Proceedings of the 2010 Industrial Engineering Research Conference

Adaptive Sequential Experimentation Methodology for Response Surface Optimization

A.Alaeddini, K.Yang*, A.E. Murat

Department of Industrial and Manufacturing Engineering, Wayne State University, Detroit, MI 48201, USA
E-mail: dz3027@wayne.edu, ac4505@wayne.edu , amurat@wayne.edu

Abstract

Response surface methodology (RSM) is a collection of statistical and operational research techniques for optimizing process and designs. Traditional RSM designs use fixed factor settings to fit appropriate mathematical model which is not very effective in terms of the number of required experiments and model precision. In this paper, we introduce the methodology of adaptive sequential experiment, which uses previous experiments information for determining factor settings of new experiments, and shrinks the factor space to smaller a region toward the optimal point. The effectiveness of the proposed methodology is examined on a case study conducted on rat brain trauma.

Keywords

Adaptive Sequential Experiment, Response Surface Optimization, Central Composite Design (CCD)

1. Introduction

Most process and design optimization approaches such as the response surface methodology (RSM) require a complete experimental design to be determined prior to the experimentation process (Spendley, Hex and Himsworth, [1]). These preset designs offer ease of implementation and good performance over a wide range of applications. However, they lack the ability to adapt the design based on the characteristics of application and experimental space so as to reduce the number of experiments necessary. This, in particular, constitutes a major disadvantage in many industrial applications where the cost of experimentation is high or when the experimentation resources are limited. These industrial experiments share the following two main characteristics: (1) prior to the experiment, the behaviour of the experimental design space is not well known; (2) the cost of each experimental trial is prohibitively high and the experimental budget is limited. An example for such industrial experiments is the combustion test for aircraft engine or turbines where prototypes are very expensive and the behaviour of different designs are highly unpredictable.

In this study, we propose an adaptive sequential experimentation approach based on response surface methodology for optimizing design and processes. This approach is a local optimization approach for physical experiments where the region of interest of two primary input factors is already determined. The shifting the region of interest close to the optimum is beyond the scope of this paper and is assumed to be performed a priori using an efficient method such as steepest descent. We therefore focus on the sequential experimentation in the region of interest which can be considered as successive series of small data collection efforts. At each step, we learn from the previous results, refine our understanding and develop a new model for the next experiment to reduce waste and improve the quality of results. The idea of adaptive experimental design is not new. Beginning with the sequential RSM experimentation with multiple blocks in Box and Wilson [2], there have been many ideas such as one-factor-at-a-time (OFAT) (Friedman and Savage [3], Daniel [4]), adaptive OFAT (Frey, Engelhardt, and Greitzer, [5]), adaptive RSM (Wang, Dong and Aitchison [6], Wang [7]), successive RSM (Stander, [8]), evolutionary operation (Box and Draper, [9]), steepest ascent based methods (Box and Wilson, [10]), and sequential and adaptive approximation methods from the engineering design discipline.

This paper differs from earlier approaches on adaptive and sequential RSM in three different ways. First, the reduction of the region of interest is optimal if the relationship between the response and input factors is quadratic and response is deterministic. Specifically, the optimal factor combination is always contained in the reduced region. Second, the method requires fewer experiments in each run as a result of inheriting previous experiments and fixed design structure. Lastly, the proposed method identifies the reduced region of interest with a ranking based method rather than the response levels obtained from each experiment. This is indeed similar to using not the value of a parameter but its rank in robust statistics (Hettmansperger and McKean, [11]).

2. Proposed Methodology

2.1 Terminology and Assumptions

Let FS_r denote the factor space at run $r = 1,2, \dots, R$ and is expressed as the Cartesian product of the chosen range of all variables. Each run r contains in total e experiments some of which are inherited from earlier runs. At each run, the experiments are ranked according to response levels where B and W denote the best and worst experiments, respectively. Further, let N_k denote the experiment with the k^{th} best result $(2 \le k \le e-1)$ in a given run. The factor space is classified as best at corner (*BCO*) or best at center (*BCE*) depending on the ranking and the location of the *B*. At run *r*, the optimal region (OR_r) containing the estimated optimal experiment (EO_r) is determined based on the factor space classification. The estimated optimal in each run is our best estimation of the optimal experiment (*O*).

