Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS

Jeffrey W. Barker,¹,²,* Ardalan Aarabi,¹,³ and Theodore J. Huppert¹,²

¹Department of Radiology, University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, PA 15260, USA
²Department of Bioengineering, University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, PA 15260, USA
³GRAMFC, Faculty of Medicine, University of Picardie-Jules Verne, 3 rue des Louvels, Amiens Cedex, 80036, France
*jwb52@pitt.edu

Abstract: Systemic physiology and motion-induced artifacts represent two major sources of confounding noise in functional near infrared spectroscopy (fNIRS) imaging that can reduce the performance of analyses and inflate false positive rates (i.e., type I errors) of detecting evoked hemodynamic responses. In this work, we demonstrated a general algorithm for solving the general linear model (GLM) for both deconvolution (finite impulse response) and canonical regression models based on designing optimal pre-whitening filters using autoregressive models and employing iteratively reweighted least squares. We evaluated the performance of the new method by performing receiver operating characteristic (ROC) analyses using synthetic data, in which serial correlations, motion artifacts, and evoked responses were controlled via simulations, as well as using experimental data from children (3-5 years old) as a source baseline physiological noise and motion artifacts. The new method outperformed ordinary least squares (OLS) with no motion correction, wavelet based motion correction, or spline interpolation based motion correction in the presence of physiological and motion related noise. In the experimental data, false positive rates were as high as 37% when the estimated p-value was 0.05 for the OLS methods. The false positive rate was reduced to 5-9% with the proposed method. Overall, the method improves control of type I errors and increases performance when motion artifacts are present.

© 2013 Optical Society of America

OCIS codes: (170.2655) Functional monitoring and imaging; (300.0300) Spectroscopy; (100.2960) Image analysis; (170.5380) Physiology.

References and links

1. F. F. Jöbsis, “Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters,” Science 198, 1264–1267 (1977).
2. M. Cope, D. T. Delpy, E. O. Reynolds, S. Wray, J. Wyatt, and P. van der Zee, “Methods of quantitating cerebral near infrared spectroscopy data,” Adv. Exp. Med. Biol. 222, 183–189 (1988).
3. I. Miyai, H. C. Tanabe, I. Sase, H. Eda, I. Oda, I. Konishi, Y. Tsunazawa, T. Suzuki, T. Yanagida, and K. Kubota, "Cortical mapping of gait in humans: a near-infrared spectroscopic topography study," Neuroimage 14, 1186–1192 (2001).
1. Introduction

Functional near infrared spectroscopy (fNIRS) is a non-invasive brain imaging technique that is based on the spectroscopic measurement of the optical absorption of cerebral blood [1]. Evoked changes in tissue hemoglobin concentrations are detected by an array of light sources and detectors on the scalp that monitor changes in optical absorption and scattering. Both oxy-hemoglobin (HbO₂) and deoxy-hemoglobin (Hb) concentration changes can be estimated from the modified Beer-Lambert law [2] using absorption measurements taken at multiple wave-
lengths of light within the optical window (650-850nm). In comparison to other cerebral vascular imaging methods such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), fNIRS imaging is considerably lower cost and more portable. Because it is non-ionizing and non-confining, fNIRS is particularly well suited for non-invasive studies in children or infants. In addition, the portability of fNIRS and its ability to record brain signals during moderate subject movement has allowed its use during various tasks such as walking [3, 4], balance [5–7], or social interaction [8], which otherwise could not be imaged by conventional fMRI.

Functional NIRS imaging is accomplished through a set of spatially arranged optical sources and detectors (optodes), which are often coupled through fiber optics between a head cap worn by the participant and the fNIRS instrument. Although fNIRS can be used to image brain activity during participant movement, fNIRS is prone to artifacts introduced by slippage of these optodes and headcap on the surface of the scalp. These artifacts are particularly common in studies involving children or infants where the fNIRS cap is more loosely fastened on the head to increase participant tolerance of the method. A great deal of effort has been spent by researchers in developing better methods for identification and removal of motion artifacts, including principal component analysis [9], spline interpolation [10], wavelet transform [11], and Kalman filtering [12] based algorithms. These methods were compared by Cooper et. al. [13], in which spline interpolation and wavelet based filtering gave the best results for two different performance metrics. As a note, these methods are applied in pre-processing of the fNIRS data and in general have some level of subjective tuning involved in the implementation of the algorithm. In addition to motion artifacts, fNIRS signals are also contaminated by superficial and systemic physiological signals such as cardiac, respiratory, and blood pressure oscillations. Because fNIRS signals are measured from sensors placed on the surface of the scalp, these measurements have a high sensitivity to superficial changes in optical absorption. These slow background oscillations contribute to serially correlated noise in fNIRS measurements. This effect is increased by the high sampling rate of fNIRS instruments (~10Hz or more) that is substantially faster than the underlying physiological signals.

