NTNU
Norwegian University of
Science and Technology

A Comparison Study of Different Optimizing Criteria and Confounding Patterns For Multi-Level Binary Replacement and Other Designs Used in Computer Experiments

Hege Grøstad Thalberg

Master of Science in Physics and Mathematics

Submission date:	July 2011
Supervisor:	John Sølve Tyssedal, MATH
Co-supervisor:	Harald Martens, CIGENE

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

As computer simulations become more popular, construction of good design of computer experiments is essential. The aim of this work is to evaluate different optimizing criteria used to obtain good designs and to compare various designs, focusing on the multi-level binary replacement (MBR) design.

This is to be carried out by conducting two comparison studies. The first should deal with the evaluation of optimizing criteria. The second should be focused on comparing various MBR designs based on different confounding patterns with other types of design of computer experiments.

The designs to be constructed and compared in addition to the MBR designs are based on latin hypercube sampling, orthogonal arrays and random sampling.

Assignment given: 14. February 2011
Supervisor: John Sølve Tyssedal

Preface

This thesis has been carried out at the Department of Mathematical Sciences at the Norwegian University of Science and Technology (NTNU), under the supervision of John Sølve Tyssedal. The thesis is the final part of the 5-years Industrial Mathematics program at NTNU and leads to the degree Master of Science.

The paper was motivated by my supervisor, John Sølve Tyssedal, and my co-supervisor Harald Martens, at The Centre for Integrative Genetics (CIGENE). It has been challenging to combine the instructions and wishes of both supervisors, as both have different backgrounds and motivations. But at the same time it has given me experience and insight in real-life working conditions, where teams often include people with different backgrounds, knowledge and interests.

I would like to thank my supervisor John Sølve Tyssedal for the guidance and help he has given on this paper. A thanks to my co-supervisor Harald Martens for his motivation and inputs on the subject. I would also like to thank Anders Nesbakken, Christian Skar and Johan N. Fatnes for reading through the paper and answering my questions during the semester.

Ås, July 11, 2011.

Hege Grøstad Thalberg

Abstract

We have constructed four different types of designs for computer experiments. The design types are based on latin hypercube sampling (LHS), orthogonal arrays (OA), random sampling and the recently proposed multi-level binary replacement (MBR) design. For each type of design we have attempted to find the best possible design out of a certain number of constructed designs using three different optimizing criteria: the alias sum of square criterion (ASSC), the L-criterion and a modified A-criterion. The chosen design has then been tested by fitting an approximate model and calculating maximum error (MAX) and root mean squared error (RMSE) values. We observed that out of the three criteria applied the ASSC performed the best.

In addition to comparing criteria for optimizing the design choice, we have also constructed non-optimized designs for comparing the different design types and the different ways of constructing MBR designs. In this setting we observed that OA designs performed well in general, whereas the MBR designs performed well when restricted to a small number of factors.

Contents

1 Introduction

When conducting experiments we want to make it less time-consuming and as cheap as possible, but at the same time extract as much information as possible from it. In an experiment we have different input variables or factors. These factors have different levels. A factor can be either quantitative, where the levels are measured on a numeric scale, for example "time" or "temperature", or they can be qualitative, where we for example can turn something on or off. Choosing the combinations of factors and levels to run in order to maximize the information obtained for experiments with time or economical restraints is the main topic for design of experiments (DoE). DoE are techniques for choosing the input combinations in which to run.

There are different types of experiments. We often distinguish between physical experiments and computer experiments. A physical experiment can be conducted in for example a laboratory, an agricultural field or a factory. These physical experiments contain random errors. Because of the random error we will not get the same result even if we do exactly the same experiment with the same input combination several times. A design for a physical experiment will often have the sample points placed along the borders of the design space.

In computer experiments we will usually get the same output, result, when the same input, combinations of factors and levels, is used in the experiment, thus we say that computer experiments are deterministic. Deterministic computer experiments will not have the random error we experience in physical experiments. Because of the lack of error in deterministic computer experiments we wish to select the combination of factors and levels differently than for physical experiments. The goal is often to minimize the bias. Bias is the difference between the true and the estimated model. Designs for computer experiments often seek to have their sample points uniformly spread on their design space; they are so called space-filling designs.

Computer simulations, or computer experiments, are popularly used as an alternative to physical experiments. Designs for computer experiments are the main focus in this thesis. We have especially focused on multi-level binary replacement (MBR) design recently presented in Martens, Måge, Tøndel, Isaeva, Høy & Sæbø (2010). Two comparison studies have been conducted:

1. We tested the use of different optimizing criteria; criteria helping us choose the best possible design out of a certain number of constructed designs. We have tested and compared three different optimizing criteria: a modified A-criterion, L-criterion and the alias sum of squares criterion (ASSC).

2. The MBR designs are constructed using fractional factorial designs. We used different ways of constructing fractional factorial designs to see if there was one giving better MBR designs than the others, and whether or not the best MBR designs are favorable to other popular designs of computer experiments.

The designs used in the above mentioned comparison studies are based on latin hypercube sampling (LHS), orthogonal array (OA) sampling and the multi-level binary replacement design. In the optimizing criteria study we also constructed random designs.

The thesis is organized in different sections, these are as follows:

2. Design of Experiments: A description of the different designs used in the comparison studies in sections 5 and 6. This section includes a short graphical description of why the sample points in computer experiments are desired to be space-filling.

3. Metamodels: After finding a design and conducting the experiment, fitting an approximated model may be desirable. For this purpose we use metamodels. There are several popular metamodels used today. In this work we have focused on polynomial models.

4. Comparison and Optimizing Criteria: In this section we present the optimizing criteria tested in the first comparison study. Also the comparison criteria used in both comparison studies are presented.

5. Comparison Study of Different Optimizing Criteria: In this section we present the comparison study of the three different optimizing criteria tested in this paper.

6. Comparison Study of Confounding Patterns for the MBR Design: In this section a comparison study of different confounding patterns for constructing fractional factorial designs, used to find MBR designs, are tested. The MBR designs are also compared to the results obtained using LHS and OA designs.

7. Discussion: A discussion of the results from the two comparison studies.

2 Design of Experiments

Design of computer experiments are as mentioned different than traditional design of experiments. In the classical designs the responses have a random error, which the responses for computer experiments do not have. Also, computer simulations are deterministic. Fang, Li & Sudjianto (2006) states a deterministic experiment as one where we get the same output when using the same input. The true model is often unknown or complex, and we seek to minimize the bias, the difference between the true model, $f(x)$, and the estimated model, $g(x)$. For these two reasons space-filling designs are popularly used for computer experiments, where all the sample points, n, are tried to be evenly spread on the design space $[0, 1]^s$, whereas the classical approach would be to place the sample points on the border of the design space. This is graphically presented in figure 1.

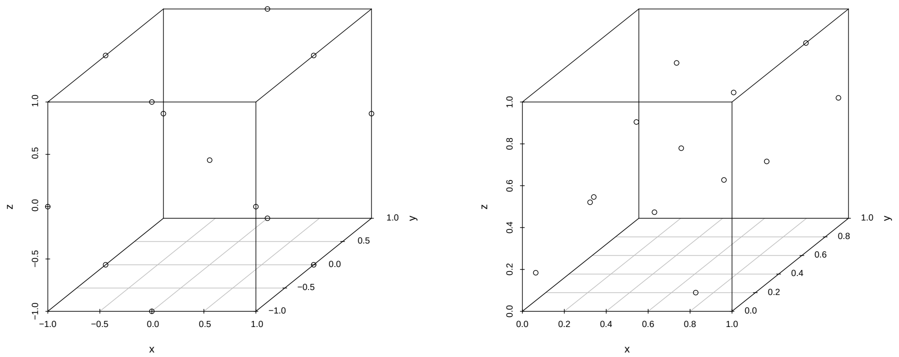
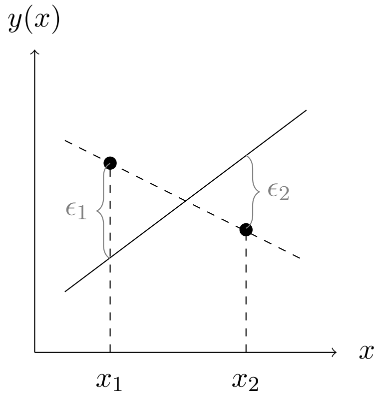


Figure 1: The plot to the left is a 3D plot of a classical Box Behnken design (BBD), while the one to the right is a LHS design. We can see the sample points in the BBD are placed in the center or on the border of the design space, while the LHS design has its samples throughout the design space, thus it is space filling.

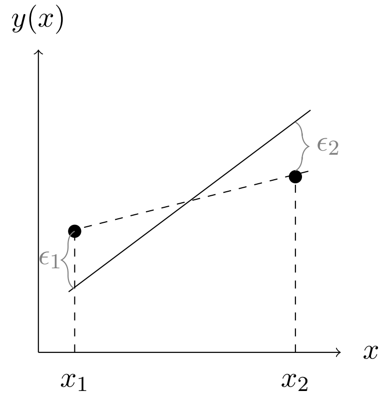
In Giunta, Wojtkiewicz Jr & Eldred (2003) a graphical explanation of why placing the sample points on the border of the design space is advantageous for designs with a random error is given, this is presented in figure 2.

A graphical presentation as to why the space-filling designs are preferable for computer experiments is given in Nesbakken (2011), this is shown in figure 3.

There already exist some methods for finding space-filling designs, such as latin hypercube sampling, orthogonal array sampling, random sampling, uniform designs and sequence designs. Further descriptions of the designs can be found in for example Fang et al. (2006) and Santner, Williams & Notz (2003). Designs used in the comparison studies presented in sections 5 and 6 are described in the following, this include designs based on latin hypercube sampling, orthogonal array sampling, random sampling and multi-level binary replacement designs.

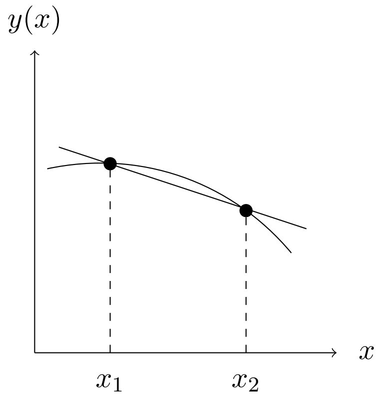


(a) This figure shows how the effect of the random error affects how well the approximated model fits the true model when the sample points are close to each other.

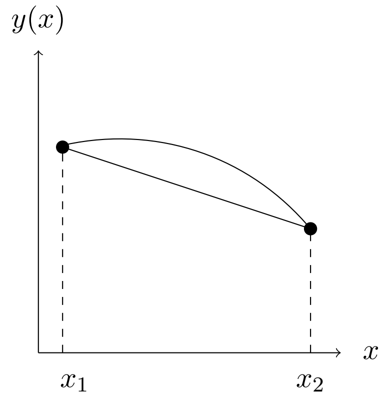


(b) The sample points are placed further to the boundaries of the design space, the effect of the random error gets smaller and the approximated model is a better fit to the true model.

Figure 2: The true linear model is represented by the solid line, and the dotted line represents the estimated linear model.



(a) This is the modern placement of the sample points, inside of the design space.



(b) This is the classical way of placing the sample points, on the border of the design space.

Figure 3: The arched line is the estimated linear model, while the solid line is the true linear model. We can see that the arched lined in the plot with sample points placed inside the design space is a better fit to the true model.

2.1 Pseudo-Monte Carlo Sampling

Pseudo-Monte Carlo sampling, or random sampling, is generated by using a pseudo random number generator algorithm to choose the desired number of samples. An advantage of the random designs is that they are easily implemented. However random designs will often leave large areas empty, especially when unable to afford many samples, since all the samples are randomly placed without any restrictions. More about pseudo-Monte Carlo sampling can be found in Giunta et al. (2003). An example is given in figure 4.

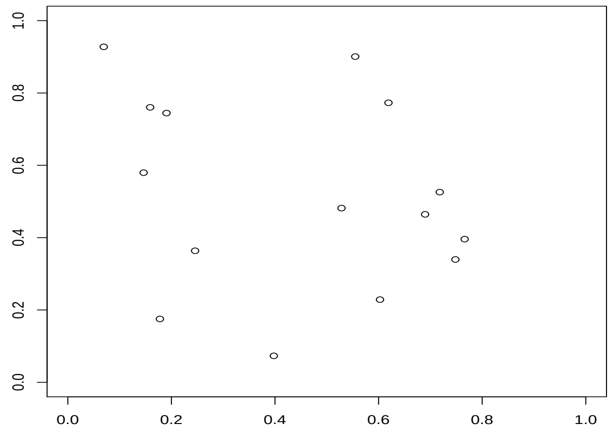


Figure 4: A plot of a two-dimensional design constructed by pseudo-Monte Carlo sampling, with $n = 15$ samples.

2.2 Fractional Factorial Design

Factorial designs have been widely used for experiments with several factors, where we want to study the joint effect of the factors, often the main and interaction effects. The factors are often assigned two levels, high and low. The levels can be represented by 1 and 0, + and −, or for factor A and B a, b and ab. By assigning each of the s factors two levels, and with each combination of high and low occurring once, we get a 2^s factorial design, where 2^s is the number of combinations. Having three factors, A, B and C we get a 2^3 design, as illustrated in table 1.

When having many factors, 2^s might combine to more runs than we wish to, or can afford to, conduct. Then we might only use a fraction of the factorial design, giving us a fractional factorial design. A fractional factorial design will in general be a $(\frac{1}{2})^p$ fraction of the full factorial design, giving 2^{s-p} runs. When having a half-fractional design $p = 1$ and we will get 2^{s-1} runs. For a one-fourth of a factorial design p is set to 2 giving 2^{s-2} runs and so on.

Run	A	B	C	Combination	A	B	C
1	$-$	$-$	$-$	(1)	0	0	0
2	$+$	$-$	$-$	a	1	0	0
3	$-$	$+$	$-$	b	0	1	0
4	$-$	$-$	$+$	c	0	0	1
5	$+$	$+$	$-$	ab	1	1	0
6	$+$	$-$	$+$	ac	1	0	1
7	$-$	$+$	$+$	bc	0	1	1
8	$+$	$+$	$+$	abc	1	1	1

Table 1: A full factorial 2^3 design, presented in three different ways of denoting the high and low levels.

When constructing and implementing a fractional factorial design we start by choosing a set of design generators. This could for three factors, A, B and C, be $C = AB$. This leads to the concept of a defining relation which in this case is $I = ABC$. Creating a one-half fractional design with these three factors, we take the $2^{3-1} = 2^2$ full factorial design, then add C by combining the two columns A and B. We can choose to use either $C = AB$ or $C = -AB$ to get a fractional design. This is illustrated in table 2. Which full factorial design we should start with depends on p. For example, if we want a one-fourth fractional design, we set $p = 2$ and we start with the full factorial design $2^{s-p} = 2^{s-2}$.

Run	A	B	A	B	$C = AB$	A	B	$C = -AB$
1	$-$	$-$	$-$	$-$	$+$	$-$	$-$	$-$
2	$+$	$-$	$+$	$-$	$-$	$+$	$-$	$+$
3	$-$	$+$	$-$	$+$	$-$	$-$	$+$	$+$
4	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$-$

Table 2: The full 2^2 factorial design is shown in the first A and B columns. The 2^{3-1} fractional design using $C = AB$ is presented in the middle part. While the last A, B and $C = -AB$ is the 2^{3-1} fractional design using the defining relation $I = -ABC$.

Traditionally the designs have been classified by a resolution and the design generators have been chosen to give the design the highest possible resolution. Resolution III, IV and V are considered to be especially important, and defined as follows:

Resolution III: In these designs two-factor interactions can be aliased with main effects or each other. But no main effects are aliased with other main effects.

Resolution IV: Two-factor interactions can be aliased with each other, but no main effects are aliased with each other or two-factor interactions.

Resolution V: For these designs two-factor interactions are aliased with three-factor interactions. While no main factor or two-factor interactions are aliased with other main factors or two-factor interactions.

In the comparison studies presented in sections 5 and 6 factorial or fractional factorial designs are not used by themselves, but as a way of finding MBR designs. The MBR designs are presented in section 2.5. The traditional way of creating fractional factorial designs may not be optimal, this is further discussed in section 2.5. More on how fractional factorial designs traditionally have been constructed and used can be found in Myers & Montgomery (1995) and R.H. Myers & Walpole (2007).

2.3 Latin Hypercube Sampling

Latin hypercube sampling (LHS) is based on Latin square sampling. A latin square is defined as:

Definition 2.1. *An $n \times n$ matrix with n symbols as its elements is called a Latin square of order n if each symbol appears in each row as well as in each column once and only once.*

An example of a latin square is:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \\ 3 & 4 & 1 & 2 \\ 2 & 3 & 4 & 1 \end{bmatrix}.$$

Latin hypercube sampling is a form of generalization of the latin square sampling to s dimensions. A latin hypercube consists of a matrix, $\mathbf{A} = n \times s$, with n rows and s columns, with n being the number of levels for each of the s factors. A design based on latin hypercube sampling will divide each of its factors into n bins of equal probability, and for each factor there will only be one sample, randomly placed, in each bin. The n samples in a LHS design is chosen following these restrictions:

1. Within each bin the sample should be randomly placed.
2. There should only be one sample in each bin for every one dimensional projection.

Fang et al. (2006) gives these equations for finding a LHS(n, s):

$$x_k^j = \frac{\pi_j(k) - U_k^j}{n} \tag{2.1}$$

where $k = 1, \ldots, n$, $j = 1, \ldots, s$ and

$$\begin{aligned} &\boldsymbol{x}_k = (x_k^1, \ldots, x_k^s) \\ &\pi_j(1), \ldots, \pi_j(n), \text{ is a permutation of the integers } 1, \ldots, n \\ &U_k^j \sim Unif(0, 1). \end{aligned}$$

In which bin the sample, $\boldsymbol{x}_k$, should be placed is determined by $\pi_j(k)$, and U_k^j determines where in the bin the sample should be placed.

When wanting to generate a LHS$(8,2)$, a LHS with 2 factors and 8 levels, we start by generating two permutations of $\{1,2,\cdots,8\}$. Then we generate a 8×2 matrix of random numbers. We could for example end up with these two matrices:

$$\boldsymbol{\pi} = \begin{bmatrix} 8 & 2 \\ 6 & 8 \\ 1 & 5 \\ 7 & 6 \\ 5 & 4 \\ 3 & 3 \\ 2 & 7 \\ 4 & 1 \end{bmatrix}, \quad \boldsymbol{U} = \begin{bmatrix} 0.4213 & 0.1180 \\ 0.1999 & 0.9688 \\ 0.2511 & 0.7912 \\ 0.9274 & 0.0448 \\ 0.1458 & 0.1037 \\ 0.0495 & 0.7447 \\ 0.5299 & 0.2753 \\ 0.7219 & 0.7936 \end{bmatrix}.$$

Using these matrices and equation (2.1) we find the LHS design to be:

$$\boldsymbol{x} = \frac{\boldsymbol{\pi} - \boldsymbol{U}}{n} = \frac{1}{8}\left[\begin{bmatrix} 8 & 2 \\ 6 & 8 \\ 1 & 5 \\ 7 & 6 \\ 5 & 4 \\ 3 & 3 \\ 2 & 7 \\ 4 & 1 \end{bmatrix} - \begin{bmatrix} 0.4213 & 0.1180 \\ 0.1999 & 0.9688 \\ 0.2511 & 0.7912 \\ 0.9274 & 0.0448 \\ 0.1458 & 0.1037 \\ 0.0495 & 0.7447 \\ 0.5299 & 0.2753 \\ 0.7219 & 0.7936 \end{bmatrix}\right] = \begin{bmatrix} 0.9473 & 0.2353 \\ 0.7250 & 0.8789 \\ 0.0936 & 0.5261 \\ 0.7590 & 0.7444 \\ 0.6068 & 0.4870 \\ 0.3688 & 0.2819 \\ 0.1838 & 0.8406 \\ 0.4098 & 0.0258 \end{bmatrix}.$$

The plot of the example LHS design is shown in figure 5.

