This is the accepted manuscript *of an article published by Taylor & Francis in IISE Transactions, available online: http://www.tandfonline.com/doi/full/10.1080/0740817X.2016.1241458.*

Frequency tables for the coding invariant quality assessment of factorial designs

Ulrike Grömping, Beuth University of Applied Sciences Berlin, Germany

Abstract

Quality assessment of factorial designs, particularly mixed level factorial designs, is a non-trivial task. Existing methods for orthogonal arrays include generalized minimum aberration, a modification thereof that was proposed by Wu and Zhang for mixed two- and four-level arrays, and minimum projection aberration. For supersaturated designs, E(s^2) or χ^2 based criteria are widely used. Based on recent insights by Grömping and Xu regarding the interpretation of the projected a_R values used in minimum projection aberration, this paper proposes three new types of frequency tables for assessing the quality of level-balanced factorial designs. These are coding invariant, which is particularly important for designs with qualitative factors. The proposed tables are used in the same way as those used in minimum projection aberration and behave more favorably when used for mixed level arrays. Furthermore, they are much more manageable than the above-mentioned approach by Wu and Zhang. The paper justifies the proposed tables based on their statistical information content, makes recommendations for their use and compares them with each other and with existing criteria. As a byproduct, it is shown that generalized minimum aberration refines the established expected χ^2 criterion for level-balanced supersaturated designs.

Key words: Design of Experiments, Squared Canonical Correlation Frequency Tables, Average R^2 Frequency Tables, Generalized Resolution, Ranking Factorial Designs

1. Introduction

Experimental design is an important tool for gaining as much information as possible from a limited number of experimental runs. Orthogonal arrays (OAs) are widely used for screening experiments and for experiments with qualitative factors with a view to more detailed modeling of factorial effects; screening experiments are sometimes even done using arrays with non-orthogonal columns, e.g. in the so-called supersaturated designs (see e.g. Georgiou 2014). This paper discusses quality metrics for level-balanced arrays (BAs) and the experimental designs based on them. Before going into the specifics, some terminology is provided: arrays are Nxn matrices of symbols. The n columns correspond to design factors, the N rows to experimental runs. A subset of k columns is called a k-factor projection or k-factor set. The symbols that occur in a column are called levels. Arrays with the same number of levels in all columns are called fixed or pure level or symmetric arrays. If different columns may have different numbers of levels, an array is called mixed level or asymmetric. All arrays considered in this paper are level-balanced, i.e., each column contains each of its symbols the same number of times. In addition, for OAs, each pair of columns contains each pair of levels the same number of times. An array is said to be of strength t, if each k-factor set, $k \leq t$ contains each k-tuple of levels the same number of times. In the statistical literature for OAs, strength t is often denoted as resolution $R=t+1$ (see e.g. Hedayat, Sloane and Stufken p. 280 for the equivalence). Obviously, OAs per definition have at least strength $t=2$ ($R=3$), while BAs have at least strength $t=1$; we extend the usual equivalence of $R=t+1$ from $R\geq3$ to $R=2$, by defining $R=2$ through $t=1$.

In the screening phase of the experimental process, the number of experimental runs is usually required to be small, while attempting to accommodate relatively many factors, and there will not be detailed knowledge on a model for which to optimize a design. Rather, the design should be model-robust. Given the reasonable and frequently-made assumption that lower order effects are more likely than higher order effects to be active, the typical screening design is requested to be able to estimate at least the factors' main effects with as little bias risk as possible from low order interactions. In cases with particularly strong requirements on run economy relative to the number of factors, one might even consider strength 1 ($R=2$) arrays as suitable for the screening task. For quantitative factors, it is common to consider two levels per factor in the screening phase. For qualitative factors, the experimental purpose often dictates the numbers of factor levels for some of the factors, which may lead to a need for the use of mixed level arrays. As the construction of parsimonious mixed level arrays is by no means straightforward for the practitioner, some collections of OAs for such situations are available in literature, web and software, e.g. Taguchi (1987), Hedayat, Sloane and Stufken (1999), Kuhfeld (2009), Eendebak and Schoen (2013), Grömping (2016). Catalogues for supersaturated designs are also available, e.g., by Gupta et al. (2008). Recently, with new algorithms for checking isomorphism of arrays, some authors have discussed the creation of complete catalogues of non-isomorphic arrays, both for pure level and mixed level cases (e.g. Stufken and Tang 2007, Evangelaras, Koukouvinos and Lappas 2007, 2011, Schoen 2009, Schoen, Eendebak and Nguyen 2010). Of course, such catalogues are useful only if there are criteria for choosing designs from them. Quality criteria commonly used for OAs include generalized minimum aberration (GMA) which is based on the generalized word length pattern (GWLP) by Xu and Wu (2001) and minimum projection aberration which is based on projection frequency tables (PFTs; Xu, Cheng and Wu 2004). While both these criteria have been applied to mixed level arrays (see e.g. Schoen 2009, Xu, Phoa and Wong 2009), their validity for that situation has been conceptually questioned as early as 1993 by Wu and Zhang (henceforth WZ). For supersaturated designs, E(s^2), ave(χ^2), E(χ^2) or E(f_{NOD}) criteria are frequently used (Booth and Cox 1962, Yamada and Matsui 2002, Ai, Fang and He 2007, Fang, Lin and Liu 2003); criteria based on the maximum contributors to those averages / expectations are also in use. This paper aims at providing tractable quality criteria for BAs that fairly treat mixed level arrays. The recent work by Grömping and Xu (2014) will be useful in obtaining three new and conceptually convincing ways to replace conventional criteria by refined versions that take care of mixed level BAs in a coding invariant way and such that WZ's concerns are also addressed. The coding invariance is particularly desirable for qualitative factors, which often occur in mixed level arrays. One of the proposed criteria will also prove useful for distinguishing pure level arrays with identical GWLP, i.e. can be used as an easy means for establishing non-equivalence of arrays. The criteria can be used for deciding between several arrays from smaller selections for concrete experiments, but also for ranking larger catalogues of arrays. For the latter application, computing effort is an issue; however, since the criteria concentrate on R-factor sets only, the effort remains manageable; of course, very large catalogues will nevertheless pose a challenge, which can, however, be expected to reduce with technical development. The methods proposed here, as implemented in the R package DoE.base (Grömping 2016), have already been used for the creation of design matrices for specific screening experiments with only little confounding among 2-factor interactions, as was e.g. reported in Vasilev et al. (2014).

Section 2 presents the existing quality criteria. Section 3 deep-dives the projected a_R values that underlie GMA and the PFTs: it sheds light on their statistical implications regarding bias and (im)precision, points out their relation to the criteria for supersaturated designs, and explains why tabulation of raw projected a_R values can be improved upon in case of mixed level arrays. Section 4 presents the new criteria and gives recommendations for their use in ranking and non-equivalence detection of arrays. Section 5 provides several examples for their performance. The paper closes with a discussion.

The following notation will be used: the letter s stands for the number of factor levels. A BA of resolution R = strength R–1 in N runs with n factors will be denoted as BA(N, s_1, …, s_n, R–1), with s_1, …, s_n, possibly but not necessarily distinct, or as BA(N, s_1^{n1}…s_k^{nk}, R–1) with s_1, …, s_k, possibly but not necessarily distinct and n_1+…+n_k=n (whichever is more suitable for the purpose at hand). For OAs, "BA" will be replaced by "OA" in this notation. The unsquared letter R always refers to the resolution of an array, while R^2 denotes the coefficient of determination. k-factor sets are denoted by an index set $\{u_1,\ldots,u_k\}$, or by the union $\{c\}\cup$C, where c denotes a single factor, C a set of k–1 additional factors. R-factor sets are of particular interest for this paper.