Motion artifacts and serial correlations due to physiology violate common statistical assumptions of the independence of repeated measurements over time. Subsequently, these sources of noise can contribute to inaccurate estimation of type I errors. In this work, we describe an iterative algorithm that utilizes an autoregressive model based pre-whitening filter and robust regression to reduce the effects of physiological and motion-induced sources of noise, respectively. We demonstrated that this algorithm has better performance than ordinary least squares with no motion correction, wavelet based motion correction [11], or spline interpolation based motion correction [10] in the presence of physiological noise and motion artifacts for both deconvolution (e.g., HOMER [9]) and canonical (e.g., SPM-NIRS [14]) based regression models. Our proposed algorithm is demonstrated for both synthetic and experimental data containing motion artifacts. In addition, we tested both short (event-related design) and long (block design) duration tasks as well as varying magnitudes of hemoglobin concentration changes (contrast to noise ratios).

2. Methods

2.1. Theory

The simplest model of the observed fNIRS signal is an evoked response with a normally distributed error term as given by

\[ y_t = h_t + \varepsilon_t \]  

\[ \varepsilon_t \sim N(0, \sigma^2) \]  

(C) 2013 OSA 1 August 2013 | Vol. 4, No. 8 | DOI:10.1364/BOE.4.001366 | BIOMEDICAL OPTICS EXPRESS 1368
where \( y_t \) is the observed signal; \( h_t \) is the evoked hemodynamic response; \( \epsilon_t \) is the noise at time \( t \); and \( \sigma^2 \) is the variance of \( \epsilon_t \). The additive noise term in the fNIRS signal is often contaminated with a high degree of physiological noise, which is sampled at the high temporal resolution of fNIRS instruments leading to serially correlated error terms. In other words, slow physiological oscillations are sampled rapidly causing nearby time points in the fNIRS signal to be highly correlated. The presence of these errors can greatly affect the false positive rate (e.g., type I error) of the model and lead to inaccurate interpretations of the data. In standard linear regression methods, serially correlated errors are often dealt with by pre-whitening based on an autoregressive (AR) model of the error terms. A first-order model (AR(1)) of the error terms is

\[
\epsilon_t = \rho_1 \epsilon_{t-1} + \nu_t \quad (2a)
\]

where the error term at the current time point \( (t) \) depends on the error term from the previous sample point \( (t-1) \). Estimating the coefficients of the AR model and designing a filter that is applied to both sides of the equation can remove the correlation between subsequent time points. For the AR(1) model, a simple filter can be constructed as \( f = [1 - \rho_1]^T \) and applied to Eq. (2a) to produce

\[
f * y_t = (y_t - \rho_1 y_{t-1}) = (h_t - \rho_1 h_{t-1}) + \nu_t, \quad (3)
\]

where the correlated error terms have been transformed by subtraction, leaving only the white noise term \( (\nu_t = \epsilon_t - \rho_1 \epsilon_{t-1}) \). This transformation of the data \( (y) \) and model \( (h) \) using an AR(1) model of the residual error terms is the Cochrane-Orcutt method [15], which has been used in several studies [16–18] to remove serial correlations in fNIRS analysis. This method can easily be generalized to any AR model order [19] as follows:

\[
\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \ldots + \rho_p \epsilon_{t-p} + \nu_t \quad (4a)
\]

\[
f = [1 - \rho_1 - \rho_2 - \ldots - \rho_p]^T \quad (4b)
\]

\[
(f * y_t) = (f * h_t) + \nu_t. \quad (4c)
\]

In practice, the AR coefficients can be estimated by fitting an AR model to the residual errors \( (\epsilon_t) \) themselves. The optimal AR model order should be sufficient to remove serially correlated errors without overfitting. The model order can be chosen objectively by minimizing an information criterion function. For example, the Bayesian information criterion (BIC) [20] function is

\[
BIC(P) = -2 LL + P \ln(n), \quad (5)
\]

where \( P \) is the model order; \( LL \) is the log-likelihood of the model fit; and \( n \) is the number of time points.

