Journal of the Royal Statistical Society
SERIES C Applied Statistics

Appl. Statist. (2012)
61, *Part* 4, *pp.* 515–534

Constrained multiobjective designs for functional magnetic resonance imaging experiments via a modified non-dominated sorting genetic algorithm

Ming-Hung Kao

Arizona State University, Tempe, USA

and Abhyuday Mandal and John Stufken

University of Georgia, Athens, USA

[Received September 2010. Final revision October 2011]

Summary. Functional magnetic resonance imaging (MRI) is an advanced technology for studying brain functions. Owing to the complexity and high cost of functional MRI experiments, high quality multiobjective functional MRI designs are in great demand; they help to render precise statistical inference and are keys to the success of functional MRI experiments. Here, we propose an efficient approach for obtaining multiobjective functional MRI designs. In contrast with existing methods, the approach proposed does not require users to specify weights for the different objectives and can easily handle constraints to fulfil customized requirements. Moreover, the underlying statistical models that we consider are more general. We can thus obtain designs for cases where brief, long or varying stimulus durations are utilized. The usefulness of our approach is illustrated by using various experimental settings.

Keywords: Design efficiency; Genetic algorithms; Haemodynamic response function; Multiobjective optimization

1. Introduction

Functional magnetic resonance imaging (MRI) is one of the most dominant brain mapping techniques. This pioneering technology has many important clinical potentials such as early identification of Alzheimer's disease (Wierenga and Bondi, 2007) and preneurosurgical planning (Bookheimer, 2007); see also Brown (2007). In a functional MRI experiment, an MRI scanner is used to measure non-invasively cerebral haemodynamic changes following brain activity due to mental tasks or stimuli (Ogawa *et al.*, 1990). Specifically, the scanner acquires functional MRI signals from a three-dimensional grid of boxes that covers (some region of) the subject's brain; these boxes are called voxels and are of size, say, $3.15 \times 3.15 \times 5\,\text{mm}^3$ (see, for example, section 2.1.1. of Lazar (2008)). At each voxel, a time series consisting of functional MRI signals collected over time is obtained and is analysed to make statistical inference about brain functions. Two common statistical goals are

(a) detection of active brain voxels and
(b) the estimation of the haemodynamic response function (HRF), which is a function of time describing the fluctuations of the functional MRI signals evoked by one single stimulus.

Address for correspondence: Ming-Hung Kao, School of Mathematical and Statistical Sciences, Arizona State University, PO Box 871804, Tempe, AZ 85287, USA.
E-mail: mhkao@math.asu.edu

 0035–9254/12/61515

Detection allows the identification of brain regions activated by the stimuli. Estimation helps to understand the effect of a stimulus on the brain. Here, we aim at obtaining good designs that help to achieve both goals while taking into account practical constraints. Other relevant issues and overviews of functional MRI can be found in Lazar (2008) and Lindquist (2008).

Well-planned designs are crucial to the success of functional MRI experiments. However, obtaining high quality functional MRI designs is not an easy task. It requires careful consideration regarding study objectives, statistical models, psychological constraints and MRI machine settings. Taking into account these practical issues, existing approaches utilize discrete optimization techniques to search over the enormous design space for good designs. Wager and Nichols (2003) proposed a genetic algorithm framework that targets multiobjective designs achieving high statistical efficiencies in detection and estimation and avoiding possible psychological confounds such as anticipation and habituation. Their genetic algorithm has been applied in many studies over the last few years (e.g. Summerfield *et al.* (2006), Wang *et al.* (2007) and Rameson *et al.* (2010)). Following this work, Kao, Mandal, Lazar and Stufken (2009) developed a more efficient approach that takes advantage of current knowledge about functional MRI designs. Their approach largely improves the efficiency in searching for good functional MRI designs and was demonstrated to outperform previous methodologies.

Previous studies deal with the multiobjective nature of functional MRI experiments by using an objective function that is a weighted sum of criteria that target each objective separately. When weights are available, the genetic algorithm of Kao, Mandal, Lazar and Stufken (2009) is most efficient, and thus recommended. However, assigning weights can be arduous in practice (Deb, 2001; Ding *et al.*, 2004). One reason is that the mapping between the assigned weights and the performance of the resulting design under the various criteria is usually unclear. The design may fail to meet the experimenter's expectation (Marler and Arora, 2004, 2010). For example, assigning equal weights is common when seeking a design with equal efficiencies for all objectives. However, the result can be far from this requirement; see, for example, Cook and Wong (1994) and Section 4 of this paper.

In this paper, we consider an alternative approach for obtaining multiobjective designs when information about weights is vague. The idea is to obtain not only one, but a class of diverse, (near) Pareto optimal designs; a design is Pareto optimal if no other designs perform better in one or more objectives while being equivalent in all other objectives. The designs in this class approximate the Pareto frontier and, hence, offer optimal trade-offs between different objectives. Experimenters can then scrutinize the characteristics of the designs obtained and select one that suits their needs best. To achieve such a design class efficiently, we propose the non-dominated sorting genetic algorithm II (NSGA II) (Deb *et al.*, 2002), but with the improvement of including well-known functional MRI designs in the first step, just as in Kao Mandal, Lazar and Stufken (2009).

The approach proposed belongs to the class of '*a posteriori*' methods, in which a set of solutions is obtained for users to select from on the basis of their preferences (Miettinen, 1999). The weighted sum method can also be used to achieve this. In our context, this is done by repeatedly using the genetic algorithm of Kao, Mandal, Lazar and Stufken (2009) to obtain designs for systematically changed weights. Compared with this weighted sum method, our proposed approach requires much less central processing unit (CPU) time and achieves good designs. More importantly, the designs that we obtain have greater diversity in the objective space. This diversity facilitates design selection and is viewed as an important property to achieve (Zitzler *et al.* (2003) and references therein). Furthermore, our proposed approach can handle constraints easily and generates designs satisfying experimenters' requirements. By contrast, the weighted sum method is clumsy in this regard. We also compare the modified NSGA II approach with

the original approach. With a slightly increased CPU time, our approach achieves designs with better ability to detect brain activation.

Moreover, the underlying statistical models that we consider are more general than those used in previous studies. Our models accommodate cases considering brief stimuli (about 1 s), stimuli lasting several seconds or stimuli with durations varying across types. Such cases are quite common. For example, Murphy *et al.* (2006) used a 1-s flashing checkerboard to detect active voxels and to estimate the HRF, whereas Birn *et al.* (2004) and Brendel *et al.* (2010) considered stimuli of varying durations. Martin *et al.* (2005) employed 5-s pictures and detected activated brain regions on stroke patients with aphasia. The models that we consider also allow cases where event-related regressors are used to replace epoch regressors as advocated by Mechelli *et al.* (2003). With these important improvements, our approach provides a flexible and powerful tool for obtaining good functional MRI designs. A MATLAB program implementing this approach is available on request from the first author.

The remainder of the paper is organized as follows. In Section 2, we briefly introduce background information about functional MRI designs. We then present our proposed approach in Section 3, including the statistical models, design criteria and search algorithm. In Section 4, our proposed approach is compared with the weighted sum method and the NSGA II approach under various experimental settings. The paper closes with a brief discussion in Section 5.

2. Background and terminology

A functional MRI design is a sequence of mental stimuli (e.g. pictures or sounds) of one or more types interlaced with a control (e.g. rest or fixation). Such a sequence is presented to an experimental subject while an MRI machine scans the subject's brain to collect functional MRI time series from each brain voxel. When being presented, each stimulus may last from several milliseconds to a few seconds. Times between consecutive stimulus onsets may vary across stimuli; they are assumed to be multiples of a prespecified time which is termed the interstimulus interval in this paper. The control fills in the time when no stimulus is being presented. Such a functional MRI design is typically written as a finite sequence of finite numbers, e.g. $\mathbf{s} = \{101210 \ldots 1\}$. A number q ($\neq 0$) at the kth position of a design indicates an onset of a qth-type stimulus at time $(k-1)\tau_{\mathrm{ISI}}$, where τ_{ISI} is the specified interstimulus interval. A '0' means no stimulus onset at that time point.

Well-known functional MRI designs include block designs, m-sequences, random designs, mixed designs, permuted block designs and clustered m-sequences. A block design is a patterned sequence in which stimuli of the same type are clustered. For example, a block design with block size 4 may consist of repetitions of $\{111122220000\}$. An m-sequence can be generated from primitive polynomials for a Galois field (MacWilliams and Sloane, 1977; Godfrey, 1993; Buračas and Boynton, 2002). Although systematically generated, m-sequences look rather random with no perceivable patterns. Random designs are randomly generated and typically have no perceivable patterns. A mixed design is obtained from concatenating a fraction of a block design with a fraction of an m-sequence or a random design. Permuted block designs are generated by repeatedly exchanging elements of a block design to make the resulting design increasingly random. In contrast, a clustered m-sequence is generated by clustering (through permutation) stimuli of the same type in an m-sequence. See Liu (2004) for detailed descriptions of these designs.

Different designs are recommended for different purposes. Friston *et al.* (1999) demonstrated a high performance of the block design in detecting activations. For estimating the HRF, Dale

(1999) advocated the use of random designs, and Buračas and Boynton (2002) recommended *m*-sequences owing to their low auto-correlation property. Buxton *et al.* (2000), Liu *et al.* (2001), Liu and Frank (2004) and Liu (2004) further investigated efficiencies of designs with respect to these two dimensions. They observed that

(a) block designs have high efficiencies in detection, but have low estimation efficiencies,
(b) random designs and *m*-sequences are efficient for estimating the HRF, but they are not as good as block designs in detection, and
(c) designs containing both 'blocky' and 'random' components can reach good compromises between the two objectives;

such designs can be selected from mixed designs, permuted block designs and clustered *m*-sequences. Liu *et al.* (2001) and Liu and Frank (2004) concluded that there is a trade-off between efficiency for detection and efficiency for estimation. In addition, Liu and colleagues emphasized the importance of perceived randomness (or unpredictability) of designs. Patterned designs such as block designs are easy to predict by the experimental subject. They may give rise to psychological confounds such as anticipation and expectation; see also Dale (1999). Designs that are not patterned tend to avoid such confounds and are thus favourable.

