Selection of Interstimulus Intervals for Event-Related fMRI Experiments

Darcie A. P. Delzell[a], Richard F. Gunst*[a], William R. Schucany[a], Wayne A. Woodward[a], Patrick S. Carmack[b], Qihua Lin[c], Jeffrey S. Spence[c], and Robert W. Haley[d]

March 20, 2009

[a] *Department of Statistical Science, Southern Methodist University, P.O. Box 750332, Dallas, TX 75275-0332, USA*

[b] *Department of Mathematics, University of Central Arkansas, 201 Donaghey Avenue, Conway, AR 72035, USA*

[c] *Department of Clinical Science, Biostatistics Division, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75390-8830, USA*

[d] *Department of Internal Medicine, Epidemiology Division, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75390-8874, USA*

* Corresponding author. Department of Statistical Science, Southern Methodist University, P.O. Box 750332, Dallas, TX 75275-0332. Fax +1–214–768–4035
*E-mail:*rgunst@smu.edu (R. F. Gunst).

Abstract

In order to properly formulate functional magnetic resonance imaging (fMRI) experiments involving interconnected mental activities, it is advantageous to permit great flexibility in the statistical components of the design of these studies. Issues that are fundamental to the selection of an appropriate design include considerations of the length of imaging sessions and the timing of stimuli. Major advances in understanding the implications of various statistical designs of fMRI experiments have taken place over the last decade. Nevertheless, new and increasingly difficult issues relating to the modeling of hemodynamic responses and the detection of activated brain regions continue to arise because of the increasing complexity of the experiments. In this article, statistical D-optimality criteria are used for the selection of interstimulus intervals (ISI) in event-related experimental designs. Emphasis is placed on the estimation of the parameters of a two-gamma hemodynamic response function. A number of constant, cyclical, and probability-based ISI selection alternatives are investigated. Several of the better-performing cyclic and probability-based ISI selection methods are identified. Applications concentrate on a single-stimulus-type model, but the methods are also applied to two-stimulus-type experiments and to experiments that have autoregressive errors. The proposed methods are easily extended to multiple stimulus types and to block designs. In the course of these investigations, the choice of design was shown to induce apparent cylic patterns in the data that are unrelated to the model that generated the data. Many of the designs investigated lessen or eliminate these cyclic patterns.

Key Words: D-optimality, optimal statistical design, sinusoidal trends

1 Introduction

Functional magnetic resonance imaging (fMRI) experiments that are designed to advance cognitive neuroscience are by their very nature unique to the individual experiments. In addition to the requisite considerations of the nature of the stimuli, the expected evoked responses, and the large number of measurements issues, the statistical components of an fMRI design can enhance or detract from the ability to detect changes in brain activity. Several recent articles have focused on the choice of stimulus times (STs) or interstimulus intervals (ISIs, the time from the initiation of one stimulus to the initiation of the next one) in an effort to improve hemodynamic response detection power or estimation efficiency. A number of general recommendations for selecting block or event-related designs and for choosing STs or ISIs have resulted, recommendations that should be considered and adapted as needed to the specifics of the proposed fMRI experiment. This article investigates the use of D-optimal design criteria for the selection of ISIs in an event-related fMRI experiment when the goal is to fit a nonlinear hemodynamic response function (HRF).

Event-related fMRI experiments are generally considered preferable to experiments with block designs when concerns over possible habituation or anticipation could seriously violate the assumption of linear-time-invariant (LTI) hemodynamic responses (Dale and Buckner 1997). Liu et al. (2001) demonstrate that block designs are advantageous for signal detection in single-trial-type fMRI experiments, while event-related designs are preferable for estimating the HRF. These findings were supported for multiple-trial-type designs in Liu and Frank (2004) and in Liu (2004).

With respect to the selection of ISIs, Josephs et al. (1997) used relatively long (with respect to the volume acquisition time, TR), equally spaced ISIs in order to allow multiple acquisitions of the complete time course of each hemodynamic response. They alternated the onset times of the stimuli so that they were either between volume acquisitions or at the midpoint of the volume acquisitions. This was done so that the peaks and troughs of the HRF could be more accurately estimated. Price et al. (1999), using a fixed ISI and TRs between 1 and 2 times the ISI, demonstrated the importance of varying the onset times of the stimuli relative to the TR. They showed clear disadvantages of having the TR be an integer multiple of the ISI, leading to acquisition times that are the same relative times during the time course of the HRF. Serious underestimation or overestimation of the HRF can occur, along with the attendant loss of statistical power in tests for activation.

For designs with a constant ISI, Bandettini and Cox (2000) found that for event-related fMRI experiments that have stimulus durations shorter than 2 seconds the optimal ISIs ranged from 10 to 12 seconds. They observed that shorter ISIs caused the HRFs to overlap, resulting in a loss of signal amplitude relative to baseline and the presence of a "pseudo pre-undershoot" due to the failure of the HRF to return to baseline prior to the next stimulus. Optimal inter-stimulus intervals were considered for designs with variable ISIs by Birn et al. (2002). They found that shorter ISIs, on the order of 2 seconds, were optimal. This was also the conclusion of Dale (1999), who found that alternating between task and rest states as quickly as possible gave the best HRF estimation. There remain widely varying opinions regarding the choice of optimal statistical designs for fMRI experiments, and Dale's (1999) comment remains valid: "It has been widely argued, based on empirical as well as theoretical evidence, that using ISIs of at least 15 sec is optimal and that using shorter ISIs results in a severe reduction in statistical power. However, numerous studies have shown highly reliable event-related fMRI response estimates using ISIs as short as seconds or less".

The present article was motivated by the desire to use statistical optimality criteria to select ISIs for a series of fMRI experiments on 1991 Persian Gulf War veterans. This extensive series of experiments is intended to better identify and define suspected brain abnormalities among veterans suffering from what is now termed the Gulf War Syndrome; see Haley et al. (2008) and the references therein. An interesting facet of these experiments is that all of them are fundamentally different and one cannot rely on a common rule of thumb such as "choose ISIs between a and b." Of particular importance in some of these experiments is the desire to fit a statistical model that describes the form of the HRF. The model might use a canonical form for the HRF and fit only scale parameters or it might be a nonlinear form such as a two-gamma model in which location, scale, and shape parameters are to be estimated. In part, interest in fitting such nonlinear models results from the potential to detect differences in parametric maps of parameter estimates that might provide important characterizations of differences in the HRFs between syndrome and control groups.

During a pilot study of the test protocols on a single subject, a perplexing issue arose while fitting models to the fMRI data. This issue was confirmed with the testing of a second subject months later using the same protocol. The experiment consisted of having the subjects silently repeat nonsense words after hearing them. Figure 1 shows the fMRI time courses from 4 voxels in the auditory cortex from the second subject. The time courses are of residuals from a quadratic drift fit to the fMRI signals, with a separate drift fit for each voxel. Superimposed is a nonparametric second-order loess curve fit (Cleveland and Devlin 1988, R Development Core Team 2008; span = 0.25) to the residuals, with a separate fit to each time course. The unexpected issue is the apparent cyclic behavior in the time courses. Cyclic behavior like this not only appeared in a large number of voxels in the auditory cortex but also in the average across 161 voxels chosen from a preliminary assessment of the presence of an HRF signal. Moreover, the cyclic pattern remained even after the fitting of a six-parameter two-gamma HRF to the drift residuals; see, for example, Figure 6. The similarity of this cyclic behavior in many of the individual voxels, in the averages, in the residuals, and for both of subjects was unexpected.

[Insert Figure 1]

Many articles on model fitting for fMRI data indicate the presence of cyclic behavior of the observed signal, often ascribed to patient cardiac and respiratory cycles or to unknown temporal confounds (e.g., Friston et al. 1995, Holmes et al. 1997, Price et al. 1999, Liu et al. 2001). Suggestive, but certainly not conclusive, in the low-frequency trend in the cyclic patterns in Figure 1 is that the patterns might be due to the statistical design itself; i.e. to the ISIs. At the bottom of the figure, the triangles indicate the stimulus times. There are clusters of relatively rapid stimulus times and gaps of less frequent stimulus times. The reason the cyclic patterns and the possibility of the design causing them are so surprising is that the ISIs were randomly and relatively uniformly selected from candidate ISIs of 2,4,...,20 seconds. That the random, uniform distribution of ISIs could cause artificial cycles in the data was a surprising result of these investigations.

The intent of this article is to investigate the selection of ISIs within the constraints imposed by the need to test a large number of subjects in an extensive series of fMRI experiments with limited magnet time available for each experiment. In such a setting statistical protocols for the selection of ISIs must be easily repeatable for each subject; they must adhere as faithfully as possible to the usual recommendations for randomization to alleviate, hopefully eliminate, unknown sources of bias; and they must be targeted to the goals of individual experimenters for detecting and modeling the hemodynamic responses.