In the context of fNIRS brain imaging, a linear model of the evoked response [21] can be written as

\[
y = X\beta + \epsilon, \quad (6)
\]

where \( X \) is the design matrix, and \( \beta \) is the estimate of the magnitude of brain activity. In the case of a finite impulse response (FIR) model, \( X \) is given by the convolution matrix of a series of delta functions describing the timing of the stimulus onsets [9]. The coefficients \( (\beta) \) then describe the magnitude of the evoked hemodynamic response at each time point. The FIR model is equivalent to block averaging in the case of long inter-stimulus intervals or deconvolution in the case of closely spaced inter-stimulus intervals [22]. Note that boxcar basis functions describing the whole duration of the task can alternatively be used in FIR models and are typically one...
option of basis function in fMRI analysis; however, we chose not to investigate this option. For canonical linear models (e.g., the statistical parametric modeling approach; SPM-NIRS [14]), $X$ is given by the convolution of the stimulus blocks with a canonical hemodynamic response. In this case, a limited number of coefficients ($\beta$) are used to allow statistical testing of differences in brain activity (e.g., the amplitude of $\beta$) between different task conditions. In both cases, additional regressors such as the time course of systemic physiology or motion terms can also be included in the design matrix to model nuisance terms (e.g., [23]). Regardless of whether an FIR or canonical design matrix is assumed, the values of $\beta$ (e.g., brain activity) can be estimated by the inversion of $X$ in Eq. (6). The Gauss-Markov equation for the unbiased ordinary least squares (OLS) estimator is

$$\beta = (X^TX)^{-1}X^Ty$$  \hspace{1cm} (7a)

$$\text{cov}(\beta) = \sigma^2(X^TX)^{-1},$$  \hspace{1cm} (7b)

where $\text{cov}(\beta)$ is the covariance matrix of $\beta$; $\sigma^2$ is the variance of the residual ($\epsilon$); and the superscript $T$ denotes the transpose. The linear model (Eq. (6)) can be corrected for serially correlated errors by pre-whitening with the filter $f$ generated via Eq. (4b) giving the expression

$$Fy = FX\beta + F\epsilon,$$  \hspace{1cm} (8)

where $F$ is the convolution matrix of $f$ and performs column-wise filtering of the measurements in $y$ and of the regressors in $X$, leading to whitened residuals ($F\epsilon$). The least squares solution to this model is

$$\beta = (X^TF^TX)^{-1}X^TF^Ty$$  \hspace{1cm} (9a)

$$\text{cov}(\beta) = \sigma^2(X^TF^TX)^{-1}.$$  \hspace{1cm} (9b)

As a practical note, the matrix $F$ is approximately the size of the vector $y$ in both dimensions. This is often too large to form explicitly; however, the matrix $F$ is sparse with only entries of the filter coefficients on the diagonals, which can be implemented as a sparse matrix. A more efficient implementation is to employ digital filtering methods as discussed in section 2.2.

The methods above can correct for serial correlations due to physiology; however, motion artifacts also affect the error terms by producing outliers to a normal distribution. In this work, we propose to use iteratively reweighted least squares [24] to deal with outliers caused by motion. In this approach, the influence of each time point is weighted based on the value of the residual by a weighting function, such as Tukey’s bisquare function [25] given by

$$w(r) = \begin{cases} 
1 - \left(\frac{r}{\kappa \sigma}\right)^2 & |r| < \kappa \\
0 & |r| \geq \kappa
\end{cases},$$  \hspace{1cm} (10)

where $\kappa$ is the tuning parameter with a default value of 4.685 that preserves 95% statistical efficiency in the case of normally distributed residuals (i.e., no outliers). Adjusting the tuning parameter based on statistical efficiency gives an objective criteria for determining the tuning parameter. For example, the tuning parameter could be decreased to the point of maintaining 85% statistical efficiency if more robustness against outliers is desired. The weighted least squares (WLS) solution is

$$\beta = (X^TF^TFX)^{-1}X^TF^TFy$$  \hspace{1cm} (11)

where $W$ is diagonal matrix of weights calculated from the filtered residuals as follows:

$$[W]_{ii} = w([Fy - FX\beta]_i).$$  \hspace{1cm} (12)
Regression is done iteratively, updating the weight matrix $W$ with each iteration until a convergence criterion is reached. As a practical note, robust regression is implemented as part of the statistics toolbox in Matlab (Mathworks R2012b, Natick MA) as the function “robustfit”. The iteratively reweighted least squares estimator has an asymptotic covariance matrix given by

$$\text{cov}(\beta) = \sigma^2 \frac{E[\psi^2(F\varepsilon/\sigma)]}{E[\psi^2(F\varepsilon/\sigma)]^2} (X^TF^FX)^{-1},$$

(13)

where $E$ denotes expectation, and $\psi$ is the influence function [26], which can be specified as $\psi(r) = r w(r)$.

