ISSN No:-2456-2165

Optimal Experimental Designs for Functional Neuroimaging Studies using Gaussian Process

Amani Alrumayh
Mathematics department. College of sciences
Northern Border University
Arar, Saudi Arabia

Abstract:-The main focus of this paper is to find optimal experimental designs for functional magnetic resonance imaging (fMRI) experiments with compound stimuli by considering uncertain error correlations. We target designs, which are robust against a misspecification of error correlation. The maximin approach was proposed in the literature to tackle this problem. Unfortunately, obtaining maximin designs is computationally very expensive and time consuming. We propose to adapt Gaussian process (Kriging) that is widely used in spatial statistics and computer experiments to reduce the computational resource needed to find such designs. The proposed method is compared with a previously used approach. We observe that, in terms of the performance of the achieved designs, the results are quite similar between the two methods. In addition, our proposed Kriging approach requires less CPU time, and it is very efficient in obtaining good fMRI designs for compound stimuli experiments. The proposed approach is demonstrated via case studies.

Keywords:- *A-Optimality, fMRI, Genetic Algorithm, Kriging, Maximin Criterion.*

I. INTRODUCTION

One of the commonly used techniques in neuroimaging for studying the reactions of the human brain as it performs mental tasks is the functional magnetic resonance imaging (fMRI). Recently, there has been rapid growth in the number of neuroimaging studies accomplished by using fMRI, which involve contributions from researchers in neuroscience, psychology, physics, and statistics (Lindquist 2008). In an fMRI experiment, an experimental subject is placed inside the MRI machine. A sequence of mental stimuli, such as pictures or sounds interlaced with periods of rest or visual fixation, is then presented to the subject. While the subject is cognitively engaging with the stimuli, the MR scanner obtains a series of brain images every few seconds. The subject's blood oxygenated level dependent (BOLD) time series is then collected from each brain voxel (a three-dimensional image unit). This time series is used to make inferences about the neuronal activities at the corresponding voxel. In many cases, the mental stimuli could be 'simple', with only one component, (e.g., a picture), or 'compound', with more than one component (e.g., a cue followed by a picture).

The main focus in this paper is on fMRI studies with compound stimuli. For simplicity, we will assume that each stimulus involves two components although this assumption is not essential. As an example, each stimulus may consist of a brief cue and a mental task for the subject to complete after some specific time interval following the cue (to be explained in Section 2). For experiments involving such compound stimuli, the interest might be on examining how the subject's brain reacts toward the mental tasks rather than the joint response to both the cue and the task. Our goal is to find the 'best' experimental designs for such studies. However, this is not an easy task. One reason for this is that an fMRI design is a long sequence of mental stimuli that determines the onset times and orders of the stimuli of one or more types. Typically, an fMRI design can include tens or hundreds of stimuli. The design space that consists of all possible sequences of stimuli is enormous and an exhaustive search over this space is impossible. Therefore, an efficient search approach is needed for obtaining designs with high efficiencies to allow valid and precise statistical inferences. There are several statistical objectives in the analysis of fMRI experiments. The most common objectives, which will also be considered in this paper, are 1) detecting the brain regions that are activated by the stimuli, and 2) estimating the hemodynamic response function (HRF). The HRF is the (noise-free) fMRI signals evoked by single, brief stimulus, and can be viewed as effects of the stimulus to the brain. It is well known that an fMRI design that is good for one study objective might perform poorly for the other objective (e.g., Liu and Frank, 2004). If the goal is to study both objectives, it follows that a good multi-objective design that strikes a good balance between the two competing study objectives is needed. In addition, some unwanted psychological confounds (e.g., anticipation and habituation), and other experimental settings that a researcher might have in mind (such as a given stimulus frequency in the design) may need to be taken into account at the design stage. Consequently, finding an optimal design best suited to the needs of the experimenter can be very challenging. To account for all these aspects, a good tool based on the genetic algorithm technique was developed by Kao et al. (2009). Kao and coauthors first rigorously formulated the statistical models considered at the design stage and the multi-objective optimality criteria for evaluating the goodness of competing designs and taking advantage of knowledge about the performance of fMRI designs to propose a very efficient approach for obtaining good multi-objective fMRI designs. Their approach is shown to outperform the previous methods. It

has been applied in obtaining efficient designs for fMRI experiments (e.g., Eck et al., 2013; Kubilius et al., 2011), and adopted or adapted in tackling several important fMRI design issues (e.g., Maus et al., 2010; Maus et al., 2012; Kao et al., 2013). In this work, we will modify this approach to efficiently obtain fMRI designs for experiments with compound stimuli.