This approach for finding designs can also provide us with designs having sample points poorly spread on the design space. For example can we end up with a design having all its samples along a diagonal, as shown in figure 6.

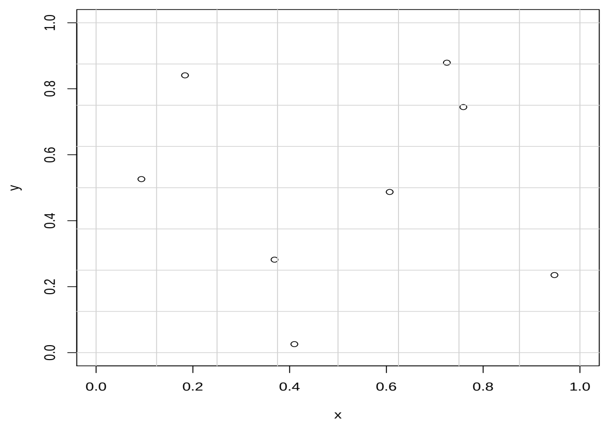


Figure 5: Plot of the example LHS. When making a projection to one dimension we can see that there is only one sample in each bin.

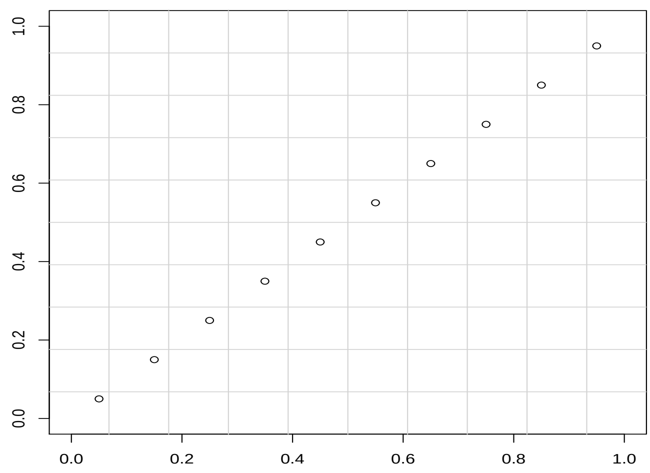


Figure 6: A plot of a bad LHS design. One can see that there is still only one sample in each bin, but they are all placed along the diagonal, leaving most of the design space empty.

2.4 Orthogonal Array Designs

Orthogonal array designs are based on orthogonal arrays, which are in Hedayat, Sloane & Stufken (1999) defined as:

Definition 2.2. *An $n \times s$ array $\boldsymbol{A}$ with entries from S is said to be an orthogonal array with p levels, strength t and index λ (for some t in the range $0 \leq t \leq s$) if every $n \times t$ subarray of $\boldsymbol{A}$ contains each t-tuple based on S exactly λ times as a row.*

S is a set containing p symbols, or levels. With s columns and n rows, there is one element per row-column pair and entries of S is a collection of ns elements of S in n rows and s columns. $\boldsymbol{A}$ is a $n \times s$ matrix, with its elements being $1, \cdots, p$ with p being the number of levels for the orthogonal array. The strength t gives information about the number of columns where all the combinations appear an equal number of times. An OA with strength $t = 1$ can be used to find a latin hypercube design. The index λ depends on the number of rows each of the n bins in the $n \times s$ matrix appears. The number of samples possible to construct for the OAs is given by the relationship $n = \lambda p^t$. An example OA with $n = 4$ and $s = 2$:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 2 & 1 \\ 2 & 2 \end{bmatrix}.$$

Each combination of 1 and 2 appear only once, and gives index $\lambda = 1$. An orthogonal design can be extended to a U-type design, an algorithm is given in Fang et al. (2006):

> For each column we replace $r = \frac{n}{p}$ ones by a permutation of $(1, \cdots, r)$, r twos by a permutation of $(r+1, \cdots, 2r)$, and so on until r ps are replaced by a permutation of $((p-1)r+1, \cdots, pr)$, giving us a U-type design.

Using the orthogonal array given above:
Then $r = \frac{n}{p} = \frac{4}{2} = 2$, and we change $r = 2$ ones by a permutation of $(1, \cdots, r) = (1, 2)$ and $r = 2$ twos by a permutation of $(r+1, \cdots, 2r) = (3, 4)$. We get the new design matrix:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 4 \\ 2 & 2 \\ 3 & 1 \\ 4 & 3 \end{bmatrix}.$$

The orthogonal array based designs perform good, but we can not use any combination of t, n, s, p and λ as we might want to - they might not exist.

2.5 Multi-Level Binary Replacement Design

The multi-level binary replacement (MBR) design was proposed by Martens et al. (2010). It uses two level fractional designs and binary coding for creating multi-level designs. And in this way the MBR design enables us to reduce the number of runs for the experiment.

If $k = 1, \cdots, s$ where s is the the number of factors, the number of levels, $L(k)$ for a variable, $\boldsymbol{x}_k$, must be set to a power of 2, such that $L(k) = 2^{M(k)}$. $M(k)$ is given when we choose the number of levels, but $M(k)$ is also the number of binary digits required for a binary representation of the $L(k)$ levels. The number of binary digits required to represent all the levels of all factors is $M_{tot} = \sum_{k=1}^{s} M(k)$.

Creating a MBR design requires three different matrices:

1. The $n \times s$ matrix $\boldsymbol{D}$ where the column, $\boldsymbol{d}_k$, containing values ranging from 0 to $(L(k) - 1)$ representing the levels of the kth factor, $\boldsymbol{x}_k$, of the s factors.
2. The $n \times M_{tot}$ matrix $\boldsymbol{F}$ containing zeros and ones which combine to be the binary representation of the matrix $\boldsymbol{D}$.
3. The $n \times M_{tot}$ matrix $\boldsymbol{G}$ where 1 and -1 represent the ones and zeros in $\boldsymbol{F}$.

An example of a multi-level binary replacement design with $s = 2$, $L(k) = 4$ for all k and $n = 4$, with the corresponding matrices given in figure 7:
The matrix $\boldsymbol{D}$ will contain digits ranging from 0 to $L(k) - 1$, that is $0 - 3$. With $L(k) = 4 = 2^{M(k)} = 2^2$, we get $M(k) = 2$ and the matrices $\boldsymbol{F}$ and $\boldsymbol{G}$ will be $n \times M_{tot} = 4 \times 4$, where each factor is represented by $M(k) = 2$ columns.

In figure 7 we see that the first row in $\boldsymbol{D}$ contains the elements 0 and 2, these elements are in $\boldsymbol{F}$ represented by $(0, 0, 1, 0)$, where $(0, 0)$ is the binary representation of

0 and $(1,0)$ is the binary representation of 2. The first row in $\boldsymbol{F}$ is further represented by $(-1,-1,1,-1)$ in the first row of the matrix $\boldsymbol{G}$. In this example $M(k)$ is the same for every factor k, but the design may also be applied when the different factors have a different number of levels, giving different $M(k)$s. When constructing a MBR design we use fractional factorial designs, from section 2.2, to find the matrix $\boldsymbol{G}$. An example using a full factorial design is shown in figure 8.

$$\boldsymbol{D} = \begin{bmatrix} 0 & 2 \\ 1 & 1 \\ 2 & 0 \\ 3 & 3 \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{F} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{G} = \begin{bmatrix} -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

(a) (b) (c)

Figure 7: An example of how a $\boldsymbol{D}$, $\boldsymbol{F}$ and $\boldsymbol{G}$ can be, for $s=2$, $n=4$ and $L(k)=4$.

$$\boldsymbol{D} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 2 & 0 \\ 2 & 1 \\ 0 & 2 \\ 0 & 3 \\ 2 & 2 \\ 2 & 3 \\ 1 & 0 \\ 1 & 1 \\ 3 & 0 \\ 3 & 1 \\ 1 & 2 \\ 1 & 3 \\ 3 & 2 \\ 3 & 3 \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{F} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{G} = \begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

(a) (b) (c)

Figure 8: An example of how a $\boldsymbol{D}$, $\boldsymbol{F}$ and $\boldsymbol{G}$ can be, for $s=2$, $L(k)=4$ and $n=16$.

When implementing a MBR design we start by choosing a set of design generators for constructing a fractional factorial design, matrix $\boldsymbol{G}$. We name the $s=4$ columns in $\boldsymbol{G}$ a_1, a_2, b_1 and b_2, and let them be the binary representations of $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$. For our example, in figure 9, we have used the design generator $b_2 = a_1 a_2 b_1$, giving us the highest possible resolution, to construct the fractional factorial design in matrix $\boldsymbol{G}$. Further we change all the -1s in $\boldsymbol{G}$ to 0s, and we obtain $\boldsymbol{F}$. We then combine columns a_1 and a_2 in $\boldsymbol{F}$ to obtain the first column in $\boldsymbol{D}$, and likewise use b_1 and b_2 to obtain the second column in $\boldsymbol{D}$.

Looking at figure 9 we see the third row of $\boldsymbol{G}$, $\boldsymbol{G}_{3\cdot} = (-1, 1, -1, -1)$ is transformed to $\boldsymbol{F}_{3\cdot} = (0, 1, 0, 0)$. We then convert the binary representation in $\boldsymbol{F}_{3\cdot}$ to $\boldsymbol{D}_{3\cdot} = (1, 0)$, by taking the first two elements in $\boldsymbol{F}_{3\cdot}$ $(0, 1)$ to be the binary representation of 1 and the last two elements $(0, 0)$ to be the binary representation of 0. A plot of the MBR design based on the half-fraction design is shown in figure 10.

One question when constructing multi-level binary replacement designs is which design generator to choose. Because we use fractional factorial designs as a way of finding MBR designs the design generators giving the best possible resolution is not necessarily the best to use. Therefore two other methods for finding confounding patterns are also used. We start by constructing a 2^{s-p} full factorial design, before we find confounding patterns:

1. The confounding pattern is found by first combining all the $s - p$ columns, then combining $s - p - 1$ of the $s - p$ columns, then $s - p - 2$ and so on. If $s = 5$ and $p = 2$, giving us a one-fourth fractional design, the 2^3 full factorial design will provide us with 3 columns, say A, B and C. We then set column 5 to be a combination of columns A, B and C, ABC. Column 4 will be a combination of $s - p - 1 = 5 - 2 - 1 = 2$ of the first three columns, so either AB, AC or BC. This will leave us with a fractional factorial design of resolution III.

2. The second way is similar to the first, but only combinations of a odd number of the first $s - p$ columns are being used. This will give us a resolution IV fractional factorial designs.

We wish to use different confounding patterns for constructing fractional factorial designs to see if there is one specific giving us better MBR designs than the others. We have earlier suggested that using fractional factorial designs having the highest possible resolution may not give the best MBR designs, so in order to compare we have also used traditional highest resolution confounding patterns in section 5 and 6 for comparison.

$$\boldsymbol{G} = \begin{bmatrix} -1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}$$

(a) The half-fraction design, $\boldsymbol{G}$.

$$\boldsymbol{F} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

(b) The half-fraction design, $\boldsymbol{G}$, changed back to $\boldsymbol{F}$.

$$\boldsymbol{D} = \begin{bmatrix} 0 & 1 \\ 0 & 2 \\ 1 & 0 \\ 1 & 3 \\ 2 & 0 \\ 2 & 3 \\ 3 & 1 \\ 3 & 2 \end{bmatrix}$$

(c) The final design matrix, $\boldsymbol{D}$.

Figure 9: How a half-fraction design and the MBR design constructed from it look like when using the design generator $b_2 = a_1 a_2 b_2$.

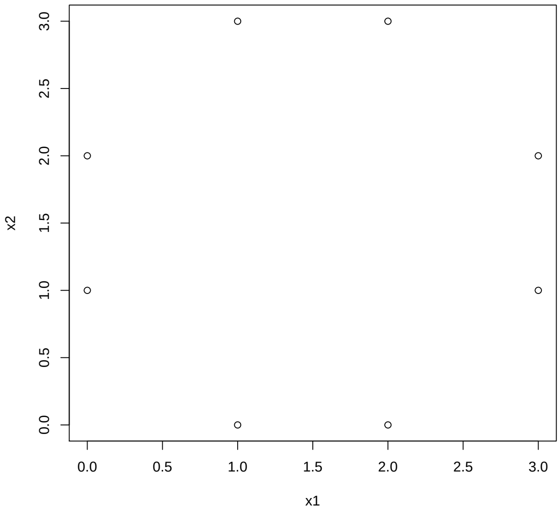


Figure 10: Plot of the example MBR design in figure 9 using the confounding $b_2 = a_1a_2b_1$ to obtain the fractional factorial design $\boldsymbol{G}$.

When making a projection of a multi-level binary replacement design the samples are often arranged in a symmetrical pattern, as shown in figure 11; if we draw a line at $x_1 = 0.5$ we can see that the pattern is symmetrical about it.

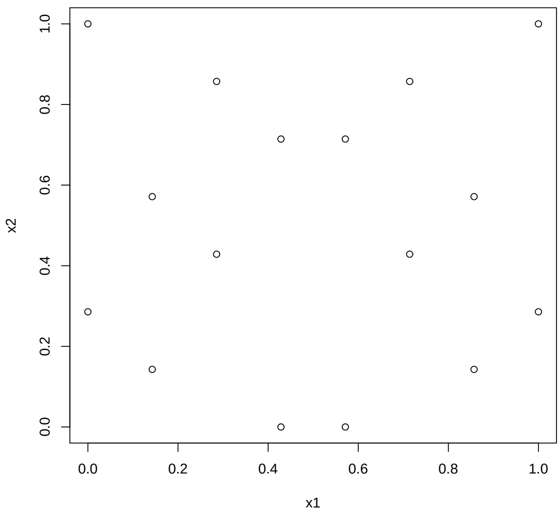


Figure 11: One of the two-dimensional projections of a five factor design with 8 levels, scaled down to $[0, 1]^2$.

3 Metamodels

After constructing a design and performing the computer experiment on the chosen sample points, we are often interested in finding an approximated model for the true model that might be complex or unknown. This is often referred to as a metamodel. In a deterministic experiment the relationship between the input and the output can be given as:

$$\text{output variables} = f(\text{input variables}). \tag{3.1}$$

In equation 3.1 f is the function we want to approximate by using a metamodel. A graphical representation of the relationship between the approximated model, $\hat{y}$, and the true model, y, is given in Fang et al. (2006) and shown figure 12.

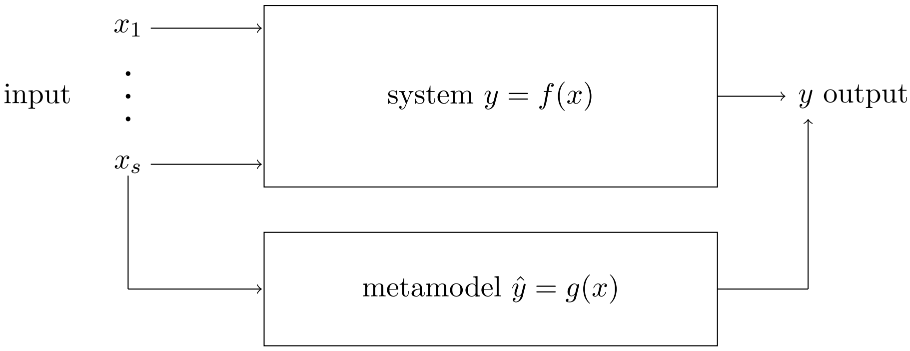


Figure 12: A graphical presentation of a metamodel in computer experiments.

The goal of a metamodel is often that the model should give insight into the relationship between the input and the output, it should be close to the real model, but faster to run, and it should give information about the untried points.

In Praveen & Duvigneau (2007) metamodels are categorized as either global or local. A global model uses all the available data, and are often used to replace the true model. A local model will only use the small set of data surrounding the point to be approximated, these models are often used as preconditioners to intensify the exploration of the area being investigated.

There are several different metamodels one can use, for example polynomial models, kriging models, splines methods and radial basis functions. In this comparison study we have chosen to use polynomial models, which are presented below. Information about other metamodels can be found in Fang et al. (2006) and Simpson, Peplinski, Koch & Allen (2001).

3.1 Polynomial Models

Response surface models, or polynomial models, have been popularly used for physical experiments and are also popular in computer experiments. A second order polynomial model:

$$g(x) = \beta_0 + \sum_{i=1}^{s} \beta_i x_i + \sum_{i=1}^{s} \sum_{j=i}^{s} \beta_{ij} x_i x_j. \tag{3.2}$$

Where the βs often are estimated by using least squares regression. Three reasons for the extensive use of polynomial models are given in Myers & Montgomery (1995):

1. The second order polynomial is flexible and can take on several different functional forms. Because of this it will be a good approximation to the true response surface.

2. The β's are easily estimated, the least squares method is often used for this estimation.

3. There are several practical experiences indicating that the second-order models work well in solving real response surface problems.

There are also drawbacks to using polynomial models in computer experiments. Fang et al. (2006) states that situations with highly nonlinear or irregular models to be unfitting when using polynomial models. In this thesis a third order polynomial model is chosen as the true model, while a second order polynomial model without squared terms is fitted as the approximated model.

4 Comparison and Optimizing Criteria

When consctructing designs, conducting experiments and fitting metamodels we need a way of retrieving information to know whether or not the design we chose was good, and if the metamodel we fitted was a good fit for the true model. Sometimes we might want a criterion to help us choose which design to use, or to optimize designs. The criterion may say something about the design with regards to one or several criteria defined by the user. Or the criterion may be a measure of how uniformly spread, space-filling, the sample points are.

