



# The depth-design: An efficient generation of high dimensional computer experiments



Martin Piffl<sup>a,\*</sup>, Ernst Stadlober<sup>b</sup>

<sup>a</sup> AVL List GmbH, Hans-List-Platz 1, 8020 Graz, Austria

<sup>b</sup> Graz University of Technology, Institute of Statistics, NAWI Graz, Kopernikusgasse 24/III, 8010 Graz, Austria

## ARTICLE INFO

### Article history:

Received 1 July 2014

Received in revised form 5 March 2015

Accepted 6 March 2015

Available online 16 March 2015

### Keywords:

Multiple linear regression

Screening designs

Total sensitivity index

Space filling designs

Median-oriented quantiles

Depth functions

## ABSTRACT

This paper provides an approach on how to generate representative experiments for the investigation of a model based system or process, depending on quantitative variables, when the number of experiments  $N$  is limited ( $25 \leq N \leq 500$ ). An exemplified overview of known screening designs that are suitable for quadratic response surfaces possibly depending on  $k \geq 50$  factors is given. The relevance of these factors is measured by a sensitivity index, which is based on corresponding sums of squares of the underlying linear, quadratic as well as the linear two way interaction effects. Bearing in mind the sparsity-of-effects principle, we expect the process or system to be dominated only by a minority of the factors ( $k_r \leq 10$ ) assumed. Among other space filling designs we especially investigate the very efficient Latin Hypercube Design in terms of its capability to represent a multidimensional distribution with its experiments. We use the theory of median-oriented quantiles and depth functions to assess this capability and to introduce our new space filling design approach, the Depth-Design. On the example of the multivariate normal distribution we demonstrate that our Depth-Design represents a multidimensional distribution with much less experiments in comparison to the Latin Hypercube design.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Computers have become an indispensable tool in all technical fields where complex processes need to be investigated and optimized. Although simulation models are becoming more and more accurate, the associated increase in complexity of these models has hesitated much faster computation times and has caused confusing frameworks over the last twenty years (cf. [Currin et al., 1991](#); [Siebertz et al., 2010](#)).

Design of experiments (DOE) minimizes the required effort of simulations (or “simply” experiments) to be run for an investigation of a response variable  $Y$  within given system boundaries that enclose the so called feature space. DOE additionally generates the experiments necessary to optionally set up a fast predicting regression model of  $Y$ , where complex system interactions can be easily tracked. These advantages make DOE a very attractive tool, which can be used to overcome the difficulties coming along with present day computer simulation models ([Montgomery, 2012](#)). Nevertheless, computer simulation results are frequently non-linearly determined by a large number ( $k \geq 50$ ) of factors  $X_1, \dots, X_k$  so that most usual DOE approaches, like fractional factorial designs or the central composite designs, would exhibit a too large number of experiments, or suffer from ambiguous possibilities of interpretation. Given that the feature space is subject

\* Corresponding author. Tel.: +43 660 657 37 38.

E-mail address: [martin@piffl.com](mailto:martin@piffl.com) (M. Piffl).

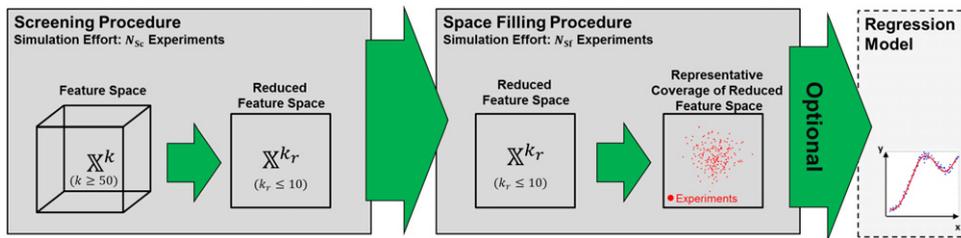


Fig. 1. Guideline to generate representative experiments.

to a multidimensional distribution, a representative experimental coverage is a challenge to be additionally mastered. For these reasons, this paper provides a guideline on how to identify the most significant factors  $X_1, \dots, X_{k_r}$  for  $Y$ , and how to represent the distribution of the remaining feature space with a limited number of experiments.

In Section 2 we compare 3-level screening designs and emphasize the power of the Definitive Screening Design of Jones and Nachtshiem (2011) in the context of computer simulation models. We recall the total sensitivity index of Homma and Saltelli (1996) in order to identify the most significant factors. Thereafter, in Section 3 we compare the capability of the Latin Hypercube design to generate representative experiments for a multidimensional distribution with the aid of median-oriented quantiles and depth functions, discussed in detail by Serfling (2010). With regard to the results obtained we develop our new space filling design approach, the Depth-Design, which is presented in Section 4. In the end we conclude our findings and present our recommendations.

## 2. Generation of representative experiments

In this paper we assume that the random vector  $\mathbf{X} = (X_1, \dots, X_k) \sim F^k$  represents all feasible combinations of input factors of a computer simulation model, whereas  $X_1, \dots, X_k$  are considered to be quantitative variables. The domain of the multidimensional distribution  $F^k$  is denoted as the feature space  $\mathbb{X}^k \subseteq \mathbb{R}^k$ , which contains all feasible experiments realized by vectors  $\mathbf{x} = (x_1, \dots, x_k)$  of  $\mathbf{X}$ . Furthermore, we are interested in the simulation results of one response variable  $Y$ , obtained by experiments  $\mathbf{x} \in \mathbb{X}^k$ .

Given that only one response variable  $Y$  is of interest, and the number of experiments is constrained (i.e.  $25 \leq N \leq 500$ ), we propose to generate representative experiments not only with regard to the feature space distribution  $F^k$ , but also with regard to the response  $Y$ . This is carried out as process (see Fig. 1) consisting of a screening and a space filling procedure. We use the rule of thumb that most processes or systems are dominated by a few factors. A preliminary screening procedure, using a portion  $N_{sc}$  of the number feasible experiments  $N$ , can cheaply reduce all considered factors  $X_1, \dots, X_k$  to the most significant factors  $X_1, \dots, X_{k_r}$  (usually  $k_r \leq 10$ ) for  $Y$ . As a consequence, it is possible to reduce the feature space  $\mathbb{X}^k$  to a subspace  $\mathbb{X}^{k_r}$ . The associated reduction in dimension of the feature space eases the experimental coverage of the resulting multidimensional distribution  $F^{k_r}$  with  $N_{sf}$  experiments during the subsequent space filling procedure. Once, representative data is gathered, there exists the option to build a regression model. Still, it remains to be clarified, which type of screening design and which type of space filling design are most appropriate for  $F^{k_r}$ , when the number of experiments  $N = N_{sc} + N_{sf}$  is limited.

### 2.1. Screening designs

Screening designs enable the identification of the most relevant factors  $X_1, \dots, X_k$  in terms of the response  $Y$  with comparable low simulation effort. Given a fixed number of factors  $k$  to investigate, screening designs mainly differ in the number of experiments  $N_{sc}$  and in the interpretability of the estimated effects. While 2-level screening designs, share the idea of comparing the results of two different levels “−1” and “+1” (extreme case scenarios), 3-level screening designs additionally consider the factors at a center level “0” in order to detect possible curvature in the relationship between  $X_1, \dots, X_k$  and  $Y$ .

For the screening procedure factors  $X_1, \dots, X_k$  are assumed as independent random variables. This simplification may be especially feasible for computer simulation models, and may only lead to a larger number of selected factors  $k_r$  obtained by the screening procedure due to neglected correlation structure. If the independent setting of two or more factors is not possible, it is suggested to successively neglect such factors until an independent consideration becomes possible. It is proposed to standardize the feasible ranges of  $X_1, \dots, X_k$ , which need to be chosen after good engineering judgment, to  $[-1, +1]$  so that eventually  $\mathbb{X}^k = [-1, +1]^k$ . Scientists and engineers do often feel more comfortable with 3-level screening designs, because they tend to expect a substantial non-linear relationship between the factors  $X_1, \dots, X_k$  and the response  $Y$ . As a result, the straightforward application of 3-level full factorial designs ( $3^k$  designs) is not possible, because the number of experiments  $N_{sc}$  explodes when  $k \geq 50$ . Therefore, we shortly discuss the following alternative 3-level screening designs.

1.  $3^{k-p}$  fractional factorial design (Montgomery, 2012)
2. Box–Behnken Design (Box and Behnken, 1960)

- 3. Face-Centered Central Composite Design (Box and Wilson, 1951)
- 4. Definite Screening Design (Jones and Nachtsheim, 2011).

At closer inspection, the  $3^{k-p}$  fractional factorial design comes along with a major disadvantage, as each effect has  $(3^p - 1) / 2$  different aliases. If the number of fractions  $p$  is increased in order to decrease the number of experiments, the computation of  $3^{k-p}$  fractional factorial designs with clear main effects becomes very challenging. Xu (2005) proposes a method based on coding theory to identify clear main effects of large  $3^{k-p}$  fractional factorial designs. However, his method is only applicable for  $k \leq 20$  factors.

The very efficient Box–Behnken design, however, does not provide the vertices of the feature space. This property may indeed be useful for applications, where extreme factor combinations are not feasible. Still, if the factors are allowed to follow any distribution, this property is not desirable at all.

An often proposed alternative to a  $3^k$  design is the 3-level Face-Centered Central Composite design, which expands a 2-level full factorial design ( $2^k$  design) by introducing  $2 \cdot k$  axial experiments and multiple experiments in the center of the feature space to cope with non-linearity. Nevertheless, the simple application of a  $2^k$  design is too expensive when the number of factors  $k$  is large. In order to become numerically efficient, it is possible to replace the  $2^k$  design by a  $2^{k-p}$  fractional factorial design with resolution IV. However, face-centered central composite designs show a very poor precision in the estimation of pure quadratic effects, which contradicts the original intended purpose of a 3-level screening design (Siebertz et al., 2010).