The use of compound stimuli is quite common in practice; see also, Huettel (2012) and Liu (2012). Rosenthal (2011) has adapted the approach by Kao et al. (2009) to obtain fMRI designs for such an important case. However, he followed previous studies to assume that the autocorrelation among the measurements in an fMRI time series can be specified at the design stage. Unfortunately, this assumption is not always valid. Here, we will address this issue and obtain good designs by considering uncertain error correlations. Our idea is to find a design that is relatively efficient for all possible values of the unknown autocorrelation coefficient(s). In particular, we will consider the maximin approach that is also considered in Berger and Tan (2004). This is to work on the worst-case scenario to find a design that maximizes the minimum relative efficiency (to be defined in Section 3), where the minimum is taken over all the possible values of autocorrelation coefficients. Maus et al., (2010) also considered this maximin approach. However, their method is computationally very expensive and time consuming. To reduce computational burden, we propose to combine the Kriging approach and the genetic algorithm approach of Kao et al. (2009) to find a maximin design. For comparison purposes, we obtain maximin fMRI designs using our approach, and the more expensive method by Maus et al., (2010). We observe that, in terms of the statistical efficiency of the achieved designs, the results are quite similar between the two methods. In addition, the proposed Kriging-based genetic algorithm approach requires much less CPU time and is very efficient in obtaining good fMRI designs with uncertain error correlation.

II. BACKGROUND

A. fMRI Designs for Simple Stimuli

The simplest situation when performing an fMRI experiment is probably to consider 'simple stimulus' that involves only one component (e.g., a picture or a sound). For such cases, we would like to prepare a sequence of mental tasks of one or more types (e.g., different pictures). The stimuli are separated by, say, periods of rest or visual fixation. While the selected design is presented to the experimental subject, the MR scanner starts to collect data from each of the subject's brain voxel at a fixed sampling rate of TR (e.g., 2) seconds. An example of an fMRI design sequence with two stimulus types (Q=2) can be written as $d = \{1002102\ldots.0\}$, where 1 denotes the first stimulus type that will be presented to the subject at 0 s,16 s, and so on since 1 appears at the first and the fifth positions of d; 2 indicates the second stimulus type that will be displayed at 12 s, 24 s, and so on; and 0 (the control) means no stimulus onset. Here, we assume that the specified time between elements in d is ISI = 4 seconds. After its onset, each stimulus appears briefly (e.g., 1 s), immediately followed by the control (e.g., a rest period) that appears until the next stimulus onset. Suppose TR = 2 s, the MR machine will scan the same voxel of the subject's brain every 2 seconds to collect fMRI signals. The fluctuation of the collected signals reflects the change in the concentration of oxygenated cerebral blood due to the neuronal activity in response to the mental stimulus. The noise–free fluctuation in the fMRI signals following a brief stimulus is typically described by a function of time called the hemodynamic response function (HRF). The HRF, which can be viewed as the effect of the stimulus, is of utmost importance to neuroscientist for understanding the inner workings of the human brain. One of the main assumptions considered when analyzing fMRI data is that, for the same voxel, the mental stimuli of the same type that are presented at different time points will have the same HRF. In addition, the heights of overlapping HRFs accumulate additively when multiple stimuli are presented in a time interval shorter than the duration of the HRF; the duration of a typical HRF is about 30 seconds (Henson & Friston, 2006).

B. A linear Model for Estimation

For the objective of estimating the HRF, we first need to define the HRF parameters associated with the heights of the HRF by using the discretization interval ΔT, where ΔT is the greatest time making both ISI/ΔT and TR/ΔT integers (Kao et al., 2009). We note that the heights of the HRF that contribute to the observed fMRI signals occur at 0 s and every ΔT s following a stimulus onset. A linear model that is commonly used for the estimation of the HRF is:

$$Y = X_1 h_1 + S\gamma + e, \qquad (1)$$

where Y is a $T \times 1$ vector represents the BOLD time series from a voxel, $X_1 = [X_{11} \ldots . X_{1Q}]$ is $0-1$ design matrix with X_{1q} representing the design matrix for *qth*-type stimulus ($q = 1, \ldots, Q$), $h_1 = (h'_{11}, \ldots . , h'_{1Q})'$ is the HFR parameter vector with $h_{1q} = (h_{1q1}, h_{1q2}, \ldots , h_{1qk})'$ where k denotes the number of the HRF heights and defined as $k = 1 + \lfloor H/\Delta T \rfloor$; H represents the duration of the HRF (e.g., 32 s), $S\gamma$ is the nuisance term describing the trend or drift of the time series with S being a known matrix and γ an unknown parameter vector, and e is a $T \times 1$ vector represents the correlated noise.