2. Basic definitions and results

This section restates the most important definitions and results from the literature in concise form.

2.1. Resolution, GWLP and GMA

The GWLP with entries $A_0, A_1, A_2, A_3, A_4,\ldots, A_n$ contains the numbers of *generalized words* of lengths 0, 1, 2, 3, 4, …, n and generalizes the word length pattern (WLP), which is known from symmetric regular fractional factorial arrays that can be specified via generators: there, each generator implies a "word", and combinations of generators imply further words. The WLP for regular 2-level arrays is a special case of the GWLP, whereas one word in the WLP for regular symmetric s-level arrays, as e.g. for the 3-level arrays listed in Chen, Sun and Wu (1993), corresponds to s–1 generalized words in the GWLP. The GWLP is reasonably defined for general, not necessarily symmetric or regular, arrays. The resolution of the array is the number R for which $A_R > 0$, but $A_k = 0$ for all k with $0 < k < R$. The GMA criterion ranks arrays by minimizing the GWLP entries from left to right, which automatically maximizes the resolution; clearly, GMA is equivalent to MA for regular symmetric arrays with arbitrary numbers of levels. According to Xu and Wu (2001), who introduced the GWLP, A_k measures the overall confounding between all k-factor interactions and the general mean (see also the formal definition below).

Formally, the GWLP of a BA(N, $s_1,\ldots,s_n$, $R-1$) is most easily defined using the model matrix $\mathbf{M}$ = ($\mathbf{M}_0$, $\mathbf{v}$, $\mathbf{M}_2$, …, $\mathbf{M}_n$) of the full model up to the n factor interaction: $\mathbf{M}_0$ is a column of "+1"s, $\mathbf{M}_1$ the matrix of the n main effects model matrices $\mathbf{X}_i$ ($N\times(s_i-1)$) in orthogonal coding, with all main effects columns normalized to mean 0 and squared length N; for $2 \le k \le n$, $\mathbf{M}_k$ is the matrix of all $\binom{n}{k}$ k-factor interaction model matrices, i.e. $\mathbf{M}_k$ = ($\mathbf{X}_{1\ldots k}$, …, $\mathbf{X}_{n-k+1\ldots n}$), with $\mathbf{X}_{u1\ldots uk}$ the $N\times((s_{u1}-1)\cdot\ldots\cdot(s_{uk}-1))$ model matrix of the interaction among factors $\{u_1,\ldots,u_k\}$ obtained as element-wise products of one column from each of the k main effects model matrices (see e.g. Table 5 below). The coding of matrix $\mathbf{M}_1$ (orthogonal, squared column length normalized to N) will be called "normalized orthogonal coding" in the following. The elements $A_0,A_1,A_2,A_3,\ldots, A_n$ of the GWLP can be calculated as the sums of squared column averages of the respective portions of $\mathbf{M}$, i.e. $A_k=\mathbf{1}^{\mathrm{T}}\mathbf{M}_k\mathbf{M}_k^{\mathrm{T}}\mathbf{1}/N^2$. These sums can be split into contributions from the separate k-factor sets, i.e.

$$A_k = \sum_{\substack{\{u_1,\ldots,u_k\}\\ \subseteq\{1,\ldots,n\}}} \mathbf{1}_N{}^{\mathrm{T}}\mathbf{X}_{u_1\ldots u_k}\mathbf{X}_{u_1\ldots u_k}{}^{\mathrm{T}}\mathbf{1}_N \Big/ N^2 = \sum_{\substack{\{u_1,\ldots,u_k\}\\ \subseteq\{1,\ldots,n\}}} a_k\left(u_1,\ldots,u_k\right); \qquad (1)$$

the $a_k(u_1,\ldots,u_k)$ are called projected a_k values and are the basis of minimum projection aberration, as presented in the next section; they contain the number of generalized words of length k that the k-factor set contributes, and measure the overall confounding of the set's k-factor interaction with the general mean. Section 3 will discuss in detail in how far the projected a_R values (with R the resolution) measure confounding between each main effect and the R–1 factor interaction among the respective other factors in the set. The most important cases for screening experiments are those of R=3 and R=2, i.e., resolution III arrays or (supersaturated) resolution II arrays are most often used for screening experiments; for these, the GWLP reduces to (A_2,)A_3, A_4, …, A_n, and projected a_3 or a_2 values are of interest. Notationally, the resolution R is kept general, however.

2.2. Minimum projection aberration

For resolution R arrays, the minimum projection aberration criterion looks at the projected a_R values $a_R(u_1,\ldots,u_R)$ of the different R-factor sets and ranks array d_2 as better than array d_1, if d_2 has fewer R-factor projections with high projected a_R values. This criterion can be assessed using the so-called projection frequency tables (PFTs), as defined below:

Table 1: The five GMA OA(16, $2^3 4^2$, 2) and array d_3 from WZ (obtained from a listing of all 17 non-isomorphic OA(16, $2^3 4^2$, 2) which was provided by Eric Schoen)

	1 (d_3)					2 (d_1)					3					4 (d_2)					5					6				
1	1	2	2	3	3	2	1	1	2	4	1	2	2	4	3	2	2	1	4	1	1	2	2	2	1	1	2	2	1	2
2	1	1	2	1	2	2	1	2	3	1	2	1	2	3	1	1	2	2	1	2	1	2	1	4	4	1	1	1	3	4
3	1	1	1	4	4	2	2	1	1	3	2	2	2	3	2	1	1	1	1	1	2	1	1	2	4	2	2	1	3	2
4	2	2	1	1	3	1	1	1	4	3	2	2	1	1	3	1	1	1	4	4	1	2	2	1	2	1	1	1	1	1
5	2	1	2	2	4	1	1	2	4	4	1	1	1	3	3	1	2	2	2	1	2	1	2	4	2	1	2	2	3	3
6	1	2	2	2	2	1	2	1	2	2	1	1	2	1	2	1	1	1	2	2	2	2	1	4	1	2	2	1	4	1
7	2	1	1	3	2	2	2	2	1	4	2	1	2	2	3	2	1	2	1	3	2	2	1	3	2	1	2	2	2	1
8	1	1	1	1	1	1	2	1	3	4	2	2	1	4	1	2	2	1	3	2	2	1	1	1	3	2	1	2	1	3
9	1	1	2	4	3	1	2	2	3	3	1	1	1	1	1	1	2	2	3	4	2	2	2	1	4	1	1	1	2	2
10	2	2	2	4	1	1	1	1	1	1	1	2	1	2	2	1	1	1	3	3	2	1	2	3	1	1	2	2	4	4
11	2	2	2	1	4	1	1	2	1	2	2	1	1	4	2	2	1	2	4	2	1	2	1	3	3	2	2	1	1	4
12	1	2	1	2	1	2	1	2	2	3	2	1	1	2	4	2	1	2	3	1	2	2	2	2	3	1	1	1	4	3
13	2	1	1	2	3	1	2	2	2	1	2	2	2	1	4	2	2	1	1	4	1	1	1	2	2	2	2	1	2	3
14	2	2	1	4	2	2	2	1	4	1	1	2	1	3	4	2	1	2	2	4	1	1	2	3	4	2	1	2	3	1
15	1	2	1	3	4	2	1	1	3	2	1	1	2	4	4	2	2	1	2	3	1	1	1	1	1	2	1	2	2	4
16	2	1	2	3	1	2	2	2	4	2	1	2	2	2	1	1	2	2	4	3	1	1	2	4	3	2	1	2	4	2

Definition 1 (Case $k=R$ in Xu, Cheng and Wu 2004)

(i) For a BA(N, $s_1,\ldots,s_n$, $R-1$), the projection frequency table PFT$_k$ ($k \geq R$) is defined as the frequency table of the $\binom{n}{k}$ values $a_k(u_1,\ldots,u_k)$, $\{u_1,u_2,\ldots,u_k\} \subseteq \{1,\ldots,n\}$.

