Projections of Definitive Screening Designs by Dropping Columns: Selection and Evaluation

Alan R. Vazquez[1, 2], Peter Goos[1, 2], and Eric D. Schoen[1, 3]

[1]KU Leuven, Belgium
[2]University of Antwerp, Belgium
[3]TNO, Zeist, Netherlands

December 21, 2018

Abstract

Definitive screening designs permit the study of many quantitative factors on a response in a few runs more than twice the number of factors. In practical applications, researchers often require a design for m quantitative factors, construct a definitive screening design for more than m factors and drop the superfluous columns. This is done when the number of runs in the standard m-factor definitive screening design is considered too limited or when no standard definitive screening design exists for m factors. In these cases, it is common practice to arbitrarily drop the last columns of the larger design. In this article, we show that certain statistical properties of the resulting experimental design depend on the exact columns to be dropped and that other properties are insensitive to these columns. We perform a complete search for the best sets of 1-8 columns to drop from standard definitive screening designs with up to 24 factors. We observed the largest differences in statistical properties when dropping four columns from 8- and 10-factor definitive screening designs. In other cases, the differences are small, or even nonexistent. Supplementary materials for the article are available online.

Keywords: Conference Matrix; D-efficiency; Isomorphism; Second-Order Model; Two-Factor Interaction.

1 Introduction

Screening designs permit the experimental study of many factors in a small number of runs. Practitioners studying quantitative factors may not feel comfortable with screening designs that restrict attention to two levels per factor. They could argue that screening also requires checking whether a factor's main effect is linear or not, and identifying active two-factor interactions. To meet these concerns, Jones and Nachtsheim (2011) developed three-level designs using a number of runs that is only one more than twice the number of factors studied. The designs are now called definitive screening designs (DSDs).

The original DSDs presented by Jones and Nachtsheim (2011) were based on a heuristic optimal design algorithm. For an odd number of factors and also for some even numbers of factors, the original DSDs were not orthogonal. Xiao et al. (2012) presented a construction of DSDs using conference matrices. A major advantage of that construction is that it guarantees that the resulting DSDs are orthogonal. A drawback is that, for certain numbers of factors, the number of runs of the resulting DSDs is larger than two times the number of factors plus one. In this article, we refer to an n-factor DSD constructed from an n-dimensional conference matrix as a standard DSD or sDSD.

As an illustration, Table 1 shows how a 10-factor sDSD is constructed from a 10×10 conference matrix $\mathbf{C}$. The first ten runs in the table show the original conference matrix. In general, a conference matrix $\mathbf{C}$ is an n-dimensional square matrix of -1s, 0s and 1s for which $\mathbf{C}^T\mathbf{C} = (n-1)\mathbf{I}_n$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Consequently, the columns of a conference matrix are orthogonal. This implies that a conference matrix is an ideal building block for an orthogonal experimental design. For the design in Table 1, it is easy to verify that $\mathbf{C}^T\mathbf{C} = 9\mathbf{I}_{10}$.

The second set of ten runs of the 10-factor sDSD in Table 1 contains the mirror images or the negatives of the first 10 runs. The sDSD's final run is a center run in which all factors are set at their middle level. Xiao et al. (2012) point out that their construction guarantees that the linear main effects (LEs) are orthogonal to all second-order effects (i.e., the quadratic main effects (QEs) and the two-factor interaction effects (TFIs)), and that the second-order effects are never completely aliased.

Conference matrices do not exist when n is odd, and when n is 22, 34 or 58 (Colbourn

Table 1: Standard definitive screening design (sDSD) with 10 factors, constructed by folding over a 10×10 conference matrix $\mathbf{C}$.

Part	Run	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
$\mathbf{C}$	1	0	1	1	1	1	1	1	1	1	1
	2	1	0	−1	−1	−1	−1	1	1	1	1
	3	1	−1	0	−1	1	1	−1	−1	1	1
	4	1	−1	−1	0	1	1	1	1	−1	−1
	5	1	−1	1	1	0	−1	−1	1	−1	1
	6	1	−1	1	1	−1	0	1	−1	1	−1
	7	1	1	−1	1	−1	1	0	−1	−1	1
	8	1	1	−1	1	1	−1	−1	0	1	−1
	9	1	1	1	−1	−1	1	−1	1	0	−1
	10	1	1	1	−1	1	−1	1	−1	−1	0
$-\mathbf{C}$	−10	−1	−1	−1	1	−1	1	−1	1	1	0
	−9	−1	−1	−1	1	1	−1	1	−1	0	1
	−8	−1	−1	1	−1	−1	1	1	0	−1	1
	−7	−1	−1	1	−1	1	−1	0	1	1	−1
	−6	−1	1	−1	−1	1	0	−1	1	−1	1
	−5	−1	1	−1	−1	0	1	1	−1	1	−1
	−4	−1	1	1	0	−1	−1	−1	−1	1	1
	−3	−1	1	0	1	−1	−1	1	1	−1	−1
	−2	−1	0	1	1	1	1	−1	−1	−1	−1
	−1	0	−1	−1	−1	−1	−1	−1	−1	−1	−1
$\mathbf{0}$	11	0	0	0	0	0	0	0	0	0	0

and Dinitz, 2006). For this reason, it is impossible to construct sDSDs for which the run size is as small as two times the number of factors plus one when the number of factors is odd, or when it is 22, 34 or 58. To deal with this problem, Xiao et al. (2012) recommend dropping columns from a sDSD with one, two or three columns more than the required number. For a design comparison, Dougherty et al. (2015) followed this recommendation and generated a 9-factor design with 21 runs by dropping one column from the 10-factor sDSD in Table 1. Fidaleo et al. (2016) used the same 9-factor DSD to investigate the electrochemical decolorization of the azo dye RV5, a compound used for textile dyeing. In this article, we refer to a DSD obtained by dropping one or more columns from a sDSD as a projected DSD or pDSD.

Dropping k columns from a sDSD with $n = m + k$ columns can result in an m-factor design with better aliasing properties than an m-factor sDSD so that the pDSD is more likely to identify the active effects. For instance, when comparing different cost-efficient screening designs, Stone et al. (2014) preferred a 6-factor pDSD with 17 runs constructed by dropping two columns from the 8-factor sDSD to a 6-factor 13-run sDSD, due to the substantial aliasing between pairs of TFIs and between a QE and a TFI in the 13-run design. Patil (2017) studied the impact of seven factors on a welding process where some TFIs were expected to be active, and observed that the 7-factor design formed by dropping one column from the 8-factor sDSD exhibited a substantial amount of aliasing among the interactions. To reduce the aliasing, he dropped three columns from the 10-factor sDSD in Table 1, and thus used a 21-run design instead of a 17-run design.

Errore et al. (2017) conducted a simulation study involving sDSDs and pDSDs with 8, 10 and 12 factors. The pDSDs were constructed by dropping two or four columns from 10-, 12-, 14- and 16-factor sDSDs. The simulations showed that an m-factor pDSD is more likely to identify the active effects than an m-factor sDSD. For this reason, Errore et al. (2017) recommend the use of m-factor pDSDs obtained by dropping two columns from n-factor DSDs.