2 Methods

2.1 ISI Design-Induced Sinusoids

ISI-induced cyclic effects on the measurement of hemodynamic responses were investigated by calculating, with and without additive random errors, theoretical values from two-gamma HRFs for a large number of alternative fixed-length designs. The functional form of the two-gamma HRF is

$$h\,(t;\theta) = c_1\left(\left[w_1\,(t)\,exp\left\{1 - w_1\,(t)\right\}\right]^{a_1} - c_2\left[w_2\,(t)\left\{exp(1 - w_2\,(t)\right\}\right]^{a_2}\right), \tag{1}$$

where $w_j(t) = t/d_j$ for $j = 1, 2$ and $h\,(t;\theta) = 0$ for $t < 0$. Using estimated parameter values from the pilot study, the parameters in (1) were set to $\boldsymbol{\theta}$ $=(a_1, a_2, c_1, c_2, d_1, d_2) = (13, 27, 5.1, .5, 5.9, 13.6)$. The influence of stimuli and random errors were obtained by simulating observable signals from the model

$$y(t_i) = f(t_i;\theta) + e_i = \sum_{j=1}^{m} h(t_i - T_j;\theta) + e_i, \tag{2}$$

where the t_i are n measurement times and the T_j are m stimulus times. Measurement times t_i were every 2 seconds. Random Gaussian noise e_i having mean 0 and standard deviation $\sigma_e = 1$ or 3 was added to the calculated signal; $\sigma_e = 3$ was estimated from the pilot study.

2.2 Design Optimality

Optimal statistical design for fMRI experiments requires the specification of three components: a statistical model that is to be fit to the data, temporal design features (session duration, minimum and maximum ISIs, etc.), and a criterion for the selection of optimal designs. In results reported in this article, the HRF model used was the two-gamma model (1) in which either (a) all 6 parameters were estimated or (b) only the scale parameter c_1 was estimated. The latter model is referred to as the *canonical* HRF model. For this model, the remaining parameter values for the two-gamma model were either those specified above or those in the fit to the auditory data of Glover (1999), converted to the form of (1). Limited investigations were conducted for multiple trial-type experiments in which several canonical models were included and only the scale parameters were estimated. This work is easily adapted to any other statistical model, including those that use basis functions with unknown scale constants.

A wide range of designs that include only a small number of distinct ISIs were investigated. In this article, the results for designs using 3 preselected ISIs are detailed, as are probability-based ISI selections. The minimum ISI evaluated was 4 seconds because of concern over the potential violation of LTI assumptions with shorter ISIs. The maximum ISI was set at 30 seconds in order to permit potential designs to have sufficient time duration for nonoverlapping HRF time courses. The duration of individual sessions was set at 300 seconds.

For an experiment in which only scale parameters are to be fit and fMRI signals from n volumes are to be obtained from a single session, a linear model can be used to represent the n measurements for each voxel. The model has the form

$$\mathbf{y} = \mathbf{X}\theta + \mathbf{e}, \tag{3}$$

where $\mathbf{y}$ is the vector of voxel fMRI measurements, $\mathbf{X}$ is the matrix whose columns are canonical HRF or basis function values, each convolved with the vector $\mathbf{T} = (T_1, \ldots, T_m)$ of stimulus times, $\boldsymbol{\theta}$ is the vector of scale parameters, and $\mathbf{e}$ is a vector of independent random errors. Additional covariates, if desired, can also be included in $\mathbf{X}$ but are beyond the scope of this work. The covariance matrix of the linear least squares estimator $\hat{\boldsymbol{\theta}} = (\mathbf{X'X})^{-1}\,\mathbf{X'y}$ is $\boldsymbol{var}\{\hat{\boldsymbol{\theta}}\} = (\mathbf{X'X})^{-1}\,\boldsymbol{\sigma_e^2} = \boldsymbol{\Sigma_\theta}$, where $\boldsymbol{\sigma_e^2}$ is the variance of the errors. Note that the variances of the estimators of the individual scale parameters are the diagonal elements of $\boldsymbol{\Sigma_\theta}$.

When an experiment is designed in which all 6 parameters of the nonlinear HRF model (1) are to be estimated, a variety of computational algorithms are available to solve for the nonlinear least squares estimate of $\boldsymbol{\theta}$. Under very broad conditions that are satisfied by model (2), Seber and Wild (1989, Chapter 12) show that the nonlinear least squares estimator exists and for large n it is approximately normally distributed with covariance matrix $\boldsymbol{var\{\hat{\theta}\} = (\mathbf{F'F})^{-1}\,\sigma_e^2 = \Sigma_\theta}$, where the elements F_{ij} of the matrix $\boldsymbol{F}$ are the partial derivatives $\frac{\partial f(t_i;\theta)}{\partial \theta_j}$. Nonlinear least squares estimates of $\boldsymbol{\theta}$ were calculated using the NLMINB function in R (R Development Core Team 2008).

Design optimality generally focuses on the precision of parameter estimators; i.e., on choosing designs that minimize some function of estimator variances. Several recent research articles (e.g., Bandettini and Cox 2000; Liu et al. 2001; Birn et al. 2002; Liu 2004; Liu and Frank 2004) evaluated competing designs using the criterion of A-optimality. A-optimality chooses stimulus vectors so that the sum of the variances of the parameter estimators is minimized; i.e., designs that minimize $\boldsymbol{tr\{\Sigma_\theta\}}$, where $\boldsymbol{tr\{\ldots\}}$ is the trace operator.

The A-optimality criterion is widely used and a very effective criterion for evaluating proposed statistical designs. Many other design criteria are available that incorporate more than the variances of the estimated parameters. Alternatives that are also widely used and highly effective make use of properties of the entire covariance matrix $\boldsymbol{\Sigma_\theta}$. One such criterion is D-optimality (Atkinson et al. 2007). D-optimality chooses stimulus vectors that minimize the volume of a simultaneous confidence ellipsoid for the vector of parameters $\boldsymbol{\theta}$. Confidence ellipsoids are geometric generalizations of confidence intervals for a single parameter. D-optimality minimizes the determinant of $\boldsymbol{\Sigma_\theta}$; i.e., minimize $|\,(\mathbf{X'X})\,|^{-1}$ or $|\,(\mathbf{F'F})\,|^{-1}$, depending on whether a linear or nonlinear HRF model is being fit. For scaling purposes, $\mathrm{d}(\mathbf{T}) = |\,(\mathbf{X'X})\,|^{-1/p}$ or $\mathrm{d}(\mathbf{T}) = |\,(\mathbf{F'F})\,|^{-1/p}$ are often calculated, where p is the number of parameters being estimated.

In practice, a number of candidate designs with different stimulus vectors $\mathbf{T}$, defined either by specifying the individual stimulus times or by specifying the successive ISIs, are compared by calculating $\mathrm{d}(\mathbf{T})$ for each. Let $\mathbf{T}^*$ designate the stimulus vector that produces the minimum $\mathrm{d}(\mathbf{T})$. Relative comparisons among the candidate designs are made by calculating each design's D-efficiency: $\mathrm{d}_{eff}(\mathbf{T}) = \frac{d(\mathbf{T}^*)}{d(\mathbf{T})}$, expressed as a percent. The stimulus vectors with d_{eff} values close to 100% are then evaluated with respect to any additional experimental restrictions or desirable features that are special to the fMRI experiment, including whether the candidate designs might induce cyclic behavior in the signal that is an artifact of the design itself.

2.3 Computation Issues

The validity of the asymptotic variance results for nonlinear least squares estimators is crucial for the efficient use of statistical fMRI design optimality criteria. The appropriateness of these asymptotic results was examined in several ways. Asymptotic variances for nonlinear parameter estimators were calculated and compared to results from simulations in order to verify that asymptotic calculations were in agreement with results that could be expected by sampling from common fMRI fixed-time-length sessions. In addition, graphs of a variety of two-gamma HRFs with differing parameter values were made and visually examined to determine the sensitivity of the theoretical curves to changes in the individual parameter values. The five stimulus vectors shown in Figure 2 were used for this evaluation. The stimulus vectors were chosen to provide a variety of reasonable ISIs that might be proposed for use. The parameter values used for this evaluation were those from the pilot study; see Section 2.1.

[Insert Figure 2]

When all 6 parameters of the two-Gamma HRF model (1) were estimated, a number of computational difficulties were encountered if nonlinear least squares calculations were made with unconstrained parameter values. As detailed in the Results section, the difficulties were mainly associated with the shape parameters α_1 and α_2 in the two-Gamma HRF model (1). Subsequent evaluations focused on either fixing the shape parameters and estimating the remaining 4 parameters, or placing reasonable constraints on the magnitudes of all 6 parameter estimates. Based on an examination of published results and on preliminary evaluations of a variety of alternatives, the following constraints were imposed on the nonlinear least squares parameter estimates: $\hat{a}_1 \in [2,20]$, $\hat{a}_2 \in [20,35]$, $\hat{c}_1 \in [1.5,10]$, $\hat{c}_2 \in [0.1,1]$, $\hat{d}_1 \in [2.5,10]$, and $\hat{d}_2 \in [6,30]$.

2.4 Candidate Designs

When one desires to estimate parameters of an HRF, several studies have shown that designs using a single ISI are not optimal; e.g., Birn et al (2002), Burock et al. (1998), Dale (1999), Friston et al (1999). Nevertheless, for comparison purposes, in this work constant ISIs of 4, 7, 10, and 14 seconds were evaluated and were designated designs $C_1 - C_4$. Preliminary investigations of a number of cyclic ISI designs (not to be confused with the cyclic behavior of observed signals) of the form {a,b,c,a,b,c,...} were conducted. Small numbers of distinct ISIs were emphasized because of investigators' desires to keep designs as simple as possible. Three of the better cyclic designs, each having 3 distinct ISIs were comprehensively evaluated: {4,5,8}, {6,9,14}, and {4,6,20}. They were designated designs $CD_1 - CD_3$. The cyclic designs were both nonrandomized and randomized.

