Title:

Estimation Efficiency and Statistical Power in Arterial Spin Labeling FMRI.

Authors:

Jeanette A. Mumford [3]

Luis Hernandez-Garcia [1,2]

Gregory R. Lee [1,2]

Thomas E. Nichols [1,3]

[1] University of Michigan Functional MRI laboratory ,

[2] University of Michigan Dept. of Biomedical Engineering ,

[3] University of Michigan Dept. of Biostatistics

Correspondence to

Luis Hernandez-Garcia

FMRI Laboratory

2360 Bonisteel Ave.

Ann Arbor, MI 48109-2108

hernan@umich.edu

tel: (734) 763 - 9254

fax: (734) 936 – 4218

Abstract

Arterial Spin Labeling (ASL) data are typically differenced, sometimes after interpolation, as part of pre-processing before statistical analysis in fMRI. While this process can reduce the number of time points by half, it simplifies the subsequent signal and noise models (i.e., smoothed box-car predictors and white noise). In this paper we argue that ASL data are best viewed in the same data analytic framework as BOLD fMRI data, in that all scans are modeled and colored noise is accommodated. The data are not differenced, but the control/label effect is implicitly built into the model. While the models using differenced data may seem easier to implement, we show that differencing models either produce biased estimates of the standard errors or suffer from a loss in efficiency. The main disadvantage to our approach is that non-white noise must be modeled in order to yield accurate standard errors, however this is a standard problem that has been solved for BOLD data, and the very same software can be used to account for such autocorrelated noise.

Introduction

Arterial Spin Labeling (ASL) techniques have been in development for over a decade since their inception (Williams 1992) but it was not until more recently that arterial spin labeling was shown to be a very powerful technique for functional imaging of low-frequency paradigms (Aguirre 2002, Wang 2003). Further improvements in the technique have made it a practical tool for functional MRI of high-frequency (i.e., event related) paradigms by overcoming issues of temporal resolution, SNR, and the ability to collect multiple slices in a single TR (Wong 2000, Hernandez 2004, Silva 1995). ASL is inherently a low signal to noise ratio (SNR) technique so it is important to maximize the accuracy and sensitivity of the analysis.

ASL techniques are very appealing for functional imaging primarily because they offer a physiologically meaningful and quantitative alternative to BOLD effect imaging, currently the dominant technique used for functional brain mapping. In summary, ASL consists of acquiring image pairs made up of a "labeled" image, in which the inflowing blood has been magnetically labeled, and a control image without labeled blood. Perfusion can be calculated from the difference of those two images, which is made up only of the labeled blood present in the imaged slice. The subtraction of image pairs results in an added benefit, namely that the subtraction of label/control pairs in ASL produces contains noise that is whiter than BOLD noise (Aguirre *et* al., 2002, Wang *et al*., 2003), depending on the specific subtraction scheme used for obtaining the labeled images from the raw ASL images (Liu *et* al., 2005).

In this article, we consider data collected using the Turbo-CASL sequence. Turbo-CASL is a spin labeling technique that takes advantage of the delay period between labeling spins at the neck and the time they reach the imaging plane to collect the control image, resulting in a more efficient use of the time. This technique is obviously quite sensitive to transit times, so one must

collect a transit time measurement and adjust the timing parameters of the sequence accordingly. The benefits of the technique are an increase in the temporal resolution while preserving some of the higher SNR characteristics of continuous ASL and the ability to obtain exaggerated activation responses by proper choice of labeling parameters. The drawbacks are that it requires knowledge of transit times and that excessive variability of those transit times over the imaged tissue can result in SNR loss in some regions (Hernandez-Garcia *et al*. 2004 & 2005, Lee *et al*. 2004).

Despite the simplicity of the subtraction analysis method, it cannot strictly be optimal for estimation and detection of brain activity using the general linear model. For a given linear model of the full (length-N) dataset, the Gauss-Markov Theorem (Graybill, 1976) dictates that optimally precise estimates are obtained from ordinary least squares (OLS) estimates for independent data, or from whitened OLS (i.e., generalized least squares, GLS) for dependent data. Hence differencing can be no more accurate than OLS or GLS on the full data is likely sub-optimal.

The goal of this work is to characterize the statistical properties of different ASL modeling methods. Starting from a general linear model that includes the alternating control-label effect, we examine several differencing schemes including a no-differencing approach. To measure goodness of the differencing schemes we calculate the bias of the variance estimators, the estimation efficiency and the estimator power for different study designs and error covariance structures. We also analyze real data to produce a comparison of Z-scores between the models and explain the results in the context of signal processing.

Theory

We first describe the signal model, then differencing methods and their frequency responses, the noise model used and, finally, the estimation methods.

Signal Model. We pose the signal model for ASL data in terms of a General Linear Model (GLM). While Liu *et al*. (2002) posed a separate GLM for control and label data, we consider a single model for the collected data,

$$Y = X\beta + \varepsilon \quad \textbf{(1)}$$

where Y is a vector of length N that contains the original experimental data ordered as acquired, including labeled and non-labeled images; X is a $N \times p$ design matrix; β is a vector of p parameters; and ε is the error vector of length N, where $\mathrm{Cov}(\varepsilon) = \sigma^2 V$.

The design matrix for our experimental conditions was built to reflect the principal contributions to the observed ASL signal. This signal is made up of two fixed baseline components and two dynamically changing components that are due to hemodynamic changes induced by the stimulation paradigm.

The two fixed components are the MR signal from static tissue, which makes up the bulk of the image, and the inflowing blood signal in the baseline state. The baseline MR signal is just constant in time, while the inflowing blood signal (or baseline blood flow) is sensitive to whether the tag is applied or not (top panel Figure 1). Hence, the baseline blood flow regressor is simply a function of alternating positive and negative values, $+a$ and $-a$, depending on whether the tag is applied or not. While alternating 1's and -1's ($a = 1$) seems natural for this predictor, instead $a = 1/2$ should be used so that the corresponding parameter expresses a unit effect in the data. Note that the presence of the arterial tag corresponds to $-a$, since the tag is made of inverted spins, and hence reduces the total signal magnitude (second panel Figure 1).

The two activation-related regressors are the the perfusion changes and the BOLD effect changes in signal, generated by convolution of the stimulation function with a gamma-variate BOLD response function (third panel Figure 1). The changes in perfusion due to activation are known to have similar temporal properties to those in the BOLD response, but they are sensitive to the presence of the arterial inversion tag. Thus, in order to capture the perfusion changes due to activation, we created the activation perfusion regressor by modulating the BOLD regressor with the baseline blood flow regressors to reflect the presence or absence of the inversion tag (fourth panel Figure 1).

The observed MR signal is thus made up of the weighted sum of these four components, or regressors, which are depicted in Figure 1.

Differencing Methods. Differencing the data can be built into the model by premultiplying both sides of the GLM equation by a generic differencing matrix, D:

$$DY = DX\beta + D\varepsilon. \quad \textbf{(2)}$$

Any differencing scheme can be encompassed in this model by specification of an appropriate D. We denote $D_1 = I$, where I is a $N \times N$ identity matrix, for the case of no differencing at all. The standard pairwise differencing can be implemented with a $N/2 \times N$ differencing matrix

$$D_2 = \begin{pmatrix} 1 & -1 & & & & 0 \\ & & 1 & -1 & & \\ & & & & \ddots & \\ 0 & & & & 1 & -1 \end{pmatrix}.$$

The other differencing approaches we study include running subtraction, with $(N-1) \times N$ differencing matrix

$$D_3 = \begin{pmatrix} 1 & -1 & & & & 0 \\ & -1 & 1 & & & \\ & & 1 & -1 & & \\ & & & \ddots & \ddots & \\ 0 & & & & 1 & -1 \end{pmatrix},$$

surround subtraction with $(N-2)\times N$ differencing matrix,

$$D_4 = \begin{pmatrix} 1 & -2 & 1 & & & \\ & -1 & 2 & -1 & & \\ & & 1 & -2 & 1 & \\ & & & \ddots & \ddots & \ddots \\ & & & & -1 & 2 & -1 \end{pmatrix}$$

and sinc subtraction (D_5). The $N\times N$ differencing matrix for sinc subtraction is best illustrated as an image of the differencing matrix, as in Figure 2.

Noise Model. It is well known that the elements of the error, ε, are not independent, with $\mathrm{Cov}(\varepsilon)=\sigma^2 V$ being non-diagonal. For example, Zarahn *et al*. (1997) found that the power spectra of fMRI noise data follow a "1/f" frequency-domain structure, which is associated with a lower order autoregressive (AR) model. In the evaluations below we will use an AR(1) plus white noise (WN) model; we found this autocorrelation structure to follow that of our data through empirical observations. Such a model has an autocorrelation structure, V, where the correlation for a lag of ℓ is defined by

$$V_{k,k-\ell} = \rho^{|\ell|}\left(\frac{\sigma_{AR}^2}{\sigma_{AR}^2+\sigma_{WN}^2}\right)$$

and the variance of each measurement is given by $\sigma^2=\sigma_{AR}^2+\sigma_{WN}^2$ where σ_{AR}^2 is the variance contributed from the AR(1) process, σ_{WN}^2 is the white noise variance and ρ is the AR(1) correlation parameter.