(ii) Minimum projection aberration ranks arrays according to their PFT$_R$, by minimizing the frequency of the largest $a_R(u_1,u_2,\ldots,u_R)$, in case of ties the frequency of the second largest $a_R(u_1,u_2,\ldots,u_R)$, and so forth. In case of identical PFT$_R$, a version of minimum projection aberration continues considering the PFT$_{R+1}$ etc.

Example 1: Table 1 shows six OA(16, $2^3 4^2$, 2), three of which are isomorphic to those investigated by Wu and Zhang (1993); these are the ones labeled with d_1 to d_3. Array d_3 is the only non-GMA array in the table. Table 2 shows the GWLP and PFT_3 for the arrays from Table 1. According to GMA, array 1 (d_3) is worst, the other five arrays are equivalent. PFT_3 further distinguishes array 3 from the other GMA arrays: it has only 3 instead of 4 triples with a projected a_3 value of "1" and is therefore considered better. Among the three Wu and Zhang arrays, d_1 and d_2 are equivalent, while d_3 is worse.

Table 2: GWLP and PFT_3 for the arrays of Table 1

	GWLP				PFT_3			
	Rank	A_3	A_4	A_5	Rank	0	0.5	1
1 (d_3)	6	5	1	1	6	5	0	5
2 (d_1)	1	4	3	0	2	6	0	4
3	1	4	3	0	1	5	2	3
4 (d_2)	1	4	3	0	2	6	0	4
5	1	4	3	0	2	6	0	4
6	1	4	3	0	2	6	0	4

2.3. Wu and Zhang (1993)

WZ proposed to treat different types of projections in mixed level arrays differently. Their solution is very specific to the arrays they studied: arrays with factors at four and two levels, and at most two factors at four levels. While their approach is interesting, it is messy to generalize it to general mixed level arrays or even arrays with more 4-level factors, and the author does not know of any such work (for other generalizations, see below). The key idea is to distinguish between k-factor sets of only 2-level factors (i.e. zero 4-level factors), k-factor sets with one 4-level factor, and k-factor sets with two 4-level factors. Accordingly, WZ partitioned the overall number of words of length k into components $A_k = A_{k0} + A_{k1} + A_{k2}$, where the second index indicates the number of 4-level factors in the k-factor set. WZ proceeded by defining "Type 0 minimum aberration" as minimum aberration based on A_{k0}, resolving ties in A_{k0} by using A_{k1} (and so forth). (Their second concept, "Type 1 minimum aberration", will not be pursued here.)

Example 1 continued: The six OA(16, $2^3 4^2$, 2) of Table 1 have the WZ patterns shown in Table 3. Thus, according to the Type 0 MA criterion, d_1 is best (equivalent to 3 and 5), followed by d_3 and d_2 (equivalent to 6) in that order. This ranking invoked a skeptical remark of Wu and Zhang regarding the universal usefulness of their criterion: they did not like the ranking of d_2 behind d_3.

Table 3: The Wu and Zhang patterns for the arrays of Table 1

	Rank	A_{30}	A_{31}	A_{32}	A_{40}	A_{41}	A_{42}	A_{52}
1 (d_3)	4	0	2	3	0	0	1	1
2 (d_1)	1	0	1	3	0	1	2	0
3	1	0	1	3	0	1	2	0
4 (d_2)	5	1	0	3	0	0	3	0
5	1	0	1	3	0	1	2	0
6	5	1	0	3	0	0	3	0

WZ restricted attention to a very specific class of arrays, for which the 4-level factor(s) can be constructed from the first two or four base factors of a regular fractional factorial 2-level array; all their arrays are therefore regular. Array 3 of Example 1 is not of that nature, but nevertheless shows the same WZ pattern as the best WZ array. For the 16 run cases, it has been investigated whether there are better arrays according to the WZ criterion within the larger set of all the non-isomorphic arrays with the respective numbers of factors and levels. (Attention was restricted to the 16 run arrays, because there is an easily manageable number of them, whereas the 32 runs arrays have not even been enumerated by Schoen et al. 2010.) The supplementary materials provide all the criterion values for all non-isomorphic OA(16, $4^1 2^a$, 2) and OA(16, $4^2 2^a$, 2). It was found that the WZ Type 0 MA arrays remain best in the overall set of arrays according to the WZ criterion (verified for up to A_{4j} values only). However, their performance regarding the new criteria is generally lacking – in many cases they are close to the worst arrays. This is not surprising because regular arrays have repeatedly been found to have undesirable projection properties, e.g. by Cheng and Wu (2001) or Xu and Deng (2005).

An obvious generalization of WZ's method for arrays with two numbers of levels of any sort, e.g. with 2- and 3-level factors, is to look at exactly the same concept, with the second index providing the number of factors with more levels in the set. For example, an OA(18, 2^1 3^7, 2) would have A_{32} and A_{33}, A_{43} and A_{44} and so forth. In this way, it is possible to handle any arrays with only two different numbers of levels. However, if both numbers of levels occur with higher frequency, the situation becomes more complex. For example, for an OA(36, 2^5 3^6, 2), there are A_{30}, A_{31}, A_{32}, A_{33}, A_{40}, A_{41}, A_{42}, A_{43}, A_{44}, and so forth. In such cases, WZ's approach of primarily ranking w.r.t. one particular type of words and using the other types of words for resolving ties only becomes more and more problematic. This has also been noted by WZ, who discussed in the end of their paper to also take second best arrays into consideration or to take weighted sums of the A_{ij}. Generalization of the WZ approach for arrays with more than two different numbers of levels becomes even more cumbersome than outlined above, because the definition and ranking of the types of numbers of words to look at has to be tackled. Section 5 contains an example with three different numbers of levels, for which a third subscript to the "A"s has been introduced and an order of the word types has been arbitrarily fixed; this underlines the complexity involved in WZ's approach.

2.4. Established criteria for supersaturated designs

The earliest quality criterion for supersaturated designs in 2-level factors was the E(s^2) criterion (Booth and Cox 1962), which is known to be refined by GMA (see e.g. Xu 2015). For mixed level supersaturated designs, two criteria related to χ^2 values are in use: for a pair (u,v) of factors in s_u and s_v levels, respectively, dependency within that pair can measured by

$$\chi^2(u,v)=\sum_{j=1}^{s_v}\sum_{i=1}^{s_u}\left(n_{ij}-\frac{n_{i\bullet}n_{\bullet j}}{N}\right)^2 \Bigg/ \frac{n_{i\bullet}n_{\bullet j}}{N} \underset{\substack{\text{in case of}\\ \text{level balance}}}{=} \frac{s_u s_v}{N}\sum_{j=1}^{s_v}\sum_{i=1}^{s_u}\left(n_{ij}-\frac{N}{s_u s_v}\right)^2, \qquad (2)$$

with n_{ij} the frequency of the level combination i and j and the usual dot notation for sums ($n_{\bullet\bullet}=N$). Yamada and Matsui (2002) and Ai, Fang and He (2007) equivalently proposed minimization of the sum, average or expectation of the $\chi^2(u,v)$, called ave(χ^2) criterion in the sequel, whereas Fang, Lin and Liu (2003) proposed minimization of E(f_{NOD}), where $f_{\text{NOD}}(u,v) = N\,\chi^2(u,v)/(s_u s_v)$. Booth and Cox

(1962) and later authors also proposed to minimize the worst contribution to their criteria, which is procedurally comparable to minimum projection aberration for 2-factor projections.