4.1 Optimality Criteria

In classical design of experiments different optimality criteria have been popularly used. In Fang et al. (2006) some are shortly explained. The relationship between the response y_k and the input factors $\boldsymbol{x}_k = (x_{k1}, \cdots, x_{ks})$ is expressed in a regression model:

$$y_k = \sum_{j=1}^{m} \beta_j g_j(x_{k1}, \cdots, x_{ks}) + \epsilon_k = \sum_{j=1}^{m} \beta_j g_j(\boldsymbol{x}_k) + \epsilon_k, \ \ k = 1, \cdots, n \tag{4.1}$$

$g_j(\boldsymbol{x}_k)$ are prespecified or known functions and ϵ is the random error. If we let:

$$\boldsymbol{G} = \begin{bmatrix} g_1(\boldsymbol{x}_1) & \cdots & g_m(\boldsymbol{x}_1) \\ \vdots & & \vdots \\ g_1(\boldsymbol{x}_n) & \cdots & g_m(\boldsymbol{x}_n) \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}.$$

the model in 4.1 can be written as:

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{\beta} + \epsilon.$$

Where the matrix $\boldsymbol{G}$ is the *design matrix* and the matrix $\boldsymbol{M} = \boldsymbol{G}'\boldsymbol{G}$ is the *information matrix*. The covariance matrix of the least squares estimator:

$$Cov(\hat{\boldsymbol{\beta}}) = \sigma^2 \boldsymbol{M}^{-1}.$$

Several optimality criteria have been suggested by different authors, some are:

1. D-optimality: maximize the determinant of $\boldsymbol{M}$.
2. A-optimality: minimize the trace of $\boldsymbol{M}^{-1}$.
3. E-optimality: minimize the largest eigenvalue of $\boldsymbol{M}$.

These criteria may favor classical designs with sample points on the border of the design space.

4.1.1 Modified A-criterion

In the comparison study in section 5.1 we use a modified A-criterion as an optimizing criterion. The A-optimality criteria seeks to minimize the trace of the inverse of the information matrix, $(\boldsymbol{X}'\boldsymbol{X})$. The matrix $\boldsymbol{X}$ is constructed from the design , $\boldsymbol{D}$, and the columns of $\boldsymbol{X}$ are based on the approximated model. If the metamodel is:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2.$$

Further, we use the LHS design constructed in section 2.3. The matrix $\boldsymbol{X}$ will then have the columns x_1 and x_2 from $\boldsymbol{D}$. The last column will be $x_1 x_2$:

$$\boldsymbol{D} = \begin{bmatrix} 0.9473 & 0.2353 \\ 0.7250 & 0.8789 \\ 0.0936 & 0.5261 \\ 0.7590 & 0.7444 \\ 0.6068 & 0.4870 \\ 0.3688 & 0.2819 \\ 0.1838 & 0.8406 \\ 0.4098 & 0.0258 \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 0.9473 & 0.2353 & 0.2229 \\ 0.7250 & 0.8789 & 0.6372 \\ 0.0936 & 0.5261 & 0.0492 \\ 0.7590 & 0.7444 & 0.5650 \\ 0.6068 & 0.4870 & 0.2955 \\ 0.3688 & 0.2819 & 0.1040 \\ 0.1838 & 0.8406 & 0.1545 \\ 0.4098 & 0.0258 & 0.0106 \end{bmatrix}.$$

We find the eigenvalues, λ_i, of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ and the modified A-criterion, $\mathrm{A}_{\mathrm{mod}}$, is:

$$\left(\sum_i \frac{1}{\lambda_i} \right) \lambda_{\max}.$$

This criterion gives us information about how many directions in the matrix $\boldsymbol{X}$ are badly spanned. We wish for it to be as many directions as possible that are well spanned, and therefore the criterion should be minimized.

4.1.2 L-criterion

Another criterion, with similar calculations as the modified A-criterion, was suggested in Tøndel, Gjuvsland, Måge & Martens (2010). In this work it is referred to as the L-criterion. We find the eigenvalues, λ_i, of the inverse of the information matrix, $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. The formula for the L-criterion is then:

$$\frac{\sum_i \lambda_i}{\lambda_{\max}}.$$

This criterion should be maximized as it gives information about how many directions in the matrix $\boldsymbol{X}$ that are well spanned. This is used as an optimizing criterion in section 5.2.

4.2 Alias Sum of Squares Criterion

The Alias Sum of Squares Criterion (ASSC) was proposed by Bursztyn & Steinberg (2004) as an alternative to already existing criteria, like the RMSE and the MAX. The criterion is based on the alias matrix for a simple approximation model.

Firstly we fit a simple regression model:

$$\boldsymbol{y}_i = \beta_0 + \sum_{j=1}^{p} \beta_j \boldsymbol{x}_{ij} = f(\boldsymbol{x}_i)'\boldsymbol{\beta}. \tag{4.2}$$

Assuming that this model is not providing a good enough description of the output data, we add extra terms to the model in 4.2, giving us:

$$\boldsymbol{y}_i = \beta_0 + \sum_{j=1}^{p} \beta_j \boldsymbol{x}_{ij} + \sum_{j=p+1} \beta_j \boldsymbol{f}_{2,j}(\boldsymbol{x}_i) = \boldsymbol{f}'(\boldsymbol{x}_i)\boldsymbol{\beta} + \boldsymbol{f}_2'(\boldsymbol{x}_i)\boldsymbol{\beta}_2. \tag{4.3}$$

Which can be written in matrix form:

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{f}'(\boldsymbol{x}_1) \\ \vdots \\ \boldsymbol{f}'(\boldsymbol{x}_n) \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{f}_2'(\boldsymbol{x}_1) \\ \vdots \\ \boldsymbol{f}_2'(\boldsymbol{x}_n) \end{bmatrix} \boldsymbol{\beta}_2 = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{X}_2\boldsymbol{\beta}_2.$$

In this model the $\boldsymbol{X}_2\boldsymbol{\beta}_2$ represents the extra terms, which were not in the original model in 4.2. The least-squares estimator for $\boldsymbol{\beta}$ is found from:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}_2\boldsymbol{\beta}_2 = \boldsymbol{\beta} + A\boldsymbol{\beta}_2.$$

Here $\boldsymbol{A}$ is the alias matrix. The alias matrix gives us information about how the original model, equation 4.2, is biased by the extra terms included in the full model, equation 4.3, and by studying the alias matrix we can say something about the design's effectiveness. Further Bursztyn & Steinberg (2004) uses a method for turning the bias from terms not included in the model into variance, enabling us to use design criteria based on variance, such as A-optimality. Assuming that $\boldsymbol{\beta}_2$ is random and normally distributed, with covariance $\sigma^2\boldsymbol{I}$ and mean 0. Using the lack of error in computer experiments, we now get:

$$Cov(\hat{\boldsymbol{\beta}}) = \sigma_\beta^2 \boldsymbol{A}\boldsymbol{A}'.$$

$\boldsymbol{A}$ is still the alias matrix, and $\boldsymbol{A}\boldsymbol{A}'$ measures the extent to which the design allows higher order bias to affect the simple approximation. Using the A-optimality criterion from $Cov(\hat{\boldsymbol{\beta}})$, we find the alias sum of squares criterion:

$$A = tr(Cov(\hat{\boldsymbol{\beta}})) = \sum_{i,j} a_{i,j}^2.$$

The alias sum of squares criterion is in section 5 used as both an optimizing and a comparison criterion. In section 6 it is used as a comparison criterion.

4.3 IMSE, RMSE and MAX criteria

Popular methods for comparing designs are the integrated mean square error (IMSE), the root mean square error (RMSE) and the maximum error (MAX) criteria. These criteria measures how well the estimated metamodel, $\hat{y}(\boldsymbol{x})$ fits the true model, $y(\boldsymbol{x})$, by using additional validation points other than the sample data used for estimating the metamodel. The IMSE is given in Bursztyn & Steinberg (2004) as:

$$IMSE = \int E\{y(\boldsymbol{x}) - \hat{y}(\boldsymbol{x})\}^2 d\boldsymbol{x}.$$

The RMSE and MAX criteria are easier to calculate and stated in Simpson, Lin & Chen (2001):

$$MAX = \max\{|y_i - \hat{y}_i|\}_{i=1,\cdots,n_{\text{error}}}$$
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_{\text{error}}}(y_i - \hat{y}_i)^2}{n_{\text{error}}}}.$$

We can use the same validation points, n_{error}, for both the MAX and the RMSE criteria, enabling us to calculate them at the same time. The RMSE criterion measures an average squared deviance, and therefore gives a good estimate for the "global error" for the chosen region. While the MAX criterion gives a good estimate for the maximum "local error", as it measures the largest difference between the estimated and the true model.

The RMSE and MAX criteria have in sections 5 and 6 been used as criteria for comparing different designs. In section 5 it has also been used for checking how well the optimizing criteria, ASSC, L-and A_{mod}-criteria, performs. The optimizing criteria have been compared by checking how well they coincide with the RMSE and MAX criteria.

5 Comparison Study of Different Optimizing Criteria

In this section a comparison of different designs, optimizing criteria and comparison criteria has been conducted. The designs, comparison criteria and optimizing criteria will be further presented after a short description on how the comparison has been performed:

1. 100 designs have been constructed, either LHS, OA or MBR designs.
2. An optimizing criterion, either ASSC, A_{mod}- or L-criterion, has been used for finding the best, optimized, design out of the 100 designs constructed.
3. A second-order model without quadratic terms has been fitted using the design chosen in step 2.
4. The comparison criteria have been calculated, to compare the different types of designs and optimizing criteria.
5. Steps $1-4$ have been repeated 1000 times.
6. The maximum, minimum, mean and median of the comparison criteria have been found and are presented in tables.

The designs constructed are LHS, OA, MBR and random designs. A short description on the construction of the different designs, all having $s = 5$ factors, $n = 32$ samples and are scaled to $[0, 1]^s$:

The MBR designs have been constructed with $L(k) = 8$ levels and using three different ways of finding confounding patterns, as explained in section 2.5:

1. all-confounding: The first confounding pattern explained in section 2.5, using combinations of $s-p$, $s-p-1$, $s-p-2$ and so on. So if we need $2^{s-p} = 2^5 = 32$ samples, we would first use the combination of all 5 original factors, $ABCDE$, then we would use a combinations of 4, for example $ABCD$ or $ACDE$. When there are no more combinations of 4 left we would use combinations of 3, for example ABC or CDE. All combinations are used in a descending order until we have constructed the desired number of columns.
2. odd-confounding: The second method presented in section 2.5, with only combinations of a odd number of the first $s-p$ columns. As above, if we need $2^5 = 32$ samples we would start by combining all 5, $ABCDE$, but we would then skip combinations of 4 and use combinations of 3 instead, for example BCE.
3. Resolution IV: Classical resolution IV confounding. This is similar to the odd-confounding, but it is the confounding pattern the statistical software Minitab would recommend. The exact combinations can be found in Appendix A.

For the all- and odd-confounding we may have more combinations of the $s-p$ columns than needed, and which combinations to use have randomly been drawn for each design constructed. For all of the confounding patterns we have mixed up the columns in the fractional design, such that the combination of columns for each MBR-factor will differ every time.

The LHS designs have been constructed as described in section 2.3.

The random designs have been constructed as described in section 2.1.

The OA designs have been constructed as described in section 2.4, using both $p=2$ and $p=4$.

Replicate observations at the same input values for a computer experiment provides us with the same output. Therefore we try to avoid replicated samples. For all the designs we have checked for replicated samples, and excluded designs having replicated sample points in one or several of the two-dimensional projections.

For each simulation there have been constructed 100 designs of which one design has been chosen by using either modified A-criterion, L-criterion or the alias sum of squares criterion. A short description of the different optimizing criteria:

Modified A-criterion: The design $\boldsymbol{D}$ has been used, including all the interaction terms in the approximated model, let us name this matrix $\boldsymbol{X}$. We found $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ and its eigenvalues, λ_i, then we calculated $\sum_i \left(\frac{1}{\lambda_i}\right)\lambda_{\max}$, this value was minimized. As presented in section 4.1.1.

L-criterion: Also here the design $\boldsymbol{D}$ and all the interaction terms in the approximated model have been used. We then found the design giving the highest value of $\frac{\sum_i \lambda_i}{\lambda_{\max}}$, where λ_i are the eigenvalues of the matrix $\boldsymbol{X}$. As presented in section 4.1.2.

Alias sum of squares criterion: As explained in section 4.2. The criterion uses the true and the approximated model for estimating the design's skewness. This criterion is also used as a comparison criterion when the A_{mod}- or L-criterion are used as optimizing criteria.

We then ran 1000 simulations for each combination of design and optimizing criteria. We used a third-order model as the "true" model, with βs chosen to be between 0 and 1:

$$y=\sum_{i=1}^{s}(\beta_i x_i)+\sum_{i=1}^{s}(\beta_{ii}x_i^2)+\sum_{i=1}^{s}\sum_{j=i}^{s}(\beta_{ij}x_ix_j)+\sum_{i=1}^{s}\sum_{j=i}^{s}\sum_{k=j}^{s}(\beta_{ijk}x_ix_jx_k).$$

The exact true model used can be found in appendix A. A second-order model without quadratic terms is used as the approximated model:

$$\hat{y}=\beta_0+\sum_{i=1}^{s}(\beta_i x_i)+\sum_{i=1}^{s}\sum_{j=i+1}^{s}(\beta_{ij}x_ix_j).$$

After fitting a first-order model we calculated the MAX and RMSE criteria, when using the A_{mod}- or L-criterion the ASSC was also calculated. The MAX and the RMSE criteria have been calculated using a grid of 10 samples per factor. Short descriptions of the comparison criteria used are presented here, more thorough descriptions are given in section 4:

The MAX criterion gives a number for the maximum deviation between the true and the approximated model, giving a measure for the maximum "local error".

The RMSE criterion estimates an average squared deviance between the true and approximated model, and therefore gives a good measure for a "global error".

The ASSC only uses the true and the approximated model for estimating the design's skewness, this makes it computationally much cheaper than the other two criteria used. The ASSC is also used as an optimizing criterion.

For each combination of design, optimizing criterion and comparison criteria, we have found the maximum, minimum, mean and median value of the comparison criteria, these are presented in tables. The values for the optimizing criterion used are given in parentheses for the max and min values. The best and worst value for the min, max, mean and median are shown in green and red.

The comparisons are presented in three sections, sections 5.1 to 5.3, with a short section summary after.

5.1 Optimizing Using the Modified A-criterion

Here we have found the optimal, or best, design, from 100 designs constructed, using the modified A-criterion, A_{mod}. We found the eigenvalues for $(\boldsymbol{X}'\boldsymbol{X})^{-1}$, with $\boldsymbol{X}$ being the design including all the terms included in the metamodel. Then the design having the smallest value for the A_{mod}, $\sum_i \frac{1}{\lambda_i}\lambda_{max}$, was used for fitting a model.

The tables 4, 5 and 6 contain the minimum, maximum, mean and median values of the comparison criteria, ASSC, MAX and RMSE.

As we see in tables 4, 5 and 6, the designs getting the best, lowest, A_{mod} value does not necessarily score the best for the MAX, RMSE or ASSC criteria. For example, for the RMSE comparison criterion, the MBR_{IV} which gets the smallest RMSE minimum value, 0.4699974, has an A_{mod}-criterion value 2197.984. While the MBR_{all} design achieving the smallest A_{mod}-criterion value 1690.121, gets a RMSE score of 0.482867, which is only the third lowest RMSE score even though its A_{mod}-criterion value is the smallest.

We can also see this for several other designs. In the table showing the RMSE scores, we see that for the $OA_{p=4}$ design achieving the smallest maximum value, its corresponding A_{mod}-value, 3038.156, is only the fourth best. We also notice that this A_{mod}-value is better than the A_{mod}-value for the its $RMSE_{min}$ value.

When looking at the correlation between the A_{mod}-criterion and the comparison criteria in table 3 we see that the correlation coefficients for the A_{mod} and the MAX and RMSE criteria are small and no larger than 0.2, indicating little correlation. In table 59 in Appendix A we see that the correlation coefficients for the $OA_{p=4}$ designs are close to zero, which corresponds to what we observed in table 5.

The observations we make when looking at tables 4, 5 and 6, with the comparison criteria, and table 3, with the correlation coefficients, indicate that the A_{mod}-criterion is not suitable as an optimizing criterion.

	MAX	RMSE	ASSC
minimum	0.066	0.024	−0.056
maximum	0.201	0.172	0.357

Table 3: The minimum, closest to 0, and maximum, the furthest from 0, correlation coefficients for the A_{mod}-criterion and the comparison criteria, MAX, RMSE and ASSC, of the different designs constructed.

	LHS	$OA_{p=2}$	$OA_{p=4}$	Random	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	4.131447	3.925025	3.432505	4.291686	3.635808	3.322843	3.765613
min	2.599731	2.652109	2.548500	2.729632	2.359470	2.314869	2.380221
	(2970.177)	(3006.951)	(2814.682)	(2929.735)	(2392.775)	(1700.983)	(2455.939)
max	8.301431	7.236691	4.946750	7.276748	6.684528	5.139390	7.846681
	(3659.466)	(3173.540)	(3060.359)	(3126.222)	(2931.731)	(1634.082)	(3213.368)
median	4.023349	3.828872	3.394110	4.187709	3.577639	3.247354	3.668319

Table 4: The MAX criterion values for designs with $s = 5$ factors and $n = 32$ samples, using the A_{mod}-criterion for finding optimized designs.

	LHS	$OA_{p=2}$	$OA_{p=4}$	Random	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	0.658116	0.642581	0.581135	0.659015	0.653164	0.620423	0.667289
min	0.518703	0.530831	0.512237	0.490878	0.473262	0.482867	0.469974
	(3314.246)	(2612.550)	(3279.480)	(3619.886)	(1879.357)	(1690.121)	(2197.984)
max	1.146706	0.966885	0.718995	1.072721	0.939626	0.783114	1.095732
	(3659.466)	(3173.540)	(3038.156)	(2992.410)	(2946.776)	(1657.097)	(3213.368)
median	0.646132	0.635631	0.576556	0.645751	0.652748	0.616332	0.661194

Table 5: The RMSE criterion values for designs with $s = 5$ factors and $n = 32$ samples, using the A_{mod}-criterion for finding optimized designs.

	LHS	$OA_{p=2}$	$OA_{p=4}$	Random	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	57.341207	57.222591	63.478888	54.413684	73.163148	54.494722	73.511939
min	45.002839	48.248426	53.980540	40.481022	50.890384	46.720602	48.064196
	(2903.790)	(2479.789)	(3158.195)	(2290.126)	(3214.947)	(1872.645)	(2769.008)
max	74.379342	72.780095	77.880264	78.736714	104.685654	65.335432	104.036678
	(4128.429)	(3094.317)	(3155.248)	(4080.463)	(2411.498)	(1685.445)	(3190.872)
median	56.919332	56.956369	63.296201	54.096216	73.196542	54.225541	73.134565

Table 6: The alias sum of squares criterion values for designs with $s = 5$ factors and $n = 32$ samples, using the A_{mod}-criterion for finding optimized designs.

5.2 Optimizing Using the L-criterion

In this section we have optimized designs using the L-criterion, which is similar to the modified A-criterion used in section 5.1. We found the eigenvalues of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ and maximized $\frac{\sum_i \lambda_i}{\lambda_{\max}}$, the values should lie between 1 and the number of columns in matrix $\boldsymbol{X}$, it tells us something about how many directions of the matrix $\boldsymbol{X}$ that are well spanned. The MAX, RMSE and ASSC scores are presented in tables 8, 9 and 10.