After regression, the null hypothesis that the hemodynamic response was zero (e.g., $\beta = 0$) over a given time window or set of regressors can be tested by defining a contrast vector ($c$) and calculating the t-statistic given by the expression

$$t = \frac{c^T \beta}{\sqrt{c^T \text{cov}(\beta) c}}.$$  

(14)

In the case of an FIR (deconvolution or block averaging) model, the contrast vector might be of value 1 for some predefined window of time (e.g., over the expected peak of the response) and 0 otherwise. In the case of the canonical design matrix, the contrast vector would be 1 for the column of the design matrix representing the regressor (task condition) of interest. A contrast vector can also be defined using positive and negative terms to test differences between two conditions.

### 2.2. Summary of algorithm

For each fNIRS channel:

1. Initialize $\beta$ via an OLS fit (Eq. (7a)).
2. Fit the residual ($y - X\beta$) to an AR(P) model where P minimizes BIC (Eq. (5)).
3. Generate the whitening filter $f$ via Eq. (4b).
4. Apply the whitening filter to the data $y$ and column-wise to the design matrix $X$.
5. Perform iteratively reweighted least squares (IRLS).
   
   (a) Solve for $\beta$ by WLS (Eq. (11a)).
   
   (b) Recalculate weight matrix $W$ via Eq. (12).
   
   (c) Repeat step 5a-b until changes in $\beta$ are sufficiently small (e.g., < 1% change).
6. Repeat steps 2-5 until changes in $\beta$ are sufficiently small (e.g., < 1% change).

Note that it is much more computationally efficient to perform the filtering in step 4 via a digital filtering algorithm, such as the “filter” function in Matlab, rather than explicitly forming the convolution matrix $F$. The filtered data and design matrix can then be passed to any software implementation of IRLS, such as the “robustfit” function in Matlab.

There are many approaches to estimating the AR model coefficients. The current default in Matlab is to minimize the squared prediction errors of the forward and reversed signals. Because motion artifacts are likely to influence the results, we modified this approach to a weighted least squares method, in which the weighted squared prediction errors of the forward and reversed signal were minimized. This was performed as follows: (1) find the ordinary least
(1) solve the weighted least squares problem. This approach was computationally efficient enough to perform a direct search of the AR model order that minimizes BIC. The entire algorithm ran in \( \sim 3s \) for the canonical regression model and \( \sim 20s \) for the deconvolution model, which required a much larger design matrix to invert, for one subject’s data (24 channels, \( \sim 1500 \) data points/channel) on a single core of a 2.9 GHz Intel Xeon E5-2690 CPU.

2.3. Simulated data

In order to test the performance of our proposed method under controlled noise structures, we first applied the method to synthetic data, in which noise, motion artifacts, and evoked response were simulated. We simulated serially correlated noise via tools in Matlab’s Econometrics toolbox with an AR(5) model resulting in a signal with standard deviation of 0.05\( \mu M \), which was comparable to the real data set discussed in the next section. The AR model order of 5 used for simulated data was close to the median optimal model order of 6 for the experimental data. The AR coefficients were taken as the median values across subjects from an AR(5) model fit to the experimental data. Two sets of AR coefficients were used in the simulations based on AR(5) model fits to the HbO\(_2\) and Hb channels, respectively. Half of the data were simulated using the HbO\(_2\) AR coefficients, and half of the data were simulated using the Hb AR coefficients.

Simulations were performed with and without simulated motion artifacts. Data that were simulated with motion artifacts contained either 5-10 spike artifacts or 1-3 shift artifacts at random times. The spike artifacts were modeled with a Laplace distribution function given by

\[
 f(t) = A \exp \left( -\frac{|t-t_0|}{b} \right),
\]