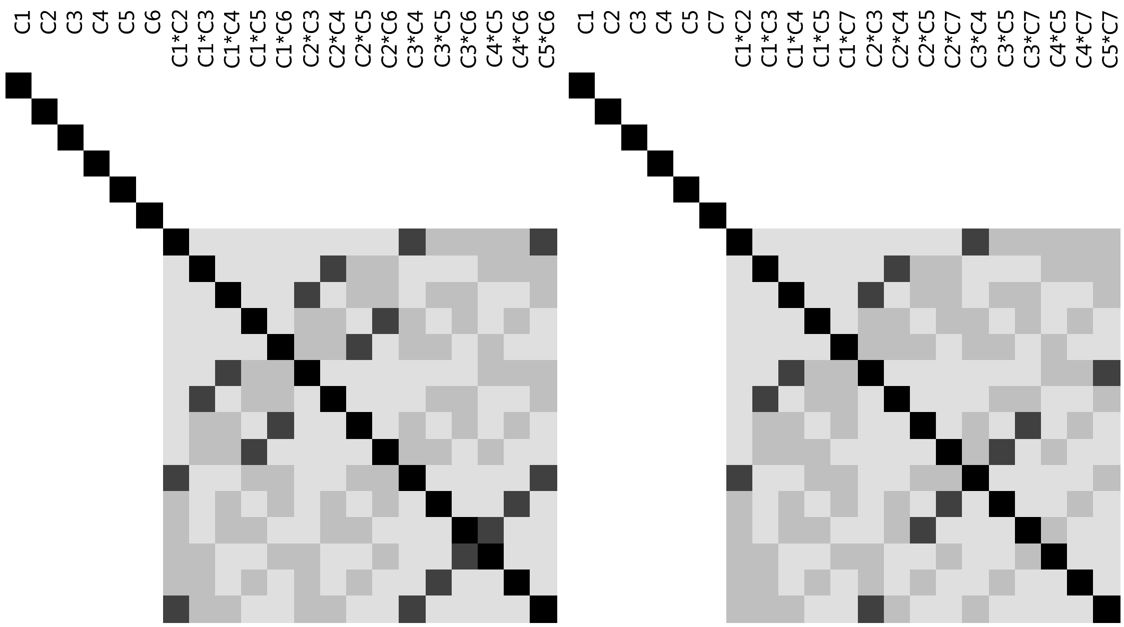
So far, no systematic study has been performed about the best subsets of k columns to drop from a sDSD with $n = m + k$ columns. In each of the applications mentioned above, the authors arbitrarily dropped the last k columns. This is also what commercial software packages do. However, a motivating example detailed in Section 2 shows that there can be marked differences in powers for detecting active effects and type-I error rates between pDSDs obtained by dropping different sets of columns.

This article has three main contributions. Its first contribution is that it identifies the best sets of k columns to drop from sDSDs. This required us to define criteria that distinguish the designs obtained by dropping columns. In Section 3, we show that several criteria from the literature are insensitive to the sets of columns dropped, and that the criteria that do depend on the sets of columns dropped are all based on correlations between TFI contrast vectors. Our detailed study of criteria that do and do not depend on the sets of columns dropped provide additional insights into the properties of DSDs, and forms the second contribution of this article. In Section 4, we report the results of a complete search for the best sets of 1-8 columns to drop from sDSDs for up to 24 factors. In Section 5, we compare m-factor sDSDs to m-factor pDSDs with larger run sizes, obtained by dropping different numbers of columns from sDSDs with more than m factors. In doing so, we highlight the fact that TFI contrast vectors are, in several cases, more severely aliased in pDSDs with larger run sizes than in sDSDs with smaller run sizes. Our detailed study of this counterintuitive phenomenon is the third contribution of this article. Finally, in Section 6, we conclude with a discussion and some suggestions for future research.

2 Motivating example

Our motivating example is inspired by the study of Stone et al. (2014) on alternative screening designs for an artificial 6-factor experiment. Using simulations, these authors compared the performance of a 16-run two-level fractional factorial design, a 12-run Plackett-Burman design, a 16- and a 24-run two-level no-confounding design (Jones and Montgomery, 2010) and a 17-run pDSD to correctly identify two to four active LEs, up to three active TFIs and up to two active QEs. In order to detect the second-order effects with the two-level design options, it was possible to augment these designs with extra runs such as center or axial runs, and to fold over the designs. The goal of the study was to identify the screening design with the lowest expected cost for the complete experiment, measured by the total number of runs used, the number of correctly identified active effects at the screening stage and the efficiency of the estimates of these effects.

Here, we consider a situation where four out of six factors are active. The nonzero effects involving these factors are their four LEs, their six TFIs and one QE. A sensible screening design for this scenario would be a 6-factor 21-run pDSD obtained by dropping four columns



(a) pDSD$_a$ utilizing columns C_1-C_6 (b) pDSD$_b$ utilizing columns C_1-C_5 and C_7

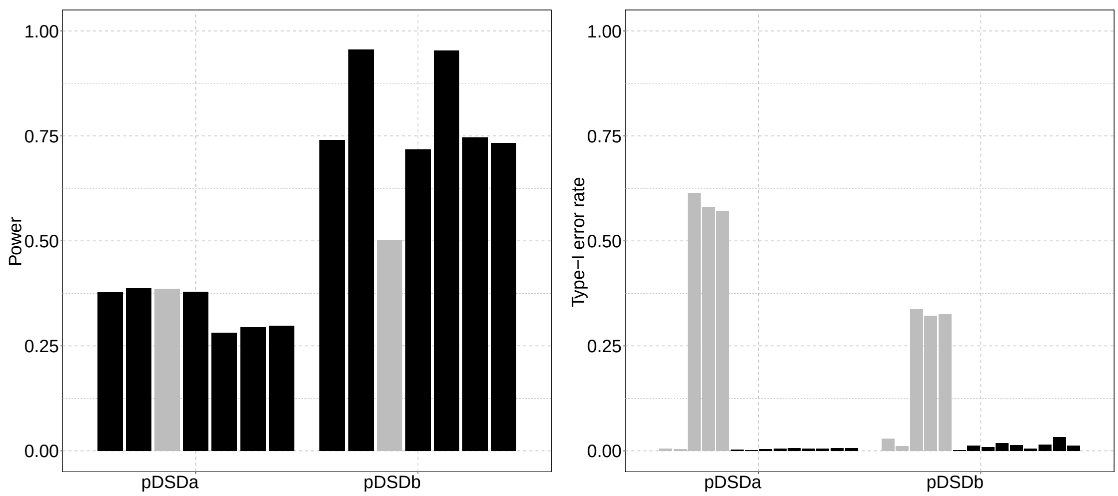
Figure 1: Color maps showing absolute correlations between the LE and TFI contrast vectors for two 6-factor pDSDs obtained from the 10-factor sDSD in Table 1.

from the 10-factor sDSD in Table 1. Let pDSD$_a$ be the design constructed by dropping the last four columns, C_7, C_8, C_9 and C_{10}, of the 10-factor sDSD and let pDSD$_b$ be the design constructed by dropping the columns C_6, C_8, C_9 and C_{10} instead. Figure 1 visualizes the absolute correlations between all pairs of contrast vectors corresponding to the LEs and the TFIs for both design options. In the color maps, the largest absolute correlations for the TFIs equal 0.75. They are visualized by the darkest off-diagonal cells. Figure 1a shows 18 of these dark off-diagonal cells (corresponding to nine pairs of TFIs), while there are only 12 such cells in Figure 1b (corresponding to six pairs of TFIs). The differences between the design options are also reflected in the average absolute correlations and the sums of squared correlations between all pairs of TFI contrast vectors. The average absolute correlation is 0.221 for pDSD$_a$ and 0.207 for pDSD$_b$. The sum of squared correlations equal 8.25 and 6.75 for pDSD$_a$ and pDSD$_b$, respectively. So, the aliasing between the TFIs is more severe when dropping the last four columns of the design in Table 1 than when dropping the columns C_6, C_8, C_9 and C_{10}.