Although study objectives are important considerations at the design stage, other experimental conditions such as time between consecutive MRI scans of the same voxel (or time to repetition), interstimulus interval and duration of the experiment are also key factors. These conditions can vary across experiments. To achieve tailor-made designs for each unique experiment, Wager and Nichols (2003) proposed a genetic algorithm framework to search for good designs with respect to user-specific experimental settings. Following this framework, Kao, Mandal, Lazar and Stufken (2009) took advantage of current knowledge about functional MRI designs to develop a more efficient approach. Their approach has also been adapted to cases involving multiple scanning sessions (Kao, Mandal and Stufken 2009), and studies where both stimulus effects and pairwise comparisons are of interest (Kao *et al.*, 2008). Maus *et al.* (2010a) also applied this approach for investigating the robustness of designs against misspecified temporal auto-correlation of functional MRI noise. They also used it to study the efficiency of *D*-optimal designs under the *A*-optimality criterion, and vice versa.

One difficulty of the previously mentioned genetic algorithms is that they require user-specific weights on study objectives for a multiobjective approach. The interpretation of these weights is almost always unclear, so there is almost never a meaningful choice. To tackle this issue, we propose an alternative approach. Our approach is described in Section 3 including the underlying statistical models, design criteria and a search algorithm; current knowledge about the mechanism behind the observed functional MRI response, which underpins the statistical models, is also briefly presented.

3. Methodology

3.1. Dual general linear models

In a functional MRI experiment, stimuli are presented to a subject in an order that is determined by a selected design. Each stimulus evokes neuronal firings at an activated brain voxel. Because of increased metabolic demands, oxygenated blood flows into the cerebral blood vessels around the active voxel and changes the ratio of oxyhaemoglobin to deoxyhaemoglobin. This change affects the local magnetic field, and leads to a fluctuation in the signal intensity that is collected by an MRI scanner; see also Cabeza and Kingstone (2006). It takes a few seconds for the MRI signal intensity to rise and decay. The process is sluggish and is typically described by an HRF.

When stimuli occur in close succession, the evoked HRFs overlap and accumulate to form the functional MRI time series, which also involves nuisance signals and noise.

Linear models are the most popular for analysing the functional MRI time series (Friston *et al.*, 1995; Worsley and Friston, 1995; Lindquist, 2008). These models commonly assume a linear time invariant system, in which each stimulus of the same type evokes the same HRF throughout the experiment, and overlapping HRFs accumulate additively. For detection problems, the HRF is typically assumed to be the product of an unknown amplitude parameter θ_q and an assumed HRF shape $\mathbf{h}_q^*$. Here, $q = 1, \ldots, Q$, and Q is the total number of types of stimulus. With these assumptions, the linear model that we consider for detection has the form

$$\mathbf{y} = \sum_{q=1}^{Q} \mathbf{X}_q \mathbf{h}_q^* \theta_q + \mathbf{S}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}. \tag{1}$$

Here, $\mathbf{y}$ is a $T \times 1$ vector representing the functional MRI time series of a voxel acquired every τ_{TR} s, where τ_{TR} is the time to repetition. $\mathbf{X}_q \mathbf{h}_q^* \theta_q$ represents the contribution to the accumulated HRFs evoked by the qth-type stimulus. $\mathbf{X}_q$ is a 0–1 design matrix corresponding to occurrences of the qth-type stimulus (see Appendix A for further details). $\mathbf{S}\boldsymbol{\gamma}$ is a nuisance term representing the drift or trend of $\mathbf{y}$ with a parameter vector $\boldsymbol{\gamma}$, and $\boldsymbol{\varepsilon}$ is noise. The noise is typically auto-correlated and is commonly assumed to follow an auto-regressive AR(1) process. Although other models may be considered, AR(1) seems to provide satisfactory results in many functional MRI studies; see also Bullmore *et al.* (1996), Worsley *et al.* (2002) and Zhang and Yu (2008).

The main focus of detection problems is on the amplitudes θ_q of model (1) or, more generally, their linear combinations $\mathbf{C}_{\boldsymbol{\theta}}\boldsymbol{\theta}$, where $\mathbf{C}_{\boldsymbol{\theta}}$ is a coefficient matrix for linear combinations of interest, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_Q)'$. For example, if $\mathbf{C}_\theta = (1, -1, 0, \ldots, 0)$, the focus is on the difference between the HRF amplitudes evoked by the first two types of stimulus.

Model (1) is general in that it accommodates many practical situations. For example, when the study involves only brief stimuli, we could choose $\mathbf{h}_q^*$ from the double-gamma function of the popular statistical parametric mapping software package for functional MRI analysis (`http://www.fil.ion.ucl.ac.uk/spm/`). Specifically, we would take the jth element of $\mathbf{h}_q^*$ as $h_{q,j}^* = g\{(j-1)\,\Delta T\}/\max_t\{g(t)\}$, where

$$g(t) = \begin{cases} \dfrac{t^5 \exp(-t)}{5!} - \dfrac{1}{6}\dfrac{t^{15}\exp(-t)}{15!}, & t \in [0, 32], \\ 0, & \text{otherwise}, \end{cases}$$

is a double-gamma function (Fig. 1(a)), ΔT is the greatest real value making both $\tau_{\mathrm{ISI}}/\Delta T$ and $\tau_{\mathrm{TR}}/\Delta T$ integers, $j = 1, 2, \ldots, 1 + \lfloor 32/\Delta T \rfloor$, and $\lfloor a \rfloor$ is the integer part of a.

When the stimulus lasts several seconds, $\mathbf{h}_q^*$ may be chosen as a convolution of a boxcar function and the function $g(t)$. In such a case, $h_{q,j}^* = h\{(j-1)\,\Delta T\}/\max_t\{h(t)\}$, where $h(t) = \int_0^t b(t-\tau)g(\tau)\,\mathrm{d}\tau$ is the convolution of $b(t)$ and $g(t)$,

$$b(t) = \begin{cases} 1, & t \in [0, \tau_{\mathrm{dur}}], \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

is a boxcar function for a stimulus with duration τ_{dur}, and $j = 1, \ldots, 1 + \lfloor (32 + \tau_{\mathrm{dur}})/\Delta T \rfloor$. Boxcar functions of 3, 6 and 12 s are presented in Fig. 1(b); their convolutions with $g(t)$, normalized to have a maximum of 1, are presented in Fig. 1(c). We note that τ_{dur} should ideally be determined by the duration of the underlying neuronal activity evoked by the stimulus. However, this duration may not be known; see also Loh *et al.* (2008). Here, we follow a common practice to set τ_{dur} to the duration of stimulus, and we assume that the neuronal activity and stimulus

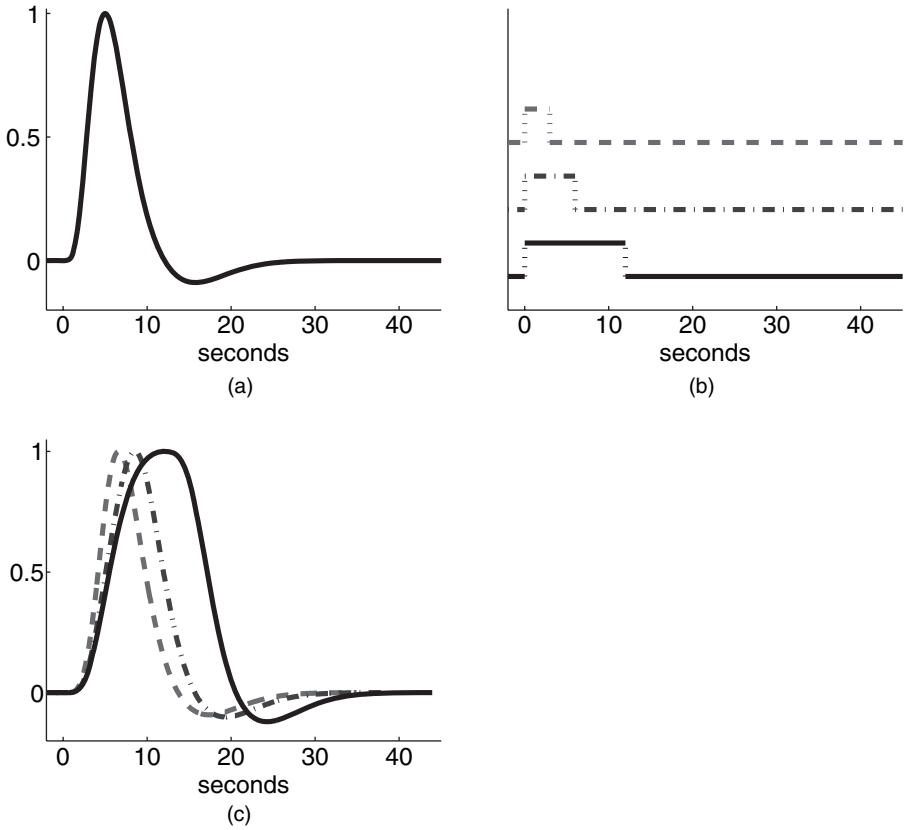

Fig. 1. (a) Function $g(t)$, normalized to have a maximum of 1, (b) boxcar functions of 3 s (– – –), 6 s (·–·–·) and 12 s (———), and (c) convolutions, normalized to have a maximum of 1, of the function $g(t)$ and the boxcar functions (– – –, 3 s; ·–·–·, 6 s; ———, 12 s)

durations are the same. In Section 4, we investigate the robustness of designs when these two durations do not match, i.e. τ_{dur} is misspecified.