with random peak amplitude, \( A \), of 0.25-1\( \mu M \) and random scale parameter, \( b \), of 0-1.5s. Shift artifacts were modeled as a random positive or negative change in DC value of 0.125-0.5M with a linear transition lasting between 0.25-1.5s. All random parameters described above were drawn from uniform distributions. In order to quantify the performance of the various regression methods, a known simulated evoked response was added to a subset (50\%) of the channels. After regression, a receiver operating characteristic (ROC) analysis was performed to quantify the true positive rate (sensitivity) as a function of false positive rate (1-specificity). We simulated both a short duration task consisting of 15 trials of 1s stimuli every 20s and a long duration task of 10 trials of 10s stimuli every 30s. Evoked responses ranging from 0.01-0.10M in peak amplitude were generated by convolving the stimulus design with the canonical hemodynamic response function and added to the simulated noise. Evoked responses were added at half magnitude to simulated deoxy-hemoglobin channels (\( \Delta[Hb] = \Delta[HbO_2]/2 \)). We simulated \( > 100,000 \) random channels of data at each magnitude of evoked response with exactly half containing an evoked response.

2.4. Experimental data

In addition to the completely synthetic dataset, we used experimental “resting state” data from a study involving 22 normal, healthy children (3-5 years old) as a source of real physiological noise and motion artifacts. All subjects provided written parental informed consent, and this data was collected as part of a larger study of functional brain activity [27]. Each “resting state” scan lasted approximately three minutes and was acquired while the children watched a short video on the computer (Disney films). Although this is not true resting state, there were no specific cued tasks during the scan, thus serving as an effective estimate of background physiology for the purpose of this study. The experimental data were acquired at 690nm and 830nm wavelengths at a sampling rate of 4Hz with a probe consisting of 24 source-detector

---

Received 25 Apr 2013; revised 7 Jun 2013; accepted 8 Jun 2013; published 17 Jul 2013

(C) 2013 OSA

1 August 2013 | Vol. 4, No. 8 | DOI:10.1364/BOE.4.001366 | BIOMEDICAL OPTICS EXPRESS 1372
measurements over the prefrontal cortex. The optical density changes were converted to oxy- and deoxy-hemoglobin prior to input into the analysis method via the modified Beer-Lambert law. No other preprocessing was performed.

In order to test the performance of the various regression methods, a known simulated evoked response was added to the experimental “resting state” data on a subset of the channels (50%), and an ROC analysis was performed. Evoked responses were generated using the event-related (1s on, 19s off, 15 trials) and block design (10s on, 20s off, 10 trials) paradigms. Each channel of data was concatenated with itself in order to accommodate the length of the experimental designs (300s). We simulated evoked responses with amplitudes ranging from 0.01-0.1M on the oxy-hemoglobin channels. Evoked responses were added at half amplitude to the deoxy-hemoglobin channels. For each of the 22 subjects, we simulated 200 trial data sets (>100,000 channels), in which exactly half (chosen randomly) of the source-detector pairs contained an added evoked response in order to perform ROC analysis.

2.5. Analysis methods

We tested the performance of the proposed AR(P)-IRLS algorithm as well as OLS with and without AR(1) pre-whitening. Additionally, we tested wavelet based motion correction [11] followed by OLS regression (Wavelet-OLS below) and spline interpolation based motion correction [10] followed by OLS regression (Spline-OLS below) in order to validate the performance of the proposed method’s ability to correct motion artifacts. The wavelet based motion correction was implemented using matlab function “hmrMotionCorrectWavelet” that is included in the HOMER2 software package (www.nmr.mgh.harvard.edu/DOT/resources/software.html) using an interquartile range of 1.5, as suggested in the function’s documentation. The spline based motion correction was implemented using the functions “hmrMotionCorrectSpline” and “hmrMotionArtifactByChannel” in HOMER2. The parameters for this method were set to AMPThresh = 0.5, tMotion = 0.5s and tMask = 2s following reference [13]. The optimal parameter for SDThresh varied from 12-20 in reference [13] depending on the subject, thus we set SDThresh = 16. Receiver operating characteristic (ROC) curves were generated for each of the tested methods by varying the estimated p-value threshold for activation from 0 to 1 and calculating the true positive rate (sensitivity) and false positive rate (1-specificity). Partial area under the ROC curve (AUC 0.05) was used as a metric for performance, in which the true positive rate vs false positive rate curve is integrated up to a false positive rate of 0.05. In this case, a random estimator would result in AUC 0.05 = 0.00125, and a perfect estimator would result in AUC 0.05 = 0.05. Partial AUC is potentially a better indicator of overall performance than AUC, which integrates the entire curve, since the threshold for statistical significance is often set to 0.05 or less. Plots of false positive rate as a function of estimated p-value were generated and examined to evaluate control of type I errors (false positive rate). We estimated standard errors of AUC 0.05 via bootstrap resampling of the data 1,000 times with replacement and calculating the mean and standard deviation of 1,000 estimates of AUC 0.05.

