Biophysically based method to deconvolve spatiotemporal neurovascular signals from fMRI data

J Pang  
*University of Sydney*

K M. Aquino  
*University of Sydney*

Peter Robinson  
*University of Sydney*, robinson@physics.usyd.edu.au

T C. Lacy  
*University of Sydney*

Mark M. Schira  
*University of Wollongong*, mschira@uow.edu.au

Follow this and additional works at: https://ro.uow.edu.au/sspapers

Part of the Education Commons, and the Social and Behavioral Sciences Commons

**Recommended Citation**

Pang, J; Aquino, K M.; Robinson, Peter; Lacy, T C.; and Schira, Mark M., "Biophysically based method to deconvolve spatiotemporal neurovascular signals from fMRI data" (2018). *Faculty of Social Sciences - Papers*. 3806.  
https://ro.uow.edu.au/sspapers/3806

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au
Biophysically based method to deconvolve spatiotemporal neurovascular signals from fMRI data

Abstract
2018 Elsevier B.V. Background: Functional magnetic resonance imaging (fMRI) is commonly used to infer hemodynamic changes in the brain after increased neural activity, measuring the blood oxygen level-dependent (BOLD) signal. An important challenge in the analyses of fMRI data is to develop methods that can accurately deconvolve the BOLD signal to extract the driving neural activity and the underlying cerebrovascular effects. New method: A biophysically based method is developed, which combines an extensively verified physiological hemodynamic model with a Wiener filter, to deconvolve the BOLD signal. Results: The method is able to simultaneously obtain spatiotemporal images of underlying neurovascular signals, including neural activity, cerebral blood flow, cerebral blood volume, and deoxygenated hemoglobin concentration. The method is tested on simulated data and applied to various experimental data to demonstrate its stability, accuracy, and utility. Comparison with existing methods: The resulting profiles of the deconvolved signals are consistent with measurements reported in the literature, obtained via multiple neuroimaging modalities. Conclusions: The method provides new testable predictions of the spatiotemporal relations of the deconvolved signals for future studies. This demonstrates the ability of the method to quantify and analyze the neurovascular mechanisms that underlie fMRI, thereby expanding its potential uses.

Disciplines
Education | Social and Behavioral Sciences

Publication Details

This journal article is available at Research Online: https://ro.uow.edu.au/sspapers/3806
Biophysically based method to deconvolve spatiotemporal neurovascular signals from fMRI data

J.C. Pang\textsuperscript{a,b,\*}, K.M. Aquino\textsuperscript{a,c,d}, P.A. Robinson\textsuperscript{a,b}, T.C. Lacy\textsuperscript{a,b}, M.M. Schira\textsuperscript{e,f}

\textsuperscript{a}School of Physics, University of Sydney, New South Wales 2006, Australia
\textsuperscript{b}Center for Integrative Brain Function, University of Sydney, New South Wales 2006, Australia
\textsuperscript{c}Sir Peter Mansfield Magnetic Resonance Center, School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, United Kingdom
\textsuperscript{d}Monash Institute of Cognitive and Clinical Neurosciences, Monash University, Clayton, Victoria 3800, Australia
\textsuperscript{e}School of Psychology, University of Wollongong, Wollongong, New South Wales 2522, Australia
\textsuperscript{f}Neuroscience Research Australia, Randwick, New South Wales 2031, Australia

\section*{Abstract}

\textbf{Background:} Functional magnetic resonance imaging (fMRI) is commonly used to infer hemodynamic changes in the brain after increased neural activity, measuring the blood oxygen level-dependent (BOLD) signal. An important challenge in the analyses of fMRI data is to develop methods that can accurately deconvolve the BOLD signal to extract the driving neural activity and the underlying cerebrovascular effects.

\textbf{New Method:} A biophysically based method is developed, which combines an extensively verified physiological hemodynamic model with a Wiener filter, to deconvolve the BOLD signal.

\textbf{Results:} The method is able to simultaneously obtain spatiotemporal images of underlying neurovascular signals, including neural activity, cerebral blood flow, cerebral blood volume, and deoxygenated hemoglobin concentration. The method is tested on simulated data and applied to various experimental data to demonstrate its stability, accuracy, and utility.

\textbf{Comparison with Existing Methods:} The resulting profiles of the deconvolved signals are consistent with measurements reported in the literature, obtained via multiple neuroimaging modalities.

\textbf{Conclusions:} The method provides new testable predictions of the spatiotemporal relations of the deconvolved signals for future studies. This demonstrates the ability of the method to quantify and analyze the neurovascular mechanisms that underlie fMRI, thereby expanding its potential uses.
1. Introduction

Functional magnetic resonance imaging (fMRI) has become a powerful and widely used tool to indirectly capture brain function through changes in blood flow, volume, and oxygenation that accompany neural activity; based on measuring the blood oxygen level-dependent (BOLD) signal (Logothetis, 2008). To interpret and exploit the BOLD signal, uncovering the dynamics of its underlying physiological processes, i.e., neural activity, astrocytic dynamics, cerebral blood flow (CBF), cerebral blood volume (CBV), and deoxygenated hemoglobin (dHb) concentration, is crucial. The conventional way to study these processes noninvasively is through other neuroimaging modalities such as electroencephalography (EEG), positron emission tomography (PET), invasive optical imaging, arterial spin labeling (ASL), vascular space occupancy (VASO), near infrared spectroscopy (NIRS), or diffuse optical tomography (DOT) (Feng et al., 2004; Newberg et al., 2005; Hillman et al., 2007; Zhang et al., 2007; Talagala et al., 2004; Boas et al., 2001). These modalities have a variety of spatial and temporal resolutions, each of which reveals particular structural and functional features of the brain. Whilst multimodal imaging with a combination of either of these modalities, such as simultaneous EEG-fMRI and PET-fMRI (Huster et al., 2012; Wey et al., 2014), are becoming more routine, there are still several technical challenges to overcome, e.g., different scales of involved physical phenomena, gas challenges interacting on the neurovasculature, and different resolutions. Also, when these techniques are used separately, the precise brain dynamics for each measurement will inevitably be different, and hence difficult to compare, because the data are collected at different times and under different conditions. The ability to simultaneously infer all these measures using minimal neuroimaging protocols would provide a new paradigm that allows direct and detailed spatiotemporal inferences of brain dynamics, complementing and advancing current multimodal neuroimaging studies.

Here, we develop a biophysically based method to deconvolve BOLD-fMRI data and simultaneously extract and image the spatiotemporal patterns of underlying neurovascular signals, as shown in Fig. 1. The deconvolution technique combines a Wiener filter with a physiological cortical hemodynamic model, derived from first principles, that has successfully linked the BOLD signal to spatiotemporal changes in neural activity, the astrocytic response and its effects on the vasculature (herein termed neuroglial drive), CBF, CBV, pressure, and dHb concentration (Drysdale et al., 2010; Aquino et al., 2012, 2014a,b; Pang et al., 2016; Lacy et al., 2016; Pang et al., 2017). The model has predicted propagating cortical hemodynamic waves, which have been experimentally validated in multiple fMRI experiments (Aquino et al., 2012; Lacy et al., 2016; Puckett et al., 2016), using two-photon microscopy (Gao et al., 2015), and via a data-driven model of effective connectivity (Gravel et al., 2017). These propagating hemodynamic waves have also been observed in a number of other recent studies.

*Corresponding author.

Email address: james.pang@sydney.edu.au (J.C. Pang)

Preprint submitted to Journal of Neuroscience Methods August 13, 2018
(Muller et al., 2018; Shatillo et al., 2018). In addition, the model has also predicted that three response modes govern the fundamental patterns of oscillations of the BOLD signal, with structures that are independent of the stimulus (Pang et al., 2016). More importantly, the model rests on parameters that can be measured independently to reflect the physiology and hemodynamics of the brain, including structural and functional abnormalities (D’Esposito et al., 2003).

![Pipeline of the biophysically based method to deconvolve BOLD-fMRI data (input) to obtain the underlying spatiotemporal neurovascular signals (output), including neural activity, CBF, CBV, and dHb concentration. The deconvolution technique is a combination of a physiological cortical hemodynamic model and a Wiener filter.](image_url)

**Figure 1:** Pipeline of the biophysically based method to deconvolve BOLD-fMRI data (input) to obtain the underlying spatiotemporal neurovascular signals (output), including neural activity, CBF, CBV, and dHb concentration. The deconvolution technique is a combination of a physiological cortical hemodynamic model and a Wiener filter.

This opens new noninvasive windows to analyze dynamics of stimulus-evoked physiological signals directly from fMRI measurements, thereby allowing new information to be obtained from BOLD-fMRI data. We first test the method on simulated data to establish its stability and accuracy in the absence of confounds. Then, it is applied to experimental datasets and the results are qualitatively compared with measurements from neuroimaging modalities, such as NIRS and DOT, to demonstrate its utility. Overall, the applications illustrated here substantiate the ability of the method to noninvasively image, quantify, and analyze the neurovascular underpinnings of fMRI data. Moreover, this work demonstrates that the method provides a solid framework that produces testable predictions for the wider neuroimaging community in order to better understand brain function.