Eighteen probability-based ISI designs were evaluated. Figures 3 and 4 show the designs. A variety of uniform, unimodal, and bimodal designs are depicted and were evaluated. The designs are designated $PD_1 - PD_{18}$. For these designs, the number of distinct ISIs chosen was not fixed; i.e., each time an ISI was selected, it was done so by sampling with replacement using the probabilities indicated in Figures 3 and 4. A number of preliminary investigations of randomly generated designs revealed that generating 25 designs randomly from each of $PD_1 - PD_{18}$ would provide sufficiently precise mean $d(\mathbf{T})$ values for comparison purposes with $d(\mathbf{T})$ values from designs with constant or cyclic ISIs.

[Insert Figures 3 and 4]

2.5 HRF Parameters

The evaluation of statistical designs requires stipulating values for the parameters of the HRF. For the two-gamma model (1), all 300 combinations of the following parameter values were evaluated with the constant and cyclic ISI designs: $\{a_1, a_2\} \in \{12, 20\}$, $\{12, 28\}$, $\{20, 20\}$, $\{20, 28\}$, $\{20, 35\}$; $c_1 \in \{1.5, 4, 6.5, 10\}$; $c_2 \in \{.2, .35, .5\}$; and $\{d_1, d_2\} \in \{2.5, 6\}$, $\{5, 13.5\}$, $\{7.5, 13.5\}$, $\{7.5, 21\}$, $\{10, 16\}$. After these designs were evaluated, the best of the constant and cyclic designs were compared to the probability-based designs using the following 4 parameter vectors: $\boldsymbol{\theta}_{min} = (12,20,10,.5,5,13.5)$, $\boldsymbol{\theta}_{median} = (12,20,4,.2,5,13.5)$, $\boldsymbol{\theta}_{max} = (12,20,1.5,.2,10,16)$, $\boldsymbol{\theta}_{pilot} = (12,28,4,.35,5,13.5)$. The first 3 parameter vectors were chosen from the 300 possible sets of parameter vectors because they were consistently at or near the minimum, median, and maximum d($\mathbf{T}$) values across the constant and cyclic ISI designs. The fourth parameter vector was chosen because it is close to the estimated parameter vector from the pilot study. Figure 5 depicts the time courses of the two-gamma HRFs using these 4 parameter vectors. These 4 parameter vectors were used in the comparisons of the constant and cyclic designs with the probability-based designs.

[Insert Figure 5]

3 Results

3.1 Methodological Issues

Figure 6 displays two sets of graphs that illustrate one of the unexpected impacts that the choice of ISIs can have on measured responses. The design for Figure 6(a) and 6(b) is the actual design used in the pilot study; however, the data are generated from model (1), where the fitted HRF is convolved with the stimulus vector and with additive independent normal errors with mean 0 and standard deviation 3. Figure 6(b) is a graph of the residuals after a two-gamma HRF model was fit to the generated data in Figure 6(a). The cyclic pattern in these graphs is very similar to those in Figure 1, yet there is no sinusoidal term in the model that was used to generate the data.

The design used in Figure 6(c) and 6(d) is a more extreme design in which the stimulus times are clustered and there are gaps between the clusters of stimuli. While the extreme design would not be a design of choice, it could result from designs that have stimulus times generated completely at random. Once again, there are apparent cyclic patterns when no such effects are included in the model used to generate the data.

[Insert Figure 6]

Nonlinear least squares estimation of HRF parameters from models like the two-gamma model (1) can be very sensitive to the specification of the form of the model. Even if great care is taken to reduce this sensitivity to the form of the model, computational difficulties can persist. Figure 7(a) shows the shape of the two-gamma HRF using the estimated auditory parameters $(a_1, a_2, c_2, d_1, d_2)$ from Glover (1999) and $c_1 = 6$. The remaining graphs in this figure illustrate the effects of changing, respectively, scale (c_1), location (d_1 and d_2), and shape (a_1 and a_2) parameters by 50%. Dramatic differences in the HRF curves occur when scale or location parameters are changed but not when shape parameters are changed. Because of this insensitivity of the HRF to the shape parameters, computations can result in wildly different values for fitted shape parameters when data differ trivially. More importantly, poor estimates of the shape parameters can result in poor estimates of the remaining parameters because the estimating equations require simultaneous estimation of all the parameters.

[Insert Figure 7]

When simulation estimates were compared to asymptotic theoretical variance and covariance values for fits to the two-gamma model (2) using the 5 vectors of stimulus times shown in Figure 2, large discrepancies between sample and asymptotic variances and covariances were observed when all the parameters were unconstrained. Using the parameters from the pilot study and an error standard deviation of 1, the average parameter variances for 5,000 simulations are shown in Table 1 along with the corresponding asymptotic values. There are large discrepancies between the sample and asymptotic values for all the parameter variances. If, however, the same simulation is rerun with the shape parameters fixed at their correct values, the sample and asymptotic variances for the scale and location parameters are in excellent agreement.

[Insert Table 1]

In practice, setting the shape parameters to constant values and only estimating the scale and location parameters would defeat the intended purpose of using these estimated parameters to possibly distinguish between syndrome and control groups. Such a procedure would likely also bias the estimated scale and location parameters and would lead to unreasonably small variances for the estimated parameters. Alternatively, reasonable constraints on all the parameter values greatly improves estimation using nonlinear least squares. When the same calculations were made with the same simulated data using the constraints in Section 2.3, dramatic improvements were observed. Table 2 compares simulated and asymptotic variances for the same 5 stimulus vectors used in Table 1. Simulations were also conducted with larger error variances. While the agreement was not as good as in Table 2, it was markedly better than if unconstrained parameter values were used. In addition, even if the shape parameters were poorly estimated, the variances and covariances of the remaining parameters were estimated well and fitted HRF curves were very reasonable. This dramatic improvement in agreement between simulated and asymptotic variance and covariances permitted the use of asymptotic values in the comparisons of ISIs for fMRI experiments.

[Insert Table 2]

3.2 Design Comparisons

Asymptotic covariance matrices were calculated for the constant ISI designs C_1 – C_4 and for the cyclic designs CD_1 – CD_3 using all the 300 parameter vectors of Section 2.5. A variety of parameter vectors must be evaluated for nonlinear fitting of the two-gamma HRF model because the asymptotic covariance matrix of the parameters contains the partial derivatives $\frac{\partial f(t_i;\theta)}{\partial \theta_j}$, which are functions of the parameter vector θ. For each design, the minimum, median, and maximum $d(\mathbf{T})$ value was determined, along with which parameter vector resulted in each.

Several conclusions were drawn from this initial evaluation. Cyclic designs CD_1 and CD_2 were clearly superior to the constant designs and the cyclic design CD_3 in terms of median and maximum $d(\mathbf{T})$ and no worse in terms of minimum $d(\mathbf{T})$. In addition, randomizing the cyclic designs produced results that were equivalent to the nonrandomized designs. This suggests that among these designs it is not the actual placement of the stimulus times that is important but the selection of the ISIs that are to be used. From both clinical and statistical perspectives, randomization is preferred and was used in all the remaining calculations reported below and in the designs for the forthcoming Gulf War study.

Mean d($\mathbf{T}$) values across all 18 probability-based designs were very consistent relative to one another for each of the 4 parameter vectors ($\boldsymbol{\theta}_{min}$, $\boldsymbol{\theta}_{median}$, $\boldsymbol{\theta}_{max}$, $\boldsymbol{\theta}_{pilot}$), even though the magnitudes of the individual d($\mathbf{T}$) values differed depending on which parameter vector was used. Moreover, D-efficiencies across all of the constant, cyclic, and probability-based designs were also very consistent across the 4 parameter vectors. Consequently, the results for $\boldsymbol{\theta}_{median}$ are used to indicate the overall performance of the various designs.

Table 3 provides a comparison of d($\mathbf{T}$) values for the constant and 3-ISI cyclic designs and for the 18 probability-based designs. The probability-based designs shown in Table 3 are those that produced the smallest mean d($\mathbf{T}$) values. Also included in the table are D-efficiency values for the designs, relative to the minimum d($\mathbf{T}$) value across these designs, design CD_1. The last column of the table indicates the number of stimuli that could be accommodated within the 300 second duration of the design, with the first stimulus always occurring at 2 seconds. For the probability-based designs, the d($\mathbf{T}$), D_{eff}, and the number of stimuli are the averages across the 25 replications of the design. Cyclic design CD_1 and the probability-based designs PD_{16} and PD_{17} performed well (D_{eff} >90%) relative to one another.

[Insert Table 3]

Investigations similar to these were conducted on experiments designed to have two different stimulus types (two-trial-type designs) and, consequently, two different HRF parameter vectors. Among those examined were the auditory HRF parameter vector used in this study coupled with a visual HRF parameter vector approximated from Dale and Buckner (1997, Figure C). A second pair of HRF parameter vectors were the visual stimulus vector and a parameter vector in which the scale parameter c_1 from the visual stimulus vector was increased by 50%. A third pair of parameter vectors was the approximate Dale and Buckner parameter vector and a parameter vector in which both location parameters and the second scale parameter vector c_2 were increased by 50%. The two stimulus types were randomly assigned throughout the time course with the constraint that each stimulus type appeared an equal number of times in the design. Errors were independent and normally distributed with mean 0 and standard deviation 1. Constant ISI designs were compared with the 18 probability-based designs. Cyclic designs were not evaluated because the purpose of this activity was only to check the feasibility of using these methods on a broader class of HRF models.