C. A linear Model for Detection

A commonly considered model for identifying the activated regions in the brain is:

$$Y = Z_1\theta_1 + S\gamma + e, \qquad (2)$$

where $Z_1 = X_1 h_0$; h_0 is a k× 1 vector representing the assumed shape of the HRF; and $\theta_1 = (\theta_{11}, \ldots . , \theta_{1Q})'$ denotes a vector of unknown response amplitudes. All the remaining terms in (2) are as in (1). The basis h_0 is commonly assumed to be the canonical HRF of the widely used software package SPM (http://www.fil.ion.ucl.ac.uk/spm/) for fMRI data analysis.

D. fMRI Designs for compound Stimuli

Studying the fMRI experiments that have compound stimuli has caught the attention for many researchers (Liu, 2012; Huttel 2012). In the case of compound stimulus, each stimulus has more than one component. For instance, a compound stimulus with two components may contain a cue followed by a mental task of interest or the presentation of a simple question followed by the subject's response.

For simplicity, we will consider two-component compound stimuli. In this case, an fMRI design can be written in the same way as designs for simple stimuli such as $d = \{11021\dots\}$, but 1 here represents the cue followed by the first type of stimulus, and 2 denotes the cue followed by the second type of stimulus. The two components are separated by a time interval which, following Rosenthal (2011), is denoted by CTSI. This time interval could be fixed throughout the experiment or could vary from one stimulus to another. As an example for fixed CTSI, let us consider the aforementioned design that has two different stimulus types $d = \{11021\dots\}$ with ISI = 6 s, and CTSI = 2 s. The first component of type-1 stimuli will be presented at 0 s, 6 s, and 24 s and so on. The second component of type-1 stimuli will be presented at 2 s, 8 s, and 26 s and so on. In addition, the first component of type-2 stimuli will be presented at 18 s and so on. The second component of type-2 stimuli will be presented at 20 s and other corresponding time points where '2' occurs. For cases where CTSI can vary across stimuli, we will consider a sequence that has the same length as d to indicate the CTSI for each stimulus in d. Specifically, for design $d = \{11021\dots\}$, the sequences for CTSI may look like CTSI = $\{43042\dots\}$. In this case, the subject will receive the second component of type-1 stimulus at 4 s, 9 s, and 26 s while that for the second component of type-2 stimulus will be shown to the subject at 22 s. The entry of the CTSI sequence is 0 when the corresponding entry in d is 0 since there is no stimulus presentation.

In a similar manner as the simple stimulus case, we assume that at an activated brain voxel, each component evokes a change in the fMRI signal, which is described by the hemodynamic response function HRF. Additionally, components of the same type evoke the same HRF throughout the experiment, and the heights of overlapping HRFs sum linearly. We now follow Rosenthal (2011) to extend models (1) and (2) to accommodate compound stimuli.

E. Linear Models for Compound Stimuli

To accommodate compound stimulus, we generalize models (1) and (2) respectively to be as follows:

$$Y = X_1 h_1 + X_2 h_2 + S\gamma + e; \qquad (3)$$

$$Y = Z_1 \theta_1 + Z_2 \theta_2 + S\gamma + e. \qquad (4)$$

Model (3) is for estimating the HRFs and model (4) is for detecting brain voxels that are activated by the components of the stimuli. Here, $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ are the design matrices for the first and the second components of the stimuli with $\boldsymbol{h_1}$ is the HRF parameter vector for the first components, $\boldsymbol{h_2}$ is the HRF parameter vector for the second components, $\boldsymbol{Z_i} = \boldsymbol{X_i h_0}$, with $\boldsymbol{h_0}$ being the assumed shape of the HRF, $\boldsymbol{\theta_1}$ represents the HRF amplitudes corresponding to the first components and $\boldsymbol{\theta_2}$ is for the second components. This formulation is for cases where the first components are different across stimuli of different types. All the remaining terms in (3) and (4) are as in (1) and (2).

F. Design Selection Criteria

The elaboration in the previous sections has discussed two very common statistical models that used respectively for estimation and detection. In this section, we would explicate the variance-covariance matrix of the generalized least square estimator of the parametric functions of interest. The variance-covariance matrix of parameter estimates when models (3) and (4) are considered can be written as follows:

$$\sigma^2 M = \sigma^2 C[\, W'V'(I - P_{VS})VW]^{-} C', \qquad (5)$$

where $W = [X_1\ X_2\ ]$ in the case of estimation, and $W = [Z_1 Z_2]$ in the case of detection. C is a user specified coefficient matrix of linear combinations. All other terms are as in (4). We would like a design that helps to yield the most precise parameter estimates by optimizing a statistically meaningful design selection criterion (e.g., A-optimality or D-optimality). The GA proposed by Kao et al. (2009) can help to achieve such designs when simple stimulus is considered. We adapt this algorithm to solve our problem. The objective of doing so is to search over the design space for highly efficient fMRI designs with compound stimuli. The abovementioned design criteria are used to evaluate the efficiencies of competing designs for the estimation and detection purposes. We describe our methodology in detail in the next section.

