

Neurophotonics

Neurophotonics.SPIEDigitalLibrary.org

Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy

Theodore J. Huppert

Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy

Theodore J. Huppert*

University of Pittsburgh, Center for the Neural Basis of Cognition, Clinical Science Translational Institute, Departments of Radiology and Bioengineering, Pittsburgh, Pennsylvania 15260, United States

Abstract. Functional near-infrared spectroscopy (fNIRS) is a noninvasive neuroimaging technique that uses low levels of light to measure changes in cerebral blood oxygenation levels. In the majority of NIRS functional brain studies, analysis of this data is based on a statistical comparison of hemodynamic levels between a baseline and task or between multiple task conditions by means of a linear regression model: the so-called general linear model. Although these methods are similar to their implementation in other fields, particularly for functional magnetic resonance imaging, the specific application of these methods in fNIRS research differs in several key ways related to the sources of noise and artifacts unique to fNIRS. In this brief communication, we discuss the application of linear regression models in fNIRS and the modifications needed to generalize these models in order to deal with structured (colored) noise due to systemic physiology and noise heteroscedasticity due to motion artifacts. The objective of this work is to present an overview of these noise properties in the context of the linear model as it applies to fNIRS data. This work is aimed at explaining these mathematical issues to the general fNIRS experimental researcher but is not intended to be a complete mathematical treatment of these concepts. ©

The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.NPh.3.1.010401]

Keywords: near-infrared spectroscopy; brain imaging; statistical analysis; linear regression.

Paper 15051NR received Nov. 16, 2015; accepted for publication Jan. 22, 2016; published online Mar. 2, 2016.

1 Introduction

Functional near-infrared spectroscopy (fNIRS) is a noninvasive brain imaging technique that uses low levels of red to near-infrared light to measure changes in oxy- and deoxy-hemoglobin in the brain.¹ Biological tissue in the range of around 650 to 900 nm (termed the “near-infrared window”) has low intrinsic absorption and allows light to remain detectable after passing through up to centimeters of tissue. In this wavelength range, tissue is highly turbid (scattering), which results in a stochastic and diffuse path as the photons migrate through the tissue. fNIRS brain imaging is accomplished by placing light sources on the scalp of the head. These transmit light into the tissue, where it diffuses through the volume and can reach around 5 to 8 mm into the cortex of the brain before exiting the tissue and being sensed by light detectors placed generally between 2 and 3.5 cm from the source. Changes in the optical absorption of the tissue along this diffuse path from source to detector will change the intensity of light received by the detector, which is recorded as a time series, usually at sample rates of several hertz or above. Measurements at two or more optical wavelengths within this near-infrared window are used to provide a distinction between oxy- and deoxy-hemoglobin changes based on the intrinsic absorption profile differences of these two chromophores and the modified Beer–Lambert law.²

During a cognitive task, there is typically an increase in local neuronal activity in the brain. This increased activity drives an

increase in oxygen metabolism and blood flow, which results in a change in blood oxygen saturation and hemoglobin concentrations in the brain. This hemodynamic response changes the optical absorption of tissue and provides signal contrast in the fNIRS measurements taken from source-to-detector pairs that cross this region of the brain. The measured fNIRS signals are then statistically compared between two or more conditions, which may be a baseline/rest condition and a specific task or to compare the differences between multiple tasks.

Compared with functional magnetic resonance imaging (MRI), fNIRS is more portable and allows for a wider range of studies that would be difficult to perform in the immobile environment of the functional MRI (fMRI) scanner. For example, fNIRS has been previously used in functional imaging of children and infants (reviewed in Lloyd-Fox et al.³), clinical and bedside imaging (reviewed in Obrig⁴), and for a number of unique tasks such as movement/walking,⁵ operating a car,⁶ or interpersonal interactions,⁷ to name only a few. While compared with fMRI, the fNIRS technique offers lower spatial resolution (usually around 2 to 3 cm), a limited depth of penetration (5 to 8 mm into the brain), and greater sensitivity to superficial physiological noise and motion artifacts, this technology has been shown to be a complementary alternative tool to conventional fMRI for these sorts of unique applications. The number of fNIRS studies has been steadily growing over the last decade (see Ferrari and Quaresima⁸ and Boas et al.⁹ for reviews).

2 The Linear Model

During a typical functional brain study, a participant will perform repeated trials of a specific task or tasks. A statistical

*Address all correspondence to: Theodore J. Huppert, E-mail: huppert@upmc.edu

model is then used to detect differences in the brain signals between pairs of tasks or task/baseline. The most common model used is a linear regression model. This model assumes that the brain response to the task(s) is linearly additive and that each task changes the signal by the same amount regardless of the level of baseline or residual brain activity at the time the task was performed. In other words, the linear assumption is that each time the task occurs, the brain changes by the same amount without saturation effects or interactions between trials. These assumptions are generally believed to be true for most studies, although there is some evidence for nonlinearities in the brain response related to the underlying neural response and the physical response of blood vessels that have to dilate each time blood flow is increased. In particular, these effects are observed when there are very short intervals between tasks related to neural refraction¹⁰ or very long sustained task performance due to neural habituation or to the biophysical Windkessel effect on blood vessel compliance.¹¹

Under the assumptions of the linear model, an fNIRS experiment composed of repeated trials of a cognitive event can be analyzed by using a linear regression model. The standard linear model of brain activity is described by the general equation^{12,13}

$$Y = X \cdot \beta + \varepsilon, \quad (1)$$

where Y represents the measurement vector. This is a column vector with one entry per measurement point. In the case of fNIRS, this is a vector containing either the discrete time-series measurements of optical absorption for each wavelength or the estimates of oxy- and deoxy-hemoglobin. X is termed the design matrix. This matrix encodes the timing of the stimulus events as regressors in the linear model. In other words, the fNIRS signal is modeled as the weighted sum of the columns of this matrix (regressors), with each column representing a different time course of interest (e.g., a task component). The design matrix may also contain additional nuisance regressors to model trends or drift in the data, motion artifacts, or other covariates of the model. This matrix should additionally have a constant (DC) column, which models the mean of the signal. The variable β defines the unknowns in the model and is a column vector of the weights (coefficients) of the regression model. The goal of the regression is to estimate these weights based on the data vector Y and the design matrix X . If the weight β associated with a particular regressor (e.g., a specific task component) is statistically nonzero, then that regressor is important in modeling the data. Thus, the statistical map of these weights associated with task components is generally interpreted as indicating the brain regions that statistically change based on the task (e.g., brain activity). Finally, the term ε defines the measurement error. The specific properties of this noise term are essential to understanding the specific application of the linear model to fNIRS data and in what ways fNIRS analysis deviates from other modalities such as fMRI. The specific statistical properties of fNIRS noise are the topic of the remainder of this work.

2.1 Solving the Linear Model

The linear regression model [Eq. (1)] is typically solved by use of the Gauss–Markov expression, which yields the solution to the least-squares fit of the data (Y) under the assumption of spherical, uncorrelated, and homoscedastic noise. These assumptions will be discussed in Sec. 3. Under these

assumptions, the best linear unbiased estimator of the coefficients is given by

$$\beta = (X^T \cdot X)^{-1} \cdot X^T \cdot Y. \quad (2)$$

The uncertainty (covariance) of these coefficients is given by the expression

$$\text{Cov}_\beta = (X^T \cdot X)^{-1} \cdot \sigma^2, \quad (3)$$

where σ^2 is the mean-squared error of the residual from the model fit of the data ($\sigma^2 = \langle \varepsilon^2 \rangle$).