The results (not shown) were consistent with those reported above for the single-trial-type model. All of the constant ISI designs were inferior to the better probability-based designs. A number of the probability-based designs performed well: PD_{12}, PD_{17}, and PD_{18}. These designs all had D_{eff} values greater than 90%.

First-order autoregressive errors were also investigated. The experimental data were simulated using the HRF parameters from the pilot study and normal errors having mean 0, standard deviation 1, and lag-1 autocorrelations of -0.12, 0.18, and 0.50. The constant ISI designs were inferior to all the probability-based designs. Designs PD_8, PD_{10}, PD_{11}, PD_{16}, and PD_{17} all had D_{eff} values greater than 90%.

4 Discussion

Design-induced cyclic patterns in fMRI time series are a function of the design, the randomization of the ISIs, the strength of the HRF, and the strength of the noise. If data are generated without error, the HFR can be fit exactly and residuals are zero regardless of the design and the randomization. In particular, there is no residual cyclic pattern. When errors are added, the HRF is not fit exactly, either directly or after deconvolution, and cyclic patterns in the residuals can occur.

Figure 8 shows time series generated from the same HRF used in previous examples using 4 of the better designs shown in Table 3. The designs and randomization resulted in fewer clusters, gaps, and caused less dramatic amplitudes in the data, as well as less regular cyclic patterns than those shown in Figure 6. Moreover, when autoregressive models were fit to residuals from two-gamma HRF fits, Figure 9 shows that the AR orders selected by AIC in over 70% of the fits using the better cyclic and probability-based designs was 0, indicating white noise residuals. In contrast, using the same errors as for the cyclic and probability-based designs the fits to HRF residuals for the pilot design resulted in 100% of the fits selecting an AR order greater than 0.

[Insert Figures 8 and 9]

It must be emphasized, however, that these designs do not guarantee that there will be no apparent cyclic patterns, in spite of their superior statistical design properties. Figure 10 shows data generated from the same designs and the same HRF as were used in Figure 8 but with different random errors. While the amplitudes in the data remain smaller than those in Figure 6, weaker cyclic patterns do appear to be present due to the random clusters and gaps in the stimuli. Autoregressive model fits to the HRF residuals for all the simulated time series in Figures 8 and 10 indicate white noise residuals. This was not the case for the pilot study residuals, even when simulated with the same errors. In addition, it must be stressed that none of the better cyclic and probability-based designs were intentionally chosen for the purpose of eliminating the cyclic patterns in the data.

[Insert Figure 10]

The importance of these results is that the choice of STs or ISIs can have a dramatic effect on the perceived signal from actual fMRI data. If these cyclic patterns are an artifact of the design and not true effects of the magnet or subjects, temporal filtering could result in residuals that have weakened fMRI signals. By choosing a design with good statistical properties, one can evaluate the likelihood of cyclic patterns by generating random errors and using an HRF that is similar to one expected in the experiment. If a suitable HRF is not known, use of published HRF forms and parameter values can provide a reasonable assessment.

From a comprehensive statistical modeling perspective, the most serious issue with design-induced cyclic patterns is the perceived need to introduce additional terms to a model in order to account for the spurious behavior. Investigations in which such a comprehensive modeling approach was attempted when errors were actually independent often resulted in high-order autoregressive models being fit to the data or to HRF-fit residuals. This is not uncommon even in linear least squares regression with independent errors. If cyclic patterns are due to magnet inhomogeneities, analyses of calibration data could reveal the presence or absence of sinusoidal instrument fluctuations. By obviating the need for very complex autoregressive fitting of HRF residuals, inferences on parameters can be simplified and potentially have more power to detect differences in parameters across groups of subjects.

The pilot design was not one of the candidate designs considered in this work. The pilot design was a uniform design with ISIs in increments of 2 seconds from 2 and 20 seconds. Each of these ISIs occurred approximately an equal number of times and were randomized. The clusters and gaps in the design were a result of the randomization. The reason this design was not studied in the work reported in this article is the possibility that ISIs as short as 2 seconds might lead to violation of the LTI assumption. There is an additional difficulty with this design that is highlighted by the better designs shown in Table 3. The two best-performing cyclic designs and all of the probability-based designs permit ISIs that are not multiples of the TR, 2 seconds, but the pilot design had all ISIs a multiple of the TR.

The HRF used in most of this work had a peak centered at 5.87 seconds and a trough centered at 13.55 seconds, neither of which is a multiple of the TR. By including ISIs at even and odd seconds, greater opportunity of fitting the location parameters (times to peak) is possible. This is especially important because in practice it is unlikely that the location parameters are integer multiples of the TR. The cyclic designs with 3 fixed ISIs were expanded (not shown) to include all possible designs for which 3 ISIs were selected from $\{4,5, ..., 12\}$ seconds. In all of the better designs one of the ISIs was either 4 or 5 and every design had an even and an odd ISI. As mentioned above, Josephs et al. (1997) and Price et al. (1999) also recognized the importance of varying ISIs relative to the TR. Finally, cyclic designs with more than 3 distinct ISIs were investigated. Emphasis in this article is on designs with 3 distinct ISIs because 4 or more distinct ISIs did not provide demonstrable improvement over designs with 3.

D-optimality is not the only criterion that could be used to select a design for an fMRI experiment. However, an advantage of D-optimality as a statistical criterion is that it uses the determinant of the entire covariance matrix, not just the diagonal elements. Moreover, D-optimality is invariant to changes in the scales of the parameters, a property not shared by A-optimality.

Ultimately, the goals of the experiment and specifics of the functional tasks determine an appropriate design. There will usually be a number of alternative designs with high D-optimality efficiencies, some of which may be preferable to others from a neurological perspective; e.g. if longer ISIs are needed to return brain activity to baseline. This is especially important if two or more stimulus types are to be included in an experiment. The reason some of the probability-based designs performed well in the two-trial investigations but not in the one-trial investigations (e.g., PD_{12} and PD_{18}) is likely that their short ISIs caused too much overlap and cancellation of signals in the one-trial simulations but not in the two-trial simulations. In the two-trial simulations the two stimulus types interleaved and each had sequences in which the other stimulus was (randomly) repeated several times. This likely allowed a sufficient number of hemodynamic response time courses to be completed prior to the occurrence of the next stimulus of that type.

These methods have been used to assist investigators who are planning fMRI experiments for the upcoming Gulf War study. A variety of adaptations were easily accommodated. One visual memorization project required short ISIs because 9 different visual stimuli were to be included. For this experiment, the investigator wished to use 9 canonical HRFs for which all but the scale parameters c_1 were pre-specified. Using the methods described above, a cyclic design with randomized ISIs similar to CD_1 was selected. A quite different sensory perception experiment required that two stimulus types be used. Both stimuli were to be presented in blocks, one 3 seconds long and the other 3, 6, or 9 seconds. Due to the nature of the experiment, longer ISIs between 12 and 18 seconds, were required so that cumulative effects would not mask activation in sensory areas of the brain. Two two-gamma HRF models in which all 12 parameters were estimated and one in which only the two scale parameters were estimated were investigated. Two additional options were requested by investigators: optimal designs for a fixed number of stimuli and optimal designs for a fixed duration of a session. The former condition resulted in longer ISIs (e.g., 14, 15, and 16 seconds) while the latter, not unexpectedly, required somewhat shorter ISIs (e.g., 12, 14, and 15 seconds) so that more stimuli were included.

Acknowledgments

This study was supported by VA IDIQ contract number VA549-P-0027 awarded and administered by the Dallas, TX VA Medical Center. The content of this paper does not necessarily reflect the position or the policy of the U.S. government, and no official endorsement should be inferred. The authors would also like to thank Dr. Sergey Cheshkov for insightful comments on this work and Joel O'Hair and Ohn-Jo Koh for their valuable computational assistance.

References

Atkinson, A.C., Donev, A.N., and Tobias, R.D. 2007. *Optimum Experimental Designs, with SAS*, Oxford Univ. Press, New York.

Bandettini, P. A. and Cox, R. W. 2000. Event-Related fMRI Contrast When Using Constant Interstimulus Interval: Theory and Experiment, *Magnetic Resonance in Medicine* **43:** 540-548.

Birn, R.M., Cox, R.W., and Bandettini, P.A. 2002. Detection versus Estimation in Event-related fMRI: Choosing the Optimal Stimulus Timing, *NeuroImage* **15:** 252-264.

Buracas, G.T. and Boynton, G.M. 2002. Efficient Design of Event-Related fMRI Experiments Using M-Sequences, *NeuroImage* **16:** 801-813.

Burock, M.A., Buckner, R.L., Woldorff, M.G., Rosen, B.R., and Dale, A.M. 1998. Randomized Event-related Experimental Designs Allow for Extremely Rapid Presentation Rates Using Finctional MRI, *NeuroReport* **9:** 3735-3739.

Buxton, R.B., Liu, T.T., Martinez, A., Frank, L.R., Luh, W.-M., and Wong, E.C. 2000. Sorting

Out Event-related Paradigms in fMRI: the Distinction Between Detecting and Activation and Estimating the Hemodynamic Response, *NeuroImage* **11:** S457.