III. METHODOLOGY

A. Uncertain Corelation Coefficients

The fMRI signals acquired from the same voxel of an experimental subject tend to be correlated (Henson, 2003). To take this correlation into account, some previous studies (e.g., Worsley et al., 2002) considered the first order autoregressive process (AR1). In Kao et al. (2009), the autocorrelation coefficient of the AR1 error is assumed to be ρ = 0.3. However, this correlation coefficient may vary from voxel to voxel, and is uncertain at the design stage. Therefore, a design that is good for a given autocorrelation of a voxel might be inefficient for another voxel (Maus et al., 2010).

Our target is thus to find robust designs that are relatively efficient over a set of possible values of the autocorrelation coefficient. To this end, we will consider a maximin approach used in Maus et al. (2010). This method helps to find designs that protect against the worst case over a specified range for ρ to obtain a design that maximizes the minimum relative efficiency in estimating

model parameters; the minimum is taken over a selected grid on the range of ρ. Specifically, Maus et al. (2010) generated the locally optimal designs $d_{\boldsymbol{\rho}}$ by the genetic algorithm proposed by Kao et al. (2009) for the given ρ values ranging from 0 to 0.5 in steps of size 0.01. Each locally optimal design is the best design for the given ρ-value on the selected grid. These locally optimal designs are necessary for calculating the relative efficiencies of each candidate design. In particular, for each ρ value, the relative efficiency of a candidate design d is:

$$RE\ (d;\ \rho)\ =\ \phi(d;\ \rho)/\phi(d_{\rho};\ \rho), \qquad (6)$$

where $\phi\ (.)$ is the optimality criterion (e.g., A-optimality or D-optimality). Our goal is to obtain a design maximizing the minimal RE over $\rho\ \in\ [0, 0.5]$. This means that, for each candidate design, we will need to obtain its RE values over all the grid points on the specified range (i.e., $[0, 0.5]$) of ρ. The minimal value of these RE-values can then be determined. We then select a design that maximizes this minimum RE-value. When the minimum RE-value of a design is very close to 1, this design performs well for the different values in the range of ρ. Clearly, the previously described process will need to be repeated for every candidate design and thus is computationally expensive. For that reason, we will adapt the widely used approach in spatial statistics and computer experiments, which is known as Kriging approach to reduce the time needed for finding the desired optimal designs.

B. Gaussian Process (Kriging)

The emphasis on designing products using computer models has been steadily on the rise in the past decade due to their capability to ease the exploration of alternative designs and reduce the need for expensive hardware patterns (Jones, Schonlau, and Welch, 1998). Typically, deterministic computer models that produce the same output for the same input parameters are widely considered. These computer models normally require much time for producing an output. Therefore, a good approximation model to 'predict' the future output based on several existing outputs is helpful (Stantner, Williams, and Notz, 2003). One of the well-known approximation models is the Kriging model that has gain much popularity in approximating deterministic computer models and optimization purposes. One advantage is that a Kriging model can interpolate the observed or known data points (Martin & Simpson, 2005). Briefly speaking, Kriging is a method to build an approximation of a target function from a given finite set of evaluations of the function. It helps to approximate the result at an unmeasured location using the observed values at some (surrounding) locations. The method is initially developed by geologists to estimate the underground concentration of a valuable mineral over an area of interest given a set of sampled sites from the area (Matheron, 1963), and is now widely used in the domain of spatial analysis and computer experiments (Goldberger, 1962). The technique is also known as Gaussian process regression.

C. Some Background Information about Kriging

A Kriging model is a generalized linear regression model that accounts for the correlation in the residuals between the regression model and the observations (Goldberger, 1962). The mathematical form of a Kriging model can be written as:

$$y\ (a) = \sum_{j=1}^{p} f_j\ (a)\beta_j + Z(a) = \boldsymbol{f}'(a)\boldsymbol{\beta} + Z\ (a), \qquad (7)$$

where $f_1\ (.), \ldots., f_p\ (.)$ are known regression functions, $\boldsymbol{\beta} = (\beta_1, \ldots., \beta_p)'$ is a vector of unknown parameters. $Z(.)$ corresponds to a stationary Gaussian with mean zeros, variance σ^2 and correlation function R (.); that is:

$$Cov\ (a_1, a_2)\ =\ \sigma^2\ R\ (a_1\ , a_2). \qquad (8)$$

Here, $R\ (a_1\ , a_2)$ is the spatial correlation function and is commonly assumed to be:

$$R\ (a_1\ , a_2)\ =\ e^{-\propto|a_2 - a_1|^2}, where \quad \propto > 0 \qquad (9)$$

The correlation function parameter $\propto$ controls the smoothness of the surface. A large $\propto$ value tends to yield a smooth surface. In this paper, we follow some previous studies to assume that the correlation function is given. This implies that $\propto$ is known and does not need to be estimated.