The effect of differencing matrices on the noise can be thought of as a filter in terms of its frequency response, essentially damping some frequencies of the signal while emphasizing others. Those frequency responses can be derived analytically for a given input y[n] whose discrete Fourier transform is given by $Y(e^{jw})$, yielding the equations in Table 1. Pairwise subtraction and sinc subtraction do not have a straightforward linear response since they both involve down-sampling the data, although sinc-subtraction subsequently upsamples the data. In both of those cases, the down-sampling process causes the top half of the frequency spectrum to alias into the bottom half before the filtering process. The frequency response of the filters can be seen in Figure 3. In terms of their frequency response, they are very similar except for the imperfections of the sinc kernel used in the implementation. In terms of detection and statistics, sinc subtraction preserves greater degrees of freedom than pairwise subtraction, which reduces the number of time points by half.

Liu *et al*. (2005) took a similar signal-processing approach in order to examine the effects of differencing on the BOLD and perfusion responses observed in ASL functional time series data, including BOLD effects (an ASL sequence in which the acquisition is carried out with a gradient echo with a long echo time would also be BOLD weighted). By modeling the acquisition scheme as a linear time invariant (LTI) system, they found that the control-label modulation shifts the activation to higher frequencies in the BOLD weighted data and the choice of differencing method is contingent on the spectral content of the time series (determined by the experimental design).

Estimation. Based on both the differenced and undifferenced data, OLS or GLS can be used to estimate and make inference on β. While OLS estimates of β are unbiased even when data

are temporally autocorrelated, the estimates do not have minimum variance, meaning they are not fully efficient; further, the estimated standard errors are biased, which can result in test statistics that are either too large or too small. The optimal approach is GLS, corresponding to OLS on the whitened data and model, and is implemented in most fMRI packages (e.g. FSL & SPM). GLS requires that the structure of the noise in the data be known or, at least, estimated with high precision. The noise covariance, $\sigma^2 V$, is estimated by a variety of means (Worsley *et al.* 2002, Friston *et al.* 2002, Woolrich *et al*. 2001); most methods use a regularized fit of a low-dimensional autocorrelation model to the OLS residuals.

Given that the covariance of the error, ε, is $\sigma^2 V$, the covariance of the error of the differenced data, $D\varepsilon$, is given by $\sigma^2 DVD^T = \sigma_D^2 V_D$. Let W be a whitening matrix such that $W(V_D)W^T = I$, the whitened version of the general linear model is then

$$WDY = WDX\beta + WD\varepsilon, \quad \textbf{(3)}$$

and the GLS estimate of β is given by

$$\hat{\beta} = (WDX)^{-} WDY, \quad \textbf{(4)}$$

where the symbol $^{-}$ denotes a pseudo-inverse operator (Graybill, 1976, p. 28). The variance of the estimate is given by

$$\mathrm{Var}\left(\hat{\beta}\right) = (WDX)^{-}\left[WV_D W^T\right]\left((WDX)^{-}\right)^T \sigma_D^2. \quad \textbf{(5)}$$

If the whitening is accurate, the middle bracketed term will be identity; however, if OLS is used then $W = I$ and this term will not vanish.

Two problems may arise with the estimation procedure. First, if $D \neq I$ the intercept predictor in X will become an all-zero predictor in DX. Since the differencing will de-mean the data, the baseline effect absent and that column can be omitted from DX. More generally, effects

nullified by D can be removed by an appropriate transformation, resulting in a reducing number of columns. Second, a poorly-chosen D may induce linear dependencies into DY, resulting in a singular covariance matrix $V_D = DVD^T$. This problem can be resolved by removing rows of D until V_D is positive definite.

The residual variance is estimated with the residual mean square of the whitened differenced data,

$$\hat{\sigma}_D^2 = \frac{1}{N-p}(WDY - WDX\hat{\beta})^T(WDY - WDX\hat{\beta}). \quad \textbf{(6)}$$

Then the estimated variance, $\widehat{\mathrm{Var}}\left(\hat{\beta}\right)$, is found by substituting $\hat{\sigma}_D^2$ into equation (5). Note that under OLS it is assumed that $V_D = W = I$, while with GLS the assumption is $W(V_D)W^T = I$.

The T-test for the null hypothesis $H_0 : c\beta = 0$ is $T = c\hat{\beta} \Big/ \sqrt{\widehat{\mathrm{Var}}(c\hat{\beta})}$, where c is a contrast used to express the effect of interest.

Methods: Model Evaluation

Model Details. For a given times series with length N = 258 (TR=1.4 sec) where the temporal autocorrelation followed an AR(1)+WN structure, a signal model for a TurboCASL perfusion experiment was created for three different experimental designs: A fixed ISI event related design (ISI = 18, SOA = 20, stimulus duration = 2 seconds), a randomized event-related design (uniform distribution, 5<ISI<12sec.), and a blocked design (30 scans ON , 30 scans OFF). The randomized event related design was repeated 100 times to verify consistency across different realized designs. Appropriate general linear models were created consisting of baseline image intensity, baseline blood flow, BOLD response and increases in perfusion due to activation. The

perfusion responses were modeled as a difference of two gamma variate functions[1]. Figure 1 graphically illustrates the first 60 seconds of the first subject's design matrix's regressors for a block design.

Efficiency and Bias. The efficiency of the estimated contrast, $c\hat{\beta}$, from the different models is given by the reciprocal of the true variance, $1/\mathrm{Var}\left(c\hat{\beta}\right)$. If one method is less efficient than another method, it is not as sensitive for the detection of effect $c\beta$.

The biases of $\hat{\sigma}_D^2$ and $\widehat{\mathrm{Var}}\left(c\hat{\beta}\right)$ for the differencing models under the assumptions of OLS were also calculated. The derivations of these quantities are given in Appendix A. If these estimates are biased, it indicates that the differenced data are correlated, hence violating the OLS assumption of independent measurements and resulting in test statistics that can be too large or too small.

We computed the efficiency of the contrast estimated and bias of the estimated variance over a range of AR(1)+WN models. Specifically, we studied a range of AR parameter values (ρ) between 0 and 0.9. Also, since all values of σ_{AR}^2 and σ_{WN}^2 that share the same $\sigma_{AR}^2/\sigma_{WN}^2$ yield the same relative efficiency and bias values, we varied the variance of the AR(1)+WN by varying the ratio , $\sigma_{AR}^2/\sigma_{WN}^2$, between 0 and 25.

Statistical power . Power is the probability of detecting a given effect of magnitude $c\beta$ with a given significance level. Power estimates are not meaningful when $\widehat{\mathrm{Var}}\left(c\hat{\beta}\right)$ is biased; for

[1] spm_hrf.m, SPM2, http://www.fil.ion.ucl.ac.uk/spm.

example negative bias leads to an overestimation of power since the test statistics are artificially large. Hence we follow what is standard practice in the statistics literature, and only considered statistical power for methods where $\widehat{\mathrm{Var}}\left(c\hat{\beta}\right)$ was found to be unbiased; this included the no differencing model estimated with GLS and the pairwise subtraction model estimated with OLS. We calculated power over a range of the signal to noise ratio (SNR=change in perfusion/σ). SNR was varied by varying the change in perfusion between 0.1 and 2 and fixing the variance at 1.45. The details of the power calculation are described in Appendix A.

Methods: Imaging data

Data Collection. All imaging was carried out using a 3.0 T Signa LX scanner (General Electric, Milwaukee, WI, USA) fitted with an additional, home-built, spin labeling system. Double-coil turboCASL time series data collected from six subjects during a finger tapping event related experiment (4 slices, FOV = 24cm, ISI=18 sec, 360 sec. duration, TR=1.4 to 1.6 sec. depending on resting transit time, GE spiral, TE = 12ms). Prior to acquisition of time series data, the sequence was optimized for each individual subject by collecting a set of Turbo-CASL images with varying TR (800, 1200, 1400, 1600, 1800, 2000, 2200 and 4000 ms). Labeling time was always 200 ms less than TR. The parameters that produced the highest SNR were chosen as the optimum TurboCASL regime as in (Hernandez-Garcia, 2004). K-space data were filtered to remove spurious RF noise that may be introduced by the labeling coil, and reconstructed using field map homogeneity correction.

Data Analysis. While there are 10 possible methods (OLS and GLS for 5 differencing methods), we only considered two to be practical with real data: GLS with no differencing (D_1) and OLS with simple subtraction (D_2). OLS is inappropriate with any method other than simple

subtraction because the errors are not independent (non-white noise). While the error autocorrelation of other differencing methods (DVD^T) could feasibly be estimated, existing fMRI software is designed to estimate V in un-differenced time series, and hence we only used GLS with the original data.

For each subject, at each voxel, a GLS model with no differencing was fit with the FEAT software tool, which is part of FMRIB's software library FSL[2]. FSL estimates an autocorrelation function (ACF) at each voxel and, after tapering and non-stationary spatial smoothing, constructs V for data and model whitening. The simple subtraction data was also fit with FEAT, but without whitening. T statistics were converted to Z statistics with a probability integral transform. The *Z* statistics from each analysis were then compared. Both methods should yield valid inferences, that is, null-hypothesis voxels should have comparable Z statistics; but when a signal is present, larger Z values are evidence of greater sensitivity.

Frequency response. Active voxels were identified by correlation analysis ***Luis: How? What threshold?***, and time courses (length = 258 points) were extracted from those active voxels (N=87), and from 87 non-active voxels in the frontal lobe. Frequency spectra were computed from the un-differenced and differenced data and compared to the frequency spectrum obtained from applying the predicted frequency response function to the raw data.