3. Results

The AR model based pre-whitening filters removed serially correlated errors in the optical data. In Fig. 1(a), an illustrative example of a simulated fNIRS signal with a few spike artifacts and a shift artifact is shown. Figure 1(b) shows the same data after pre-whitening with an AR(2) model (determined optimal by BIC) generated from Eq. (4b). Figure 1(c) shows the autocorrelation function of the original (Fig. 1(a)) and whitened (Fig. 1(b)) signals, in which the whitened signal showed significantly reduced signal autocorrelations. After filtering, the simulated artifacts were visually identifiable as outliers in the whitened signal. These time points are subsequently given little or zero weight during the iteratively reweighted least squares
regression. Similarly, an illustrative example of an experimental fNIRS signal is shown in Fig. 1(d). After filtering with an AR(2) model (determined optimal by BIC), the whitened signal (Fig. 1(e)) showed significantly reduced autocorrelations (Fig. 1(f)). For the experimental data in this study, the lower quartile, median, and upper quartile of the AR model order chosen to minimize BIC were 5, 7, and 9 for HbO$_2$ channels and 3, 5, and 8 for Hb channels. In order to briefly test the model order dependence on sampling rate, we downsampled (without filtering) the experimental data to 4/5Hz, 1Hz, 4/3Hz, and 2Hz. We observed that the median optimal model order increased with increasing sampling rate.

Fig. 1. A simulated fNIRS signal generated from an AR(5) process with simulated motion artifacts is shown in (a). After generating an optimal pre-whitening filter via fitting an AR(2) model, the whitened signal (b) has significantly reduced autocorrelations (c). An experimental fNIRS signal is shown in (d). After generating an optimal pre-whitening filter via fitting an AR(2) model, the whitened signal (e) has significantly reduced autocorrelations (d).

Figure 2 shows example hemodynamic response functions recovered via deconvolution analysis for the block task (a-c) and event task (d-f). These examples used the experimental data as a baseline signal with a simulated evoked response of 0.04µM amplitude. The first column (Figs. 2(a,d)) shows examples when hemodynamic response function produced by OLS appears to be free of artifacts. In this case, the hemodynamic response functions produced by all methods appeared to be of high quality. The middle column (Figs. 2(b,e)) shows examples when the hemodynamic response recovered by OLS appeared to be mildly contaminated by motion artifacts. In this case, Wavelet-OLS and AR(P)-IRLS completely removed the suspected artifacts, while Spline-OLS provided partial attenuation of the artifacts. The last column (Figs. 2(c,f)) shows examples, in which the hemodynamic response recovered by OLS was severely corrupted by artifacts and/or noise. Wavelet-OLS and AR(P)-IRLS showed significant improvement in the visual appearance of the recovered hemodynamic response, while Spline-OLS failed to improve the results visually.
Figure 2. Examples of recovered hemodynamic response functions for simulated block design (a-c) and event-related (d-f) design using the experimental data as baseline physiology/noise.

Figure 3 shows the results of ROC analysis of the deconvolution model performance for the block task (Fig. 3(a)) and event tasks (Fig. 3(b)). The data is shown for simulated evoked responses of 0.04µM amplitude. In the case of synthetic data (AR(5) process) with no simulated motion artifacts (AR/None), the performance was similar for all methods tested with AR(1)-OLS and AR(P)-IRLS showing small, but statistically significant increases in AUC over OLS for both the block and event task. In the case of synthetic data with simulated spike artifacts (AR/Spike), AR(P)-IRLS gave the best performance, followed by Wavelet-OLS. The AR(1)-OLS results also showed a small, but statistically significant increase in AUC over OLS for both the block and event task. For synthetic data with simulated shift artifacts (AR/Shift), AR(P)-IRLS showed a much larger performance over the other methods for both block and event tasks. The Spline-OLS method showed a small, but statistically significant increase in AUC over OLS for both the block and event task. For experimental data, only AR(P)-IRLS showed a significant increase in performance over OLS.