Based on the value of the coefficients and their covariance, statistical T -tests can be used to test if the coefficients are nonzero or if multiple coefficients differ from each other (e.g., the difference is nonzero). The T -test on these coefficients is then given by the expression

$$T = c \cdot \beta / \sqrt{c \cdot \text{Cov}_\beta \cdot c^T}, \quad (4)$$

where c is the contrast vector. The contrast vector denotes the specific test being performed. For example, to test if the first coefficient is nonzero in a model that has, e.g., four total coefficients, the contrast vector would be $c = [1, 0, 0, 0]$. Likewise, to test the difference of two coefficients (e.g., the difference of the first and third regressor) in the model, the contrast vector would be of the form $c = [1, 0, -1, 0]$.

3 Noise in Functional Near-Infrared Spectroscopy Is Nonspherical

The Gauss–Markov unbiased estimator [Eq. (2)] assumes zero-mean, spherical noise. Noise is said to be spherical when the errors have uniform variance (homoscedasticity) and are uncorrelated. As we will discuss in Secs. 3.1–3.4, fNIRS data violates these assumptions.

3.1 The Noise in Functional Near-Infrared Spectroscopy Is Correlated

Typical fNIRS data have strong physiological noise due to signals such as cardiac, respiratory, and blood pressure changes. These physiological signals are typically much slower than the sample rate of an fNIRS system. This implies that the noise in fNIRS data has a colored structure and serial correlations. Colored noise refers to the fact that specific temporal (or spatial) frequencies are present in the noise spectrum. In contrast, white noise is characterized by a flat (uniform) noise spectrum. Colored noise results in serial correlations in the error term ε , which means that each sample point is not independent from the others and depends on its self history. Thus, the effective degrees of freedom of the linear model are often much less than the number of sample points. This can also result in a potential bias in the estimate of the coefficients if the noise is not mean zero. While fMRI also has colored noise due to the same systemic physiology, since the sample rate of fNIRS (up to tens to hundreds of Hz) is typically higher than fMRI (generally around 0.5 Hz), and since fNIRS is more sensitive to superficial physiology in the scalp, this noise tends to affect fNIRS measurements more strongly. Figure 1(a) shows an example of an fNIRS-recorded oxy-hemoglobin signal demonstrating physiological noise from both cardiac and blood pressure oscillations. In Fig. 1(c), the autocorrelation of this data is

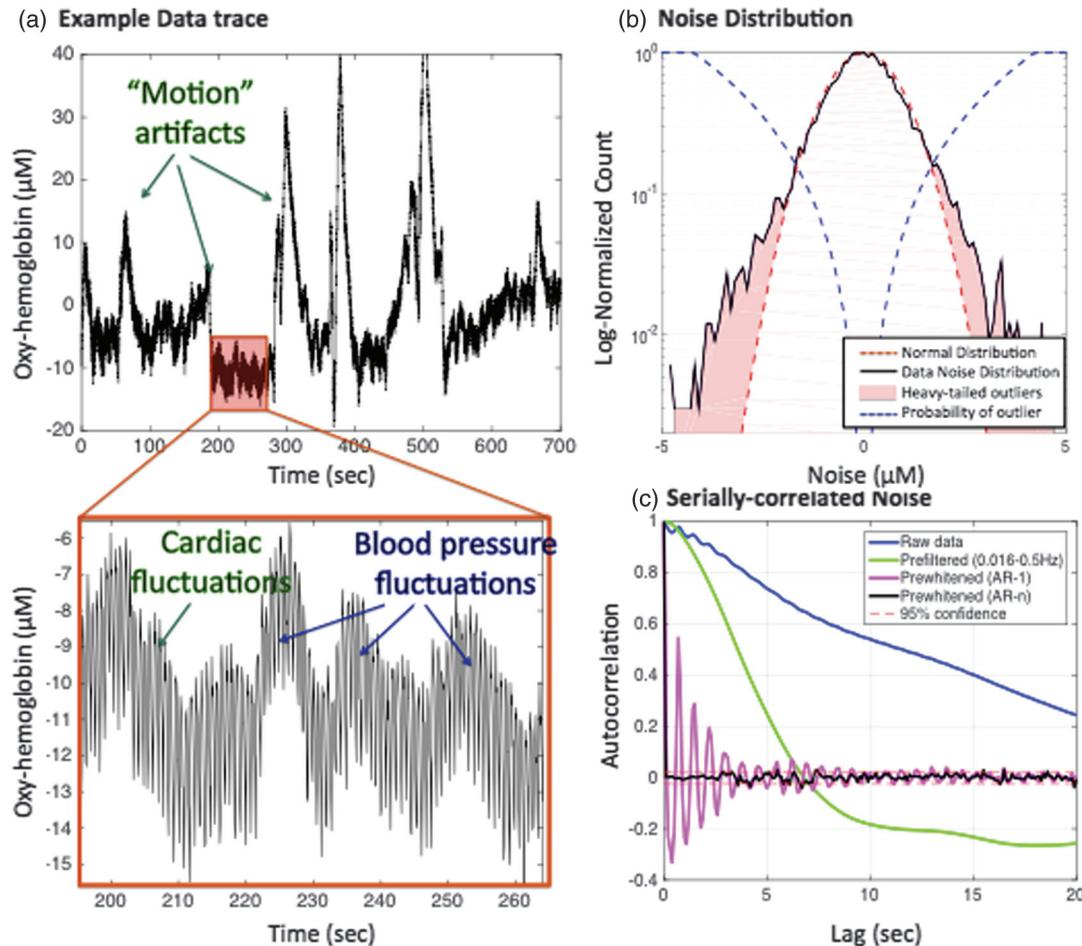


Fig. 1 Example of noise in fNIRS data. This figure shows an example of an fNIRS time course of oxy-hemoglobin. The zoomed view (a, inset) shows physiological noise due to cardiac and blood pressure changes. Typical motion artifacts (a) are clearly visible on this data as well. (b) A histogram noise distribution of the data and the best-fit ideal distribution (dotted red line). This is shown in logarithm scale for the y-axis to highlight the deviations at the tails of this distribution (highlighted in red), which constitute heavy-tailed noise associated with the motion artifacts. (c) The temporal autocorrelation of this raw data along with the autocorrelation after filtering and AR prewhitening (described in Sec. 4), which are used to reduce these serially correlated errors.

demonstrated. As can be seen, the autocorrelation of the raw signal remains significantly above chance well beyond the 20 s of lag time shown in the figure. This shows that two data points sampled 10 s apart in this file still share on average 60% of the information (down to 20% after 20 s) and cannot be considered independent measurements. Thus, in this sample data in Fig. 1, although there are over 7000 time points (700 s at 10 Hz) in this data trace, there are far fewer than this number of independent degrees of freedom, thus corrections must be made to account for this in the statistical model and estimate of the false-discovery rates for reports of brain activity, which we will further detail in Sec. 6 of this work.

