004; Friston, Josephs, Rees, & Turner, 1998) showed strong biases. As might be expected, the derivative models and related methods (e.g., Liao (Liao et al., 2002); Henson (Henson, Price, Rugg, Turner, & Friston, 2002)) may be quite accurate for very short shifts in latency ( $< 1$  s) but become progressively more inaccurate as the shift increases. The IL model and the smooth FIR model did not show large biases, and the IL model showed by far the least amount of confusability of all the models that were examined.

The strongest biases were found for all models when the response width was manipulated by extending the HRF at its peak by 4 s. No model was bias-free, but the IL model showed no bias in H and only a slight bias in T (Table 4). This feature may be useful in comparing task conditions that have processes that are extended in time over a number of seconds, such as working memory and expectation/anticipation paradigms and tasks with long separation between phases of trials (e.g., cue – target). Thus, the FIR

model sacrifices some interpretability, particularly in dealing with prolonged stimulation periods, for the benefit of power. It may be an excellent choice for modeling shorter-duration events, whereas the IL model may fare better with longer and more variable epochs. In fact, the ability to model both events and extended epochs is a design feature that motivated our development of the IL model.

Notably, the smooth FIR model had the highest power for estimating true effects of all the models (Table 4). The canonical HRF did not fare well because the empirical HRFs on which our study was based tended to peak earlier than the canonical HRF, and because individual differences in the shape and timing of activity were translated into differences in H. The IL and smooth FIR models can account for individual differences in timing and delay without affecting H, which increases power in H estimation. The nonlinear gamma and derivative-based models have a limited ability to do this, and power is lower on average across H and T estimates. Interestingly, the derivative model has high power for estimating H but not T, and vice versa for the nonlinear gamma model. The IL and smooth FIR models are both consistently high in power and less biased than either of the other methods, with the FIR model having higher power, but increased bias compared to the IL model. As for the individual model fits, both the FIR and IL models are able to accurately fit HRF A (Tables 1-3). However, the IL model is far more effective at modeling HRF B in all three simulation types, and thereby gives rise to less cross-talk than the FIR model.

*Relationships between neural activity and activation parameters*

As mentioned in the introduction, problems with parameter interpretability can come from two major sources. This paper addresses the simpler issue of whether differences in evoked HRF shape can be accurately captured by a variety of linear models. The best models (IL and smooth FIR) were able to accurately capture changes in HRs with high sensitivity and specificity; that is, changes in one estimate were seldom confused for another. Ultimately, researchers may want to interpret parameter changes in terms of underlying neuronal activity. This is a much more complex problem that involves building physiological models of the sources of BOLD signal (Buxton, Wong, & Frank, 1998; Logothetis, 2003; Mechelli, Price, & Friston, 2001; Vazquez & Noll, 1998; Wager, Vazquez, Hernandez, & Noll, 2005).

Based on preliminary analysis using a simple nonlinear model (Wager 2005) it appears that estimated latency differences are not uniquely attributable to neuronal onset delays, but could be caused by true differences in firing rate, delay or duration. Estimated width differences may generally be attributable to increases in the duration of neuronal activity. For brief events, estimated height differences could be caused by either duration increases or activity magnitude increases. For longer epochs ( $> 8$  s) estimated height differences are caused only by increases in firing rate. These results do not render models of the HRF useless; finding differences in HRF time to peak among conditions would constitute scientific evidence that may correspond with behavioral performance or distinguish the responses of one brain region from another. In addition, finding a significant difference in T but no difference in W (for brief events) or no difference in H (for long events) may be sufficient evidence to make a claim about differences in

neuronal onset latency. Other combinations of significant results may be similarly interpretable depending on the specifics of the study.

However, this simulation has many limitations, including that it does not attempt to model physiological parameters, and second, that the nonlinearity estimates used do not take into account differences in stimulation density. In these simulations, all models use trains of brief stimuli repeated at 1 s intervals, consistent with the density used in the experiments from which the nonlinearity estimates were derived (Wager et al., 2005). In addition, the nonlinear model here provides a rough characterization of nonlinearities, which may vary both with brain region and with task. Thus, these results are suggestive, but cannot provide definitive guidelines on the complex issue of how evoked HRF shapes may be related to underlying neuronal activity.

### ***Choosing a hemodynamic response model***

When determining which HRF model to use, the first question one is faced with is how strongly assumptions should be made *a priori*. Models with few assumptions and many variable parameters have the flexibility to model a large variety of shapes and are able to handle unexpected behavior in the underlying response. However, as the number of parameters in the model increases, the number of degrees of freedom in the statistical tests of the parameters decreases. In addition, it is also much simpler and more statistically powerful to test contrasts across event types (e.g. A – B) on a single parameter such as height than it is to test for differences in multiple parameters (e.g. A1A2A3 – B1B2B3). An ANOVA F-test will accomplish the goal of testing for multiple parameters, but the statistical power of the test decreases sharply as a function of the

number of parameters included in the test, and then the problem remains of interpreting which parameters are carrying the difference.

Critically, free parameters in most flexible basis sets are not directly interpretable (e.g. as the response magnitude or latency). Consider, for example, the TD model. Let us denote  $A_1$  and  $B_1$  as the responses to the canonical HRF for conditions A and B,  $A_2$  and  $B_2$  the temporal derivatives, and  $A_3$  and  $B_3$  the dispersion derivatives. One cannot simply fit the basis set and compute the contrast  $A_1 - B_1$ , ignoring the other parameters, and interpret the result as the difference in magnitude between A and B. This is because the amplitude of the fitted response depends on a combination of all three parameters, and so each one is only interpretable in the context of the others.

This suggests that perhaps using a single canonical HRF may be the best choice. If, in fact, the actual shape of the HRF matches the model perfectly and that the shape is invariant across the brain, using a single canonical HRF offers optimal power. However, it is reasonable to assume that the shape of the HRF varies as a function of both task and brain region, and therefore any fixed model will undoubtedly to be wrong in much of the brain, and will be wrong to different degrees across individuals. If the model is incorrectly specified, then statistical power decreases and the model may also produce invalid and biased results, as was shown in our study. As is well known in statistics, the fact that a linear model explains a significant amount of the variance in the data is no guarantee that the underlying model is correct. For example, imagine that one conducts an experiment with trials spaced 15 s apart. A canonical HRF such as that used in SPM, consisting of a positive-going gamma function peaking at 6 s and a negative-going gamma function peaking at 16 s, is used to model the response at the onset of each trial.

Now imagine that a particular brain region shows activity increases not in response to the trial onset, but in the inter-trial interval in preparation for the predictable onset of the next trial. Such a region would be likely to show a *negative* activation, leading the researchers to erroneously infer that the region was deactivated by the task. In fact, in our example, it is activated in anticipation of the task. Such potential problems require the checking of assumptions, including that the model is correctly specified, which is difficult to do in brain imaging due to the massive number of tests involved (though methods have been developed; (Luo & Nichols, 2003)). Finally, using a canonical HRF provides no way to assess latency and duration and differences between conditions in response latency will be confused for differences in amplitude (Calhoun, Stevens, Pearlson, & Kiehl, 2004).

In this work we introduced a new HRF modeling technique, based on the superposition of three inverse logit functions, which attempts to balance flexibility and ease of interpretation. Our study showed the efficiency of the fitting procedure compared with four other commonly used models. In particular, the IL model was by far the most effective at modeling the combination of HRF types A and B for each of the three types of simulations, and therefore gave rise to significantly less cross-talk than the other models. The mayor drawback of our method is that it is relatively time-consuming using a non-linear fitting procedure. The ultimate speed of the IL model will depend on whether deterministic (e.g. Gauss-Newton, L-M algorithms) or stochastic (simulated annealing) are used. The deterministic algorithms take on the order of 5 times longer than the FIR model, while the simulated annealing algorithm roughly doubles that time. As an alternative to non-linear least-squares fitting, one could instead use *a priori* knowledge to specify each parameter in the model, except for the three amplitude parameters, and use

the three resulting inverse logit functions as temporal basis functions in the GLM framework. Alternatively, one could follow the methodology outlined in Woolrich et. al. (Woolrich, Behrens, & Smith, 2004) and generate a large number of plausible HRF shapes, by randomly sampling values for the parameters from an appropriate range. Using singular value decomposition one can thereafter find the optimal basis set that spans the space of generated functions and use this set as our temporal basis functions.