Suppose that the 'training set' of data is obtained at given input sites $\{a_1, a_2, \ldots, a_n\}$. The resulting outputs are $\boldsymbol{Y}$ = $(y\ (a_1),\ y\ (a_2) \ldots,\ y\ (a_n))'$. Given these sampled outputs we may predict the output at another location *a* by:

$$\hat{y}(a) = \lambda'(a)\ \boldsymbol{Y}. \qquad (10)$$

When using the Kriging approach, $\lambda\ (a)$ is selected by minimizing the mean square error (MSE) for prediction,

$$\text{MSE}\ [\hat{y}(a)]\ =\ E[\lambda'(a)\ \boldsymbol{Y} - y\ (a)]^2, \qquad (11)$$

with the unbiasedness constraint,

$$E\ [\hat{y}(a)\ -\ y\ (a)] =\ E\ [\lambda'(a)\ \boldsymbol{Y} - y\ (a)]\ =\ 0. \qquad (12)$$

If $\lambda\ (a)$ solves the minimization problem of (11) subject to the unbiasedness constraint of (12), then $\lambda'(a)\ \boldsymbol{Y}$ is called the best linear unbiased predictor (BLUP) of $y(a)$. By solving for $\lambda(a)$ and substituting into (10), the BLUP of $y(a)$ is given by

$$\hat{y}(a) = \boldsymbol{f}'\widehat{\boldsymbol{\beta}} + \boldsymbol{r}'\ \boldsymbol{R}^{-1}\ \left(\boldsymbol{Y} - \boldsymbol{F}\widehat{\boldsymbol{\beta}}\right), \qquad (13)$$

where $\boldsymbol{F}$ is the expanded n$\times\ p$ matrix of regressors having (i,j)th element $f_j(a_i)$ for $1 \leq i \leq n$, $1 \leq j \leq p$, $\boldsymbol{r}$ is the n$\times$ 1 vector of correlations between the sample points and an untried point a , which is defined as $\boldsymbol{r}(a)$ = $\{R(a, a_1), R(a, a_2), \ldots, R(a, a_n)\}'$, $\boldsymbol{R}$ is the correlation matrix, which is composed of spatial correlation functions evaluated at each possible combination of the known points, and the remaining terms as defined earlier in (7).

Usually, the results of Kriging include the expected value ("Kriging mean") and variance ("Kriging variance") computed for every point within a given region. The maximum likelihood estimate of $\widehat{\boldsymbol{\beta}}$ is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{Y}, \quad (14)$$

and the MSE or variance of the estimate $\hat{y}(a)$ is as follows:

$$MSE[\hat{y}(a)] = \sigma^2\{1 - [\boldsymbol{f}'(a)\ \boldsymbol{r}'(a)]\begin{bmatrix}\boldsymbol{0} & \boldsymbol{F}' \\ \boldsymbol{F} & \boldsymbol{R}\end{bmatrix}\begin{bmatrix}\boldsymbol{f}(a) \\ \boldsymbol{r}(a)\end{bmatrix}. \quad (15)$$

The maximum likelihood estimate of σ^2 is

$$\widehat{\sigma^2} = \left(\frac{1}{n}\right)(\boldsymbol{Y} - \boldsymbol{F}\,\widehat{\boldsymbol{\beta}}\,)'\,\boldsymbol{R}^{-1}\,(\boldsymbol{Y} - \boldsymbol{F}\,\widehat{\boldsymbol{\beta}}) \quad (16)$$

D. Ordinary Kriging

Ordinary Kriging is probably the most commonly used form of Kriging that leads to satisfactory results in many cases. An ordinary Kriging model is a special case of model (7) by taking $p = 1$ and $f_1(a) = 1$.

Consequently, the mathematical form of the ordinary Kriging model is:

$$y(a) = \beta_1 + Z(a). \quad (17)$$

We will make use of this ordinary Kriging model to facilitate the search for maximin fMRI designs.