3.2 Functional Near-Infrared Spectroscopy Data Is Not Independent Across Measurement Channels

In fNIRS, correlations are also found between different source-detector pairs due to the typically low spatial frequencies of systemic and superficial physiological signals. Motion artifacts can also give rise to strong spatial noise structures. In addition,

measurements are correlated between optical wavelengths and estimates of oxy- and deoxy-hemoglobin. Oxy- and deoxy-hemoglobin are partially correlated variables, which are together related to underlying changes in neural activity and the resulting blood flow and oxygen metabolism changes in the brain. Thus, unlike fMRI, fNIRS analysis generally involves a multivariate problem with partially correlated variables. More formally, oxy- and deoxy-hemoglobin are nonlinear correlated as related to the underlying dynamics of blood vessel dilation and oxygen transport. However, this topic is considered beyond the scope of this discussion. Spatial covariance and spatially structured noise have an impact on the assumptions of group level, region-of-interest, and image reconstruction procedures used in fNIRS.

3.3 Functional Near-Infrared Spectroscopy Data Exhibits Heteroscedasticity

Noise is said to have homoscedasticity when all the samples arise from the same noise distribution. In other words, the variance of the noise is uniform. In general, this is not true for fNIRS data, which is therefore said to be heteroscedastic. Over the time

dimension, motion artifacts often confound fNIRS data. A motion artifact, which may often be considerably larger than physiological noise, constitutes a heavy-tail to the noise distribution. These artifacts arise from the movement of the fNIRS sensors on the head and transient changes in the contact of these sensors with the scalp. A heavy-tailed noise distribution is one in which there are a few strong noise components compared to the normal noise background. These strong noise terms are outliers in the noise model. Such noise is heteroscedastic because these outlier samples are from a different distribution than the rest of the noise. The example data trace shown in Fig. 1(a) shows several motion artifacts. The variance of this oxy-hemoglobin signal evaluated over the time window free of motion-artifacts from 200 to 260 s [inset; ($\sigma^2 = 3.5 \mu\text{M}^2$)] is a factor of 25 less than the period from 300 to 360 s ($\sigma^2 = 95 \mu\text{M}^2$), which has several large motion artifacts. Thus, this data demonstrates noise from at least two distinct distributions. In Fig. 1(b), the noise distribution of this data is shown on a logarithm scale. The best fit of the normal distribution for this data is shown in red. Heavy-tailed noise elevates the tails of this distribution and is highlighted in red/shaded on this figure, which deviates from the assumed normal statistical model. We note that not all motion artifacts are necessarily outliers from the normal distribution of noise, however, and since these specific types of artifacts do not necessarily violate the homoscedastic assumption of the noise model, these are not of specific concern to the solution to the linear model. Although such noise would lower effect sizes and reduce the power to detect brain activity, homoscedastic noise would not necessarily result in an inaccurate statistical model or uncontrolled false-discovery.

In addition, the noise over the spatial dimension across fNIRS channels (source–detector pairs) is also typically nonuniform. Specific source–detector pairs or measurements at a particular wavelength may vary considerably in the level of noise compared with the other measurements. This noise is largely determined by how well the sensors are placed on the head, contact with the skin, the presence of hair under the fNIRS probe, and so on. In practice, it is not uncommon to have an order of magnitude difference or more in the signal-to-noise ratio between channels with good and poor contact on the head. Thus, the noise across fNIRS channels should not be assumed to be normal and exhibits heteroscedasticity. Due to the small number of fNIRS measurement channels compared with voxels in an fMRI scan, the central limit theorem generally does not apply in the fNIRS spatial domain. In the analysis methods used in the program NIRS-SPM,¹⁴ noise covariance estimates are performed by pooling the data across channels with the iterative restricted maximal likelihood (ReML) method. The ReML method is used to estimate both the noise covariance and the autoregressive (AR) coefficients within the covariance structure used in the prewhitening procedure (discussed in Sec. 4 of this work). Because this method obtains only one set of hyperparameters for oxy- and deoxy-hemoglobin that define the covariance structure applied to the whole channel set (e.g., across channels with varied qualities of optode-head contact), this estimate can be skewed by having one or a few channels that have very different noise from the other channels, as demonstrated in Sec. 6 of this work. As a result, the use of ReML in the NIRS-SPM software requires an additional preprocessing step, described in the user manual, to remove “bad” channels prior to computing the ReML covariance models. This step is not

generally required in fMRI analysis, since the spatial noise of fMRI is more uniform and can otherwise be spatially smoothed (spatial precoloring). However, in fNIRS data, the noise per optical source–detector pair can vary significantly, thus the same noise model is often not sufficient for analysis (we will return to this concept in Sec. 6 of this paper). In addition, spatial heteroscedastic noise across the fNIRS measurements is of concern in group-level analyses, where corresponding channels are analyzed across subjects, and in analysis pooling data from multiple channels on the probe such as region-of-interest or image reconstruction methods.

3.4 Functional Near-Infrared Spectroscopy Data May Be Nonergodic and Nonstationary

fNIRS noise due to systemic physiology may also appear to change over the length of a scan or study as a reflection of very slow underlying physiological changes (e.g., due to posture, blood pressure, and so on). Of more concern in analysis, however, systemic noise may also change corresponding to the stimulus task. During physical tasks such as motor movement (e.g., finger tapping) or tasks such as walking or exercise, changes in respiration, heart-rate, and blood pressure may be altered in a task-dependent manner. Thus, the noise during these time blocks may have different properties (mean, variance, or spectral content). This type of noise is said to be nonstationary or nonergodic. Specifically, because systemic physiology can change in a task-dependent manner, the noise structure of the “baseline” can be different from the “task,” which violates assumptions of the T -task and the estimation of brain activity under these conditions. The issue of nonstationary noise in fNIRS has been addressed in a number of publications in the context of dynamic estimation models such as a Kalman filter,^{15–18} but since there is no simple solution to dealing with these noise errors, we will simply note that this is a potential source of error that could cause fNIRS signals to violate the common assumptions in most statistical models used to date in analysis, and very little is known about how much of an effect this has on the sensitivity and specificity of fNIRS methods.

In Fig. 1(a), we show an example of an fNIRS data trace of oxy-hemoglobin from an experimental study (unpublished). This is a fairly representative trace of the quality of data that we often see in our own studies. This particular time course was from a child imaging study and shows a few more motion artifacts than one might see in an adult fNIRS study since often for adults, the head cap can be placed tighter and there is a bit better subject compliance. While a zoomed view (inset figure) shows clear physiological signals related to cardiac and respiratory fluctuations, the full time course also shows larger but less frequent noise fluctuations from the motion of the fNIRS head cap relative to the head (indicated in the figure). These motion artifacts are almost an order of magnitude larger than the variance of other parts of the data (e.g., physiological noise shown in the zoomed view) and much greater than expected from physiological fluctuations. These infrequent, larger fluctuations deviate from uniform normal statistical distribution and create a heavy-tailed noise distribution, as shown in Fig. 1(b), which shows the best fit of a normal distribution to this data and highlights the deviations due to these outliers at the tail of this distribution (red/shaded). A metric such as the Tukey’s bisquare weight (also shown in panel B as the dotted blue line) can be used to estimate the probability of these values at the tails of the normal distribution as outliers, which is

iteratively used in robust regression methods, discussed in Sec. 5 of this paper. As further shown in Fig. 1(c), the structures in the noise also generate serially correlated errors in which the current sample point depends on previous samples and the assumption of independent noise is violated. In this example trace, the raw data maintains autocorrelation for more than 20 s of lag (200 sample points at 10 Hz).