As in most RSM approaches, the proposed methodology relies on a number of simplifying assumptions. The extensions due to the relaxation of these assumptions are beyond the scope of this paper and some of these extensions discussed in the conclusion. For the proposed methodology we consider the following assumptions: (1) There are two significant and controllable factors; (2) The underlying relation between a single response and two factors can be represented by a quadratic model; (3) The response is convex in the region of interest; (4) The factor space in the region of interest is feasible.

2.2 Algorithm and Initial Run Design

Figure 1 illustrates the structure of the proposed methodology. The procedure is initialized with the region of interest, e.g., a factor space which guaranteed to contain the optimal experiment. The goal is reach to the vicinity of O in a finite set of runs (R). Each run is setup with a modified version of the factorial design augmented with a center point. Once the experimentation is completed, the approach follows two concurrent strategies, e.g., non-parametric ranking strategy and parametric model fitting strategy. According to the ranking of experiments and the estimated optimal point from quadratic model fitting, a reduced factor space containing the estimated optimal experiment is determined for the next run. This procedure continues until the convergence criteria based on estimated optimal experiment or coefficient of determination of the fitted model is attained. The justification for the dual parametric and non-parametric strategies is that, while the information from ranking strategy is accurate but not precise, the information from model fitting is precise but not accurate.