All the methods and tools developed in this work are provided in an accompanying freely available toolbox called “BrainPalimpsest”, which can be found at [http://github.com/BrainDynamicsUSYD/BrainPalimpsest](http://github.com/BrainDynamicsUSYD/BrainPalimpsest); details are discussed in Sec. 2.5.

2. Theory and Methods

In this section, we describe the principles of our cortical hemodynamic model, the features of the new deconvolution method, and the datasets to test and apply the method on.

2.1. Physical model of cortical hemodynamics and the BOLD signal

To formulate a technique that can simultaneously deconvolve the spatial and temporal aspects of BOLD-fMRI data to image the underlying neurovascular signals, a model that
can quantitatively link the physical processes governing the brain’s hemodynamics is needed. Our recent spatiotemporal hemodynamic model derived from the physiological properties of cortical hemodynamics satisfies this criterion (Drysdale et al., 2010; Aquino et al., 2012, 2014a,b; Lacy et al., 2016; Pang et al., 2016, 2017).

2.1.1. Approximations and processes involved in the model

In order to physically model spatiotemporal hemodynamics, three well-founded approximations are needed: (i) neural activity and hemodynamics are treated on a two-dimensional (2D) sheet as local averages, denoted as mean field; (ii) cortical tissue is treated as a porous elastic medium, where the pores are the dense vasculature that is embedded in incompressible tissue; and (iii) small-scale spatial distribution of underlying vasculature is averaged, considering only dynamics over larger scales (∼1 mm) that complements current fMRI voxel resolutions. These approximations allow a set of dynamical equations for the neuroglial drive, CBF, CBV, and dHb concentration to be derived subject to conservation of blood mass and momentum, and of oxygen.

The underlying physical processes in the model are illustrated in Fig. 2a and summarized as follows. Due to a rise in neural activity, \( \phi(r, t) \), at time \( t \) and position \( r \) on the cortical sheet, neurotransmitters released from synapses activate nearby astrocytes, which in turn produce a response called the neuroglial drive, \( \zeta(r, t) \). This drive affects the surrounding vasculature via neurovascular coupling (Koehler et al., 2006), leading to an increase in CBF, \( F(r, t) \). This results in an increase in CBV, \( \Xi(r, t) \), that deforms the surrounding tissue, thus exerting a pressure, \( P(r, t) \), on nearby vessels. Due to the resulting change in vessel volume, the influx of arterial blood increases the content of oxygenated hemoglobin (oHb), hence reducing the local concentration of dHb, \( Q(r, t) \). Finally, relative changes in CBV and the concentration of paramagnetic dHb lead to changes in the BOLD signal, \( Y(r, t) \). A block diagram of the causal chain of physical processes from neural activity to BOLD is shown in Fig. 2b (see Appendix A.1 for full details of the model’s mathematical equations).

We have previously shown that the BOLD signal can be mathematically decomposed into response modes as \( Y(r, t) = W(r, t) + L(r, t) + D(r, t) \), where \( W(r, t) \), \( L(r, t) \), and \( D(r, t) \) are the wave, local-oscillating, and local-decaying response modes, respectively (Pang et al., 2016); these response modes are schematically shown in Fig. 2c and are included in our set of quantities that describe the dynamics of the model. This method of signal decomposition has been immensely useful across numerous fields of physics, such as electromagnetism, quantum mechanics, and plasma physics, revealing the fundamental properties of the system under study.

It is important to note that the neural activity we consider here only represents the low-frequency component of the synaptic activity, which the BOLD signal is most sensitive to. This is based on studies showing that the amplitude and power of BOLD are directly related to low-frequency electrical activity as measured by EEG (Lu et al., 2007; He et al., 2008; Portnova et al., 2018). We are currently unable to capture high-frequency neural activity because the processes of neurovascular coupling responsible for the conversion of fast synaptic/spiking responses to slow dynamics remain to be unclear.
Figure 2: Physical processes in the cortical tissue that are described in our model. (a) Schematic of a small cortical region showing the relay of activity from neuron via astrocyte to surrounding vasculature. The key physical quantities that are calculated in our model are the neural activity $\phi$, neuroglial drive $\zeta$, CBF $F$, CBV $\Xi$, and dHb concentration $Q$. (b) Block diagram showing how the physical quantities interact to produce the BOLD signal $Y$. Moreover, $Y$ can be decomposed into wave ($W$), local-oscillating ($L$), and local-decaying ($D$) response modes. (c) BOLD-response modes for an impulsive source. The $W$ mode propagates outward as a wave from a source. The $L$ and $D$ modes (see inset) are both local responses that are restricted to very near the source; the $L$ mode has a damped temporal oscillation, while the $D$ mode is negative and decays monotonically.

2.1.2. Linearization and Fourier transformation

The physical quantities in the model obey dynamical equations that can be used to predict the nonlinear hemodynamic response to arbitrary neural activity. However, experiments have shown that the hemodynamic response is approximately linear for a wide array of stimuli that induce low-amplitude neural responses (Boynton et al., 1996; Cohen, 1997; Robinson et al., 2006; Aquino et al., 2012; Gaglianese et al., 2017); for example, responses due to fast moving stimuli (Lacy et al., 2016) or rapid event-related stimuli (Hinrichs et al., 2000; Wager et al., 2005). Moreover, we have previously determined the limits of applicability of the linear approximation of the theory and showed that all but the strongest responses can be approximated as linear (Lacy et al., 2016). This serves as our rationale for linearizing the equations of the model. Linearization is performed by assuming that each hemodynamic variable $\theta$, in our case $\theta = \phi, \zeta, F, \Xi, P, Q, Y$, can be written as a sum of the steady-state value $\theta^{(0)}$ and a linear perturbation $\theta^{(1)}$ from the steady state; i.e., $\theta(r, t) = \theta^{(0)}(r, t) + \theta^{(1)}(r, t)$, where higher order terms are ignored; Appendix A.2 summarizes the results of this procedure.

Linearization allows the analyses of the dynamics of the hemodynamic variables via Fourier transform. Fourier transform is a powerful tool that operates in the frequency domain, inside which certain mathematics become simple as compared to staying in the coordi-
nate domain (Arfken and Weber, 2005). One important use of Fourier transform is its ability to parse the different frequency components of a signal, e.g., intrinsic oscillations/rhythms of EEG, providing a better understanding of its properties. In addition, because of the elementary properties that make the mathematics involved in Fourier transform simple, some of its other advantages are the following: (i) its magnitude is directly related to the power of a signal at a particular frequency; (ii) it is energy preserving, i.e., the total energy in a signal across all space and time is equal to the total energy in the transform across all frequencies; (iii) it turns convolution into a simple multiplication; and (iv) it transforms a differential equation into an algebraic equation, which allows solutions of coupled differential equations to be easily solved.

2.1.3. Transfer function and forward calculation

Once the model equations are linearized and transformed to Fourier space, we can use a variety of well-established physics-based methods to understand their properties (Arfken and Weber, 2005). Most importantly, we use transfer functions to conveniently represent their input–output responses (Shinners, 1998). The transfer function $T_{AB}(k,\omega)$ describes the change in quantity $A$ per unit change in quantity $B$ at the same spatial frequency, $k$, and temporal frequency, $\omega$, with $A(k,\omega) = T_{AB}(k,\omega)B(k,\omega)$. To convert $A(k,\omega)$ to a signal in coordinate space (i.e., in $r$ and $t$), an inverse Fourier transformation can be implemented as

$$A(r,t) = \mathcal{F}^{-1}\{T_{AB}(k,\omega)B(k,\omega)\}$$  \hspace{1cm} (1)

where $k = (k_x,k_y)$, $r = (x,y)$, and $\mathcal{F}^{-1}$ is the inverse Fourier transform operator for both space and time.

If the neural activity, $\phi(r,t)$, is known or assumed, the hemodynamic model can directly yield the resulting patterns of hemodynamic quantities. This is accomplished using transfer functions via the following steps: we first Fourier transform $\phi(r,t)$ to obtain $\phi(k,\omega)$; we then calculate $A$ from $\phi$ via $A(k,\omega) = T_{A\phi}(k,\omega)\phi(k,\omega)$, where $A$ is $\zeta$, $F$, $\Xi$, $Q$, $W$, $L$, $D$, or $Y$ (see Appendix A.3 for explicit forms of these transfer functions); and finally, inverse Fourier transformation yields the corresponding responses in coordinate space. We term this chain of calculations from $\phi$ to the hemodynamic responses the forward technique, as summarized in Fig. 3a. Arguably, in most cases, $\phi(r,t)$ is not accurately known and only the BOLD signal is available. Thus, we develop a deconvolution technique in the next section to obtain $\phi(r,t)$ and the intermediate quantities from the BOLD signal.