The difference in the average absolute correlations and in the sum of squared correlations between the two pDSD options may have major consequences for any data analysis using the two screening designs. To illustrate this, we conducted a simulation study assuming that the signal-to-noise (SN) ratio for the active LEs, TFIs and QEs equals two, which is the smallest SN ratio considered in the simulation study of Stone et al. (2014). The model used for the simulations is formally described in supplementary Section A. In this section, we consider the scenario in which the active factors were allocated to the last columns of pDSD_a and pDSD_b. We simulated 1,000 responses for each design and determined the set of effects declared active using the two-step approach of Jones and Nachtsheim (2017); see supplementary Section A for details about this method as well as the selection of its tuning parameters. We assessed the performance of the pDSDs in terms of powers and type-I error rates. We calculated the power as the fraction of the simulations for which the active effects were correctly declared active, and the type-I error rate as the fraction of the simulations for which the inactive effects were incorrectly declared active. The supplementary materials of this article include an R implementation of our simulation study.

Figure 2a shows the powers for the seven active second-order effects given our assignment of the active factors to the designs' columns. Both for pDSD_a and pDSD_b, the powers for the active LEs all equal 1 (not shown). The figure shows that the powers for the active second-order effects for pDSD_b are uniformly larger than those for pDSD_a. More specifically, the powers for pDSD_a are in the range 0.28-0.39, while, for pDSD_b, the powers for the active TFIs range from 0.71 to 0.96, and the power for the active QE equals 0.50. Figure 2b shows the type-I error rates for the 14 inactive second-order effects. The type-I error rates for the inactive LEs were very close to zero for both pDSD_a and pDSD_b (not shown). Figure 2b shows that the type-I error rates tend to be smaller for pDSD_b than for pDSD_a, since the maximum type-I error rate for pDSD_a and pDSD_b are 0.62 and 0.34, respectively. In summary, Figure 2 demonstrates that, in the presence of several active TFIs, the reduced aliasing of pDSD_b when compared to pDSD_a results in a more correct detection of the active effects.

We conducted the above simulation study for all 15 possible assignments of the active factors to the designs' columns and provide a comprehensive discussion of the results in supplementary Section A. The results show that pDSD_a has the unfavorable powers and



(a) Power (b) Type-I error rate

Figure 2: Powers for the active second-order effects and type-I error rates for the inactive second-order effects for pDSD_a, constructed by dropping the last columns of the 10-factor 21-run sDSD in Table 1, and pDSD_b, constructed by dropping the columns C_6, C_8, C_9 and C_{10}. Black bars: powers and type-I error rates for TFIs; gray bars: powers and type-I error rates for QEs.

type I error rates from Figure 2 for three of the assignments, and the favorable ones for the 12 other assignments. For pDSD_b, two of the assignments have the unfavorable powers and type I errors, while the other 13 have the favorable ones. Given our results, we conclude that pDSD_b is preferable to pDSD_a.

3 Criteria to evaluate projected definitive screening designs.

To evaluate all possible pDSDs as comprehensively as possible, we considered all the statistical criteria that have been used to evaluate DSDs in the literature, and express these criteria as functions of the numbers of columns dropped from the sDSDs. The criteria either do not depend on the exact columns dropped from the sDSDs or they do depend on these columns. Obviously, the criteria that do not depend on the columns dropped are not helpful to select the best possible pDSDs. The criteria that do depend on the columns dropped allow us to determine which pDSDs ought to be preferred, and play a central role in Section 4. We discuss the two groups of criteria in separate subsections.

3.1 Criteria that do not depend on the columns dropped

Supplementary Section B shows analytical expressions for (i) relative D-efficiencies to estimate the model with all LEs and the model with all LEs and QEs; (ii) relative standard errors for LE and QE estimates in these models; (iii) correlations between pairs of QE contrast vectors; (iv) correlations between the contrast vector of a QE and a TFI; and, (v) correlations between pairs of TFI contrast vectors involving three factors. Each of these properties only depend on the numbers of factors in the pDSDs and the numbers of columns that were dropped from the sDSDs. So, these measures do not depend on the exact set of columns dropped from a sDSD. They all improve with the run size of the pDSD, except for the correlation between pairs of QE contrast vectors which increases to 1/3 for large run sizes. Expressions similar to (i)-(v) were given by Jones and Nachtsheim (2011), Xiao et al. (2012) and Georgiou et al. (2014). Our expressions differ from the ones derived in the earlier articles in that they are written as a function of the number of columns dropped from sDSDs.

In this section, we provide expressions for powers of significance tests in a null model, in a model including all LEs and in a full response surface model in two or three factors. To the best of our knowledge, our analytical expressions for these properties are new to the literature on DSDs.

3.1.1 Null model and model with all linear effects

Departing from a model including only the intercept or from a model containing the intercept and all LEs, we can conduct t tests for individual LEs, QEs and TFIs. Four pertinent tests are shown in Table 2. The first two columns in the table identify the hypothesis to be tested, while the third column shows the terms that appear in the model besides the effect to be tested. The fourth column shows the degrees of freedom ν for the t statistic. The last column is the non-centrality parameter λ of the non-central t-distribution needed to calculate the power of the test. In the table's second column, β_i, β_{ii} and β_{ij} represent the LE of factor i, the QE of factor i and the interaction between the factors i and j (where $i \neq j$), respectively. The derivations for the degrees of freedom and non-centrality parameters are included in supplementary Section B.

Table 2: Degrees of freedom ν and non-centrality parameters λ for various significance tests using a pDSD, assuming a signal-to-noise ratio of 1 for the effect tested. Setting $k = 0$ shows the results for a sDSD.

Label	Hypothesis	Model terms	ν	λ
L_1	$\beta_i = 0$	Intercept only	$2(m+k)-1$	$\sqrt{2(m+k)-2}$
L_{me}	$\beta_i = 0$	Intercept + all LEs	$2k+m$	$\sqrt{2(m+k)-2}$
Q_{me}	$\beta_{ii} = 0$	Intercept + all LEs	$2k+m-1$	$\sqrt{\frac{6(m+k-1)}{2(m+k)+1}}$
I_{me}	$\beta_{ij} = 0$	Intercept + all LEs	$2k+m-1$	$\sqrt{2(m+k)-4}$

The first test, labeled L_1, is useful for testing whether adding a LE to a model containing only the intercept has added value. The second test in Table 2, labeled L_{me}, is useful in a scenario where the experimenter first fits a model including all m LEs and then tests whether one LE can be removed. The third and fourth test, labeled Q_{me} and I_{me}, are relevant in situations where the experimenter first fits a model including the intercept and all m LEs, and then tests whether adding a single QE or a single TFI improves the model significantly.

All non-centrality parameters λ listed in Table 2 are increasing functions of k. As a result, the powers for the four significance tests increase with k and with the number of runs. The powers for the four significance tests can all be calculated as

$$1 - \text{Prob}(-\text{t}_{\nu,\alpha/2} < T_{\nu,\lambda} < \text{t}_{\nu,\alpha/2}),$$

where $T_{\nu,\lambda}$ is a random variable following a non-central t-distribution with ν degrees of freedom and non-centrality parameter λ, and $-t_{\nu,\alpha/2}$ and $t_{\nu,\alpha/2}$ are the critical values based on a central t-distribution with ν degrees of freedom for a significance level equal to α. The non-centrality parameters and the resulting powers are independent of the sets of k columns dropped from an $(m+k)$-factor sDSD, and from the values of i and j in the effects tested (i.e., β_i, β_{ii} and β_{ij}); see supplementary Section B for details.