Results:

Power and Efficiency Calculations.

Three experimental design types were considered, as described in the Methods section:

[2] http://www.fmrib.ox.ac.uk/fsl

blocked, randomized event related and fixed event related. For the reasons described above, OLS was used to estimate all differencing models and GLS was only used in the no differencing case. Figure 4 illustrates the % biases in both $\hat{\sigma}_D^2$ and $\widehat{\mathrm{Var}}\left(c\hat{\beta}\right)$ when OLS is used to estimate the differencing models. When bias is present, it is an indication that there is autocorrelation or variability in the data that the model is ignoring. The top panel shows the bias in $\hat{\sigma}_D^2$, where the no subtraction model results in strong negative bias while the running, surround and sinc subtraction models have a slightly negative bias and the pairwise differencing model has almost no bias. The bottom panel illustrates that $\widehat{\mathrm{Var}}\left(c\hat{\beta}\right)$ is biased in all cases except for the simple subtraction model. To understand the effect of bias on a p-value, consider an example where there is -50% bias in $\widehat{\mathrm{Var}}\left(c\hat{\beta}\right)$. If the biased variance is used in the test statistic, a biased p-value of 0.01 will be found when the true p-value is actually 0.05; similarly a 0.0001 biased p-value would be found when 0.004 is actually correct. So the biased variance inflates significance and can lead to an incorrect conclusion of significant activation.

The study design used in Figure 4 was block design, but results were similar for the event related designs also. Different values of ρ were also considered and as ρ decreased, the bias in the undifferenced model approached 0, but the percent bias in $\widehat{\mathrm{Var}}\left(c\hat{\beta}\right)$ was similar to that shown in Figure 4 for the other differencing methods (results not shown).

The linear models considered were able to estimate the amplitude of the responses with varying statistical efficiency. Figure 5 shows the relative efficiency of the model estimation under the block design study, where the differencing models estimated with OLS are compared to no differencing with GLS. The pairwise differencing method shows the greatest loss in efficiency,

up to 24% or more, while the other differencing methods are more efficient. The results were similar for the event related designs, with a slightly larger loss in efficiency for the pairwise subtraction method but the relative efficiency for the other methods remained near 1 (results not shown).

For the power study we only consider models where $\widehat{\mathrm{Var}}\left(c\hat{\beta}\right)$ has little or no bias: simple subtraction using OLS and no differencing with GLS. Figure 6 shows the power for no differencing with GLS and and that of pairwise differencing with OLS for the 3 study designs over SNR values ranging between 0.1 and 1.4. The AR(1)+WN model used had a correlation parameter $\rho = 0.90$, AR variance, $\sigma^2_{AR} = 0.11$ and white noise variance, $\sigma^2_{WN} = 2$. The dotted lines on the random event related figure indicate ± 2 standard deviations of the average power over the 100 realizations. As expected, the power is similar between the two methods for the block design, but the no differencing GLS model is shown to have larger power for the event related study designs. Another view of this result is shown in Figure 7 which shows the ratio of OLS power to GLS power. The random event related design can have up to 35% (s.d. 1.5%) lower power when OLS is used compared to GLS. Note that the U-shape to the curves in Figure 7 is to be expected: When SNR is very small, power is 0 for any method, and hence there is no percent difference; likewise, when SNR is very large, power is 1 for any method and there is again no difference.

Experimental Data.

Undifferenced time courses from the experimental ASL time series were extracted from 87 active and 87 resting voxels (identified through simple correlation analysis). The effects of the differencing schemes on the frequency spectra of data can be seen in figure 8. The frequency spectra of time courses extracted from resting and active voxels are shown in the top and

bottom panels, respectively. The blue line shows the spectral content of the raw (undifferenced) data, the other lines show the effects of simple, running, surround, and sinc subtraction schemes. The appearance of the raw data spectra indicate that our data contain primarily white noise but also a strong AR(1) component ($\rho = 0.90$, $\sigma^2_{AR} = 0.11$, $\sigma^2_{WN} = 2$) .

The activation effects, which are typically low frequency by nature, were modulated by the Nyquist frequency because of the alternating acquisition pattern of the control and labeled images. Hence, the activation effects appeared in the higher frequency range in the spectrum. The theory and data agreed that the un-differenced, running and surround subtraction methods attenuated the frequency content at the higher frequencies, while the simple and sinc subtraction methods aliased the high-frequency content into the low-frequency range of the spectrum because of the sub-sampling step. The sinc interpolation kernel produced the smooth roll-off seen at the half-Nyquist frequency. Obviously, the un-differenced data retained all its frequency content. Crucially, only the pairwise subtraction yielded a flat spectrum (for its halved sampling rate), and all other methods had spectra that were more colored than the original data, consistent with our bias calculations above.

Statistical Mapping. Figure 9 shows a comparison of Z statistics obtained using GLS with no differencing and OLS with pairwise subtraction in all subjects. These figures show that the Z values for the full model tend to be larger when the statistics are positive and smaller when the statistics are negative, when compared to the subtracted model, which would result in more significant voxels found without differencing and using GLS. Figure 10 displays the boxplots of the difference in *Z* values from the two methods when the *Z* value from either of the methods was larger than 2. Boxplots display the distribution of data by showing the median (horizontal black line in box), the first and third quartiles of the data (edges of the blue box), range of points

included within 1.5*interquartile range (whiskers) and outliers (points). In all cases, the median is larger than 0, indicating that typically the *Z* value from the undifferenced model using GLS is larger than the differenced model estimated with OLS.

Discussion:

Our analysis indicates that the most powerful analysis of ASL data is obtained by using the full data, with no differencing and building the control/label effect into the model implicitly. Using direct calculations and real data we've found increased sensitivity with this approach relative to the standard approach of pairwise differencing. Pairwise differencing reduces the dimensionality of the data in half, yet should yield white noise. Our calculations found that differencing methods that did not severely reduce dimensionality had good efficiency even with OLS (Fig. 5), but such methods colored the noise (Fig. 8) leading to biased variance estimates when using OLS (Fig. 4). So while surround and sinc subtraction methods may appear attractive efficiency-wise, they induce yet more spectral structure which would have to be accounted for in subsequent modeling (in particular, to obtain accurate standard errors). Hence, we are drawn to the modeling of the full data with standard BOLD fMRI methods, with an ASL-tailored design matrix X and BOLD noise models for V. While noise modeling is a challenge, it is now a standard feature in BOLD fMRI modeling software.

As there are substantial differences between pairwise differencing and the other methods in terms of efficiency, it is tempting to infer that this is due to pairwise differencing yielding half as many observations. In fact, a simple example shows this can't be. Consider the case of pairwise differencing with independent noise and a rest-only experiment; that is, the design matrix consists just of an intercept and the control/tag effect (-½, ½, -½, …). Even though the

residual variance in the differenced model is doubled ($\sigma_D^2 = Var(\varepsilon_i - \varepsilon_{i+1}) = 2\sigma^2$), this is exactly countered by DX having twice the relative efficiency as X. That is, under iid errors, pairwise differencing is fully efficient for the baseline perfusion effect, despite having half the observations of the full model. This suggests that the colored noise plays a role; in particular, the other differencing methods are providing some sort of approximate whitening which then improves their efficiency. Any such whitening is definitely approximate, however, as marked by the bias in the variance estimates. Specifically, in Figure 5, these differencing methods have greater efficiency than no-differencing as the autocorrelation increases.

When the ultimate goal is group modeling, that is, making inferences for a whole population and not just a single subject, the same conclusion prevails. Optimal inferences are obtained by using the best intrasubject estimates, no differencing with GLS, and taking the intrasubject variances $Var(c\hat{\beta})$ to the second level (Beckman *et al*., Woolrich *et al*.). An alternative, less optimal approach is only to take the contrast estimates to the second level (Holmes & Friston, Friston, *et al.*, 1999); with this method the combined between and within subject variance estimate is made implicitly and intrasubject standard errors are not needed. In this instance, only the precision of the intrasubject estimates matters and hence any of the high-efficiency methods (no differencing and full-dimensionality differencing methods; see Figure 5) could be used. However, the simplicity of no-differencing and its unity with BOLD modeling methods makes it our method of choice, even if the simple group method is used.

Although the data presented here were collected using Turbo-CASL, there are large number of ASL techniques that can be used to perform perfusion based fMRI. In terms of statistical properties, using Turbo-CASL presents primarily the same issues as other ASL techniques: the signal is modulated by a control-label acquisition pattern and it contains temporally

autocorrelated noise like all MRI data. Because of the faster acquisition, though, aliasing of respiratory and cardiac effects will occur in different locations in frequency than in standard continuous ASL techniques. The transit time sensitivity of Turbo-CASL likely introduces regionally specific spatial correlations, depending on the vascular network feeding the region. Otherwise, the temporal characteristics of the signal are the same as in other ASL techniques. Hence, the issues discussed in this article can be generalized to all ASL techniques.

Summary:

In our comparison of techniques for modeling ASL data, we found that when using OLS to estimate a model, pairwise subtraction is less efficient than running, surround, sinc and no subtraction, but pairwise subtraction produces variance estimates with almost no bias. The no differencing model estimated with GLS also yields unbiased variance estimates, but with more efficiency than pairwise subtraction. Therefore the two methods with valid test statistics are the no subtraction estimated with GLS and pairwise subtraction estimated with OLS, and with real data we found larger test statistics using no subtraction estimated with GLS, reflecting the gain in efficiency.