Model (1) can also be considered when one chooses to follow Mechelli *et al.* (2003) to consider event-related regressors in lieu of epoch-related regressors. This can be done by replacing the previously described boxcar function $b(t)$ by a spike train $\mathbb{1}_\Gamma(t)$ with, say, $\Gamma = \{0, 3, 6, \ldots, 3\lfloor \tau_{\text{dur}}/3 \rfloor\}$; here, $\mathbb{1}_\Gamma(t)$ is 1 for $t \in \Gamma$ and 0 otherwise. We also note that $\mathbf{h}_q^*$ can be different across types of stimulus. This allows varying stimulus durations across types. Although the double-gamma function is popular, other HRF shapes can also be considered.

For estimating the HRF, the following linear model is considered (see, also, Dale (1999)):

$$\mathbf{y} = \sum_{q=1}^{Q} \mathbf{X}_q \mathbf{h}_q + \mathbf{S}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}. \tag{3}$$

Here, $\mathbf{h}_q$ is a vector of parameters describing the HRF evoked by a qth-type stimulus; the jth element of $\mathbf{h}_q$ represents the height of the HRF at time $(j-1)\Delta T$ following a stimulus onset. All other terms are defined under model (1). The focus here is on estimating $\mathbf{C}_h\mathbf{h}$; $\mathbf{C}_h$ is a linear combination matrix, and $\mathbf{h} = (\mathbf{h}_1', \ldots, \mathbf{h}_Q')'$. When $\mathbf{C}_h = (\mathbf{I}_k, -\mathbf{I}_k, \mathbf{0}, \ldots, \mathbf{0})$, the interest lies in the difference between the HRFs incurred by the first two types of stimulus; $\mathbf{I}_k$ is the $k \times k$ identity

matrix, and k is the length of the vector $\mathbf{h}_q$, which is set to be the same as that of $\mathbf{h}_q^*$ in model (1).

3.2. Optimal design criteria

Following Kiefer (1959), the performance of a design for detecting activation is evaluated by some function of the covariance matrix $\text{cov}(\mathbf{C}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}})$ of the generalized least squares estimator, $\mathbf{C}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}$. Similarly, the estimation efficiency is measured by some function of $\text{cov}(\mathbf{C}_h\hat{\mathbf{h}})$, where $\mathbf{C}_h\hat{\mathbf{h}}$ is the estimator of $\mathbf{C}_h\mathbf{h}$. Specifically, with models (1) and (3), these covariance matrices are

$$\text{cov}(\mathbf{C}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}) = \mathbf{C}_{\boldsymbol{\theta}}(\mathbf{H}'\mathbf{X}'\mathbf{V}'(\mathbf{I}_T - \mathbf{P}_{\mathbf{VS}})\mathbf{VXH})^{-}\mathbf{C}_{\boldsymbol{\theta}}';$$
$$\text{cov}(\mathbf{C}_h\hat{\mathbf{h}}) = \mathbf{C}_h(\mathbf{X}'\mathbf{V}'(\mathbf{I}_T - \mathbf{P}_{\mathbf{VS}})\mathbf{VX})^{-}\mathbf{C}_h'.$$

Here, $\mathbf{H}$ is a diagonal matrix with the vectors $\mathbf{h}_1^*, \ldots, \mathbf{h}_Q^*$ along the diagonal, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_Q)$ and $\mathbf{V}$ is an assumed whitening matrix so that $\mathbf{V}\boldsymbol{\varepsilon}$ is white noise. $\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ is the orthogonal projection matrix onto the column space of $\mathbf{A}$, and $\mathbf{A}^{-}$ is a generalized inverse of $\mathbf{A}$.

We consider the A-optimality criterion in our case-studies. Designs with smaller average variances for the estimators of the linear combinations of the parameters of interest are said to be (A) better (Bailey, 2007). For a design $\mathbf{s}$, the A-efficiency for detection is defined as

$$F_d^*(\mathbf{s}) = \frac{F_d(\mathbf{s})}{F_d(\mathbf{s}_d^*)},$$

where

$$F_d(\mathbf{s}) = \text{tr}\{\text{cov}(\mathbf{C}_{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}})\}^{-1},$$

and $\mathbf{s}_d^*$ is a design maximizing F_d. We use the genetic algorithm of Kao, Mandal, Lazar and Stufken (2009) to maximize F_d, and to obtain such a design. For estimation, the A-efficiency of $\mathbf{s}$ is

$$F_e^*(\mathbf{s}) = \frac{F_e(\mathbf{s})}{F_e(\mathbf{s}_e^*)},$$

where

$$F_e(\mathbf{s}) = \text{tr}\{\text{cov}(\mathbf{C}_h\hat{\mathbf{h}})\}^{-1}.$$

The design $\mathbf{s}_e^*$ is a design maximizing F_e and can be approximated via the genetic algorithm of Kao, Mandal, Lazar and Stufken (2009). Although we present results for A-optimality, our approach can also accommodate other optimality criteria, such as D-optimality. Note that the design criteria that are presented here are for a subset of the model parameters. Such criteria are sometimes termed A_s- (or D_s-)optimality; see also Atkinson *et al.* (2007). For clarity, we omit the subscript and use the term A- (or D-)optimality in this paper.

In addition to statistical efficiencies, experimenters may want to use a desired frequency for each type of stimulus for practical reasons (see Section 4.3). We take this requirement into account by considering the F_f^*-criterion of Kao, Mandal, Lazar and Stufken (2009):

$$F_f^*(\mathbf{s}) = 1 - \frac{F_f(\mathbf{s})}{F_f(\mathbf{s}^0)},$$

where

$$F_f(\mathbf{s}) = \sum_{i=1}^{Q} \lfloor |n_i - nP_i| \rfloor, \tag{4}$$

n_i is the number of the type i stimulus in the subdesign of $\mathbf{s}$ excluding all zeros, n is the length of the subdesign, P_i is the desired proportion of the ith stimulus type, $\mathbf{s}^0$ is the design containing only the stimulus type i with the smallest P_i and $\lfloor |a| \rfloor$ is the greatest integer less than or equal to the absolute value of a. To fulfil the experimenter's requirement, we regard '$F_f^* \geqslant c_f$' as a constraint for a given c_f while optimizing (F_d^*, F_e^*). In Section 3.3, we present an efficient algorithm to achieve this goal.

3.3. Search algorithm

Our proposed algorithm is built on the NSGA II algorithm of Deb *et al.* (2002) with a modification in the first step of the algorithm. Specifically, we follow Kao, Mandal, Lazar and Stufken (2009) to include well-known functional MRI designs in the first step of the algorithm to facilitate the search for good functional MRI designs. The algorithm proposed mimics Darwin's theory of evolution to move through generations. Good parents are selected to reproduce offsprings and, with the survival of the fittest principle, individuals of better fit survive to the next generation. The process, when repeated, ensures preservation of good traits and high quality designs can be expected.

To determine the fitness of designs with respect to multiple objectives, two measures, namely non-domination rank and crowding distance, are considered. These measures are functions of design efficiencies (i.e. F_i^*-values). Non-domination ranks are assigned to designs within each generation of the algorithm. The first rank is assigned to designs that are not dominated by any other designs in the same generation; a design $\mathbf{s}_1$ is said to be dominated by another design $\mathbf{s}_2$ if $F_i^*(\mathbf{s}_1) \leqslant F_i^*(\mathbf{s}_2)$ for both $i = d$ and $i = e$ and $F_i^*(\mathbf{s}_1) < F_i^*(\mathbf{s}_2)$ for $i = d$ or $i = e$ or both. Designs of the second non-domination rank are dominated by one or more first-ranked designs, but not by others. The subsequent ranks are assigned accordingly. When comparing designs, this measure is of the primary concern; it helps to move towards Pareto optimality. An efficient way for rank assignments can be found in Deb *et al.* (2002) and is not repeated here.

When the constraint $F_f^* \geqslant c_f$ is used, the 'constrained domination' of Deb *et al.* (2002) is utilized for assigning non-domination ranks. A design $\mathbf{s}_1$ is said to be constrained dominated by $\mathbf{s}_2$ if

(a) both designs satisfy the constraint, and $\mathbf{s}_1$ is dominated by $\mathbf{s}_2$, or
(b) at least one $\mathbf{s}_i$ fails to satisfy the constraint and $F_f^*(\mathbf{s}_1) < F_f^*(\mathbf{s}_2)$.

We note that the constrained case is equivalent to the unconstrained case if $c_f = 0$.

As the secondary measure, the crowding distance is used to compare designs of the same non-domination rank. It helps to maintain diversity in the objective space in terms of design efficiencies (F_d^* and F_e^*). Designs with small crowding distances are close to their neighbours, and we would like to leave them out. To obtain the crowding distance, we first use F_d^* to sort the designs. If a design $\mathbf{s}$ has the smallest or the largest F_d^*, we follow Deb *et al.* (2002) to set $d_d(\mathbf{s}) = \infty$ to avoid leaving out the designs on the boundaries. Otherwise, $d_d(\mathbf{s})$ is the absolute difference between the F_d^*-values of the first neighbours of $\mathbf{s}$. Repeat the procedure again by using F_e^* to obtain $d_e(\mathbf{s})$. The crowding distance of $\mathbf{s}$ is $d_d(\mathbf{s}) + d_e(\mathbf{s})$; see also Deb *et al.* (2002).

With these two measures, we describe the proposed algorithm below.