For the L-criterion we observe that the best, highest, L-criterion does not necessarily give the best values for the comparison criteria, MAX, RMSE and ASSC. We see that also here the $OA_{p=4}$ designs achieve low maximum values for the RMSE and MAX criteria, but the L-criterion score is best for the maximum values, and not for the minimum values as we might expect.

When looking at the correlation coefficients for the L-criterion and the comparison criteria in table 7 we see that the values are close to 0 which suggests that they are not dependent on each other. The correlation coefficients for the L-criterion actually indicates that it is slightly less correlated with the comparison criteria than what the A_{mod}-criterion is, and is therefore less suitable as an optimizing criterion.

	MAX	RMSE	ASSC
minimum	−0.043	−0.030	−0.008
maximum	−0.174	−0.143	−0.263

Table 7: The minimum, closest to 0, and maximum, the furthest from 0, correlation coefficients for the L-criterion and the comparison criteria, MAX, RMSE and ASSC, out of all the different designs constructed.

	LHS	$OA_{p=2}$	$OA_{p=4}$	Random	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	4.275472	3.937303	3.497350	4.443785	3.768577	3.361846	3.843266
min	2.782363	2.616182	2.538301	2.394242	2.437575	2.219867	2.397051
	(5.164)	(4.771)	(4.654)	(4.011)	(4.486)	(4.774)	(4.326)
max	8.596571	6.542996	4.969484	9.267712	7.885090	5.761226	6.189537
	(4.167)	(4.781)	(4.982)	(4.273)	(3.557)	(4.585)	(3.825)
median	4.149602	3.827858	3.451766	4.312303	3.665046	3.278401	3.750165

Table 8: The MAX criterion values for designs with $s = 5$ factors and $n = 32$ samples, using the L-criterion for finding optimized designs.

	LHS	$OA_{p=2}$	$OA_{p=4}$	Random	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	0.681680	0.644157	0.587358	0.711792	0.662565	0.627841	0.667940
min	0.527069	0.525353	0.513642	0.509215	0.472632	0.488461	0.488824
	(4.467)	(4.962)	(4.684)	(4.529)	(4.226)	(4.519)	(4.014)
max	1.187969	0.894430	0.767802	1.266943	1.066760	1.097359	1.005561
	(4.167)	(5.407)	(4.805)	(4.273)	(3.557)	(4.405)	(3.696)
median	0.668025	0.633511	0.581310	0.696098	0.661441	0.624002	0.665869

Table 9: The RMSE criterion values for designs with $s = 5$ factors and $n = 32$ samples, using the L-criterion for finding optimized designs.

	LHS	$OA_{p=2}$	$OA_{p=4}$	Random	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	58.969301	58.184858	63.923020	60.194694	73.852831	55.382570	74.429249
min	47.055662	48.325657	54.700145	42.654354	51.901325	47.603427	50.658065
	(4.611)	(5.074)	(4.899)	(4.218)	(4.573)	(4.975)	(4.331)
max	89.802471	74.746357	73.281151	98.029878	107.153832	72.781493	107.187589
	(4.167)	(4.937)	(4.714)	(4.345)	(3.900)	(4.622)	(3.593)
median	58.260139	57.792906	63.829025	58.966887	73.305786	55.042384	74.105979

Table 10: The alias sum of squares criterion values for designs with $s = 5$ factors and $n = 32$ samples, using the L-criterion for finding optimized designs.

5.3 Optimizing Using the Alias Sum of Squares Criterion

In the last part of this section on optimizing criteria, we have used the alias sum of squares criterion, presented in section 4.2, as the criterion for optimizing designs, thus the ASSC has not been used as a comparison criterion in this section.

When looking at the results in tables 12 and 13, we see that the comparison criteria and the optimizing criterion's values coincide better for each design than the L-and A_{mod}-criteria did. All designs' minimum comparison criteria values have a corresponding ASSC value smaller than the ASSC value corresponding to the maximum value, except for the LHS design combined with the RMSE criterion. When studying the tables closer we see that the designs with the best ASSC does not necessarily score the best for the MAX or RMSE criterion value when compared to the other design types constructed.

Looking at the correlation values in table 11 we see that the ASSC correlates better with both the RMSE and MAX criteria than the L-and A_{mod}-criteria did. However, also for the ASSC as an optimizing criterion we can see that the smallest correlation is close to 0. In table 61 in Appendix A we see that the MBR designs optimized using the alias sum of squares criterion, independent of which confounding pattern used, correlate better with both the MAX and RMSE than the other four designs does, especially for the RMSE criterion.

	MAX	RMSE
minimum	0.019	0.020
maximum	0.244	0.427

Table 11: The minimum, closest to 0, and maximum, the furthest from 0, correlation coefficients for the ASSC and the comparison criteria, MAX and RMSE, out of all the different designs constructed.

	LHS	$OA_{p=2}$	$OA_{p=4}$	Random	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	4.035824	3.770230	3.416689	4.366887	3.580889	2.891271	3.764362
min	2.729728	2.599056	2.639426	2.796449	2.351988	2.088877	2.423959
	(49.182)	(49.735)	(54.745)	(48.173)	(54.724)	(49.466)	(55.292)
max	6.527050	5.679290	5.201257	7.963170	5.455832	4.560252	5.575570
	(52.555)	(52.633)	(59.724)	(50.068)	(64.537)	(50.370)	(59.907)
median	3.987381	3.726373	3.368958	4.259059	3.516050	2.855529	3.712280

Table 12: The MAX criterion values for designs with $s = 5$ factors and $n = 32$ samples, using the ASSC for finding optimized designs.

	LHS	$OA_{p=2}$	$OA_{p=4}$	Random	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	0.639345	0.611610	0.588885	0.667700	0.587158	0.533421	0.585866
min	0.518048	0.506519	0.508767	0.487658	0.468214	0.469541	0.466042
	(52.932)	(53.341)	(55.840)	(48.040)	(53.172)	(46.826)	(62.328)
max	0.835455	0.768908	0.727545	1.185298	0.818280	0.708567	0.821757
	(52.001)	(53.525)	(59.724)	(50.068)	(64.506)	(50.370)	(64.352)
median	0.631031	0.607548	0.585909	0.660929	0.575255	0.527599	0.573378

Table 13: The RMSE criterion values for designs with $s = 5$ factors and $n = 32$ samples, using the ASSC for finding optimized designs.

5.4 Section Summary

In this section we used three different criteria for optimizing designs. The criteria we used were the L-criterion, a modified A-criterion and the alias sum of squares criterion. We constructed 100 designs and found the criterion value for each design, before choosing the design with the best criterion score. We approximated a model using the optimized design, before computing the values for the comparison criteria. We then computed and compared the minimum, maximum, mean and median values for the comparison criteria and the corresponding optimizing criterion value, we also found their correlation coefficients.

We found that neither the A_{mod}- nor the L-criterion correlated well with the RMSE, MAX or the ASSC. When using the ASSC as an optimizing criterion we found that it corresponds better with the comparison criteria than the other two optimizing criteria. Thus the ASSC is probably better to use, but its maximum correlation coefficient is 0.4273 which is still not as correlated as one might wish for it to be.

If we compare the worst maximum and best minimum values of the comparison criteria for each of the ASSC, L- and A_{mod}-criterion, we see that the designs optimized using the ASSC score the smallest minimum values for both the MAX and RMSE criteria. The L-criterion's scores are worse than the A-criterion, and is therefore less suitable as an optimizing criterion. These results were also indicated by the correlation coefficients.

In Bursztyn & Steinberg (2004) the correlations for the ASSC and IMSE-criterion ranged from 0.60 to 0.88, which is substantially higher than our results. This might be because we use different equations for the true and approximated models, or because we choose out of 100 constructed designs, which they are not doing in Bursztyn & Steinberg (2004).

The correlation coefficients were not consistent for the different types of design, which may indicate that we need to choose the optimizing criterion dependent on which design we want to use.

6 Comparison Study of Confounding Patterns for the MBR Design

In this section we have taken a closer look at non-optimized designs. Especially the multi-level binary replacement designs have been studied. As in the previous section we have used three different confounding patterns to construct fractional factorial designs used for obtaining MBR designs. We now want to see if we can gain more knowledge about which, or what kind of, confounding pattern provides us with the best MBR design, and whether or not this design is favorable to the other two designs studied here, LHS and $\mathrm{OA}_{p=4}$.

The procedure is similar to the one in section 5. We have constructed 1000 designs, but without optimizing. We constructed LHS designs, OA designs with $p = 4$ and MBR designs using the same three different confounding patterns as in the previous section. We used the same true and fitted model as in section 5 when using $s = 5$ factors. We have also constructed designs with 4, 3 and 2 factors. Where possible we have used a third-order model as the "true" model with βs between 0 and 1, see Appendix A for the exact models. We found an approximated model using a second-order model without quadratic terms. After fitting the model we found the values for the ASSC, MAX and RMSE criteria.

6.1 Non-Optimized Designs with Two Factors

In this section all the designs have $s = 2$ factors and $n = 32$ samples. We have only constructed LHS designs other than the MBR designs, as there was no OA design with 2 factors and 32 samples. We have also in this section used three different confounding patterns, but they are slightly alterated. As we want 32 sample points, we need a 2^{6-1} fractional factorial design. We only need one column which is constructed by a combination of the first five. The 3 confounding patterns are:

$\mathrm{MBR}_{\mathrm{VI}}$: We used a combination of the 5 original columns, so $ABCDE$, giving us a resolution VI fractional design.

MBR_4: We used a combination of 4 of the 5 original columns, so for example $ABCD$ or $BCDE$. Which combination to use was randomly chosen.

MBR_3: Here we used 3 of the 5 original factors, so for example ABC or BDE. Which combination to use was randomly chosen.

We still mixed up the columns before combining the binary numbers, and in that way we got different MBR designs.

In tables 14, 15 and 16 we observe that the $\mathrm{MBR}_{\mathrm{VI}}$ scores the best for all comparison criteria, only for the minimum value it only achieves the second or third best score. The LHS designs score poorly for all criteria, except for the minimum value of the alias sum of squares criterion. We see that the MBR_4 got the smallest minimum value for the MAX criterion, while the MBR_3 achieved the best minimum value for the RMSE criterion.

The confounding patterns used for the best or worst MBR designs are shown in tables 17 to 22. Some of their plots are presented in figures 13, 14 and 15. When looking at figure 15, the two-dimensional plots of the MBR_3 designs, we see that the plots scoring well for the RMSE criterion is more space-filling than the design getting a bad score. The two top plots have some areas undiscovered, but they cover the edges and the corners fairly well. The bottom right plot, representing the MBR_3, which got a bad RMSE score has several areas without any sample points, and not all of the borders are well covered.

Looking at the confounding patterns we see that for the good MBR_3 designs two of the original columns used in the combination are used as binary bits in the opposite factor of where the combination is a binary bit. So if the combination used is $-CDE$, and $-CDE$ is placed so that it will be a binary bit of factor 1 of the MBR design, then only one of C, D or E can be one of the other binary bits and A has to be the last binary bit. As we see in table 17. We can notice from table 22 that using two of C,D or E as a binary bit in the same MBR-factor as $-CDE$ leads to a bad MBR design.

We can see some of the same tendencies of how the factors for the MBR_4 are constructed using the fractional design's factors, but as the difference between its min and max values are small it is hard to find a consistent pattern.

In figure 13 two of the MBR_4 designs scoring well for the minimum value of the MAX criterion is plotted. We can see that these plots are covering the design space well, having symmetrical, circular patterns covering the design space and only leaving small circles empty. It is similar to the MBR_{VI} designs shown in figure 14, which only scored marginally less.

When constructing MBR designs with $s = 2$ factors there were no designs having overlapping sample points.

	LHS	MBR_3	MBR_4	MBR_{VI}
mean	0.763061	0.573627	0.535747	0.533472
min	0.555948	0.502638	0.520685	0.533472
max	1.283950	0.742621	0.559047	0.533472
median	0.739337	0.546273	0.534787	0.533472

Table 14: The MAX criterion values for designs with $s = 2$ factors and $n = 32$ samples.

	LHS	MBR_3	MBR_4	MBR_{VI}
mean	0.218349	0.199293	0.197104	0.197039
min	0.202990	0.197006	0.197039	0.197039
max	0.303534	0.237498	0.211366	0.197039
median	0.213232	0.197370	0.197056	0.197039

Table 15: The RMSE criterion values for designs with $s = 2$ factors and $n = 32$ samples.

	LHS	MBR_3	MBR_4	MBR_{VI}
mean	7.350143	7.103666	6.918175	6.911703
min	6.170827	6.807537	6.669325	6.911703
max	12.158061	8.567461	7.202436	6.911703
median	7.072390	6.916297	6.846576	6.911703

Table 16: The alias sum of squares criterion values for designs with $s = 2$ factors and $n = 32$ samples.

$$(A,\ -CDE,\ E) \quad (C,\ D,\ B)$$

Table 17: The confounding pattern used to construct the MBR_3 design getting the best value for the RMSE criterion's minimum value, 0.197.

$$(C,\ D,\ B) \quad (-BCE,\ A,\ E)$$

Table 18: The confounding pattern used to construct another MBR_3 design getting the best value for the RMSE criterion's minimum value, 0.197.

$$(-ACD,\ D,\ E) \quad (B,\ C,\ A)$$

Table 19: The confounding pattern used to construct another MBR_3 design getting the best value for the RMSE criterion's minimum value, 0.197.

$$(C,\ E,\ D) \quad (A,\ -ABDE,\ B)$$

Table 20: The confounding pattern used to construct MBR_4 design getting the minimum value for the RMSE criterion, 0.197.

$$(A,\ B,\ E) \quad (C,\ D,\ BCDE)$$

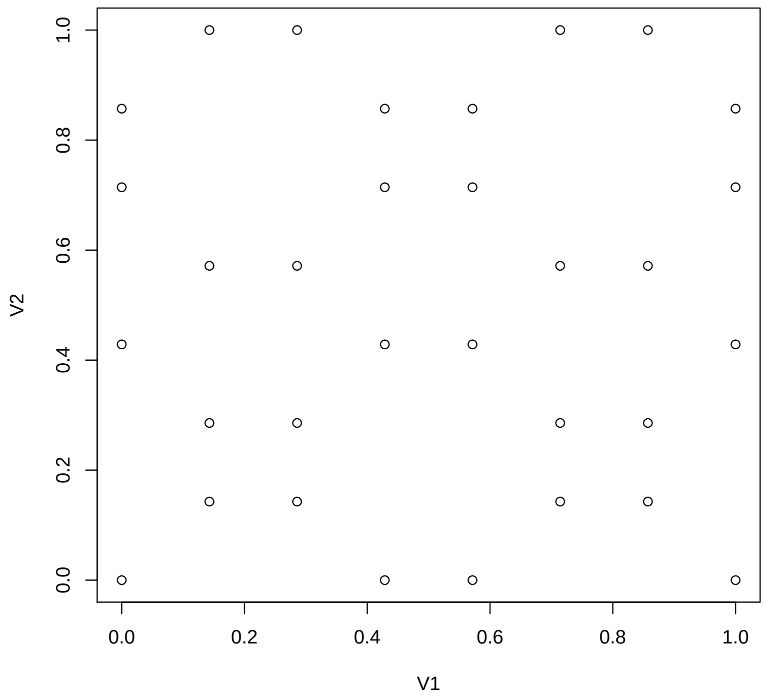
Table 21: The confounding pattern used to construct a MBR_4 design getting its best minimum value for the MAX criterion, 0.521.

$$(B,\ E,\ D) \quad (C,\ A,\ ABC)$$

Table 22: The confounding pattern used to construct the MBR_3 design getting the second largest maximum value for the RMSE criterion, 0.211.

$$(-BCDE,\ D,\ A) \quad (B,\ E,\ C)$$

Table 23: The confounding pattern used to construct the MBR_4 design getting the second largest maximum value for the RMSE criterion, 0.559.



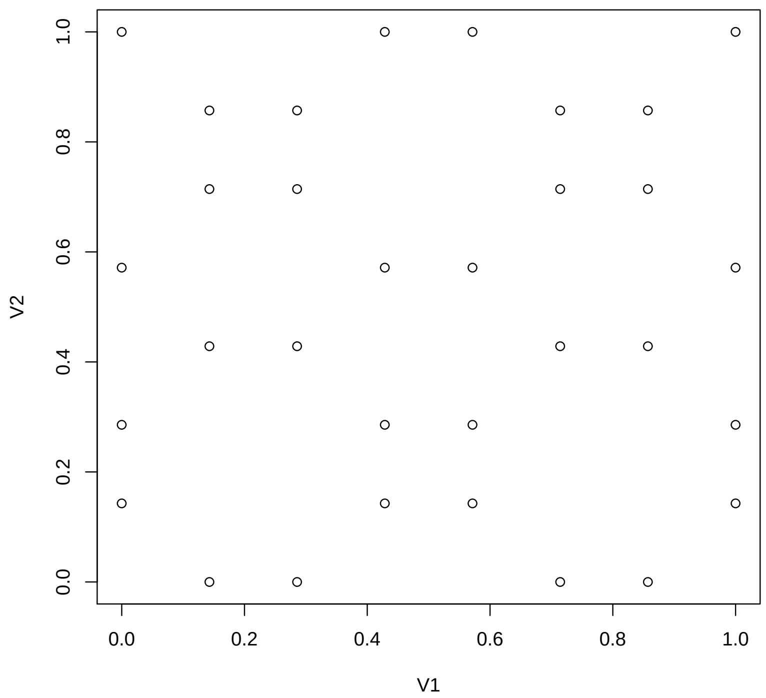


Figure 13: Plots of two MBR_4 designs scoring the best for the minimum value of the MAX criterion. The top plot belongs to the confounding pattern in table 20, while the design in the bottom plot is constructed using the confounding pattern in table 21.