Cheng and Wu (2001), Tsai et al. (2000) and Jones and Nachtsheim (2011) address their studies to non-regular 3-level designs for screening experiments. The work of Jones and Nachtsheim and their Definitive Screening Design (DS-Design) with  $N_{Sc} = 2 \cdot k + 1$  experiments should be emphasized at this point. Based on the strategy to minimize the correlation structure among second order effects by the coordinate exchange algorithm and the target to achieve uncorrelated main effects by the fold-over technique, they produced 3-level orthogonal designs with notable properties. Jones and Nachtsheim itemize their characteristics in the following way:

1. The number of required experiments is only one more than twice the number of factors ( $N_{Sc} = 2 \cdot k + 1$ ).
2. Unlike resolution III designs, main effects are completely independent of two-factor interactions.
3. Unlike resolution IV designs, two-factor interactions are not completely confounded with other two-factor interactions, although they may be correlated.
4. Unlike resolution III, IV and V designs with added center experiments, all quadratic effects are estimable in models comprised of any number of linear and quadratic main-effects terms.
5. Quadratic effects are orthogonal to main effects and not completely confounded (though correlated) with interaction effects.

However, Jones and Nachtsheim constructed their designs with a complicated algorithm, which was not able to find orthogonal designs for  $k = 12$ . Furthermore, this algorithm suffered from slow convergence speed in cases where  $k$  was large. In the paper of Xiao et al. (2012), the generation of the DS-Design was completed by a well applicable construction method based on conference matrices. Although conference matrices do not exist for  $k$  factors, if  $k \equiv 2 \pmod 4$  and  $k - 1 \neq a^2 + b^2$  for  $a$  and  $b$  integers (Van Lint and Seidel, 1966), it is always possible to increase  $k$ , construct a DS-Design and finally delete the dispensable columns. As long as no design rows are deleted, orthogonality is not affected by this procedure. Altogether, these designs provide the possibility to estimate pure main effects, quadratic effects and interaction effects with one run-through. Hence, the DS-Design seems to be a very competitive screening tool for all kinds of processes, where the number of factors  $k$  is large and where a non-linear relationship between input factors and response is expected.

### 2.2. Evaluation of the Definitive Screening Design

By construction, the non-regular DS-Design provides  $2 \cdot k + 1$  degrees of freedom. In principle, this allows the estimation of  $2 \cdot k$  different effects. Still, the assessment of all main effects, 2nd order interactions and quadratic effects requires  $2 \cdot k + \binom{k}{2} + 1$  degrees of freedom. The full power of the DS-Design can only be examined if all of these effects are estimated. Pair-by-pair comparisons using orthogonal contrasts can be a possible remedy to estimate all effects (Wu and Hamada, 2009). For two arbitrary factors  $A := X_j$  and  $B := X_{j'}$  ( $1 \leq j < j' \leq k$ ) the following linear (“L”) and quadratic (“Q”) effects can be estimated.

$$\underbrace{A_L, A_Q, B_L, B_Q, AB_{L \times L}}_{\text{of interest}}, \overbrace{AB_{L \times Q}, AB_{Q \times L}, AB_{Q \times Q}}^{\text{not considered}} \cdot \tag{1}$$

Due to the linear dependencies of the DS-Design, estimates of  $AB_{L \times Q}, AB_{Q \times L}, AB_{Q \times Q}$  can be evaluated independently only once per model. Furthermore, these interactions are hard to interpret and should be rather used in a fine tuning step than during the screening procedure. The associated square sums of the contrasts are compared with the total variation of the response outlined by the total square sum. To guarantee that all effects of interest can be estimated, the ANOVA procedure is conducted sequentially for all  $\binom{k}{2} = \frac{k \cdot (k+1)}{2}$  factor pairs.

The computer simulation results of the experiments run shall be denoted by  $y_n$ , where  $n = 1, \dots, N_{Sc}$  for  $N_{Sc} = 2 \cdot k + 1$  under the DS-Design. Using the contrast weights of [Hinkelman and Kempthorne \(1994\)](#), the linear contrast  $C_{A_L}$  of factor  $A$  is computed in the following way.

$$C_{A_L} = -1 \cdot \sum_{n=1}^{N_{Sc}} y_n \cdot \mathbf{1}_{\{A=-1\}} + 0 \cdot \sum_{n=1}^{N_{Sc}} y_n \cdot \mathbf{1}_{\{A=0\}} + 1 \cdot \sum_{n=1}^{N_{Sc}} y_n \cdot \mathbf{1}_{\{A=+1\}} = -1 \cdot y_- + 0 \cdot y_0 + 1 \cdot y_+. \tag{2}$$

The wrapped fold-over technique of the DS-Design enables the independent estimation of pure main effects. The corresponding square sum of the linear effect is obtained by the fraction of the squared contrast and the number of experiments in each group weighted by its squared contrast coefficient (cf. [Montgomery \(2012\)](#), p. 93),

$$SS_{A_L} = \frac{C_{A_L}^2}{\sum_{n=1}^{N_{Sc}} \mathbf{1}_{\{A=-1\}} \cdot (-1)^2 + \sum_{n=1}^{N_{Sc}} \mathbf{1}_{\{A=+1\}} \cdot (+1)^2} = \frac{C_{A_L}^2}{|y_-| \cdot (-1)^2 + |y_+| \cdot (+1)^2}. \tag{3}$$

Recording the pure quadratic effect of a factor requires experiments at the center level “0”. As the Definitive Screening Design only provides three center level experiments per factor, but  $(k - 1)$  experiments with levels “-1” and “+1”, the design is in general not balanced. By definition of a contrast, the scalar product between vector  $m$ , which contains the observations per level

$$m = (|y_-|, |y_0|, |y_+|), \tag{4}$$

and the contrast weight vector  $c = (c_{-1}, c_0, c_{+1})$  must be zero (cf. [Montgomery \(2012\)](#), p. 94). Hence, to comply with  $m^T c \stackrel{!}{=} 0$ , it is necessary to standardize to the smallest coefficient of  $m$ , which corresponds to  $|y_0| = 3$  for the DS-Design. The contrast weight vector for the quadratic contrast  $c = (1, -2, 1)$  needs to be replaced by  $c = \left(1 \cdot \frac{|y_0|}{|y_-|}, -2 \cdot \frac{|y_0|}{|y_0|}, 1 \cdot \frac{|y_0|}{|y_+|}\right)$  so that

$$m^T c = (k - 1) \cdot \frac{3}{(k - 1)} - 2 \cdot 3 + (k - 1) \cdot \frac{3}{(k - 1)} = 0 \tag{5}$$

holds for the DS-Design. The quadratic contrast  $C_{A_Q}$  of factor  $A$  can be estimated by

$$[C_{A_Q}] = c_{-1} \cdot \frac{|y_0|}{|y_-|} \cdot y_- + c_0 \cdot \frac{|y_0|}{|y_0|} \cdot y_0 + c_{+1} \cdot \frac{|y_0|}{|y_+|} \cdot y_+ = +1 \cdot \frac{3}{(k - 1)} \cdot y_- + (-2) \cdot y_0 + 1 \cdot \frac{3}{(k - 1)} \cdot y_+. \tag{6}$$

Even though orthogonal contrasts are used, it has to be noted that the contrast  $C_{A_Q}$  is biased due to the correlation structure of the DS-Design. Restriction of the design matrix to second order effects gives the correlation coefficients  $r(\cdot, \cdot)$ , which describe the coherence between  $[C_{A_Q}]$  and  $C_{A_Q}$  as follows

$$[C_{A_Q}] = C_{A_Q} + \sum_{E \neq A} r_{(A_Q, E_Q)} \cdot C_{E_Q} + \sum_{F \neq E} r_{(A_Q, EF_{L \times L})} \cdot C_{EF_{L \times L}}, \tag{7}$$

where factors  $E$  and  $F$  differ from factors  $A$  and  $B$ . [Jones and Nachtshiem \(2011\)](#) showed that the absolute value of correlation between pure quadratic effects converges towards  $\frac{1}{3}$  and between quadratic effects and 2nd order interaction effects towards  $\frac{1}{6}$  for  $k \rightarrow \infty$ .

For pure quadratic effects, the corresponding square sum is estimated by

$$[SS_{A_Q}] = \frac{([C_{A_Q}])^2}{|y_{-1}| \cdot \left(\frac{3}{(k-1)}\right) \cdot (-1)^2 + |y_0| \cdot (-2)^2 + |y_{+1}| \cdot \left(\frac{3}{(k-1)}\right) \cdot (+1)^2}. \tag{8}$$

Finally, the contrasts of the two-factor interaction effects have to be computed. Here, the contrast weights origin a  $(3 \times 3)$ -matrix derived from the matrix multiplication of the contrast weights of the involved effects

$$c_{AB_{L \times L}} = c(A_L)^T \times c(B_L) = (-1, 0, 1)^T \times (-1, 0, 1) = \begin{pmatrix} +1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & +1 \end{pmatrix}. \tag{9}$$

As a result, the contrast  $C_{AB_{L \times L}}$  of the two-factor interaction is estimated by

$$[C_{AB_{L \times L}}] = 1 \cdot y_{--} + (-1) \cdot y_{-+} + (-1) \cdot y_{+-} + 1 \cdot y_{++} \tag{10}$$

whereas

$$[C_{AB_{L \times L}}] = C_{AB_{L \times L}} + \sum_E r_{(AB_{L \times L}, E_Q)} \cdot C_{E_Q} + \sum_{\substack{E \neq F \\ EF \neq AB}} r_{(AB_{L \times L}, EF_{L \times L})} \cdot C_{EF_{L \times L}}. \tag{11}$$

The desired square sum is

$$[SS_{AB_{L \times L}}] = \frac{([C_{AB_{L \times L}}])^2}{|y_{--}| \cdot (+1)^2 + |y_{-+}| \cdot (-1)^2 + |y_{+-}| \cdot (-1)^2 + |y_{++}| \cdot (+1)^2}. \tag{12}$$

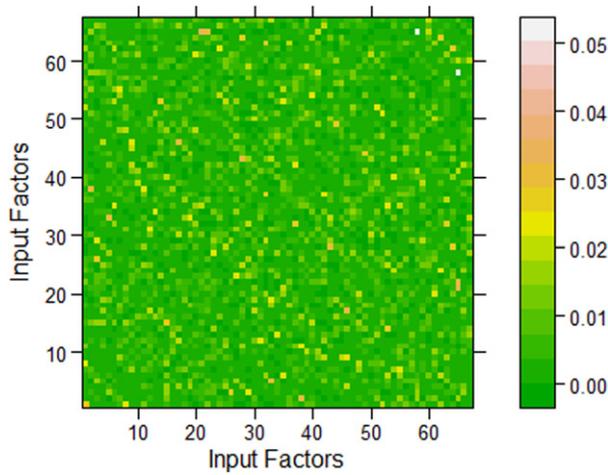


Fig. 2. Share of effect square sums in the total square sum.