## CONCLUSIONS

In this work, we introduce a new technique for modeling the HRF, based on the superposition of three inverse logit functions (IL), which balances the need for interpretability and flexibility of the model. In simulations based on actual HRFs, measured on a group of 10 participants, we compare the performance of this model to four other popular choices of basis functions. We show that the IL model can capture magnitude, delay, and duration of activation with less error than the other methods tested, and therefore provides a promising way to flexibly but powerfully test the magnitude and timing of activation across experimental conditions.

## ACKNOWLEDGEMENTS

The authors wish to thank Andrew Gelman for his helpful suggestions in preparing this manuscript.

## APPENDIX

### *Conditions to ensure minimal overlap between the IL functions*

The interpretability of the parameters in the IL model are increased if the first and second and the second and third IL functions are made as orthogonal as possible to one another. This implies that the rise in the first function needs to stabilize prior to the decrease in the second function. In principal the first function will not reach its maximum value of 1 until  $t = \infty$ . However, one can set a constraint to the effect that the first function needs to complete 99% of its rise prior to the second function completing 1% of its decrease, i.e. assuming  $L_1(t_1) = 0.99$  and  $L_2(t_2) = 0.01$  then we need to derive constraints that ensure that  $t_1 < t_2$  holds.

To find these constraints we need to re-express  $t_1$  and  $t_2$  in terms of the parameters of the model. Define  $t_1$  as the time point when  $L_1(t_1) = c$ , where  $c = 0.99$  in the example above, but can reasonably be set to take other values as well. This implies that,

$$\left(1 + \exp\left(-\frac{(t_1 - T_1)}{D_1}\right)\right) = \frac{1}{c} \quad (30)$$

Through simple algebra, this equation can be rewritten as:

$$t_1 = T_1 + D_1 \log\left(\left(\frac{1}{c} - 1\right)^{-1}\right)$$



$$= T_1 + D_1 k \quad (31)$$

where  $k = \log\left(\left(c^{-1} - 1\right)^{-1}\right)$ .

In a similar manner we can rewrite  $t_2$  as,

$$t_2 = T_2 - D_2 k . \quad (32)$$

Combining these two expressions, the condition  $t_1 < t_2$  can be written as

$$T_2 - T_1 > (D_1 + D_2)k \quad (33)$$

Using exactly the same reasoning an equivalent condition for minimizing the overlap between the second and third IL function is given by,

$$T_3 - T_2 > (D_2 + D_3)k . \quad (34)$$

### ***Parameter estimates***

Assuming the two constraints (33) and (34) hold, the estimates for height, time-to-peak and width are easily expressed as functions of the parameters of the model.

#### ***(i) Height***

Assuming  $c \approx 1$ , the first and second IL function will have minimal overlap and the height can be reasonably estimated as the amplitude of the first logit function, i.e.

$$H = \delta_1.$$

(ii) *Time-to-peak*

Again, assuming  $c \approx 1$ , the time-to-peak can be estimated, using Eq. 31, as

$$T = T_1 - D_1 k.$$

(iii) *Width*

To find the full-width at half-maximum, we need to determine (a) the time point when the first IL function reaches half of its height and (b) the time point when the second IL function crosses  $0.5\delta_1$ . The time point (a) is simply given by  $T_1$ , so the problem boils down to finding time point (b), i.e. we want to find the time  $t^*$  when

$\delta_2 L_2(t^*) = 0.5\delta_1$ . This implies that,

$$\left(1 + \exp\left(-\frac{(t^* - T_2)}{D_2}\right)\right)^{-1} = \frac{\delta_1}{2\delta_2} \quad (35)$$

which can be rewritten as:

$$t^* = T_2 - D_2 \log\left(\frac{2\delta_2}{\delta_1} - 1\right) \quad (36)$$

Hence, the FWHM is the distance between  $t^*$  and  $T_1$ , i.e.

$$\begin{aligned} W &= t^* - T_1 \\ &= T_2 - T_1 - D_2 \log\left(\frac{2\delta_2}{\delta_1} - 1\right). \end{aligned} \quad (37)$$

## REFERENCES

- Aguirre, G. K., Singh, R., & D'Esposito, M. (1999). Stimulus inversion and the responses of face and object-sensitive cortical areas. *Neuroreport*, *10*(1), 189-194.
- Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *Neuroimage*, *8*(4), 360-369.
- Bellgowan, P. S., Saad, Z. S., & Bandettini, P. A. (2003). Understanding neural system dynamics through task modulation and measurement of functional MRI amplitude, latency, and width. *Proc Natl Acad Sci U S A*, *100*(3), 1415-1419.
- Birn, R. M., Saad, Z. S., & Bandettini, P. A. (2001). Spatial heterogeneity of the nonlinear dynamics in the fmri bold response. *Neuroimage*, *14*(4), 817-826.
- Buckner, R. L. (2003). The hemodynamic inverse problem: making inferences about neural activity from measured MRI signals. *Proc Natl Acad Sci U S A*, *100*(5), 2177-2179.
- Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med*, *39*(6), 855-864.
- Calhoun, V. D., Stevens, M. C., Pearlson, G. D., & Kiehl, K. A. (2004). fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *Neuroimage*, *22*(1), 252-257.
- Ciuciu, P., Poline, J.-B., Marrelec, G., Idier, J., Pallier, Ch. and Benali, H. (2003). Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Trans. Med. Imag.*, *22*(10):1235--1251.

- Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping, 5*, 329-340.
- Formisano, E., & Goebel, R. (2003). Tracking cognitive processes with functional MRI mental chronometry. *Curr Opin Neurobiol, 13*(2), 174-181.
- Formisano, E., Linden, D. E., Di Salle, F., Trojano, L., Esposito, F., Sack, A. T., et al. (2002). Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. *Neuron, 35*(1), 185-194.
- Friston, K. J., Josephs, O., Rees, G., & Turner, R. (1998). Nonlinear event-related responses in fMRI. *Magn Reson Med, 39*(1), 41-52.
- Gelman, A. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, Fla.: Chapman & Hall/CRC.
- Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage, 9*(4), 416-429.
- Goutte, C., Nielsen, F. A., & Hansen, L. K. (2000). Modeling the haemodynamic response in fMRI using smooth FIR filters. *IEEE Trans Med Imaging, 19*(12), 1188-1201.
- Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage, 21*(4), 1639-1651.
- Henson, R. N., Price, C. J., Rugg, M. D., Turner, R., & Friston, K. J. (2002). Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. *Neuroimage, 15*(1), 83-97.