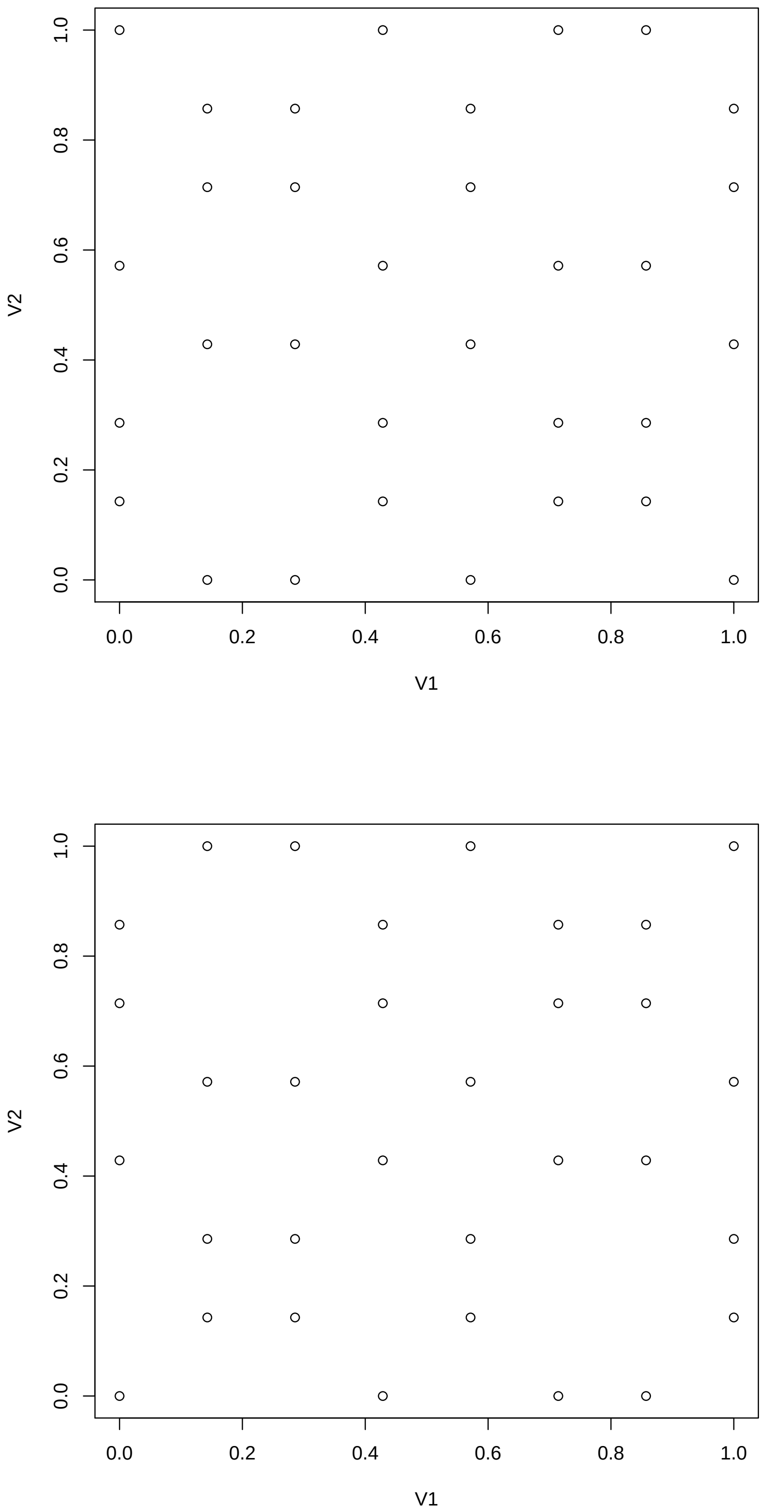


Figure 14: Plots of the MBR_{VI} design, the designs achieving the best mean, median and max values for all comparison criteria.

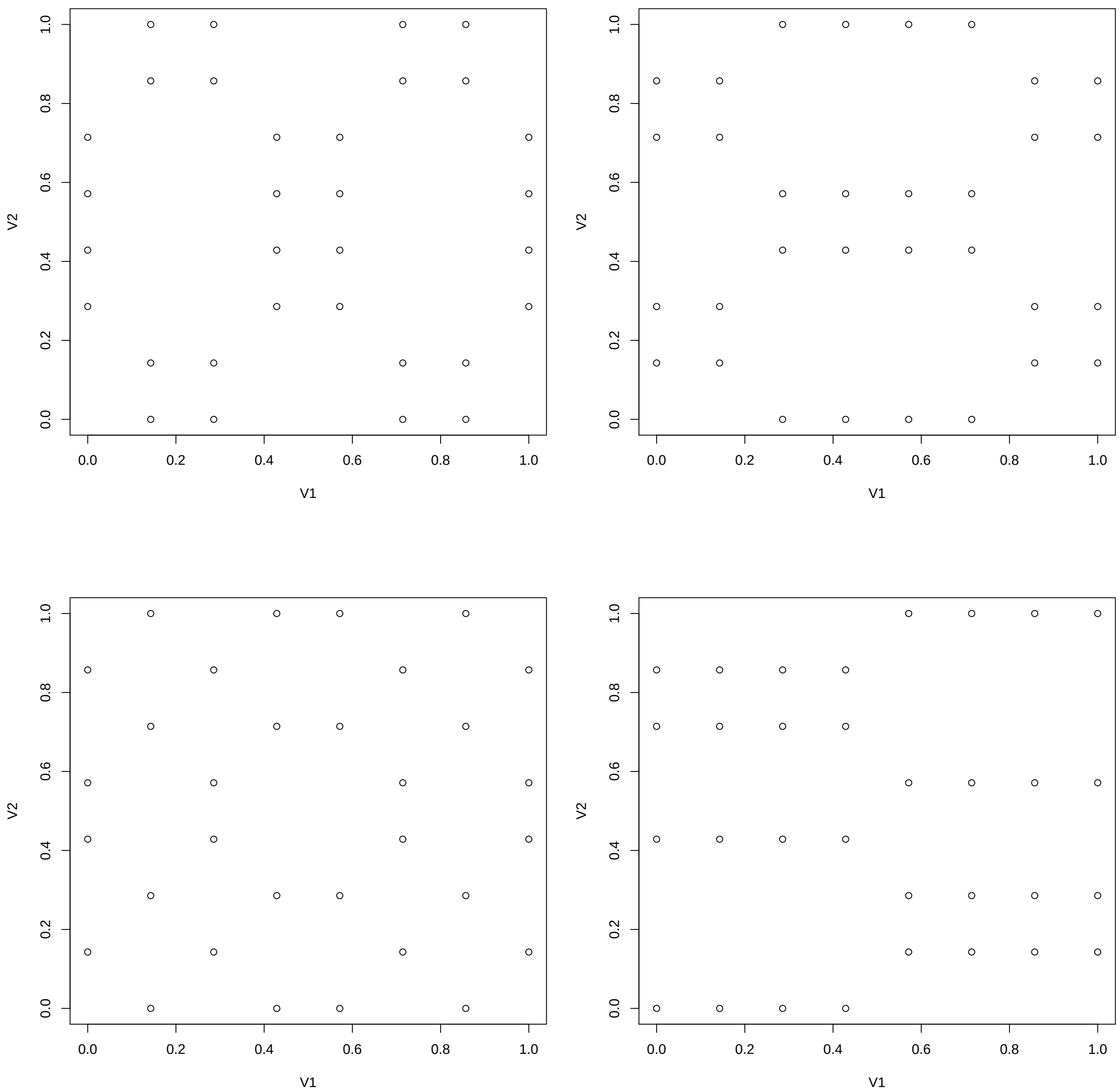


Figure 15: The MBR_3 designs from tables 17, 18, 19 and 22. The top two plots and bottom left corresponds with the designs having minimum RMSE values. While the last plot, bottom right, belongs to the design obtaining the maximum RSME score, table 22.

6.2 Non-Optimized Designs with Three Factors

In this section we have constructed designs with $s = 3$ factors and $n = 32$ samples. The confounding patterns used are the same as in section 5, but without using the optimizing criteria, ASSC, L- or A_{mod}-criteria. The max, min, mean and median values for the comparison criteria, MAX, RMSE and ASSC, can be found in tables 26, 27 and 28.

Looking at the maximum and minimum values for the MAX and RMSE criteria we see that the MBR designs score well, for both criteria MBR scores better than both LHS and $OA_{p=4}$ designs. While for the ASSC the MBR designs score worse. The MBR_{odd} scores similarly to the MBR_{all} and MBR_{IV} for the minimum values of the RMSE and MAX criteria, but for the max, mean and median values it scores worse. MBR_{odd} has a bigger difference between its max and min values, which could make it more risky to use when we are not optimizing. All three MBR designs have a bigger difference between max and min than $OA_{p=4}$ for the ASSC and RMSE criterion.

As for designs with two factors LHS scores badly for all criteria. It achieves the best minimum value for the ASSC, but all other values are poor.

In tables 25 to 34 some of the confounding patterns for the best and worst MBR designs are presented. The corresponding two-dimensional plots are shown in figures 16 and 17. We can easily see from the two-dimensional plots in figure 16 that the designs scoring good for the comparison criteria are covering the design space well. While the MBR designs in figure 17 have larger areas, corners or edges of the design space poorly covered, and therefore achieves higher values for the MAX and RMSE criteria.

When comparing the confounding patterns in tables 25 to 34 we see that using only the original factors, A, B, C, D and E, as the binary bits for one of the MBR-factors results in a bad design. We can also notice that three out of the four MBR designs scoring well for the comparison criteria have used negative negative defining relations.

In table 24 we see that to construct MBR designs having three factors we need to construct more designs than we desire to avoid replicated sample points. Especially when using all-combinations confounding we need to construct extra designs.

Design	Extra designs constructed
MBR_{odd}	342
MBR_{all}	835
MBR_{IV}	187

Table 24: This table shows how many designs were constructed, but discarded because of overlapping samples. For non-optimized designs.

$$(-ABDE, -ABCE, C) \quad (-ABCDE, B, A) \quad (-BCDE, E, D)$$

Table 25: The confounding pattern used to obtain the MBR_{all} design getting the lowest minimum value for the MAX criterion, 0.920.

	LHS	$OA_{p=4}$	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	1.839613	1.499082	1.384439	1.211289	1.247989
min	1.154111	1.134832	1.007812	0.919663	1.014340
max	3.307980	2.084974	2.278844	1.695228	1.763329
median	1.778546	1.484776	1.355244	1.183678	1.206884

Table 26: The MAX criterion for designs with $s = 3$ factors and $n = 32$ samples.

	LHS	$OA_{p=4}$	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	0.387092	0.347532	0.338767	0.334495	0.330060
min	0.328498	0.331202	0.314296	0.315021	0.314549
max	0.582960	0.395890	0.435993	0.393443	0.389337
median	0.376896	0.345602	0.334254	0.329926	0.326387

Table 27: The RMSE criterion for designs with $s = 3$ factors and $n = 32$ samples.

	LHS	$OA_{p=4}$	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	19.729832	18.767350	18.867087	18.178308	18.114551
min	16.174935	16.368278	16.802530	16.493380	16.291267
max	33.492133	22.175260	27.585828	23.722000	23.323185
median	19.214386	18.629220	18.440369	17.927701	17.900909

Table 28: The ASSC for designs with $s = 3$ factors and $n = 32$ samples.

$$(C, -ACDE, D) \quad (E, -ABCDE, -BCDE) \quad (-ABCE, B, A)$$

Table 29: The confounding pattern used to construct the MBR_{all} design getting minimum value for the RMSE criterion, 0.315.

$$(ABCDE, B, ABC) \quad (ACD, A, E) \quad (D, C, ABD)$$

Table 30: The confounding pattern used to achieve the MBR_{odd} design getting the minimum value for the MAX criterion, 1.001.

$$(-ABE, E, -BCD) \quad (D, -ACE, B) \quad (A, C, -ABCDE)$$

Table 31: The confounding pattern used to construct the MBR_{odd} design getting the lowest minimum value for the RMSE criterion, 0.314.

$$(-ABCE, C, -ABDE) \quad (-ACDE, -ABCDE, E) \quad (D, B, A)$$

Table 32: The confounding pattern used to construct the MBR_{all} design getting the lowest maximum value for the MAX criterion, 1.695.

$$(ABCD, E, ABCE) \quad (BCDE, D, B) \quad (A, ABCDE, C)$$

Table 33: The confounding pattern used to construct the MBR_{all} design getting maximum value for the RMSE criterion, 0.393.

$$(A,\, E,\, C) \quad (-CDE,\, -ACD,\, -ABE) \quad (B,\, -ABCDE,\, D)$$

Table 34: The confounding pattern used to find the $\mathrm{MBR}_{\mathrm{odd}}$ design achieving its maximum value for both MAX and RMSE criteria.

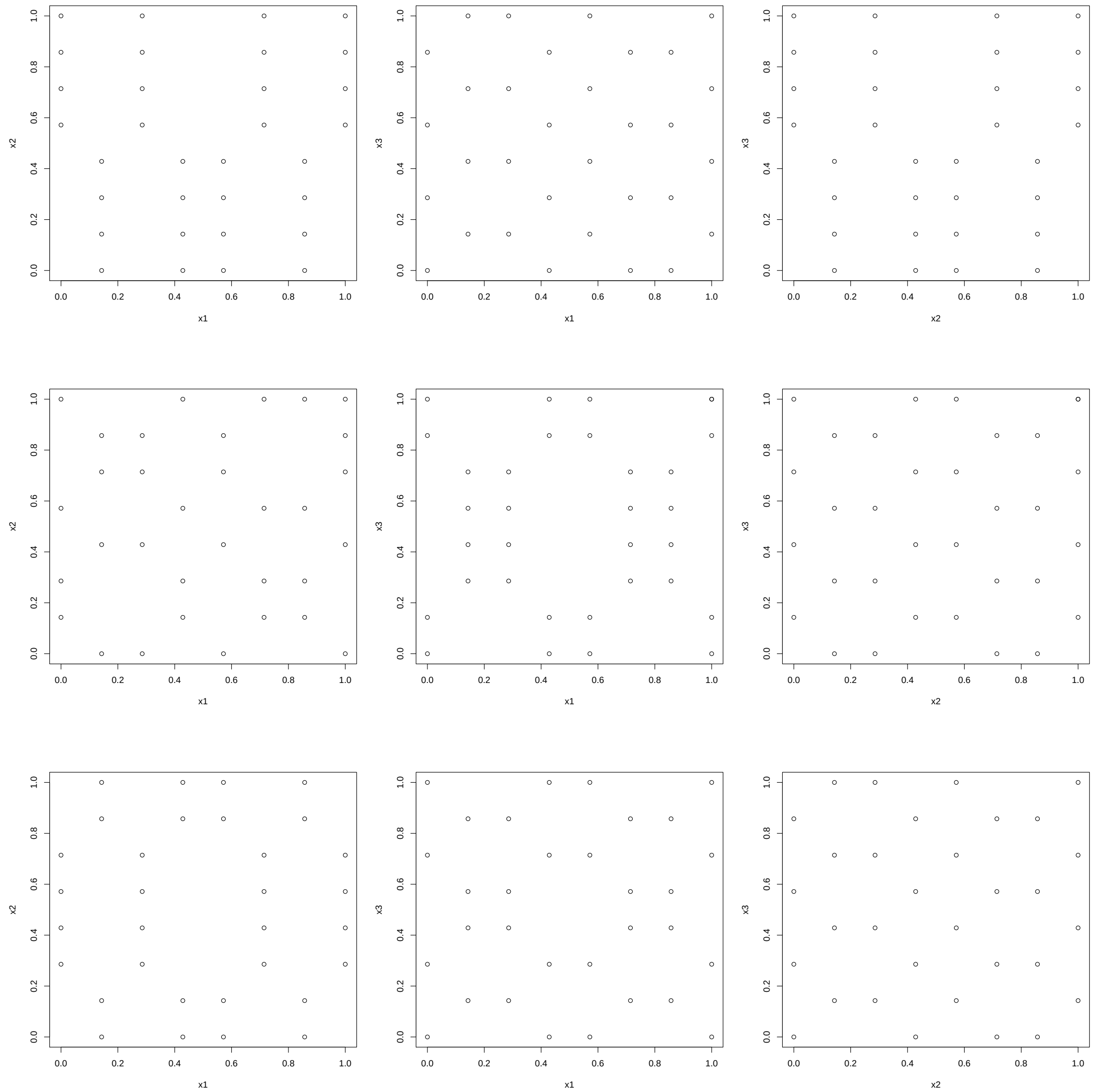


Figure 16: Two-dimensional plots of the MBR designs achieving the minimum values of the comparison criteria. With the top plots belonging to the design in table 25, the middle plot corresponds to the design in table 29 and the bottom is the two-dimensional plot of the design in table 31.

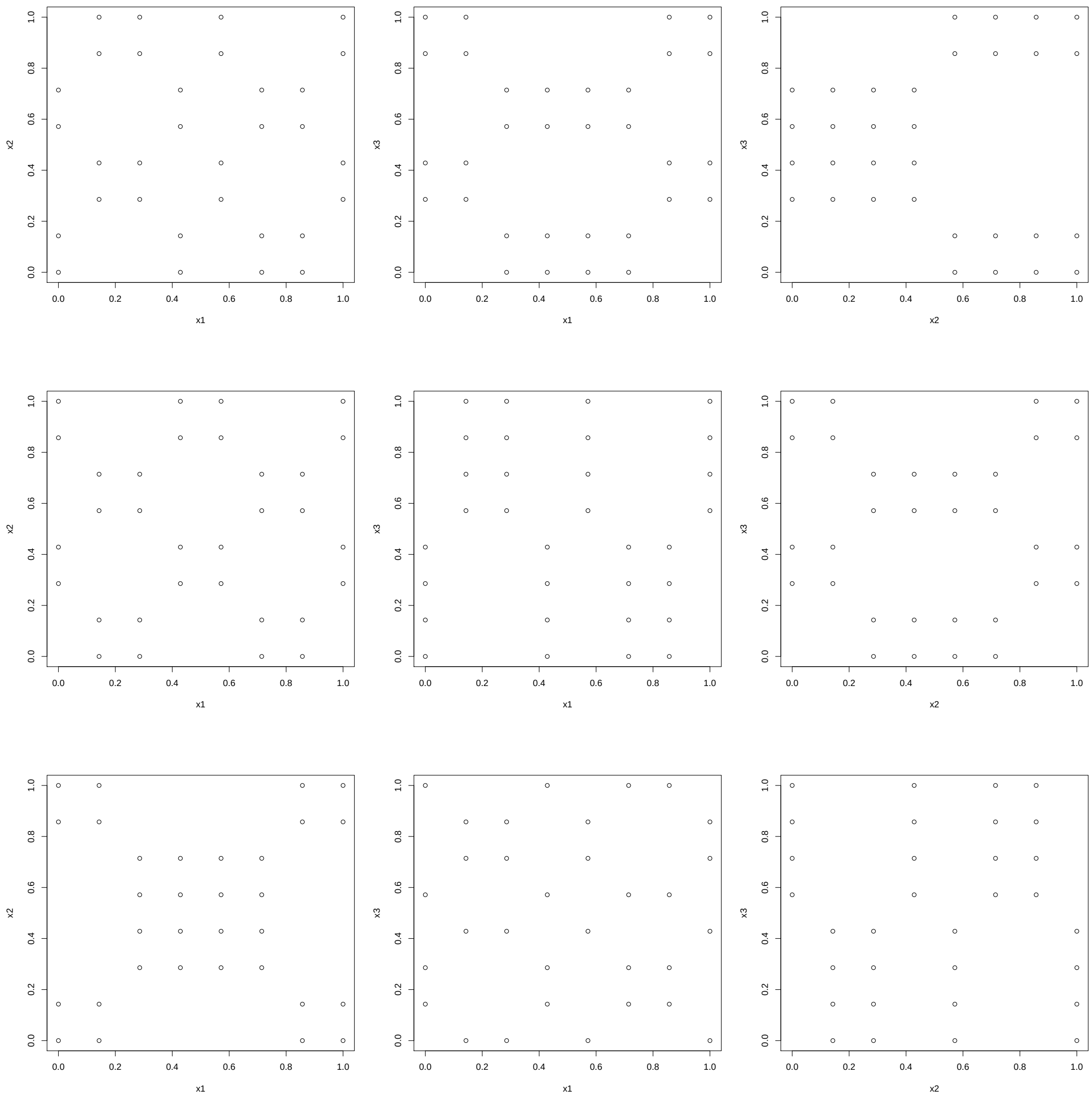


Figure 17: Two-dimensional plots of the MBR designs obtaining the largest values for the comparison criteria. With the top plots belonging to the design in table 33, the middle plot corresponds to the design in table 32 and the bottom is the two-dimensional plot of the design in table 34.