2.3. Total sensitivity index

The total sum of squares  $SST = \sum_{n=1}^{N_{Sc}} (y_n - \bar{y})^2$  reflects the artificially generated deviation of the response  $Y$  from the average screening result  $\bar{y}$ . Whenever two factor pairs,  $A$  and  $B$ ,  $A$  and  $C$ , are evaluated by the introduced contrast approach,  $SST$  is partitioned into the following components

$$SST = SS_{A_L} + [SS_{A_Q}]_B + SS_{B_L} + [SS_{B_Q}]_A + [SS_{AB_{L \times L}}] + SSR_{AB}, \tag{13}$$

$$SST = SS_{A_L} + [SS_{A_Q}]_C + SS_{C_L} + [SS_{C_Q}]_A + [SS_{AC_{L \times L}}] + SSR_{AC}, \tag{14}$$

where the error sums of squares  $SSR_{AB}$ ,  $SSR_{AC}$  reflect the portion of  $SST$  not explained by the effects considered. The combination of factor  $A$  with another factor  $E \neq A$  entails  $k - 1$  different square sums  $[SS_{A_Q}]_E$ . Hence, it is proposed to take their mean value

$$\overline{[SS_{A_Q}]} = \frac{1}{k - 1} \cdot \sum_{E \neq A} [SS_{A_Q}]_E \tag{15}$$

as the desired statistic. With the total sensitivity index (TSI), introduced by Homma and Saltelli (1996), the portion of the total variance caused by factor  $A$  can be finally outlined as

$$ST_A = \frac{SS_{A_L} + \overline{[SS_{A_Q}]} + \sum_{E \neq A} [SS_{AE_{L \times L}}]}{SST}. \tag{16}$$

Originally, this index demands unbiased estimators of the applied square sums. In the worst case  $\overline{[SS_{A_Q}]}$  or  $[SS_{AE_{L \times L}}]$  are affected by an overall noise, which may obscure relevant effects or erroneously increase the significance of unimportant effects (cf. Fig. 2). A possible remedy is to consider a bias corrected sensitivity index, where the overall noise is removed with respect to a relevance level  $0 < \alpha \leq 0.1$ .

$$ST_{A_{corr(\alpha)}} = \frac{SS_{A_L} + \overline{[SS_{A_Q}]_{corr(\alpha)}} + \sum_{E \neq A} [SS_{AE_{L \times L}}]_{corr(\alpha)}}{SST} \tag{17}$$

with

$$\overline{[SS_{A_Q}]_{corr(\alpha)}} := \begin{cases} \delta_{A_Q} & \alpha \leq \frac{\overline{[SS_{A_Q}]}}{SST} \\ 0 & \text{else,} \end{cases} \tag{18}$$

where

$$\delta_{A_Q} = \left| \overline{[SS_{A_Q}]} - \text{mean}_{E \neq A} \left\{ \overline{[SS_{EQ}]} \cdot \mathbb{I} \left\{ \frac{\overline{[SS_{EQ}]}}{SST} < \alpha \right\} \right\} \right|. \tag{19}$$

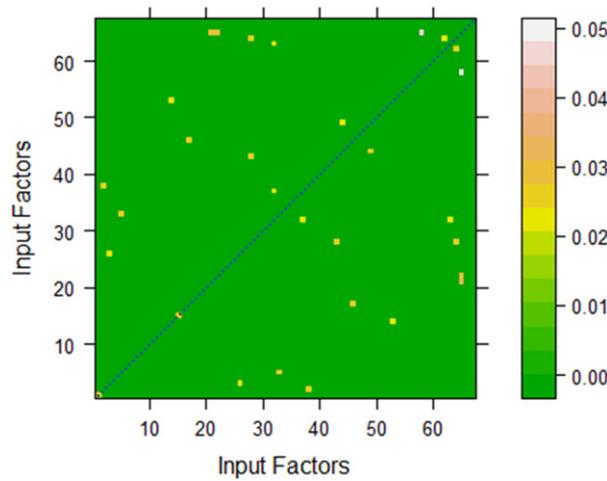


Fig. 3. Share of corrected effect square sums in the total square sum.

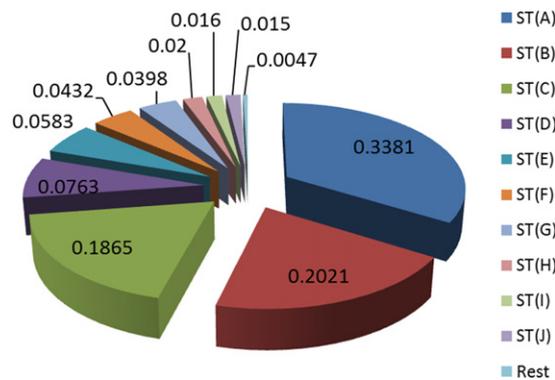


Fig. 4. Pie plot—corrected TSIs.

Accordingly, the same formulas apply to  $[SS_{AB_{L \times L}}]_{\text{corr}(\alpha)}$ . The impact of this correction becomes visible when Figs. 2 and 3 are compared. Fig. 2 contains the fractions  $[SS_{A_0}]/SST$  and  $[SS_{AB_{L \times L}}]/SST$ , determined by simulation results of a corresponding DS-Design. The results origin a computer simulation model, which takes into account  $k = 69$  factors. As visible, the total sensitivity index of many factors are likely to be influenced by a multiplicity of interaction effects. On the basis of the sparsity-of-effects principle, which is a well accepted rule of thumb in DOE (cf. Wu and Hamada, 2009; Mee, 2009; Montgomery, 2012), we expect that only a few decisive main effects and second order effects dominate the simulated process. As a consequence, we assume correlations of Eq. (11) to be the main reason that almost all investigated interactions pretend relevance. Fig. 3 illustrates the corresponding fractions after having applied the corrections outlined in Eq. (18) with  $\alpha = 0.025$ . As visible, two quadratic effects and only a few interaction effects are relevant. After a consultation with the model experts these effects have been confirmed. Moreover, no further effects than the identified effects have been expected to be significant by the experts. The results of the concerning sensitivity indices, calculated after definition (17), are visible in Figs. 4 and 5. We propose pie plots or bar diagrams to make clear which factors are significant and which factors can be neglected respectively. As expected, only  $k_r = 10$  of the  $k = 69$  considered factors dominate the investigated computer model based process in terms of the response  $Y$ .

We made the experience that the DS-Design of Jones and Nachtsheim (2011) with only  $N_{Sc} = 2 \cdot k + 1$  experiments is the perfect screening tool for computer simulation models where a non-linear relationship between factors and response is expected. On this account, we propose the application of the DS-Design as the first step in generating representative data, when the number of experiments is limited.

#### 2.4. Space-filling-designs

In the last subsection we propose the DS-Design of Jones and Nachtsheim (2011) as the first step of the process, outlined by Fig. 1, for the generation of representative experiments considering a multidimensional distribution. With the aid of the DS-Design it is efficiently possible ( $N_{Sc} = 2 \cdot k + 1$ ) to identify the most significant factors  $X_1, \dots, X_{k_r}$  for a response  $Y$ . As a result, a representative experimental coverage is facilitated, because the originally assumed feature space  $\mathbb{X}^k$  reduces to a

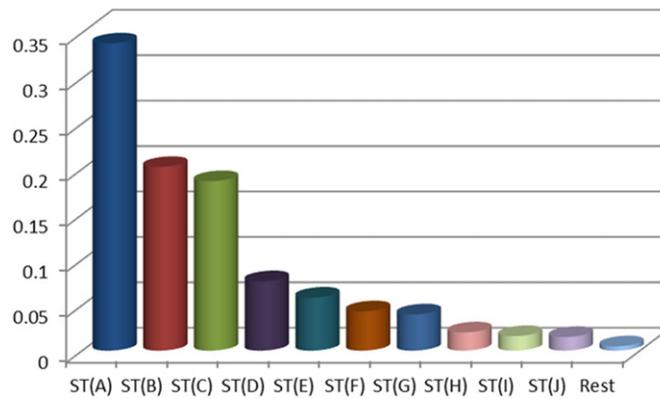


Fig. 5. Bar plot—corrected TSIs.

Table 1  
Standard space filling designs.

Design type	Background	Non-uniform distributions possible	Source
1. Sphere Packing	$\max \min_{x,y \in C^k} \ x - y\ _2$	No	Johnson et al. (1990)
2. Uniform	$\min_{C^k}$ Discrepancy	No	Fang (1980)
3. Maximum entropy	$\max_{C^k}$ Shannon entropy	Yes	Shewry and Wynn (1987)
4. IMSE	$\min_{C^k}$ MSE (residuals)	Yes	Sacks et al. (1989)
5. (Quasi) Monte Carlo	Law of large numbers	Yes	Johnson (1987)
6. Latin Hypercube	Permutation matrix	Yes	Mckay et al. (1979)

subspace  $\mathbb{X}^{k_r}$ . While the feature space distribution played a minor role during the screening procedure, for the space filling procedure we want to apply the theory of space filling designs in order generate representative experiments with regard to the corresponding distribution  $F^{k_r}$ .

The first space filling designs were constructed to uniformly distribute experiments over a feature space in an optimal manner. Hence, the experiments of these designs represent a multidimensional uniform distribution. As Montgomery (2012) points out, this is desirable, if the experimenter is not aware of the relationship between the input factors and the response, and if interesting phenomena are likely to be found in different regions of the feature space. The most popular design approaches, included in commercial software packages, are listed in Table 1.

For our purpose, a space filling design needs to provide experiments that represent the feature space distribution  $F^{k_r}$ , which generally differs from a multidimensional uniform distribution. The p.d.f.  $f^{k_r}$  of  $F^{k_r}$  is determined by

$$f^{k_r}(x_1, \dots, x_{k_r}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f^k(\mathbf{x}) dx'_1 \dots dx'_{k-k_r+1}, \tag{20}$$

where  $x'_1, \dots, x'_{k-k_r+1}$  correspond to the insignificant factors  $X'_1, \dots, X'_{k-k_r+1}$  eliminated by the screening procedure. Independent of their capability to represent other than uniform distributions, due to their long computation times in multidimensional spaces, we do not recommend space filling designs 1–4 of Table 1. Hence, they will not be considered anymore.