- Hernandez, L., Badre, D., Noll, D., & Jonides, J. (2002). Temporal sensitivity of event-related fMRI. *Neuroimage*, *17*(2), 1018-1026.
- Jain, R. K. (1985). Ridge regression and its application to medical data. *Comput Biomed Res*, *18*(4), 363-368.
- Kirkpatrick, S., Gerlatt, C. D. Jr., and Vecchi, M.P., (1983). Optimization by Simulated Annealing, *Science*, *220*, 4598, 671-680.
- Kruggel, F., & von Cramon, D. Y. (1999). Temporal properties of the hemodynamic response in functional MRI. *Hum Brain Mapp*, *8*(4), 259-271.
- Kruggel, F., Wiggins, C. J., Herrmann, C. S., & von Cramon, D. Y. (2000). Recording of the event-related potentials during functional MRI at 3.0 Tesla field strength. *Magn Reson Med*, *44*(2), 277-282.
- Lange, N. and Zeger, S. L. (1997). Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with discussion). *Applied Statistics, Journal of the Royal Statistical Society, Series C*, *46*(1):1-29.
- Liao, C. H., Worsley, K. J., Poline, J. B., Aston, J. A., Duncan, G. H., & Evans, A. C. (2002). Estimating the delay of the fMRI response. *Neuroimage*, *16*(3 Pt 1), 593-606.
- Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *J Neurosci*, *23*(10), 3963-3971.
- Luo, W. L., & Nichols, T. E. (2003). Diagnosis and exploration of massively univariate neuroimaging models. *Neuroimage*, *19*(3), 1014-1032.

- Maccotta, L., Zacks, J. M., & Buckner, R. L. (2001). Rapid self-paced event-related functional MRI: feasibility and implications of stimulus- versus response-locked timing. *Neuroimage*, *14*(5), 1105-1121.
- Marrelec, G., Benali, H., Ciuciu, P., Pelegrini-Issac, M., & Poline, J. B. (2003). Robust Bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. *Hum Brain Mapp*, *19*(1), 1-17.
- Mechelli, A., Price, C. J., & Friston, K. J. (2001). Nonlinear coupling between evoked rCBF and BOLD signals: a simulation study of hemodynamic responses. *Neuroimage*, *14*(4), 862-872.
- Menon, R. S., Luknowsky, D. C., & Gati, J. S. (1998). Mental chronometry using latency-resolved functional MRI. *Proc Natl Acad Sci U S A*, *95*(18), 10902-10907.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M. N., Teller, A.H. and Teller, E., (1953). Equation of State Calculations by Fast Computing Machines, *J. Chem. Phys.*, *21*, 6, 1087-1092.
- Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., & Buckner, R. L. (2000). Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage*, *11*(6 Pt 1), 735-759.
- Noll, D. C., Cohen, J. D., Meyer, C. H., & Schneider, W. (1995). Spiral K-space MR imaging of cortical activation. *J Magn Reson Imaging*, *5*(1), 49-56.

- Ollinger, J. M., Shulman, G. L., & Corbetta, M. (2001). Separating processes within a trial in event-related functional MRI. *Neuroimage*, *13*(1), 210-217.
- Rajapakse, J. C., Kruggel, F., Maisog, J. M., & von Cramon, D. Y. (1998). Modeling hemodynamic response for analysis of functional MRI time-series. *Hum Brain Mapp*, *6*(4), 283-300.
- Richter, W., Somorjai, R., Summers, R., Jarmasz, M., Menon, R. S., Gati, J. S., et al. (2000). Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *J Cogn Neurosci*, *12*(2), 310-320.
- Riera, J. J., Watanabe, J., Kazuki, I., Naoki, M., Aubert, E., Ozaki, T., et al. (2004). A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signals. *Neuroimage*, *21*(2), 547-567.
- Saad, Z. S., DeYoe, E. A., & Ropella, K. M. (2003). Estimation of fMRI response delays. *Neuroimage*, *18*(2), 494-504.
- Vazquez, A. L., & Noll, D. C. (1998). Nonlinear aspects of the BOLD response in functional MRI. *Neuroimage*, *7*(2), 108-118.
- Wager, T. D., Vazquez, A., Hernandez, L., & Noll, D. C. (2005). Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. *Neuroimage*, *25*(1), 206-218.
- Woolrich, M. W., Behrens, T. E., & Smith, S. M. (2004). Constrained linear basis sets for HRF modelling using Variational Bayes. *Neuroimage*, *21*(4), 1748-1761.
- Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited--again. *Neuroimage*, *2*(3), 173-181.



Zarahn, E. (2002). Using larger dimensional signal subspaces to increase sensitivity in fMRI time series analyses. *Hum Brain Mapp*, 17(1), 13-16.

**Table 1.**

|             | Event Type A  |               |               | Event Type B  |               |               |
|-------------|---------------|---------------|---------------|---------------|---------------|---------------|
|             | H             | T             | W             | H             | T             | W             |
| <b>IL</b>   | 1.0257        | 4.8695        | 5.0287        | 0.5006        | 4.8723        | 5.0165        |
| <b>NL</b>   | 0.9461        | 4.9180        | 4.5268        | 0.4229        | 5.1465        | 4.3002        |
| <b>TD</b>   | 0.9952        | 4.8641        | 4.5236        | 0.4859        | 4.9456        | 4.4780        |
| <b>FIR</b>  | 1.0116        | 5.0860        | 4.9078        | 0.5479        | 5.5385        | 4.3674        |
| <b>Gam</b>  | 0.9401        | 5.5000        | 5.5000        | 0.4776        | 5.5000        | 5.5000        |
| <b>True</b> | <b>1.0000</b> | <b>5.0000</b> | <b>5.0000</b> | <b>0.5000</b> | <b>5.0000</b> | <b>5.0000</b> |

Note. Simulation 1 - The average height (H), time-to-peak (T) and width (W) over all the “participants” and repetitions for each of the five models and event-types together with the true values.

**Table 2.**

|             | Event Type A  |               |               | Event Type B  |               |               |
|-------------|---------------|---------------|---------------|---------------|---------------|---------------|
|             | H             | T             | W             | H             | T             | W             |
| <b>IL</b>   | 0.9978        | 4.9135        | 5.0092        | 0.9942        | 8.0700        | 5.0521        |
| <b>NL</b>   | 0.9723        | 4.6631        | 4.4406        | 0.7706        | 8.0635        | 3.9813        |
| <b>TD</b>   | 0.9700        | 4.8016        | 4.4446        | 0.9003        | 7.1894        | 4.9835        |
| <b>FIR</b>  | 1.0114        | 5.0756        | 4.9192        | 1.0142        | 8.0786        | 4.8996        |
| <b>Gam</b>  | 0.9887        | 5.5000        | 5.5000        | 0.7074        | 5.5000        | 5.5000        |
| <b>True</b> | <b>1.0000</b> | <b>5.0000</b> | <b>5.0000</b> | <b>1.0000</b> | <b>8.0000</b> | <b>5.0000</b> |

Note. Simulation 2 - The average height (H), time-to-peak (T) and width (W) over all the “participants” and repetitions for each of the five models and event-types together with the true values.

**Table 3.**

|             | Event Type A  |               |               | Event Type B  |               |               |
|-------------|---------------|---------------|---------------|---------------|---------------|---------------|
|             | H             | T             | W             | H             | T             | W             |
| <b>IL</b>   | 1.0157        | 4.9183        | 4.7790        | 0.9969        | 5.2824        | 8.9002        |
| <b>NL</b>   | 0.9476        | 4.6632        | 4.5472        | 0.8157        | 6.5994        | 6.0840        |
| <b>TD</b>   | 1.0016        | 4.7811        | 4.4341        | 1.2410        | 6.1079        | 5.3303        |
| <b>FIR</b>  | 1.0092        | 5.0823        | 4.9243        | 1.1023        | 7.0621        | 8.5430        |
| <b>Gam</b>  | 0.9786        | 5.5000        | 5.5000        | 1.2573        | 5.5000        | 5.5000        |
| <b>True</b> | <b>1.0000</b> | <b>5.0000</b> | <b>5.0000</b> | <b>0.5000</b> | <b>5.0000</b> | <b>9.0000</b> |

Note. Simulation 3 - The average height (H), time-to-peak (T) and width (W) over all the “participants” and repetitions for each of the five models and event-types together with the true values.