6.3 Non-Optimized Designs with Four Factors

In this section we have constructed MBR, OA and LHS designs with $s = 4$ factors and $n = 32$ sample points. The confounding patterns used for constructing MBR designs are the same as we used for three factors and in section 5.

The maximum, minimum, mean and median values of the comparison criteria, ASSC, MAX and RMSE, are presented in tables 36, 37 and 38. We see that the $\mathrm{MBR}_{\mathrm{IV}}$ having three factors scored better compared to the other designs than it does when having four factors. The $\mathrm{OA}_{p=4}$ scores better compared to the MBR designs with four factors, especially for the ASSC. The difference between maximum and minimum values are also for MBR designs having four factors larger than for the $\mathrm{OA}_{p=4}$. The LHS scores badly.

When comparing the different MBR designs we see that the $\mathrm{MBR}_{\mathrm{all}}$ scores well, especially for the MAX criterion. For the RMSE criterion the $\mathrm{MBR}_{\mathrm{all}}$ design has the smallest mean and median values, its difference between the maximum and minimum values is small compared to the other two MBR design types.

The odd-combination and all-combination confounding patterns for the MBR designs used when achieving their maximum or minimum values for the RMSE and MAX criteria are shown in tables 39 to 46. The two-dimensional plots for the $\mathrm{MBR}_{\mathrm{all}}$ design achieving the best minimum value for the MAX criterion is shown in figure 18. We see that some of its two dimensional plots cover the design space badly. Especially the one in the top right corner is less space-filling as it leaves two of its corners and much of its borders undiscovered. This plot has all of its sample points in three areas.

We can see the same for the $\mathrm{MBR}_{\mathrm{odd}}$, some of the two dimensional plots cover the design space more poorly than what we wish for it to in order to be space-filling. When looking at figure 20, the $\mathrm{MBR}_{\mathrm{odd}}$ design scoring a high maximum value for the RMSE criterion, we see that all of its two-dimensional plots are covering the design space badly. The plots in figure 20 look less space-filling than the ones in figure 18, as we expect since the MAX and RMSE values for the design in figure 18 are better than for the design in figure 20.

Both designs in figure 18 and 20 are covering their design spaces badly when compared to the two-dimensional plots of the design achieving the best minimum value of the RMSE criterion, figure 19.

In table 35 we see the number of extra designs constructed, but discarded because of overlapping sample points, This table presents a big drawback to the MBR designs, and especially to the $\mathrm{MBR}_{\mathrm{all}}$ design. We have constructed 1669 extra designs using all-combination confounding, compared to a little more than 900 when using the other two confounding patterns. For LHS and OA designs no extra designs were constructed.

Design	Extra designs constructed
$\mathrm{MBR}_{\mathrm{odd}}$	991
$\mathrm{MBR}_{\mathrm{all}}$	1669
$\mathrm{MBR}_{\mathrm{IV}}$	954

Table 35: This table shows how many designs were constructed and discarded because of overlapping sample points. For non optimized designs with $s = 4$ factors and $n = 32$ sample points.

	LHS	$\mathrm{OA}_{p=4}$	$\mathrm{MBR}_{\mathrm{odd}}$	$\mathrm{MBR}_{\mathrm{all}}$	$\mathrm{MBR}_{\mathrm{IV}}$
mean	2.914421	2.263318	2.332529	1.931600	2.272945
min	1.599464	1.736542	1.467653	1.329916	1.466754
max	6.280772	3.551684	4.403467	3.136148	4.654013
median	2.782094	2.239603	2.259380	1.894507	2.207479

Table 36: The MAX criterion for designs with $s = 4$ factors and $n = 32$ samples.

	LHS	$\mathrm{OA}_{p=4}$	$\mathrm{MBR}_{\mathrm{odd}}$	$\mathrm{MBR}_{\mathrm{all}}$	$\mathrm{MBR}_{\mathrm{IV}}$
mean	0.513357	0.437391	0.453569	0.431730	0.452287
min	0.402553	0.394253	0.369865	0.375783	0.371413
max	0.907655	0.555617	0.760158	0.563563	0.735282
median	0.498110	0.433297	0.445851	0.428123	0.446304

Table 37: The RMSE criterion for designs with $s = 4$ factors and $n = 32$ samples.

	LHS	$\mathrm{OA}_{p=4}$	$\mathrm{MBR}_{\mathrm{odd}}$	$\mathrm{MBR}_{\mathrm{all}}$	$\mathrm{MBR}_{\mathrm{IV}}$
mean	34.936746	30.660230	35.944555	31.454301	35.680102
min	25.801940	26.872616	26.856439	25.737101	26.841935
max	63.425420	35.634771	66.949584	47.800687	67.779881
median	33.855296	30.531093	35.395885	30.827943	34.947704

Table 38: The alias sum of squares criterion for designs with with $s = 4$ factors and $n = 32$ samples.

$$(-ACDE,\, A,\, -ABD) \quad (-ABCE,\, E,\, C) \quad (B,\, -ABCD,\, D) \quad (-ABDE,\, -ABCDE,\, -BCDE)$$

Table 39: The confounding pattern and the combinations of binary bits to construct the $\mathrm{MBR}_{\mathrm{all}}$ design getting its minimum value for the RMSE criterion, 0.376.

$$(-ABCDE,\, D,\, -ACDE) \quad (B,\, A,\, -ACD) \quad (-ABCE,\, C,\, E) \quad (-BCDE,\, -ABDE,\, -ABCD)$$

Table 40: The confounding pattern and the combinations of binary bits to construct the $\mathrm{MBR}_{\mathrm{all}}$ design achieving the lowest minimum value for the MAX criterion, 1.330.

$(-BDE, -BCE, D) \quad (E, -ABCDE, -ADE) \quad (-ABC, -BCD, B) \quad (A, -ACE, C)$

Table 41: The confounding pattern and the combinations of binary bits to construct the $\mathrm{MBR}_{\mathrm{odd}}$ design achieving the lowest minimum value for the RMSE criterion, 0.370.

$(A, E, -ACE) \quad (-ABCE, -BCE, -ABCDE) \quad (-CDE, B, D) \quad (C, -BDE, -ADE)$

Table 42: The confounding pattern and the combinations of binary bits to construct the $\mathrm{MBR}_{\mathrm{odd}}$ design scoring its minimum value for the MAX criterion, 1.468.

$(ABCDE, ABC, E) \quad (C, ABCD, B) \quad (BCDE, ACDE, ABCE) \quad (ABDE, D, A)$

Table 43: The confounding pattern and the combinations of binary bits to construct the $\mathrm{MBR}_{\mathrm{all}}$ design getting the maximum value for the RMSE criterion, 0.564.

$(ACD, BCDE, ABCDE) \quad (D, ACDE, C) \quad (ABCE, A, B) \quad (ABDE, E, ACDE)$

Table 44: The confounding pattern and the combinations of binary bits to construct the $\mathrm{MBR}_{\mathrm{all}}$ design getting its maximum value for the MAX criterion, 3.136.

$(ACE, CDE, E) \quad (D, C, ABCDE) \quad (ABE, B, ABD) \quad (ADE, BCD, A)$

Table 45: The confounding pattern and the combinations of binary bits to construct the $\mathrm{MBR}_{\mathrm{odd}}$ design getting the maximum value for the RMSE criterion, 0.760.

$(-BCD, C, A) \quad (-BCE, D, E) \quad (-ABD, B, -CDE) \quad (-ABCDE, -BDE, -ADE)$

Table 46: The confounding pattern and the combinations of binary bits to construct the $\mathrm{MBR}_{\mathrm{odd}}$ design getting the maximum value for the MAX criterion, 4.403.

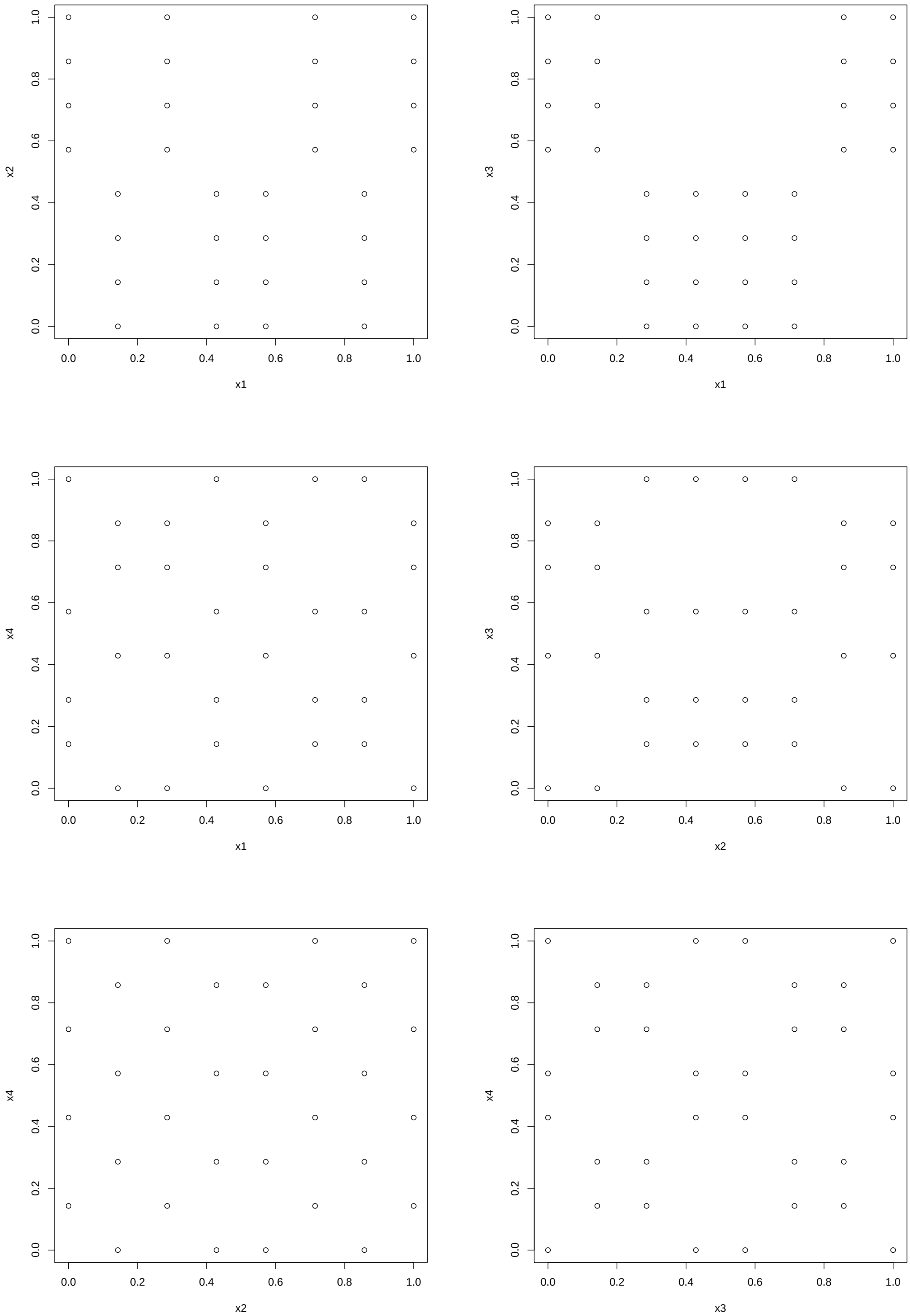


Figure 18: The MBR_{all} design scoring the best for the MAX criterion's minimum value shown in table 40.

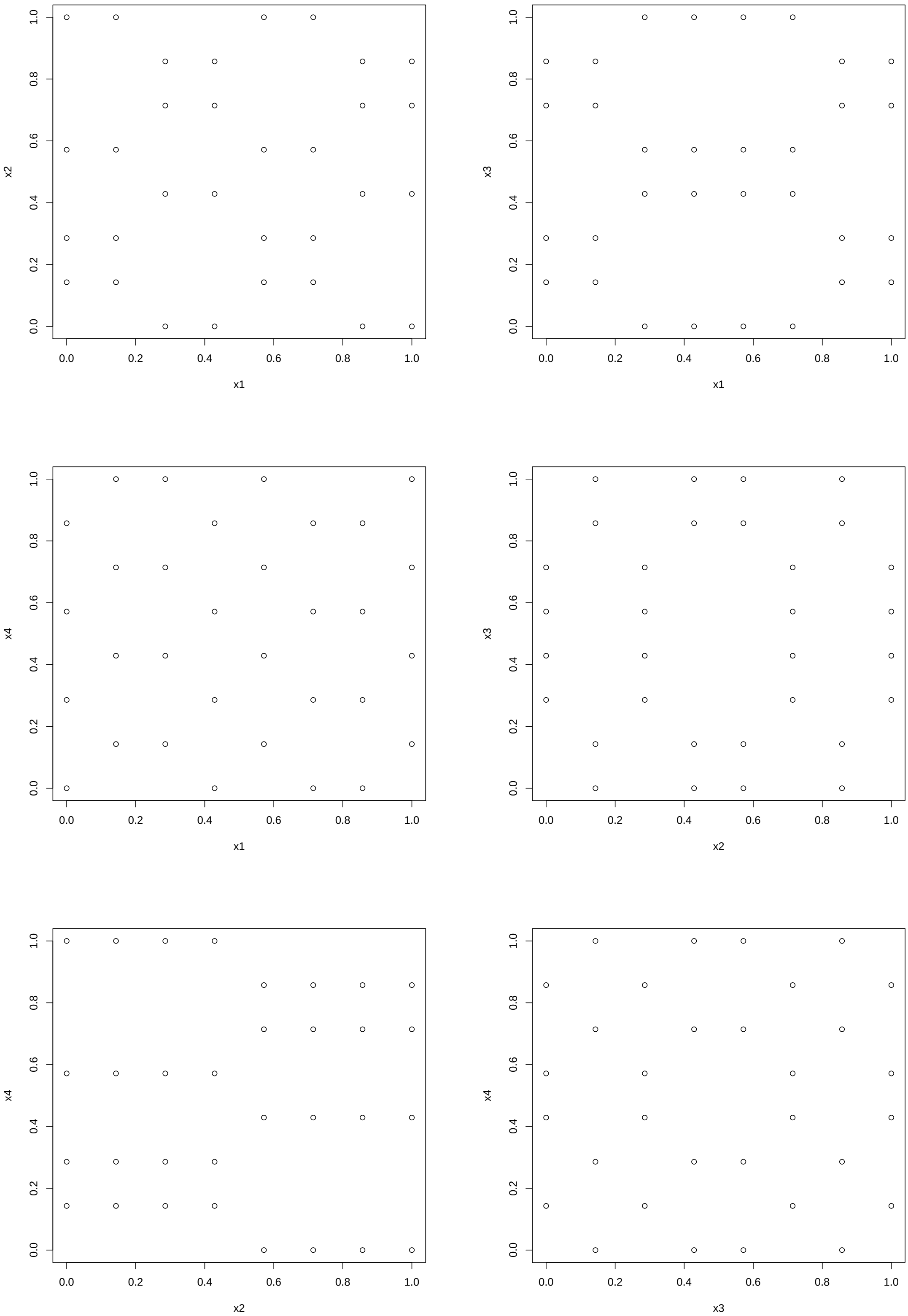


Figure 19: The MBR_{odd} design scoring the best for the RMSE criterion's minimum value, the confounding pattern used is shown in table 41.

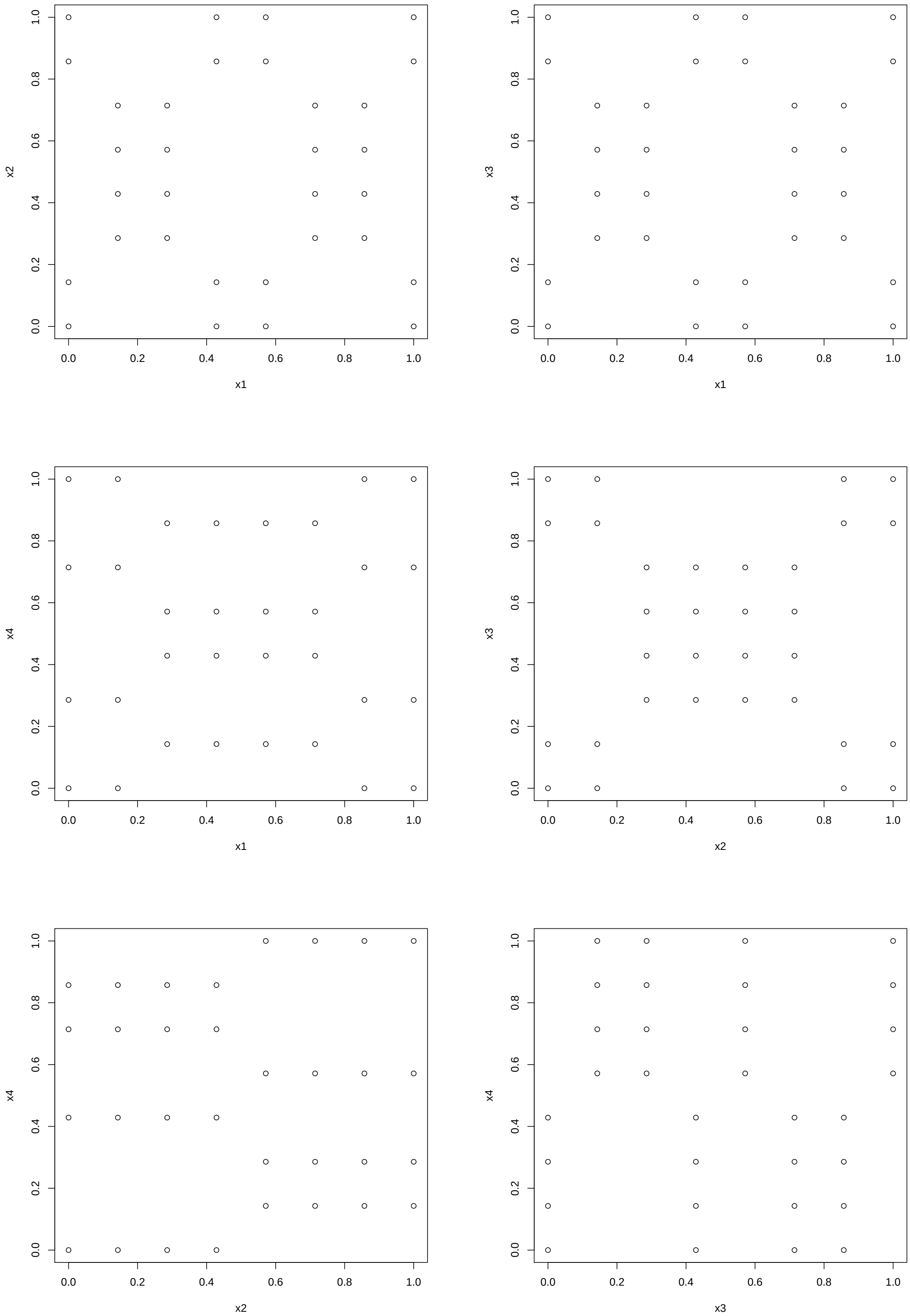


Figure 20: The MBR_{odd} design with the design type's highest maximum value for the RMSE criterion, the confounding pattern is shown in table 45.