Given that the factors  $X_1, \dots, X_{k_r}$  are independent (i.e.  $X_j \stackrel{\text{ind}}{\sim} F_j$  for  $j = 1, \dots, k_r$ ), Fishman (1996) motivates Monte Carlo methods by directly merging pseudo random samples of the marginal distributions  $F_j$  to multidimensional vectors in order to get representative experiments (pseudo MC-Design). If the assumption of independent factors is too restrictive, there exists an easy transformation based on the Cholesky decomposition to generate representative experiments for a multivariate normal distribution  $F^{k_r}$  (Dagpunar, 1988). Moreover, Johnson (1987) presents several techniques on how to generate samples for other than multivariate normal distributions. If  $F^{k_r}$  is a multidimensional uniform distribution, a more efficient experimental coverage can be achieved by quasi Monte Carlo sampling techniques (Siebertz et al., 2010). Because the theory of non-uniform random numbers does not directly apply to the quasi Monte Carlo approach, however, we advise against this technique for other than uniform feature space distributions. Probably because of their simple applicability, the pseudo MC-Design has gained a considerable amount of recognition in engineering. Still, it has to be mentioned that the efficiency of such designs is heavily influenced by the assumed distribution. Especially when computer simulation models are used, regions of the feature space with small variation may lead to replicate experiments with similar simulation results.

The Latin Hypercube Design (LH-Design) of McKay et al. (1979) is also a well applied design in industry. The general idea behind the LH-Design is to avoid replicate experiments in order to increase simulation efficiency. Although the LH-Design is the first proposed space filling design, and it was originally intended, like other pioneer space filling designs, to generate experiments in an uniform manner, Stein (1987) introduced an approach to generate LH-Designs for general multidimensional distributions. If factors  $X_1, \dots, X_{k_r}$  are independent, the LH-Design is generated by partitioning the feature space  $\mathbb{X}^{k_r}$  into

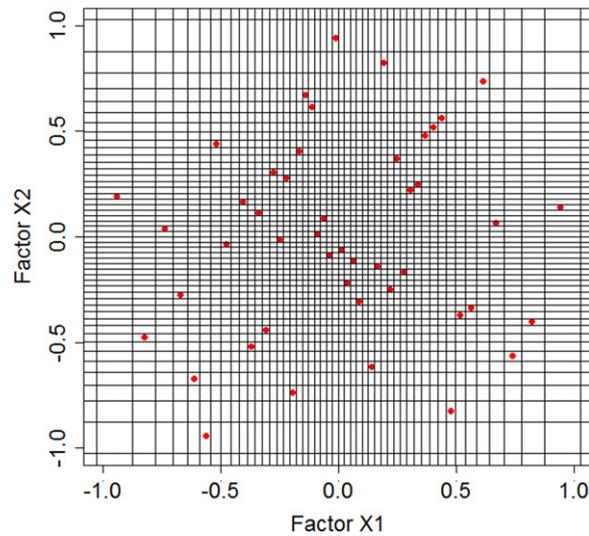


Fig. 6. LH-Design for independent factors.

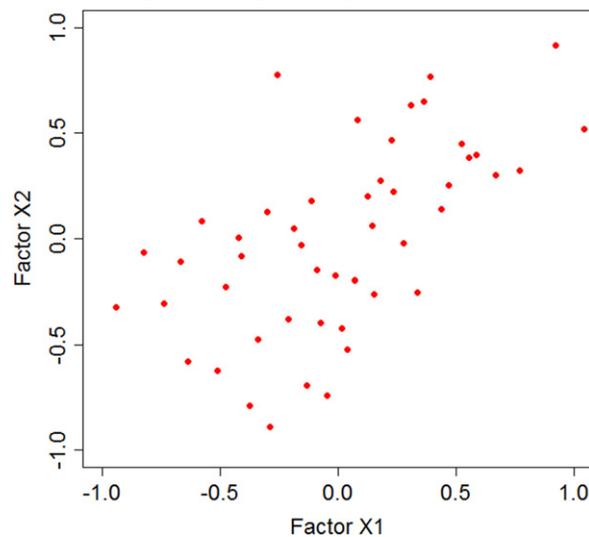


Fig. 7. LH-Design for dependent factors.

hyper-rectangles with equal probability. Eventually, the experiments are distributed over the resulting lattice so that the projection to any one dimension yields exactly one experiment per cell. For cases, where factors  $X_1, \dots, X_{k_r}$  are not independent, Stein (1987) proposes a technique using inverse marginal cumulative distribution functions and permutation matrices to generate appropriate LH-Designs. Fig. 6 illustrates the lattice approach for two independent factors  $X_1 \sim N(0, 1/4)$  and  $X_2 \sim N(0, 1/4)$  and Fig. 7 depicts the LH-Design, generated by approach of Stein (1987), for dependent factors  $(X_1, X_2)$  following a bivariate normal distribution  $N^2(\mu, \Sigma)$ , where

$$\mu = c(0, 0) \quad \text{and} \quad \Sigma = \begin{pmatrix} 1/4 & 1/5 \\ 1/5 & 1/4 \end{pmatrix}. \quad (21)$$

The LH-Design approach is a simple method to involve the feature space distribution and to avoid replicates at the same time. Still, Stein (1987) requires large sample sizes in order to represent the multidimensional distribution in a satisfactory manner. The question remains whether the Latin Hypercube experiments are descriptive enough or not to represent the assumed distribution, when the number of possible experiments is limited. In the next section we may answer this question employing multidimensional quantiles.

### 3. Multidimensional quantiles and depth functions

In one dimension the adaptation towards a dataset can usually be illustrated by a Quantile–Quantile (Q–Q) plot, which scatters the empirical sample quantiles against the theoretical quantiles of the distribution. However, the concept of

one-dimensional quantiles with their natural linear order is not directly transferable to multidimensional distributions. Serfling (2010) overcomes this issue by using median-oriented quantiles implying nested contours that enclose regions containing amounts of inner probability around the  $k_r$ -dimensional median of a distribution  $F^{k_r}$ . Median-oriented quantiles are defined for  $\mathcal{F}$  that is the class of distributions on the Borel sets of  $\mathbb{R}^{k_r}$ .

**Definition 3.1** (*Median-Oriented Quantile Function*). For  $\mathbf{u} \in \mathbb{B}^{k_r-1}(\mathbf{0}) = \{\mathbf{u} \in \mathbb{R}^{k_r} : \|\mathbf{u}\| \leq 1\}$  the median-oriented quantile function  $Q(\cdot, \cdot) : \mathbb{B}^{k_r-1}(\mathbf{0}) \times \mathcal{F} \rightarrow \mathbb{R}^{k_r}$  generates nested contours  $\{\mathbf{x} = Q(\mathbf{u}, F^{k_r}) : \|\mathbf{u}\| = c, 0 \leq c \leq 1\}$  in  $\mathbb{R}^{k_r}$  that enclose the  $k_r$ -dimensional median, defined by  $\mathbf{M}_{F^{k_r}} = Q(\mathbf{0}, F^{k_r})$ .

The following example shows how to determine the median-oriented quantiles for a bivariate normal distribution with independent components (trivial covariance matrix  $\Sigma$ ).

**Example 3.1** (*Median-Oriented Quantiles of a Bivariate Normal Distribution*). Let us consider the bivariate normal distribution  $N^2(\boldsymbol{\mu}, \Sigma)$ , well defined by the parameters

$$\boldsymbol{\mu} = (0, 0) \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 = \sigma^2 & 0 \\ 0 & \sigma_2^2 = \sigma^2 \end{pmatrix} \quad \text{for } 0 < \sigma < \infty. \tag{22}$$

Then, the circles  $A_l$ , defined by radii  $r_l$  and by the center point  $\boldsymbol{\mu} \in \mathbb{R}^{k_r}$ , are contours possessing inner probabilities  $0 < p_l < 1$ . These are given by

$$\mathbb{P}(\mathbf{X} \in A_l) = p_l \tag{23}$$

for  $l \in \mathbb{R}$ , and we claim  $p_l < p_{l'}$ , if  $l < l'$  (or  $A_l \subset A_{l'}$ ). For  $l \in \mathbb{R}$  and  $\mathbf{x} \in \bar{A}_l$  the transformation

$$R(\mathbf{x}, N^2(\boldsymbol{\mu}, \Sigma)) = p_l \cdot \frac{\mathbf{x}}{\|\mathbf{x}\|} = \mathbf{u} \tag{24}$$

yields the mapping  $R : \mathbb{R}^{k_r} \times \mathcal{F} \rightarrow \mathbb{B}^{k_r-1} \times \mathcal{F}$ , which satisfies  $\|\mathbf{u}\| = p_l < 1$ . The probability  $p_l$  can be determined for a circle  $A_l$  with radius  $r_l$  as follows.

$$p_l = \int_{A_l} f_{N^2(\mathbf{0}, \Sigma)}^2(x_1, x_2) d(x_1, x_2) = \frac{1}{2\pi\sigma^2} \cdot \int_{A_l} \exp\left(-\frac{1}{2\sigma^2} \cdot (x_1^2 + x_2^2)\right) d(x_1, x_2).$$

Substitution of  $x_1 = r \cdot \cos \phi$  and  $x_2 = r \cdot \sin \phi$  by polar coordinates  $(r, \phi)$  leads to

$$p_l = \frac{1}{2\pi\sigma^2} \cdot \int_{\phi=0}^{2\pi} \int_0^{r_l} r \cdot \exp\left(-\frac{1}{2\sigma^2} \cdot (r^2 \cdot \cos^2 \phi + r^2 \cdot \sin^2 \phi)\right) dr d\phi.$$

Noting that  $\cos^2 \phi + \sin^2 \phi = 1$ , we receive the *Circular Gaussian* probabilities

$$p_l = 1 - \exp\left(-\frac{r_l^2}{2\sigma^2}\right).$$

The inverse of transformation (24) yields the median-oriented quantile function, with which the median-oriented quantiles are obtained for all inner probabilities  $0 < p_l < 1$ . They are given by

$$\left\{ \mathbf{x} = \frac{\mathbf{u} \cdot \overbrace{\|\mathbf{x}\|}^{=r_l}}{p_l} : \sqrt{u_1^2 + u_2^2} = p_l \right\} \tag{25}$$

for  $\mathbf{u} \in \mathbb{B}^{k_r-1}(\mathbf{0})$  and  $p_l \neq 0$ . The  $k_r$ -dimensional median  $\mathbf{M}_{F^{k_r}} = Q(\mathbf{0}, F^{k_r})$  corresponds to  $\boldsymbol{\mu}$ .