Table 4.

| <u>True A-B effect</u> | <u>Power (at p &lt; .0001)</u><br>Inverse Logit estimates |              |              | <u>Estimated effects (t-values)</u><br>Inverse Logit estimates |              |              |
|------------------------|---|--------------|--------------|--|--------------|--------------|
|                        | <i>Height</i>   | <i>Delay</i> | <i>Width</i> | <i>Height</i>  | <i>Delay</i> | <i>Width</i> |
| S1: <i>Height</i>      | 1.00  |              |              | 69.27  | n.s.         | n.s.         |
| S2: <i>Delay</i>       |   | 1.00         |              | n.s.   | -58.33       | n.s.         |
| S3: <i>Width</i>       |   | 0.05         | 1.00         | n.s.   | -4.14        | -42.36       |
|                        | <b>Nonlinear Gamma Estimates</b>                          |              |              | <b>Nonlinear Gamma Estimates</b>                               |              |              |
|                        | <i>Height</i>   | <i>Delay</i> | <i>Width</i> | <i>Height</i>  | <i>Delay</i> | <i>Width</i> |
| S1: <i>Height</i>      | 1.00  | 0.05         |              | 15.55  | -4.13        | n.s.         |
| S2: <i>Delay</i>       | 0.95  | 1            |              | 7.88   | -69.66       | 3.95         |
| S3: <i>Width</i>       | 0.17  | 1.00         | 0.99         | 5  | -22.08       | -8.73        |
|                        | <b>TD Estimates</b>                                       |              |              | <b>TD Estimates</b>  |              |              |
|                        | <i>Height</i>   | <i>Delay</i> | <i>Width</i> | <i>Height</i>  | <i>Delay</i> | <i>Width</i> |
| S1: <i>Height</i>      | 1.00  |              |              | 51.76  | n.s.         | n.s.         |
| S2: <i>Delay</i>       | 0.44  | 1.00         |              | 5.86   | -15.71       | -2.91        |
| S3: <i>Width</i>       | 1.00  | 1.00         | 0.12         | -13.5  | -13.18       | -4.75        |
|                        | <b>Smooth FIR Estimates</b>                               |              |              | <b>Smooth FIR Estimates</b>                                    |              |              |
|                        | <i>Height</i>   | <i>Delay</i> | <i>Width</i> | <i>Height</i>  | <i>Delay</i> | <i>Width</i> |
| S1: <i>Height</i>      | 1.00  |              | 0.22         | 192.72   | -3.63        | 5.2          |
| S2: <i>Delay</i>       |   | 1.00         |              | n.s.   | -188.84      | n.s.         |
| S3: <i>Width</i>       | 1.00  | 1.00         | 1.00         | -46.17   | -69.8        | -65.95       |
|                        | <b>Gamma Estimates</b>                                    |              |              | <b>Gamma Estimates</b>   |              |              |
|                        | <i>Height</i>   | <i>Delay</i> | <i>Width</i> | <i>Height</i>  | <i>Delay</i> | <i>Width</i> |
| S1: <i>Height</i>      | 1.00  | N/A          | N/A          | 24.47  | N/A          | N/A          |
| S2: <i>Delay</i>       | 0.97  | N/A          | N/A          | 8.24   | N/A          | N/A          |
| S3: <i>Width</i>       | 0.64  | N/A          | N/A          | -6.38  | N/A          | N/A          |

Note. An overall summary of statistical power for estimating both modulated and non-modulated (cross-talk) effects across all the simulations. Power estimates for detecting A - B differences at  $p < .0001$  are shown in the left columns, and average t-values for A - B

estimates are shown in the right columns. For clarity of presentation, cells with power < 5% are left empty. The absolute magnitudes of the t-values and power estimates depend on the signal-to-noise ratio in the simulations, but it is informative to compare across analysis types and to assess whether modulations in some parameters reliably induce effects in other parameters. The diagonal elements show the power for estimates (columns) when the corresponding effect is modulated (rows). High power in these diagonal elements indicates more sensitivity to experimental effects. The off-diagonal elements show power in estimates when other effects are modulated. High power in these elements is undesirable, as it indicates bias in the estimates that decreases the interpretability of parameter estimates. n.s., not significant at  $p < .05$  uncorrected.

## Figure Captions

**Figure 1:** Estimates of response height (H), time-to-peak (T), and full-width at half-max (W) from a simulated HRF.

**Figure 2:** Relationship between neural activity, evoked changes in the BOLD response, and estimated parameters. Solid lines indicate expected relationships, and dashed lines indicate relationships that, if they exist, create problems in interpreting estimated parameters. For task-induced changes in estimated time-to-peak to be interpretable in terms of the latency of neural firing, for example, estimated time-to-peak must vary only as a function of changes in neural firing onsets, not firing rate or duration. The relationship between neural activity and true BOLD responses determines the *theoretical* limits on how interpretable the parameter estimates are. The relationship between true BOLD changes and estimated BOLD changes using a model introduce additional model-dependent constraints on the interpretability of parameter estimates.

**Figure 3:** The functions  $\delta L(A(t-T))$  with parameters: (A)  $\delta = 1.0$ ,  $T = 15$  and  $A = 0.75$ , (B)  $\delta = -1.3$ ,  $T = 27$  and  $A = 0.4$  and (C)  $\delta = 0.3$ ,  $T = 66$  and  $A = 0.5$ . (D) The three functions in (A)-(C) superimposed (bold line) together with actual HRF function (Dotted line).

**Figure 4:** Results for Simulation 1 - In the top row, the true effects are shown by horizontal lines, and means and error bars for each of the 10 “participants” are shown by

the vertical lines. In the bottom panels the between subject means and standard errors are shown. The blue bars imply that no estimate is available for T and W using the Gam model.

**Figure 5:** Results for Simulation 2 - In the top row, the true effects are shown by horizontal lines, and means and error bars for each of the 10 “participants” are shown by the vertical lines. In the bottom panels the between subject means and standard errors are shown.

**Figure 6:** Results for Simulation 3 - In the top row, the true effects are shown by horizontal lines, and means and error bars for each of the 10 “participants” are shown by the vertical lines. In the bottom panels the between subject means and standard errors are shown.

**Figure 7:** (A) Bias vs. variance plots for the estimated A-B difference. Each row represents a simulation (S1 – S3) and each column represents an estimated parameter (H, T and W). (B) Scatter plots of the change in bias for the two non-modulated parameters, induced by the change in the modulated parameter, for each simulation type. For clarity the point (0,0) is marked as the cross between the dotted lines in the x and y-axis. Points that lie close to the origin imply that the method induces little confusability.



**Figure 8:** Typical fits for IL, NL, TD, FIR and Gam (Rows 1-5 respectively) for simulations S1, S2 and S3 (Columns 1-3 respectively) are shown in bold, while the underlying empirical HRFs are depicted using dotted lines.

**Figure 9:** Results from an application of the 5 fitting procedures to 89 single voxel time courses. (A) The means and error bars for the estimates of H, T and W for each of the 5 methods. (B) HRF estimates for each method extracted from a representative single voxel time course. (C) The model fit using the IL method extracted from a representative single voxel time course.

**Figure 10:** Exploration of the relationship between changes in trains of neural events and changes in height (H), time to peak (T), and width (W) of activation. A) a train of events (18 s. one burst of simulated neural activity per second) and the predicted activation accounting for nonlinear neuro/vascular responses (see text for details). Each event represents a collection of action potentials that occur in response to a cognitive event. Analyses were conducted using a linear activation model as well, but are less physiologically plausible and are not shown for space reasons. B) Effects of increasing the amplitude of neural events, a proxy for neural firing rate. For short-duration (3 s) and long-duration (18 s) trains, both H and W are affected to some degree. C) Increasing the onset latency of event trains affected only T, but not H or W. D) Increasing the duration of event trains affected H, T, and W for short trains (3 to 6 s durations), but only affected W for long trains (12 to 15 s durations and longer). Thus, W is most interpretable for long stimulation epochs (> 12 s), but may reflect increases in either duration or intensity

of firing. These results are illustrative rather than exhaustive, and all activation parameters should be interpreted with caution. E-G) Parameter differences (A-B) for H, T and W for each of the three types of simulated neuronal effects. Each figure shows epoch duration on the x-axis, and parameter differences on the y-axis.