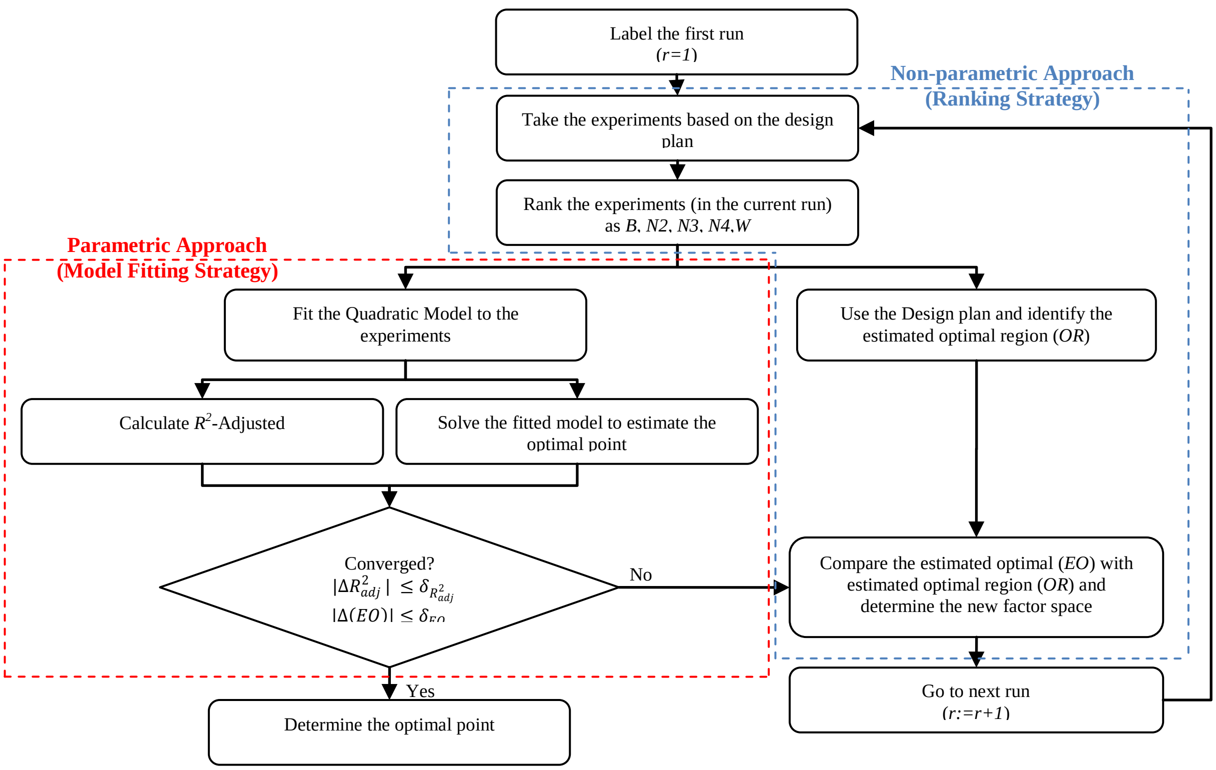


Figure 1: Scheme of the proposed methodology

2.3 Non-parametric Approach: Ranking Strategy

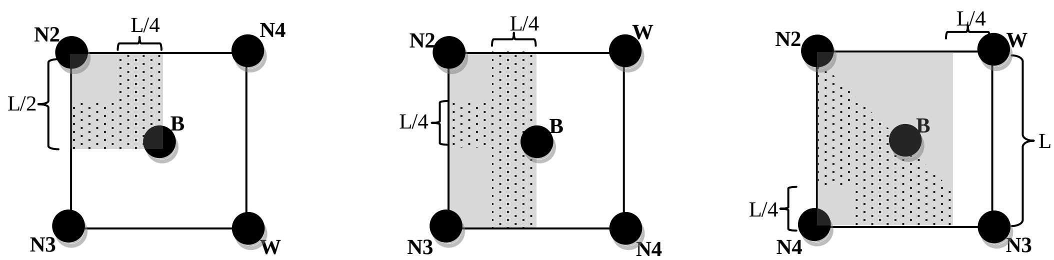
At each run *r* of the proposed approach, we first rank the experiments as $B, N2, N3, N4$ and W according to the response levels. Throughout the process, we maintain the same experimental design and consider a constant number of experiments, e.g. e=5. Based on the ranking, we identify the implied optimal region which contains the EO_r. This region is a polygon contained in FS_r and can be convex or non-convex. We then identify a rectangle which contains the implied optimal region and denote it as the optimal region (OR_r), which determines the factor space of the next run. This process of encapsulating the implied optimal region with a rectangle is a form of relaxation and is not efficient in terms of factor space reduction. However, there are valid reasons which motivate this relaxation. The foremost reason is the reduced need for new experiments due to the inheritance of experiments from the previous run. Secondly, the rectangular *FS* preserves the orthogonality of factorial experimental design. Further, this rectangular form facilitates the recursive characterization of the same rectangular structure throughout the process. In addition, we can use the same experimental design structure, e.g. full factorial with a center point. Lastly, the relaxation reduces the risk of selecting an optimal region which excludes the optimal experiment.

An alternative to the rectangular envelope is the convex hull of implied optimal region. Due to its convexity, it also allows for easier tessellation of the FS. While the convex hull reduces the optimal region more than the rectangular envelope, it does not reduce the number of new experiments as much. Furthermore, the experimental design used in each run will be different since the convex hulls of the implied optimal regions will vary in shape. Clearly the choice of the right form is a trade-off between the rate of contraction of the optimal region and the total number of experiments conducted.

We express the factor space of each run (FS_r) as a mapping (φ_r) of the factor space of the preceding run (FS_{r-1}). The output of this mapping φ_r depends on the current factor space, the outcome of ranking, the experimentation design as well as the result of parametric strategy described in subsection 2.4. In most general form, the proposed methodology generates a series of factor spaces which are nested, e.g. $FS_r = \varphi_r\left(\varphi_{r-1}\left(\dots\varphi_0(FS_1)\right)\right)$. With rectangular envelope, the mapping across runs will be identical, e.g. $\varphi(\cdot) \equiv \varphi_r(\cdot)$ for $\forall r$. This is because we maintain the same experiment design structure and there is a finite number of optimal region as a result of ranking outcomes. In what follows, we present the optimal region alternatives based on the ranking of the experiments and the location of B in the current FS.