6.4 Non-Optimized Designs with Five Factors

In this section we have constructed designs having $s = 5$ factors and $n = 32$ samples. This is the part of this section which is most similar to what we did in section 5, but without using the optimizing criteria, ASSC, A_{mod}- or L-criterion. The comparison criteria's values, MAX, RMSE and ASSC, are presented in tables 48, 49 and 50.

Looking at the results presented in the tables we see that $OA_{p=4}$ performs well, for all three comparison criteria it achieved the smallest maximum value, its mean and median values are also good compared to the other designs. However, the smallest minimum values for the three criteria are all achieved by the MBR design, with either odd-combination or all-combination confounding. The MBR_{all} design scores well for the mean and median value of all three criteria, but its max values are a little high. The MBR_{odd} design has the largest maximum value for all three criteria, also its mean and median values are quite high compared to the other designs.

We see that the LHS design performs poorly for both the RMSE and MAX criteria, but it performs average for the ASSC.

If we look at the different confounding patterns for the MBR design, we see that the designs constructed using all-combination and odd-combination confounding scores the best for the MAX and RMSE minimum values. We see that both MBR_{odd} and MBR_{IV} has high maximum values for all three criteria, they also have similar minimum, mean and median values, which we might expect as they both are constructed using resolution IV fractional factorial designs. Tables 51 to 58 show how the factors of some of the MBR designs are constructed.

The MBR_{odd} design scoring the best for the RMSE criterion's minimum value is plotted in figure 21. We see that the two-dimensional plots cover the design space well. The two-dimensional projection which covers its design space worst is the middle plot to the right, x_2 and x_4, which covers the center and the corners badly. This less space-filling two-dimensional projection is also included in the two-dimensional plots of the MBR_{all}, shown in figure 22, which scores the best for the RMSE minimum value for designs using all-combinations confounding. Most of the two-dimensional plots in figure 21 and 22 cover the design space well, some are leaving circular areas empty. When comparing these two figures to figure 23, the MBR_{odd} design scoring the worst maximum value for the RMSE criterion, we see that the good designs have a much better coverage of the design space, they are more space-filling than the bad design.

We saw for three and four factors that we discarded MBR designs having overlapping sample points, table 47 shows that for five factors this is an even bigger problem, and we might assume that the problem will increase as we need more factors. This is a drawback to the MBR design.

Design	Extra designs constructed
MBR_{odd}	1689
MBR_{all}	9066
MBR_{IV}	1695

Table 47: This table shows how many designs were constructed and discarded because of overlapping samples. For non-optimized designs with $s = 5$ factors and $n = 32$ sample points.

	LHS	$\text{OA}_{p=4}$	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	4.931089	3.581583	4.592674	3.514203	4.693552
min	2.935927	2.623204	2.191684	2.296659	2.710690
max	12.105273	5.239734	14.786403	7.043198	12.336697
median	4.741866	3.524037	4.260931	3.426173	4.421390

Table 48: The MAX criterion for designs with $s = 5$ factors and $n = 32$ samples.

	LHS	$\text{OA}_{p=4}$	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	0.742696	0.603179	0.773210	0.635904	0.754106
min	0.529672	0.514487	0.474715	0.483289	0.489318
max	1.550762	0.771001	2.807374	1.040195	2.260235
median	0.718268	0.596468	0.725096	0.624814	0.702931

Table 49: The RMSE criterion for designs with $s = 5$ factors and $n = 32$ samples.

	LHS	$\text{OA}_{p=4}$	MBR_{odd}	MBR_{all}	MBR_{IV}
mean	68.740075	65.898310	92.325573	61.441368	91.995095
min	47.518958	55.060520	57.672853	44.954157	54.291716
max	138.130082	77.989154	387.950475	119.171823	343.70959
median	66.487986	65.686747	84.173782	60.016666	84.249792

Table 50: The alias sum of squares criterion for designs with $s = 5$ factors and $n = 32$ samples.

$$(-CDE,\, -ACDE,\, -BCE)\,(C,\, -ABCE,\, -BCDE)\,(B,\, -ABCDE,\, -ABDE)$$
$$(E,\, D,\, -ABD)\ (A,\, -ABCD,\, -BDE)$$

Table 51: The confounding pattern and binary bit combinations used to construct the MBR_{all} design getting the minimum value for the MAX criterion, 2.297.

$$(A,\, -ABD,\, C) \quad (D,\, -ACD,\, -ACDE) \quad (-ABDE,\, -BCDE,\, -ABCDE)$$
$$(-BCE,\, E,\, -ABCDE)\; (B,\, -ABE,\, -ABCD)$$

Table 52: The confounding pattern and binary bit combinations used to construct the $\mathrm{MBR}_{\mathrm{all}}$ design getting the minimum value for the RMSE criterion, 0.483.

$$(C,\, A,\, ABE) \quad (BCD,\, BCE,\, B) \quad (D,\, E,\, ABD)$$
$$(ADE,\, BDE,\, ABCDE)\; (BCE,\, ABC,\, ACD)$$

Table 53: The confounding pattern and binary bit combinations used to construct the $\mathrm{MBR}_{\mathrm{odd}}$ design getting the best value for the MAX criterion's minimum value, 2.192.

$$(A,\, -CDE,\, -ABCDE) \quad (D,\, -BCE,\, C) \quad (-BCD,\, -ACE,\, E)$$
$$(-ACD,\, -ABE,\, -ADE)\; (-BDE,\, B,\, -ABD)$$

Table 54: The confounding pattern and binary bit combinations used to construct the $\mathrm{MBR}_{\mathrm{odd}}$ design getting the smallest minimum value for the RMSE criterion , 0.475.

$$(C,\, D,\, E) \quad (ABD,\, B,\, ABCDE) \quad (ADE,\, A,\, BCD)$$
$$(ACDE,\, ABCE,\, ABDE)\; (ABCD,\, ACE,\, BCDE)$$

Table 55: The confounding pattern and binary bit combinations used to construct the $\mathrm{MBR}_{\mathrm{all}}$ design getting the largest maximum value for the MAX criterion's minimum value, 7.043.

$$(ABC,\, B,\, D) \quad (BCDE,\, ABE,\, ABCD) \quad (ABDE,\, ABCE,\, ABD)$$
$$(C,\, A,\, E)\; (ABCDE,\, ABE,\, BDE)$$

Table 56: The confounding pattern and binary bit combinations used to construct the $\mathrm{MBR}_{\mathrm{all}}$ design getting the designs largest maximum value for the RMSE criterion, 1.040.

$$(A,\, BCD,\, BDE)\; (C,\, ABC,\, ABD)\; (ADE,\, ACE,\, CDE)$$
$$(ABCDE,\, ABE,\, ACD)\; (B,\, D,\, E)$$

Table 57: The confounding pattern and binary bit combinations used to construct the $\mathrm{MBR}_{\mathrm{odd}}$ design getting the worst value for the MAX criterion's maximum value, 14.786.

$$(-ABCDE,\, E,\, -CDE) \quad (-BCE,\, -ABD,\, B) \quad (-BCD,\, A,\, D)$$
$$(-ABC,\, -ACD,\, -ACE)\; (-ABE,\, C,\, -BDE)$$

Table 58: The confounding pattern and binary bit combinations used to construct the $\mathrm{MBR}_{\mathrm{odd}}$ design getting largest maximum value for the RMSE criterion, 2.807.

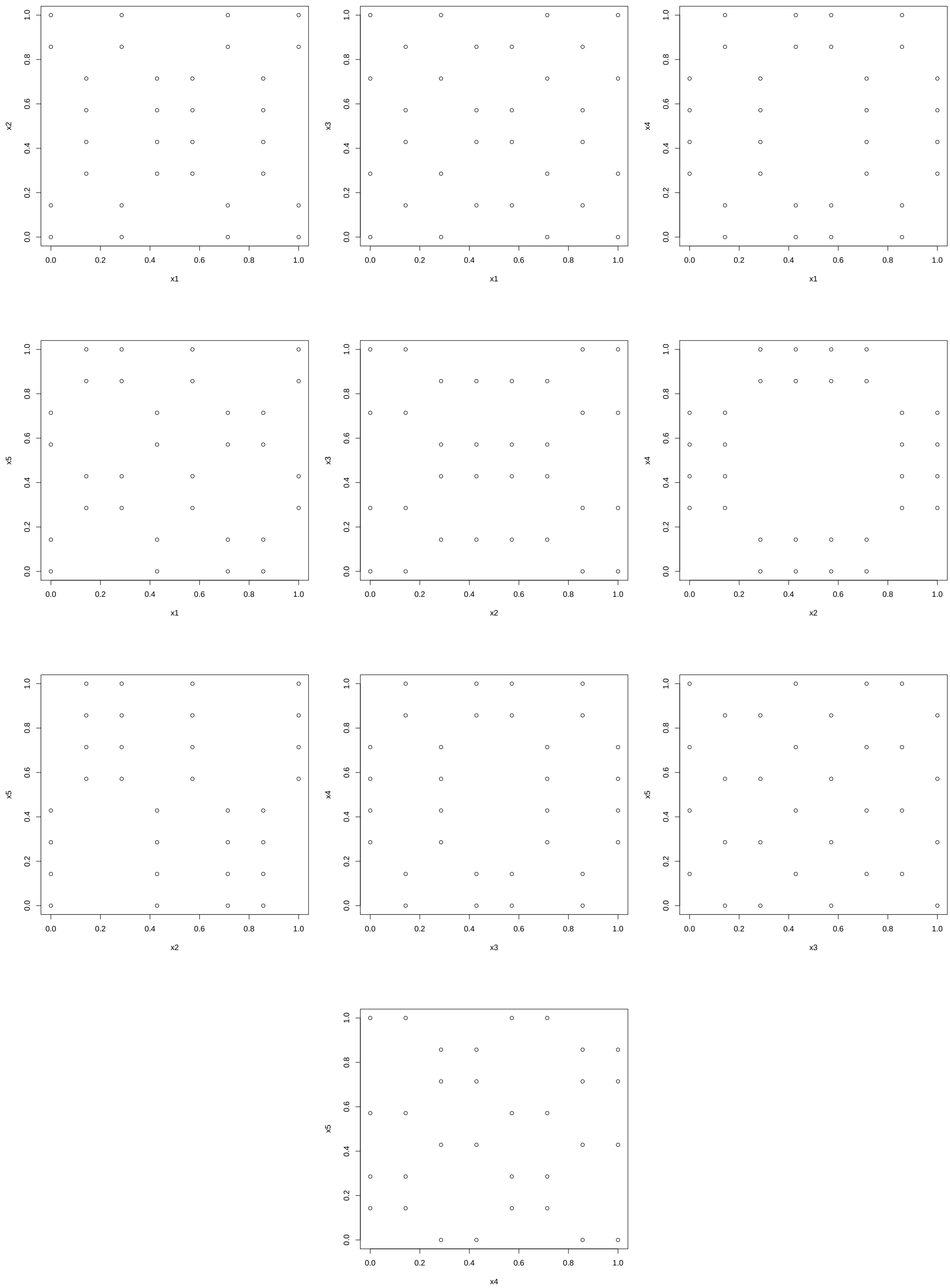


Figure 21: The MBR_{odd} design which got the best minimum value for the RMSE criterion. Its confounding pattern and how it combined the binary bits to construct the MBR design is shown in table 54.

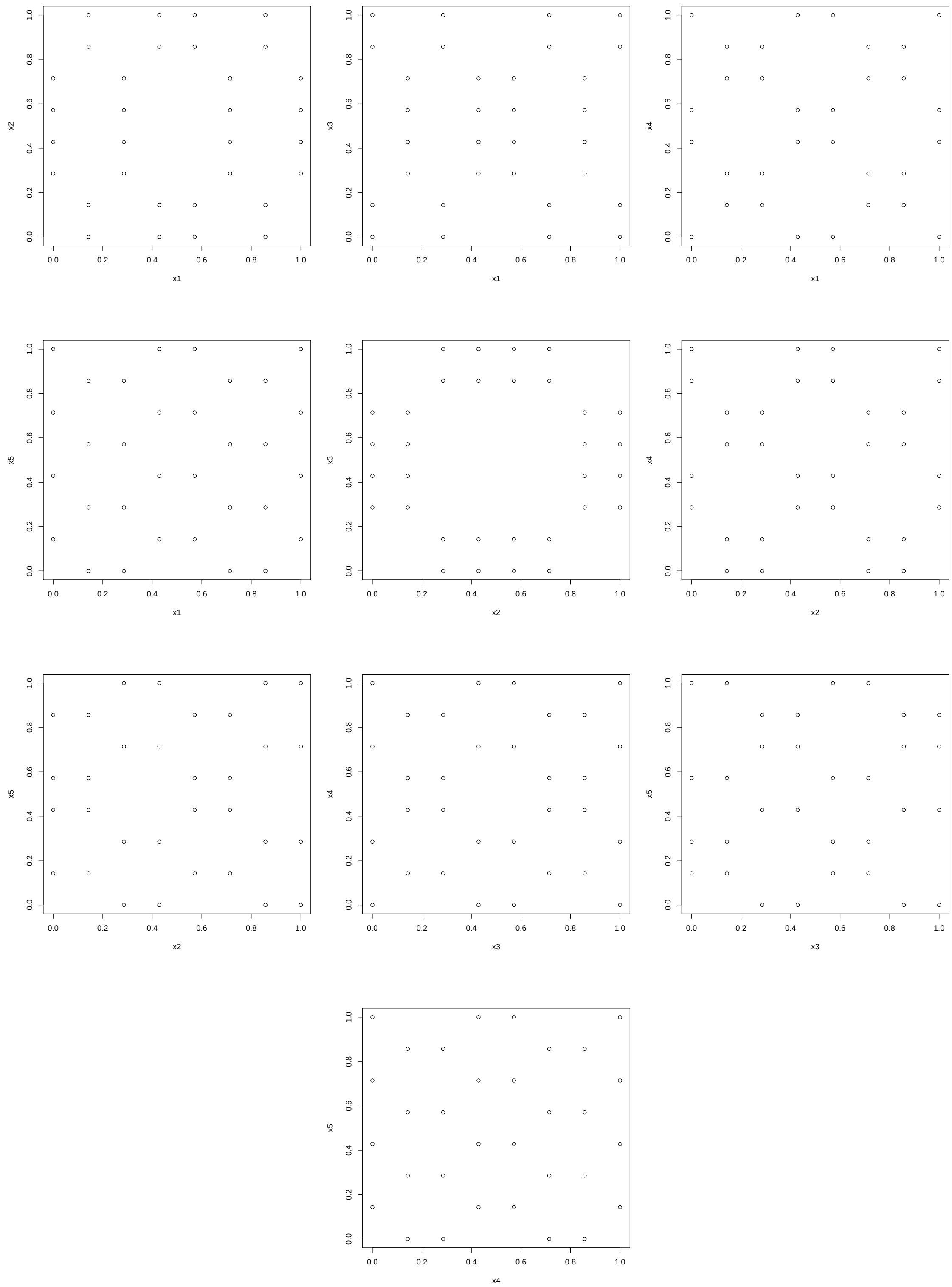


Figure 22: The MBR_{all} design which got the best minimum value for the RMSE criterion. Its confounding pattern and how it combined the binary bits to construct the MBR design is shown in table 52.

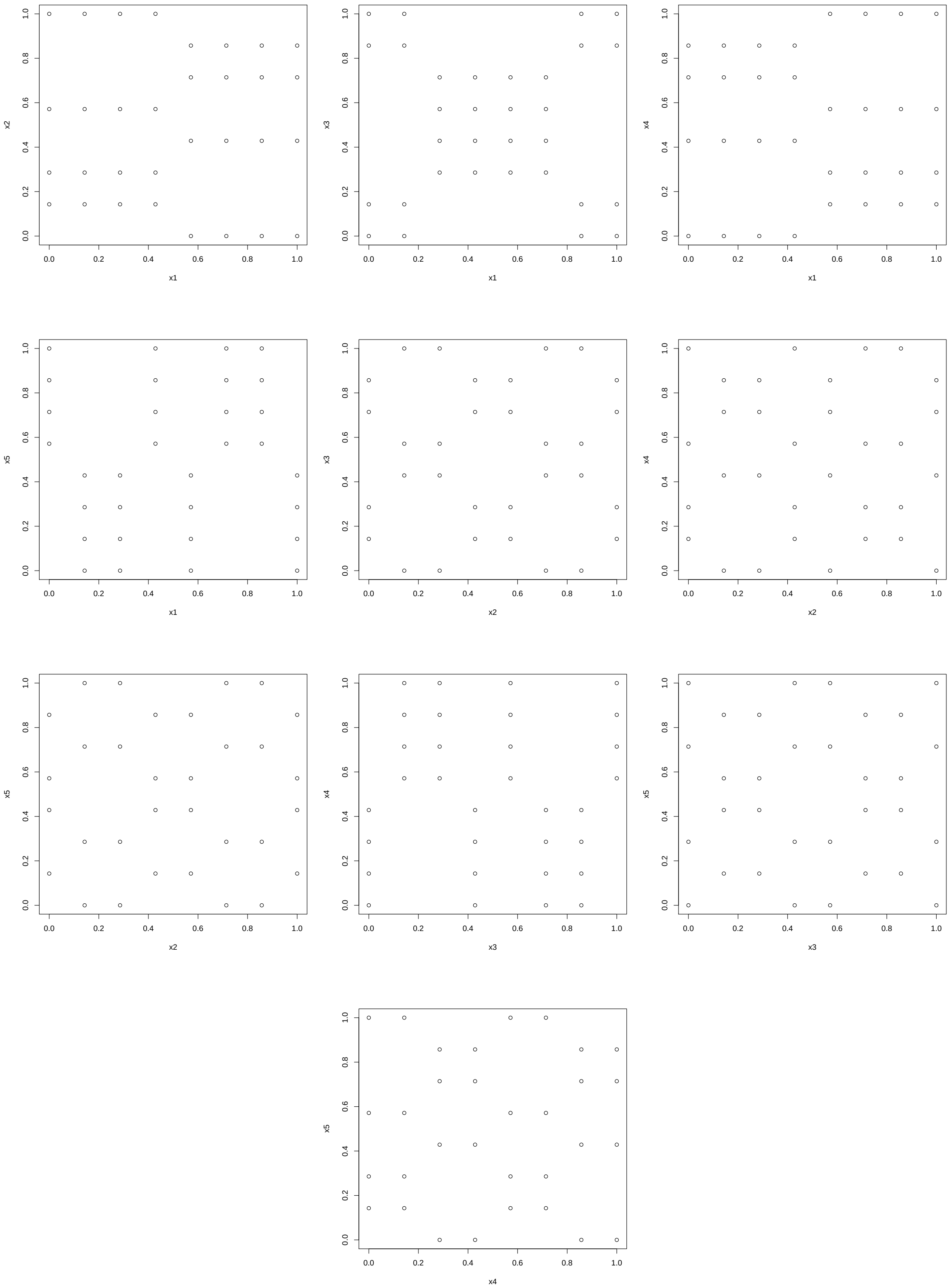


Figure 23: The MBR_{odd} design which got the worst maximum value for the RMSE criterion. Its confounding pattern is shown in table 58.