Cleveland, W.S. and Devlin, S.J. 1988. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting, *J. Amer. Stat. Assn.* **83:** 596-610.

Dale, A.M. 1999. Optimal Experimental Design for Event-related fMRI, *Hum. Brain Map.* **8:** 109-114

Dale, A.M. and Buckner, R.L. 1997. Selective Averaging of Rapidly Presented Individual Trials Using fMRI, *Hum. Brain Map.* **5:** 329-340.

Friston, K.J., Zarahan, Z., Josephs, O., Henson, R.N.A., and Dale, A.M. 1999. Stochastic Designs in Event-Related fMRI, *NeuroImage* **10:** 607-619.

Friston, K.J., Holmes, A.P., Poline, J-B, Grasby, P.J., Williams, S.C.R., Frackowiak, R.S.J., and Turner, R. 1995. An Analysis of fMRI Time-Series Revisited, *NeuroImage* **2:** 45-63.

Glover, G.H. 1999. Deconvolution of Impulse Response in Event-Related BOLD fMRI, *NeuroImage* **9:** 416-429.

Hagberg, G.E., Zito, G., Patria, F., and Sanes, J.N. 2001. Improved Detection of Event-Related Functional MRI Signals Using Probability Functions, *NeuroImage* **14:** 1193-1205.

Haley, R.W., Spence, J.S., Carmack, P.S., Gunst, R.F., Schucany, W.R., Petty, F., Devous, M.D. Sr., Bonte, F.J., and Trivedi,M.H. 2008. Abnormal Response to Cholinergic Challenge in Chronic Encephalopathy from the 1991 Gulf War, *Psychiatry Research:Neuroimaging* **171:** 207-220.

Holmes, A.P., Josephs, O., Buchel, C., and Friston, K.J. 1997. Statistical Modeling of Low-Frequency Confounds in fMRI, *NeuroImage* **5:** S480.

Josephs, O., Turner, R., and Friston, K. 1997. Event-Related fMRI, *Hum. Brain Map.* **5:** 243-248.

Liu, T.T. 2004. Efficiency, Power, and Entropy in Event-related fMRI with Multiple Trial Types, Part II: Design of Experiments, *NeuroImage* **21:** 401-413.

Liu, T.T. and Frank, L.R. 2004. Efficiency, Power, and Entropy in Event-related fMRI with Multiple Trial Types, Part I: Theory, *NeuroImage* **21:** 387-400.

Liu, T.T., Frank, L.R., Wong, E.C., and Buxton, R.B. 2001. Detection Power, Estimation Efficiency, and Predictability in Event-Related fMRI, *NeuroImage* **13:** 759-773.

Price, C.J., Veltman, D.J, Ashburner, J., Josephs, O., and Friston, K.J. 1999. The Critical Relationship Between the Timing of Stimulus Presentation and Data Acquisition in Blocked Designs with fMRI, *NeuroImage* **10:** 36-44.

R Development Core Team 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Seber, G.A.F. and Wild, C.J. 1989. *Nonlinear Regression*. Wiley, New York.

Wager, T.D. and Nichols, T.E. 2003. Optimization of Experimental Design in fMRI: A General Framework Using a Genetic Algorithm. *NeuroImage* **18:** 293-309.

Table 1. Comparison of Asymptotic and Sample Variances for Unconstrained Parameter Values.

Stimulus Vector	Variance Calculation	$\boldsymbol{a_1}$	$\boldsymbol{a_2}$	$\mathbf{d}_1$	$\mathbf{d}_2$	$\mathbf{c}_1$	$\mathbf{c}_2$
1	**Asymptotic**	1.7	27	.007	.059	.042	.001
	Sample	6.1	78	.252	3.627	202.457	.014
2	**Asymptotic**	1.8	33	.009	.086	.048	.002
	Sample	10.3	127	.714	6.565	422.140	.029
3	**Asymptotic**	1.5	32	.009	.116	.056	.002
	Sample	2.3	44	.094	.542	31.885	.004
4	**Asymptotic**	3.9	29	.006	.059	.058	.001
	Sample	10.4	83	2.068	2.153	317.075	.040
5	**Asymptotic**	3.5	60	.021	.133	.102	.003
	Sample	4.5	84	.112	.782	6.068	.006

Table 2. Comparison of Asymptotic and Sample Variances for Constrained Parameter Values.

Stimulus Vector	Variance Calculation	$\boldsymbol{a_1}$	$\boldsymbol{a_2}$	$\mathbf{d}_1$	$\mathbf{d}_2$	$\mathbf{c}_1$	$\mathbf{c}_2$
1	**Asymptotic**	1.74	26.7	.007	.059	.042	.001
	Sample	1.67	21.0	.007	.061	.042	.001
2	**Asymptotic**	1.76	32.9	.009	.086	.048	.002
	Sample	1.73	23.4	.009	.086	.046	.001
3	**Asymptotic**	1.45	31.6	.009	.116	.056	.002
	Sample	1.44	23.4	.010	.121	.056	.002
4	**Asymptotic**	3.86	29.2	.006	.059	.058	.001
	Sample	3.52	21.9	.006	.060	.054	.001
5	**Asymptotic**	3.54	59.8	.021	.133	.102	.003
	Sample	3.42	30.1	.017	.143	.100	.003

Table 3. Comparison of Most Efficient D-optimal Designs Using the Parameter Vector $\boldsymbol{\theta_{median}}$.

Design	d(**T**)	D_{eff}	$\mathbf{N}_{stim}$
C_1	1.18	12%	75
C_2	.50	28%	43
C_3	.35	40%	30
C_4	.29	48%	22
CD_1	.14	100%	54
CD_2	.17	82%	32
CD_3	.20	70%	30
PD_8	.16	88%	35
PD_{10}	.16	88%	32
PD_{11}	.16	88%	38
PD_{16}	.15	93%	39
PD_{17}	.15	93%	45

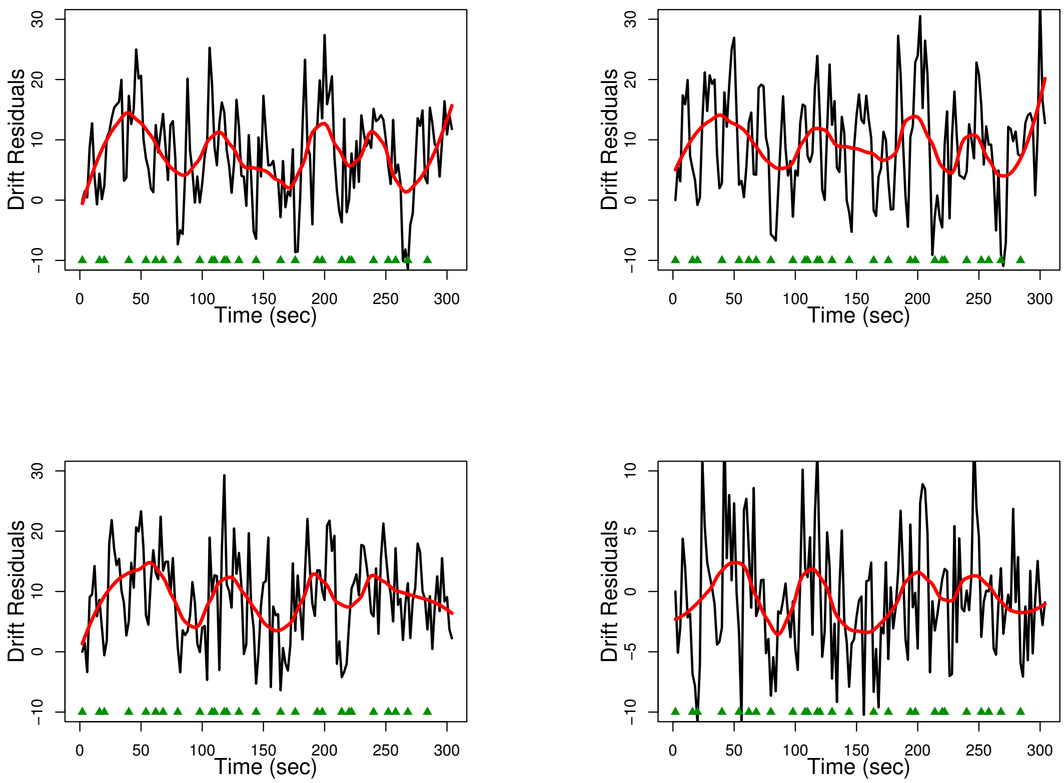

Fig. 1: Quadratic drift residuals from 4 voxels in the auditory cortex. Nonparametric loess curve fit in red. Stimulus activiation times indicated by the green triangles.