3. The case for new projection based metrics for mixed level arrays

Projection frequency tables tabulate the projected a_R values for all R-factor sets. Xu, Cheng and Wu (2004) introduced PFTs for pure 3-level arrays only. For mixed level arrays, the projected a_R values correspond to R-factor sets of different patterns of numbers of factor levels. This issue was raised by WZ for mixed 2-level and 4-level arrays, and by Grömping (2011, 2013) in general. The following two sections shed light on the nature of the projected a_R values, regarding how they can be decomposed into statistically meaningful components, how they relate to (2), and how these insights give rise to new tabulations for mixed level arrays.

3.1. Projected a_R values: decomposition

The following two lemmata state two results from Grömping and Xu (2014) that are important for this paper; all proofs of Grömping and Xu (2014) generalize to R=2; the results are therefore stated for BAs. Before formally stating them, they are briefly given in words: according to Lemma 1, each R-factor set in a resolution R array contributes to A_R the sum of the R^2 values from explaining the s_c-1 main effects columns (in orthogonal coding) of an arbitrary singled out factor in s_c levels by a full model in the other $R-1$ factors; alternatively, according to Lemma 2, the contribution can be thought of as the sum of squared canonical correlations (SCCs) between the main effects model matrix for one of the factors (in arbitrary coding) and the full model matrix for the other $R-1$ factors. Note that orthogonal coding encompasses orthogonality to the intercept column, i.e. the classical dummy coding, although yielding main effects model columns with zero pair wise scalar products, is not considered to be orthogonal coding. Furthermore, note that the individual R^2 values depend on the particular choice of orthogonal coding, while their sum or average is independent of that choice, and that canonical correlations are generally coding invariant.

<u>Lemma 1 (Grömping and Xu 2014).</u>

Consider a BA(N, $s_1 \ldots s_n$, $R-1$) and an R factor set $\{c\} \cup \mathrm{C} \subseteq \{1,\ldots,n\}$. Denote by $\mathbf{X}_c$ the $N \times (s_c-1)$ main effects model matrix in orthogonal coding for the factor c. Then, the projected a_R value

$a_R(\{c\} \cup C)$ is the sum of the R^2 values from the s_c-1 linear models that explain the columns of $\mathbf{X}_c$ using a full model of the factors in C.

Lemma 2 (Grömping and Xu 2014).

Consider a BA(N, $s_1 \dots s_n$, R–1) and an R factor set $\{c\} \cup C \subseteq \{1,\dots,n\}$. Denote by $\mathbf{X}_c$ the $N \times (s_c-1)$ main effects model matrix in arbitrary coding for the factor c, by $\mathbf{F}_C$ the model matrix of a full model of the factors in C. Then, the projected a_R value $a_R(\{c\} \cup C)$ is the sum of the SCCs between $\mathbf{X}_c$ and $\mathbf{F}_C$.

Note that, under normalized orthogonal coding, all columns of the full model matrix in the lemmata can be omitted, except for the $R-1$ factor interaction matrix $\mathbf{X}_C$ for the factors in C. Lemma 2 makes use of canonical correlation analysis (CCA) (Hotelling 1936). Details on CCA can e.g. be found in Härdle and Simar (2003). In brief, CCA partitions the linear relation between an Nxp matrix $\mathbf{S}$ and an Nxq matrix $\mathbf{T}$ into uncorrelated pairs of linear combinations ($\mathbf{u}_i=\mathbf{S}\mathbf{a}_i$, $\mathbf{v}_i=\mathbf{T}\mathbf{b}_i$), $i=1,\dots,\min(p,q)$, such that ($\mathbf{u}_1$, $\mathbf{v}_1$) maximizes the correlation among all possible linear combinations, and the subsequent pairs ($\mathbf{u}_j$, $\mathbf{v}_j$) maximize the remaining correlation among all pairs that are uncorrelated to previous pairs. The i-th canonical correlation is the correlation of the pair ($\mathbf{u}_i$, $\mathbf{v}_i$). The canonical correlations are invariant to nonsingular affine transformations of the matrices $\mathbf{S}$ and $\mathbf{T}$, which implies that the SCCs provide a coding invariant way of partitioning the projected a_R values into s_c–1 contributions. In fact, they partition the overall R^2 obtained from modeling a factor's main effects df based on $R-1$ other factors in the most concentrated way that is obtainable from an orthogonal factor coding. Note that the SCCs of Lemma 2 are closely related to the canonical efficiencies introduced by James and Wilkinson (1971), which are used in the literature on incomplete block designs.

As the projected a_R values are very important for all that follows, their calculation for a resolution III 3 factor array (Table 4) according to the definition and the lemmata is pointed out in detail below:

Table 4: An OA(8, $2^2 4^1$, 2) (transposed)

A	1	1	1	1	2	2	2	2
B	1	1	2	2	1	1	2	2
C	1	3	2	4	4	2	3	1

Table 5: The full model matrix for the array of Table 4

$$
\left(\begin{array}{cccccccccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\
 & \mathrm{A} & \mathrm{B} & \mathrm{C}_1 & \mathrm{C}_2 & \mathrm{C}_3 & \mathrm{AB} & \mathrm{AC}_1 & \mathrm{AC}_2 & \mathrm{AC}_3 & \mathrm{BC}_1 & \mathrm{BC}_2 & \mathrm{BC}_3 & \mathrm{ABC}_1 & \mathrm{ABC}_2 & \mathrm{ABC}_3 \\
\hline
1 & -1 & -1 & -\sqrt{2} & -\sqrt{2/3} & -\sqrt{1/3} & 1 & \sqrt{2} & \sqrt{2/3} & \sqrt{1/3} & \sqrt{2} & \sqrt{2/3} & \sqrt{1/3} & -\sqrt{2} & -\sqrt{2/3} & -\sqrt{1/3} \\
1 & -1 & -1 & 0 & \sqrt{8/3} & -\sqrt{1/3} & 1 & 0 & -\sqrt{8/3} & \sqrt{1/3} & 0 & -\sqrt{8/3} & \sqrt{1/3} & 0 & \sqrt{8/3} & -\sqrt{1/3} \\
1 & -1 & 1 & \sqrt{2} & -\sqrt{2/3} & -\sqrt{1/3} & -1 & -\sqrt{2} & \sqrt{2/3} & \sqrt{1/3} & \sqrt{2} & -\sqrt{2/3} & -\sqrt{1/3} & -\sqrt{2} & \sqrt{2/3} & \sqrt{1/3} \\
1 & -1 & 1 & 0 & 0 & \sqrt{3} & -1 & 0 & 0 & -\sqrt{3} & 0 & 0 & \sqrt{3} & 0 & 0 & -\sqrt{3} \\
1 & 1 & -1 & 0 & 0 & \sqrt{3} & -1 & 0 & 0 & \sqrt{3} & 0 & 0 & -\sqrt{3} & 0 & 0 & -\sqrt{3} \\
1 & 1 & -1 & \sqrt{2} & -\sqrt{2/3} & -\sqrt{1/3} & -1 & \sqrt{2} & -\sqrt{2/3} & -\sqrt{1/3} & -\sqrt{2} & \sqrt{2/3} & \sqrt{1/3} & -\sqrt{2} & \sqrt{2/3} & \sqrt{1/3} \\
1 & 1 & 1 & 0 & \sqrt{8/3} & -\sqrt{1/3} & 1 & 0 & \sqrt{8/3} & -\sqrt{1/3} & 0 & \sqrt{8/3} & -\sqrt{1/3} & 0 & \sqrt{8/3} & -\sqrt{1/3} \\
1 & 1 & 1 & -\sqrt{2} & -\sqrt{2/3} & -\sqrt{1/3} & 1 & -\sqrt{2} & -\sqrt{2/3} & -\sqrt{1/3} & -\sqrt{2} & -\sqrt{2/3} & -\sqrt{1/3} & -\sqrt{2} & -\sqrt{2/3} & -\sqrt{1/3} \\
\hline
8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\sqrt{32} & \sqrt{32/3} & -\sqrt{4/3}-\sqrt{12}
\end{array}\right)
$$