2.2. Method to deconvolve the BOLD signal

Following Sec. 2.1, we can derive the transfer function $T_{YA}(k,\omega)$ to obtain BOLD $Y$ from quantity $A$ (see Appendix A.3 for explicit forms of these transfer functions). To perform the deconvolution of BOLD, we can, in principle, get $A$ from $Y$ by taking the inverse of $T_{YA}$ to get $T_{AY} = T_{YA}^{-1}$ and then implement Eq. (1) by replacing $B$ with $Y$. However, for real data with intrinsic noise and/or measurement errors, performing the deconvolution this way would lead to corrupted solutions (Frank et al., 2001; Krüger and Glover, 2001) because this
naive method would amplify noise effects at high spatial and temporal frequencies where the noise exceeds the signal (Aquino et al., 2014b). One of the most recognized ways to solve the above inverse problem is to employ Wiener deconvolution, which uses a filter whose design is informed by the likely noise structure of the raw signal (Peacock and Treitel, 1969; Kerr et al., 2009). The Wiener filter effectively cuts off the signal at frequencies higher than \((k_c, \omega_c)\) where \(|T_{YA}(k_c, \omega_c)| \approx |\text{NSR}(k_c, \omega_c)|\) and NSR is the noise-to-signal ratio. This removes high-frequency components and limits the signal to within the effective bandwidth where the transfer function dominates the noise, as illustrated in Fig. 3b. In principle, the cutoff frequencies can be determined using a data-driven approach. For example, in the context of fMRI, since the BOLD signal generally has a low-pass character with significant power at temporal frequencies \(\leq 0.2\) Hz (Robinson et al., 2006; Wang et al., 2014), a cutoff value of \(f_c = \omega_c / 2 \pi \approx 0.2\) Hz is a viable choice to increase the signal-to-noise ratio.

Wiener deconvolution is mathematically implemented as \(A(r, t) = F^{-1} \{ D(k, \omega)Y(k, \omega) \}\), where the Wiener filter is

\[
D(k, \omega) = T_{YA}^{-1}(k, \omega) \frac{|T_{YA}(k, \omega)|^2}{|T_{YA}(k, \omega)|^2 + |\text{NSR}(k, \omega)|^2}.
\] (2)

Assuming that the noise is white and the signal is an impulse function in space and time, NSR\((k, \omega)\) can be approximated by a constant \(\sigma\); this choice has been shown to be effective in previous deconvolution studies (Glover, 1999; Kerr et al., 2009; Aquino et al., 2014b). The value of \(\sigma\) can be estimated using the technique of Aquino et al. (2014b). Note that changing \(\sigma\) only changes the degree of smoothing, but not the overall features of the deconvolved signal.

Hence, to deconvolve BOLD-fMRI data and image the underlying neurovascular signals, we develop a biophysically based Hybrid Wiener deconvolution technique, which combines the model’s forward technique in Fig. 3a and the Wiener deconvolution mentioned above. The technique is as follows: we first Fourier transform \(Y(r, t)\) to obtain \(Y(k, \omega)\); we next use the Wiener deconvolution, via the transfer function \(T_{Y\phi}^{-1}(k, \omega)\), to calculate \(\phi(k, \omega)\); forward calculations are then implemented, exactly as in Fig. 3a, to obtain quantities \(A\) from the calculated \(\phi\), where \(A\) is \(\zeta, F, \Xi, Q, W, L, \) or \(D\); and finally, we inverse Fourier transform the calculated quantities to obtain the corresponding neurovascular responses in coordinate space. This chain of calculations is illustrated in Fig. 3c. The technique outperforms directly inverting each transfer function because the effective bandwidth of each filter must be tuned independently.

2.3. Datasets

Here, we describe four datasets; one of which is simulated and three are empirical obtained from different and independent visual experiments.

2.3.1. Simulated data

To test the deconvolution method, we produce synthetic noisy BOLD data due to simulated neural activity

\[
\phi(r, t) = \delta(y)\phi(x, t),
\] (3)
Figure 3: Deconvolution schemes. (a) Forward technique to calculate physiological quantities in our model for given neural activity, $\phi(r,t)$, as follows: $\phi(r,t)$ is Fourier transformed (FT) to get $\phi(k,\omega)$; the result is fed into forward calculations (red box) where the Fourier responses $A(k,\omega)$ are calculated using the transfer functions $T_A\phi(k,\omega)$, with $A$ being $\zeta$, $F$, $\Xi$, $Q$, $W$, $L$, $D$, or $Y$; and finally, inverse Fourier transformations yield $A(r,t)$. (b) Wiener filter to invert a transfer function $T$ by minimizing its response at high frequencies where it no longer dominates the NSR; this limits $T$ to the effective bandwidth shown. (c) Biophysically based Hybrid Wiener deconvolution technique to deconvolve the BOLD signal, $Y(r,t)$, to calculate physiological quantities, as follows: $Y(r,t)$ is Fourier transformed to get $Y(k,\omega)$; the result is fed into a Wiener deconvolution (blue box) that uses the Wiener filter in (b) to calculate $\phi(k,\omega)$; the result is then used in forward calculations (red box), exactly as in (a), to obtain $A(k,\omega)$; and finally, inverse Fourier transformations yield $A(r,t)$.

where $\delta(y)$ is the Dirac-delta function along the $y$-dimension and $\phi(x,t)$ is the component of the neural activity that varies along the $x$-dimension, which we choose to have a stationary 1D Gaussian structure of the form

$$\phi(x,t) = e^{-x^2/\sigma_x^2}e^{-(t-t_0)^2/\sigma_t^2} + \mathcal{N}(x,t),$$

(4)

where $\sigma_x$ and $\sigma_t$ are the spatial and temporal widths, respectively, and $t_0$ is the time of peak neural activity; for simplicity, we fix their values to $\sigma_x = 1$ mm, $\sigma_t = 1$ s, and $t_0 = 2$ s. The quantity $\mathcal{N}(x,t)$ represents spatiotemporal white noise with autocorrelation $\langle \mathcal{N}(x,t), \mathcal{N}(x',t') \rangle = \delta(x-x')\delta(t-t')$. Then, the corresponding simulated BOLD signal is obtained by substituting Eq. (3) into the model equations to perform a forward calculation.

2.3.2. Experimental dataset 1: Stationary ring data

This dataset was previously collected from a study of evoked response in the visual cortex (Aquino et al., 2012). The visual stimulus was an image of three stationary isoeccentric rings.
(with checkerboard patterns) at 0.6°, 1.6°, and 3° eccentricity that oscillated in contrast at a rate of 2 Hz for 8 s during the “on” phase, and disappeared for 12 s during the “off” phase. The experiment was performed on a healthy subject.

The data were acquired on a Phillips 3T Achieva MRI scanner using an 8-channel head coil to resolve fMRI with 1.5 mm cubic voxels and a repetition time (TR) of 2 s with coverage restricted to the occipital pole. The BOLD response was projected onto a flattened representation of the visual cortex using a surface reconstruction from a gray and white matter segmentation of a T1-weighted image sampled with 0.75 mm isotropic voxels. The average BOLD response was then measured at various distances, \( x \), from the centerline of the evoked neural response, restricted to the primary visual cortex (V1), as a function of time \( t \). This resulted in a BOLD signal that was sampled on the interval \( |x| \leq 5 \text{ mm} \) and \( 0 \text{ s} \leq t \leq 20 \text{ s} \). Finally, the temporal time series was low-pass filtered with a third order Butterworth filter below 0.1 Hz to obtain a temporally smoothed version of the data (Aquino et al., 2012).

2.3.3. Experimental datasets 2 and 3: Expanding ring and expanding arc data

For the first experiment, the visual stimulus was an image of a ring (with multicolored pattern) that expands from 0.3° to 5.5° eccentricity at a rate corresponding to a neural response in V1 that travels at roughly 0.5 mm s\(^{-1}\). During this expansion, the subject performed a simple fixation task to detect when a small square turned red. This is a standard retinotopic stimulus designed to map eccentricities in the visual cortex (Schira et al., 2009). Sample snapshots of the stimulus are shown in Fig. 4. The solid lines are always present to aid fixation.

![Figure 4: Snapshots of the expanding ring visual stimulus. The center of each snapshot corresponds to the center of the fovea in the visual field where the subject fixated.](image)

In another experiment in the same recording session, the subject was presented with an image of an arc (with checkerboard pattern) that expands from 0.3° eccentricity (spanning 15° of polar angle) to end at 5.5° eccentricity (spanning 7° of polar angle). The rate of expansion was the same as that for the expanding ring data and was designed to retinotopically map to a bar moving at a constant speed in V1 where its height and width were unchanging (Lacy et al., 2016). Sample snapshots of the stimulus are shown in Fig. 5. The solid lines are always present to aid fixation.