The solution approach shown above on how to determine the median-oriented quantiles of a distribution  $F^{k_r}$  is a special application of the *D-O-Q-R paradigm* (Depth–Outlyingness–Quantile–Rank paradigm), described in detail by Serfling (2010). In a nutshell, it says that three functions, the *Depth Function*, the *Outlyingness Function* and the *Centered Rank Function*, closely relate to the median-oriented quantile function. Serfling (2010) uses this equivalence in order to show that it suffices to find a depth function  $D(\mathbf{x}, F^{k_r})$ , which possesses nested contours enclosing a point  $\mathbf{M}_{F^{k_r}} \in \mathbb{R}^{k_r}$  and bounding regions

$$A_l = \{\mathbf{x} \in \mathbb{R}^{k_r} : D(\mathbf{x}, F^{k_r}) \geq l\} \quad (l \in \mathbb{R}) \tag{26}$$

to induce a valid quantile function  $Q(\mathbf{u}, F^{k_r})$  for a continuous distribution  $F^{k_r}$ . As already indicated in Example 3.1, where we implicitly assumed the nested contours  $A_l$  in (23), the median-oriented quantile function  $Q(\mathbf{u}, F^{k_r})$ ,  $\mathbf{u} \in \mathbb{B}^{k_r-1}$ , has an inverse at each  $\mathbf{x} \in \mathbb{R}^{k_r}$ , given by the centered rank function  $R(\mathbf{x}, F^{k_r})$ . The magnitude  $O(\mathbf{x}, F^{k_r}) = |R(\mathbf{x}, F^{k_r})|$  of the

centered rank function is called *Outlyingness Function*. The outlyingness function, in turn, is related with the depth function by  $D(\mathbf{x}, F^{kr}) = 1 - O(\mathbf{x}, F^{kr})$ .

After Serfling (2010) a valid quantile function is induced by setting  $\mathbf{M}_{F^{kr}} = Q(\mathbf{0}, F^{kr})$  and  $\mathbf{u} = R(\mathbf{x}, F^{kr}) = p \cdot \mathbf{v} \in \mathbb{B}^{kr-1}$ , where  $\mathbf{v}$  the unit vector that points towards  $\mathbf{x}$  from  $\mathbf{M}_{F^{kr}}$ , and  $p_l$  is the probability of the region  $A_l$  with  $\mathbf{x}$  on its boundary (in Example 3.1: transformation (24)). Hence,

$$\mathbb{P}(\mathbf{X} \in A_l) = p_l \quad \text{for } \mathbf{X} \sim F^{kr}. \tag{27}$$

Furthermore, the probability  $p_l, l \in \mathbb{R}$ , equals the outlyingness of all  $\mathbf{x}$  on the closure  $\bar{A}_l$  of  $A_l$ , i.e.  $O(\mathbf{x}, F^{kr}) = \|\mathbf{u}\| = p_l$ . Consequently,  $\bar{A}_l$  corresponds to the median-oriented quantile possessing an inner probability  $p_l$ . Zuo and Serfling (2000) define the depth function in a more general way.

**Definition 3.2 (Depth Function).** Let the mapping

$$D(\cdot, \cdot) : \mathbb{R}^{kr} \times \mathcal{F} \rightarrow \mathbb{R}_+ \tag{28}$$

be bounded and non-negative satisfying the subsequent conditions.

- (i)  $D(\mathbf{A}\mathbf{x} + \mathbf{b}, F_{\mathbf{A}\mathbf{x}+\mathbf{b}}^{kr}) = D(\mathbf{x}, F_{\mathbf{x}}^{kr})$  holds for any random vector  $\mathbf{X} \sim F^{kr}$ , and nonsingular matrix  $\mathbf{A} \in \mathbb{R}^{kr \times kr}$ , and any vector  $\mathbf{b} \in \mathbb{R}^{kr}$  (*affine invariance property*);
- (ii)  $D(\boldsymbol{\theta}, F^{kr}) = \sup_{\mathbf{x} \in \mathbb{R}^{kr}} D(\mathbf{x}, F^{kr})$  holds for any  $F^{kr} \in \mathcal{F}$  having a “center”  $\boldsymbol{\theta}$  in  $\mathbb{R}^{kr}$  (*maximality at center*);
- (iii) For any  $F^{kr} \in \mathcal{F}$  having a “center”  $\boldsymbol{\theta} \in \mathbb{R}^{kr}$ ,  $D(\mathbf{x}, F^{kr}) \leq D(\boldsymbol{\theta} + \delta(\mathbf{x} - \boldsymbol{\theta}), F^{kr})$  holds for all  $\delta \in [0, 1]$  (*monotonicity relative to the deepest point*);
- (iv)  $D(\mathbf{x}, F^{kr}) \rightarrow 0$  as  $\|\mathbf{x}\| \rightarrow \infty$  for arbitrary  $F^{kr} \in \mathcal{F}$  (*vanishing at infinity*).

Then  $D(\cdot, F^{kr})$  is called statistical depth function.

A random vector  $\mathbf{x} \in \mathbb{R}^{kr}$  is usually denoted as centrally symmetric about a “center”  $\boldsymbol{\theta}$ , if  $\mathbf{x} - \boldsymbol{\theta} \stackrel{d}{=} \boldsymbol{\theta} - \mathbf{x}$ , whereas “ $\stackrel{d}{=}$ ” means equal in distribution. Zuo and Serfling (2000) introduce a more general notion, which defines  $\mathbf{X}$  to be halfspace symmetric about  $\boldsymbol{\theta}$ , if  $\mathbb{P}(\mathbf{X} \in H) \geq 1/2$  for every closed halfspace  $H \subset \mathbb{R}^{kr}$  that contains  $\boldsymbol{\theta}$ .

With condition (ii) of Definition 3.2, the D–O–Q–R paradigm implies that the center  $\boldsymbol{\theta}$  of  $F^{kr}$  equals the multidimensional median, i.e.  $\boldsymbol{\theta} \equiv \mathbf{M}_{F^{kr}}$ . Thus, it follows that the  $k_r$ -dimensional median possesses maximal depth.

With moderate effort it is possible to determine the probabilities  $p_l$  in (27) of the  $k_r$ -dimensional normal distribution  $N^{kr}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  with parameters

$$\boldsymbol{\mu} = (\mu_1, \dots, \mu_{k_r}) \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{k_r}^2 \end{pmatrix} \quad \text{for } |\boldsymbol{\Sigma}| < \infty. \tag{29}$$

As highlighted by Waugh (1961), the determination procedure requires the integration of the p.d.f.  $f^{kr}$  over a centered  $k_r$ -dimensional ellipse, whose  $k_r$  semiaxes correspond to the diagonal elements of  $\boldsymbol{\Sigma}$ . The analytical resolvability of the median-oriented quantile task, as presented in Example 3.1, becomes challenging for  $F^{kr} = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  ( $\boldsymbol{\Sigma}$  not trivial), because the orientation of the  $k_r$ -dimensional ellipse changes. Moreover, unless a multivariate normal distribution is considered, the analytical resolvability of probability  $p_l$  of region  $A_l$  is no longer ensured.

We want to use the D–O–Q–R paradigm and the Depth Function approach to estimate median-oriented quantiles of a general distribution  $F^{kr}$  by realizations  $\mathbf{x}_1, \dots, \mathbf{x}_{N'}$  of  $\mathbf{X}_n \stackrel{iid}{\sim} F^{kr}$ . Then the depth function  $D(\mathbf{x}_n, F^{kr})$  yields a center-outward ordering of  $\mathbf{x}_1, \dots, \mathbf{x}_{N'}$ . For the region  $A_l$ , as defined in (26), the probability  $p_l$  is estimated as follows.

$$\mathbb{P}(\mathbf{X} \in A_l) = p_l \approx \frac{|A_l|}{N'} \quad \text{for } l \in \mathbb{R} \text{ and } \mathbf{X} \sim F^{kr}. \tag{30}$$

The median oriented quantiles, which correspond to probabilities  $p_l$ , are estimated by the convex hulls of  $A_l$ . In the papers of Liu et al. (1999) and Zuo and Serfling (2000) various depth functions are presented, compared and discussed in terms of the four desired properties *affine invariance*, *maximality at center*, *monotonicity relative to the deepest point* and *vanishing at infinity*. Zuo and Serfling (2000) conclude by recommending the halfspace- and the projection depth, because they fulfill the four desired properties of Definition 3.2. Although the D–O–Q–R paradigm can be applied for every valid depth function, from our point of view, an appropriate choice would be the affine invariant version of the  $L_2$  depth. This recommendation is threefold:

1. The affine invariant version of the  $L_2$  depth fulfills all properties of Definition 3.2.
2. The affine invariant version of the  $L_2$  depth can be approximated with controllable effort, as shown in (33).
3. The affine invariant version of the  $L_2$  depth can be determined for  $F^{kr} \equiv N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

**Definition 3.3** ( $L_2$  Depth). When considering a distribution  $F^{k_r}$ , the  $L_2$  depth of every  $\mathbf{x} \in \mathbb{R}^{k_r}$  is given by

$$L_2 D(\mathbf{x}, F^{k_r}) = (1 + \mathbb{E}_{F^{k_r}} \|\mathbf{x} - \mathbf{X}\|_2)^{-1}, \quad (31)$$

where  $\mathbf{X} \sim F^{k_r}$ .

It can be easily shown that the nested contours possessed by (31), bound convex regions  $A_l$  (Mosler, 2013). In addition to this desirable characteristic, Zuo and Serfling (2000) show that the  $L_2$  depth provides all desired properties with the exception of affine invariance. Nonetheless, by replacing the euclidean norm with  $\|\mathbf{x}\|_M \equiv \sqrt{\mathbf{x}^T \mathbf{M} \mathbf{x}}$ , where  $\mathbf{M} \in \mathbb{R}^{k_r \times k_r}$  is a positive definite matrix, an affine invariant version of Definition 3.3 can be obtained (Rao, 1988). That is

**Definition 3.4** (Affine Invariant Version of the  $L_2$  Depth).

$$\tilde{L}_2 D(\mathbf{x}, F^{k_r}) = (1 + \mathbb{E}_{F^{k_r}} \|\mathbf{x} - \mathbf{X}\|_{\Sigma^{-1}})^{-1}, \quad (32)$$

where  $\mathbf{X} \sim F^{k_r}$  and  $\Sigma^{-1}$  the covariance matrix of distribution  $F^{k_r}$ .