Table 6: Ways to obtain A_3=1 for the array of Table 4

The LHS columns are the responses for a regression model or the **S** variables for CCA, the RHS columns are the explanatory columns for a regression model or the **T** variables for CCA; columns in parentheses can be omitted in case of normalized orthogonal contrasts, like in Table 5

Applying the definition of A_3

Sum of squared column averages of $\mathbf{M}_3$	Model matrix columns
(i.e. ½ + 1/6 + 1/3)	14 to 16

Applying R^2 sums or sums of SCCs

LHS	LHS columns (**S**)	RHS columns (**T**)	Metric
A	2	(1, 3 to 6,)	single R^2 value (1) or
(2 levels)		11 to 13	single SCC (1)
B	3	(1, 2, 4 to 6,)	single R^2 value (1) or
(2 levels)		8 to 10	single SCC (1)
C	4 to 6	(1 to 3,)	sum of three R^2 values (1/2, 1/6 and 1/3) or
(4 levels)		7	sum of three SCCs (1, 0, 0)

Table 5 shows the full model matrix for the following normalized orthogonal coding for the main effects: factors A and B are coded in -1/+1 coding (1 coded with -1), factor C in normalized Helmert coding. All columns of the matrix have squared length N=8, and the main effects model matrix columns are orthogonal to each other, implying uncorrelated estimation of main effect coefficients in the absence of 2-factor interactions. The bottom row shows the column sums of the model matrix. Table 6 shows the different ways in which the number of generalized words of length 3 ($A_3 = a_3(1,2,3) = 1$) can then be obtained, referencing the model matrix column numbers provided in Table 5.

3.2. Statistical interpretations and connections to other criteria

We now consider the R-factor set $\{c\}\cup$C and denote by $\mathbf{X}_c$ the main effects model matrix of factor c in s_c levels, by $\mathbf{F}_{\mathrm{C}} = (\mathbf{F}_{\mathrm{C\ lower}} \vdots \mathbf{X}_{\mathrm{C}})$ the full model matrix of the factors in C, by $\mathbf{X}_{\mathrm{C}}$ the part of $\mathbf{F}_{\mathrm{C}}$ that relates to the $R-1$ factor interaction, and by $a_R(\{c\}\cup\mathrm{C})$ the projected a_R value when emphasizing that factor c is singled out for the main effects role. Furthermore, we assume that all factors are coded in normalized orthogonal coding. The orthogonality assumption ensures that confounding between $\mathbf{F}_{\mathrm{C}}$ and $\mathbf{X}_{\mathrm{c}}$ reduces to confounding between $\mathbf{X}_{\mathrm{C}}$ and $\mathbf{X}_{\mathrm{c}}$, which is unavoidable in case of resolution R; the normalization assumption allows convenient technical simplifications, which are exploited below. In addition, we denote by R_i^2 the R^2 from regressing the ith column $\mathbf{x}_{ci}$ of $\mathbf{X}_c$ on $\mathbf{F}_{\mathrm{C}}$, which is equivalent to regressing on $\mathbf{X}_{\mathrm{C}}$ because of resolution and coding. Then, the projected a_R value of the factor set is the sum of these R_i^2s. Furthermore, with an argument similar to the proof in the appendix of Grömping and Xu (2014), $R_i^2 = \mathbf{x}_{ci}^{\mathrm{T}}\mathbf{X}_{\mathrm{C}}\mathbf{X}_{\mathrm{C}}^{\mathrm{T}}\mathbf{x}_{ci}/N^2$, and $\mathbf{X}_c^{\mathrm{T}}\mathbf{X}_{\mathrm{C}}\mathbf{X}_{\mathrm{C}}^{\mathrm{T}}\mathbf{X}_c/N^2$ is the diagonal matrix of the R_i^2s; the trace of this matrix thus coincides with the projected a_R value $a_R(\{c\}\cup\mathrm{C})$; as a preparatory remark, this trace is exactly the squared Frobenius norm of $\mathbf{X}_c^{\mathrm{T}}\mathbf{X}_{\mathrm{C}}/N$, i.e. $a_R(\{c\}\cup\mathrm{C}) = \|\mathbf{X}_c^{\mathrm{T}}\mathbf{X}_{\mathrm{C}}/N\|_{\mathrm{F}}^2$. In the model

$$Y = \beta_0 + \mathbf{X}_c\boldsymbol{\beta}_c + \mathbf{F}_{\mathrm{C}}\boldsymbol{\beta}_{\mathrm{C\ all}} + \varepsilon = \beta_0 + \mathbf{X}_c\boldsymbol{\beta}_c + \mathbf{F}_{\mathrm{C\ lower}}\boldsymbol{\beta}_{\mathrm{C\ lower}} + \mathbf{X}_{\mathrm{C}}\boldsymbol{\beta}_{\mathrm{C}} + \varepsilon, \tag{3}$$

$\mathbf{X}_c^{\mathrm{T}}\mathbf{X}_{\mathrm{C}}\boldsymbol{\beta}_{\mathrm{C}}/N$ is the bias vector of the main effect coefficients for factor c, if the interaction portion $\mathbf{X}_{\mathrm{C}}$ is omitted from the model. The above-mentioned squared Frobenius norm $\|\mathbf{X}_c^{\mathrm{T}}\mathbf{X}_{\mathrm{C}}/N\|_{\mathrm{F}}^2$ and thus the projected a_R value provides an upper bound for the squared Euclidean norm of this bias vector: $\|\mathbf{X}_c^{\mathrm{T}}\mathbf{X}_{\mathrm{C}}\boldsymbol{\beta}_{\mathrm{C}}/N\|_2^2 \le \|\mathbf{X}_c^{\mathrm{T}}\mathbf{X}_{\mathrm{C}}/N\|_{\mathrm{F}}^2\ \|\boldsymbol{\beta}_{\mathrm{C}}\|_2^2$. An exact upper bound for a specific coding would require application of the squared Euclidean induced norm, which is given by the largest eigen value of $\mathbf{X}_c^{\mathrm{T}}\mathbf{X}_{\mathrm{C}}\mathbf{X}_{\mathrm{C}}^{\mathrm{T}}\mathbf{X}_c/N^2$, i.e. the largest individual R_i^2, and is thus coding dependent. A coding invariant Euclidean norm based bound is given by the largest SCC, which is the largest individual R_i^2 in case the most extreme orthogonal coding is chosen. Those exact upper bounds are individual summands of the conservative Frobenius norm based bounds. For a more realistic assessment of the *typical* bias risk for factor c's coefficients from the projection $\{c\}\cup$C, the average R^2 or average SCC, i.e. $a_R(\{c\}\cup\mathrm{C})/(s_c-1)$, appears reasonable. This average for the typical bias and the largest SCC for worst case consideration supplement each other for obtaining an overall assessment of the bias risk from omitting $\mathbf{X}_\mathrm{C}$. Note that, within a mixed level projection, singling out different factors for the role of c yields different average values: for example, in a resolution III 3-factor set with two 2-level factors and one 4-level factor, for the 2-level factors, a sum of "1" is also an average of "1" and implies that the factors are completely aliased in the set, whereas a sum of "1" for a 4-level factor is an average of 1/3 and means less severe aliasing for that factor. Regardless of fixed or mixed level projections, singling out different factors may also yield different maximum SCCs.