Step 1: *initials (first generation)*—obtain $2G$ initial designs to form the first generation, where G is an even number. The designs that we use include block designs, m-sequences, random designs, a design maximizing F_d, a design maximizing F_e and combinations of the last two. Calculate design efficiencies for these designs.

Step 2: *fitness and mating pool*—compute the non-domination ranks and crowding distances for the designs in the current generation. On the basis of these two measures, select the G best designs from the current generation to form a mating pool.
Step 3: *stopping rule*—if a given number of generations is reached, terminate the search and report the G designs in the mating pool. Otherwise, continue to the next step.
Step 4: *selecting parents*—obtain $G/2$ pairs of parents from the mating pool via a tournament selection. Specifically, four distinct designs are randomly selected from the mating pool to form two pairs of designs. Choose the better one from each of the two design pairs; the two winners continue as a pair of parents. This process is then repeated $G/2$ times to form $G/2$ pairs of parents, each time starting with the selection of four designs from the original mating pool of G designs.
Step 5: *reproduction*—use the paired parents to generate offspring via crossover and mutation; each pair gives birth to two offspring designs. The crossover operator exchanges the corresponding fractions of each pair of designs on the basis of a randomly selected cut point. The mutation operator then perturbs a randomly selected portion α_m of elements of the G resulting designs. The selected elements are replaced by new elements generated from the discrete uniform distribution $\mathcal{U}\{0, 1, \ldots, Q\}$.
Step 6: *next generation*—obtain the design efficiencies of the G offspring designs. The G offspring along with the G designs in the mating pool form the next generation. Go back to step 2 and repeat the process until the stopping rule in step 3 is met.

We follow Kao, Mandal, Lazar and Stufken (2009) to include well-known functional MRI designs in the initial generation. These designs provide 'building blocks' to facilitate the search. They can be generated systematically but are not easy to achieve via random mechanisms (crossover and mutation). We also include designs maximizing F_d and F_e, which are required for calculating F_d^* and F_e^* respectively. Including them does not significantly increase the computation time and is helpful. When implementing the algorithm, we set G to 100 and α_m to 1%. The search is terminated after 2500 generations. Although these algorithmic parameters may not be optimal, the achieved designs are satisfactory as presented in Section 4.

We note that the NSGA II algorithm has a different procedure in the first iteration. In that algorithm, G designs are first randomly generated to produce another G designs via tournament selection, crossover and mutation. The $2G$ designs are combined to form 'initial designs' in step 1. Other steps are then implemented to generate subsequent generations; see also Deb *et al.* (2002). When implementing this algorithm, we also set $G = 100$, and $\alpha_m = 1\%$, and terminate the search after 2500 generations.

4. Case-studies

In this section, we use our proposed approach to generate multiobjective functional MRI designs for various experimental settings. The results are compared with those of the weighted sum method and the NSGA II approach in terms of the diversity of the obtained designs in the objective space, achieved trade-offs and CPU time spent. In addition, designs are selected from the obtained design classes by using various selection criteria. These selected designs are compared to provide information about the ability of the three approaches in fulfilling the experimenter's needs. We also study the robustness of designs to misspecified τ_{dur} in equation (2).

Moreover, our proposed approach is applied to obtain designs for the experiment of Brendel *et al.* (2010), who used a design obtained through random permutations. Although such designs are quite common in practice, we demonstrate that our designs can significantly outperform them.

4.1. Multiobjective designs for brief stimuli

We first consider an experimental setting that was also studied by Kao, Mandal, Lazar and Stufken (2009). The number of stimulus types is 3 ($Q=3$) and all stimuli have a short duration. The length of the design sequence is 255. The interstimulus interval is 2 s and so is the time to repetition. Two study objectives of interest include detecting activation and estimating the HRF, and the linear combination matrices $\mathbf{C}_{\boldsymbol{\theta}}$ and $\mathbf{C}_h$ are identity matrices. The response is assumed to have a second-order polynomial drift. The noise follows an AR(1) process with a correlation coefficient $\rho=0.3$. On the basis of the results of Maus *et al.* (2010a), A-optimal designs obtained with $\rho=0.3$ are quite robust against other $\rho\in[0,0.5]$. Alternatively, if a prior estimate of ρ is available, one could use that value; see also Wager and Nichols (2003). For $q=1,2,3$, the HRF shape $\mathbf{h}_q^*$ of model (1) is the normalized $g(t)$ that was discussed in Section 3.1.

Fig. 2(a) presents the F_e^*- against F_d^*-values of the 21 designs obtained by the weighted sum method under this scenario. These designs are achieved by repeatedly implementing the MATLAB program of Kao (2009) to maximize $F=wF_d^*+(1-w)F_e^*$ with w increased from 0 to 1 in steps of 0.05. Except for the stopping rule, the algorithmic parameters of the MATLAB program, including the population size (20), mutation rate (1%) and number of immigrants (4), are set to their default values (see Table 1 of Kao (2009)). The stopping rule that we choose is also built in the program and it terminates the search if there is no significant improvement in the F-value. In our experience, this stopping rule can save CPU time without sacrificing much in the efficiency achieved. Note that the mesh size of 0.05 was also considered in Kao, Mandal, Lazar and Stufken (2009).

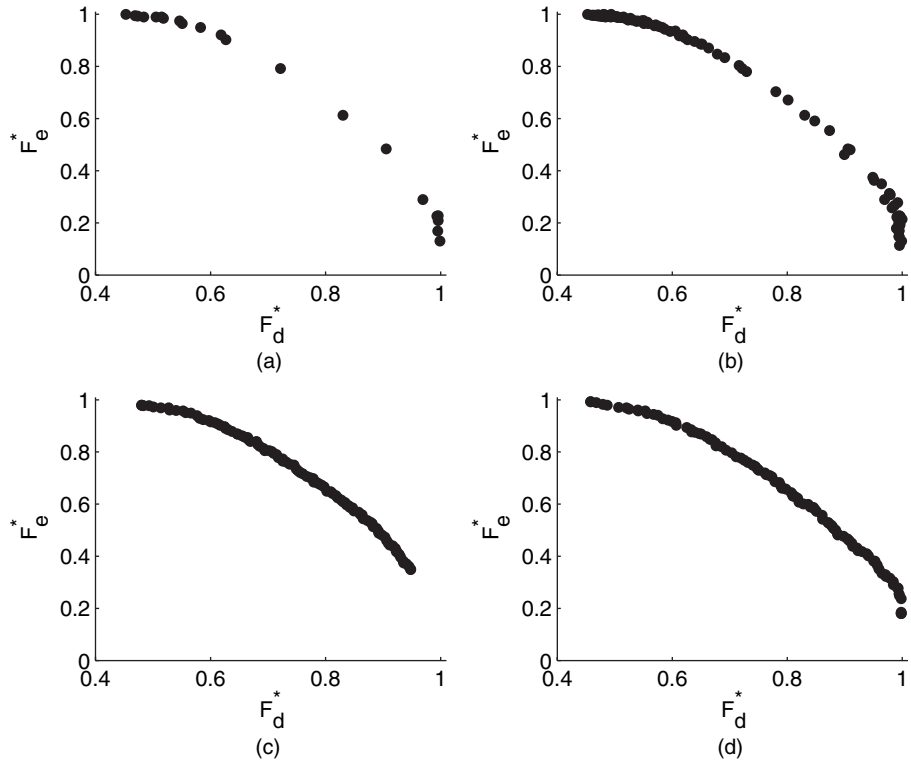

Fig. 2. F_e^*- against F_d^*-values of the designs obtained from (a) the weighted sum method with mesh size 0.05 (21 designs), (b) the weighted sum method with mesh size 0.01 (101 designs), (c) the NSGA II approach (100 designs) and (d) our approach (100 designs)

A disadvantage of the weighted sum method is observed in Fig. 2(a)—the designs are clustered at the two ends of the approximate Pareto frontier and are sparse in the middle. Users do not have many choices if designs with intermediate efficiencies are desired. The use of a smaller mesh size of 0.01 slightly improves the situation. However, clusters and gaps are still observed (see Fig. 2(b)). Our approach and the NSGA II approach yield designs with better diversity in the objective space. As shown in Figs 2(c) and 2(d), no large distances between consecutive designs are observed, and there are designs almost everywhere on the approximate Pareto frontier. Experimenters can select a suitable design on the basis of their preferences for a trade-off between the different objectives.

Figs 2(c) and 2(d) also reveal that the maximal F_d^*-value that is achieved by the NSGA II approach is slightly smaller than that attained by our approach. This is better manifested in Fig. 3 where Figs 2(b)–2(d) are overlaid; note that the design class in Fig. 2(a) is a subset of that in Fig. 2(b). As shown in Fig. 3, the NSGA II approach does not provide any design with $F_d^* > 0.948$. With the help of good initial designs, our approach is free from this drawback. Fig. 3 also indicates that the weighted sum method can achieve designs with slightly better trade-offs; these designs are located slightly to the upper right of some designs of the other two approaches. Although the weighted sum approach is slightly advantageous in this regard, it requires much more CPU time as also shown in Fig. 3.