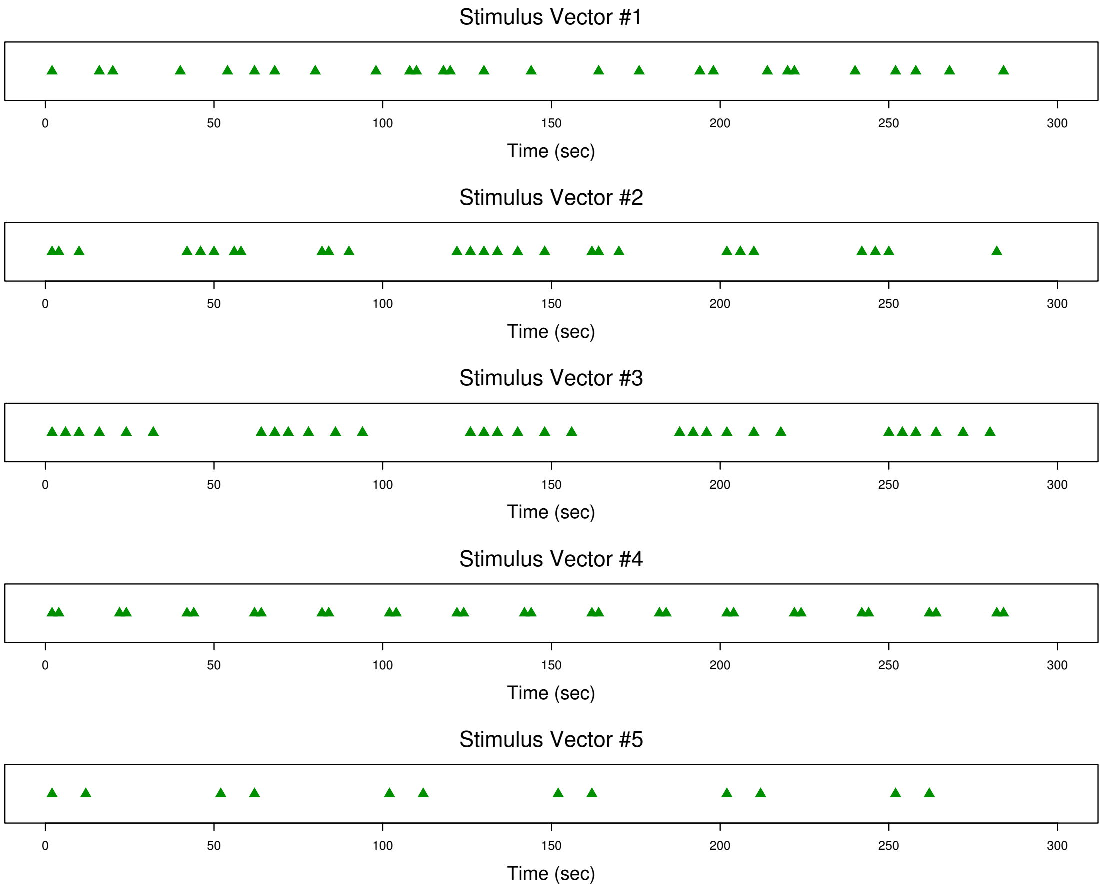

Fig. 2: Stimulus activation times for comparing simulated and asymptotic variances. Stimulus times indicated by the green triangles.

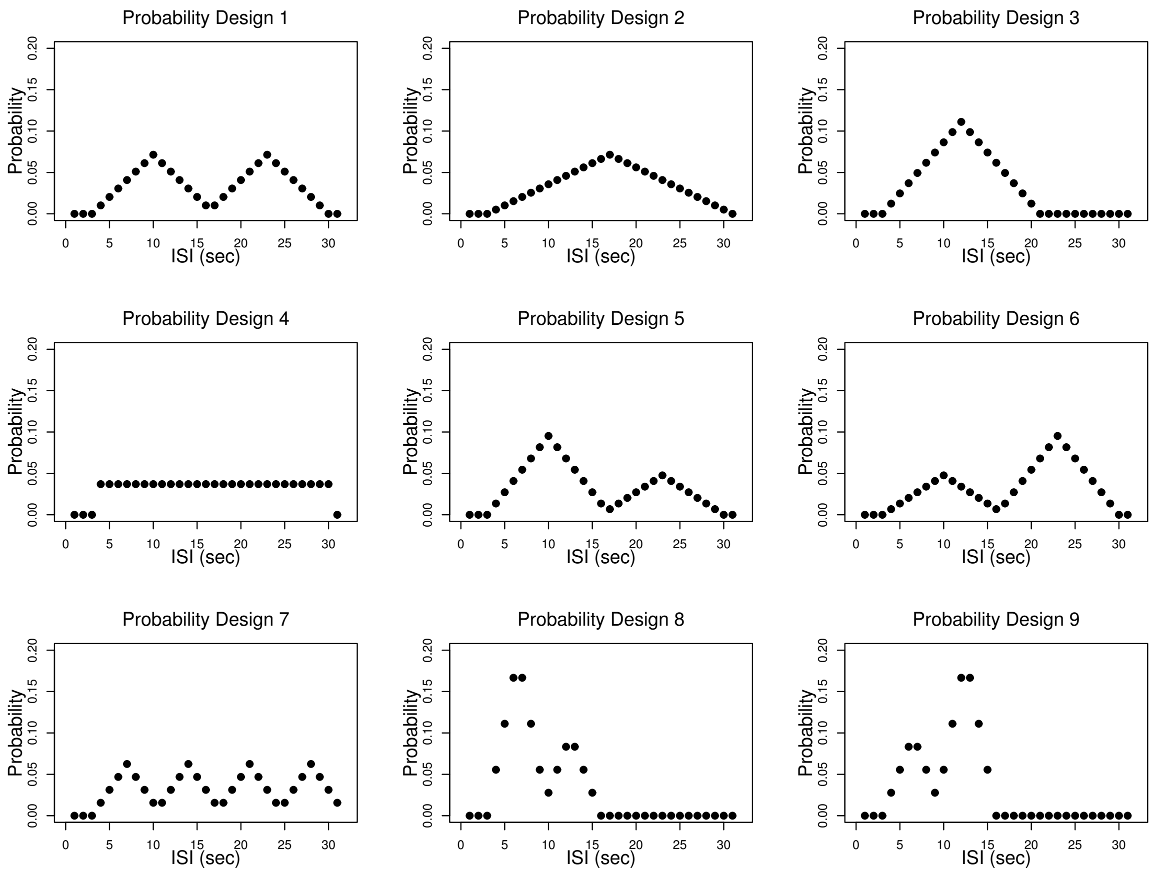

Fig. 3: Discrete probability-based designs for selecting ISIs.

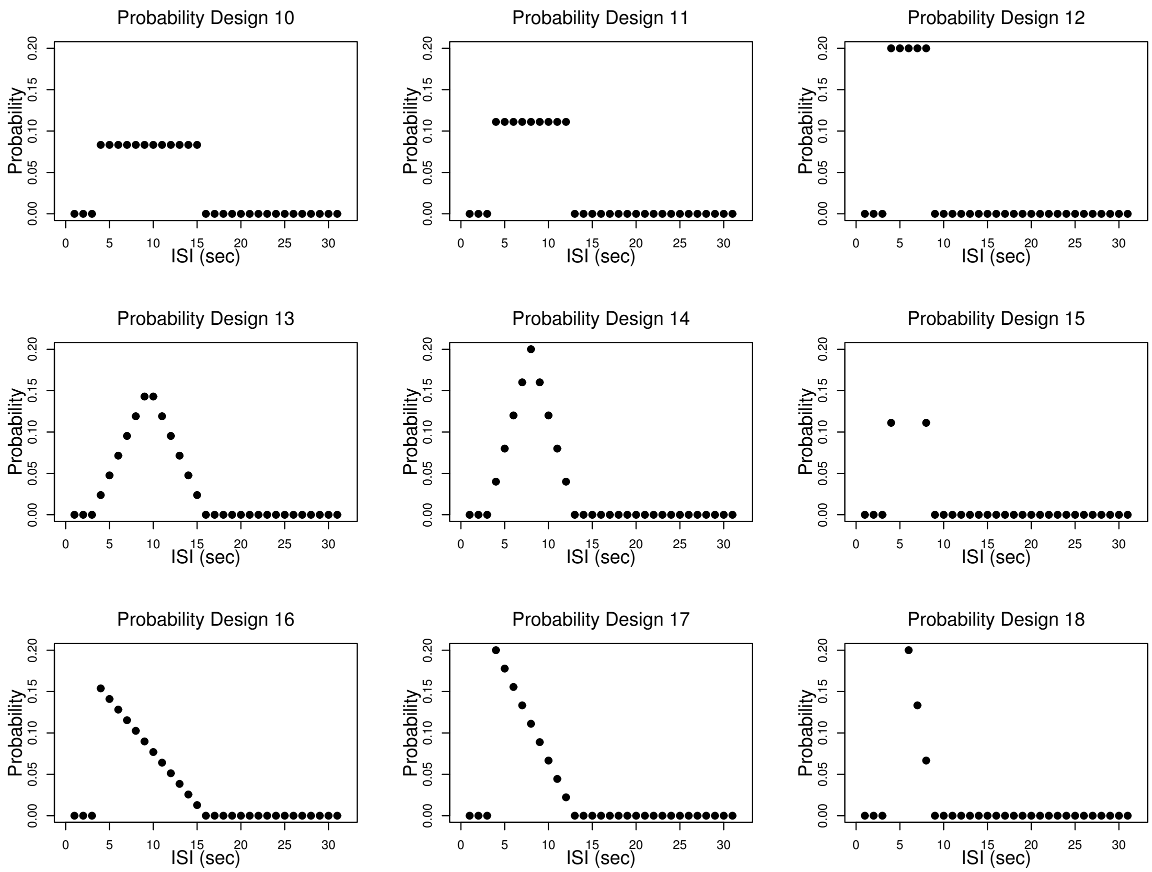

Fig. 4: Additional discrete probability-based designs for selecting ISIs.