Figure 1

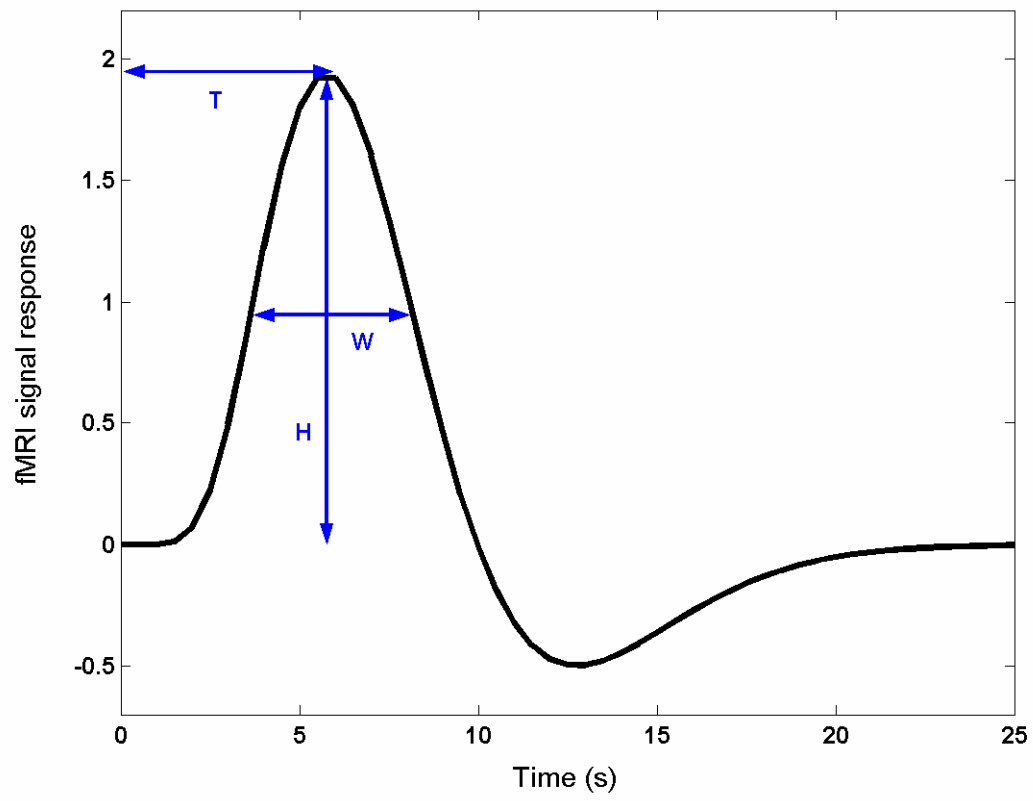


Figure 2

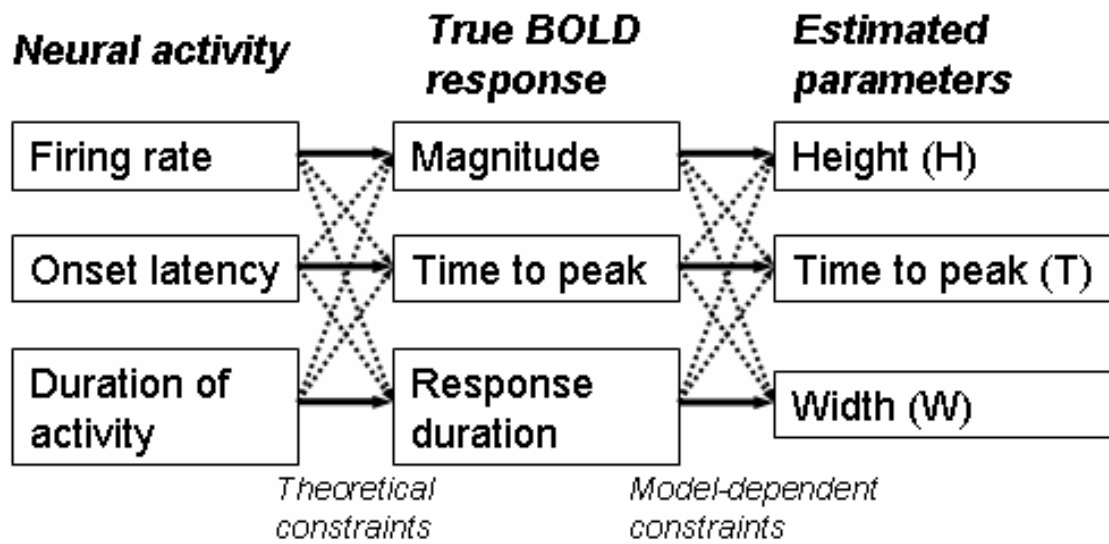


Figure 3