The CPU times that are presented in Fig. 3 are obtained from implementing the different methods on a desktop computer with a 3.0 GHz Intel Pentium 4 quad-core processor. The weighted sum method is the most time consuming. With the mesh size of 0.05, this method takes about 19 min to obtain 21 designs. About 98 min are required for the weighted sum

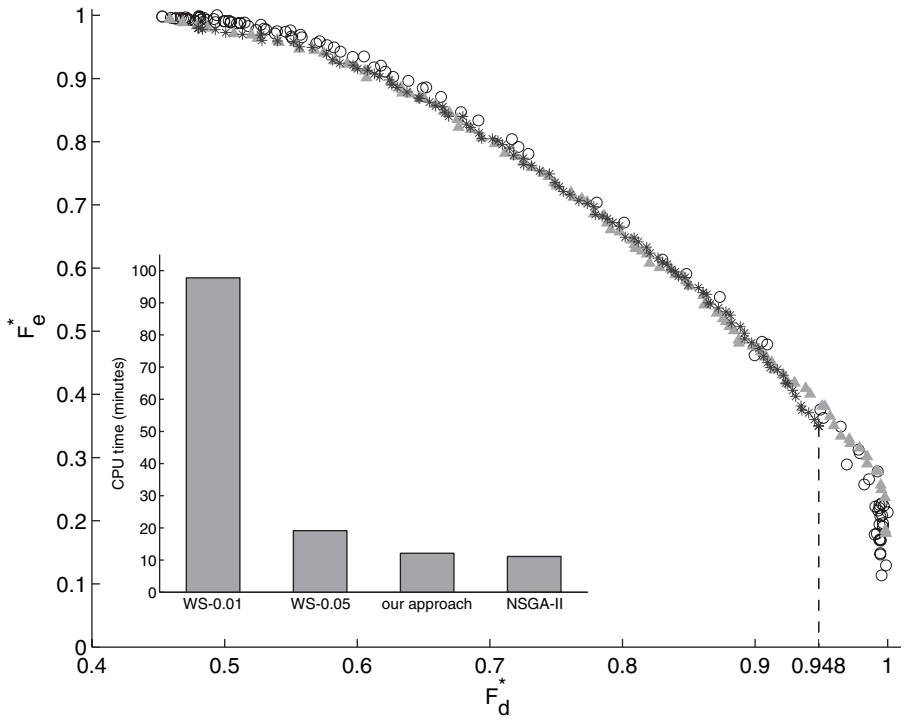

Fig. 3. F_e^*- against F_d^*-values of the designs of the various approaches and CPU times spent for obtaining these designs (WS-0.05 and WS-0.01 represent the weighted sum methods with mesh sizes 0.05 and 0.01 respectively): ○, weighted sum; ▲, our approach; ∗, NSGA II approach; ⋮, reference line

method to achieve 101 designs. The NSGA II algorithm spends about 11 min and our approach uses about 12 min to obtain 100 designs. The last two approaches are more efficient.

In addition to the above comparisons, it is also of interest to see whether the three approaches can fulfil the experimenter's needs in design selection. One possible selection criterion is to find a design with balanced efficiencies ($F_d^* = F_e^*$). Such a design works equally well, relative to the best designs, in both dimensions. With this selection criterion, a design with $(F_d^*, F_e^*) = (0.744, 0.745)$ can be selected from the 100 designs of our approach. The NSGA II approach achieves a design with $(F_d^*, F_e^*) = (0.745, 0.749)$. The weighted sum method yields $(F_d^*, F_e^*) = (0.729, 0.781)$, which is achieved by using the mesh size of 0.01, and the objective function $F = 0.62F_d^* + 0.38F_e^*$. Our approach and the NSGA II approach produce designs with closer F_d^*- and F_e^*-values and better satisfy the selection criterion. We also note that, when equal weights are assigned ($w = 0.5$), the weighted sum method does not yield a design with equal efficiencies for detection and estimation but results in an (F_d^*, F_e^*) value of $(0.618, 0.92)$.

Selecting constrained designs that maximize F_d^* (or F_e^*) subject to $F_e^* \geqslant c$ (or $F_d^* \geqslant c$) for a given c is also quite common. Fig. 4 presents the $\max(F_i^*)$ that is achieved by different approaches when $F_j^* \geqslant 0.75, 0.85, 0.95$. The weighted sum method with mesh size 0.05 consistently performs the worst, and the use of a finer mesh size improves the result. Although our approach and the NSGA II algorithm use much less CPU time, Fig. 4 demonstrates that these two methods achieve designs that are better or not much worse than those of the weighted sum method. Although the NSGA II approach performs similarly to our approach in most cases, it fails to produce designs with $F_d^* \geqslant 0.95$ as indicated in the rightmost group of Fig. 4(b). Our approach is thus recommended.

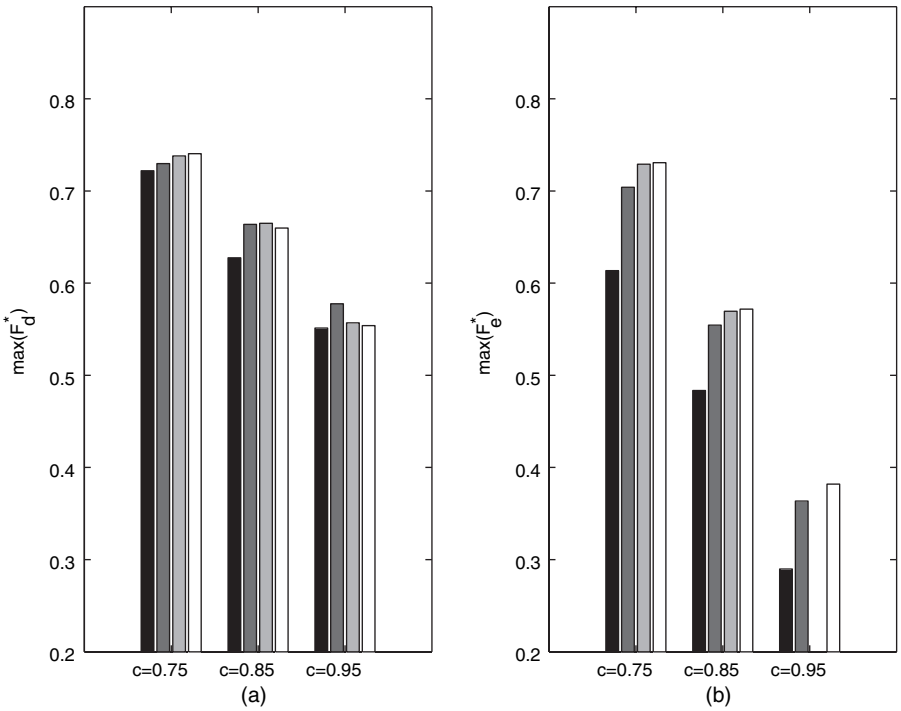

Fig. 4. (a) Maximum F_d^* achieved when $F_e^* \geqslant c$ and (b) maximum F_e^* achieved when $F_d^* \geqslant c$, $c = 0.75, 0.85, 0.95$: ■, weighted sum method, mesh size 0.05; ■, weighted sum method, mesh size 0.01; ■, NSGA II approach; □, our approach

4.2. Multiobjective designs for longer durations of stimulus

For the second setting, stimuli that last several seconds are considered. We start with the case of two types ($Q = 2$) of 4-s stimuli. The length of the design is 242. The interstimulus interval is 4 s and the time to repetition is 2 s. The HRF shape is the convolution of a 4-s boxcar function with the normalized $g(t)$; see Section 3.1. Other conditions are the same as for the experimental setting in Section 4.1.

Comparisons between our approach, the NSGA II approach and weighted sum method are again conducted. The results convey information similar to that in Section 4.1 and have thus been omitted. Here, we focus on another important issue, namely the robustness of designs to misspecification of the duration τ_{dur} of the boxcar function $b(t)$ in equation (2). This misspecification may occur when there is a difference between the duration of stimulus, which is commonly used to specify τ_{dur}, and the duration of the neuronal activity evoked, which is almost always uncertain and may vary across brain voxels. For example, in an experiment involving tasks requiring memory or decision making, the neuronal activity might last longer than the stimulus presentation. In other instances, the neuronal activity that is evoked by a simple task might not last as long as the presentation duration of the task. Our results suggest that our obtained designs remain efficient if the difference between the misspecified neural activity duration and the real duration is no more than 2 s, especially when the interstimulus interval is 6 s or less. A loss of efficiency is observed when the misspecified duration is more than 4 s apart from the real duration. The amount of loss in the design efficiency increases with the interstimulus interval.

Fig. 5 presents the performance of two sets of designs, namely designs for brief stimuli and those for 2-s stimuli ($\tau_{\text{dur}} = 2$) when the interstimulus interval is 4 s. In Fig. 5(a), we assume that the neuronal activity is brief; the τ_{dur} of 2 that was used to generate the second set of designs is thus misspecified, and the first set of designs is suited to this particular case. As shown in Fig. 5, the Pareto fronts approximated by the two design sets are nearly indistinguishable. This indicates that the designs that were obtained for 2-s stimuli are very efficient when the neuronal activity is brief. In Fig. 5(b), we assume that the actual neuronal activity duration is 2 s. Again, the two designs sets perform similarly in this latter case. The designs generated for brief stimuli are therefore very efficient when the actual τ_{dur} is 2 s.