Appendix A

Bias. We evaluated the bias in the estimate of $\mathrm{Var}\left(c\hat{\beta}\right)$ for each of the differencing models. When OLS is used to estimate the model ($W = I$), then the expected value of $\hat{\sigma}_D^2$ is given by

$$E\left(\hat{\sigma}_D^2\right) = \frac{\sigma_D^2}{N\text{-}p}\left(tr\left(V_D\right) - tr\left(\left(DX\right)^T V_D DX\left(\left(DX\right)^T DX\right)^{-1}\right)\right), \quad \textbf{(A1)}$$

And so the % bias of $\hat{\sigma}_D^2$ is given by

$$100 \times \frac{E\left(\hat{\sigma}_D^2\right) - \sigma_D^2}{\sigma_D^2} \quad \textbf{(A2)}$$

(Watson, 1955). Therefore, the % bias of the estimated variance of a contrast, $c\hat{\beta}$, is given by

$$100 \times \frac{E\left(\widehat{\mathrm{Var}}\left(c\hat{\beta}\right)\right) - \mathrm{Var}\left(c\hat{\beta}\right)}{\mathrm{Var}\left(c\hat{\beta}\right)} = 100 \times \left(\frac{E\left(\hat{\sigma}_D^2\right) c\left(\left(DX\right)^T DX\right)^{-1} c^T}{\sigma_D^2 c\left(DX\right)^- V_D \left(\left(DX\right)^-\right)^T c^T} - 1 \right) \quad \textbf{(A3)}$$

Power. Power is the probability of detecting a given effect of magnitude $c\beta$ with a given significance level α

$$P\left(T \geq t_\alpha\right) = 1 - \Phi\left(t_\alpha - c\beta \Big/ \sqrt{Var(c\hat{\beta})}\right), \quad \textbf{(A4)}$$

where t_α is the T-statistic critical value and $\Phi(\cdot)$ is the cumulative density function of a standard Normal distribution (where we have assumed the degrees of freedom, $N - p$, to be large, as would be typical for a ASL study). The AR(1)+WN autocovariance $\sigma^2 V$ was assumed known, and was used to find estimator variance with equation (5).

References:

Aguirre, G. K., Detre, J. A., Zarahn, E., & Alsop, D. C. (2002). Experimental design and the relative sensitivity of BOLD and perfusion fMRI. Neuroimage, 15(3), 488-500.

Beckmann, C.F., Jenkinson M. and Smith, S.M. (2003). General multilevel linear modeling for group analysis in fmri, NeuroImage, 20, 1052-1063.

Friston, K. J., Holmes, A. P., Price, C. J., Buchel, C., & Worsley, K. J. (1999). Multisubject fMRI studies and conjunction analyses. Neuroimage, 10(4), 385-96.

Friston KJ, Penny W, Phillips C, Kiebel S, Hinton G, and Ashburner J. (2002) Classical and Bayesian inference in neuroimaging: theory. Neuroimage, 16, 465-483.

Graybill, F.A. (1976). Theory and Application of the Linear Model. Duxbury press, Belmont, CA.

Hernandez-Garcia, L., Lee, G. R., Vazquez, A. L., & Noll, D. C. (2004). Fast, pseudo-continuous arterial spin labeling for functional imaging using a two-coil system. Magn Reson Med, 51(3), 577-85.

Hernandez-Garcia L., Lee G. R. , Vazquez A. L., Noll D. C. (2005). Quantification of Perfusion FMRI using a numerical model of Arterial Spin Labeling accounting for dynamic transit time effects. Magn Reson Med, in press.

Holmes A. and Friston, K. (1998). Generalisability, random effects and population inference. NeuroImage,7, S754.

Lee, G. R., Hernandez-Garcia, L., & Noll, D. C. (2004). Effects of Activation Induced Transit Time Changes On Functional Turbo ASL Imaging. In P. ISMRM (Ed.), Kyoto, Japan:

Liu, T. T., & Wong, E. C. (2005). A signal processing model for arterial spin labeling functional MRI. Neuroimage, 24(1), 207-15.

Liu, T. T., Wong, E. C., Frank, L. R., & Buxton, R. B. (2002). Analysis and design of perfusion-based event-related fMRI experiments. Neuroimage, 16(1), 269-82.

Silva, A. C., Zhang, W., Williams, D. S., & Koretsky, A. P. (1995). Multi-slice MRI of rat brain perfusion during amphetamine stimulation using arterial spin labeling. Magn Reson Med, 33(2), 209-14.

Wang, J., Aguirre, G. K., Kimberg, D. Y., & Detre, J. A. (2003). Empirical analyses of null-hypothesis perfusion FMRI data at 1.5 and 4 T. Neuroimage, 19(4), 1449-62.

Watson, G.S. (1955). Serial correlation in regression analysis I, Biometrika, 42, 327-342.

Williams, D. S., Detre, J. A., Leigh, J. S., & Koretsky, A. P. (1992). Magnetic resonance imaging of perfusion using spin inversion of arterial water. Proc Natl Acad Sci U S A, 89(1), 212-16.

Wong, E. C., Buxton, R. B., & Frank, L. R. (1998). Quantitative imaging of perfusion using a single subtraction (QUIPSS and QUIPSS II). Magn Reson Med, 39(5), 702-08.

Wong, E. C., Luh, W. M., & Liu, T. T. (2000). Turbo ASL: arterial spin labeling with higher SNR and temporal resolution. Magn Reson Med, 44(4), 511-15.

Woolrich MW, Ripley BD, Brady M and Smith, S. Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage, 14, 1370-1386 (2001).

Worsley KJ, Liao CH, Aston J, Petre V, Duncan GH, Morales F, and Evans AC. A general statistical analysis for fMRI data. Neuroimage, 15, 1-15 (2002).

Zarahn, E., Aguirre, G. K., & D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. Neuroimage, 5(3), 179-97.

Acknowledgements: This work is supported by NIH grants R01 DA15410 and R01 EB004346-01A1

Table 1. – Frequency Responses of the differencing schemes

Type of Differencing		Fourier Domain Expression
D_1	No subtraction	$Y_1(e^{j\omega}) = Y(e^{j\omega})$
D_2	Simple subtraction	$Y_2(e^{j\omega}) = Y(e^{j\omega/2})(1-e^{-j\omega/2})$
D_3	Running subtraction	$Y_3(e^{j\omega}) = Y(e^{j\omega})(1-e^{-j(\omega+\pi)})$
D_4	Surround subtraction	$Y_4(e^{j\omega}) = Y(e^{j\omega})(2-e^{-j(\omega+\pi)}-e^{j(\omega+\pi)})$
D_5	Sinc subtraction	$Y_5(e^{j\omega}) = Y(e^{j\omega/2})(1-e^{-j\omega/2})S(e^{j\omega})^*$

*In an ideal Infinite Impulse Response filter, S is a perfect *rect* function but, depending on the implementation, it typically is the Fourier transform of a truncated *sinc* function instead.