The R_i^2 values can also be used to indicate variance inflation of the ith coefficient estimator b_i for $\boldsymbol{\beta}_{ci}$ by including $\mathbf{X}_\mathrm{C}$ in model (3): $1/(1-R_i^2) = \mathrm{VIF}_i$ is the variance inflation factor. Thus, large R_i^2 values are harmful not only in terms of bias risk from omitting an active $R-1$ factor interaction but also in terms of losing precision for estimating $\boldsymbol{\beta}_{ci}$ when including the $R-1$ factor interaction of the factors in C. Again, the most extreme and coding invariant case is characterized by the largest SCC between $\mathbf{X}_c$ and $\mathbf{X}_\mathrm{C}$; if this is "1", there is a rank defect, and at least one coefficient is not estimable in model (3). The quantity

$$\left(1-\frac{a_R\left(\{c\}\cup\mathrm{C}\right)}{s_c-1}\right)^{-1} \tag{4}$$

is the harmonic mean of the VIF_i for estimation of the $\boldsymbol{\beta}_{ci}$, $i=1,\ldots,s_c-1$ in model (3). Note that the harmonic mean is the only one of the popular means that can include infinity values into an average without becoming infinite itself; it only becomes infinite if *all* averaged values are infinite, which in the present application allows to distinguish between complete aliasing of all main effects df and complete aliasing of some but not all main effects df. Thus, (4) is a reasonable measure for the average loss of precision for factor c main effects coefficients from confounding in model (3).

For R=2, the projected $a_2(u,v)$ value equals an Nth of $\chi^2(u,v)$ as defined in (2) (for the proof, see the supplementary material). Consequently, minimization of A_2 is equivalent to minimizing the ave(χ^2)

criterion, and GMA can be used as a refinement in case of ties. This insight underlines the qualitative character of GMA. For R>2, $a_R(c,\mathrm{C})$ within the R-factor set $\{c\}\cup\mathrm{C}$ can still be seen as an Nth of the $\chi^2(c,\mathrm{C})$ value, where the latter is obtained from a two-dimensional contingency table of factor c vs. the $s_\mathrm{C} = \prod_{j\in\mathrm{C}} s_j$ level cross-product factor for all factors in C. Furthermore, the average R^2 value for a factor with $s_{\min} = \min(s_u, s_v)$ levels in a resolution II pair u,v coincides with the square of Cramér's V, which is $\chi^2(u,v)/(N(s_{\min}-1))$. The analogous result holds for factor c with $s_c \le s_\mathrm{C}$. Thus, the average R^2 value can also be interpreted as a well-known measure of qualitative association.

As mentioned before, the projected a_R values uniquely determine the average R^2 value for any factor with s_c levels, but can come with different decompositions into SCCs. The statistical meaning of the latter is worth a detailed example. The supplementary materials provide a further resolution III example in addition to the resolution II example given below.

Example 2: There are two types of GMA BA(8, 4^2, 1) with $A_2 = a_2(1,2) = 1$. An instance of each is shown in Table 7. Of course, both average R^2 values are 1/3 for both arrays, while the SCCs differ: in the concentrated case, the one generalized word of length 3 can be concentrated on a single df for each factor in the main effect role ($\{c\}\cup\mathrm{C}$ either $\{1\}\cup\{2\}$ or $\{2\}\cup\{1\}$): here, the contrast of levels 1,2 vs. 3,4 for factor A is completely confounded with that same contrast for factor B; thus, in both cases, there is one SCC of "1" and two zeroes (implying in total two ones and four zeroes, when simply combining both variants). In the more even array, the one generalized word of length 3 is as evenly distributed over the df as possible for a BA: in both decompositions, there are two SCCs 0.5 and one 0. Contrary to the concentrated array, the array with the more even distribution of SCCs partially confounds two contrasts of each factor, and there is no coding for which it would be possible to concentrate the entire confounding on one df; consequently, when used for an experimental design in the two factors, the more even array allows estimation of all main effect coefficients, whereas for the concentrated array one of the main effect coefficients is not estimable. As was mentioned before, in model (3) an SCC of "1" in the $\{c\}\cup\mathrm{C}$ decomposition implies a rank defect; this is undesirable for any resolution, but particularly detrimental for resolution II arrays, for which a rank defect in model (3) implies a pair of factors for which main effects coefficients are not estimable even if only these two factors are active.

Table 7: Two GMA BA(8, 4^2, 1) and their SCCs

The table lists the levels of factor B that occur with each level of factor A.

even					concentrated				
A	**1**	**2**	**3**	**4**	**A**	**1**	**2**	**3**	**4**
B	2,4	1,3	3,4	1,2	**B**	1,2	1,2	3,4	3,4
SCCs: two times 0, four times 0.5					SCCs: four times 0, two times 1				

3.3. Coding invariant mixed level friendly refinements of PFTs

PFTs simply tabulate the $a_R(\{u_1,\ldots,u_R\})$ for all R-factor sets. The previous section pointed out several properties of the projected a_R values, if decomposed into R_i^2 or SCCs considering the set $\{u_1,\ldots,u_R\}=\{c\}\cup\mathrm{C}$. The various interpretations for the average R^2 or average SCC show that division by the df leads to a natural metric that does more justice to mixed level arrays than the simple $a_R(\{c\}\cup\mathrm{C})$. The conceptually cleanest approach to handle the factor specific $a_R(\{c\}\cup\mathrm{C}) / (s_c - 1)$ within a projection is to refrain from aggregating over the different choices for c in $\{c\}\cup\mathrm{C}$: average R^2 frequency tables (ARFTs) tabulate the $a_R(\{c\}\cup\mathrm{C})/(s_c - 1)$ values for each choice of c, i.e. a total of $R\binom{n}{R}$ average R^2 values are tabulated. If a single value per projection is insisted upon, it is possible to average the average R^2 values over the different singled out c, which leads to projection average R^2 tables (PARFTs). One version of such PARFTs will be formally defined in Section 4, while two others are only touched upon.

The previous section also pointed out the relation of SCCs to estimability and bias risk, and in particular the severe consequences of SCCs “1”. The SCCs are the single entities that allow coding invariant assessment of individual df confounding for main effects. Therefore, SCC frequency tables (SCFTs) are the third proposed quality criterion, which should be brought into play in case of comparable performance regarding ARFT (or PARFT). One might also think of tabulating individual df R^2 values, possibly in relation also to only a selection of individual df also for the right-hand side of the model; this would be the route of choice, would one want to decompose the b_k values considered by Tsai and Gilmour (2010), or any other criterion relating to only a part of the s–1 df of an s-level factor's main effect. As such quantities would necessarily be coding dependent for $s > 2$, one would have to decide on a coding and apply this tabulation for an array / coding combination; it might also be possible to apply a different weight to different df of a factor, e.g., linear, quadratic or cubic. This is, however, beyond the scope of this paper. The three new types of tables and the quality criteria based on them are defined and exemplified in Section 4.

4. The new criteria

The following definitions will be demonstrated with the worked example of the 8 run array of Table 4, before applying them all to the arrays from Table 1 and to further examples.

4.1. Average R^2 frequency tables (ARFTs)

In this section, the unit of tabulation is the factor-projection combination.

Definition 2:

(i) For a BA(N, $s_1,\ldots,s_n$, R–1), the average R^2 frequency table (ARFT$_R$) is the frequency table of the $R\binom{n}{R}$ values $a_R\left(u_1,\ldots,u_R\right)\big/\left(s_{u_i}-1\right)$, $\{u_1,\ldots,u_R\} \subseteq \{1,\ldots,n\}$, i=1,…,R.

(ii) Minimum average R^2 aberration ranks arrays according to their ARFT$_R$, in complete analogy to minimum projection aberration.

For the worked example, n=R=3, $a_3(1,2,3)$=1, s_1=s_2=2, s_3=4, so that ARFT$_3$ from the only projection is a table of the three values 1/1, 1/1 and 1/3, i.e.