To accommodate stimuli with duration longer than 2 s, we increase the interstimulus interval to 6 s and 8 s. The F_e^*- against F_d^*-values of designs generated with 6-s interstimulus interval are presented in Figs 6(a) and 6(b), and those with 8-s interstimulus interval are in Figs 6(c) and

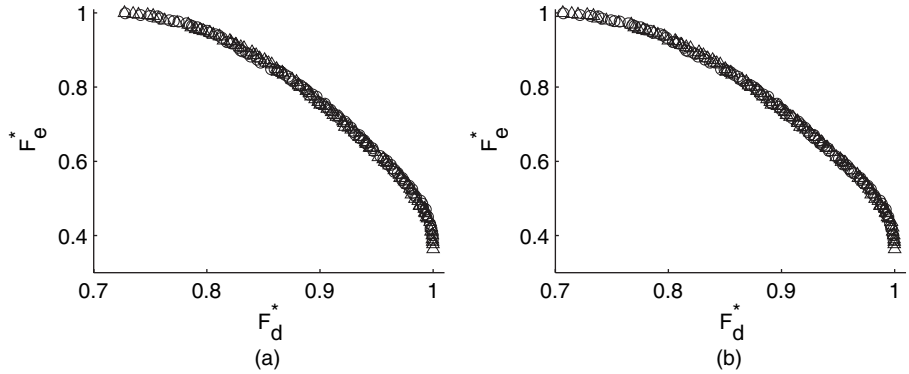

Fig. 5. F_e^*- against F_d^*-values of designs when the interstimulus interval is 4 s and the actual neuronal activity duration is (a) brief or (b) 2 s: ○, brief design; △, 2-s design

6(d). As shown in Fig. 6(a), designs for 4-s stimuli yield F_d^*- and F_e^*-values that are similar to those of designs for brief stimuli when the actual neuronal activity duration is brief. In contrast, Fig. 6(b) indicates that designs for brief stimuli remain efficient when the actual neuronal activity duration is 4 s. Similar results are observed when the interstimulus interval is 8 s. The 4-s stimulus designs perform quite well for brief neuronal activity; see Fig. 6(c). We also observe that the designs for brief stimuli and those for 6-s stimuli are quite efficient when the actual neuronal activity is 4 s (not shown). Fig. 6(c) also suggests that designs for 6-s stimuli can suffer a loss of efficiency if the neuronal activity that is evoked by the stimulus is brief. Fig. 6(d) indicates that the brief stimulus designs do not perform well when the underlying neuronal activity that is evoked by each stimulus lasts 6 s. However, the 4-s stimulus designs achieve a good performance in this last scenario in Fig. 6(d). Designs for 2-s stimuli have also been studied and are observed to perform similarly to the brief stimulus designs in all cases presented in Fig. 6. For clarity, these 2-s stimulus designs have been omitted from Fig. 6.

4.3. Constrained multiobjective designs for varying durations of stimulus

Following Brendel *et al.* (2010), we consider here three types of stimulus with durations 6, 3.6 and 2 s. Specifically, the three types of stimulus are

(a) a 2-s preparatory auditory tone signal plus a 4-s, 2.5-Hz isochronous click train,

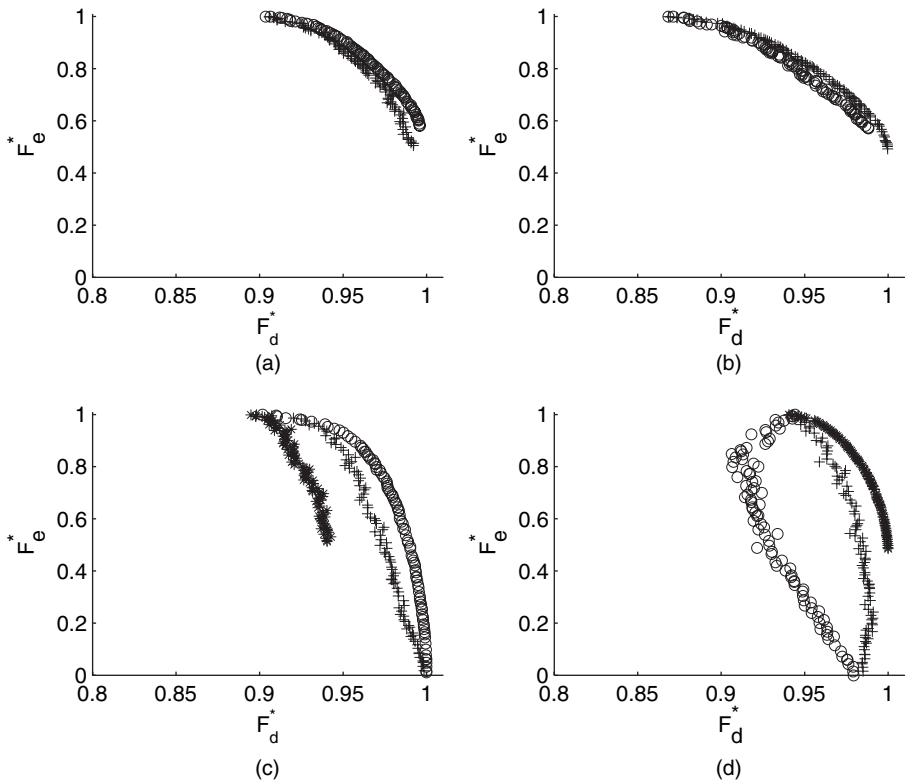

Fig. 6. F_e^*- against F_d^*-values of designs when (a) the actual neuronal activity duration is brief and the interstimulus interval is 6 s, (b) the actual neuronal activity duration is 4 s and the interstimulus interval is 6 s, (c) the actual neuronal activity duration is brief and the interstimulus interval is 8 s and (d) the actual neuronal activity duration is 6 s and the interstimulus interval is 8 s: ○, brief design; +, 4-s design; ∗, 6-s design

(b) a 2-s preparatory auditory tone signal plus a 1.6-s 2.5-Hz isochronous click train and
(c) a 2-s preparatory auditory tone signal alone.

They are respectively termed the long click train, short click train and no-click train. The interstimulus interval is set to 9 s and the design length to 67; it takes about 10 min to present the entire design. The MRI scanner scans each voxel every 1.5 s to collect functional MRI time series for detecting active brain voxels and estimating the HRFs that are evoked by the three types of stimulus.

Each click train involves clicks of sounds, namely strokes of a pen against a desk. During the experiment, the subject is asked to synchronize the syllable /ta/ to the clicks as closely as possible in time. The experimenters decide to include more long click trains, and expect that, with a higher number of this type of stimulus, the subject will be more engaged throughout the experiment. The number of occurrences of long click trains is about three times as many as the other two types of stimulus. Therefore, in the notation of equation (4), $P_1 = \frac{3}{5}$ for long click trains, $P_2 = \frac{1}{5}$ for short click trains and $P_3 = \frac{1}{5}$ for trains with no clicks. We consider '$F_f^* \geqslant 0.95$' as the constraint when seeking designs with good (F_d^*, F_e^*) values.

Our approach does not have trouble in accommodating the constraint to obtain good multiobjective functional MRI designs. The efficiencies of the designs obtained are shown in Fig. 7 (the cluster of dots that are labelled with $F_f^* \geqslant 0.95$). From these designs, we could, as an example, select the design with approximately equal F_d^*- and F_e^*-values. One can obviously make other choices on the basis of needs or preferences. The design that we select is presented in Fig. 8. Among the 67 elements, there are 27 ones (long click trains), 12 twos (short click trains) and 12

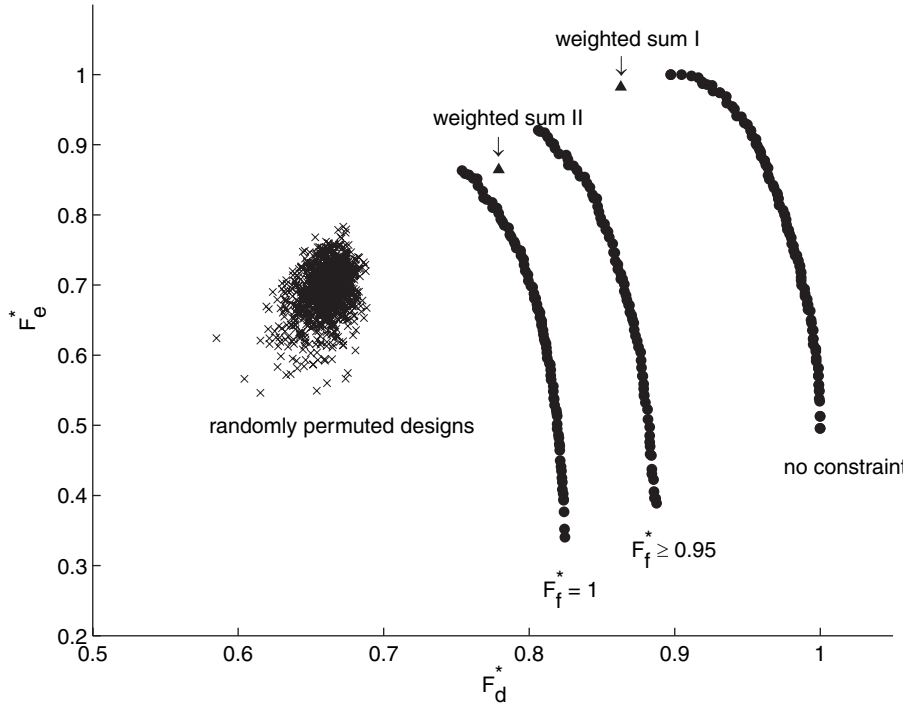

Fig. 7. F_e^*- against F_d^*-values of 1000 random permutations of designs with 22 ones, eight twos, eight threes and 29 zeros, multiobjective designs of our approach with $F_{f_1}^* = 1$, $F_f^* \geqslant 0.95$ and $F_f^* \geqslant 0$ and multiobjective designs of the weighted sum approach maximizing I, $\frac{1}{3}F_d^* + \frac{1}{3}F_e^* + \frac{1}{3}F_f^*$, and II, $\frac{1}{6}F_d^* + \frac{1}{6}F_e^* + \frac{2}{3}F_f^*$

threes (no-click trains); the remaining elements are zeros. Although the frequency of the first type of stimulus is not quite three times as much as that of the second or the third type of stimulus, (we did after all not insist that $F_f^* = 1$), the design meets the spirit of the experimenter's requirement and achieves high efficiencies for estimation and detection. In addition, although the unpredictability of designs is not explicitly taken into account, the design obtained does not seem to have a perceivable pattern (except for a larger number of long click trains). This agrees with Liu (2004), where designs with high estimation efficiencies are observed to be unpredictable; see also a discussion in Section 5.