The λ values in Table 2 assume that the absolute values of β_i, β_{ii} and β_{ij} equal the standard deviation of the responses, σ. In other words, the non-centrality parameters we report correspond to signal-to-noise ratios of 1. To calculate the power for β_i, β_{ii} and β_{ij} values equal to $\delta\sigma$, the non-centrality parameter λ has to be multiplied by δ.

Finally, expressions for the powers of the tests L_1, L_{me}, Q_{me} and I_{me} for sDSDs can be obtained from those in Table 2 by setting $k = 0$.

3.1.2 Response surface model in two or three factors

For any two factors, an $(m+k)$-factor sDSD projects into a face-centered central composite design, in which the four factorial points each appear $(m+k)/2$ times, and the center point as well as the four axial points occur only once. This is also true for any m-factor pDSD obtained from an $(m+k)$-factor sDSD, independent of which k columns are dropped from the sDSD. As a result, all two-factor projections from a sDSD and any pDSD obtained from it are identical. All statistical properties of two-dimensional projections of sDSDs and pDSDs are therefore also identical.

Schoen et al. (2018) shows that all three-factor projections from a sDSD and from any pDSD derived from it are isomorphic. The isomorphism implies that the D-efficiency for a second-order model in three factors is the same for each three-factor projection of a sDSD and for any pDSD derived from it. Similarly, the I-efficiency is the same for each three-factor projection of a sDSD or a pDSD obtained from it.

When fitting full second-order models in two or three quantitative factors, it is common to perform significance tests for the individual QEs and the individual TFIs. Table 3 lists the four tests, the degrees of freedom ν for the tests as well as the values for the non-centrality parameter λ needed for calculating the powers of the tests. The tests labeled Q_2 and I_2 are concerned with a QE and a TFI in a two-factor response surface model, while the tests labeled Q_3 and I_3 are concerned with a QE and a TFI in a three-factor response surface model. In the expressions for the non-centrality parameters for the latter two tests, we replaced $m + k$ by n to save space. Jones and Nachtsheim (2011) also considered the tests Q_2 and I_2 for evaluating DSDs, but they did not provide analytical expressions. Expressions for the powers of these tests for sDSDs can be obtained from those in Table 3 by setting $k = 0$.

The power calculations for the hypotheses Q_3 and I_3 include two cases each because the correlations between contrast vectors involving three factors can take different signs, depending on whether $n = m+k$ is a multiple of 4 or not (recall that, due to the construction of sDSDs using conference matrices, $n = m + k$ is always even). Due to these differences in signs, the expressions for the non-centrality parameters for the hypotheses Q_3 and I_3 also differ depending on whether $n = m + k$ is a multiple of 4 or not.

As in the tests in Table 2, the non-centrality parameters in Table 3 correspond to signal-to-noise ratios of 1. To calculate the power for β_{ii} and β_{ij} values equal to $\delta\sigma$ in the tests in Table 3, the non-centrality parameter λ has to be multiplied by δ.

Table 3: Degrees of freedom ν and non-centrality parameters λ for various significance tests when using a pDSD for fitting a full second-order response surface model in two or three factors, assuming a signal-to-noise ratio of 1. Setting $k = 0$ or $n = m$ shows the results for a sDSD.

Label	Hypothesis	ν	λ	Comment
Q_2	$\beta_{ii} = 0$	$2(m+k)-5$	$\sqrt{\frac{2(4m-7)}{3(m+k-1)}}$	Any $n = m + k$
I_2	$\beta_{ij} = 0$	$2(m+k)-5$	$\sqrt{2(m+k)-4}$	Any $n = m + k$
Q_3	$\beta_{ii} = 0$	$2(m+k)-9$	$\sqrt{\frac{2(5n^3-33n^2+51n+4)}{4n^3-21n^2+24n+2}}$	$n = m + k$ is a multiple of 4
			$\sqrt{\frac{2(5n^3-43n^2+109n-86)}{4n^3-29n^2+54n-26}}$	$n = m + k$ is an odd multiple of 2
I_3	$\beta_{ij} = 0$	$2(m+k)-9$	$\sqrt{\frac{2(5n^3-33n^2+51n+4)}{5n^2-19n+14}}$	$n = m + k$ is a multiple of 4
			$\sqrt{\frac{2(5n^3-43n^2+109n-86)}{5n^2-29n+36}}$	$n = m + k$ is an odd multiple of 2

3.2 Criteria that do depend on the columns dropped

Different sets of k columns dropped from a sDSD can result in different correlations between the contrast vectors of two TFI effects β_{ij} and β_{kl}, corresponding to four different factors i, j, k and l. For an $(m + k)$-factor sDSD and any m-factor pDSD obtained from it by dropping k columns, the absolute values of the correlations between the contrast vectors of β_{ij} and β_{kl} can take the values

$$r_{ij,kl} = \frac{n - 4\tau}{n - 2},$$

where $1 \leq \tau \leq \lfloor n/4 \rfloor$, and $n = m + k$. This follows from Corollary 1 in Schoen et al. (2018). The maximum possible absolute correlation therefore is

$$1 - \frac{2}{n - 2}.$$

This expression tends to 1 as n increases, but, even for small n, it can take a fairly large value. For instance, for the 10-factor sDSD in Table 1, the maximum correlation is $1 - 2/(10 + 0 - 2) = 3/4$, while the only other possible correlation value equals 1/4. None of the correlations can be zero. Similarly, the 8-factor sDSD also only involves two different values for the absolute correlation, namely 0 and 2/3. For that sDSD, certain pairs of interactions β_{ij} and β_{kl} have uncorrelated contrast vectors, while other pairs of interactions have contrast vectors that have the maximum absolute correlation of 2/3. The 16-factor sDSD involves the absolute correlations 0, 2/7, 4/7 and 6/7 for pairs of interactions β_{ij} and β_{kl}. Note that the absolute correlation of 6/7 is the maximum one possible.

In any case, the maximum absolute correlations of 2/3, 3/4 and 6/7 for the 8-, 10- and 16-factor sDSDs show that two interactions of the types β_{ij} and β_{kl} can be strongly aliased when sDSDs or pDSDs are used, especially when k is large. A consequence of this result is that a broad range of $r_{ij,kl}$ values is possible when an m-factor pDSD is obtained from a large $(m + k)$-factor sDSD. The challenge then is to drop the k columns that result in a pDSD that avoids as many large correlations of the type $r_{ij,kl}$ as possible.

4 Best sets of k columns to drop

We performed a complete search for the best sets of k columns to drop from an $(m + k)$-factor sDSD for $m + k \in \{6, 8, \ldots, 20, 24\}$ and $k \leq 8$. To this end, we considered several statistical criteria that summarize the correlation between pairs of TFI contrast vectors with four different factors. The n-dimensional conference matrices used to construct the n-factor sDSDs were obtained from Xiao et al. (2012) for n = 8, 10, 12, 14, 16 and 18. For n = 20 and 24, sDSDs were obtained from JMP 12. We introduce the criteria in Section 4.1 and identify the best sets of k columns to drop from the sDSDs in Section 4.2. The supplementary materials of this article include R programs to reproduce all the results in this section.

4.1 Classification criteria

The pDSDs (and also the sDSDs) resemble two-level orthogonal designs of strength three. For instance, orthogonal designs of strength three provide LE contrast vectors which are neither correlated with each other nor with the TFI contrast vectors. For this reason, two-level strength-3 orthogonal designs are commonly classified in terms of the correlation between pairs of TFI contrast vectors. Two well-known criteria in this context are the maximum J_4-characteristic (Deng and Tang, 1999) and the B_4 count (Tang and Deng, 1999). The maximum J_4-characteristic measures the maximum absolute correlation between pairs of TFI contrast vectors while the B_4 count measures the sum of squared correlations. Inspired by these criteria, we considered the maximum absolute correlation and the sum of squared correlations between pairs of TFI contrast vectors involving four factors, to classify the pDSDs obtained by dropping different sets of columns. We also considered the average absolute correlation between pairs of TFI contrast vectors involving four different factors, since this criterion was discussed in Jones and Nachtsheim (2011) when evaluating sDSDs.