The expected norm in Definition 3.4 can be estimated by a moderate sized subset  $\{\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_{N_{L_2}}\}$  of  $\{\mathbf{x}_1, \dots, \mathbf{x}_{N'}\}$  in  $N' \cdot N_{L_2} < N'^2$  operations as follows.

$$\tilde{L}_2 D(\mathbf{x}_n, F^{k_r}) \approx \left( 1 + \frac{1}{N_{L_2}} \cdot \sum_{n=1}^{N_{L_2}} \sqrt{(\mathbf{x} - \tilde{\mathbf{x}}_n)^T \Sigma^{-1} (\mathbf{x} - \tilde{\mathbf{x}}_n)} \right)^{-1} \quad \text{for } n = 1, \dots, N'. \quad (33)$$

If  $F^{k_r} \equiv N^{k_r}(\boldsymbol{\mu}, \Sigma)$ , the expected value in (32) can be determined. For  $\mathbf{x}_n \in \mathbb{R}^{k_r}$ , the exact solution requires the consideration of the positive random variable

$$W_{\mathbf{x}_n} = (\mathbf{x}_n - \mathbf{X})^T \Sigma^{-1} (\mathbf{x}_n - \mathbf{X}) \in \mathbb{R}_+, \quad \text{where } \mathbf{X} \sim N^{k_r}(\boldsymbol{\mu}, \Sigma). \quad (34)$$

Given that the distribution  $F_{W_{\mathbf{x}_n}}$  of  $W_{\mathbf{x}_n}$  is known, we can apply the well known formula

$$\mathbb{E}(W_{\mathbf{x}_n}^\gamma) = \gamma \cdot \int_0^\infty t^{\gamma-1} \cdot \mathbb{P}(W_{\mathbf{x}_n} > t) dt \quad (35)$$

for  $\gamma = 1/2$ . Now it holds that distribution  $F_{W_{\mathbf{x}_n}}$  corresponds to a noncentral chi-squared distribution with  $k_r$  degrees of freedom and non-centrality parameter  $\lambda^2$  (Anderson, 1984).

$$W_{\mathbf{x}_n} \sim \chi_{k_r}^2(\lambda^2) \quad \text{where } \lambda^2 = (\mathbf{x}_n - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_n - \boldsymbol{\mu}). \quad (36)$$

Still, it has to be pointed that the resolvability of the affine invariant version of the  $L_2$  depth must not be confused with the resolvability of the median-oriented quantile task. The product of  $N'$  and the computational effort, necessary to evaluate the integral in (35), dominates the operations required to estimate the median-oriented quantiles of  $F^{k_r} \equiv N^{k_r}(\boldsymbol{\mu}, \Sigma)$ . For  $F^{k_r} \neq N^{k_r}(\boldsymbol{\mu}, \Sigma)$  the computation speed mainly depends on  $N'$  and the approximation quality  $N_{L_2}$  in (33).

Fig. 8 illustrates the median-oriented quantiles of a bivariate normal distribution  $N^2(\boldsymbol{\mu}, \Sigma)$  with parameters

$$\boldsymbol{\mu} = (1/2, 1/2) \quad \text{and} \quad \Sigma = \begin{pmatrix} 1/8 & 1/10 \\ 1/10 & 1/8 \end{pmatrix} \quad (37)$$

that have been estimated by  $N' = 10.000$  pseudo MC samples for  $p_l = 0.1, \dots, 0.9$ . Representatively for a distribution  $F^{k_r} \neq N(\boldsymbol{\mu}, \Sigma)$ , Fig. 9 shows the non-parametric estimation approach for the versatile Beta distribution. The estimation was based on  $N' = 10.000$  pseudo MC samples  $(x_1, x_2) \in \mathbb{R}^2$ , which were realizations of

$$\frac{X_1}{X_1 + X_2 + X_3} \quad \text{and} \quad \frac{X_2}{X_1 + X_2 + X_3}, \quad (38)$$

where

$$X_1 \sim \gamma(2, 5) \quad \text{and} \quad X_2 \sim \gamma(1/2, 5) \quad \text{and} \quad X_3 \sim \gamma(2, 5). \quad (39)$$

$\gamma(\kappa, \theta)$  denotes a gamma distribution with shape parameter  $\kappa > 0$  and scale parameter  $\theta > 0$ . The computation of the convex hulls was performed with the algorithm of Barber et al. (1996).

In the next step, for us the question is: Which sample size  $N'$  is necessary to estimate the median-oriented quantiles of a distribution  $F^{k_r}$  in an adequate manner? By means of the law of large numbers, one can count on stochastic sampling strategies like the LH-Design (or pseudo MC sampling) to depict the true distribution provided that enough samples are

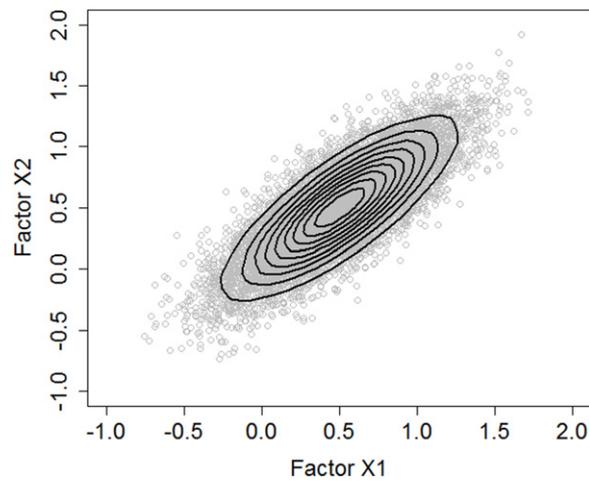


Fig. 8. Median-oriented quantiles of  $N^2(\mu, \Sigma)$ .

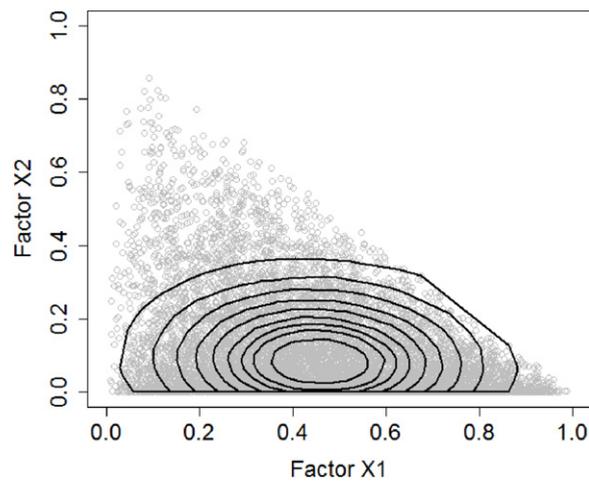


Fig. 9. Median-oriented quantiles of bivariate beta distribution.

generated. The estimated median-oriented quantiles can be easily validated with a test set  $\{\mathbf{x}_1, \dots, \mathbf{x}_{N_{\text{test}}}\}$  ( $N_{\text{test}} > N'$ ), because the estimated probabilities  $p_l$  of all regions  $A_l$  can be compared to the fractions  $p_l^{\text{test}}$  of the test set  $A_l^{\text{test}}$ , given by

$$A_l^{\text{test}} = \{\mathbf{x} \in \{\mathbf{x}_1, \dots, \mathbf{x}_{N_{\text{test}}}\} : D(\mathbf{x}, F^{k_r}) \geq l\}, \quad \text{and} \quad p_l^{\text{test}} = \frac{|A_l^{\text{test}}|}{N_{\text{test}}}. \quad (40)$$

We assess the performance of the estimated median-oriented quantiles by the root mean squared error

$$RMSE_p = \sqrt{\frac{1}{N'} \cdot \sum_l (p_l - p_l^{\text{test}})^2}. \quad (41)$$

$RMSE_p$  can also be seen as a measure of representativeness of the realizations  $\mathbf{x}_1, \dots, \mathbf{x}_{N'}$  for the distribution  $F^{k_r}$ . We apply the Latin-Hypercube sampling approach of Stein (1987) to generate the concerning realizations so that we are able to directly answer the outstanding question, stated at the end of Section 2.4. That is, is the LH-Design descriptive enough or not to represent a distribution  $F^{k_r}$ , when the number of experiments is limited?

**Example 3.2** (Validation of Estimated Median-Oriented Quantiles: LH-Design). We want to illustrate the validation procedure for the example of the multivariate normal distribution  $F^{k_r} \equiv N^{k_r}$ . This is because random distributions are easily generated and the affine invariant version of the  $L_2$  depth can be analytically determined. Several experiments with different normal distributions  $N^{k_r}(\mu, \Sigma)$  showed that  $N' = 150.000$  is a good choice for  $k_r \leq 100$  dimensions. As an example, Fig. 10 illustrates the validation of median-oriented quantiles of  $N^{100}(\mu, \Sigma)$  estimated by  $N' = 150.000$  experiments and validated by  $N_{\text{test}} = 250.000$  experiments.

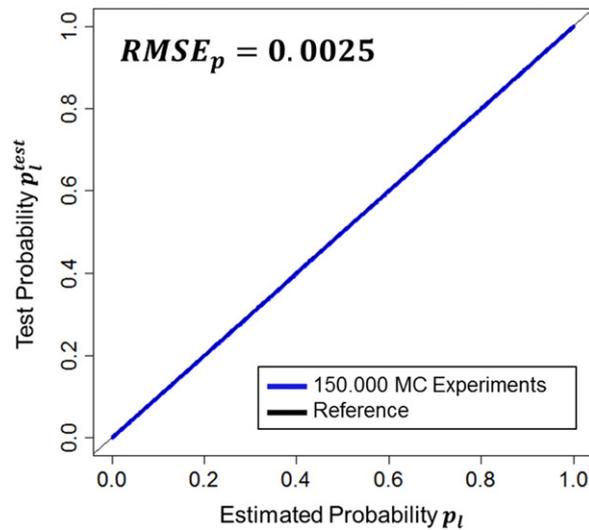


Fig. 10. Validation of estimated median-oriented quantiles of  $N^{100}(\mu, \Sigma)$ .