2.3.1 Best at Center (BCE) Optimal Regions

An important information obtained from the ranking of experiments is the location of B. When the B is located at the center, the current FS is then classified as having a *BCE* optimal region. Depending on the location of $N2, N3, N4,$ and W, there are three possible ORs as illustrated in Figure 2. We first determine the implied optimal region as illustrated as dotted regions in Figure 2. Next we characterize the *OR* as the rectangle which contains the this implied optimal region.



(a) Central Design with N2 cornered with N3 and N4 (b) Central Design with N2 cornered with N3 and W (c) Central Design with N2 cornered with N4 and W

Figure 2: BCE ORs (dotted region: implied OR, shaded region: OR)

The implied optimal regions are guaranteed to contain the optimal experiment in the absence of random noise. The mathematical proofs of the optimality of these implied optimal regions is involved and thus excluded. Instead, we provide a general proof sketch of the rectangular optimal regions and illustrate it for the *OR* in Figure 2a. The proofs are accomplished through the following steps: (1) Divide the non-optimal region into smaller rectangular sub-regions using factor centerlines; (2) Assume that the optimal point falls in one of these sub-regions; (3) Relocate the origin to that region and formulate the responses at B, N_1,..., and W based on their displacement from the new origin; (4) Show that at least one pairwise comparison of the responses violate the initial ranking (5) Replicate the steps (2-5) until all sub-regions are evaluated. The proof of *OR* in Figure 2a. is as follows.

Proposition: *For BCE optimal region with ranking in Figure 2a, the optimal experiment is located in the optimal region characterized as the quadrant with corners at B and N2 when there is no random noise.*

Proof. Consider that the *FS* is divided into equal quadrants (I,II,III,IV) which have (N2,N4,W,N3) as the corners, respectively. Further suppose that *O* is located in *FS* outside the *OR*. For the case, where *O* is in II, we consider the responses at N2$\equiv (x_{N2}, y_{N2})$, N3$\equiv (x_{N3}, y_{N3})$ and N4$\equiv (x_{N4}, y_{N4})$ as z_{N2}, z_{N3} and z_{N4}, respectively. We have $z_{N2} = A(dx_2)^2 + B(dy_2)^2 + C(dx_2)(dy_2)$, $z_{N3} = A(dx_3)^2 + B(dy_3)^2 + C(dx_3)(dy_3)$, anf $z_{N4} = A(dx_4)^2 + B(dy_4)^2 + C(dx_4)(dy_4)$ where $(dx_2,dy_2)=(x_{N2}-x_O, y_{N2}-y_O)$, $(dx_3,dy_3)=(x_{N3}-x_O, y_{N3}-y_O)$ and $(dx_4,dy_4)=(x_{N4}-x_O, y_{N4}-y_O)$ and optimal experiment location O$\equiv (x_O, y_O)$. We consider

$$dz_{N2N4} = z_{N2} - z_{N4} = A[(dx_2)^2 - (dx_4)^2] + B[(dy_2)^2 - (dy_4)^2] + C[(dx_2)(dy_2) - (dx_4)(dy_4)]$$
$$dz_{N3N4} = z_{N3} - z_{N4} = A[(dx_3)^2 - (dx_4)^2] + B[(dy_3)^2 - (dy_4)^2] + C[(dx_3)(dy_3) - (dx_4)(dy_4)].$$