4.2 Results

We identified the best and worst sets of columns to drop according to each of the three classification criteria (maximum absolute correlation, average absolute correlation and sum of squared correlations). It turned out that the rankings produced by the three criteria agreed for most combinations of $m + k$ factors and k columns to drop. Detailed results on the best and worst sets of 1-8 columns to drop are given in Table S3 in supplementary Section C. Here, we present an overview of our most important results, restricting attention to $k \leq 4$. We refer to a set of columns that is best/worst in terms of all criteria as the *overall* best/worst set.

Table 4 shows the overall best sets of 1-4 columns to drop from each $(m + k)$-factor sDSD. Sometimes, there are multiple overall best sets of columns that give rise to equally good pDSDs. In that case, we report the overall best set of columns that involves the largest indices. For all cases where the set of columns dropped does not affect the quality of the resulting design, we inserted the entry 'Any' in the table.

Table 4 shows that we can drop any single column, any pair of columns and any triplet

Table 4: Overall best sets of k columns to drop from an $(m+k)$-factor sDSD, in terms of the maximum absolute correlation, average absolute correlation and sum of squared correlations between pairs of TFI contrast vectors.

# factors in sDSD	# columns dropped (k)			
$(m+k)$	1	2	3	4
6	Any	Any	Any	Any
8	Any	Any	Any	Last four
10	Any	Any	Any	6, 8, 9, 10
12	Any	Any	Any	7, 8, 10, 12
14	Any	Any	Any	Last four
16	Any	8, 16	Last three	Last four
18	Any	Any	Any	Last four
20	Any	Any	Any	14, 17, 18, 20
24	Any	Any	Any	20, 22, 23, 24

of columns from an $(m+k)$-factor sDSD without affecting the TFI contrast vectors' correlations (in other words, without affecting the aliasing among the TFIs), except when starting from the 16-factor sDSD. As a result, Dougherty et al. (2015), Fidaleo et al. (2016), Patil (2017), and Stone et al. (2014) coincidentally used the best possible pDSD for their experiment. Any other choice of columns to drop for this particular case would have resulted in an equivalent pDSD for their experiments.

Dropping different sets of four columns from a sDSD generally results in pDSDs with different values of the correlation criteria for TFIs. To construct pDSDs for 4, 10, 12 and 14 factors from 8-, 14-, 16- and 18-factor sDSDs, respectively, the overall best option is to drop the last four columns. However, the motivating example in Section 2 showed that a 21-run 6-factor pDSD obtained by dropping columns 6, 8, 9 and 10 from a 10-factor sDSD is better than the one obtained by dropping the last four columns. Dropping columns 6, 8, 9 and 10 is in fact the overall best option. Similar results hold for dropping four columns from 12-, 20- and 24-factor sDSDs.

Table 5 shows to what extent the choice of the set of columns dropped from a sDSD affects the correlations among TFI contrast vectors in a pDSD. More specifically, it shows a comparison of the average and maximum absolute correlations for TFI contrast vectors, as well as of the sum of the squared correlations, for the best and the worst set of columns dropped from a sDSD. Table 5 only covers the 10 cases in which the set of columns dropped matters when $k \leq 4$ (i.e., the cases for which Table 4 does not have the entry 'Any'). For each combination of number of factors and number of columns dropped, the results for the overall best set are shown first followed by the results for the overall worst set, except for $m + k = 16$ and $k = 4$. For this case, we report two worst sets of columns as there is no overall worst set. The first set is worst according to the average absolute correlation, whereas the second set is worst according to the sum of squared correlations. Both of the sets have the same maximum absolute correlation.

The largest difference between the best and the worst sets of k columns is for the case in which four columns are dropped from the 8-factor sDSD. For that case, the overall best set provides a maximum absolute correlation as small as 0.167, an average absolute correlation of 0.133, and a sum of squared correlations of 0.333. In contrast, the overall worst set has a maximum absolute correlation of 0.667, an average absolute correlation of 0.267, and a sum of squared correlations of 1.667. As explained in the previous section, the value of 0.667 is the maximum possible value for the correlations between pairs of TFI columns involving four different factors when $m + k = 8$.

In all other cases, the maximum correlation is not affected by the columns dropped. However, the average correlations and the sum of squared correlations are smaller when the overall best sets of columns are dropped. For $m+k = 10$, there is an appreciable difference in average absolute correlation and in the sum of squared correlations. We illustrated the impact of this difference in average absolute correlation and sum of squared correlations in Section 2. For $m+k \geq 12$, the best and worst sets of columns to drop result in only minor differences in terms of the average absolute correlation and the sum of squared correlations. In conclusion, dropping the last few columns from a sDSD is generally a good strategy, except when leaving out four columns from an 8 or a 10-factor sDSD.

Table 5: Comparison of the average and maximum correlations among TFI contrast vectors and of the sum of the squared correlations between the best and worst sets of k columns dropped from an $(m+k)$-factor sDSD.

# Factors $(m+k)$	Set size k	Run size $2(m+k)+1$	Average correlation	Maximum correlation	Sum of squared correlations
8	4	17	0.13333	0.167	0.3333
			0.26667	0.667	1.6667
10	4	21	0.20714	0.75	6.7500
			0.22143	0.75	8.2500
12	4	25	0.19048	0.4	23.7600
			0.19365	0.4	24.2400
14	4	29	0.19394	0.5	58.0000
			0.19495	0.5	58.6667
16	4	33	0.12747	0.857	115.0408
			0.13467	0.857	116.0204
			0.12867	0.857	117.2449
16	3	33	0.13173	0.857	166.0102
			0.13458	0.857	166.0102
16	2	33	0.13333	0.857	231.8571
			0.13585	0.857	231.8571
18	4	37	0.18159	0.375	201.1875
			0.18178	0.375	201.5625
20	4	41	0.17292	0.444	322.2222
			0.17311	0.444	322.8148
24	4	49	0.13479	0.364	693.3471
			0.13485	0.364	693.7438

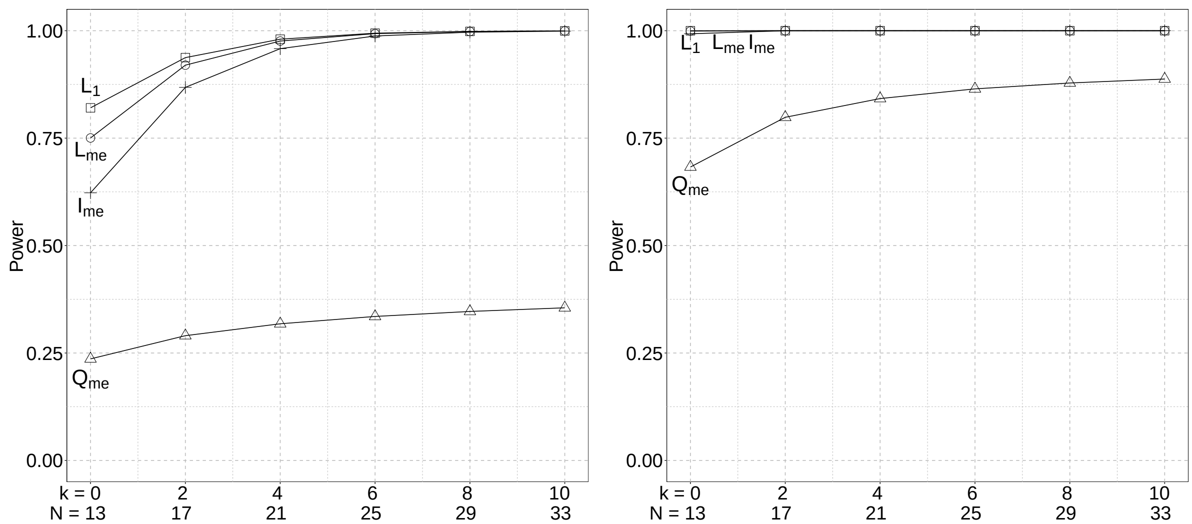
5 Comparing DSDs with different run sizes