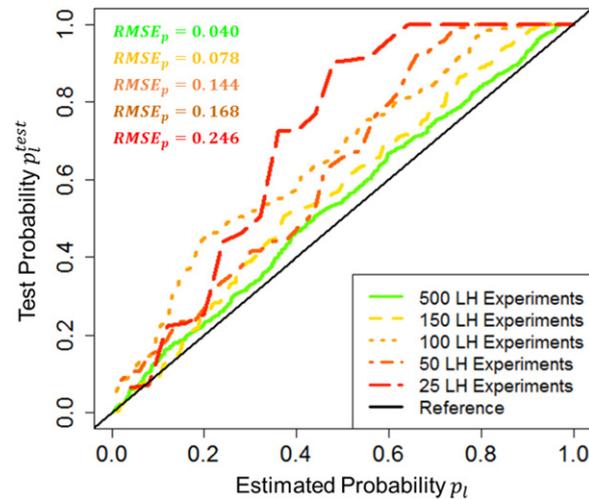


Fig. 11. Validation of estimated median-oriented quantiles of  $N^{10}(\mu, \Sigma)$ .

Hence, a LH-Design with  $N' = 150.000$  experiments represents  $N^{100}(\mu, \Sigma)$  in an adequate manner. Even if  $N' = 150.000$  are far beyond the considered boundary of  $N_{Sf} = N - N_{Sc}$  experiments,  $\mathbf{x}_1, \dots, \mathbf{x}_{150.000}$  can be definitely used to validate median-oriented quantiles, estimated by less experiments. On this account, we set

$$N' \in \left\{ \underbrace{25, 50, 75, 100, 150, 200, 300, 400, 500}_{\text{possible values for } N_{Sf}}, \underbrace{1.000, 10.000, 50.000, 100.000}_{\text{for illustration purpose}} \right\}$$

and  $N_{\text{test}} = 150.000$ . As outlined in Section 2, the dimension of the initial experimental space  $\mathbb{X}^k$  will practically reduce to  $k_r \leq 10$  after the screening procedure. Fig. 11 highlights the performance of the median-oriented quantiles, when estimated on the basis of a LH-Design consisting of  $N' \in \{25, 50, 100, 150, 500\}$  experiments. It is clearly visible that the larger the dimension the more fails the LH-Design to represent the assumed feature space distribution for a given  $N'$ . This is due to the construction method after Stein (1987), which “only” considers marginal inverse distribution functions. Fig. 12 illustrates that for the multivariate normal distribution, considered with different dimensions  $k_r \in \{2, 5, 10, 20, 50\}$ . For each dimension, the outlined  $RMSE_p$  are based on an average of 20 different distributions. The LH-Design fails to accurately estimate the median-oriented quantiles, when  $k_r > 10$  and the  $N_{Sf} \leq 500$ .

Although the LH-Design of Stein (1987) is a very efficient design and includes the feature space distribution, it lags in transmitting a high dimensional distribution to the generated experiments, when less than 500 experiments. As a

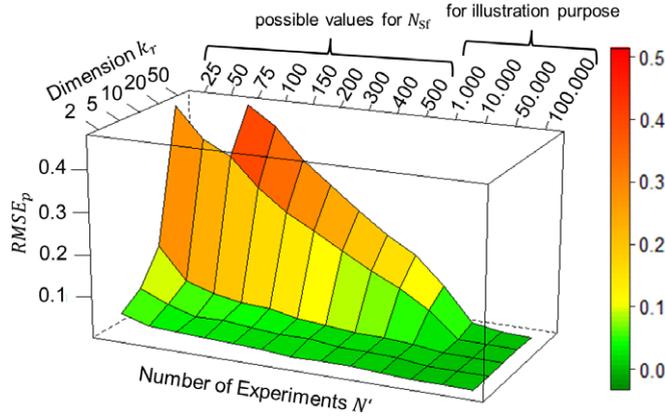


Fig. 12. LH-design: validation of estimated median-oriented quantiles of  $N^{k_r}(\mu, \Sigma)$ .

consequence, we introduce a new space filling design approach, which expands the LH-Design approach of Stein (1987) by considering median-oriented quantiles and depth functions by Serfling (2010).

#### 4. The depth-design

In Section 3 we showed that for the LH-Design of Stein (1987) a high feature space dimension combined with a limited number of experiments can significantly decrease the estimation performance of the median-oriented quantiles. In such cases ( $N_{sf} \leq 500$  and  $k_r \approx 10$ ) we infer that the experiments generated by the LH-Design do rather not provide the information necessary to represent the associated distribution  $F^{k_r}$  in an accurate manner. This disadvantage of the LH-Design can be eliminated by our new Depth-Design, which expands the LH-Design of Stein (1987) by the median-oriented quantile approach of Serfling (2010).

In the first step of constructing the Depth-Design, realizations  $\mathbf{x}_1, \dots, \mathbf{x}_{N'}$  of  $\mathbf{X} \sim F^{k_r}$  need to be generated by a stochastic sampling procedure (i.e. pseudo MC sampling, LH-Design, etc.) in order to estimate the median-oriented quantiles of the underlying distribution. In doing so, we propose to choose  $N'$  in correspondence with Fig. 12 so that  $RMSE_p$  is kept at a minimum level. This means, if  $k_r = 50$ , we suggest to set the number of experiments at least to  $N' = 50.000$ . If  $N_{sf}$  experiments are feasible we are particularly interested in the median-oriented quantiles  $\bar{A}_l$  that correspond to the probabilities

$$p_l = \frac{\bar{A}_l}{N'} \in \left\{ \frac{0}{N_{sf} - 1}, \frac{1}{N_{sf} - 1}, \dots, \frac{N_{sf} - 1}{N_{sf} - 1} \right\}. \tag{42}$$

The regions  $A_l$  are received by sorting  $\mathbf{x}_1, \dots, \mathbf{x}_{N'}$  in an ascending order by their depth value. Then, the estimated median-oriented quantiles  $\bar{A}_l$  can be approximated by

$$\bar{A}_l \approx \{ \mathbf{x} \in \{ \mathbf{x}_1, \dots, \mathbf{x}_{N'} \} : D(\mathbf{x}, F^{k_r}) - \epsilon \leq l \leq D(\mathbf{x}, F^{k_r}) + \epsilon \} \quad (\epsilon > 0). \tag{43}$$

Now the idea is to adjust a LH-Design with  $N_{sf}$  experiments by the estimated median-oriented quantiles, obtained by condition (42) for  $\mathbf{x}_1, \dots, \mathbf{x}_{N'}$ . By construction each LH experiment  $\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_{N_{sf}}$  is linked to a specific  $\bar{A}_l$ . Starting with the smallest probability  $p_l$  and successively proceeding to the largest probability, the experiments of the Depth-Design are generated as follows.

$$DD(\mathbf{x}_n) = \operatorname{argmin}_{\mathbf{x} \in \bar{A}_l} \|\mathbf{x} - \tilde{\mathbf{x}}_n\|_2 \quad \text{for } n = 1, \dots, N_{sf}. \tag{44}$$

The computational effort of the Depth-Design is dominated by the estimation procedure of the median-oriented quantiles, as described in Section 3.

**Example 4.1** (Generation of the Depth-Design for Dimension  $k_r = 2$ ). As an example, we make once more use of the affine invariant version of the  $L_2$  depth to estimate the median-oriented quantiles. The reader may reconsider the bivariate normal distribution in (37) and respectively the bivariate beta distribution in (38), where the  $L_2$  depths were estimated from  $N' = 10.000$  pseudo MC samples. For visualization purposes we set  $N_{sf} = 10$  and generate the according LH-Designs. Figs. 13 and 14 show the Depth-Design approach by arrows, where LH experiments are adjusted by (44). When Fig. 11 is compared with Fig. 15, the improvement achieved by the Depth-Design becomes instantly visible. According to the  $RMSE_p$  obtained, already 25 Depth Design experiments contain more information about the considered distribution ( $k_r = 10$ ) as 500 Latin Hypercube experiments. Again, the median-oriented quantiles have been validated with a test set of size  $N_{test} = 150.000$ .

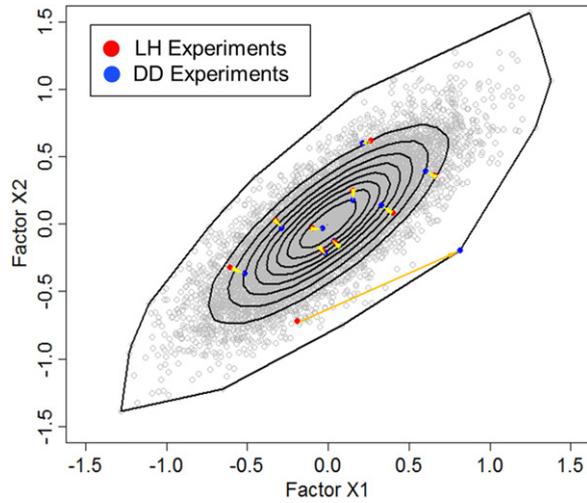


Fig. 13. Depth-Design for bivariate normal distribution  $N^2(\mu, \Sigma)$ .

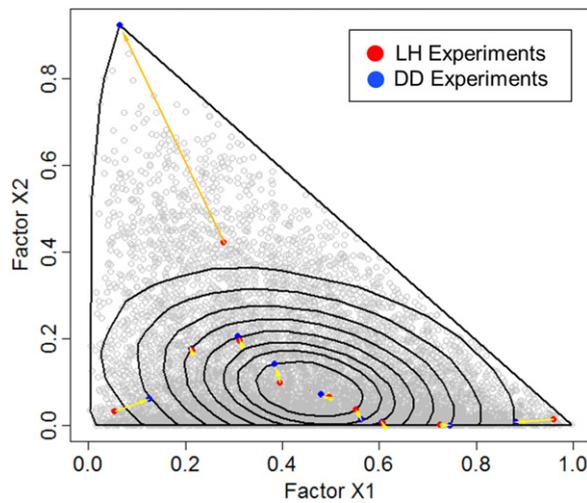


Fig. 14. Depth-Design for bivariate beta distribution.

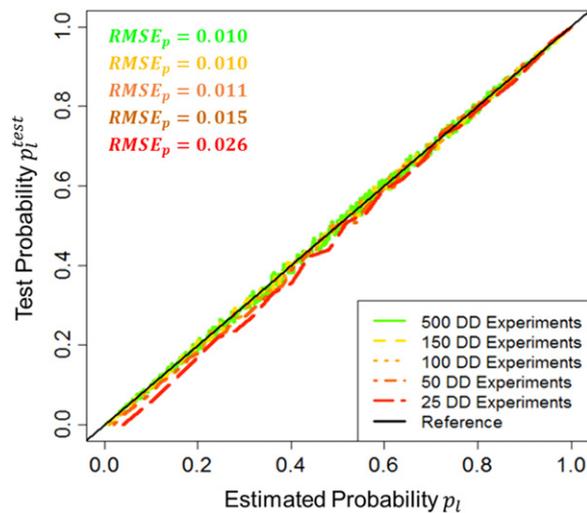


Fig. 15. Depth-Design for bivariate normal distribution  $N^2(\mu, \Sigma)$ .