4 Correcting Nonspherical Correlations in Functional Near-Infrared Spectroscopy Noise

As previously stated, Eqs. (2) and (3) make assumptions about the noise in the linear model being uncorrelated that are violated in most fNIRS data sets by the presence of systemic physiology and to a lesser extent by motion artifacts. A standard solution to deal with nonideal colored noise is to apply a filter to the model that transforms the noise into a normal distribution, then to apply Eq. (2) to this filtered model.^{13,14,19–22} These modifications to the linear regression model attempt to make the statistical approach applicable to more general noise structures and are termed the general linear model (GLM) approach. They were developed for fMRI data analysis within the program SPM^{12,13,21,23} and later adapted to fNIRS analysis.^{14,19,24,25}

4.1 Noise Prewhitening

Prewhitening of a regression model is a process in which the data and model are reweighted to remove colored noise from the residual ε . The prewhitening expression is given by

$$W \cdot Y = W \cdot X \cdot \beta + W \cdot \varepsilon, \quad (5)$$

where both the left- and right-hand sides of Eq. (1) have been multiplied by a whitening matrix W . W is selected such that the residual of the model $W\varepsilon$ is now white and the assumptions of uncorrelated noise in the model are no longer violated. As an example, if the fNIRS measurements contain a slow physiological component (e.g., 1/10 Hz; the so-called Mayer waves), then W is designed to downweight these frequencies in the noise ε and remove this correlation. This matrix is applied to both the measurement Y and the design matrix X , which means that when this whitened model is solved, the down-weighted components contribute less to the estimate of β . In a sense, this can be thought of as a high-pass filter applied to both sides of the equation. The solution to Eq. (5) can be obtained by substitution into Eqs. (2) and (3) with the new model (e.g., $Y_w = WY$ and $X_w = WX$), and is given by

$$\beta = (X_w^T \cdot X_w)^{-1} \cdot X_w^T \cdot Y_w. \quad (6)$$

As a further note, prewhitening is not the same as prefiltering. Prefiltering is a preprocessing step in which a high-pass filter is applied only to the data. For example, a 1/60 Hz high-pass filter is often used in fNIRS to remove slow drifts from measurements. As demonstrated in Fig. 1, some of the observed temporal autocorrelation can be reduced by high-pass filtering, but this may introduce anticorrelations in the data, as shown in Fig. 1(c) by the negative correlation at a 5-s lag and beyond. In comparison to prewhitening, prefiltering is applied only to the measurements (Y and ε terms), but not the design matrix X . Since this filtering is not applied to X , the estimate of β is no longer unbiased. Thus, prefiltering should only be used if one has confidence that the frequencies being removed were not relevant to

model of the data (e.g., $X\beta$). Since this is not the case in prewhitening, which is applied to both sides of the expression, prewhitening is generally recommended over prefiltering to avoid a biased estimate of β .

The whitening matrix W can be specified either based on prior expectations about the frequency content of the noise in the data or empirically by iteratively fitting the model and examining the residual. The prior case requires the user to specify a specific frequency filter. For example, a 0 to 1/60 Hz discrete cosine transform (DCT) might be used to whiten slow drifts in the data akin to a high-pass filter. In this case, W is constructed from the residual forming matrix of the filtering regressors [e.g., $W = I - D \cdot (D^T \cdot D)^{-1} \cdot D^T$, where I is an identity matrix and D is the matrix constructed from the DCT regressors]. The advantage of using filters based on prior expectations on the noise structure is that it is fast and computationally simple. In fMRI, e.g., iterative estimation of a whitening filter on a per-voxel basis would be time consuming. Prior defined whitening filters such as the DCT filter also have the advantage of having a more straightforward interpretation in terms of control over exactly which frequencies are affected in the model. Prewhitening using a DCT filter [or similar, such as the minimum descriptive length (MDL) wavelet method²⁵] is used in programs such as SPM¹² and NIRS-SPM.¹⁴ Predefined whitening filters, however, may not be entirely efficient. For example, if there were no low frequencies in the noise, one is unnecessarily discarding information from the data, but in principle does not change the estimate of β , although its uncertainty will likely increase because less information is being used in the estimate.

An alternative is an iterative approach that estimates W directly from the residual noise in the data, which can offer a more efficiently whitened noise spectrum, but is more computationally intensive. In the iterative methods, the linear model is first fit using no whitening (e.g., $W = \text{identity}$). The residual noise is then estimated from the model fit via Eq. (6), and the whitening matrix is found based on the current residual. The whitened model is refit and continued until convergence. A common approach to estimating the whitening filter is to fit an autoregressive model to the residual.^{19,20,22,24} An autoregressive model will remove periodic noise structures (e.g., the dependence of a current time point to its self history). Additional components such as moving-average (ARMA) or integrative (ARIMA) models can also be used to additionally model drift and nonstationarities in the noise. The whitening model (AR or other) is fit to the residual of the linear regression, then used to construct the whitening matrix W . As a slight variation of this approach, in the SPM¹² and NIRS-SPM¹⁴ programs, the AR model is currently implemented as an off-diagonal structure in the covariance components used in the iterative ReML method and estimated using an expectation-maximization algorithm. Here, the same AR model is applied across all spatial channels and is often used in addition to the DCT or MDL prewhitening methods.

Figure 1(c) shows the effect of prewhitening on the temporal autocorrelation of a sample NIRS data file. As demonstrated in this figure, AR models can further reduce this autocorrelation, but require a fairly high model order to completely whiten the noise. In this case of 10 Hz data, a model order of $n = 33$ was determined by Bayesian information criteria for optimal whitening. This model order is notably larger than typically used for similar corrections in fMRI research. In comparison, the default AR model order used in the ReML prewhitening step of

NIRS-SPM¹⁴ is $n = 1$, which, as we will show in Sec. 5 of this paper, corrects some but often not all of these serially correlated errors. The first order AR model ($n = 1$), shown in Fig. 1(c), reduces the significant autocorrelation to below a 5-s lag but introduces some ringing in the autocorrelation and negative correlations in the structure of the noise.

4.2 Noise Precoloring

Noise precoloring is an alternative to prewhitening that has been used in fMRI analysis including SPM¹² and NIRS-SPM.¹⁴ In noise precoloring, a known noise color spectrum is applied to the data to mask over the natural but unknown noise.²⁶ Although there is now more colored structure in the noise, it has a known profile and can be accounted for by changing the statistical model. Similar to prewhitening, the temporal precoloring matrix is applied on both sides of Eq. (1) to yield an expression similar to Eq. (5). While prewhitening is akin to a high-pass filter, precoloring is similar to a low-pass filter and imposes additional autocorrelation structure on the noise that overpowers the intrinsic (unknown) structure.