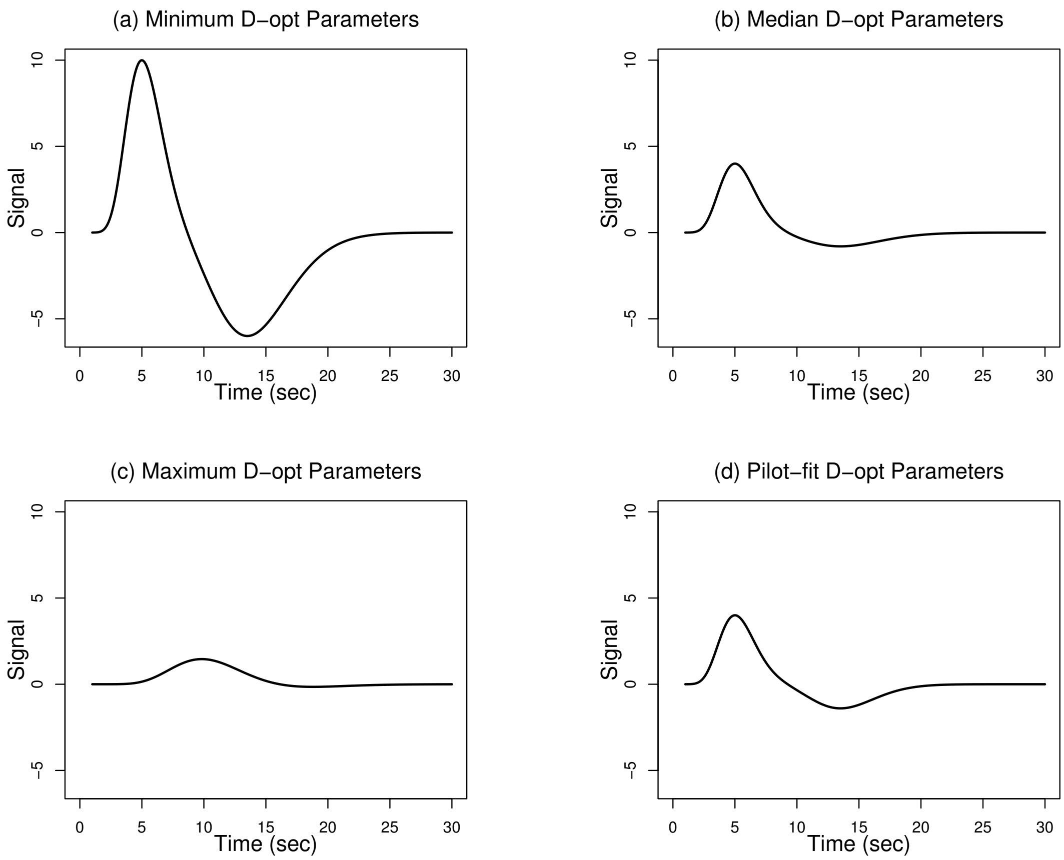

Fig. 5: Theoretical HRFs that produce different D-optimal values across constant and cyclic ISIs.

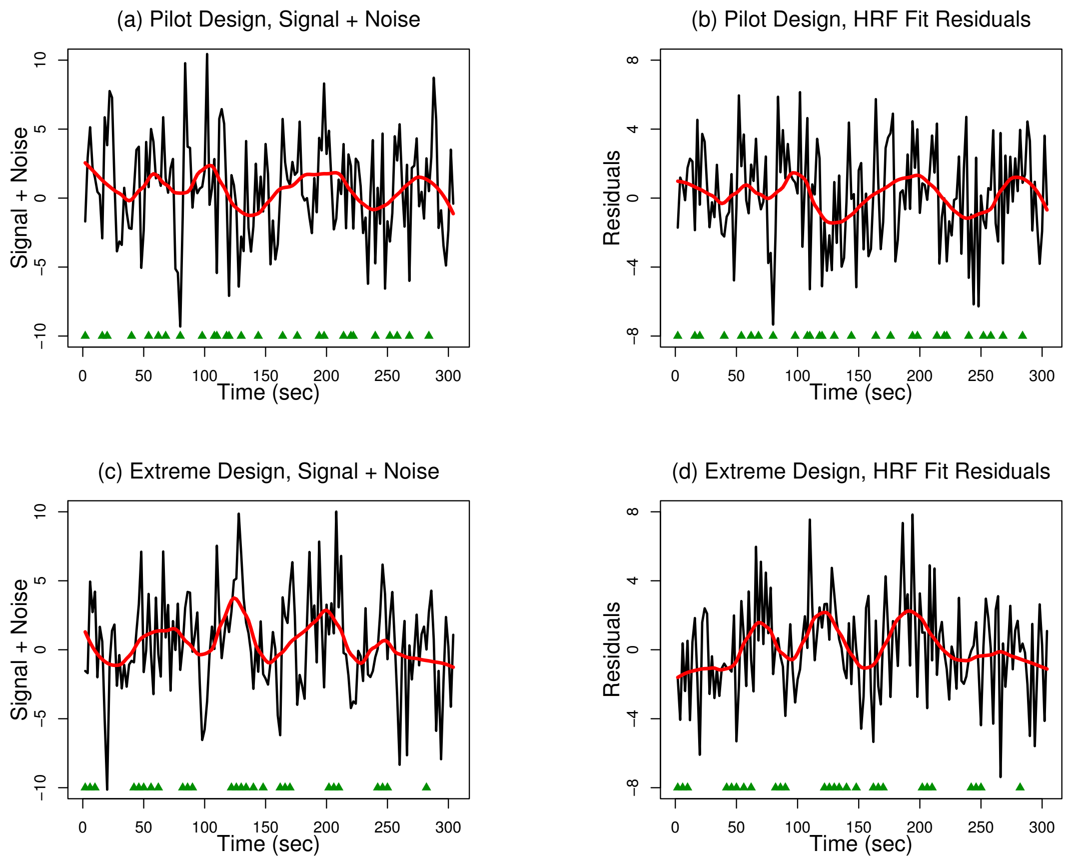

Fig. 6: Simulated two-gamma HRF + noise from the pilot design and an extreme design. Non-parametric loess curve fit in red. Stimulus activation times indicated by the green triangles.

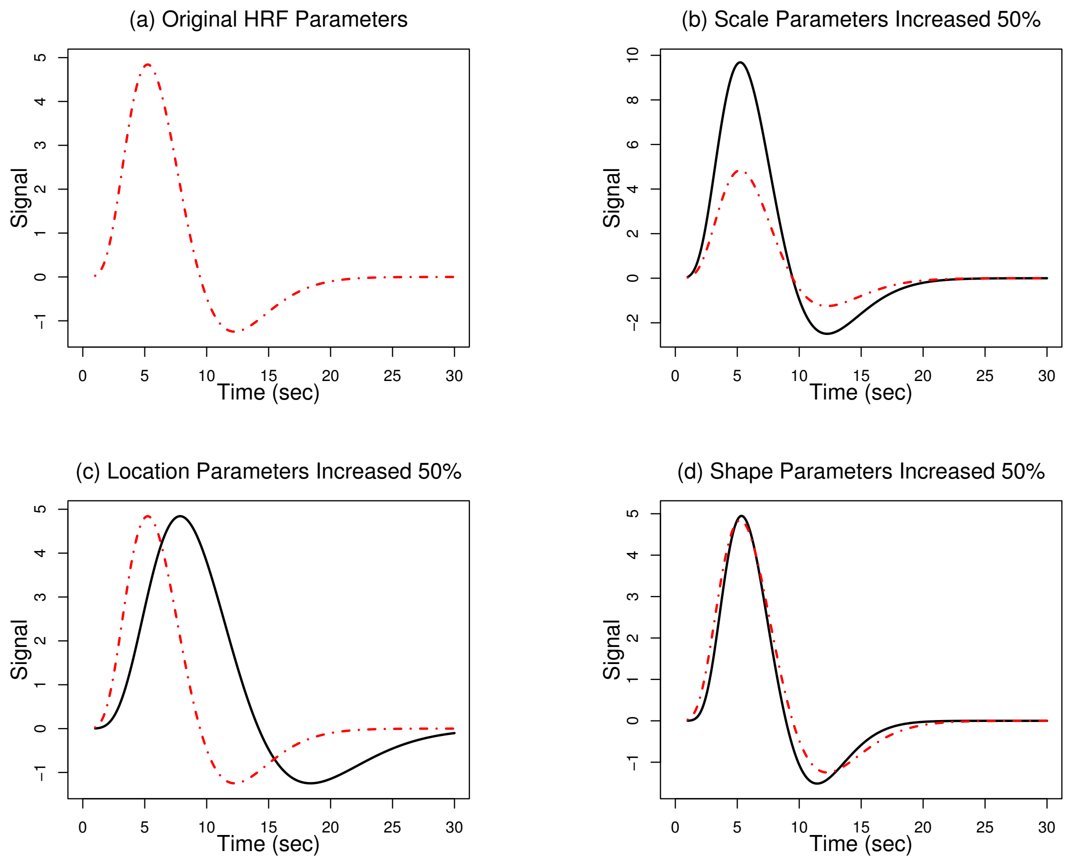

Fig. 7: Effects of changing the HRF parameter values by 50% with the original HRF (dashed) superimposed in (b)-(d).