Creating pDSDs by dropping columns from sDSDs is useful because it increases the number of runs for a given number of factors under investigation. This results in smaller standard errors and larger numbers of estimable TFIs, for instance. We demonstrate the benefits of pDSDs by looking at power curves for 6-factor designs, namely the 6-factor sDSD and five 6-factor pDSDs obtained by dropping 2, 4, 6, 8 and 10 columns from 8-, 10-, 12-, 14- and 16-factor sDSDs, respectively. Our findings support those of the simulation study in Errore et al. (2017), in that increasing the run size of the design generally leads to larger powers. However, in contrast to these authors, we use the analytical expressions in Tables 2 and 3 to determine the power curves and show separate results for the different kinds and sizes of effects. In this way, we provide more insight into the capabilities of sDSDs and pDSDs to detect active LEs, TFIs and QEs.

5.1 Power for significance tests

Figure 3 shows the power curves for the four types of tests in Table 2, assuming a significance level of 0.05 and signal-to-noise ratios of 1 and 2. The horizontal axis in the figure shows the numbers of columns dropped from the larger sDSDs and the number of runs in the resulting pDSD. The figure shows that the powers for the four tests increase with k and thus with the run size. The powers for the QEs' significance tests are much lower than the powers for the other tests. Figure 3a shows that QEs with the same size as σ are unlikely to be detected, as the power is only about 25%. Figure 3b shows that QEs twice as large are more likely to be detected. However, the power for the QEs is still markedly lower than the powers for LEs and TFIs of that size. The powers for active LEs and TFIs that are twice as large as σ equal one for the 6-factor sDSD and any pDSD constructed from a larger sDSD.

Based on Figure 3a, we cannot recommend the 6-factor sDSD involving 13 runs (and having $k = 0$) when effects as large as the noise are of interest. Instead, we recommend the 17-run pDSD used by Stone et al. (2014) and constructed by dropping two columns from an 8-factor sDSD. For this option, the powers of the tests for the LEs and the TFIs are larger than 0.86. Larger designs only marginally improve the powers. When the signal-to-noise



(a) Signal-to-noise ratio = 1

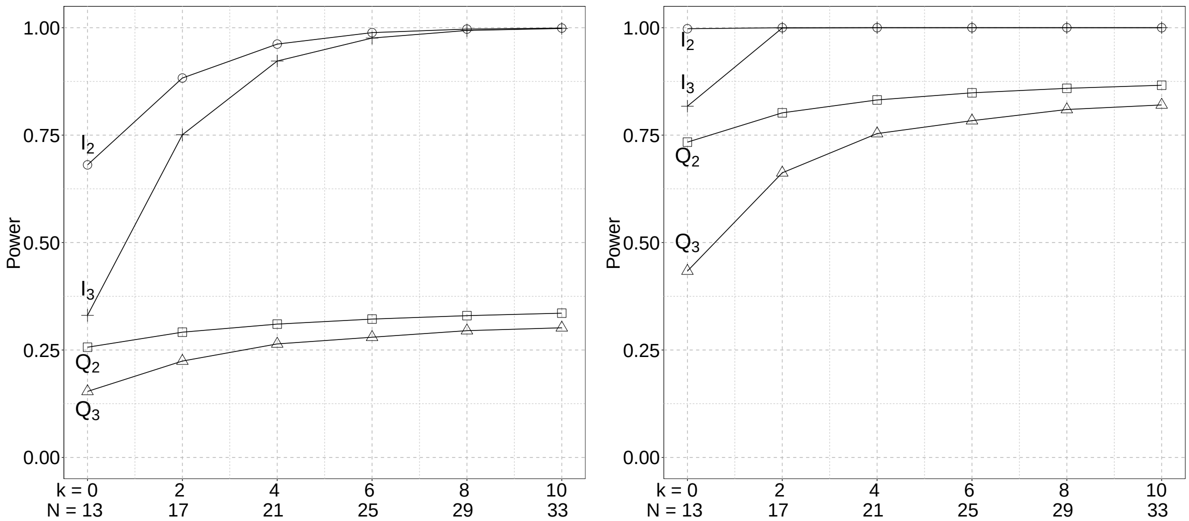
(b) Signal-to-noise ratio = 2

Figure 3: Statistical power for testing the hypotheses in Table 2 for the sDSD with $m = 6$ factors ($k = 0$) and several pDSDs ($k > 0$). N: run size of the design. $\square : L_1$; $\bigcirc : L_{\text{me}}$; $\triangle : Q_{\text{me}}$; $+ : I_{\text{me}}$.

ratio is greater than or equal to 3, the powers for all tests in Table 2 are larger than 0.90, even for the 6-factor sDSD. We conclude that it is worth considering a pDSD with four extra runs (and thus $k = 2$) when the interest is in detecting small effects. This agrees with the recommendation of Errore et al. (2017).

Figure 4 shows the power curves for the tests in Table 3 for the 6-factor sDSD and the 6-factor pDSDs constructed by dropping 2, 4, 6, 8 and 10 columns from sDSDs with 8, 10, 12, 14 and 16 factors, respectively. The signal-to-noise ratios assumed to construct the curves were again 1 and 2. Comparing this figure with Figure 3, we observe that the powers for hypothesis Q_2 hardly differ from those for hypothesis Q_{me}, while the powers for hypothesis I_2 lie between those for hypotheses I_{me} and L_{me}. The powers for hypotheses Q_3 and I_3 in the context of a three-factor model are lower than those for the hypotheses Q_2 and I_2 in the context of a two-factor model.

Figure 4a shows that QEs with the same size as σ are unlikely to be detected when a second-order model in three factors is estimated. The powers for the QEs are only about 25% in that case. The figure also shows that, for TFIs, powers of 75% or more are achieved only when a 6-factor pDSD is formed with at least four more runs than the sDSD (by



(a) Signal-to-noise ratio = 1

(b) Signal-to-noise ratio = 2

Figure 4: Statistical power for testing the hypotheses in Table 3 for the sDSD with $m = 6$ factors ($k = 0$) and several pDSDs ($k > 0$). N: run size of the design. $\square$: Q_2; $\bigcirc$: I_2; $\triangle$: Q_3; $+$: I_3.

dropping two or four columns from an 8- or 10-factor sDSD). Figure 4b shows that the powers for effects that are twice as large as the standard deviation of the noise are much larger than those for effects that are as large as the standard deviation of the noise. The power for hypothesis test Q_3, however, remains substantially smaller than 1 for any of the run sizes considered here. Signal-to-noise ratios greater than or equal to three times the noise's standard deviation result in powers larger than 0.90 for all tests listed in Table 3, except for hypothesis Q_3, in the event the sDSD is used. In conclusion, when testing QEs end TFIs in 2- or 3-factor second-order models, it pays off to use a pDSD involving more runs than the sDSD to detect effects with sizes equal to or twice the standard deviation of the noise.