Fig. 7 also presents the efficiencies of designs that are random permutations of a sequence with 22 ones, eight twos, eight threes and 29 zeros, which is how Brendel *et al.* (2010) obtained their design. Obtaining designs via random permutation is quite common in functional MRI. As presented in Fig. 7, random permutation does not typically result in designs with high efficiencies. Specifically, the F_d^*-values of the 1000 randomly permuted designs do not exceed 0.7 and their F_e^*-values are less than 0.8. Designs that we obtain are better.

Moreover, we modified the MATLAB program of Kao (2009) and used it to obtain designs for the current experimental setting. Appropriate weights of objectives are unknown, and we simply consider equal weights. The efficiencies of the design obtained are presented in Fig. 7, where it is labelled 'weighted sum I'. This design achieves an F_f^*-value of 0.877; it fails to satisfy the requirement of $F_f^* \geqslant 0.95$. We then consider maximizing $F = \frac{1}{6}F_d^* + \frac{1}{6}F_e^* + \frac{2}{3}F_f^*$, i.e. putting more weight on F_f^*. The obtained design is labelled 'weighted sum II' in Fig. 7. It achieves an F_f^*-value of 0.991. Although the constraint on stimulus frequency is satisfied, the (F_d^*, F_e^*) value is sacrificed for such a high F_f^*-value. Designs with better (F_d^*, F_e^*) values are obtained by our approach under the constraint; these designs are to the upper right of the second weighted sum design. Moreover, assigning equal weights on estimation and detection tends to produce designs that are in favour of estimation. This was also observed in Section 4.1. Although one can keep altering the weights for a more suitable design, this trial-and-error method is tedious and time consuming. Our approach is much easier to use and it saves time.

The efficiencies of designs that are obtained with a more strict constraint, $F_f^* = 1$, are also presented in Fig. 7. As expected, these designs perform worse in estimation and detection owing to a higher value of F_f^*. These designs can be considered when the experimenter demands a design completely satisfying the specified relative stimulus frequencies. In addition, the (F_d^*, F_e^*) values that were achieved without imposing the constraint are also plotted in Fig. 7. The stimulus frequency of these designs is close to the optimal stimulus frequency that was approximated by Liu and Frank (2004). Each symbol (0, 1, 2 or 3) occurs nearly equally often. Further investi-

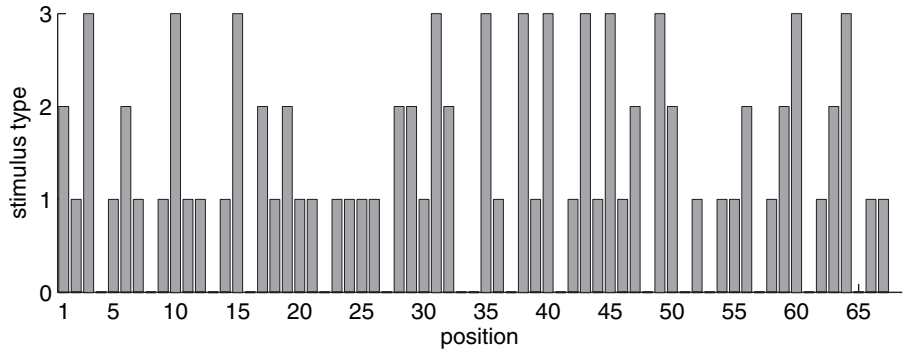

Fig. 8. Design obtained by our approach with $(F_d^*, F_e^*, F_f^*) = (0.847, 0.845, 0.953)$; this design has 27 long click trains (stimulus type 1), 12 short click trains (stimulus type 2) and 12 no-click trains (stimulus type 3)

gations are required to confirm the optimal stimulus frequency for designs with long or varying durations of stimulus.

5. Discussion

In this paper, we present an efficient approach for obtaining multiobjective designs for functional MRI experiments. Our proposed approach is built on the NSGA II approach of Deb *et al.* (2002). As in Kao, Mandal, Lazar and Stufken (2009), we include well-known functional MRI designs in the first step to improve the approach. This approach can accommodate many real life experimental settings to find a class of multiobjective functional MRI designs. With this class of designs, the trade-offs between study objectives can be explored and experimenters can choose a design from the class on the basis of their needs.

We demonstrate the advantages of our approach over the popular weighted sum method and the NSGA II approach by using various experimental settings. Compared with the weighted sum method, our approach uses much less CPU time and achieves good designs that are nearly evenly distributed on the approximated Pareto front. Our case-studies also show that designs that are obtained by the NSGA II approach may not achieve F_d^*-values as high as those for designs that are obtained by our approach (e.g. Fig. 4(b)). This drawback of the NSGA II approach can be alleviated by increasing the number of generations. For example, using the same scenario as in Section 4.1, the NSGA II approach achieves a maximal F_d^*-value of 0.994 after 100 000 generations. However, this F_d^*-value is still slightly smaller than that achieved by our approach with 2500 generations. Instead of investing much more CPU time, we take advantage of current knowledge about the performance of functional MRI designs and utilize good initial designs to improve greatly the efficiency of the search for high quality functional MRI designs. Our proposed approach largely saves resources.

We believe that the observed drawback of the NSGA II approach is due to the lack of variability in the design population that is searched by this method. Specifically, good designs for detection usually require multiple blocks of stimuli of the same type. With only random initial designs, it seems difficult for the NSGA II approach to achieve such good designs with this special feature. By contrast, our approach uses not only random designs, but also well-known functional MRI designs that are easy to generate. These designs facilitate the search by providing good building blocks to increase the variability of the design population.

In addition, although we consider A-optimality in our case-studies, our approach can also accommodate the D-optimality criterion. The selection of the optimality criterion should be guided by the needs and preferences of the experimenter. The A-optimality criterion may be considered when minimizing the average variance of parameter estimators is of interest. In contrast, D-optimality aims at minimizing the volume of a simultaneous confidence ellipsoid of the parameters; see Kao, Mandal, Lazar and Stufken (2009) and Maus *et al.* (2010a, b) for formulations of the D-optimality criterion for functional MRI. By changing the objective function, our approach can also be used for other optimality criteria.

When evaluating the performance of a design, we focus on the design's ability to detect activation, to estimate the HRF and to meet requirements for the relative stimulus frequencies. Our approach can also accommodate additional performance criteria such as the counterbalancing criterion F_c^* for avoiding patterned, predictable designs; see Kao, Mandal, Lazar and Stufken (2009) and Wager and Nichols (2003). In our experience, designs with high F_e^*-values tend to have high F_c^*-values and, hence, are not easy to predict. This observation agrees with Liu (2004). We also observe a trade-off relationship between F_c^* and F_d^*.

Moreover, we assume that the HRF shape h_q^* in model (1) is a double-gamma function or its

convolution with a boxcar function. The duration τ_{dur} of the boxcar function is typically set to the duration of a stimulus, but it should ideally reflect the duration of the underlying neuronal activity, which is usually uncertain. As shown by our simulations in Section 4.2, a small change (2s or less) of τ_{dur} does not seem to have a large effect on the efficiency of the designs that we obtain. Our designs allow the researcher to modify τ_{dur} at the analysis stage to improve the analysis results; these designs remain very efficient.

The double-gamma function that we consider is very popular and has a proven track record in functional MRI research (Lindquist, 2008; Loh *et al.*, 2008). However, our approach is not restricted to this particular function. By changing the function $g(t)$ that is described under model (1), our proposed approach can be used to search for multiobjective designs that are suited to other HRF models. If the HRF model is uncertain at the design stage, one may follow the suggestion of Maus *et al.* (2010a) of using a maximin approach for finding designs that are robust against misspecification of the HRF model. Pursuing this is beyond the scope of this present work.

We also assume the linear time invariant system; see Section 3.1. Some studies suggest that the assumption might not hold and that the functional MRI response might have a non-linear effect for some situations, e.g. when the stimuli are too close, say less than 4 s (Wager *et al.*, 2005; Soltysik *et al.*, 2004). Wager and Nichols (2003) proposed a simple way to deal with this non-linear effect. They set a ceiling for the accumulated HRFs. Similarly to Wager and Nichols (2003) and Kao (2009), our approach can be modified to find designs for cases where such a ceiling is imposed.

Acknowledgements

The research of Abhyuday Mandal was in part supported by National Science Foundation grant DMS-09-05731, and that of John Stufken by National Science Foundation grants DMS-07-06917 and DMS-10-07507. We thank Professor Dibyen Majumdar for helpful discussions, and the referees for raising questions that resulted in an improvement of this paper.

Appendix A

The design matrix $\mathbf{X}_q$ in model (1) is a 0–1 matrix for the qth-type stimuli, $q = 1, \ldots, Q$. To obtain $\mathbf{X}_q$, we first construct a working matrix $\mathbf{W}_q$ having entries 0 and 1. The first column of $\mathbf{W}_q$ indicates the times of the onsets of the qth-type stimuli in terms of multiples of ΔT. The next columns are obtained by shifting the elements of the previous column one position down and adding a 0 at the top. The matrix $\mathbf{X}_q$ is then obtained by deleting rows from $\mathbf{W}_q$, keeping only rows $1 + (i-1)m_{\text{TR}}$, where $m_{\text{TR}} = \tau_{\text{TR}}/\Delta T$ and $i = 1, 2, \ldots, 1 + [(T-1)/m_{\text{TR}}]$. Note that the dimension of $\mathbf{X}_q$ is $T \times k$, where $k = 1 + \lfloor 32/\Delta T \rfloor$ for brief (of the order of 1 s) stimuli or $k = 1 + \lfloor (32 + \tau_{\text{dur}})/\Delta T \rfloor$ for stimuli with a longer duration τ_{dur} s (Section 3.1).