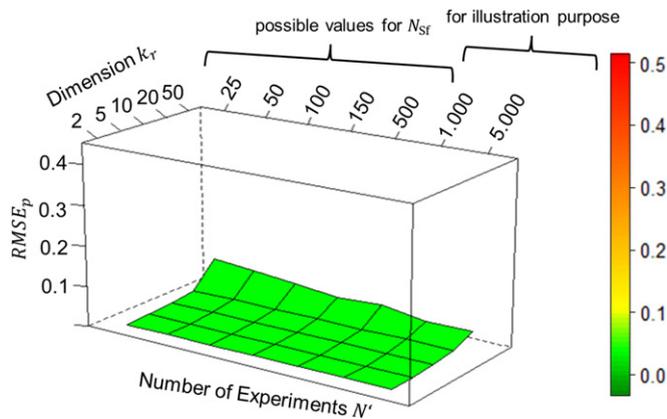


Fig. 16. LH-Design: validation of estimated median-oriented quantiles of  $N^{k_r}$  ( $\mu$ ,  $\Sigma$ ).

In line with Fig. 12, we want to analyze the performance of the Depth-Design, when the dimension  $k_r$  is increased and the number of experiments is varied. Fig. 16 clearly shows the power of the Depth-Design in terms of representing a multidimensional distribution with a limited number of experiments for the multivariate normal distribution. Individual tests showed that the performance for non-normal distribution is comparable to the results obtained in Fig. 16. Additionally, due to the fact that the Depth-Design is based on the LH approach, it nearly shares all advantages of the LH-Design, like good projection properties or the efficient distribution of the experiments within the feature space. Even if the effort to generate a Depth-Design is more expensive than as for the LH-Design, the experiments to be conducted often take much longer so that it is worth allocating a bit of time for the DOE.

## 5. Conclusion

In this paper, we provided a guideline on how to generate representative experiments for computer simulation models, when the number of feasible experiments is limited ( $N \leq 500$ ). We assume the simulation models to have a large number ( $k \geq 50$ ) of controllable quantitative inputs, which are subject to a multidimensional distribution  $F^k$ .

As the first step of generating representative experiments we suggest to spend a portion  $N_{Sc}$  of  $N$  for a screening procedure, to identify the most decisive regions of the feature space  $\mathbb{X}^k$  in terms of a response  $Y$ . For the screening procedure we propose the very efficient Definitive Screening Design of Jones and Nachtsheim (2011), which provides information on pure linear and 2nd order effects of  $k$  factors with only  $N_{Sc} = 2 \cdot k + 1$  experiments. Due to the undersized and unbalanced Definitive Screening Design we established a bias corrected pair-by-pair comparison strategy, which finally enabled us to estimate the total sensitivity index of Homma and Saltelli (1996) giving a relative significance measure for each input factor. We share the generally accepted heuristics that most systems are only dominated by a few significant factors so that the feature space  $\mathbb{X}^k$  reduces to  $\mathbb{X}^{k_r}$  (usually  $k_r \leq 10$ ).

In the second step, a space filling design with  $\mathbf{x}_1, \dots, \mathbf{x}_{N_{Sf}}$  experiments ( $N_{Sf} = N - N_{Sc}$ ) shall be applied, capable of representing distribution  $F^{k_r}$  of  $\mathbb{X}^{k_r}$ . As a measure of representativeness (concerning  $F^{k_r}$ ) we introduce the mean squared error between the inner probabilities of the estimated median-oriented quantiles on basis of  $\mathbf{x}_1, \dots, \mathbf{x}_{N_{Sf}}$  and the inner probabilities of estimated median-oriented quantiles on basis of pseudo MC samples  $\mathbf{x}_1, \dots, \mathbf{x}_{N'}$  ( $N' = 150.000$  for  $k_r \leq 50$ ). For the multivariate normal distribution and the affine invariant version of the  $L_2$  depth we showed that the Latin Hypercube Design of Stein (1987) lags in transmitting high dimensional distributions to the generated experiments when less than 500 experiments are considered. Thereupon, we introduced our new space filling design approach, the “Depth-Design”, that expands the Latin Hypercube Design by the concept of multidimensional quantiles and depth functions. The general idea of the Depth-Design is to adjust the Latin Hypercube experiments in a way that the median-oriented quantiles are accurately estimated. We showed that already  $N_{Sf} = 25$  Depth-Design experiments may deliver more information about a feature space distribution than 500 experiments, generated by the LH-Design. Just as well as for the LH-Design, we analyzed its performance for different dimensions  $k_r$  and for various number of experiments  $N_{Sf}$  exemplarily by the multivariate normal distribution and the affine invariant version of the  $L_2$  depth. The corresponding results are obvious as only a few Depth-Design experiments are already carrying almost all information about the median-oriented quantiles of the underlying distribution.

For these reasons, we propose the application of the Depth-Design to produce representative experiments in the space filling procedure. If the screening procedure is omitted, we showed that the Depth-Design remains robust for high dimensional spaces ( $k > 10$ ).

## Acknowledgments

All research necessary to create this paper was supported by Graz University of Technology and AVL List GmbH Graz. The authors want to thank all people who directly and indirectly contributed to our ideas and findings. Particularly, special thanks go to the reviewers, whose comments significantly improved this paper.

## References

- Anderson, T.W., 1984. *An Introduction to Multivariate Statistical Analysis*, second ed. John Wiley & Sons, New York, NY, pp. 72–73.
- Barber, C.B., Dobkin, D.P., Huhdanpaa, H.T., 1996. The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* 22, 469–483.
- Box, G.E.P., Behnken, D.W., 1960. Some new three level designs for the study of quantitative variables. *Technometrics* 2, 455–476.
- Box, G.E.P., Wilson, K.B., 1951. On the experimental attainment of optimum conditions. *J. R. Stat. Soc.* 13, 1–45.
- Cheng, S.-W., Wu, C.F.J., 2001. Factor screening and response surface exploration (with discussion). *Statist. Sin.* 11, 553–604.
- Curran, C., Mitchell, T., Morris, M., Ylvisaker, D., 1991. Bayesian prediction of deterministic functions: With applications to the design and analysis of computer experiments. *J. Amer. Statist. Assoc.* 86, 953–963.
- Dagpunar, J., 1988. *Principles of Random Variate Generation*. Clarendon Oxford Science Publications, Oxford, UK, pp. 157–169.
- Fang, K.T., 1980. The uniform design: Application of number-theoretic methods in experimental design. *Acta Math. Appl. Sinica.* 3, 363–372.
- Fishman, G.S., 1996. *Monte Carlo: Concepts, Algorithms and Applications*, first ed. Springer, New York, NY, pp. 145–254.
- Hinkelmann, K., Kempthorne, O., 1994. *Design and Analysis of Experiments*. John Wiley & Sons, Chichester, pp. 202–207.
- Homma, T., Saltelli, A., 1996. Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Saf.* 52, 1–17.
- Johnson, M.E., 1987. *Multivariate Statistical Simulation*. John Wiley & Sons, New York, Chichester, pp. 43–46. 116–117, 161.
- Johnson, M.E., Moore, L.M., Ylvisaker, D., 1990. Minimax and maximin distance designs. *J. Statist. Plann. Inference* 26, 131–148.
- Jones, B., Nachtsheim, C.J., 2011. A class of three-level designs for definitive screening in the presence of second-order effects. *J. Qual. Technol.* 43, 1–15.
- Liu, R.Y., Parelius, J.M., Jesse, M., Singh, K., 1999. Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Ann. Statist.* 27, 783–858.
- Mckay, D.J., Conover, W.J., Beckman, R.J., 1979. A Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21, 239–245.
- Mee, R.W., 2009. *A Comprehensive Guide To Factorial Two-Level Experimentation*. Springer, Dordrecht, Heidelberg, London, New York, p. 154.
- Montgomery, D.C., 2012. *Design and Analysis of Experiments*, eighth ed. John Wiley & Sons, New York, NY, pp. 11–21. 80–91, 233–444.
- Mosler, K., 2013. Depth statistics. In: *Robustness and Complex Data Structures*. Springer, Berlin–Heidelberg, pp. 17–34.
- Rao, C.R., 1988. Methodology based on the  $L_1$  norm in statistical inference. *Sankhya Ser. A* 50, 289–313.
- Sacks, J., Schiller, S.B., Welch, W., 1989. Designs for computer experiments. *Technometrics* 31, 41–47.
- Serfling, R., 2010. Equivariance and invariance properties of multivariate quantile and related functions and the role of standardization. *J. Nonparametric Statist.* 22, 915–936.
- Shewry, M.C., Wynn, H.P., 1987. Maximum entropy sampling. *J. Appl. Stat.* 14, 165–170.
- Siebertz, K., Van bebber, D., Hochkirchen, T., 2010. *Statistische Versuchsplanung*. Springer, Berlin–Heidelberg, p. 38. 151–158.
- Stein, M., 1987. Large sample properties of simulations using latin hypercube sampling. *Technometrics* 29, 143–151.
- Tsai, P.-W., Gilmour, S.G., Mead, R., 2000. Projective three-level main effects designs robust to model uncertainty. *Biometrika* 87, 467–475.
- Van lint, J.H., Seidel, J.J., 1966. Equilateral point sets in elliptic geometry. *Math. Sci.* 69, 335–348.
- Waugh, J.G., 1961. Evaluation of integral of elliptic Gaussian distribution over a centered ellipse. *Naval Ordnance Test Station China Lake CA TP-2766*, 1–18.
- Wu, J.C.F., Hamada, M.S., 2009. *Experiments: Planning, Analysis, and Optimization*. John Wiley & Sons, New York, NY, pp. 57–424.
- Xiao, L., Lin, D.K., Bai, F., 2012. Constructing definitive screening designs using conference matrices. *J. Qual. Technol.* 44, 1–7.
- Xu, H., 2005. A catalogue of three-level fractional factorial designs. *Metrika* 62, 259–281.
- Zuo, Y., Serfling, R., 2000. General notions of statistical depth function. *Ann. Statist.* 28, 461–482.