Average R^2	1/3	1
frequency	1	2

The interpretation of this table is straightforward: For two factor-projection combinations the average R^2 of the main effects model matrix columns is "1", i.e. main effects of the respective factors in the only 3-factor projection are completely aliased; for one factor-projection combination, the average R^2 is 1/3, i.e. the respective factor (the 4-level factor) is partially aliased in the only 3-factor projection.

Grömping and Xu (2014) defined generalized resolution (GR) as a generalized version of Deng and Tang's (1999) definition. In terms of average R^2 values, their definition can be written as $GR = R+1-\sqrt{\text{ave}R^2_{\text{worst}}}$, i.e. the next larger resolution is reduced by the square root of the worst case average R^2 value. Consequently, GR can be obtained from ARFT$_R$ by subtracting the square root of the largest ARFT$_R$ value from R+1. In the worked example, GR is thus 3+1–1=3.

4.2. Projection average R^2 frequency tables (PARFTs)

In this section, the unit of tabulation is the projection again, like for PFT. Now, a decision for a weighting approach is needed in order to aggregate the several average R^2 values into one number for each projection. It appears natural to obtain an unweighted average of the factor wise average R^2 values for each R-factor projection. PARFT$_R$ tabulates these averages:

Definition 4:

(i) For a BA(N, $s_1,\dots,s_n$, R–1), the projection average R^2 frequency table (PARFT$_R$) is the frequency table of the $\binom{n}{R}$ values $a_R\left(u_1,\dots,u_R\right)\frac{1}{R}\sum_{i=1}^{R}\frac{1}{s_{u_i}-1}$, $\{u_1,\dots,u_R\} \subseteq \{1,\dots,n\}$.

(ii) The respective minimum projection average R^2 aberration ranks arrays according to their PARFT$_R$ in complete analogy to minimum projection aberration.

For the worked example, n=R=3, so that there is only one projection. The multiplier for $a_3(1,2,3)$=1 is the average of the inverse factor df's $1/(s_{ui}–1)$ (i.e., (1+1+1/3)/3=7/9). Thus, the array of Table 4 has a PARFT$_3$ with the frequency "1" for the only value "7/9". For only 2-level factors, PARFT would use unmodified projected a_R values, for only 4-level factors, PARFT would divide the projected a_R value by 3, and for triples with one 2-level and two 4-level factor, the multiplier would be 5/9.

It would also be possible to average all individual df R^2 values within each projection, without prior averaging per factor (while the individual df R^2 values are coding dependent, their average is not). This would imply weighting $a_R(u_1,\dots,u_R)$ with $R/(s_{u_1}+\dots+s_{u_R})$; these weights would be only driven by the overall number of df in a projection, while PARFT$_R$ from the definition also focuses on the distribution of the df over the factors. For many practically relevant situations, a df-based weighting behaves almost the same as the PARFT$_R$ from the definition; there are big differences in case of a few factors with many levels, where the behavior of PARFT$_R$ from the definition seems more adequate. A further and even more extreme approach weights each projection proportionally to the inverse *product* of level numbers; the result would be closely related to the E(f_{NOD}) criterion for supersaturated designs. These alternative weightings have not been pursued.

4.3. Squared canonical correlation frequency tables (SCFTs)

PARFT$_R$ aggregated several entries of ARFT$_R$ into one projection wise entry, SCFT does the opposite: ARFT$_R$ did not differentiate between situations for which the average "1/3" is the result from e.g. three main effects columns each of which has an R^2 value of 1/3 or from one column with an R^2 of 1 and two columns with an R^2 of 0. This does matter for the SCFTs considered in this section: Instead of the factor – projection combination considered by ARFTs, SCFTs consider the df – projection combination as the unit of tabulation. As motivated in Section 3, a more even distribution of SCCs is more favorable for screening experiments than a more concentrated one, see Table 7. As mentioned before, the SCCs provide the R^2 values for individual df that one obtains with the worst case orthogonal coding, where "worst case" means that the sum $a_R(\{c\}\cup C)$ of individual R^2 values is distributed over individual main effects df of factor c as unequally as possible.

<u>Definition 3</u>:

(i) For a BA(N, $s_1,\ldots,s_n$, R–1), the squared canonical correlation frequency table (SCFT$_R$) is the frequency table of the $\sum_{i=1}^{n}(s_i-1)\binom{n-1}{R-1}$ SCCs between the main effects model matrix $\mathbf{F}_c$ for a factor $c \in \{u_1,\ldots,u_R\} \subseteq \{1,\ldots,n\}$ and the model matrix $\mathbf{X}_C$ of the full model in the factors of $C = \{u_1,\ldots,u_R\} \setminus \{c\}$.

(ii) Minimum SCC aberration ranks arrays according to their SCFT$_R$, in complete analogy to minimum projection aberration.

For the worked example, the single SCC from each 2-level factor's main effects column in the role of $\mathbf{X}_c$ is 1 (has to be equal to the R^2), and the canonical correlations with the 4-level factor main effects matrix in the role of $\mathbf{X}_c$ are a 1 and two zeroes, as was discussed in the beginning of this section. The table thus shows three ones and two zeroes:

SCC	0	1
frequency	2	3

In Section 4.1, we saw that ARFT_R is related to the generalized resolution GR. SCFT_R is related to a different type of generalized resolution, GR_{ind}: Grömping and Xu (2014) introduced GR_{ind} as the stricter version of generalized resolution that reacts to the most severe aliasing in individual main effects df. Thus, $GR_{\text{ind}} = R + 1 - CC_{\max}$ with $CC_{\max}$ denoting the largest canonical correlation occurring in any R-factor projection. GR_{ind} can thus be obtained from SCFT_R by subtracting the square root of the largest value from R+1, i.e. in the same way in which GR can be obtained from ARFT_R.

SCFT_R has a relation to array regularity; the concentrated array of Table 7 is a regular array in the following sense: with appropriate coding, all its df are either completely aliased or independent. It can be shown that all regular arrays in this sense have SCFT_R with values "0" and "1" only (completely aliased df are reflected by SCCs of "1", independent effects by SCCs of "0"). Grömping and Bailey (2016) introduced three new regularity definitions, two of which were derived from SCFT and ARFT. As was already mentioned, regular arrays often have undesirable projection properties, which is also seen here, as – for a given number of words of length R – the "0"-"1" type SCFT_R are necessarily worst; resolution II arrays should even be strictly non-regular in the sense of avoiding any ones among the SCCs, which can be guaranteed by keeping the stricter version of generalized resolution, GR_{ind}, larger than 2. Catalogued supersaturated arrays usually fulfill strict non-regularity; on the other hand, unfortunately, for setups where regular arrays are possible, catalogued OAs are often regular (e.g. many of the arrays in the Kuhfeld 2009 catalogue).

4.4. Usage recommendations

A simple use case for projection based criteria is their finer gridding as compared to the GWLP, which makes them useful for quickly ruling out equivalence between many designs whose criteria differ. With the methods given in Xu and Wu (2001), the GWLP can be determined with good computational efficiency. For cases with identical GWLP, further criteria can be used for ruling out equivalence, leaving only very few arrays to be checked with computationally very demanding equivalence searches. The examples in the next section will show that there is no single criterion that outperforms the others for all sets of arrays. SCFT is finer than the others, in that it allows differences even for fixed level arrays where all the other criteria coincide with PFTs; for mixed level arrays, PARFT and ARFT also deviate from PFTs because of different handling of the different projection types. Within a set of only regular arrays, SCFT is a weak discriminator, because it can only produce 0/1 patterns.