![Figure 5: Snapshots of the expanding arc visual stimulus.](image)
Both experiments were performed on a single healthy subject and data were acquired on a Phillips 3T Achieva MRI scanner using a 32-channel head coil to resolve fMRI with 1.5 mm cubic voxels and a TR of 2 s with coverage restricted to the occipital pole. The study protocols were approved by the ethics board of the University of New South Wales and Neuroscience Research Australia.

The data were automatically segmented, guided by anatomical templates, and cortical surfaces were reconstructed from the segmentation using FreeSurfer (http://surfer.nmr.mgh.harvard.edu). The data were detrended, motion corrected, and coregistered to an anatomical T1-weighted image (acquired at a resolution of $0.75 \times 0.75 \times 0.75$ mm$^3$) that was acquired on the same day (Lacy et al., 2016). The flattened surface representation of the data was obtained around the occipital pole and was interpolated to a regular grid. To improve signal-to-noise ratio, the data were averaged across all scan repetitions. Finally, the data were projected onto a smoothed cortical surface representation that smoothens out the gyri and sulci of the brain’s convolutions. In addition, the boundaries of V1 and the secondary visual cortex (V2), which were obtained based on a group-averaged parcellation by Glasser et al. (2016), were annotated in the resulting cortical surfaces for visualization purposes.

For the expanding arc data, the following additional processing steps were performed to obtain responses that only vary in eccentricity and time. A canonical atlas by Benson et al. (2012) was first fitted to the data to obtain eccentricity and polar angle maps in V1. Then, for each chosen eccentricity value, the maps were used to average isoeccentric responses that span polar angles from 70° to 110°.

It is important to note that the above data processing steps are employed to provide clear visualization of the results. They do not have any effect on the framework of the deconvolution method and predictions made thereafter. If desired, other techniques can be used to postprocess the model predictions for comparison with other data types.

2.4. Parameter choice

The novelty of our deconvolution method is that it involves a hemodynamic model that rests on biophysical parameters (see Table A.1) that can be measured independently to reflect the physiology and hemodynamics of the brain. Hence, to obtain the most accurate deconvolution, these parameters will need to be calibrated for different brain areas of different subjects. For purposes of demonstration, here, we use fixed nominal values of most of
the parameters, which are taken from the literature (Aquino et al., 2014a), whereas the remaining parameters, i.e., blood propagation speed ($v_\beta$), blood propagation damping rate ($\Gamma$), natural frequency of flow response ($\omega_f$), blood flow signal decay rate ($\kappa$), and astrocytic delay ($\tau_d$), are inferred in our previous studies for the primary visual cortex of the same individual subject using data-driven fitting procedures (Aquino et al., 2012; Pang et al., 2016). Note that the parameters can all be varied but only the accuracy of the deconvolution results and not the actual framework of the method crucially depends on the values of these parameters.

2.5. Deconvolution software

We have developed a Matlab toolbox called “BrainPalimpsest” to implement the methods of this work; this can be found at http://github.com/BrainDynamicsUSYD/BrainPalimpsest. In the context of ancient manuscripts, a palimpsest refers to a reused writing surface that has multiple layers of writing hidden below the surface text; these layers can now often be read by new techniques such as x-ray and fluorescence spectroscopy. This analogy best fits our work, which is to uncover the multiple layers of information that lie underneath fMRI measurements. Hence, we name the toolbox BrainPalimpsest.

The toolbox accepts both 1D and 2D BOLD-fMRI data and performs the biophysically based deconvolution method to get quantities as desired by the user. The user has options to vary the parameters and several functions of the model. Users can familiarize themselves with the toolbox by running the scripts “Results_StationaryRing.m” and “Results_ExpandingRingAndExpandingArc.m” to reproduce the results of this work. Finally, the toolbox contains the function “Test_SimulatedData.m” that shows how a simulated dataset can be produced. The experimental stationary ring, expanding ring, and expanding arc datasets are also provided. The open source nature of the toolbox allows researchers to customize individual functions according to their needs and to incorporate the functions into other analysis protocols.

3. Results and Discussion

3.1. Test: Deconvolution of simulated data

To test the accuracy of the method, we use 1D simulated neural activity with a known Gaussian structure plus noise (see Fig. 6a) to produce the synthetic BOLD response seen in Fig. 6b via the forward equations of the model. For a straightforward analysis, the spatiotemporal neural activity is simulated to vary in only one spatial dimension to highlight the BOLD response and its arterial dynamics (Aquino et al., 2012, 2014a; Pang et al., 2016). See Sec. 2.3.1 for mathematical details of the simulated data. The dashed line in Fig. 6b marks the wavefront of the spatiotemporal wave component of the BOLD response, while the solid line in Fig. 6b marks the response at $x = 0$, with its profile shown in Fig. 6c. The time profile in Fig. 6c matches the temporal hemodynamic response function conventionally used in neuroimaging studies (Friston et al., 2000). We have shown here that our model not only reproduces the well-accepted temporal characteristics of the BOLD response, but
also predicts additional dynamics, i.e., waves and response modes, from the physiological coupling of space and time that are unaccounted for in previous studies.

Since we know the ground truth, we can implement the forward technique in Fig. 3a to get the various physiological quantities from the given neural activity, which then serve as the reference against which the results of the deconvolution are compared. The spatiotemporal profiles of the responses that result from the forward technique are shown in Fig. 6d. On the other hand, the results of the deconvolution are shown in Fig. 6e.

In general, the deconvolved responses closely reproduce the dynamics of the responses obtained by the forward technique; their peak values coincide and their spatiotemporal
shapes are similar. However, the responses in Fig. 6e are smoother and overestimate the spatial and temporal extent of those in Fig. 6d. These artifacts can be attributed to the Wiener filter because higher frequencies outside the effective bandwidth are filtered out, leaving responses dominated by large and slower oscillations. Nonetheless, the results prove that the method is well founded and is sufficient to ensure that the correct spatiotemporal profiles of the responses are obtained.

It is important to note that one can easily change the noise level in the Wiener filter to improve the method, but this will only change the degree of smoothing and not the overall features of the resulting patterns; see for example our previous work on probing the effect of varying the filter’s noise amplitude on the deconvolution of neural signal (Aquino et al., 2014b). In addition, we find that the results for a simulated neural activity spatially varying in 2D space are very similar to those for the 1D counterpart, thus we excluded it in the text to avoid redundancy.

3.2. Application 1: Deconvolution of stationary ring data

To demonstrate the applicability of the method in deconvolving actual 1D BOLD signals, we deconvolve an experimental BOLD response evoked in the right visual cortex by a visual stimulus consisting of three stationary isoeccentric rings (Aquino et al., 2012). The retinotopically mapped stimulus induces straight lines of neural activity in the visual cortex, allowing the resulting BOLD signal to be measured as a function of 1D distance, \( x \), from the center of the stimulus and time after stimulus onset, as detailed in Appendix A.4.

The spatiotemporal profile and time series at \( x = 0 \) of the data are shown in the top panels of Fig. 7. Note that the time series at \( x = 0 \) can be compared more directly with voxel-based time-series data commonly found in the literature than the full spatiotemporal profile in \( x-t \) space.

The spatiotemporal profiles of the deconvolved responses shown in the left column of Fig. 7 are consistent with those in Fig. 6e, especially in their transitions between positive and negative values as the BOLD signal evolves; note that a positive or negative value represents an increase or decrease of level from baseline, respectively. The correspondence between the results of Figs 6e and 7 demonstrates that the method provides reliable and effective deconvolutions.

The time series in the right column of Fig. 7 clearly show that the peak of the deconvolved neural activity precedes the BOLD signal by approximately 3 to 4 s, which is in accordance with their physiological relationship resulting from the mechanisms in Fig. 2 (Atwell et al., 2010). We also find significant negative neural activity that can either be due to neural inhibition or reduced excitation with respect to baseline, whereas negative neuroglial drive can be interpreted as a reduction in the vasodilatory action of glia to vessels with respect to baseline. The deconvolved CBF and CBV profiles are approximately similar to that of the BOLD signal, but CBF is more localized and CBV is more spatially extended, and their shapes agree with theoretical calculations (Bennett et al., 2008) and experiments (Yang et al., 2004; Chen and Pike, 2009; Filosa et al., 2016). For example, at the most negative BOLD signal, both CBF and CBV are also negative, similarly to the simultaneous BOLD-ASL-VASO results of Yang et al. (2004) and the 7 T VASO measurement of Huber et al.
(2014); note that negative CBF and CBV values represent reduced flow and volume in the vessels with respect to baseline, respectively. The deconvolved dHb concentration becomes very negative at the peak of the BOLD signal, with the reverse at its minimum, which accords with experimental measurements using optical imaging, NIRS, and DOT (Hillman et al., 2007; Boas et al., 2001). The deconvolved $W$ mode has double peaks roughly occurring at the corresponding minimum of the BOLD signal. The deconvolved $L$ and $D$ modes have similar features to the CBV and dHb concentration, respectively.