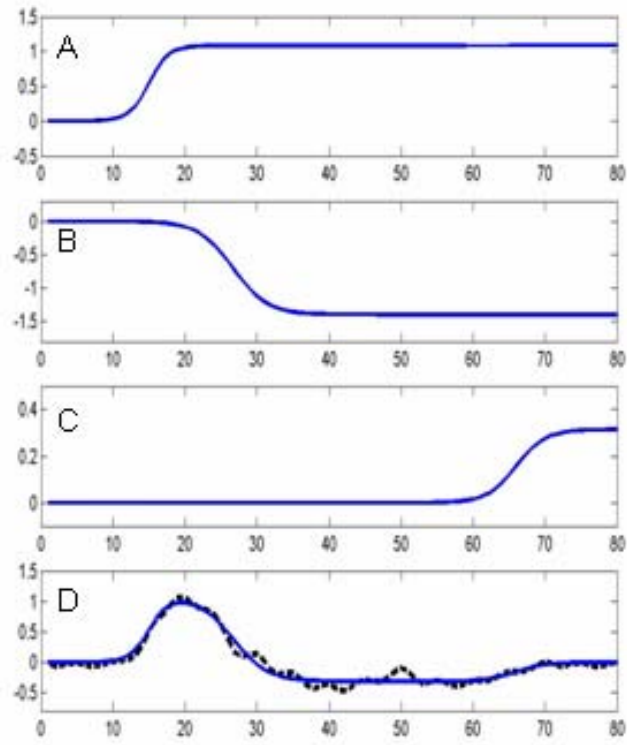


Figure 4

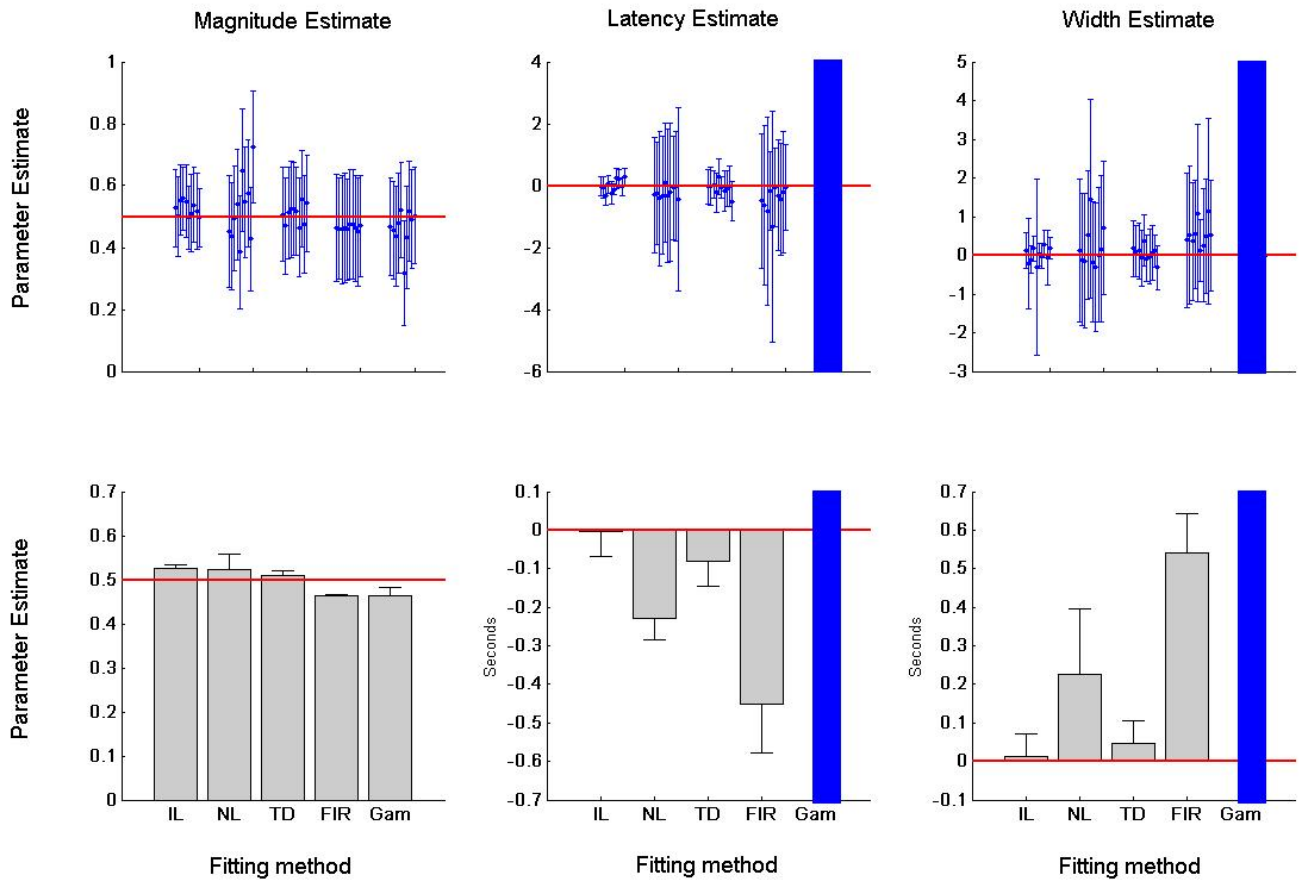


Figure 5