Figure 4 illustrates the type I error control of OLS, AR(1)-OLS, and AR(P)-IRLS for the same data as Fig. 3. The type I errors for Wavelet-OLS and Spline-OLS were similar to OLS and are not shown. For all cases, OLS showed severely inflated type I errors, with false positive rates of 33–43% when the p-value was estimated to be 0.05. In other words, false positive rates were 7–8 times higher than predicted. Prewhtening with an AR(1) filter (AR(1)-OLS) significantly reduced type I errors; however, false positive rates were still slightly higher than predicted for the experimental data, in which false positive rates were 17% and 11% for the block and event task, respectively, when the p-value was estimated to be 0.05. The AR(P)-IRLS method provided excellent type I error control for the synthetic data with no artifacts and the experimental data for both the block and event task. For synthetic data with simulated spike or shift artifacts, AR(P)-IRLS overestimated type I errors.

Figure 5 shows the results of ROC analysis of the canonical regression model performance for the block task (Fig. 5(a)) and event tasks (Fig. 5(b)). The data is shown for simulated evoked responses of 0.04µM amplitude. In the case of synthetic data with no simulated motion artifacts (AR/None), the performance was similar for all methods tested. In the case of synthetic data...
Fig. 3. Partial AUC ($AUC_{0.05}$) for detection of evoked responses with a deconvolution/FIR model for simulated block (a) and event (b) tasks using an AR model as baseline signal with no artifacts (AR/None), spike artifacts (AR/Spike), or shift artifacts (AR/Shift) or with experimental data as a baseline signal containing motion artifacts. Error bars indicate 99% confidence interval.

Fig. 4. False positive rate of detection as a function of estimated p-value (i.e., estimated false positive rate) with the deconvolution/FIR model for simulated block (top row) and event (bottom row) tasks using a simulated AR model as baseline signal with no artifacts (AR/None), spike artifacts (AR/Spike), or shift artifacts (AR/Shift) or with experimental data as a baseline signal containing motion artifacts.
with simulated spike artifacts (AR/Spike), AR(P)-IRLS gave the best performance, followed by Wavelet-OLS. For synthetic data with simulated shift artifacts (AR/Shift), AR(P)-IRLS showed a large increase in performance over the other methods for both the block and event task, while AR(1)-OLS showed a small increase in performance for the block task and a moderate increase in performance for the event task. For the experimental data, AR(1)-OLS increased performance over OLS for the event task, but decreased performance for the block task. The AR(P)-IRLS method showed a moderate increase in performance over OLS for both the block and event task.

Figure 5. Partial AUC for detection of evoked responses with the canonical regression model for simulated block (a) and event (b) tasks using a simulated AR model as baseline signal with no artifacts (AR/None), spike artifacts (AR/spike), or shift artifacts (AR/shift) or with experimental data as a baseline signal containing motion artifacts. Error bars indicate 99% confidence interval.

Figure 6 illustrates the type I error control of OLS, AR(1)-OLS, and AR(P)-IRLS for the same data as Fig. 5. The type I errors for Wavelet-OLS and Spline-OLS were similar to OLS and are not shown. For all cases, OLS showed severely inflated type I errors, with false positive rates of 32-38% when the p-value was estimated to be 0.05. Prewhitening with an AR(1) filter (AR(1)-OLS) significantly reduced type I error estimation. The AR(P)-IRLS method provided slightly better type I error estimation than AR(1)-OLS for the synthetic data with no artifacts or with spike artifacts. For the case of synthetic data with shift artifacts, the AR(P)-IRLS method overestimated type I errors. For the experimental data, AR(1)-OLS and AR(P)-IRLS performed similarly in terms of type I error estimation.

The results shown in Figs. 3-6 were for one magnitude of evoked response ($\Delta[HbO_2] = 0.04 \mu M, \Delta[Hb] = 0.02 \mu M$). We tested a range of evoked response magnitudes ($\Delta[HbO_2] = 0.01-0.10 \mu M$). In general, the false positive rates were not dependent on the magnitude of the evoked response, but on the underlying noise structure. The gain in performance (AUC$_{0.05}$) by using AR(P)-IRLS over OLS was dependent on the amplitude of the evoked response, or more specifically the contrast to noise ratio (CNR), in which all methods approach random performance (AUC$_{0.05} = 0.00125$) as CNR decreases and ideal performance (AUC$_{0.05} = 0.05$) as CNR increases. In general, correcting for motion artifacts will offer the most benefits in the middle range of CNR. The CNR of a typical NIRS experiment will vary greatly by experimental paradigm, number of trials, and design and location of the probe; however, for all of
the simulations tested in this work, the AUC<sub>0.05</sub> for AR(P)-IRLS was greater than or equal to AUC<sub>0.05</sub> of OLS.