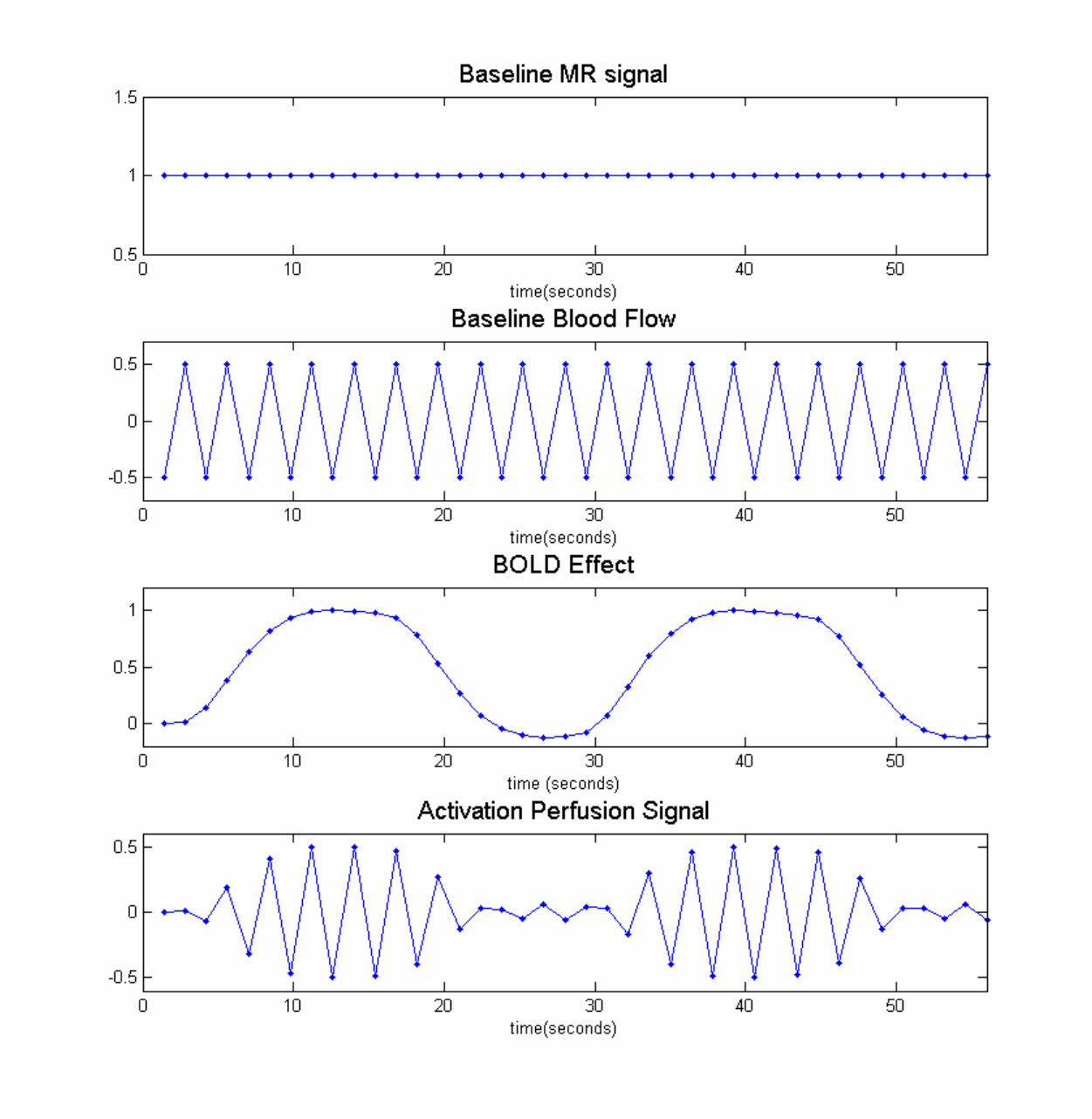

Figure 1. Predictors used in GLM for collected data (first 40 time points). The top two figures show baseline MR signal and baseline blood flow regressors and the bottom two figures show the BOLD effect and activation perfusion signal regressors. Note that the baseline blood flow predictor ranges from -0.5 to 0.5; this is done so that the corresponding parameter in the model represents a unit effect in the data.

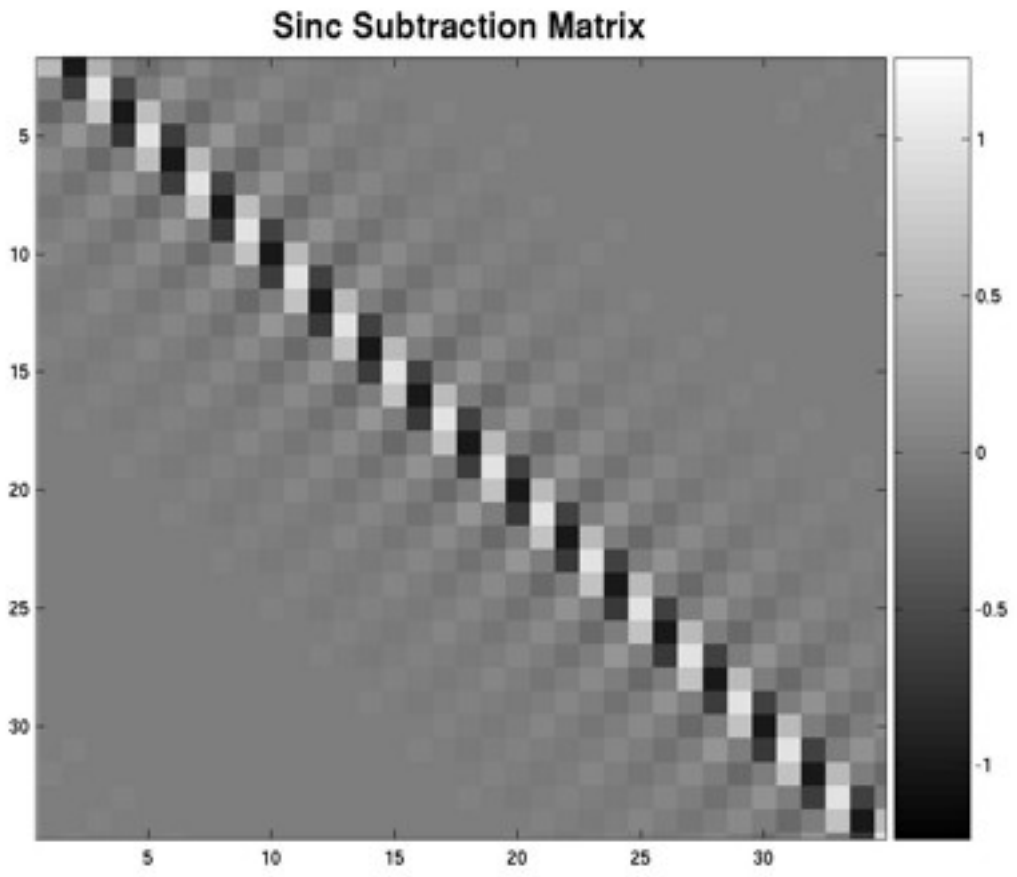

Figure 2. Example of a differencing matrix that implements a sinc subtraction (differencing matrix D_5).

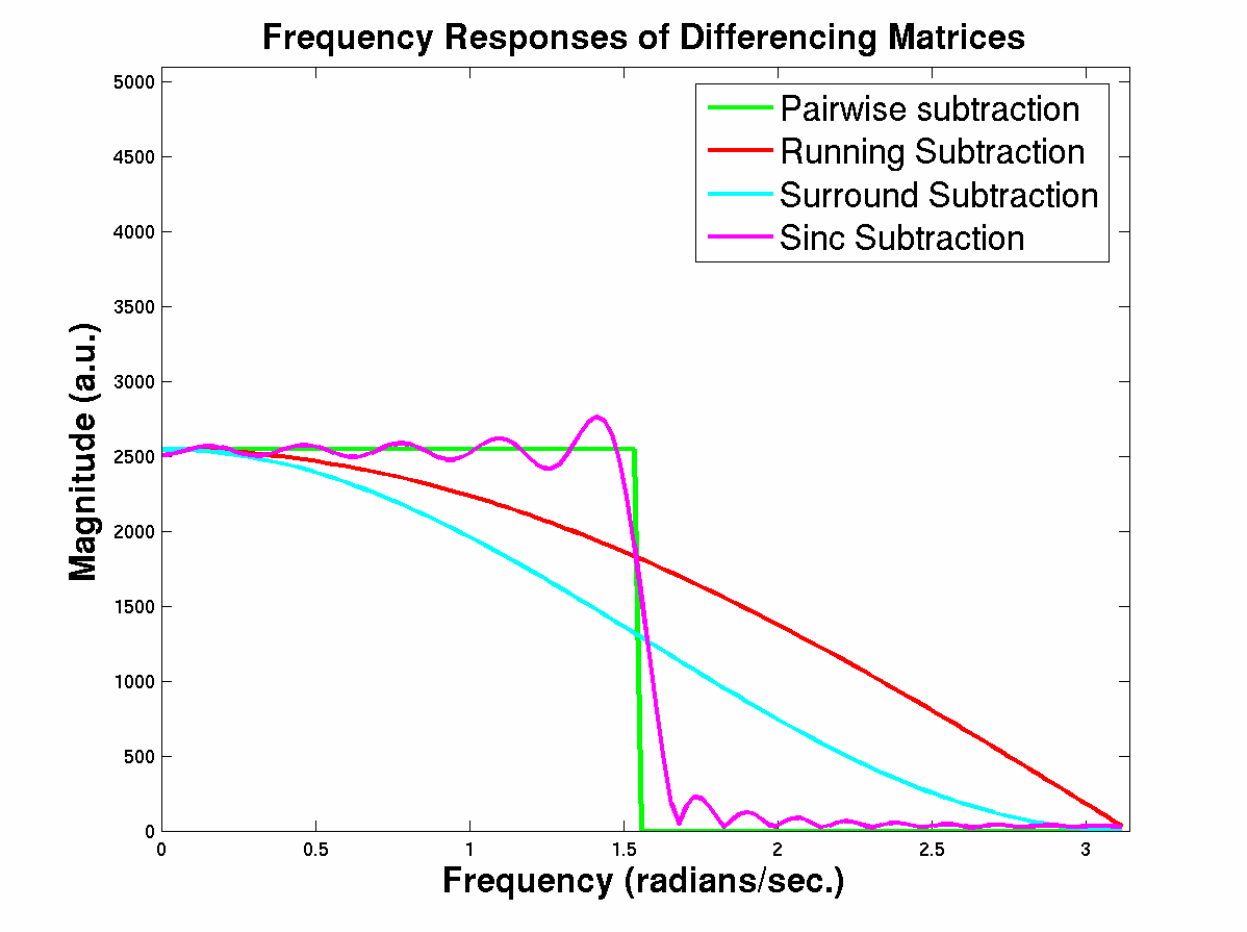

Figure 3. Expected Frequency responses as predicted by the equations in table 1. They correspond to each of the differencing matrices. These responses are in agreement with (Liu 2005).