Figure 7: Experimental BOLD-fMRI stationary ring data and deconvolved responses. The left column shows spatiotemporal profiles of BOLD and responses. Color bars indicate levels from negative (blue) to positive (red). The features of the deconvolved responses are similar to those in Fig. 6e. However, the wave of propagation of the $W$ mode is cropped due to the limited spatial range of the experimental data (10 mm vs. 30 mm in the simulation). The right column shows time series profiles at $x = 0$.

However, we have to note that there are differences in the fine-scale properties of our deconvolved responses, such as undershoot time, when compared with those in the literature, which may be attributed to different data processing; our results are single-subject responses, while most measurements in the literature are group-averaged and tend to reduce variability. Nonetheless, the macroscopic properties of the results are consistent with those
in the abovementioned literature comparison, thus supporting the plausibility of applying the deconvolution method to experimental BOLD data.

3.3. Application 2: Deconvolution of expanding ring data

To demonstrate the applicability of the method in deconvolving 2D BOLD signals, we now deconvolve an experimental BOLD response evoked in the visual cortex by a visual stimulus consisting of a ring that expands from an eccentricity of $E = 0.3^\circ$ to $E = 5.5^\circ$ (see Fig. 4). This is a stimulus very commonly used in retinotopic mapping (Sereno et al., 1995; DeYoe et al., 1996; Engel et al., 1997; Schira et al., 2007), which can efficiently map the eccentricity direction in the visual cortex. Here, we demonstrate how our method can add to the interpretation of retinotopic mapping data.

Figure 8a shows the responses projected onto a smoothed left cortical hemisphere at $t = 8, 12, \text{ and } 16 \text{ s}$. The first row shows the theoretical mapping of the visual stimulus onto V1 (black solid lines) (Schira et al., 2010; Benson et al., 2012). The BOLD signal in the second row shows propagation from the fovea (lower left corner of the black solid lines) to the periphery in the direction of expansion of the visual stimulus, but delayed by $\sim 4 \text{ s}$; this estimate will be improved via smooth interpolation later in this section. This delay is expected since the visual stimulus does not directly cause the BOLD signal; instead, the causal chain of physical processes described in Fig. 2a must be followed.

To establish that the results are physiologically plausible, the abovementioned delay must be accounted for by the deconvolved neural activity such that its peak must consistently match the visual stimulus, both spatially and temporally. This prediction is validated in the neural panels of Fig. 8a where the neural response coincides with the evolution of the visual stimulus, which is a novel result and strongly supports the validity of the model.

The deconvolved neuroglial drive in Fig. 8a has similar dynamics to the neural activity but lags it by $\sim 1.3 \text{ s}$. This agrees with theoretical and empirical observations (Masamoto et al., 2015; Pang et al., 2017) and supports the hypothesis that glial cells induce additional time delays as they facilitate neurovascular coupling, the intercellular communication from neurons to microvessels that mediates functional hyperemia (Atwell et al., 2010).

The deconvolved CBF and CBV in Fig. 8a closely match the timing and spatial localization of the BOLD signal, with high CBF and CBV regions exhibiting markedly higher BOLD signals. However, the CBF and CBV have higher signal-to-noise ratios than the BOLD signal, which have potential advantages for studying fine-scale functional brain organization. It is important to mention that the CBF and CBV we deconvolved here are not specific to the individual contributions of arterioles, capillaries, or veins to flow and volume, respectively, but instead represents a mean-field approximation.

The deconvolved dHb concentration in Fig. 8a shows opposite behavior to the BOLD signal, with regions of very negative dHb concentration exhibiting higher BOLD signal, consistent with previous sections and experimental findings (Hillman et al., 2007).

Finally, the response modes $W$, $L$, and $D$ in Fig. 8a show the wave- and localized-mode features of the BOLD signal and quantitatively follow the relationship $\text{BOLD} = W + L + D$. These modes have never been explored before and provide an understanding of the fundamental patterns of oscillations of the BOLD signal.
Figure 8: Experimental BOLD-fMRI expanding ring data and deconvolved responses. (a) Responses on a smoothed left cortical hemisphere. Colors indicate levels from negative (blue) to positive (red). Black and brown solid lines represent the boundaries of V1 and V2, respectively. Rows from top to bottom show the retinotopically mapped visual stimulus, BOLD signal, and deconvolved responses, as labeled. Note that the stimulus in retinal space is also shown on top of the mapped visual stimulus panels. In each row, columns show the responses at $t = 8, 12$, and $16$ s with fixed color bar. All the responses follow the timing, extent, and direction of propagation of the mapped visual stimulus in V1. (b) Cross-correlations of the responses with the BOLD signal as a function of $(t_{\text{response}} - t_{\text{BOLD}})$. The symbols represent mean values, the error bars show one standard deviation from the mean, and the solid lines are interpolations. The dashed lines are the times of peak correlation magnitudes with their estimated values shown to the right.
We emphasize at this point that our method provides two complementary constructs for characterizing the components of the BOLD signal. First is through its physiological constituents; i.e., the neural activity, neuroglial drive, CBF, CBV, and dHb concentration, which are directly related to actual physiology and hemodynamics and can be compared to measurements of existing neuroimaging modalities. Second is through its $W$, $L$, and $D$ response modes, which have analytic properties based on physiology that can be exploited to highlight or suppress specific wave and/or local responses of interest (Pang et al., 2016). In principle, the response modes can better describe the fundamental properties of physical propagating and nonpropagating components of BOLD-fMRI data than common signal analysis methods such as independent component analysis. This is because such methods do not account for the nature of the system that produces the signals and are not based on brain physiology, so their results may mix the dynamics of fundamental brain modes and statistical effects. However, actual comparison of the general utility of the response modes and other signal analysis methods requires further study beyond the scope of this work.

It is also important to note that the peak values of the responses are not synchronized with that of the BOLD signal. This can be due to temporal delays between responses, indicating the causal relationship of their dynamics, as dictated by physiology. We quantitatively confirm this by taking the average cross-correlation of the deconvolved responses and the BOLD signal in 2D as a function of $(t_{\text{response}} - t_{\text{BOLD}})$, as shown in Fig. 8b. Note that $(t_{\text{response}} - t_{\text{BOLD}}) < 0$ means that the response leads BOLD, while $(t_{\text{response}} - t_{\text{BOLD}}) > 0$ means that the response lags BOLD. To increase the resolution of the results, we interpolated the calculated cross-correlations using a cubic spline interpolation algorithm.

From Fig. 8b, we can see that the neural activity, neuroglial drive, CBF, CBV, and $L$ mode are generally positively correlated with the BOLD signal, and the dHb concentration, $W$ mode, and $D$ mode are anticorrelated with BOLD. The times of peak correlation magnitudes of the neural activity and neuroglial drive lead the BOLD signal by 3.4 s and 2.1 s, respectively. This sequence of response from neural activity to neuroglial drive to BOLD signal is in accordance with the causal chain of physical processes discussed in Fig. 2a where neurons activate astrocytes to produce a neuroglial drive that ultimately leads to BOLD (Atwell et al., 2010; Pang et al., 2017). We also observe that the times of peak correlation magnitudes of CBF, CBV, dHb concentration, $W$ mode, and $L$ mode are near zero, which suggests that they are almost synchronized with BOLD. Interestingly though, we find that the time of peak correlation magnitude of the $D$ mode leads the BOLD signal by 2 s, giving a potentially useful insight regarding the early phase of the BOLD signal. Closer inspection of the delays shows that they are smaller than current fMRI temporal resolutions, which is the reason why these delays are not yet observed in empirical findings.

Finally, we highlight that several model parameters used in the deconvolution are calibrated specifically for our subject (Aquino et al., 2012; Pang et al., 2016), as discussed in Sec. 2.4. Therefore, the consequences of the model, such as the delays demonstrated in Fig. 8b, are well suited to understand the mechanisms of the brain of our subject and provide quantifiable hypotheses that can be tested in future studies by the wider neuroimaging community.
3.4. Application 3: Deconvolution of expanding arc data

To further demonstrate the applicability of the method, we next deconvolve an experimental BOLD response evoked by a visual stimulus consisting of an arc that expands outward from an eccentricity of \( E = 0.3^{\circ} \) to \( E = 5.5^{\circ} \) (see Fig. 5). The responses projected onto a smoothed left cortical hemisphere at \( t = 6, 12, \) and \( 18 \) s are shown in Fig. 9a. From this representation, we can clearly see the propagation of the responses across V1, starting from the fovea. Moreover, all the responses follow the timing and direction of propagation of the mapped visual stimulus. Their patterns are generally the same as in Fig. 8a but the peak activations are spatially limited to the size of the stimulus.