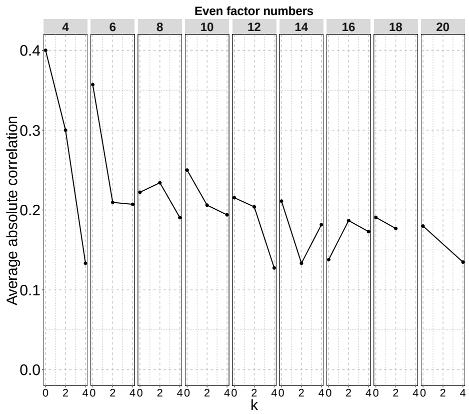
5.2 Aliasing of TFIs

Based on the results reported in Section 4, we investigated whether pDSDs have the potential to improve the aliasing pattern of TFIs in sDSDs. We studied pDSDs involving 4-18 and 20 factors constructed by dropping the overall best sets of 1-4 columns from sDSDs with up to 24 factors. For even numbers of factors up to 16, we consider the three designs obtained by dropping 0, 2 and 4 columns. For 18 and 20 factors, we consider only two different designs because there exists no 22-factor sDSD. So, 18- and 20-factor pDSDs can only be constructed starting from the 20- and 24-factor sDSDs. For odd numbers of factors up to 17, we consider the pDSDs constructed from sDSDs with one and three extra factors. For 19 factors, this is impossible, again because because there is no 22-factor sDSD. For this reason, we do not discuss the 19-factor case here.

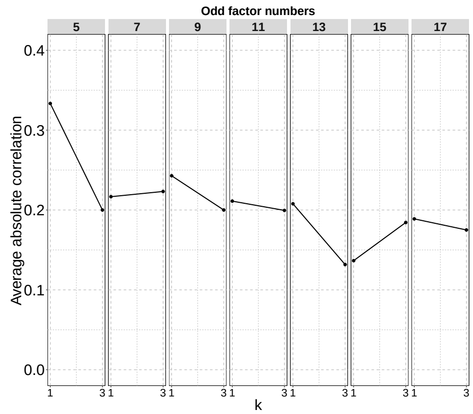
Figure 5 shows the average absolute correlation, maximum absolute correlation and sum of squared correlations between pairs of TFI contrast vectors for the designs under study. Figures 5a, 5c and 5e show the results for even numbers of factors m, while Figures 5b, 5d and 5f show the results for odd numbers of factors m. Figures 5a and 5b show the average absolute correlations, Figures 5c and 5d show the maximum absolute correlations, and Figures 5e and 5f show the logarithm of the sum of squared correlations.

Figures 5a shows that pDSDs with 4, 6, 10, 12, 18 and 20 factors have a smaller average absolute correlation between pairs of TFI contrast vectors than the corresponding sDSDs. Similarly, Figure 5b shows that 5-, 9-, 11-, 13- and 17-factor pDSDs with six extra runs and thus $k = 3$ also have a smaller average absolute correlation between pairs of TFI contrast vectors than the corresponding pDSDs with only two extra runs. The largest decrease in average correlation is for the 4-factor designs where the 9-run sDSD provides an average absolute correlation of 0.4, while the 17-run pDSD obtained from the 8-factor sDSD has an average as low as 0.13.

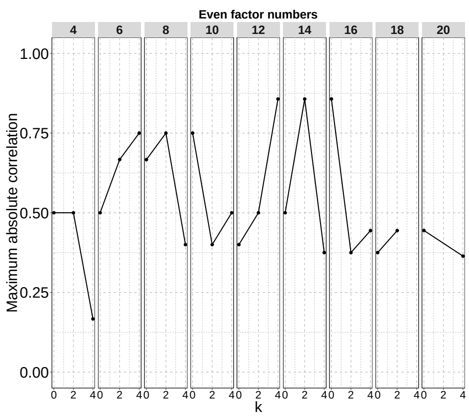
For 7 and 15 factors, increasing the run size by four (i.e., using $k = 3$ instead of $k = 1$) causes the average absolute correlation between pairs of TFI contrast vectors to go up. For eight factors, the pDSD with four extra runs (corresponding to $k = 2$) has a larger average absolute correlation than the sDSD ($k = 0$), and it is the pDSD with eight extra runs ($k = 4$) which has the smallest average absolute correlation. For 14 and 16 factors, the best design options in terms of the average absolute correlation are the pDSD obtained by dropping two columns from the 16-factor sDSD and the 16-factor sDSD itself, respectively. The 16-factor sDSD turns out to perform well in terms of the average correlation, as the best 12-, 13-, 14-, 15- and 16-factor designs in terms of that criterion are all based on it.



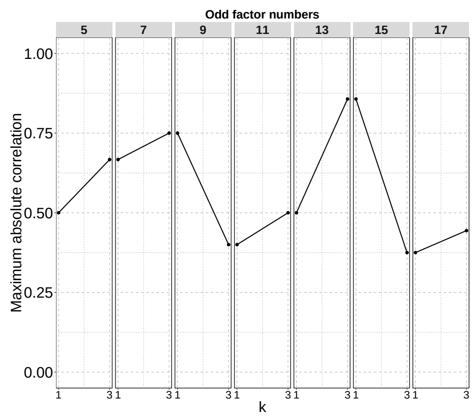
(a) Even factor numbers: average correlation



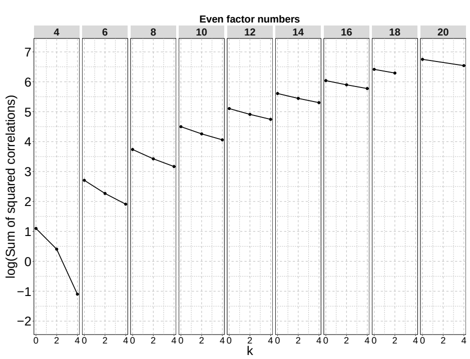
(b) Odd factor numbers: average correlation



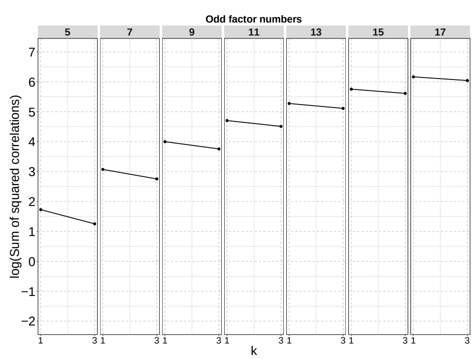
(c) Even factor numbers: maximum correlation



(d) Odd factor numbers: maximum correlation



(e) Even factor numbers: sum of squared correlations



(f) Odd factor numbers: sum of squared correlations

Figure 5: Correlations between pairs of TFI contrast vectors.

Figures 5c and 5d show that the patterns in the maximum absolute correlations are quite different from those in the average absolute correlations. More specifically, both figures show that, for designs with 5-7, 11-13, 17 and 18 factors, the maximum absolute correlation increases with the run size. The largest increase in maximum absolute correlation is for designs with 12 factors. The 12-factor sDSD provides a maximum absolute correlation of 0.4, while the pDSD with eight extra runs (corresponding to $k = 4$) exhibits a maximum of 0.857. For 8-10, 15, 16 and 20 factors, however, there are pDSDs which have smaller maximum absolute correlations than the corresponding sDSDs.

Regarding the sum of squared correlations between pairs of TFI contrast vectors, Figures 5e and 5f show that increasing the run size of pDSDs reduces the sum of squared correlations for all numbers of factors. So, although the average and maximum absolute correlations between TFI contrast vectors are exacerbated when using specific pDSDs instead of sDSDs, the sum of squared correlations of the larger options is always smaller than for the sDSDs.