Pre-coloring assumes homoscedastic noise and is adversely affected by the presence of heavy-tailed and outlier noise points. By imposing strong autocorrelation structures on the noise, outlier noise points can be smeared, affecting large portions of the data. This is similar to the effect a low-pass filter would have on an outlier spike in the data (e.g., due to motion). Thus, precoloring in the presence of heavy-tailed noise distributions can increase the overall noise of the data and be counterproductive. This blurring also reduces the effectiveness of outlier regression (robust regression) methods, as described in Sec. 5. Thus, in general, prewhitening is recommended over precoloring for fNIRS data.

5 Correcting Heteroscedastic Noise in Functional Near-Infrared Spectroscopy

Prewhitening of the linear model via Eq. (5) reduces the effect of correlations in the noise of the fNIRS data. However, even once the correlations are effectively dealt with, noise will still have

outliers due to motion artifacts within the time dimension. After prewhitening, these outlier artifacts will typically be sharpened in the noise. Thus, prewhitening is necessary, but not sufficient, to correct nonspherical errors in fNIRS data.

Heteroscedasticity is when the noise comes from a nonuniform distribution, as in the case of outliers. Robust regression is an iterative reweighting method to correct these errors by down-weighting outliers in the noise.²⁷ In this method, the model is first solved, yielding an estimate of the residual noise. The noise is then studentized (divided by the standard deviation of the noise) to provide a measure of how far each individual noise term is from the mean and variance of the remaining noise. For example, strong outliers will be several standard deviations away from the distribution of the remaining noise points. As a note, not all motion is necessarily a strong outlier. However, only outliers bias the model estimates, thus motion artifacts that are not heteroscedastic compared to the rest of the noise do not necessarily need to be corrected. A weighting function such as the bisquare or Huber loss function is then applied to downweight time points that are outliers. The resulting weight matrix S is a diagonal matrix. Similar to the prewhitening and precoloring operations, this matrix is applied to both the left and right sides of the regression model, yielding

$$(S \cdot W) \cdot Y = (S \cdot W \cdot X) \cdot \beta + (S \cdot W) \cdot \epsilon \quad (7)$$

in the case of the prewhitened and unweighted regression model. The model is solved for β and the residual $S \cdot W \cdot \epsilon$. This process is repeated until convergence and is described in detail in Barker et al.²⁴

6 Comparison of Methods

In Fig. 2, we offer a comparison of the prewhitening, precoloring, and robust regression methods described in Secs. 4.1–4.2 as implemented in currently available fNIRS software programs. To generate this comparison, known simulated “brain activity” was added to a set of experimental baseline fNIRS data files containing both physiological noise and motion artifacts as described in Barker et al.²⁴ The simulated activity was added

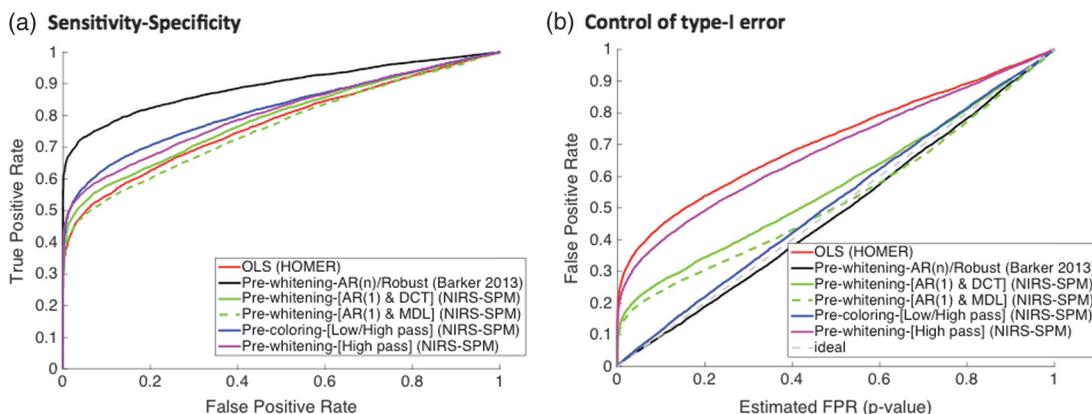


Fig. 2 Comparison of statistical model correction methods. Simulated activity was added to resting-state experimental fNIRS data and analyzed using several models described in the text. (a) A sensitivity-specificity (receiver operator curve) for the comparison of these methods. Models with better performance are closer to the upper-left corner of the plot. (b) The control of type-I errors (false-discovery rates) for the various methods. This plot shows the actual false-discovery rate (y -axis) against the reported p -value obtained by the method (x -axis). Deviations from the line at slope of unity show the tendency toward overestimation of the significance of results and uncontrolled false-discoveries by the method. Similar simulations have been described in Barker et al.²⁴

to half of the optical channels and repeated 1000 times. For each simulation, the same data were analyzed using several models including: (i) ordinary least-squares model (OLS) (e.g., as used in HOMER¹⁰ software), (ii) AR filter and iterative robust least-squares model described in Barker et al.²⁴ [prewhitening-AR(n) and robust], (iii) prewhitening using the default AR(1) covariance component model (ReML) and additional 128-coefficient DCT in NIRS-SPM¹⁴ [prewhitening-AR(1) and DCT], (iv) prewhitening using the AR(1) and MDL wavelet model²⁵ in NIRS-SPM¹⁴ [prewhitening-AR(1) and MDL], (v) noise prewhitening using the DCT high-pass filter implemented in NIRS-SPM¹⁴ [prewhitening-DCT], and finally (vi) noise precoloring using the “hrf” low-pass filter and a DCT high-pass filter implemented in NIRS-SPM¹⁴ [precoloring low/high pass]. The NIRS-SPM code¹⁴ was implemented using release 4.1 along with SPM-12. All models and data used in this example are included in a demonstration script as part of our MATLAB NIRS-toolbox (available at or from the corresponding author). The ordinary least-squares model was implemented using the Gauss-Markov equation [Eq. (2) in this work] which is the default method implored in the HOMER²⁸ and HOMER-2 programs but was coded in MATLAB-2015b outside of the actual HOMER program.

In Fig. 2(a), we show the sensitivity-specificity [receiver operator curves (ROC)] comparing the various methods described in the last paragraph. The true positive rate and false-positive rates are plotted for each method calculated from the simulations. An ideal model would approach the upper left corner of this plot indicating perfect estimation (100% sensitivity and 100% specificity). The overall sensitivity of the model in an ROC analysis depends on the contrast-to-noise of the simulated “brain activity” amplitude (or quality of the data for experimental data) and was simulated at 0.7 μ M amplitude to achieve (corresponding to approximately a medium effect sized Cohen’s-d value of 0.3 to 0.4 for this data set). Thus, for tasks that have larger amplitude changes or data with less noise, the performance of all these models should improve.