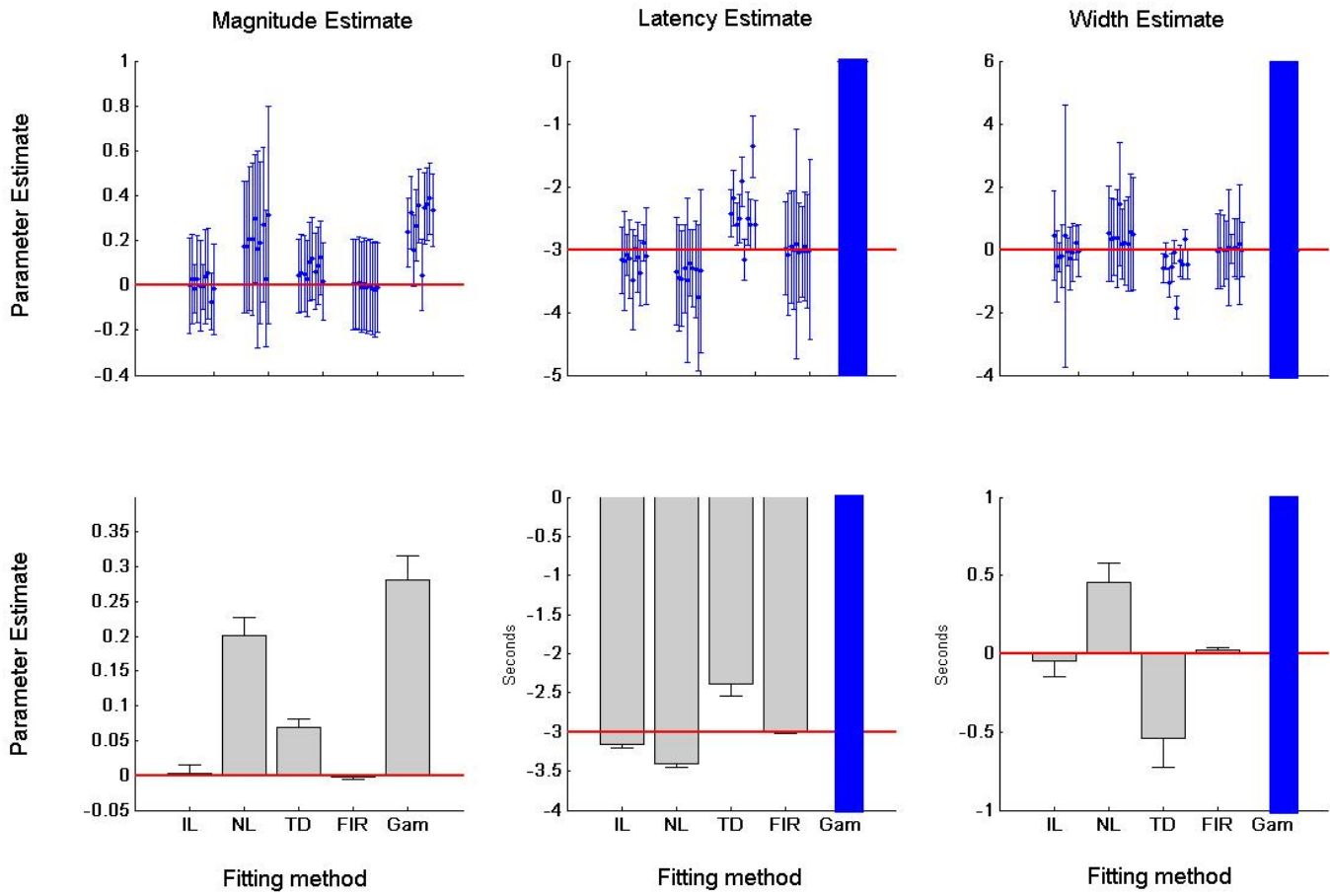


Figure 6

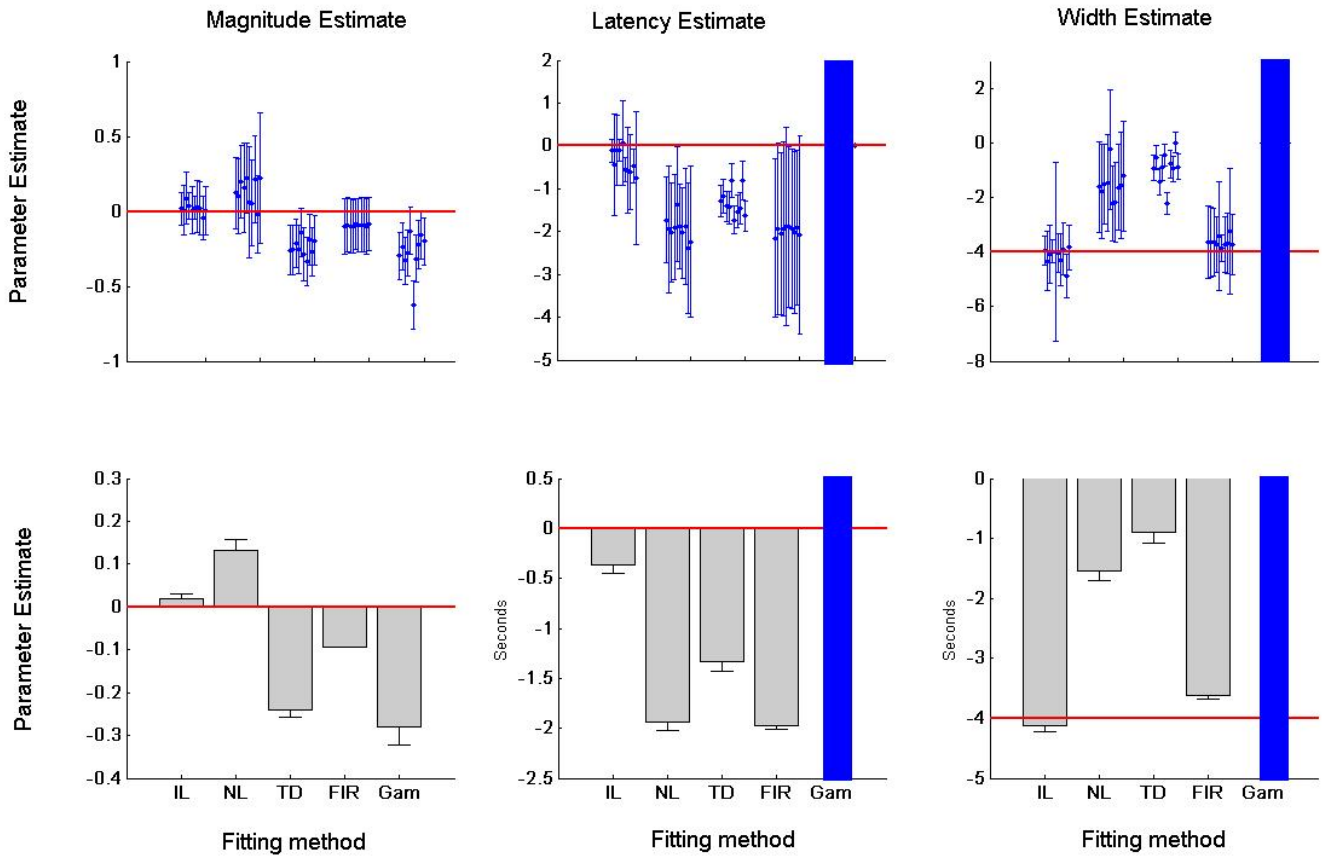
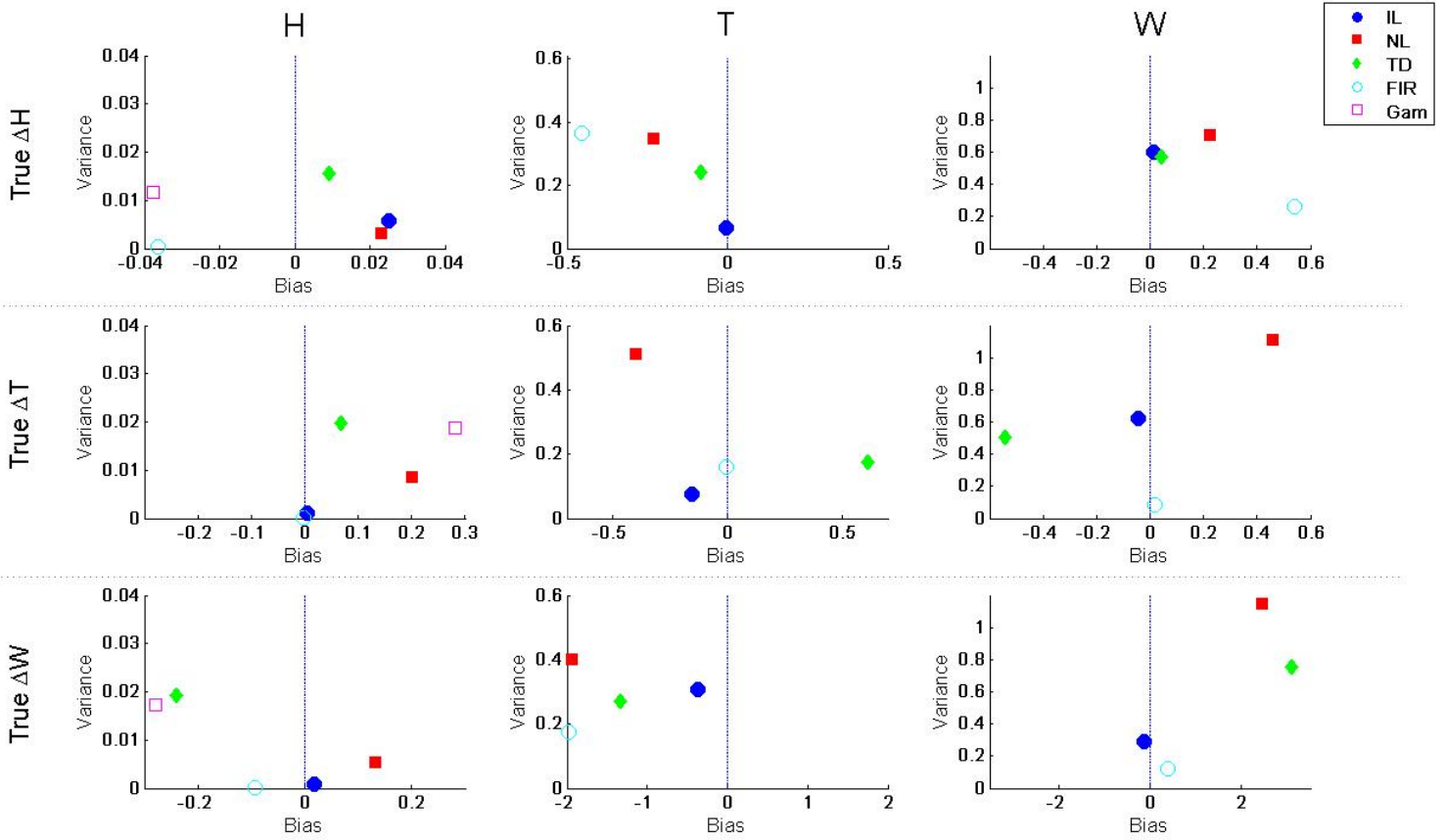