6.5 Section Summary

In this section we have constructed LHS, OA and MBR designs with $n = 32$ sample points and $s = 2$, 3, 4 and 5 factors. We constructed 1000 designs of each combination of design type, number of factors and 32 sample points, without optimizing.

For all number of factors there were some tendencies. We saw that the LHS designs performed poorly compared to the other design types for all number of factors. But as we constructed designs with an increasing number of factors the LHS design performed better. The MBR_{odd} and MBR_{all} designs performed good for the RMSE and the MAX criterion. For two factors the MBR design did especially good, but as we needed more factors the $\text{OA}_{p=4}$ did better.

The MBR designs have a large and increasing difference between the minimum and maximum values for the comparison criteria. The OA designs also have an increasing difference, but not as much. The $\text{OA}_{p=4}$ has a smaller difference between the designs' maximum and minimum values for the comparison criteria and may be less risky to use.

When comparing the different ways of constructing fractional factorial designs for the MBR designs, we see that the MBR_{all} scores well for the MAX criterion while the MBR_{odd} scores well for the RMSE criterion. The MBR_{IV} scores much like the MBR_{odd}, but without the extremes for the minimum and maximum values. This we could have expected since the two ways of finding confounding patterns are similar.

A problem which increased with a higher number of factors was the construction of MBR designs having replicated sample points. For two factors this was not a problem, but for five factors we had to construct 10 times as many MBR_{all} designs as we desired.

7 Discussion

In this paper two comparison studies have been conducted:

1. We constructed and optimized designs using three different optimizing criteria, the ASSC, L- and A_{mod}-criteria. We compared the optimizing criteria by checking how correlated they were with the comparison criteria, MAX and RMSE. The comparison was carried out on designs based on OA, LHS and MBR using different confounding patterns, and in addition random designs. All designs had five factors and 32 sample points.

2. We constructed and compared different designs with two, three, four and five factors all having 32 sample points, to see which designs were best we compared them by using the comparison criteria, MAX, RMSE and ASSC. We constructed LHS, OA and MBR designs using different confounding patterns.

The L- and A_{mod}-optimizing criteria were not well correlated with the MAX and RMSE criteria. The A_{mod}-criterion was a little better than the L-criterion, but neither of them were well correlated with the comparison criteria, indicating that they are not suited to use for optimizing.

The alias sum of squares criterion was also used as an optimizing criterion. The ASSC was more correlated with the MAX and RMSE comparison criteria than the L-and A_{mod}-optimizing criteria. Especially for the MBR designs the ASSC was more correlated with the comparison criteria than the other two optimizing criteria. The difference in correlation values for the different designs may indicate that we should consider the type of design we wish to construct when choosing which optimizing criterion to use.

The correlation coefficients for the alias sum of squares criterion and the comparison criteria, MAX and RMSE, in section 5.3 were smaller than the values Bursztyn & Steinberg (2004) obtained, which scored between 0.6 and 0.88. If we look at the correlation coefficients for the non-optimized designs, the correlation between the ASSC and the RMSE criterion ranges from 0.175 to 0.675, where the OA design has the minimum score, the second lowest correlation coefficient was 0.413. Also for the MAX criterion the correlation coefficients were substantially higher when we used non-optimized designs. The OA had the smallest correlation, 0.186, the second lowest was 0.390 and the highest was 0.579. These values are closer to the results in Bursztyn & Steinberg (2004) and may suggest that we got different values because we chose from 100 designs and used different true and estimated models. We used different methods for constructing designs than Bursztyn & Steinberg (2004) did. All our design types had varying correlation coefficients, this may also be part of the reason for us to obtain different values.

When we for the non-optimized designs compare the different design types we see that for a small number of factors the MBR designs perform well. For two factors the MBR design constructed using a fractional factorial design of the highest possible resolution, VI, scores well, its minimum values are just marginally higher than for the best designs. The MBR_{VI} design's only variation comes from the use of the fractional factorial design factors as binary bits in the MBR-factors. In the end we only have two different MBR designs which both score the same for the comparison criteria.

For designs having 3, 4 and 5 factors the $\text{OA}_{p=4}$ did better as the number of factors increased. The MBR_{odd} and MBR_{IV} designs performed similarly, as we might expect since they are both resolution IV designs. The MBR_{all} performed well for the MAX criterion, while the MBR_{odd} performed good for the RMSE criterion. For the ASSC the OA designs performed over-all the best.

The difference between a design type's minimum and maximum value for the comparison criteria was substantially higher for the MBR designs than for the OA and LHS designs. Especially the MBR_{odd} designs have a high difference between a comparison criterion's min and max values. This could make the use of the design risky, especially when an optimizing criterion is not used.

When we compare the tables in section 5 and the tables in section 6.4, in both sections the designs have 5 factors, we see that the maximum values in section 6.4 are for most designs and comparison criteria substantially higher than for the optimized designs in section 5. This implies that even though the optimizing and comparison criteria used in this study do not comply well, the use of an optimizing criterion is favorable. The design having the smallest difference between the maximum values for the comparison criteria when using and not using an optimizing criterion is the $\text{OA}_{p=4}$ design. This combined with the small difference between the minimum and maximum values for $\text{OA}_{p=4}$ designs' comparison criteria indicates that OA designs are for most part equally space-filling.

A problem that occurred when constructing MBR designs having an increasing number of factors was replicated sample points. As we needed more factors a higher number of designs were constructed with, and needed to be discarded because of, replicated sample points in one or several of the designs' two-dimensional projections. This occurred especially often for the MBR_{all} designs, and can be seen as a drawback to the multi-level binary replacement designs.

When constructing MBR designs we have restrictions to the number of levels possible to use, they need to be a power of 2. Also the OA design has restrictions, the number of possible sample points is given by the relationship $n = \lambda p^t$ and the OA having the desired number of levels and strength may not exist. These restrictions are drawbacks, especially when compared to the random or LHS designs.

In the two-dimensional plots for the MBR designs in section 6 all have symmetrical patterns. We see that this can sometimes leave large areas empty, which is not desirable. The MBR designs need to be constructed with a well chosen confounding pattern or using an optimizing criterion. Since the different confounding patterns do well for different criteria they may be suitable for different types of true and approximated models.

References

Bursztyn, D. & Steinberg, D. (2004), 'Comparison of designs for computer experiments', *Journal of Statistical Planning and Inference* .

Fang, K., Li, R. & Sudjianto, A. (2006), *Design and modeling for computer experiments*, CRC Press.

Giunta, A., Wojtkiewicz Jr, S. & Eldred, M. (2003), 'Overview of modern design of experiments methods for computational simulations', *AIAA* **649**, 6–9.

Hedayat, A., Sloane, N. & Stufken, J. (1999), *Orthogonal arrays*, Springer New York.

Martens, H., Måge, I., Tøndel, K., Isaeva, J., Høy, M. & Sæbø, S. (2010), 'Multi-level binary replacement (MBR) design for computer experiments in high-dimensional nonlinear systems', *Journal of Chemometrics* **24**(11-12), 748–756.

Myers, R. & Montgomery, D. (1995), 'Response Surface Methodology: Process and Product Optimization Using Designed Experiments', *New York: John Wiley & Sons, INC* pp. 280–318.

Nesbakken, A. (2011), 'Evaluation of Modern Design Methods for use in Computer Experiments', *Masters' thesis* .

Praveen, C. & Duvigneau, R. (2007), 'Radial basis functions and Kriging metamodels for aerodynamic optimization', *Research Report* **6151**.

R.H. Myers, S. M. K. Y. & Walpole, R. (2007), *Probability and statistics for engineers and scientists*, Pearson.

Santner, T., Williams, B. & Notz, W. (2003), *The design and analysis of computer experiments*, Springer Verlag.

Simpson, T., Lin, D. & Chen, W. (2001), 'Sampling strategies for computer experiments: design and analysis', *International Journal of Reliability and Applications* **2**(3), 209–240.

Simpson, T., Peplinski, J., Koch, P. & Allen, J. (2001), 'Metamodels for computer-based engineering design: survey and recommendations', *Engineering with computers* **17**(2), 129–150.

Tøndel, K., Gjuvsland, A., Måge, I. & Martens, H. (2010), 'Screening design for computer experiments: Metamodelling of a deterministic mathematical model of the mammalian circadian clock', *Journal of Chemometrics* **24**(11-12), 738–747.

A Appendix

A.1 True and Approximated Models Used in Sections 5 and 6

A.1.1 Five Factors

The true model used for designs constructed having $s = 5$ factors and $n = 32$ sample points.

$$
\begin{aligned}
y = {} & 0.59345729x_1 + 0.60503367x_2 + 0.96858356x_3 + 0.92468047x_4 + 0.38682111x_5 \\
& + 0.83248335x_1^2 + 0.26167091x_2^2 + 0.37436778x_3^2 + 0.22220435x_4^2 + 0.39863743x_5^2 \\
& + 0.67992135x_1x_2 + 0.42341648x_1x_3 + 0.04550819x_1x_4 + 0.28722105x_1x_5 \\
& + 0.75918545x_2x_3 + 0.42294244x_2x_4 + 0.36999552x_2x_5 + 0.71543520x_3x_4 \\
& + 0.45206398x_3x_5 + 0.79786643x_4x_5 + 0.06391047x_1^3 + 0.82238476x_1^2x_2 \\
& + 0.98251443x_1^2x_3 + 0.43446233x_1^2x_4 + 0.77772362x_1^2x_5 + 0.16910887x_1x_2^2 \\
& + 0.63712073x_1x_2x_3 + 0.68123379x_1x_2x_4 + 0.81245634x_1x_2x_5 + 0.06325685x_1x_3^2 \\
& + 0.02288149x_1x_3x_4 + 0.96875223x_1x_3x_5 + 0.11234526x_1x_4^2 + 0.57285054x_1x_4x_5 \\
& + 0.58808068x_1x_5^2 + 0.89780525x_2^3 + 0.45962887x_2^2x_3 + 0.44138936x_2^2x_4 \\
& + 0.84764645x_2^2x_5 + 0.82670836x_2x_3^2 + 0.88098623x_2x_3x_4 + 0.52923310x_2x_3x_5 \\
& + 0.02009340x_2x_4^2 + 0.98971092x_2x_4x_5 + 0.18697230x_2x_5^2 + 0.8908394x_3^2x_3 \\
& + 0.5352430x_3^2x_4 + 0.3060528x_3x_4x_5 + 0.7620614x_3x_4^2 + 0.3004369x_3x_4x_5 \\
& + 0.9446652x_3x_5^2 + 0.2745175x_4^3 + 0.1682710x_4^2x_5 + 0.1726328x_4x_5^2 + 0.2723786x_5^3
\end{aligned}
$$

The approximated model used for designs constructed having $s = 5$ factors and $n = 32$ sample points:

$$
\begin{aligned}
\hat{y} = {} & \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3c + \beta_4x_4 + \beta_5x_4 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{14}x_1x_4 \\
& + \beta_{15}x_1x_5 + \beta_{23}x_2x_3 + \beta_{24}x_2x_4 + \beta_{25}x_2x_5 + \beta_{34}x_3x_4 + \beta_{35}x_3x_5 + \beta_{45}x_4x_5
\end{aligned}
$$

A.1.2 Four Factors

The true model used for designs constructed with $s = 4$ factors and $n = 32$ sample points.

$$\begin{aligned}
y = {} & 0.59345729x_1 + 0.60503367x_2 + 0.96858356x_3 + 0.92468047x_4 + 0.83248335x_1^2 \\
& + 0.26167091x_2^2 + 0.37436778x_3^2 + 0.22220435x_4^2 + 0.67992135x_1x_2 + 0.42341648x_1x_3 \\
& + 0.04550819x_1x_4 + 0.75918545x_2x_3 + 0.42294244x_2x_4 + 0.71543520x_3x_4 \\
& + 0.06391047x_1^3 + 0.82238476x_1^2x_2 + 0.98251443x_1x_1x_3 + 0.43446233x_1^2x_4 \\
& + 0.16910887x_1x_2^2 + 0.63712073x_1x_2x_3 + 0.68123379x_1x_2x_4 + 0.06325685x_1x_3^2 \\
& + 0.02288149x_1x_3x_4 + 0.11234526x_1x_4^2 + 0.89780525x_2^2 + 0.45962887x_2^2x_3 \\
& + 0.44138936x_2^2x_4 + 0.82670836x_2x_3^2 + 0.88098623x_2x_3x_4 + 0.02009340x_2x_4^2 \\
& + 0.8908394x_3^3 + 0.5352430x_3^2x_4 + 0.7620614x_3x_4^2 + 0.2745175x_4^3
\end{aligned}$$

The approximated model used for designs constructed having $s = 4$ factors and $n = 32$ sample points:

$$\begin{aligned}
\hat{y} = {} & \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{14}x_1x_4 \\
& + \beta_{23}x_2x_3 + \beta_{24}x_2x_4 + \beta_{34}x_3x_4
\end{aligned}$$

A.1.3 Three Factors

The true model used for designs constructed with $s = 3$ factors and $n = 32$ sample points.

$$\begin{aligned}
y = {} & 0.59345729x_1 + 0.60503367x_2 + 0.96858356x_3 + 0.92468047x_4 + 0.83248335a^2 \\
& + 0.26167091x_2^2 + 0.37436778x_3^2 + 0.22220435x_4^2 + 0.67992135x_1x_2 \\
& + 0.42341648x_1x_3 + 0.04550819x_1x_4 + 0.75918545x_2x_3 + 0.42294244x_2x_4 \\
& + 0.71543520x_3x_4 + 0.06391047x_1^3 + 0.82238476x_1^2x_2 + 0.98251443x_1^2x_3 \\
& + 0.43446233x_1^2x_4 + 0.16910887x_1x_2^2 + 0.63712073x_1x_2x_3 + 0.68123379x_1x_2x_4 \\
& + 0.06325685x_1x_3^2 + 0.02288149x_1x_3x_4 + 0.11234526x_1x_4^2 + 0.89780525x_2^2 \\
& + 0.45962887x_2^2x_3 + 0.44138936x_2^2x_4 + 0.82670836x_2x_3^2 + 0.88098623x_2x_3x_4 \\
& + 0.02009340x_2x_4^2 + 0.8908394x_3^3 + 0.5352430x_3^2x_4 + 0.7620614x_3x_4^2 + 0.2745175x_4^3
\end{aligned}$$

The approximated model used for designs constructed having $s = 3$ factors and $n = 32$ sample points:

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3$$

A.1.4 Two Factors

The true model used for designs constructed with $s = 2$ factors and $n = 32$ sample points.

$$\begin{aligned} y =& 0.59345729x_1 + 0.605033676 + 0.83248335x_1^2 + 0.26167091x_2^2 + 0.67992135x_1x_2 \\ &+ 0.06391047x_1^3 + 0.82238476x_1^2x_2 + 0.16910887x_1x_2^2 + 0.89780525x_2^3 \end{aligned}$$

The approximated model used for designs constructed having $s = 2$ factors and $n = 32$ sample points:

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2$$

A.2 The Confounding Patterns Chosen by Minitab

In sections 5 and 6 we used resolution IV confounding patterns to construct MBR designs with 3, 4 and 5 factors, these were found by using Minitab. These confounding patterns are as follows:
For MBR designs having five factors and resolution IV:

$$\begin{aligned} F &= ABC & G &= ABD \\ H &= ABE & J &= ACD \\ K &= ACE & L &= ADE \\ M &= BCD & N &= BCE \\ O &= BDE & P &= CDE \end{aligned}$$

For MBR designs having four factors and resolution IV:

$$\begin{aligned} F &= ACE & G &= ACD \\ H &= ABD & J &= ABE \\ K &= CDE & L &= ABCDE \\ M &= ADE \end{aligned}$$

For MBR designs having three factors and resolution IV:

$$\begin{aligned} F &= BCDE & G &= ACDE \\ H &= ABDE & J &= ABCE \end{aligned}$$

Lastly the confounding patterns used for contructing MBR designs having two factors and resolution VI:

$$F = ABCDE$$

A.3 Correlation Coefficients from Sections 5 and 6

A.3.1 The Correlations Coefficients for the Optimizing and Comparison criteria

	LHS	$OA_{p=2}$	$OA_{p=4}$	Random	MBR_{JT}	MBR_{H}	MBR_{IV}
MAX	0.0904	0.1026	0.0665	0.1005	0.1560	0.1772	0.2009
RMSE	0.0837	0.1335	0.0802	0.1717	0.0242	0.1311	0.0843
ASSC	0.2606	0.1335	0.1184	0.3574	−0.0556	0.2674	−0.0986

Table 59: The correlation coefficients for the A_{mod}-criterion values and the comparison criteria values.

	LHS	$OA_{p=2}$	$OA_{p=4}$	Random	MBR_{JT}	MBR_{H}	MBR_{IV}
MAX	−0.1665	−0.0539	−0.0427	−0.1537	−0.1563	−0.1182	−0.1741
RMSE	−0.1426	−0.0300	−0.0865	−0.1329	−0.0887	−0.1346	−0.0390
ASSC	−0.2283	−0.1430	−0.0962	−0.1633	−0.0105	−0.2631	0.0077

Table 60: The correlation coefficients for the L-criterion values and the comparison criteria values.

	LHS	$OA_{p=2}$	$OA_{p=4}$	Random	MBR_{JT}	MBR_{H}	MBR_{IV}
MAX	0.1071	0.1132	0.0727	0.0186	0.2438	0.1051	0.1944
RMSE	0.1763	0.1406	0.0197	0.0876	0.4273	0.3773	0.4210

Table 61: The correlation coefficients for the alias sum of squares criterion values and the comparison criteria values.

A.3.2 The Correlation Coefficients for the ASSC and the Comparison Criteria for Non-Optimized Designs

	LHS	$OA_{p=4}$	MBR_{JT}	MBR_{H}	MBR_{IV}
MAX	0.5186	0.1859	0.5792	0.3975	0.5789
RMSE	0.4769	0.1754	0.6716	0.4786	0.5976

Table 62: The correlation coefficients for the alias sum of squares criterion values and the comparison criteria values when constructing designs without optimizing.