E. The proposed approach

We now describe our idea on utilizing the Kriging approach to reduce the time needed for obtaining maximin robust fMRI designs with uncertain AR1 autocorrelation coefficient ρ. First, we generate the locally optimal designs for a small sample of ρ-values; e.g., ρ = 0, 0.1,…, 0.5. Then, we calculate the RE-values of each candidate design for these six values of ρ. An ordinary Kriging model is then fitted to these RE-values to approximate the min-RE of the design d. We then search for a design that maximizes the approximated min-RE. This calls for an effective search algorithm. For this purpose, we consider to adapt the genetic algorithm by Kao et al. (2009). Our approach is thus a combination of a Kriging method that helps to approximate the objective function, and the genetic algorithm, that helps to efficiently search over the enormous design space for a design optimizing the approximated objective function. We demonstrate the usefulness of the proposed approach via case studies.

IV. CASE STUDIES

For all simulations, we assume that a compound stimulus consists of a brief cue and a mental task for the subject to complete after some specific time interval following the cue. The interest is on examining how the subject reacts toward the mental tasks (the second component) rather than the joint response to both the cue and the task. For evaluating the goodness of competing designs, we consider the A-optimality criterion. Model (3) and (4) are considered respectively for estimating the HRF of the second component of each compound stimulus, and for detecting brain voxels activation of this second component. The duration of the HRF is set to 32 seconds, which is the duration of the canonical HRF in SPM. The nuisance term $S\gamma$ of models (3) and (4) is assumed to allow for a second-order polynomial drift. The noise is assumed to follow a stationary AR (1) process with uncertain autocorrelation coefficients $\rho \in [0, 0.5]$. For the following simulations, we consider designs with Q = 1, 2 and 3 stimulus types corresponding to design lengths of L = 255, 242 and 255 events, respectively. For all designs, The ISI is set to 8s, the TR is set to 2s, and each individual CTSI is allowed to equal 2, 4, or 6 seconds.

Algorithmic parameters of the genetic algorithm (GA) used in the simulations are: G (size of generation) = 20, q (percentage of mutation) = 1%, and N (number of immigrants per generation) = 4. The algorithm is run until a stopping rule is met, e.g., no significant improvement is made (for more details see Kao, 2009). The algorithm keeps track of the design with the best fit over all generations.

We modify the MATLAB program provided by Kao (2009), and combine it with the software package DACE for Kriging approximations (Lophaven, Nielsen, and Sondergaard, 2002). We implement our simulations by using MATLAB (version R2012b) on a desktop computer with a 3.0 GHz Intel Pentium 4 quad-core processor.

A. CaseI: Estimation

The first set of case studies focuses on estimating the HRF of the second component of each compound stimulus by using model (3). We first consider the approach of Maus et al. (2010) for obtaining maximin designs. This is to generate 51 locally optimal designs for each stimulus type for $\rho = 0, 0.01, \ldots, 0.5$ by using the genetic algorithm of Kao et al. (2009). These 51 locally optimal designs allow to provide an approximated min-RE for each candidate. We then adapt the genetic algorithm to search for a design maximizing the min-RE approximated by this time-consuming 'grid method'.

We then use our proposed method to obtain maximin designs. Instead of obtaining the abovementioned 51 locally optimal designs, we obtain only six locally optimal designs for ρ = 0, 0.1, …, 0.5. For each candidate design, we then can calculate the six RE-values using the six locally optimal designs. By using the Kriging approximation, these six RE-values then help to approximate the 51 RE-values for ρ = 0, 0.01, …, 0.5. This latter approximation is done by using the abovementioned DACE software package. The default settings of the DACE package are considered. This includes a default value for $\propto = 10$ that controls the smoothness of the surface to be approximated. The genetic algorithm is then adapted to search for a maximin design that maximizes the approximated min-RE. Hereinafter, we will call this latter method the 'Kriging method'.

ISSN No:-2456-2165

For comparing the maximin designs obtained by the Kriging and grid methods, we re-evaluate the min-RE of the designs obtained by the Kriging method by using the 51 locally optimal designs for ρ = 0, 0.01, … ,0.5. Table I represents the (re-evaluated) min-RE values of the designs obtained by the two methods for Q=1, 2 and 3. We also report the CPU time needed by the two methods in hours.