4. Discussion

In this study, we demonstrated a general algorithm for solving the GLM in the context of deconvolution (FIR) and canonical regression models for fNIRS that combines two well established statistical methods: AR model based data transformations (filtering) and iteratively reweighted least squares. These techniques were combined in a simple, computationally efficient algorithm to correct for two major sources of confounding noise in fNIRS analyses: serial correlations caused by systemic physiology and large outliers caused by motion artifacts, respectively. Furthermore, we rigorously tested the algorithm by performing ROC analysis using synthetic data, in which serial correlations, motion artifacts, and evoked responses could be controlled via simulations. Finally, we validated our simulation results by using experimental data as a source of real baseline physiological noise and motion artifacts.

Although this approach to pre-whitening via an AR model based transformation is common practice in linear regression, this exact approach is not widely used in neuroscience applications. In fMRI analysis (which typically deals with data sets on the order of 64x64x30 = 122,880 voxels) the estimation of an AR model for each voxel is not practical due to the high computational time required. For example, in the fMRI analysis program SPM8 [21] a global estimate of an AR(1) model is included in the parameterization of the covariance matrix in the context of restricted maximum likelihood (ReML) rather than fitting each voxel independently. In fMRI analysis, this approximation is justified because the time between subsequent measurements is much longer (~2-3s) than fNIRS, such that serial correlations are less of a concern. In the case of fNIRS data, which has both higher sample rates and lower number of time series measurements, these approximations are no longer appropriate.

The use of simulated AR noise with simulated artifacts allowed us to control the type, number, and amplitude of motion artifacts. The simulated artifacts were relatively large compared to experimental data and provided a good test of the AR(P)-IRLS method’s ability to correct...
for motion artifacts. The AR(P)-IRLS method outperformed spline interpolation based motion correction [10] and wavelet based motion correction [11] methods when motion artifacts were present. Despite showing visual improvements in the recovered hemodynamic response functions, such as in Fig. 2, Wavelet-OLS and Spline-OLS did not offer any statistical advantage in the ROC analysis with the experimental data. It is possible that the parameters used for these methods were not optimal for the experimental data set in this study. Further ROC analyses should be performed for data collected under a wide range of experimental conditions and subject populations to validate and generalize the comparison of these methods.

In simulations without artifacts the algorithm did not negatively affect the analyses. The practical implication of this is that the algorithm can be applied to every data set without making a subjective judgment about whether artifact correction is needed or not. The method will improve performance when data is contaminated by motion and preserve normal performance when data are free of artifacts. Overall, our results suggest that application of robust regression methods after pre-whitening represents a reliable method for correction of motion artifacts.

An AR(1) model of the errors was insufficient for controlling type I errors in some cases. This effect is likely to be a greater issue at higher sampling rates than in this study (>4Hz). Using an adaptive AR model order for pre-whitening is simple to implement and computationally efficient enough to be added to standard fNIRS analysis methods to improve type I error control.

For the deconvolution model, AR(P)-IRLS method provided accurate estimation of false positive rates for the synthetic data with no artifacts and the experimental data, while overestimating type I errors for synthetic data with simulated spike or shift artifacts. For the canonical regression model AR(P)-IRLS method provided accurate estimation of false positive rates for the synthetic data with no artifacts or spike artifacts and the experimental data, while overestimating type I errors for synthetic data with shift artifacts. While it is potentially better to overestimate than underestimate type I errors, this may lead to reduced sensitivity near the threshold of statistical significance (estimated p-value < 0.05). This overestimation is possibly related to violation of stationarity (assumed by AR models) caused by severe motion artifacts, especially shift artifacts, or to the breakdown of the asymptotic covariance estimate in Eq. (13). This issue did not appear to be a problem for the simulations using experimental data with real motion artifacts, in which the estimated p-values were relatively accurate. The statistics may potentially be improved if shift artifacts can be removed or identified during preprocessing.

5. Conclusions

In conclusion, we have developed a general algorithm for both the deconvolution (FIR) and canonical regression analysis models that is robust to two major sources of noise in fNIRS: systemic physiological signals and motion artifacts. The algorithm preserves the performance of standard OLS methods when data are free of motion artifacts, thus freeing the user from making a subjective choice about artifact correction. Lastly, the algorithm contains a single tuning parameter for the weighting function used in IRLS that can be set objectively.

Acknowledgments

This study was funded by the NIH (grant no. 5R01EB013210). The authors would like to thank Dr. Susan Perlman for providing the experimental data set.