Since the response is convex (e.g., $A, B > 0$), we consider three response scenarios: C=0, C<0, C>0. Note that when O is in II, we have $|dx_2| > |dx_4|$, $dx_2 \leq 0$ and $dy_2 = dy_4$ making second term dz_{N2N4} zero. For C=0, we have the first term in dz_{N2N4} positive, thus $dz_{N2N4} > 0$ which is a contradiction to the ranking $z_{N2} < z_{N4}$. For C<0, the third term in dz_{N2N4} is positive since $dx_2 \leq 0$ thus making $dz_{N2N4} > 0$ which is also a contradiction. Lastly, for C>0, first and second terms in dz_{N3N4} are positive since $|dx_3| > |dx_4|$ and $|dy_3| > |dy_4|$. Last term in dz_{N3N4} is also positive since $dx_3 dy_3 > 0$ and $|dx_3 dy_3| > |dx_4 dy_4|$. Thus $dz_{N3N4} > 0$ which is a contradiction to the ranking $z_{N3} < z_{N4}$. For the case, where O is in III, we consider the responses at N2, N3 and W as z_{N2}, z_{N3} and z_w, respectively. Let's define the dz_{N3W}, dz_{N2W}, and dz_{N2N3} as before. For case C=0, it can be shown that $dz_{N3W} > 0$ which is a contradiction for $z_{N3} < z_W$. Similarly, for C<0 and C>0, we have $dz_{N2W} > 0$ and $dz_{N2N3} > 0$ are contradictions for $z_{N2} < z_W$ and $z_{N2} < z_{N3}$. Last case is where *O* is in IV. We consider the responses at N2, N3, N4 and W. Let define dz_{N4W} as before. For case C=0, it can be shown that $dz_{N2N3} > 0$ which is a contradiction for $z_{N2} < z_{N3}$. Similarly, for C<0 and C>0, we have $dz_{N2N3} > 0$ and $dz_{N4W} > 0$ are contradictions for $z_{N2} < z_{N3}$ and $z_{N4} < z_W$.

2.3.2 Best at Corner (BCO) Optimal Regions

The case when the *B* is located at a corner is referred as a *BCO* optimal region. In *BCO*, either *N2* or *N3* can occur at the center. For *N2* at center, there are three possible ORs based on the location of $B, N3, N4,$ and W (Figure 3).

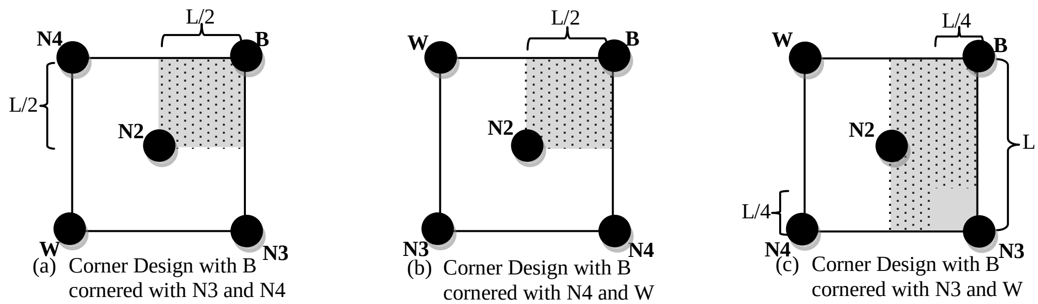


Figure 3: BCO ORs when *N2* is at center (dotted region: implied OR, shaded region: OR)

In case with *N3* at center, there are two possible ORs based on the location of $B, N2, N4,$ and W (Figure 4).

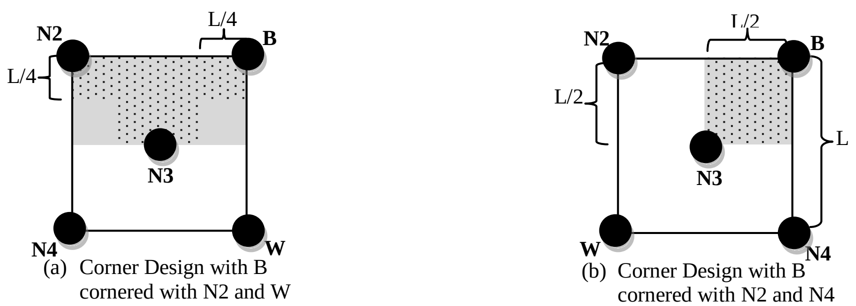


Figure 4: BCO ORs when *N3* is at center (dotted region: implied OR, shaded region: OR)