To better analyze the responses in V1, we process the data, following the details in Sec. 2.3.3, to focus on each response’s evolution near the centerline of the visual stimulus. This produces 1D responses that vary in eccentricity \( E \) and time \( t \). Representative time profiles of the normalized 1D responses for \( E = 1.5^{\circ} \) are shown in Fig. 9b. The times of peak response magnitudes have the same delays as those found in Fig. 8b and discussed in the previous section, which implies the generality of our observations. Spatiotemporal maps of the 1D responses in eccentricity-time space are shown in Fig. 9c, where the dashed line corresponds to \( E = 1.5^{\circ} \). We can clearly see that the peak response magnitudes move toward higher eccentricities with time, highlighting the propagation of the responses, as driven by the moving visual stimulus, consistent with an analysis of shock-like hemodynamic waves (Lacy et al., 2016). Averaging the data over a range of polar angles to yield 1D responses and visualizing them as in Figs 9b and 9c gives a more convenient picture of the dynamics. This is especially useful for studying data driven by a visual stimulus with an expanding pattern.

3.5. Implications for existing and future studies

Previously, Glover (1999) developed a deconvolution method to obtain the temporal variation of the neural activity from the BOLD response of event-related fMRI. Even though the method has its merits, it assumed that the spatial and temporal dynamics of the BOLD response are separable, which contradicts the propagating waves experimentally found in the cortex (see the Introduction). Moreover, it does not include the biophysics of the brain, hence its parameters cannot be tied to actual physiological mechanisms.

Here, we go beyond the work of Glover (1999) by having a more advanced and a more biophysically informed deconvolution method, with the added benefit that we can simultaneously obtain both the spatial and temporal dynamics of neurovascular signals linked to the BOLD response and not just those of the neural activity. Moreover, our method can be robustly used to analyze the BOLD response due to different types of stimuli and recorded using different fMRI field strength and resolution, provided that the model parameters are calibrated properly. It could also be used to facilitate experimental explorations of neurovascular delays, to acquire maps of different brain areas with increased sensitivity and specificity, and to quantify the contributions of the neurovascular signals to BOLD for different fMRI acquisition protocols.

However, we stress that our method does not take into account physiological confounds of nonneuronal origin, including cardiac and respiratory noise, vasomotion effects, and changes
Figure 9: Experimental BOLD-fMRI expanding arc data and deconvolved responses. (a) Responses on a smoothed left cortical hemisphere. Colors indicate levels from negative (blue) to positive (red). Black and brown solid lines represent the boundaries of V1 and V2, respectively. Rows from top to bottom show the retinotopically mapped visual stimulus, BOLD signal, and deconvolved responses, as labeled. Note that the stimulus in retinal space is also shown on top of the mapped visual stimulus panels. In each row, columns show the responses at $t = 6, 12, \text{ and } 18 \text{ s}$ with fixed color bar. (b) Normalized time profiles of the responses in V1 for an eccentricity of $E = 1.5^\circ$. The first row shows the profiles of the BOLD signal, neural activity, and neuroglial drive. The second row shows the profiles of CBF, CBV, and dHb concentration. The third row shows the profiles of the $W, L, \text{ and } D$ modes. (c) Evolution of the responses in V1 with time (ordinate) vs. eccentricity (abscissa). Color bars indicate levels and are the same to those in (a). The dashed lines are located at $E = 1.5^\circ$ for which the time profiles of the responses are shown in (b).
in blood pressure and cerebral autoregulation mechanisms (Murphy et al., 2013), or pre-cleaned data after artifact removal. This requires additional modeling and statistical analysis to quantify them, which is beyond the scope of this work. Hence, the method is currently more appropriate to be used for the analysis of data collected from task-based fMRI than those from resting-state fMRI.

3.6. Future improvements

There are a number of possible future improvements to our method to refine inferences of the neurovascular signals discussed in this work. (i) The results could be directly compared with those using data-driven methods [e.g., Glover (1999)] or using simultaneous multimodal neuroimaging techniques, as they become available, for cross-validation. (ii) The averaging of the model equations could be relaxed to achieve better resolution (<1 mm) by including the actual spatial distribution of the underlying anatomical vasculature. (iii) The deconvolution filter could be refined by optimizing the filter noise spectrum to minimize overestimations produced by the current method (see Sec. 3.1). Non-uniform effective cutoff frequencies could also be explored for different conditions, brain regions, subjects, and datasets. (iv) The values of the parameters could be calibrated for different brain areas of different subjects by employing Bayesian algorithms, culminating to a subject-specific whole-brain deconvolution analysis. (v) The hemodynamic model could be extended to study laminar-specific fMRI data and fine-scale structures such as ocular dominance and orientation preference columns. (vi) A model of precise population synaptic activity, e.g., the neural field model described by Sanz-Leon et al. (2018), in relation to hemodynamic effects could be incorporated in the method to better complement high-frequency neural measurements such as those by EEG.

4. Conclusions

We have developed a biophysically based method to deconvolve BOLD-fMRI data and simultaneously extract and image the spatiotemporal patterns of underlying neurovascular signals, i.e., neural activity, neuroglial drive, CBF, CBF, dHb concentration, and response modes. The deconvolution technique combines a Wiener filter with a physiological hemodynamic model that can accurately represent the properties of cortex. The conventional way to estimate these neurovascular signals is to either use neuroimaging modalities separately, which yield measurements that are difficult to compare because they inevitably capture different brain dynamics at different times, or combine multimodal equipment that relies on complicated setups and protocols. We have also demonstrated the applicability of the method to simulations and three different experimental datasets to illustrate its generality and utility, which are sufficient to establish the validity of the principles of the method. Our method thus enables a wide range of new capabilities of fMRI for imaging brain activity, hemodynamics, and physiology. This opens new noninvasive windows to understand and analyze dynamics of numerous stimulus-evoked physiological signals directly from fMRI measurements, thereby extracting additional information from BOLD-fMRI data. Moreover, this work demonstrates that the method provides a solid framework that produces testable predictions for the wider neuroimaging community in order to better understand brain function.
Appendix A. Spatiotemporal cortical hemodynamic model

This appendix summarizes the relevant equations of the spatiotemporal cortical hemodynamic model and the resulting transfer functions in Fourier space.

Appendix A.1. Equations of the hemodynamic model

The model is expressed as a set of interdependent equations that capture the dynamics of physical processes underlying the hemodynamic basis of blood oxygen level-dependent-functional magnetic resonance imaging (BOLD-fMRI) data. The key equations are summarized below and readers are advised to refer to our previous articles for detailed discussions and derivations (Drysdale et al., 2010; Aquino et al., 2012, 2014a,b; Pang et al., 2016; Lacy et al., 2016; Pang et al., 2017).

When neurons in a certain region of the brain are activated due to an external stimulus, a rise in neural activity, \( \phi(\mathbf{r},t) \), at position \( \mathbf{r} \) and time \( t \) can be quantified. It is important to note that \( \phi(\mathbf{r},t) \) only represents the low-frequency component of the synaptic activity, which the BOLD signal is most sensitive to. We are currently unable to capture high-frequency neural activity because the processes of neurovascular coupling responsible for the conversion of fast synaptic/spiking responses to slow dynamics remain to be unclear.

The signal from neural activity is then processed by surrounding astrocytes, resulting in a response called the neuroglial drive, \( \zeta(\mathbf{r},t) \); this drive affects the surrounding vasculature via the neurovascular coupling (NVC) (Koehler et al., 2006). The relation of \( \phi(\mathbf{r},t) \) and \( \zeta(\mathbf{r},t) \) is mathematically described as a spatiotemporal convolution (Pang et al., 2017)

\[
\zeta(\mathbf{r},t) = h(\mathbf{r},t) \otimes \phi(\mathbf{r},t) = \int \int h(\mathbf{r} - \mathbf{r}',t - t')\phi(\mathbf{r}',t')d\mathbf{r}'dt',
\]

(A.1)

where \( \otimes \) is the spatiotemporal convolution operator. The term \( h(\mathbf{r} - \mathbf{r}',t - t') \) is a weight function that quantifies how NVC spreads spatiotemporally. In our previous work (Pang et al., 2017), we approximated NVC as an impulse function in space and a delay function in time such that \( h(\mathbf{r},t) = \delta(\mathbf{r}) \delta(t - \tau_d) \), where \( \tau_d \) is an astrocytic time delay parameter that accounts for the stimulus-to-cerebral blood flow (CBF) delays observed in experiments (DeYoe et al., 1994; Buckner et al., 1996; Masamoto et al., 2015). Thus, the neuroglial drive becomes

\[
\zeta(\mathbf{r},t) = [\delta(\mathbf{r}) \delta(t - \tau_d)] \otimes \phi(\mathbf{r},t) = \phi(\mathbf{r},t - \tau_d),
\]

(A.2)

which means that the neuroglial drive is spatially the same as the neural activity but is delayed by \( \tau_d \).