Q	*Grid method*	*Kriging method*
1	min-RE = 0.9747	min-RE = 0.9790
	CPU=4.21 hours	CPU=1.65 hours
2	min-RE = 0.9796	min-RE = 0.9745
	CPU=15.43 hours	CPU=1.43 hours
3	min-RE = 0.9747	min-RE = 0.9765
	CPU=15.46 hours	CPU=3.16 hours

Table 1:- The results for estimating HRF using the grid method vs. the Kriging method

When comparing the min-RE value for both methods, we observe that the values are quite similar and close to 1. This indicates that all the obtained designs are very efficient. From Table I, we also observe a huge reduction in the CPU time needed for generating a maximin design when the Kriging approach is employed. We note that, the reported CPU does not include the time needed for generating locally optimal designs. When using the grid method, a total of 51 locally optimal designs are needed. This requires about 5, 10, and 18 hours for Q = 1, 2, and 3, respectively. For the Kriging method, we only need six locally optimal designs. Our proposed method saves much computational resource without sacrificing the efficiency of the obtained designs. *Fig. 1* provides a similar comparison result between the two methods. There, the blue bars correspond to $min\{RE(d_{Kriging};\rho)\}/min\{RE(d_{grid};\rho)\}$, where $d_{Kriging}$ and d_{grid} are maximin designs obtained by the Kriging and grid methods, respectively. The relative CPU time for obtaining $d_{Kriging}$ to d_{grid} is also shown in *Fig. 1*.

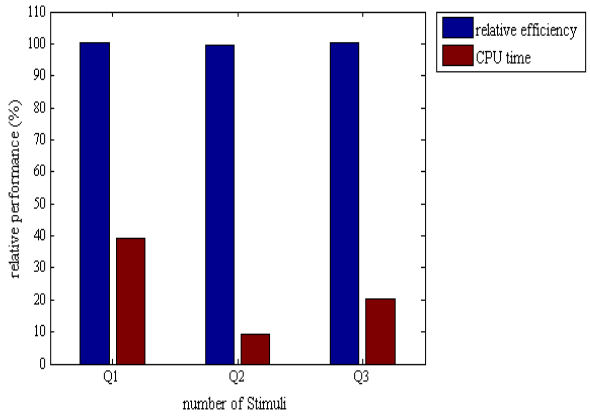

Fig 1:- Grid vs. Kriging methods for estimation purpose.

The figure suggests that the efficiencies of the achieved designs by the Kriging method are similar to those obtained by the grid method with a greatly reduced CPU time. In particular, the reductions in the CPU times are 61%, 91%, and 80% for Q=1,2 and 3, respectively.

B. CaseII: Detection

In a similar manner as we did for the estimation case, we obtain maximin designs for detecting brain activations by using the grid method and our proposed Kriging-based genetic algorithm approach. We then compare the two methods. Table II represents the min-RE values and CPU times of the grid and Kriging methods for Q=1, 2 and 3. The results suggest that the obtained maximin designs yield high statistical efficiencies in detecting brain activations and the maximal min-RE values achieved by the two methods are quite similar. In addition, the CPU time required by the Kriging method is much less than that of the grid method.

Q	*Grid method*	*Kriging method*
1	min-RE = 0.9859	min-RE = 0.9708
	CPU=3.49 hours	CPU=0.44 hours
2	min-RE = 0.9843	min-RE = 0.9801
	CPU=14.82 hours	CPU=0.95 hours
3	min-RE = 0.9918	min-RE = 0.9912
	CPU=10.77 hours	CPU=1.25 hours

Table 2:- The results for detecting the activated brain regions using the grid method vs. the Kriging method

Fig. 2, which is to be read as *Fig. 1*, suggests that the efficiencies of the achieved designs by the Kriging method are similar to those obtained by the grid method. In addition, the Kriging method helps to reduce the CPU times by 88%, 94%, and 88% respectively for Q = 1,2 and 3. Again, our proposed method uses much less CPU time to obtain high-quality maximin designs.

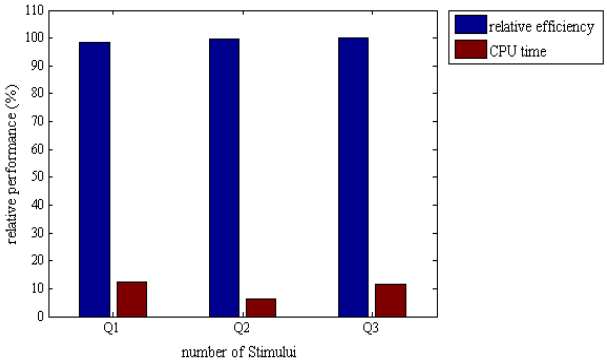

Fig 2:- Grid vs. Kriging methods for detection purpose.

V. CONCLUSION AND DISCUSSION

In this paper, we propose the use of a Kriging approximation method combined with a knowledge-based genetic algorithm in order to reduce the computational resource needed for obtaining good fMRI designs involving compound stimuli with uncertain autocorrelation coefficients.