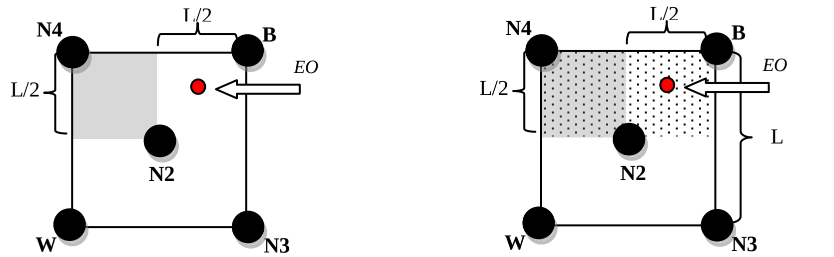
The proving strategy for BCOs' ORs is the same as BCE and is thus excluded. Note that the implied optimal regions are identical to the *ORs* in Figures 3a, 3b, and 4b, thus there is no relaxation due to rectangular envelope.

2.4 Parametric Approach: Model Fitting Strategy

We use a parametric approach based on model fitting in addition to the ranking approach. This strategy not only allows us to increase the precision of *EO* but also supports backtracking through *FS* correction as explained in section 2.5. Beginning with the completion of all first run experiments, this parametric approach is used after each experiment. In this approach we fit a quadratic model $z = [a_1 \quad a_2]\begin{bmatrix} x \\ y \end{bmatrix} + [x \quad y]\begin{bmatrix} q_1 & q_2 \\ q_2 & q_3 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + c + \varepsilon$, with $\varepsilon \sim N(0, \sigma^2)$, to the experimental data to analyze the underlying function of data and efficacy of conducted experiments. In fitting the quadratic model, two objectives are being sought in particular: (1) estimating the estimated optimal point (EO); (2) calculating the adjusted coefficient of determination (R^2_{adj}). EO, the minimum of the fitted model, not only shows the predicted optimal solution, but can also be used for correcting the FS of the next run. Furthermore, the EO changes in consecutive runs is also used as a stopping criterion. In comparison, the R^2_{adj} shows how well the information gained from the experiments explains the behavior of the underlying system (Nagelkerke [12]). We also use this measure as a stopping rule in the proposed methodology and for comparing the explanatory power of different methods.

2.5 Design Structure for the Next Runs

Following the characterization of *OR* through ranking approach and estimation of the *EO* from parametric approach, we determine the design structure for the next run. In particular, we compare the EO from the model fitting with the OR from the experiment ranking. If EO is contained in the OR, then we use the region as the factor space of the next run. If EO is not contained in the OR, we then expand the optimal region to a larger rectangle envelope containing the EO and use the region as FS of the next run (Figure 5). Next we conduct experiments on the un-experimented corners and the center of the new FS. After each experiment, we fit the quadratic model and check whether EO is contained in OR. If *EO* is outside *OR*, then we expand the *OR* as before. This expansion serves as a backtracking step. These steps are repeated monitor the change in R^2_{adj} and the EO using the fitted model. The stopping condition for the proposed approach is the convergence of R^2_{adj} or EO with thresholds δ_{EO} and $\delta_{R^2_{adj}}$.



(a) Initial FS not including EO (b) Increased FS based on EO location

Figure 5: Expansion of the *OR* when the *EO* from model fitting falls outside

3. Numerical Examples: Traumatic Brain Injury (TBI)

TBI continues to be a serious societal problem that affects more than 1.4 million Americans each year (Mao [13]). The controlled cortical impact (CCI) rat model is one of the most frequently used animal models. This model is used to correlate real world injuries with predictions from a validated finite element (FE) model in order to establish injury threshold. In CCI model, the impact depth (potentially ranging 1-3 mm) and the impactor diameter (potentially ranging 2.5-7.5 mm) are believed to be two main factors in determining injury severity. However, the percent of increase/decrease in size of rat brain contributes to variances observed in post-impact tissues. Since the effect of this external parameter is largely unknown, it can be considered as noise. In CCI studies, one common problem is to find the specific levels of factors that result in specific percent of injury in animal brain. However, these experiments are not only very expensive, but are also very time consuming.