The neuroglial drive affects the surrounding vasculature, leading to changes in CBF, represented by the arterial blood inflow rate, \( F \), which is governed by (Friston et al., 2000; Drysdale et al., 2010)

\[
\frac{\partial^2 F(\mathbf{r},t)}{\partial t^2} + \kappa \frac{\partial F(\mathbf{r},t)}{\partial t} + \left( \frac{1}{4} \kappa^2 + \omega_f^2 \right) [F(\mathbf{r},t) - F_0] = \zeta(\mathbf{r},t),
\]

(A.3)

where \( \kappa \) is the signal decay rate, \( \omega_f \) is the natural frequency of flow response, and \( F_0 \) is the resting blood inflow rate.
Changes in CBF and pressure result in an increase in cerebral blood volume (CBV), represented by the local blood mass density, $\Xi$, modeled by the nonlinear hemodynamic wave equation (Aquino et al., 2012, 2014a)

$$\frac{\partial^2 \Xi(r, t)}{\partial t^2} + \frac{D}{\rho_f} \frac{\partial \Xi(r, t)}{\partial t} - c_1 \nabla^2 P(r, t) = \rho_f \delta(z) \left( \frac{\partial}{\partial t} + \frac{D}{\rho_f} \right) \left[ F(r, t) - c_P P(r, t) \right], \quad (A.4)$$

where $D$ is the effective blood viscosity, $\rho_f$ is the blood density, $c_1$ is the pressure coupling constant, $\delta(z)$ is a Dirac delta function that expresses flows at the cortical surface at $z = 0$, $c_P$ is the blood outflow constant, and $P(r, t)$ is the blood pressure whose equation of state is approximated as a power law in $\Xi$ with an exponent $\beta$ related to the elasticity of cerebral vessels; $\beta$ is the reciprocal of Grubb’s exponent (Grubb et al., 1974; Drysdale et al., 2010).

In addition to the dynamics of CBF and CBV, the model also quantifies changes in the concentration of deoxygenated hemoglobin (dHb), $Q$, described by the dynamical equation (Lacy et al., 2016)

$$\frac{\partial Q(r, t)}{\partial t} = \eta \left[ \psi \Xi(r, t) - Q(r, t) \right] - \frac{Q(r, t)}{\Xi(r, t)/\rho_f} \delta(z) c_P P(r, t), \quad (A.5)$$

where $\eta$ is the fractional rate of oxygen consumption and $\psi$ is the ratio of hemoglobin concentration to blood density.

Finally, the BOLD signal, $Y$, is determined by the semi-empirical relation (Stephan et al., 2007)

$$Y(r, t) = V_0 \left[ k_1 \left( 1 - \frac{Q(r, t)}{Q_0} \right) + k_2 \left( 1 - \frac{Q(r, t) \rho_f V_0}{\Xi(r, t) Q_0} \right) + k_3 \left( 1 - \frac{\Xi(r, t)}{\rho_f V_0} \right) \right], \quad (A.6)$$

where $V_0$ is the resting blood volume fraction, $Q_0$ is the resting dHb concentration, and $k_1$, $k_2$, and $k_3$ are magnetic field parameters that depend on the properties of the fMRI scanner and experimental protocol used (Obata et al., 2004; Stephan et al., 2007).

**Appendix A.2. Linearized model**

Following Sec. 2.1.2 and our previous derivations (Aquino et al., 2012, 2014a), the resulting linearized equations of the model are

$$\frac{\partial^2 F(r, t)}{\partial t^2} + \kappa \frac{\partial F(r, t)}{\partial t} + \left( \frac{1}{4} \kappa^2 + \omega_f^2 \right) F(r, t) = \zeta(r, t), \quad (A.7)$$

$$\frac{\partial^2 \Xi(r, t)}{\partial t^2} + 2 \beta \frac{\partial \Xi(r, t)}{\partial t} - \nu_0^2 \nabla^2 \Xi(r, t) + \nu_0^2 \kappa^2 \Xi(r, t) = C_\beta \rho_f \left( \frac{\partial}{\partial t} + \frac{D}{\rho_f} \right) F(r, t), \quad (A.8)$$

$$\frac{\partial Q(r, t)}{\partial t} = \frac{Q_0}{\rho_f} \frac{\partial \Xi(r, t)}{\partial t} + \left[ C_{\beta} \frac{Q_0}{\Xi_0} \left( \eta + \tau^{-1} \right) \Xi(r, t) - \eta Q(r, t) \right]$$

$$- \tau^{-1} \left[ Q(r, t) + C_{\beta} \left( \beta - 1 \right) \frac{Q_0}{\Xi_0} \Xi(r, t) \right],$$

$$Y(r, t) = V_0 \left[ (k_2 - k_3) \frac{\Xi(r, t)}{\Xi_0} - (k_1 + k_2) \frac{Q(r, t)}{Q_0} \right], \quad (A.10)$$
where the superscript of the linear perturbations $\zeta^{(1)}$, $F^{(1)}$, $\Xi^{(1)}$, $Q^{(1)}$, and $Y^{(1)}$ have been dropped for notational brevity. In Eqs (A.7) to (A.10), $\Gamma$ is the wave damping rate, $\nu_\beta$ is the wave propagation speed, $k_z$ is the effective spatial frequency, $C_z$ is the outflow normalization constant, $\Xi_0$ is the resting blood mass volume, and $\tau$ is the hemodynamic transit time.

All parameters are detailed in Table A.1. For purposes of demonstration, here we use fixed nominal values of most of the parameters, which are taken from the literature (Aquino et al., 2014a), whereas the remaining parameters, i.e., $\nu_\beta$, $\Gamma$, $\omega_f$, $\kappa$, and $\tau_d$, are inferred in our previous studies for the primary visual cortex of the same individual subject (Aquino et al., 2012; Pang et al., 2016). Note that the parameters can all be varied but only the accuracy of the deconvolution results and not the actual framework of the method crucially depends on the values of these parameters.

Table A.1: Model parameters. In each row, columns are ordered from first to last to detail the variable, its symbol or formula, its mean or chosen value, and its units, respectively (Drysdale et al., 2010; Aquino et al., 2012, 2014a; Pang et al., 2017).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Symbol</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure coupling constant</td>
<td>$c_1$</td>
<td>$6 \times 10^{-8}$</td>
<td>-</td>
</tr>
<tr>
<td>Blood mass density</td>
<td>$\rho_f$</td>
<td>1062</td>
<td>kg m$^{-3}$</td>
</tr>
<tr>
<td>Average effective blood viscosity</td>
<td>$D$</td>
<td>591.1</td>
<td>kg m$^{-3}$ s$^{-1}$</td>
</tr>
<tr>
<td>Mean elasticity exponent of cortical vessels</td>
<td>$\beta$</td>
<td>3.2</td>
<td>-</td>
</tr>
<tr>
<td>Baseline CBF</td>
<td>$F_0$</td>
<td>0.01</td>
<td>s$^{-1}$</td>
</tr>
<tr>
<td>Hemodynamic transit time</td>
<td>$\tau$</td>
<td>1</td>
<td>s</td>
</tr>
<tr>
<td>Fractional oxygen consumption rate</td>
<td>$\eta$</td>
<td>0.4</td>
<td>s$^{-1}$</td>
</tr>
<tr>
<td>Hemoglobin concentration-blood density ratio</td>
<td>$\psi$</td>
<td>$1.8 \times 10^{-3}$</td>
<td>mol kg$^{-1}$</td>
</tr>
<tr>
<td>Resting blood volume fraction</td>
<td>$V_0$</td>
<td>0.03</td>
<td>-</td>
</tr>
<tr>
<td>Baseline CBV</td>
<td>$\Xi_0$</td>
<td>31.86</td>
<td>kg m$^{-3}$</td>
</tr>
<tr>
<td>Baseline dHb concentration</td>
<td>$Q_0$</td>
<td>$1.64 \times 10^{-2}$</td>
<td>mol m$^{-3}$</td>
</tr>
<tr>
<td>Magnetic field parameters at 3T, TE = 30 ms</td>
<td>$k_1$, $k_2$, $k_3$</td>
<td>4.2, 1.7, 0.41</td>
<td>-</td>
</tr>
<tr>
<td>Blood outflow constant</td>
<td>$c_p$</td>
<td>$1 \times 10^{-7}$</td>
<td>s$^{-1}$ Pa$^{-1}$</td>
</tr>
<tr>
<td>Average cortical thickness</td>
<td>$L$</td>
<td>$3 \times 10^{-3}$</td>
<td>m</td>
</tr>
<tr>
<td>Blood flow signal decay rate</td>
<td>$\kappa$</td>
<td>0.57</td>
<td>s$^{-1}$</td>
</tr>
<tr>
<td>Natural frequency of flow response</td>
<td>$\omega_f$</td>
<td>0.49</td>
<td>s$^{-1}$</td>
</tr>
<tr>
<td>Astrocytic delay</td>
<td>$\tau_d$</td>
<td>1.2</td>
<td>s</td>
</tr>
<tr>
<td>Average wave propagation speed</td>
<td>$\nu_\beta$</td>
<td>$2 \times 10^{-3}$</td>
<td>m s$^{-1}$</td>
</tr>
<tr>
<td>Average wave damping rate</td>
<td>$\Gamma$</td>
<td>0.8</td>
<td>s$^{-1}$</td>
</tr>
<tr>
<td>Perpendicular spatial frequency</td>
<td>$k_0 = \cos^{-1}(0.8)/L$</td>
<td>214</td>
<td>m$^{-1}$</td>
</tr>
<tr>
<td>Outflow normalization constant</td>
<td>$C_z = k_0 L / [3 \sin(k_0 L)]$</td>
<td>0.36</td>
<td>-</td>
</tr>
<tr>
<td>Average effective spatial frequency</td>
<td>$k_z = \sqrt{k_0^2 + \frac{1}{\nu_\beta^2} C_z^2 D^2}$</td>
<td>260</td>
<td>m$^{-1}$</td>
</tr>
<tr>
<td>Noise-to-signal ratio (NSR)</td>
<td>$\sigma$</td>
<td>0.5</td>
<td>-</td>
</tr>
</tbody>
</table>