We consider two common statistical objectives in the analysis of fMRI experiments, which are the estimation of the HRF and detection of the activated brain regions. For these two common objectives, our proposed method can obtain maximin designs that perform similarly to designs obtained by the grid method considered by Maus et al., (2010). But what makes using Kriging method very attractive is that the CPU time required by the Kriging approach is much less than that of the grid method.

When approximating the objective function to be optimized, we consider the ordinary Kriging model. This selection leads to results that are quite satisfactory for both estimation and detection purposes. There might be other approximation methods that can be applied to the current situation and may provide slightly better results which is a future research of interest.

REFERENCES

[1]. Berger, M.P.F., Tan, F.E.S. (2004). Robust Designs for Linear Mixed Effects Models. Appl. Stat.53 (4), 569-581.
[2]. Buračas, G. T., & Boynton, G. M. (2002). Efficient Design of Event-Related fMRI Experiments using M-sequences." NeuroImage, 16(3), 801-813. Part1..
[3]. Eck, J., Kaas, A. L., and Goebel, R. (2013). Crossmodal Interactions of Haptic and Visual Texture Information in Early Sensory Cortex. NeuroImage 75, 123-135.
[4]. Goldberger, A. S., Best Linear Unbiased Prediction in the Generalized Linear Regression Model. Journal of the American Statistical Association, Vol. 57, No. 298, 1962, pp. 369–375.
[5]. Henson, R. (2003). Analysis of fMRI Time series: Linear Time-Invariant Models, Event-related fMRI and Optimal Experimental Design. Human Brain Function. (793 - 822). Elsevier: London. 1-41.
[6]. Henson, R., & Friston, K. (2006). Convolution Models for fMRI. 178-192.
[7]. Huettel, S. A. (2012). Event-Related fMRI in Cognition. NeuroImage 62, 1152-1156.
[8]. Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient Global Optimaization of Expensive Black-Box Functions. Journal of global optimaization, 455-492.
[9]. Kao, M-H. (2009, June). Multi-Objective Optimal Experimental Designs for ER-fMRI using MATLAB. Journal of Statistical Sofrware, 30(11), 1-14.
[10]. Kao, M-H., Majumdar, D., Mandal, A. and Stufken, J. (2013). Maximin And Maximin Efficient Event-Related fMRI Designs under A Nonlinear Model, Annals of Applied Statistics.
[11]. Kao, M-H., Mandal, A., Lazar, N., Stufken, J., 2009. Multi-Objective Optimal Experimental Designs for Event-Related fMRI Studies. NeuroImage 44, 849-856.
[12]. Kubilius, J., Wagemans, J., and Op de Beeck, H. P. (2011). Emergence of Perceptual Gestalts in the Human Visual Cortex: The Case of the Configural-Superiority Effect. Psychological Science 22, 1296-1303.
[13]. Lindquist, M. A. (2008), The Statistical Analysis of fMRI Data. Statistical Science, 23, 439–464.
[14]. Liu, T. T. (2012). The Development of Event-Related fMRI Designs. NeuroImage 62, 1157-1162.
[15]. Liu, T. T., & Frank, L. R. (2004). Efficiency, Power, and entropy in event-related fMRI with multiple trial types: Part I: theory. NeuroImage, 21(1) 387-400.
[16]. Lophaven, S. N., Nielsen, H. B., & Sondergaard, J. (2002, August 1). DACE AMATLAB Kriging Toolbox. 1-26.
[17]. Martin, J. D., & Simpson, T. W. (2005, April 4). Use of Kriging Models to Approximate Deterministic Computer Models. AIAA JOURNAL, 43, 853-863.
[18]. Matheron, G. Principles of Geostatistics. Economic Geology, Vol. 58, 1963, pp. 1246–1266
[19]. Maus, B., van Breukelen, G. J. P., Goebel, R., and Berger, M. P. F. (2012). Optimal Design for Nonlinear Estimation of the Hemodynamic Response Function. Human Brain Mapping 33, 1253-1267.
[20]. Maus, B., van Breukelen, G. J. P., Goebel, R., and Berger, M. P. F. (2010). Robustness of Optimal Design of fMRI Experiments with Application of a Genetic Algorithm. NeuroImage 49, 2433-2443.
[21]. Rosenthal, A. (2011). Experimental Designs for Functional MRI with Compound Stimulus. Unpublished report for M.S. applied project, School of Mathematical and Statistical Sciences, Arizona State University.
[22]. Santner, Thomas J., Brian J. Williams, and William I. Notz. The Design and Analysis of Computer Experiments. New York: Springer, 2003.
[23]. Worsley, K. J. and Liao, C. H. and Aston, J. and Petre, V. and Duncan, G. H. and Morales, F. and Evans, A. C. (2002), A General Statistical Analysis for fMRI Data, NeuroImage, 15, 1–15.