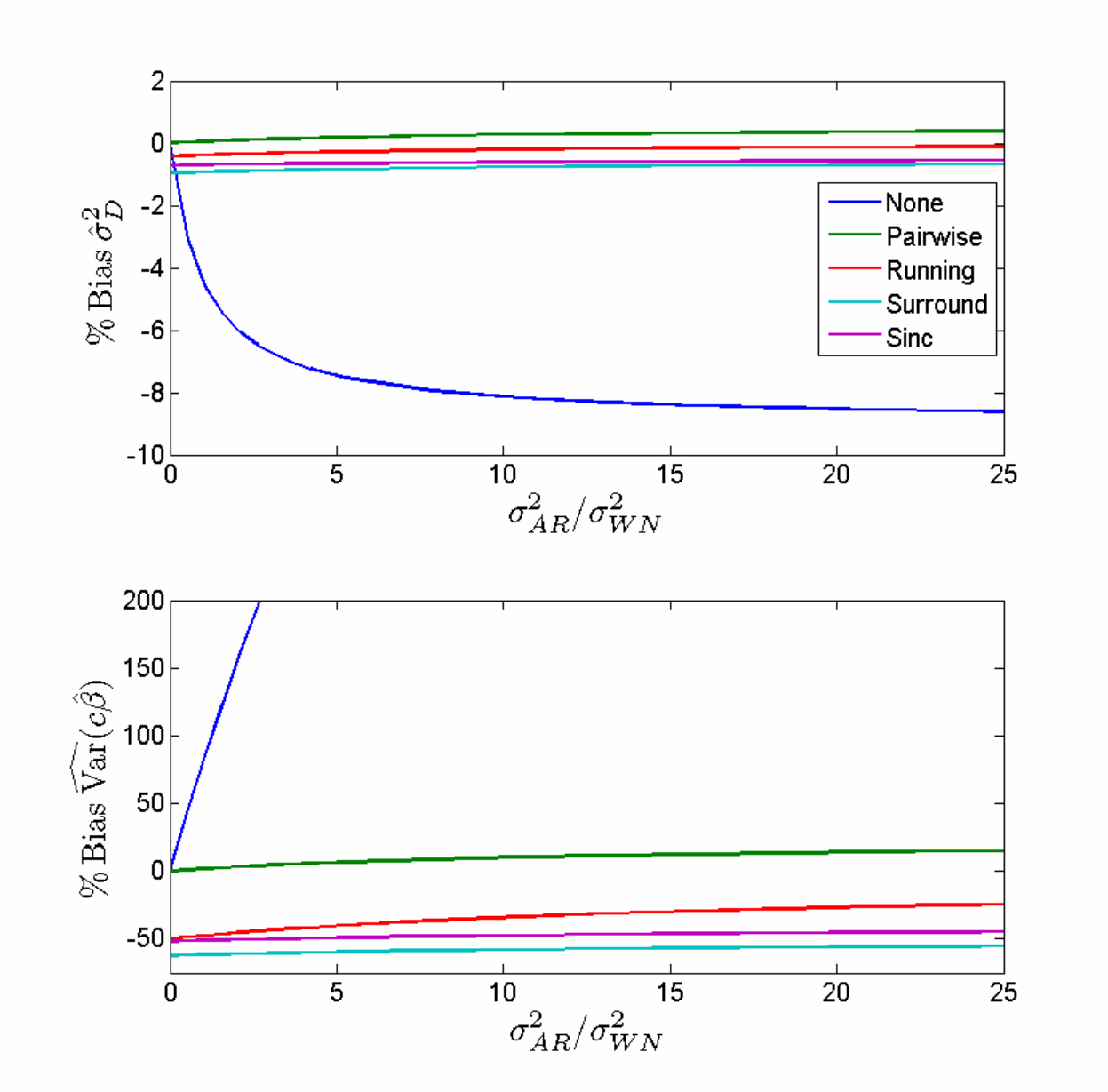

Figure 4. Bias of $\widehat{\mathrm{Var}}\left(\hat{\beta}\right)$, expressed as percent of true variance, for subtraction methods using OLS for different AR(1)+WN variance models. The study design used was the block design and the AR parameter, ρ, was fixed at 0.9.

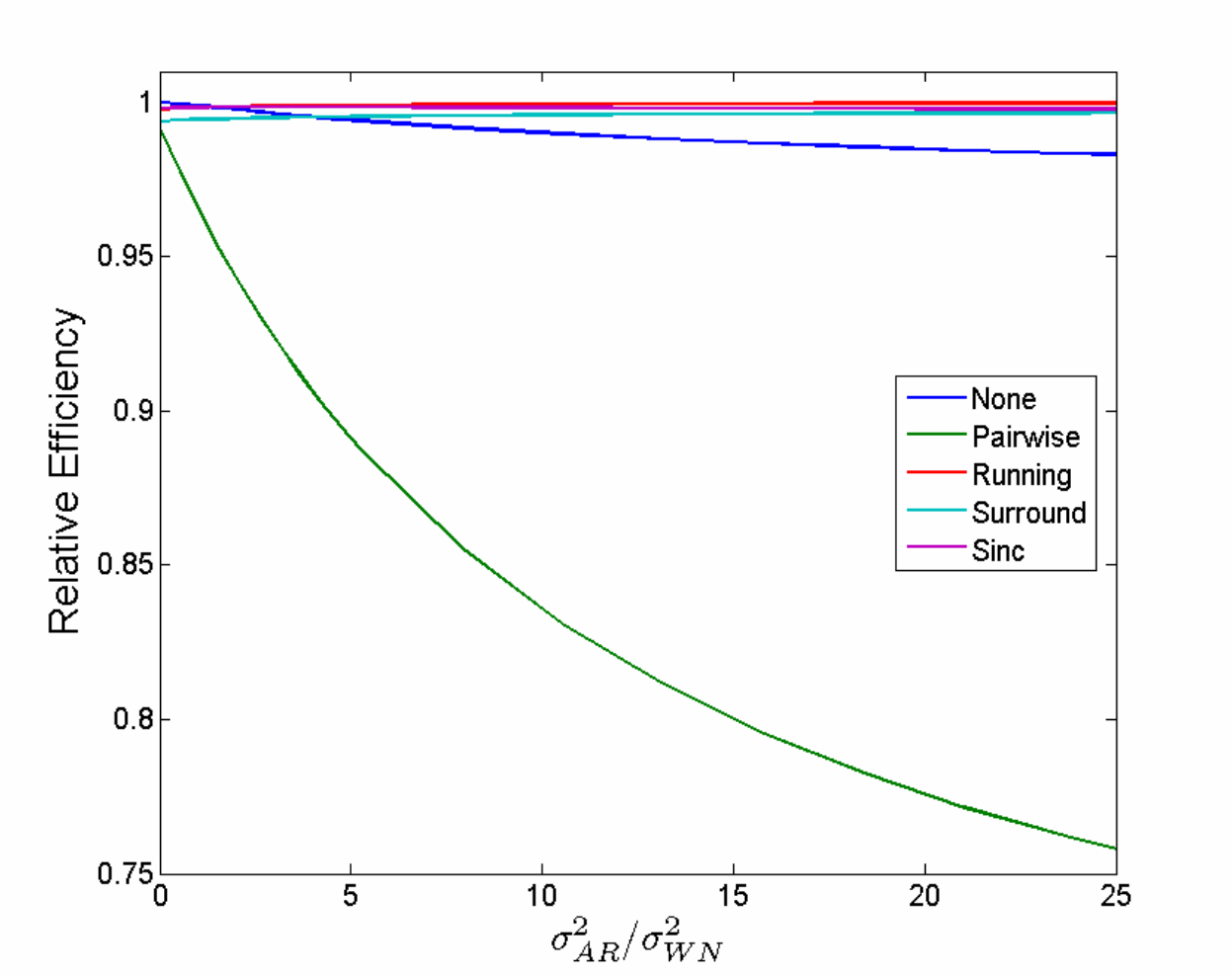

Figure 5. Estimation efficiency (relative to no differencing, GLS analysis) for subtraction methods using OLS for different AR+WN variance models for a block design study