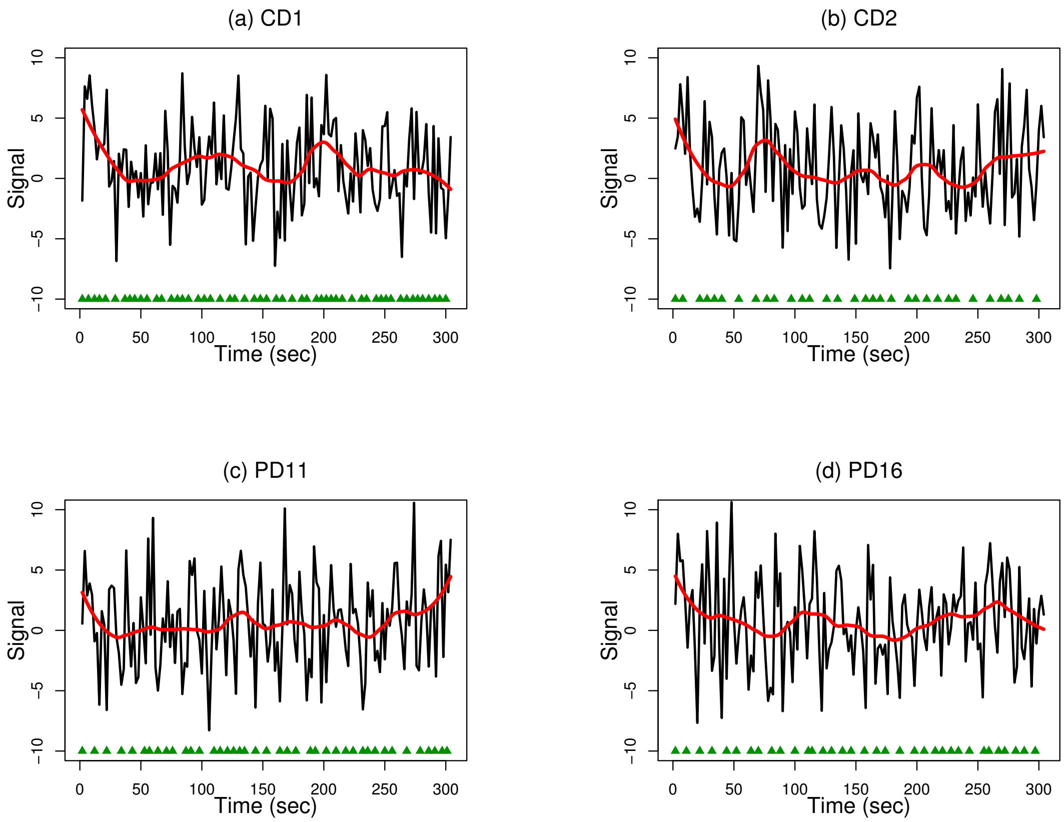

Fig. 8: Signal + Error for 4 designs, first set of random errors. Nonparametric loess curve fit in red. Stimulus times indicated by the green triangles.

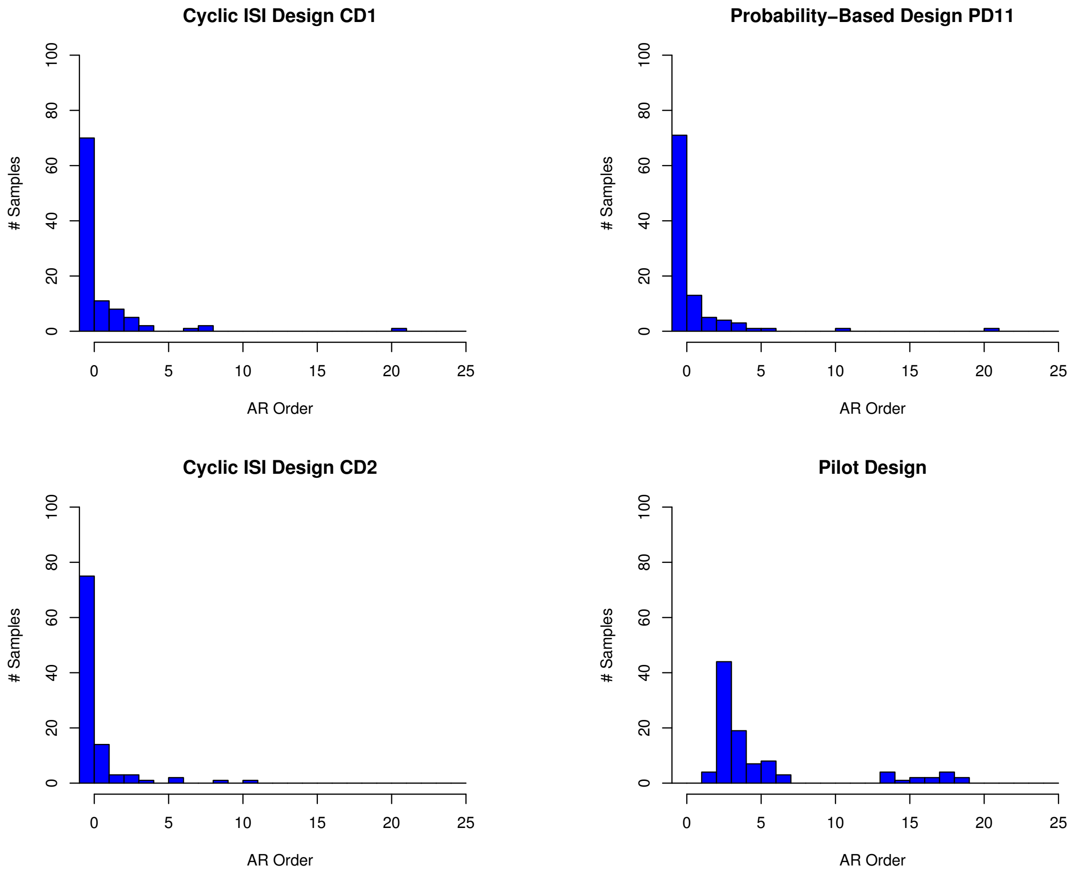

Fig. 9: Selected autoregressive orders from 100 simulations of 4 ISI designs. The same random errors were used for each sample in all of the designs.

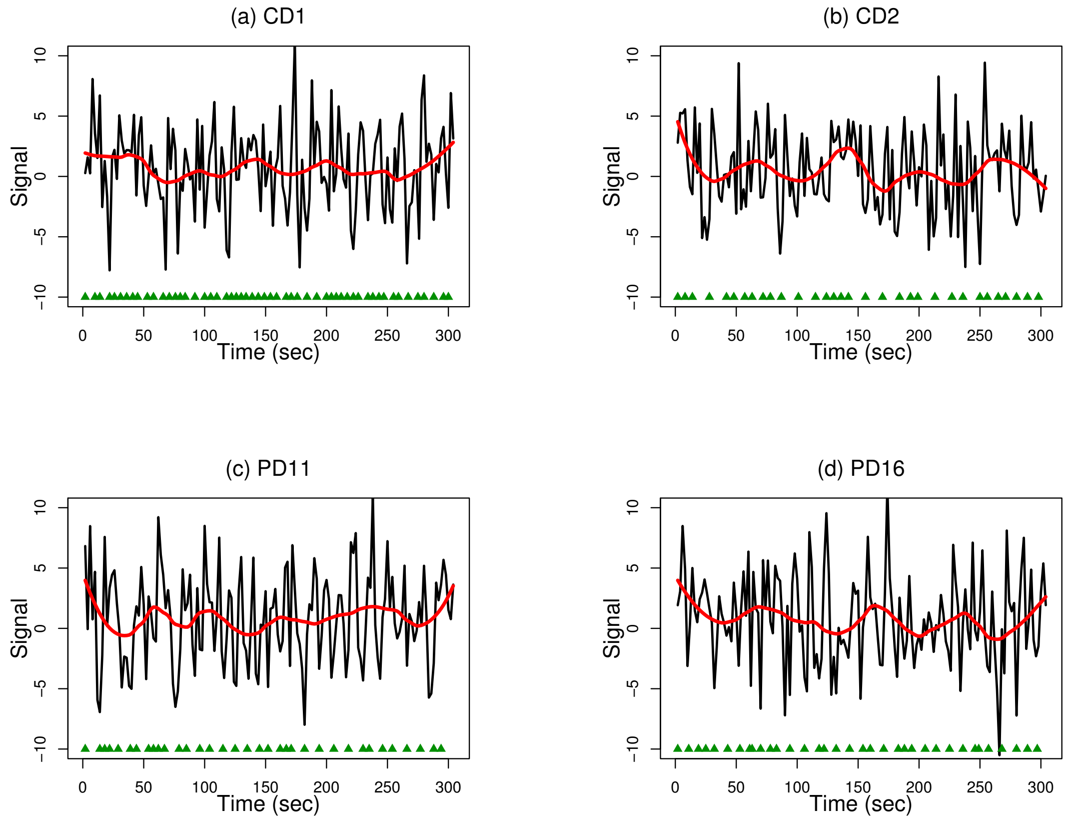

Fig. 10: Signal + error for 4 designs, second set of random errors. Nonparametric loess curve fit in red. Stimulus times indicated by the green triangles.