In this case study, we used the proposed approach along to find the parameter setting that result in 30% injury in the rat brain. We also conducted CCD experiments to compare the performance with the proposed approach. The technical details of the experiments can be found in Mao [13] and Mao, Zhang, Yang, and King [14]. Tables 1 and 2 show the conducted experiments of CCD, and the proposed approach at different runs.

Table 1: Experiments for CCD approach

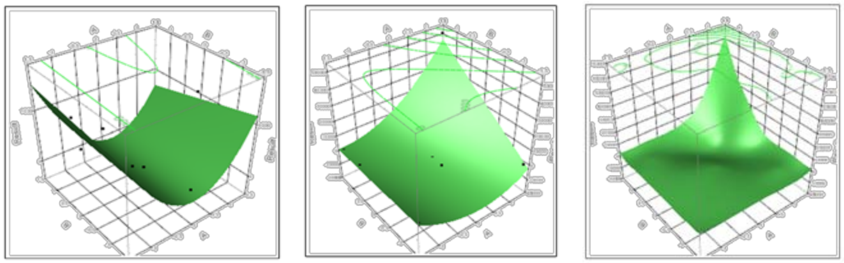
Controllable Factors				Random Factor	Response
Coded		Original			
impact depth	impactor diameter	impact depth (mm)	impactor diameter (mm)	brain size variation	Brain Injury
1	-1	0.7	1.8	0%	900.00
1	-1	2.1	1.8	1%	306.37
-1	1	0.7	5.3	-2%	900.00
1	1	2.1	5.3	0%	11.54
-1.41	0	1.0	5.0	1%	894.44
1.41	0	3.0	5.0	1%	1080.03
0	-1.41	2.0	2.5	1%	280.58
0	1.41	2.0	7.5	0%	206.08
0	0	2.0	5.0	0%	6.57
0	0	2.0	5.0	-1%	3.40
0	0	2.0	5.0	1%	10.19
0	0	2.0	5.0	2%	16.29
0	0	2.0	5.0	-2%	0.875

Table 2: Experiments for adaptive sequential experimentation based on RSM

Run	Controllable Factors				Random Factor	Response
	Coded		Original			
	impact depth	impactor diameter	impact depth (mm)	impactor diameter (mm)	brain size variation	Brain Injury
1	-1.41	-1.41	1	2.5	-1%	877.92
	1.41	1.41	3	7.5	0%	51595.49
	1.41	-1.41	3	2.5	0%	78.93
	-1.41	1.41	1	7.5	0%	761.34
	0	0	2	5	1%	10.19
2	-1.41	0.71	1	6.25	0%	852.10
	1.41	0.71	3	6.25	1%	19476.43
	0	-0.35	2	4.375	1%	51.51

Using 13 experiments, the CCD fits a quadratic surface with $\mathrm{R}^2_{\mathrm{adj}}$ =68.31% and identifies EO = (0.1857, 0.3286). Figure 6a shows the 3D plot of CCD estimated surface. In comparison, using 8 experiments in 2 runs, the proposed approach fits the quadratic model shown in Figure 6b with R^2_{adj} =82.59% and EO = (0, 0.0505).

As shown in the Figure 6, although the estimated optimal points of both approaches are close, the estimated contours of the two methods are significantly different. To compare the estimated functions, we aggregate the experimental data from both approaches (e.g. Tables 1 and 2) and used the Radial Basis Function (RBF) to find the best fit. Figure 5c illustrates the model fit using RBF and the aggregated experimental data.