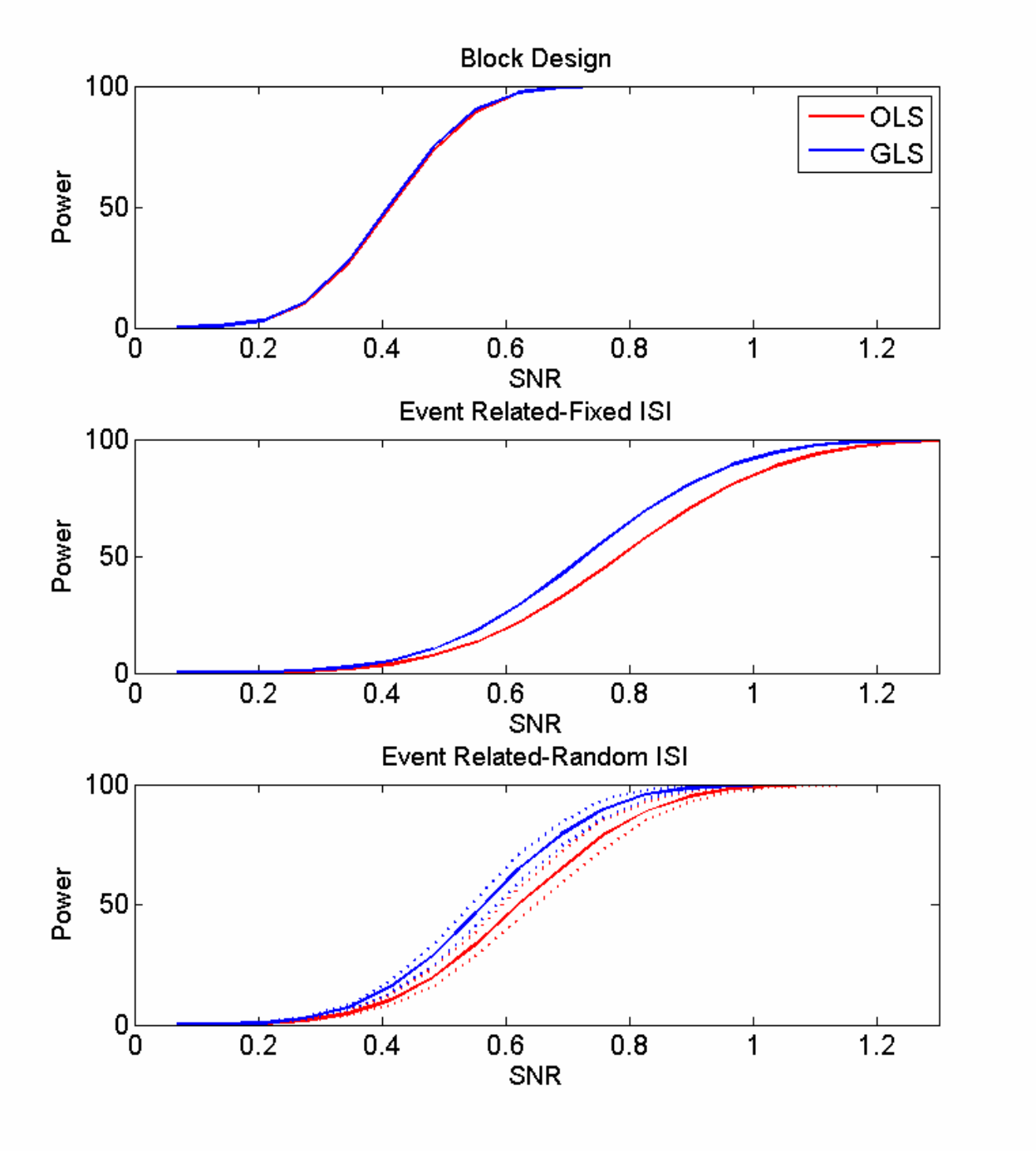

Figure 6. Power of the no differencing model estimated with GLS and pairwise subtraction estimated with OLS for 3 study designs. Random ISI event related design indicates average power over 100 iterations (solid) and ±2 standard deviations (dashed).

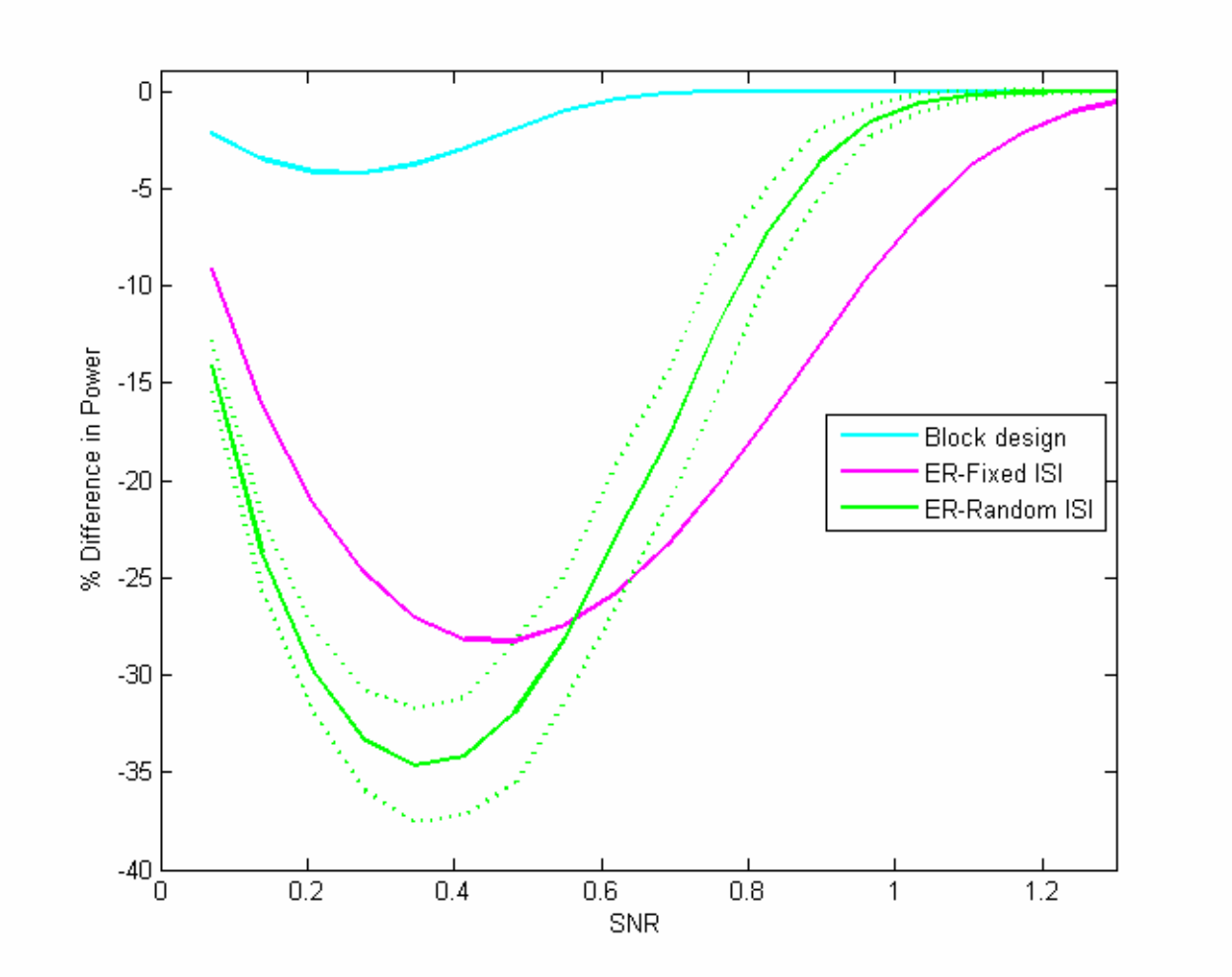

Figure 7. Relative Power 100×(OLS-GLS)/GLS for the different study designs . Random ISI event related design indicates average relative power over 100 iterations (solid) and ±2 standard deviations (dashed).

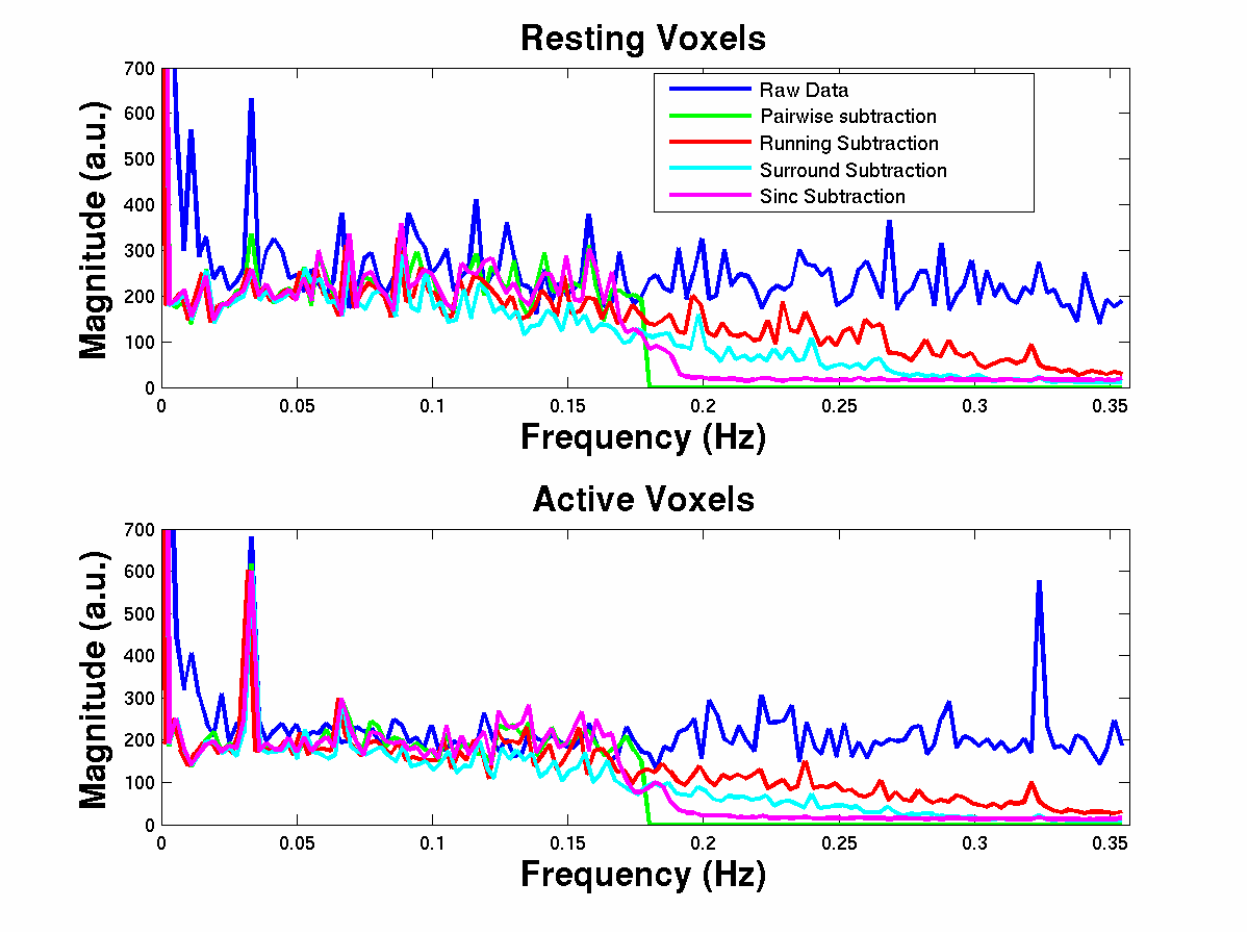

Figure 8. Frequency spectra of raw and differenced data in Resting (top) and Active (bottom) voxels.