Figure 5 shows that the pDSD options used by Patil (2017) and Stone et al. (2014) were not optimal in terms of the maximum absolute correlation between pairs of TFI contrast vectors. While the 6-factor design with 17 runs of Stone et al. (2014) provides a smaller average absolute correlation than the 6-factor sDSD, it has a larger maximum absolute correlation (0.67 versus 0.5). The 7-factor design with 21 runs and $k = 3$ of Patil (2017) has larger maximum and average absolute correlations between its pairs of TFI contrast vectors than the 7-factor design with 17 runs and $k = 1$.

If there is one thing that Figure 5 makes clear, it is that certain sDSDs and pDSDs involve very large absolute correlations between pairs of TFIs, indicating close to complete aliasing. Particularly unfavorable in this respect are the pDSDs constructed by dropping columns from the 16-factor sDSD, because all of these designs have quite a number of absolute correlations of 0.857 (despite the fact that the average correlations for this design are small). However, these large correlations only become problematic if many TFIs are active. Figure 5 also shows that, if a large maximum absolute correlation is a major concern, alternative design options with a maximum absolute correlation below 0.5 are available for all numbers of factors, except 7. For applications involving seven factors in which many TFIs are expected to be active, we recommend dropping five columns from the 12-factor sDSD because the absolute correlations between the TFI contrast vectors for the resulting design are smaller than or equal to 0.4.

6 Discussion

In this article, we studied projected DSDs or pDSDs for m factors constructed by dropping sets of k columns from sDSDs with $m + k$ factors. We considered sDSDs with 6-24 factors, and studied the pDSDs resulting from dropping sets of 1-4 columns. Table S3 in supplementary Section C includes additional results we found on dropping up to eight columns.

The sDSDs used in this study were constructed from conference matrices. This allowed us to derive analytical expressions for several criteria from the literature on DSDs. In supplementary Section B, we derive expressions for the relative D-efficiency to estimate the LEs model and the LEs-plus-QEs model for pDSDs with different run sizes, as well as analytical expressions for the relative standard errors for the LE and QE estimates. We also derived expressions for the non-centrality parameter required for calculating the power of various significance tests. We showed that the correlations between two QE contrast vectors, a QE and a TFI contrast vector, and between two TFI contrast vectors involving a common factor are independent of the set of k columns dropped from the $(m + k)$-factor sDSD. Supplementary Section B also includes analytical expressions for these correlations.

How well multiple TFIs can be estimated at the same time depends on the selection of the sets of k columns to drop from the sDSDs. Using a complete search, we identified the best sets of columns to drop in terms of the average absolute correlation, the maximum absolute correlation, and the sum of squared correlations between pairs of TFI contrast vectors. The differences between the best and worst sets were largest when dropping four columns from the 8- and 10-factor sDSDs. Table S3 in supplementary Section C shows moderate or small differences when dropping more than four columns from a sDSD except when dropping columns from the 10-factor sDSD or eight columns from the 12-factor sDSD. For these cases, the maximum absolute correlation of the worst option is more than three times as large as that of the best option. We conclude that dropping the last few columns from a sDSD constructed using the method of Xiao et al. (2012) is generally a good strategy.

We also compared designs with different run sizes constructed by dropping columns from sDSDs with different run sizes in terms of the average absolute correlation, the maximum absolute correlation, and the sum of squared correlations between pairs of TFI contrast vectors. We found that increasing the run size for a given number of factors, which is equivalent to dropping more columns from larger sDSDs, improves the sum of squared correlations between pairs of TFI contrast vectors. However, the average and maximum absolute correlations do not necessarily improve. In fact, these values may even increase with the run size of the pDSD. Thus, in order to limit the amount of aliasing between TFIs, a careful design selection is needed. For certain sDSDs and pDSDs, quite large numbers of TFIs are nearly completely aliased.

The sDSDs have also been adapted to deal with two-level categorical factors and with blocking factors. The methods developed by Jones and Nachtsheim (2013) and Nguyen and Pham (2016) to include k two-level categorical factors in a DSD transform the last k columns into two-level columns. Picking other columns than the last k may yield better designs. Similarly, the blocking schemes of Jones and Nachtsheim (2016) convert the last k columns of DSDs into blocking factors. Possibly, better designs can be obtained by using other columns to create the blocking factor. Investigating these issues would be an interesting avenue for future research too.

SUPPLEMENTARY MATERIAL

Supplementary sections.pdf Details on the simulation protocol and results; discussion and derivations of design properties; and, table with detailed results about dropping from 1-8 columns from standard definitive screening designs.

Supplementary files.zip Simulation protocol and programs to identify the best sets of columns to drop from standard definitive screening designs.

ACKNOWLEDGMENTS

The research that led to this article was financially supported by the Flemish Fund for Scientific Research FWO.

References

Colbourn, C. J. and Dinitz, J. H. (2006). *Handbook of Combinatorial Designs*. CRC Press.

Deng, L.-Y. and Tang, B. (1999). Generalized resolution and minimum aberration criteria for Plackett-Burman and other nonregular factorial designs. *Statistica Sinica*, 9:1071–1082.

Dougherty, S., Simpson, J. R., Hill, R. R., Pignatiello, J. J., and White, E. D. (2015). Effect of heredity and sparsity on second-order screening design performance. *Quality and Reliability Engineering International*, 31:355–368.

Errore, A., Jones, B., Li, W., and Nachtsheim, C. J. (2017). Using definitive screening designs to identify active first- and second-order factor effects. *Journal of Quality Technology*, 49:244–264.

Fidaleo, M., Lavecchia, R., Petrucci, E., and Zuorro, A. (2016). Application of a novel definitive screening design to decolorization of an azo dye on boron-doped diamond electrodes. *International Journal of Environmental Science and Technology*, 13:835–842.

Georgiou, S. D., Stylianou, S., and Aggarwal, M. (2014). Efficient three-level screening designs using weighing matrices. *Statistics*, 48:815–833.

Jones, B. and Montgomery, D. C. (2010). Alternatives to resolution IV screening designs in 16 runs. *International Journal of Experimental Design and Process Optimization*, 1:285–295.

Jones, B. and Nachtsheim, C. J. (2011). A class of three-level designs for definitive screening in the presence of second-order effects. *Journal of Quality Technology*, 43:1–15.

Jones, B. and Nachtsheim, C. J. (2013). Definitive screening designs with added two-level categorical factors. *Journal of Quality Technology*, 45:121–129.

Jones, B. and Nachtsheim, C. J. (2016). Blocking schemes for definitive screening designs. *Technometrics*, 58:74–83.

Jones, B. and Nachtsheim, C. J. (2017). Effective design-based model selection for definitive screening designs. *Technometrics*, 59:319–329.

Nguyen, N.-K. and Pham, T.-D. (2016). Small mixed-level screening designs with orthogonal quadratic effects. *Journal of Quality Technology*, 48:405–414.

Patil, M. V. (2017). Multi response simulation and optimization of gas tungsten arc welding. *Applied Mathematical Modelling*, 42:540–553.

Schoen, E. D., Eendebak, P. T., and Goos, P. (2018). A classification criterion for definitive screening designs. *The Annals of Statistics*. To appear.

Stone, B. B., Montgomery, D. C., Hassler, E., and Silvestrini, R. T. (2014). An expected cost methodology for screening design selection. *Quality Engineering*, 26:139–153.

Tang, B. and Deng, L.-Y. (1999). Minimum G_2-aberration for nonregular fractional factorial designs. *The Annals of Statistics*, 27:1914–1926.

Xiao, L., Lin, D. K., and Bai, F. (2012). Constructing definitive screening designs using conference matrices. *Journal of Quality Technology*, 44:1–7.