The AR-IRLS model,²⁴ which uses both prewhitening and robust regression had the best performance in these simulations. Second, the precoloring method using both a low and high-pass filter implemented in NIRS-SPM¹⁴ had the next best performance. This was followed by the prewhitening method using only the DCT high-pass filter. The ReML-based AR(1) models implemented in NIRS-SPM¹⁴ had slightly lower performance and similar to the uncorrected (OLS) model. Prewhitening with the AR(1) and MDL model had slightly worse performance than the uncorrected OLS model. As previously stated, the ReML-based implementation of the AR prewhitening models in NIRS-SPM computes one set of noise and AR hyperparameters across all channels (separately for oxy- and deoxy-hemoglobin), which can be affected by nonuniform spatial noise across the channels and can cause a few bad channels to contaminate others. This likely contributed to the lower sensitivity of these models in the simulations.

In Fig. 2(b), we show the control of the type-I errors in these models. As previously described in this paper, physiological noise (serial-correlations) and motion artifacts can introduce a high level of uncontrolled false-discovery. Our previous work published in Barker et al.²⁴ details these effects in more detail and rigor and the reader is directed here for more details. As a demonstration of uncontrolled false-discovery, in Fig. 2(b), the

actual false-discovery rate estimated from the ROC analysis is plotted against the corresponding p -value estimated from the method (e.g., this plots the actual FDR versus what MATLAB reported). An ideal model would lie along the line of slope unity (dotted gray) where the estimated p -value associated with the regression agrees with the false-discovery rate as determined by a ROC analysis. As shown in Fig. 2(b), the ordinary least-squares model (e.g., with neither corrections for serial correlations nor noise heteroscedasticity) had the worst uncontrolled false-discovery rates. For example, in this data set, at a reported estimated p -value of $p < 0.05$ (x -axis), the actual false-discovery (y -axis) was close to 30%. This means that when data is displayed at a threshold of $p < 0.05$ reported by the respective code, 30% of the channels with no added “brain activity” were falsely reported as significant. These errors were similar in the prewhitened NIRS-SPM model using only the DCT high-pass filter. We found that the use of the AR(1) model in addition to the DCT or MDL filters, produced a better control of type-I errors, however, this was still considerably above the ideal response (about 15% FDR at $p < 0.05$). The NIRS-SPM software also applies the Welch-Satterwaite estimate of the degrees-of-freedom for these models,¹⁴ which was used to report the p -values in this figure. However as shown, this correction is not entirely sufficient to correct the model. The precoloring method using both a low and high-pass filter in NIRS-SPM and the AR-IRLS model using both AR prewhitening and robust regression had the best control of the type-I errors and nearly ideal response profiles. As a further note, this uncontrolled false-discovery rate issue is different from corrections for multiple comparisons. Methods like Bonferroni or Benjamini-Hochberg²⁹ correct for false-discovery that results from performing multiple tests, but assume that the p -value is reliable for any single test. The type-I errors that arise from serially correlated noise result in unreliable and incorrect estimates of the reported p -value for a single test.

7 The Design Matrix for Functional Near-Infrared Spectroscopy Studies

The design matrix (X) encodes the timing of the task events and any additional trends in the time series of the data. It is a matrix with dimensions of the number of measurement points by the number of regressors in the model. Two common variations of this matrix are generally used in fNIRS or fMRI analysis—the deconvolution [or finite impulse response (FIR)] model and the canonical model. In the deconvolution model, no assumptions are made, and the unbiased shape of the hemodynamic response is estimated from the regression model. In the canonical model, which is widely used in fMRI analysis, a predefined shape models the response.^{30,31} Equations (1)–(7) represent both types of models and the methods for solving the equations are the same for both approaches.

Historically, fNIRS analysis has largely followed the FIR/deconvolution model, whereas fMRI analysis has tended to favor the canonical models. From a pragmatic point of view, it is not practical to view 10,000+ response curves from each voxel of an fMRI scan, so directly testing the hypothesis by compressing the data using a canonical model and displaying a statistical parametric map became the early standard offered in software like SPM.³¹ In contrast, the structured physiological noise and motion artifacts as just described can greatly affect fNIRS results if this noise is not properly accounted for and

assumptions are violated in the statistical model. Thus, the ability to visually examine the fNIRS response subjected fNIRS results to a human quality assurance test. Using a deconvolution allowed researchers to visualize the shape of the curve from their data. Since fNIRS is often used in unique populations such as infants and children or in unique tasks, there has been debate on if the standard hypothesis (e.g., canonical response) developed for fMRI is appropriate for fNIRS. For these reasons, most of the early fNIRS work used deconvolution models. Although in recent years there have been many more studies using the canonical model in fNIRS, these are still looked upon with skepticism, particularly in certain areas of fNIRS research.

7.1 The Deconvolution Finite-Impulse Response Model

In a deconvolution model, the design matrix encodes the timing of the stimulus events but makes no prior assumptions about the shape of the response. The design matrix is constructed from shifted columns of a binary vector (e.g., 0's and 1's) constructed from the onset of the stimulus. Each column of X contains a shifted version of the stimulus onsets. For example, the first (unshifted) column models the response at time = 0 from when the stimulus onset occurred. The second column (shifted by +1) models the response one time point after the onset of the stimulus. Columns shifted backward (e.g., shift = -1) model the response before the stimulus onset (e.g., prebaseline periods), and so on. Thus, in this design matrix, there is one regressor for each time point over the window in which one is trying to estimate the response. As an example, if the data was sampled twice a second (2 Hz), the design model for a 30-s-long task with a 14-s-long poststimulus (recovery) period would have 88 ($30 \text{ s} \times 2 \text{ Hz} + 14 \text{ s} \times 2 \text{ Hz}$) columns. When the model was estimated, there would be 88 corresponding coefficients β that model the amplitude of the response over this entire time period from stimulus onset until 14 s after the end of the task. If these coefficients were plotted as a time course, one would hope to see a response rising with the stimulus, staying elevated for around 30 s, then falling back to baseline around 6 to 10 s after the task ended. In this model, the entire time course for the response is estimated, and the only assumption that is made about the response is that it has recovered before the end of the estimation window. One limitation of the full deconvolution model, however, is that all events of the same type must have the same duration since the same estimation time window is used for all events of the same type.

In order to define the statistical contrast, a contrast vector must then be specified to compute the average contrast over some time window via Eq. (4). Continuing the same example, if the estimated response were 44 s long and estimated at 2 Hz, then the T -statistic for the window from 10 to 30 s could be found by setting the contrast vector c to 1's for the 20th to 60th entries and 0's for everything else, as defined in Eq. (3).

A variation on the full deconvolution model is the impulse response model. In the impulse response model, the design matrix is constructed with knowledge of the duration of events, and the model of the response is the convolution of the task duration with an impulse response. An impulse response is the theoretical response to a short task only one time point in duration. To construct this design matrix, each column of the matrix is a binary vector of 1's during the duration of the stimulus and 0's for baseline periods. In this case, the columns are shifted to model the impulse response. In the same example

discussed above, each single column of the design matrix would have 30-s blocks ($30 \text{ s} \times 2 \text{ Hz} = 60$ points) for each stimulus event. This would be shifted 28 times to model the 14-s window of the impulse response ($14 \text{ s} \times 2 \text{ Hz} = 28$). This model has fewer coefficients than the full deconvolution model, but makes the assumption that the response is sustained at the same level for the example 30-s duration of the task. The advantage of this model, however, is that it can be used to describe experiments where stimulus blocks of the same type have different durations (e.g., in the case of a self-paced study) since the duration of each block is encoded in the design matrix and the impulse response function does not vary with the stimulus duration.