As an illustrative example, we consider a design $\mathbf{s} = \{101210\ldots1\}$ of brief stimuli with time to repetition $\tau_{\text{TR}} = 2$ s and interstimulus interval $\tau_{\text{ISI}} = 3$ s; thus, $\Delta T = 1$ s, $m_{\text{TR}} = 2$ and $m_{\text{ISI}} = \tau_{\text{ISI}}/\Delta T = 3$. The stimuli of the first type occur at time 0, $2\tau_{\text{ISI}} = 6\,\Delta T$, $4\tau_{\text{ISI}} = 12\,\Delta T$, and so on (see Section 2). The first column of $\mathbf{W}_1$ is therefore $\mathbf{w}_1 = (100000100000100000...100)'$ with 1 occurring at positions 1 $(= 0+1)$, 7 $(= 6+1)$, 13 $(= 12+1)$, and so on to indicate the onsets of the first-type stimuli. The jth column of $\mathbf{W}_1$ is $\mathbf{L}^{j-1}\mathbf{w}_1$, where $j = 1, \ldots, 33$, and

$$\mathbf{L} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Similarly, the jth column of $\mathbf{W}_2$ is $\mathbf{L}^{j-1}\mathbf{w}_2$, where $\mathbf{w}_2 = (000000000100000000...000)'$. That the 10th element

of $\mathbf{w}_2$ is 1 is a consequence of the fact that the second-type stimulus occurs at time $3\tau_{\text{ISI}} = 9\,\Delta T$. We then obtain $\mathbf{X}_q$ by deleting even rows and keeping odd rows of $\mathbf{W}_q$, $q = 1, 2$.

References

Atkinson, A. C., Donev, A. and Tobias, R. (2007) *Optimum Experimental Designs, with SAS*. Oxford: Oxford University Press.

Bailey, R. A. (2007) Designs for two-colour microarray experiments. *Appl. Statist.*, **56**, 365–394.

Birn, R. M., Cox, R. W. and Bandettini, P. A. (2004) Experimental designs and processing strategies for fMRI studies involving overt verbal responses. *NeuroImage*, **23**, 1046–1058.

Bookheimer, S. (2007) Pre-surgical language mapping with functional magnetic resonance imaging. *Neurpsychol. Rev.*, **17**, 145–155.

Brendel, B., Hertrich, I., Erb, M., Lindner, A., Riecker, A., Grodd, W. and Ackermann, H. (2010) The contribution of mesiofrontal cortex to the preparation and execution of repetitive syllable productions: an fMRI study. *NeuroImage*, **50**, 1219–1230.

Brown, G. (2007) Functional magnetic resonance imaging in clinical practice: look before you leap. *Neurpsychol. Rev.*, **17**, 103–106.

Bullmore, E., Brammer, M., Williams, S. C. R., Rabehesketh, S., Janot, N., David, A., Mellers, J., Howard, R. and Sham, P. (1996) Statistical methods of estimation and inference for functional MR image analysis. *Magn. Resnce Med.*, **35**, 261–277.

Buračas, G. T. and Boynton, G. M. (2002) Efficient design of event-related fMRI experiments using m-sequences. *NeuroImage*, **16**, 801–813.

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 an activation and estimating the hemodynamic response. *NeuroImage*, **11**, suppl., S457.

Cabeza, R. and Kingstone, A. (2006) *Handbook of Functional Neuroimaging of Cognition*, 2nd edn. Cambridge: MIT Press.

Cook, R. D. and Wong, W. K. (1994) On the equivalence of constrained and compound optimal designs. *J. Am. Statist. Ass.*, **89**, 687–692.

Dale, A. M. (1999) Optimal experimental design for event-related fMRI. *Hum. Brain Mappng*, **8**, 109–114.

Deb, K. (2001) *Multi-objective Optimization using Evolutionary Algorithms*, 1st edn. New York: Wiley.

Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evolnry Computn*, **6**, 182–197.

Ding, R., Lin, D. K. J. and Wei, D. (2004) Dual-response surface optimization: a weighted mse approach. *Qual. Engng*, **16**, 377–385.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J. and Turner, R. (1995) Analysis of fMRI time-series revisited. *NeuroImage*, **2**, 45–53.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. A. and Dale, A. M. (1999) Stochastic designs in event-related fMRI. *NeuroImage*, **10**, 607–619.

Godfrey, K. (1993) *Perturbation Signals for System Identification*. New York: Prentice Hall.

Kao, M.-H. (2009) Multi-objective optimal experimental designs for ER-fMRI using MATLAB. *J. Statist. Softwr.*, **30**, 1–13.

Kao, M.-H., Mandal, A., Lazar, N. and Stufken, J. (2009) Multi-objective optimal experimental designs for event-related fMRI studies. *NeuroImage*, **44**, 849–856.

Kao, M.-H., Mandal, A. and Stufken, J. (2008) Optimal design for event-related functional magnetic resonance imaging considering both individual stimulus effects and pairwise contrasts. *Statist. Applic.*, **6**, 235–256.

Kao, M.-H., Mandal, A. and Stufken, J. (2009) Efficient designs for event-related functional magnetic resonance imaging with multiple scanning sessions. *Communs Statist. Theor. Meth.*, **38**, 3170–3182.

Kiefer, J. (1959) Optimum experimental designs (with discussion). *J. R. Statist. Soc.* B, **21**, 272–319.

Lazar, N. A. (2008) *The Statistical Analysis of Functional MRI Data*. New York: Springer.

Lindquist, M. A. (2008) The statistical analysis of fMRI data. *Statist. Sci.*, **23**, 439–464.

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.

Loh, J. M., Lindquist, M. A. and Wager, T. D. (2008) Residual analysis for detecting mis-modeling in fMRI. *Statist. Sin.*, **18**, 1421–1448.

MacWilliams, F. J. and Sloane, N. J. A. (1977) *The Theory of Error Correcting Codes*. Amsterdam: North-Holland.

Marler, R. T. and Arora, J. S. (2004) Survey of multi-objective optimization methods for engineering. *Struct. Multidisc. Optimizn*, **26**, 369–395.

Marler, R. T. and Arora, J. S. (2010) The weighted sum method for multi-objective optimization: new insights. *Struct. Multdisc. Optimizn*, **41**, 853–862.
Martin, P. I., Naeser, M. A., Doron, K. W., Bogdan, A., Baker, E. H., Kurland, J., Renshaw, P. and Yurgelun-Todd, D. (2005) Overt naming in aphasia studied with a functional MRI hemodynamic delay design. *NeuroImage*, **28**, 194–204.
Maus, B., van Breukelen, G. J. P., Goebel, R. and Berger, M. P. F. (2010a) Robustness of optimal design of fMRI experiments with application of a genetic algorithm. *NeuroImage*, **49**, 2433–2443.
Maus, B., van Breukelen, G. J. P., Goebel, R. and Berger, M. P. F. (2010b) Optimization of blocked designs in fmri studies. *Psychometrika*, **75**, 373–390.
Mechelli, A., Henson, R. N. A., Price, C. J. and Friston, K. J. (2003) Comparing event-related and epoch analysis in blocked design fMRI. *NeuroImage*, **18**, 806–810.
Miettinen, K. (1999) *Nonlinear Multiobjective Optimization*. Boston: Kluwer.
Murphy, K., Dixon, V., LaGrave, K., Kaufman, J., Risinger, R., Bloom, A. and Garavan, H. (2006) A validation of event-related fMRI comparisons between users of cocaine, nicotine, or cannabis and control subjects. *Am. J. Psychiatr.*, **163**, 1245–1251.
Ogawa, S., Lee, T. M., Nayak, A. S. and Glynn, P. (1990) Oxygenation-sensitive contrast in magnetic-resonance image of rodent brain at high magnetic-fields. *Magn. Resnce Med.*, **14**, 68–78.
Rameson, L. T., Satpute, A. B. and Lieberman, M. D. (2010) The neural correlates of implicit and explicit self-relevant processing. *NeuroImage*, **50**, 701–708.
Soltysik, D. A., Peck, K. K., White, K. D., Crosson, B. and Briggs, R. W. (2004) Comparison of hemodynamic response nonlinearity across primary cortical areas. *NeuroImage*, **22**, 1117–1127.
Summerfield, C., Egner, T., Mangels, J. and Hirsch, J. (2006) Mistaking a house for a face: neural correlates of misperception in healthy humans. *Cerebr. Cort.*, **16**, 500–508.
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.
Wager, T. D., Vazquez, A., Hernandez, L. and Noll, D. C. (2005) Accounting for nonlinear bold effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. *NeuroImage*, **25**, 206–218.
Wang, Y., Xue, G., Chen, C., Xue, F. and Dong, Q. (2007) Neural bases of asymmetric language switching in second-language learners: an ER-fMRI study. *NeuroImage*, **35**, 862–870.
Wierenga, C. and Bondi, M. (2007) Use of functional magnetic resonance imaging in the early identification of Alzheimer's disease. *Neurpsychol. Rev.*, **17**, 127–143.
Worsley, K. J. and Friston, K. J. (1995) Analysis of fMRI time-series revisited—again. *NeuroImage*, **2**, 173–181.
Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F. and Evans, A. C. (2002) A general statistical analysis for fMRI data. *NeuroImage*, **15**, 1–15.
Zhang, C. M. and Yu, T. (2008) Semiparametric detection of significant activation for brain fMRI. *Ann. Statist.*, **36**, 1693–1725.
Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M. and da Fonseca, V. G. (2003) Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Trans. Evolnry Computn*, **7**, 117–132.