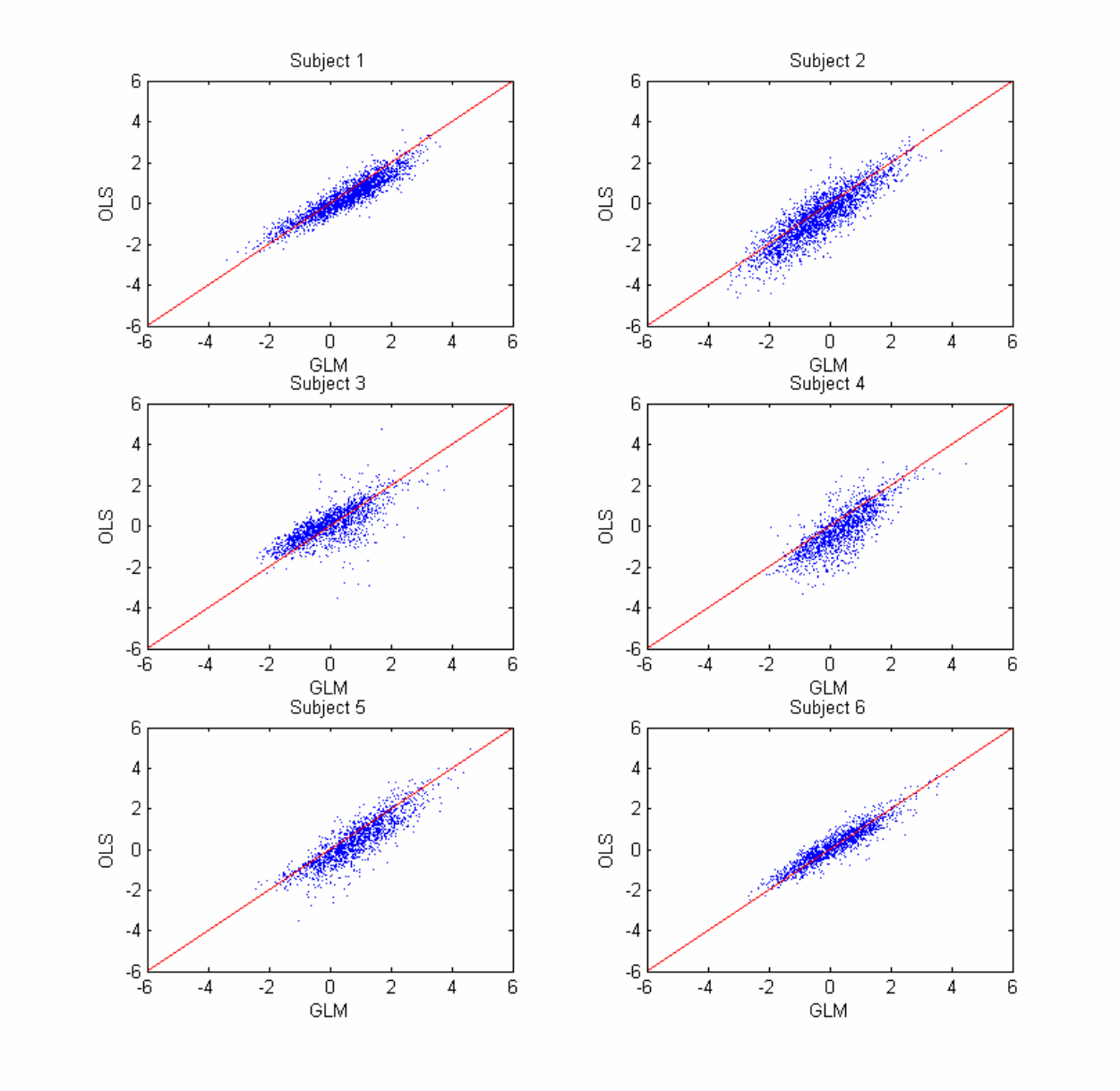

Figure 9. Comparison of Z values from pairwise differencing for all subjects using OLS and full data using GLS. Note that for most positive values GLS is larger and for most negative values GLS is smaller, indicating greater sensitivity.

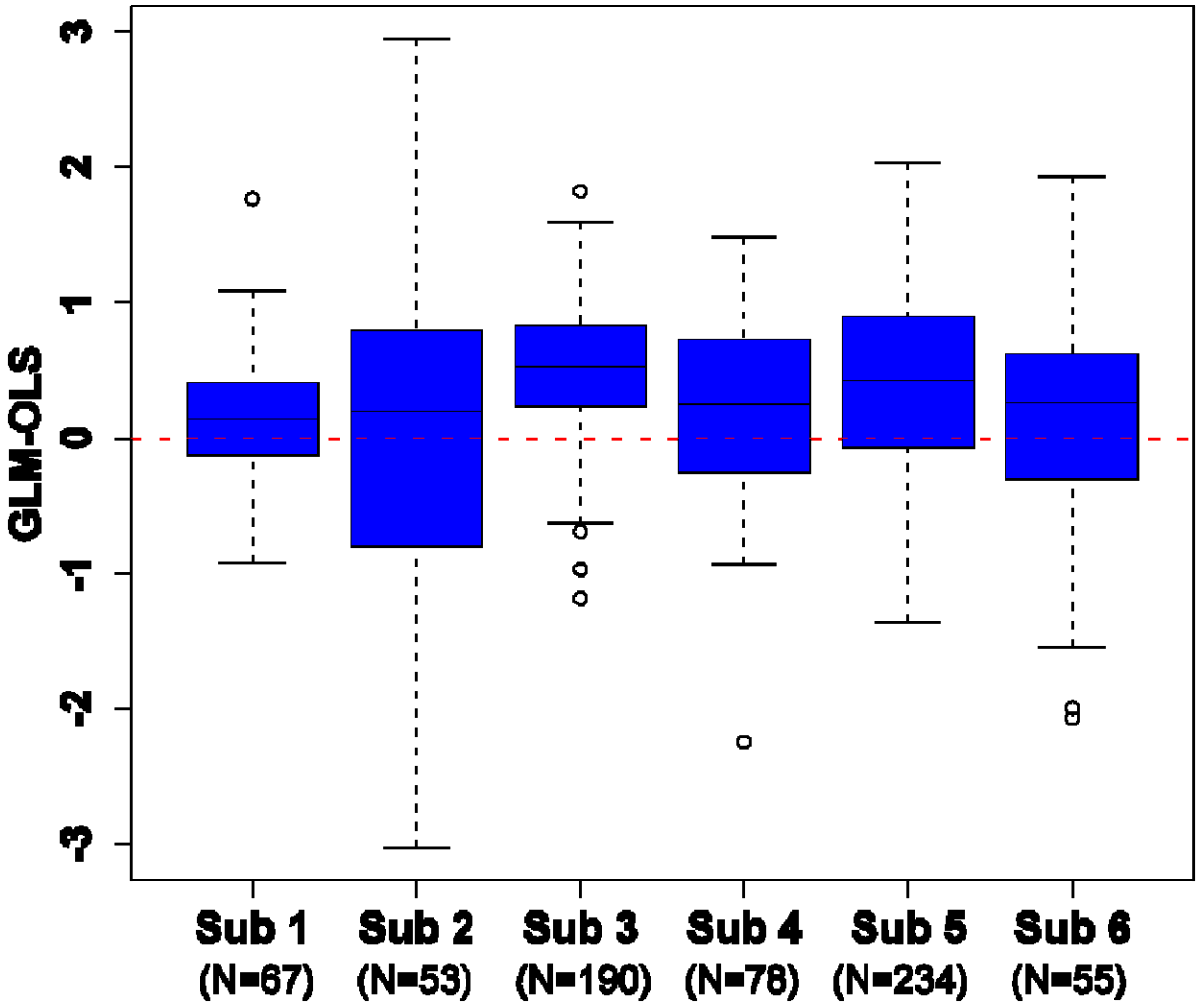

Figure 10. Boxplots of difference in Z values when the Z value from either method was larger than 2. The number of voxels included in each subject's boxplot are indicated beneath the subject number. Boxplots display the distribution of data by showing the median (horizontal black line in box), the first and third quartiles of the data (edges of the blue box), range of points included within 1.5*interquartile range (whiskers) and outliers (points). Positive values indicate greater sensitivity with full data using GLS.