(a) **(b)** **(c)**

Figure 6: 3D plots of conducted experiments ((a) CCD, (b) Proposed Approach, (c) RBF)

From Figure 6 it can be seen that, the estimated function of the proposed approach is much more similar to RBF with aggregated experimental data. Meanwhile, we also note that the EO of the proposed approach is very close to that of RBF based on the contour plots in Figure 6b and 6c. Given that the underlying response model is potentially highly nonlinear, the outcomes are very encouraging. Also, in another study based on 12 simulated examples which is not reported here the proposed approach demonstrated better performance in terms of both the number of experiments and goodness of fit. Such results clearly show the effectiveness of the proposed approach in real world applications.

4. Conclusions

In this paper we developed and presented an adaptive methodology for response surface optimization. In this approach, we combined the procedures from numerical optimization and response surface methodology along with the concept of the design of experiments. We also showed the discussed approach performs superior to one of the most popular methods in practice. The proposed methodology uses both model free and model based approaches for estimating the optimal point of the response surface at each run. Further, it inherits the results of the preciding experiments and the knowledge of the underlying response surface to conduct the subsequent experiments only on the most critical regions of the factor space. As a result, it drastically decreases the number of experiments. For future studies the proposed methodology will be extended to higher dimensional problems as well as higher orders of underlying response functions.

References

1 Spendley, G. R.; Hext G. R. and Himsworth, F. R. (1962) “ Sequential Application of Simplex Designs in Optimization and Evolutionary Operation” technometrics,4, 441-461.

2 Box. G.E.P. and Wilson. K. B. (1951). “On The Experimental Attainment of Optimum Conditions”. Journal of the Royal Statistical Society B 13. 1-15.

3 Friedman, M., and Savage, L. J. (1947), “Planning Experiments Seeking Maxima,” in Techniques of Statistical Analysis, eds. C. Eisenhart, M. W. Hastay, and W. A. Wallis, New York: McGraw-Hill, 365–372.

4 Daniel, C. (1973), “One-at-a-Time Plans,” Journal of the American Statistical Association, 68, 353–360.

5 Frey, D. D., F. Engelhardt, and E. Greitzer, 2003, “A Role for One-factor-at-a-time Experimentation in Parameter Design”, Research in Engineering Design 14(2), 65-74.

6 Wang, G., Dong, Z., and Aitchison, P., 2001, “Adaptive Response Surface Method — A Global Optimization Scheme for Computation-intensive Design Problems,” Journal of Engineering Optimization, 33(6) 707-734.

7 Wang, G., 2003, “Adaptive Response Surface Method Using Inherited Latin Hypercube Designs,” ASME Journal of Mechanical Design, 125(2), 210-220.

8 Stander, N. (2001), “The Successive Response Surface Method Applied to Sheet-Metal Forming”, Proceedings of the First MIT Conference on Computational Fluid and Solid Mechanics, Boston, June 12-14, 2001. Elsevier Ltd., Oxford.

9 Box, G.E.P. and Draper, N. R. (1969). “Evolutionary Operation: A Method for Increasing Industrial Productivity”. Wiley, New York, 1969.

10 Box, G. E. P. and Wilson, K.B. 1951, “On the Experimental Attainment of Optimum Conditions (with discussion)”. Journal of the Royal Statistical Society Series B 13(1),1–45.

11 T.P. Hettmansperger and J.W. McKean. 1998,” Robust Nonparametric Statistical Methods” London and New York, N.Y.: Arnold/Wiley.

12 Nagelkerke, N., 1991, “A Note on a General Definition of the Coefficient of Determination,” Biometrika, 78(3), 691-692.

13 Mao, H., 2009,”Computational Analysis of In Vivo Brain Trauma. Biomedical Engineering” Wayne State University, Detroit.

14 Mao, H., Zhang, L., Yang, K.H., King, A.I., 2006, “Application of a Finite Element Model of the Brain to Study Traumatic Brain Injury Mechanisms in the Rat”, Stapp Car Crash J 50, 583-600.