7.2 The Canonical Model

The canonical model is widely used in fMRI studies. Instead of modeling the shape of the response by allowing each time point in the response to be described by a separate coefficient β , the canonical model uses a prior hypothesis on the shape of the response. This idealized response shape is then used to construct the design matrix such that only one coefficient (or a few) is needed to model the amplitude of the response. In this model, the shape of the curve is fixed. This shape represents a prior on the expected response. A number of different variations of this model have been proposed.³⁰

8 Relationship of Finite-Impulse Response and Canonical Models

Combining Eqs. (3), (4), and (7), the statistical contrast for the prewhitened and weighted regression model is given by the expression ($X_{WS} = S \cdot W \cdot X$)

$$T = c \cdot \frac{(X_{WS}^T \cdot X_{WS})^{-1} \cdot X_{WS}^T}{\sqrt{\sigma^2 \cdot c \cdot (X_{WS}^T \cdot X_{WS})^{-1} c^T}} \cdot Y. \quad (8)$$

This expression applies to both the FIR and canonical models (which we will now denote as X_{FIR} and X_{CAN} for the weighted, preweighted design matrices). For the example of a single canonical regressor, by comparing the canonical and FIR expressions and setting the equivalent Eq. (11)s equal to each other, we find that

$$C_{\text{FIR}} = (X_{\text{CAN}}^T \cdot X_{\text{CAN}})^{-1} \cdot X_{\text{CAN}}^T \cdot X_{\text{FIR}}. \quad (9)$$

This states that to the extent that the estimate of the noise σ^2 is the same for both models, then the canonical and FIR models will produce the same statistical T -contrast when a contrast vector is given by Eq. (9). When the contrast vector of the FIR model is given by Eq. (9), we can see that

$$T_{\text{FIR}}/T_{\text{CAN}} = \sigma_{\text{CAN}}/\sigma_{\text{FIR}}, \quad (10)$$

where σ^2 is the mean squared error of the two models. Since the FIR model has more unknowns, σ^2 for the FIR model will generally be equal to or less than that for the canonical model. If we examine Eq. (9), we can see that this expression approaches the approximate limit

$$C_{\text{FIR}} \approx \Gamma_{\text{CAN}} \quad (11)$$

under the conditions when $X_{\text{CAN}}^T \cdot X_{\text{CAN}} = I$ (e.g., optimally designed stimulus timing; see Ref. 32) where Γ_{CAN} is assumed

to be the canonical response shape. Thus, by setting the contrast vector c of the FIR model equal to the canonical response, the two models approach each other in Eq. (10). More formally, Eqs. (9) and (10) show that this happens when the contrast vector is equal to the projection of the (weighted and prewhitened) design matrix for the FIR model onto the space of the (weighted and prewhitened) canonical model. Deviations from the approximate limit in Eq. (11) occur because of the weighting and prewhitening and any efficiencies (e.g., overlapping responses) in the design matrix, which result in differences between defining the contrast in the original measurement space (canonical model) and defining it in the estimation space (FIR model). These deviations are explained in Eqs. (9) and (10).

We can also see, following a geometric interpretation of the regression model, that Eq. (8) is maximized when the contrast vector is a unit normal vector pointing in the direction of β . In other words, the T -statistic is maximized when the shape of the contrast vector matches the shape of the estimated response β . Thus, using a tapered contrast window that matches the shape of the response will produce a higher T -statistic compared to computing a mean over a time window (e.g., c is a binary vector of 1's and 0's). Thus, the canonical response is a prior expectation of what this response looks like and will maximize the T -statistic when the actual response resembles this assumption. Following the opposite prospect, we can also see that specifying a binary contrast vector in the FIR model (e.g., c is uniform 1's over some time window and else 0's) is equivalent to using a shifted canonical boxcar model. It also follows from this interpretation that misspecification of the canonical model (e.g., using the wrong shape) is akin to using the wrong contrast window in the FIR model. In the context of fNIRS, the type-II errors for the estimate of the two chromophores (oxy- and deoxy-hemoglobin) may differ for either the FIR or canonical models because the contrast window does not necessarily (equally) match both responses. In general, the responses of oxy- and deoxy-hemoglobin are shifted relative to each other by up to a few seconds.³³

Finally, Eq. (11) provides a potential compromise between those researchers who favor the canonical and deconvolution methods because it shows how to convert from the FIR to the canonical model. Thus, using this expression, one can perform a full deconvolution, visualize the experimentally measured hemodynamic response, but then report a statistical contrast (T -statistic) based on the canonical model that allows it to be directly compared to other research studies.

9 Conclusion

This paper offered a review of the noise structures that are often observed in fNIRS data and the impact that this noise has on common statistical tests. More specifically, we discussed the assumptions that the statistical regression models often make and how these assumptions can be violated by the properties of fNIRS such as serially correlated noise due to physiology or outliers to the normal noise distribution due to motion artifacts. To restate the problem, all statistical linear models (whether block averaging, deconvolution, canonical linear regression models, and so on) make certain mathematical assumptions about noise properties, and the noise in fNIRS data often violates those assumptions.

There are two ways to deal with noise that violates the assumptions of the statistical model. One approach is to

somehow remove the noise/artifacts from the data. To date, this is how the vast majority of NIRS analysis has been done (reviewed in Huppert et al.²⁸), including many of the methods implemented in the fNIRS HOMER software program. For example, in the case of the motion artifacts shown in Fig. 1, one could identify these periods of time and remove this chunk of data from analysis or try to correct it using a principal or independent component analysis or wavelet filters (see Cooper et al. for review³⁴). After correction, one has obtained a “cleaned” time course, which is used in the statistical analysis. If the filter/noise removal worked perfectly, the noise will longer violate the statistical assumptions of the model and the results will be reliable. However, a common problem with noise reduction methods is that there is no perfect method to remove all types of artifacts; therefore, it becomes subjective and often relies on the user's expertise to select and use these signal processing tools properly. While in general, removing or correcting noise from the data could increase the effect size of the brain activity, increasing detection power and potentially lowering the required population sample size of the study, if the noise is incompletely removed, the assumptions of the model can still be violated, which can lead to inaccurate reporting of the statistical model. In some cases, such as low-pass filtering of the data, this can introduce more violations to the statistical model (in this case serial correlations).

A second approach to dealing with noise that violates the assumptions of the statistical model is to leave the data alone and instead change the assumptions of the statistical model so that it is more generalized and can handle the properties of the artifact without violating any assumptions. Specifically, the “G” in GLM refers to generalizing the statistical model to correctly deal with nonnormally/nonspherically distributed noise. This paper specifically dealt with the methods and options that use this second approach to dealing with noise. The prewhitening/precoloring and robust regression methods that were described here are not noise correction methods. Instead, these approaches attempt to correct the statistical model to make it more generalized. The benefit of this second approach is that as long as you make the model general enough, you can use the same model on any type of artifact. That is, not all fNIRS data has motion artifacts and not all motion artifacts necessarily violate the assumptions of the statistical model. Even if there was no artifact or the type of motion was such that no assumptions were violated in the original model, using the generalized model will produce a more statistically reliable result compared to a model in which the assumptions are violated.