Figure 7

**A Bias vs. Variance Plots**



**B Confusability Plots**

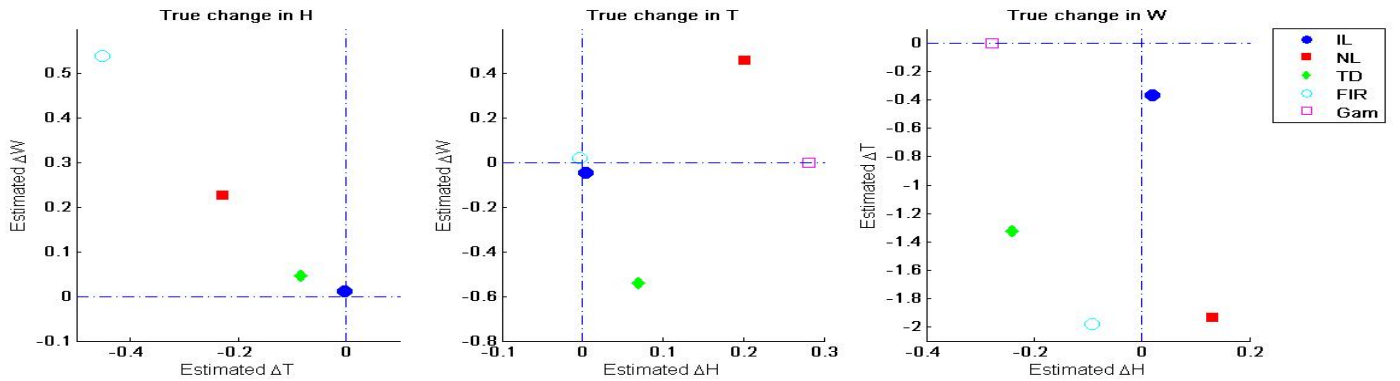


Figure 8

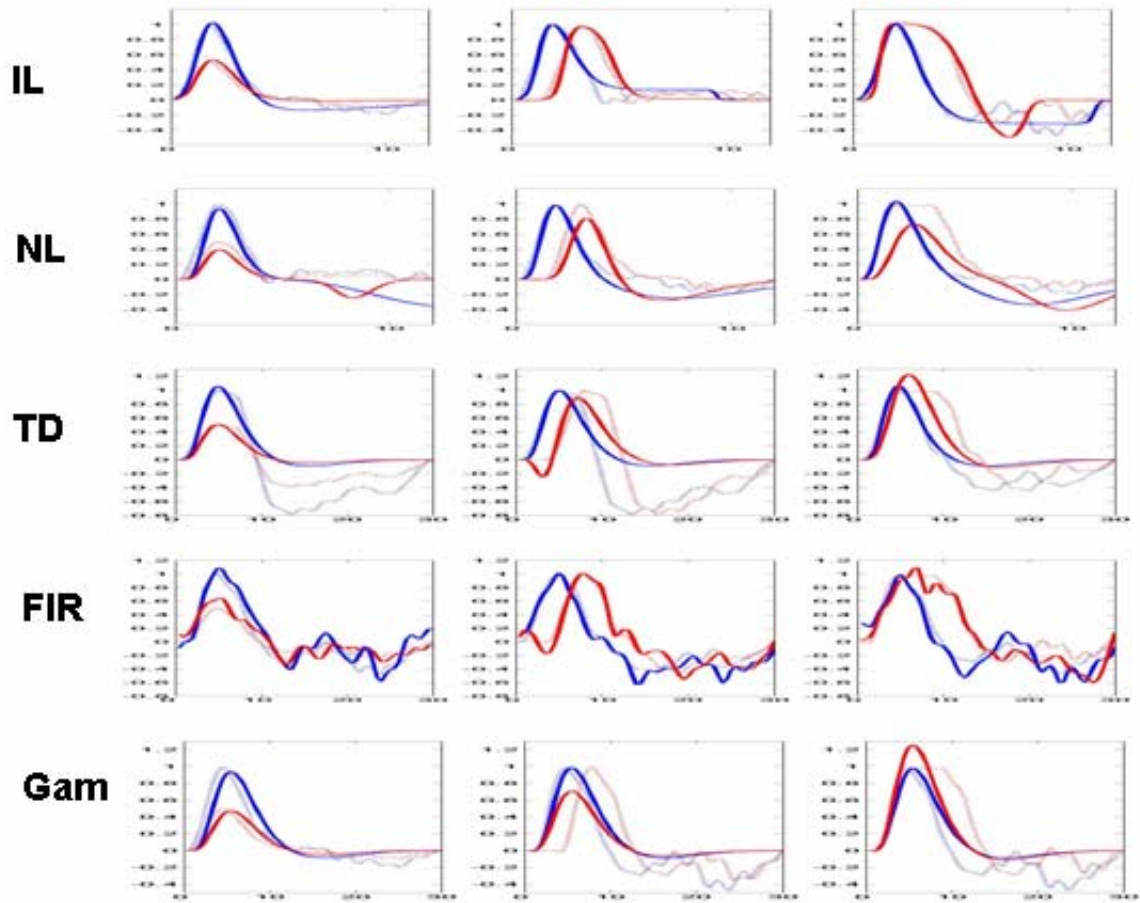


Figure 9

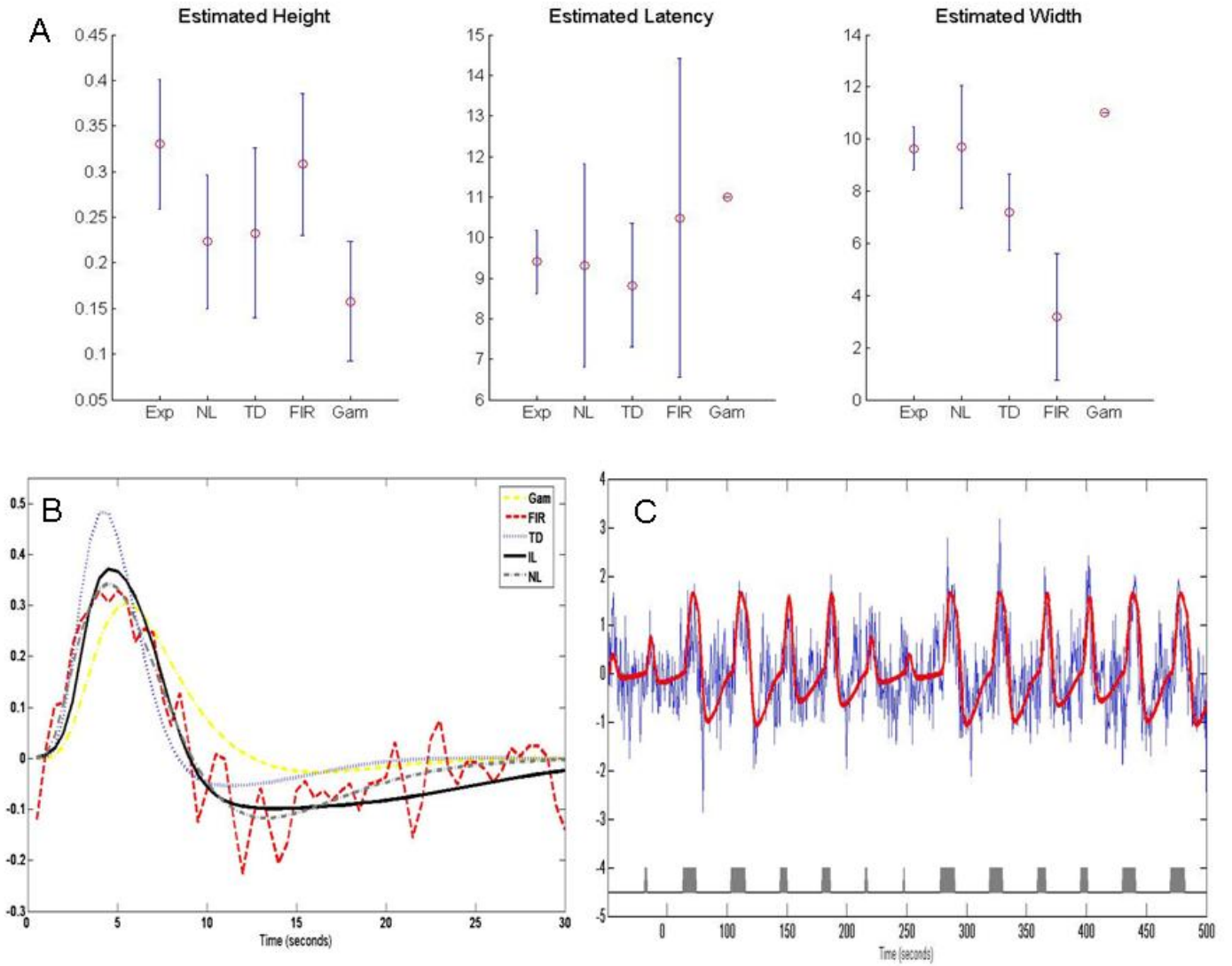


Figure 10