**Appendix A.3. Transfer functions**

Following the linearization of the model equations, we can obtain the resulting transfer function $T_{AB}(k, \omega)$ that relates the change in hemodynamic quantity $A$ to the change in hemodynamic quantity $B$ in Fourier space. This is achieved by taking the Fourier transform of Eq. (A.2) and the linearized Eqs (A.7) to (A.10), which transforms the set of differential
equations into algebraic equations that can be solved for the transfer functions, which are

\[
T_{\zeta\phi}(k, \omega) = e^{-i\omega \tau_d}, \tag{A.11}
\]

\[
T_{F\zeta}(k, \omega) = \frac{1}{-\omega^2 - i\kappa \omega + \frac{1}{4}k^2 + \omega_f^2}, \tag{A.12}
\]

\[
T_{\Xi F}(k, \omega) = \frac{C_z \rho_f (D \rho_f - i\omega)}{-\omega^2 - 2i\Gamma \omega + \nu_\beta^2 k^2 + \nu_\beta^2 k^2_2}, \tag{A.13}
\]

\[
T_{Q\Xi}(k, \omega) = \frac{Q_0}{\Xi_0} \left\{ -iV_0 \omega + C_z [\eta - \tau^{-1} (\beta - 2)] \right\}, \tag{A.14}
\]

\[
T_{V\Xi}(k, \omega) = \frac{k_2 - k_3}{\rho_f} \left[ 1 - \frac{\Xi_0 k_1 + k_2}{Q_0 k_2 - k_3} T_{Q\Xi}(k, \omega) \right], \tag{A.15}
\]

where \(|k|^2 = k^2\) is the spatial wave number in the plane of the cortex and \(\omega\) is the temporal frequency.

We have previously shown that the BOLD response can be decomposed into five response modes (Pang et al., 2016). Thus, we can obtain five more transfer functions for the model that relate these modes of the BOLD response to \(\zeta\), with

\[
T_{n\zeta}(k, \omega) = \frac{a_n(k)}{\omega - \omega_n(k)} \tag{A.16}
\]

for the \(n\)th mode \((n = 1, \ldots, 5)\), where \(a_n(k)\) is a \(k\)-dependent amplitude and \(\omega_n(k)\) is the \(n\)th frequency pole; explicit expressions for \(a_n(k)\) and \(\omega_n(k)\) are given in Pang et al. (2016). We previously showed that these five modes can be rewritten in terms of three real-valued response modes: a propagating wave mode \(W\), a local-oscillating mode \(L\), and a local-decaying mode \(D\), as illustrated in Fig. 3c.

Using the definition of transfer functions, we can easily multiply them to obtain a new transfer function that relates a different pair of quantities. For example, if we have the transfer functions \(T_{AB}\) and \(T_{BC}\), we can straightforwardly relate \(A\) and \(C\) via the transfer function \(T_{AC} = T_{AB} T_{BC}\). Thus, the transfer functions to get \(A\) from \(\phi\), where \(A\) is \(\zeta\), \(F\), \(\Xi\), \(Q\), \(Y\), \(W\), \(L\), or \(D\), are

\[
T_{\zeta\phi}(k, \omega) = e^{-i\omega \tau_d}, \tag{A.17}
\]

\[
T_{F\phi}(k, \omega) = T_{F\zeta}(k, \omega) T_{\zeta\phi}(k, \omega), \tag{A.18}
\]

\[
T_{\Xi\phi}(k, \omega) = T_{\Xi F}(k, \omega) T_{F\zeta}(k, \omega) T_{\zeta\phi}(k, \omega), \tag{A.19}
\]

\[
T_{Q\phi}(k, \omega) = T_{Q\Xi}(k, \omega) T_{\Xi F}(k, \omega) T_{F\zeta}(k, \omega) T_{\zeta\phi}(k, \omega), \tag{A.20}
\]

\[
T_{V\phi}(k, \omega) = T_{V\Xi}(k, \omega) T_{\Xi F}(k, \omega) T_{F\zeta}(k, \omega) T_{\zeta\phi}(k, \omega), \tag{A.21}
\]

\[
T_{W\phi}(k, \omega) = [T_{1\zeta}(k, \omega) + T_{3\zeta}(k, \omega)] T_{\zeta\phi}(k, \omega), \tag{A.22}
\]

\[
T_{Y\phi}(k, \omega) = [T_{3\zeta}(k, \omega) + T_{4\zeta}(k, \omega)] T_{\zeta\phi}(k, \omega), \tag{A.23}
\]

\[
T_{D\phi}(k, \omega) = T_{5\zeta}(k, \omega) T_{\zeta\phi}(k, \omega). \tag{A.24}
\]

24
Conversely, the transfer functions to get $Y$ from $A$, where $A$ is $\phi$, $\zeta$, $F$, $\Xi$, $Q$, $W$, $L$, or $D$, are

\begin{align*}
T_{Y\phi}(k, \omega) &= T_{Y\Xi}(k, \omega)T_{YF}(k, \omega)T_{F\zeta}(k, \omega)T_{\zeta\phi}(k, \omega), \\
T_{Y\zeta}(k, \omega) &= T_{Y\Xi}(k, \omega)T_{YF}(k, \omega)T_{F\zeta}(k, \omega), \\
T_{YF}(k, \omega) &= T_{Y\Xi}(k, \omega)T_{YF}(k, \omega), \\
T_{Y\Xi}(k, \omega) &= \frac{k_2 - k_3}{\rho_f} \left[ 1 - \frac{\Xi_0 k_1 + k_2}{Q_0 k_2 - k_3} T_{Q\Xi}(k, \omega) \right], \\
T_{YQ}(k, \omega) &= T_{Y\Xi}(k, \omega)T_{Q\Xi}^{-1}(k, \omega), \\
T_{YW}(k, \omega) &= T_{Y\zeta}(k, \omega) \left[ T_{1\zeta}(k, \omega) + T_{2\zeta}(k, \omega) \right]^{-1}, \\
T_{YL}(k, \omega) &= T_{Y\zeta}(k, \omega) \left[ T_{3\zeta}(k, \omega) + T_{4\zeta}(k, \omega) \right]^{-1}, \\
T_{YD}(k, \omega) &= T_{Y\zeta}(k, \omega)T_{D\zeta}^{-1}(k, \omega),
\end{align*}

where the superscript $-1$ corresponds to taking the reciprocal in Fourier space.

Appendix A.4. One-dimensional (1D) version of the equations

The equations of the model generally represent responses on a 2D cortical surface. However, experiments have shown that a linear section of the primary visual cortex will be stimulated if presented with a ring stimulus centered in the subject’s visual field, due to the retinotopic mapping between the visual field and visual cortex (Aquino et al., 2012; Schira et al., 2009). This stimulus produces responses that vary only in the direction perpendicular to its locus on the cortex. To analyze responses due to variants of the abovementioned stimulus, we can assume that the responses are impulsive with respect to one spatial dimension (i.e., $y$-dimension). Thus, the relevant equations and transfer functions in 1D coordinate space become spatially dependent only on $x$ and $k_x$, respectively, where $x$ is the coordinate perpendicular to the centerline of the stimulus and $k_x$ is the 1D spatial wave number. The 1D version is especially useful in Secs 3.1 and 3.2.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

This work was supported by the Australian Research Council Center of Excellence for Integrative Brain Function (ARC Center of Excellence grant CE140100007), the Australian Research Council Laureate Fellowship Grant FL140100025, and the Australian Research Council Discovery Project Grant DP170101778.
References


26


Logothetis, N. K., 2008. What we can do and what we cannot do with fMRI. Nature 453 (7197), 869–878.


dynamics. J. R. Soc. Interface 13 (118), 20160253.