The two approaches to dealing with noise are not exclusive of each other, but our claim is that using the right statistical model is to controlling false discovery in the results (e.g., making sure your conclusions are valid). The control for type-I errors shown in Fig. 2(b) demonstrates this point. There is an important distinction to be made between noise that exists and therefore makes it harder to detect brain activity by lowering the effect size of the activity and noise that violates the statistical assumptions of the statistical model, which results in unreliable statistical tests. While noise correction methods, when used correctly, can remove noise and increase effect sizes, statistical model correction methods like the ones described in this paper help to obtain more reliable results.

It is the opinion of the authors of this paper that the following represent the minimum current best practices for fNIRS analysis:

- Noise prewhitening should be used to remove the effects of structured noise and serially correlated errors in the fNIRS measurements. The presence of physiological noise in fNIRS data causes violations of the assumption of independent noise in the statistical model and can result in high false-discovery rates if uncorrected.
- Prefiltering (e.g., a bandpass filter applied to the fNIRS data) should be avoided as a separate step. Instead, filtering should be applied within the regression model and applied to both sides of the expression [as in Eq. (5)] to avoid bias in the estimated response.
- Noise precoloring is not generally recommended for fNIRS data because of the frequent presence of heavy-tailed noise both in time (e.g., due to motion artifacts) and across channels (e.g., bad coupling of sensors to the head). Precoloring can result in exaggeration of this heavy-tailed noise by spreading it across channels or time.
- Robust regression methods or similar outlier rejection methods are recommended in the context of the linear regression model to deal with heteroscedastic noise.
- A tapered contrast window [c in Eq. (4)] will maximize the T-statistic when the shape of c matches the hemodynamic response. The canonical hemodynamic response is an expectation of this shape.
- If noise reduction or filtering methods are used in addition to the GLM methods of prewhitening and robust regression, care should be exercised to ensure that these methods do not introduce new violations of the statistical models that might require additional generalizations.

References

1. F. F. Jobsis, "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Science* **198**(4323), 1264–1267 (1977).
2. J. S. Wyatt et al., "Quantification of cerebral oxygenation and haemodynamics in sick newborn infants by near infrared spectrophotometry," *Lancet* **328**(8515), 1063–1066 (1986).
3. S. Lloyd-Fox, A. Blasi, and C. E. Elwell, "Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy," *Neurosci. Biobehav. Rev.* **34**(3), 269–284 (2010).
4. H. Obrig, "NIRS in clinical neurology—a 'promising' tool?" *NeuroImage* **85**(Pt 1), 535–546 (2014).
5. I. Miyai et al., "Cortical mapping of gait in humans: a near-infrared spectroscopic topography study," *NeuroImage* **14**(5), 1186–1192 (2001).
6. H. Harada et al., "A comparison of cerebral activity in the prefrontal region between young adults and the elderly while driving," *J. Physiol. Anthropol.* **26**(3), 409–414 (2007).
7. X. Cui, D. M. Bryant, and A. L. Reiss, "NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation," *NeuroImage* **59**(3), 2430–2437 (2012).
8. M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," *NeuroImage* **63**(2), 921–935 (2012).
9. D. A. Boas, A. M. Dale, and M. A. Franceschini, "Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy," *NeuroImage* **23**(Suppl. 1), S275–S288 (2004).
10. A. F. Camestra et al., "Refractory periods observed by intrinsic signal and fluorescent dye imaging," *J. Neurophysiol.* **80**(3), 1522–1532 (1998).
11. J. B. Mandeville et al., "Evidence of a cerebrovascular postarteriole windkessel with delayed compliance," *J. Cerebral Blood Flow Metab.* **19**(6), 679–689 (1999).
12. K. J. Friston, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, 1st ed., Elsevier/Academic Press, Amsterdam, Boston (2007).
13. K. J. Friston et al., "Statistical parametric maps in functional imaging: a general linear approach," *Hum. Brain Mapping* **2**(4), 189–210 (1994).
14. J. C. Ye et al., "NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy," *NeuroImage* **44**(2), 428–447 (2009).
15. A. F. Abdelnour and T. Huppert, "Real-time imaging of human brain function by near-infrared spectroscopy using an adaptive general linear model," *NeuroImage* **46**(1), 133–143 (2009).
16. S. G. Diamond et al., "Physiological system identification with the Kalman filter in diffuse optical tomography," *Med. Image Comput. Assist. Interv.* **8**(Pt 2), 649–656 (2005).
17. S. G. Diamond et al., "Dynamic physiological modeling for functional diffuse optical tomography," *NeuroImage* **30**(1), 88–101 (2006).
18. X. S. Hu et al., "Kalman estimator- and general linear model-based online brain activation mapping by near-infrared spectroscopy," *Biomed. Eng. Online* **9**, 82 (2010).
19. M. M. Plichta et al., "Model-based analysis of rapid event-related functional near-infrared spectroscopy (NIRS) data: a parametric validation study," *NeuroImage* **35**(2), 625–634 (2007).
20. M. L. Schroeter et al., "Towards a standard analysis for functional near-infrared imaging," *NeuroImage* **21**(1), 283–290 (2004).
21. K. J. Friston et al., "Analysis of fMRI time-series revisited," *NeuroImage* **2**(1), 45–53 (1995).
22. P. L. Purdon and R. M. Weisskoff, "Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI," *Hum. Brain Mapping* **6**(4), 239–249 (1998).
23. K. J. Friston et al., "Characterizing evoked hemodynamics with fMRI," *NeuroImage* **2**(2), 157–165 (1995).
24. J. W. Barker, A. Aarabi, and T. J. Huppert, "Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS," *Biomed. Opt. Express* **4**(8), 1366–1379 (2013).
25. J. E. Jang et al., "Wavelet minimum description length detrending for near-infrared spectroscopy," *J. Biomed. Opt.* **14**(3), 034004 (2009).
26. K. J. Friston et al., "To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis," *NeuroImage* **12**(2), 196–208 (2000).
27. P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Commun. Stat.* **A6**, 813–827 (1977).
28. T. J. Huppert et al., "HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain," *Appl. Opt.* **48**(10), D280–D298 (2009).
29. Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. R. Statist. Soc. B* **57**(1), 289–300 (1995).
30. M. A. Lindquist et al., "Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling," *NeuroImage* **45**(1 Suppl), S187–S198 (2009).
31. K. J. Friston et al., "Event-related fMRI: characterizing differential responses," *NeuroImage* **7**(1), 30–40 (1998).
32. A. M. Dale, "Optimal experimental design for event-related fMRI," *Hum. Brain Mapping* **8**(2–3), 109–114 (1999).
33. T. J. Huppert et al., "A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans," *NeuroImage* **29**(2), 368–382 (2006).
34. R. J. Cooper et al., "A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy," *Front. Neurosci.* **6**, 147 (2012).

Theodore J. Huppert is an associate professor of radiology and bioengineering at the University of Pittsburgh. He works on the development of signal processing and statistical methods for functional near-infrared spectroscopy and multimodal brain imaging technologies.