Another important application is the quality assessment of arrays, which is needed both for selecting an array for particular experimental situations and for ranking catalogues of arrays. Generally, it is most desirable to rank arrays according to their behavior w.r.t. the most severe confounding, i.e. w.r.t. confounding from R-factor projections in resolution R arrays. That was already the rationale of the proposal to look at PFTs, and the mixed level approaches follow the same logic. ARFTs, PARFTs and SCFTs are more suitable for mixed level arrays than PFTs and less complicated than WZ for complex array structures (see e.g. Example 3); also, WZ is not always perceived as appropriate with its stepwise approach (e.g. Wu and Zhang's own criticism of the ranking of d_2 and d_3 in Example 1). As was mentioned before, the author prefers ARFT over PARFT due to its clean concept that does not require arbitrary weight decisions within projections. The following artificial example explains another advantage of ARFT over PARFT: suppose a six factor array of three 2-level and three 4-level factors, for which $a_3(1,2,3) = a_3(4,5,6) = 1$, while all other projected a_3 values are zero. With ARFT, each average R^2 enters the assessment at face value, which implies three average R^2 values of 1 (for the 2-level factors) and three average R^2 values of 1/3 (for the 4-level factors). To the contrary, PARFT depends on the distribution of the factors over the two sets: if the first three factors are 2-level, PARFT contains a "1" from the first set and a "1/3" from the second set. If there is one 4-level factor among the first three factors, PARFT contains a "7/9" for the first set and a "5/9" for the second set. The author considers the behavior of ARFT as more desirable, not least because it reveals that there are completely confounded main effects in the array.

As SCFT concentrates on the detail within each factor and assumes a worst case factor parameterization, it is considered as a secondary criterion, after using ARFT (or, if preferred in spite of Section 3.2 in conjunction with the above reasoning, PARFT) as the primary one. The examples of the next section will demonstrate the criteria in action in various smaller sets of arrays, and in one very large group of arrays.

If ARFT_R and SCFT_R cannot distinguish between arrays, higher dimensions can be considered. As the new criteria are not reasonably defined for higher dimensions, GWLP and PFT have to be used in most cases; in some situations, an array has factor sets of higher resolution than R. For these, it also makes sense to consider their ARFT and SCFT (e.g., Table 3 of Grömping and Bailey 2016). Another reasonable approach is to use A_R or even GMA as the primary criterion and to only resolve ties from this ranking with ARFT and SCFT, and possibly higher dimensional PFT.

5. Examples

Example 1 demonstrates the calculation of the new criteria in detail, comparing them also to the previously calculated ones; it shows that the ARFT criterion is superior to the WZ criterion in ranking the arrays considered by Wu and Zhang (1993), while PARFT gives the ranking obtained by Wu and Zhang. Example 2 illustrated the benefit of considering SCFT for two resolution II arrays that are identical w.r.t. all other criteria; a resolution III and a resolution IV example of the same nature, for which SCFT provides a ranking of designs that are otherwise of the same quality can be found in the supplemental materials. Example 3 illustrates a complex application for the WZ criterion to regular mixed level arrays; as all arrays are regular, SCFT does not discriminate among them, while WZ, ARFT and PARFT do. Example 4 shows a set of non-regular mixed level arrays for which PFT shows a very diverse pattern of projected a_3 values, which yields the same ranking that would be obtained by the recommended strategy. The final Example 5 shows the power of SCFT for detecting non-equivalence of arrays.

Table 8: The new criteria for the six non-isomorphic OA(16, $2^3 4^2$, 2) of Table 1

	WZ_3				PFT_3				$ARFT_3$						$SCFT_3$				$PARFT_3$					
	A_{30}	A_{31}	A_{32}	Rank	0	1/2	1	Rank	0	1/6	1/3	1/2	1	Rank	0	1/2	1	Rank	0	7/18	5/9	7/9	1	Rank
PARFT weight	*1*	*7/9*	*5/9*																					
1 (d_3)	0	2	3	4	5	0	5	6	15	0	8	0	7	6	39	0	15	6	5	0	3	2	0	4
2 (d_1)	0	1	3	1	6	0	4	2	18	0	7	0	5	2	42	0	12	4	6	0	3	1	0	2
3	0	1	3	1	5	2	3	1	15	2	6	4	3	1	35	14	5	1	5	2	3	0	0	1
4 (d_2)	1	0	3	5	6	0	4	2	18	0	6	0	6	4	42	0	12	4	6	0	3	0	1	5
5	0	1	3	1	6	0	4	2	18	0	7	0	5	2	38	8	8	2	6	0	3	1	0	2
6	1	0	3	5	6	0	4	2	18	0	6	0	6	4	38	8	8	2	6	0	3	0	1	5

Example 1 revisited: Table 8 shows all criteria regarding 3-factor projections of the six arrays from Table 1, including those already presented in Tables 2 and 3. Before interpreting the criteria, we will consider the calculation of the new criteria from WZ_3 and PFT_3 and technical relations among the criteria: First of all, the sum of the WZ_3 entries is (of course) A_3 (5 for array 1, 4 for the other arrays). For PFT_3, the sum of the products between frequencies and values is A_3, the same is true for a third of the sum of the products between frequencies and values for $SCFT_3$. The sum of the frequencies themselves is $3\binom{5}{3}=30$ for $ARFT_3$, $9\binom{4}{2}=54$ for $SCFT_3$ and $\binom{5}{3}=10$ for $PARFT_3$ (like for PFT_3). For most arrays in the table, a projection has either 0 or 1 words of length 3. $ARFT_3$ and $PARFT_3$ can be worked out from WZ_3, as each type of projection has a specific composition in terms of numbers of levels and a specific PARFT weight: For (2,4,4) projections, the PARFT weight is the average of 1, 1/3 and 1/3, i.e. 5/9. Thus, in $PARFT_3$, the "1" values from such projections become "5/9"; analogously, the "1" values from (2,2,4) projections become "7/9". The remaining projections (up to the overall total of 10) contribute the "0" entries. For array 3, there are two projections with projected a_3 values "½". These are from (2,2,4) triples (not obvious from the table, found out by inspection), i.e. the "0.5" has to be multiplied with the PARFT weight "7/9", which yields the $PARFT_3$ value "7/18" with frequency 2 for array 3. For obtaining $ARFT_3$ from WZ_3 and PFT_3, note that any 2-level factor in the projection simply contributes the number of words of the projection as average R^2, while any 4-level factor contributes a third of that number. Consequently, apart from the zeroes, there are the ARFT values "1" and "1/3" for most arrays, and the additional "1/2" and "1/6" for array 3. Let us conclude this technical portion with the detailed derivation of $ARFT_3$ for array 3: A_{32}=3 comes from three (2,4,4) projections with one word each and translates into the entry "3" for the average R^2 "1" from the single 2-level factor and the entry "6" for the average R^2 "1/3" from the two 4-level factors in these projections; A_{31}=1 comes from two (2,2,4) projections with projected a_3 values of "1/2" and translates into the entry "4" for the value "1/2" from the two 2-level factors and the entry "2" for the value "1/6" from the single 4-level factor in these projections. The further tables in this section will still report enough detail for such cross-comparisons, but the detail will not be spelled out at such length.

Turning to interpretation of Table 8, all criteria agree that the best array is array 3, like it was with

PFT$_3$ and WZ (for WZ tied with arrays 2 and 5): the array has the fewest factor-projection combinations with complete aliasing (average R^2 of 1), it has only 5 df-projection combinations with SCC 1, and the worst projection average R^2 for this array is smallest. All criteria also single out array 1 (=d_3), which is worst under PFT$_3$, ARFT$_3$ and SCFT$_3$, but not worst for PARFT$_3$ and WZ$_3$. The four arrays that were equivalent under PFT$_3$ are divided into groups of two by the other criteria: ARFT, PARFT and WZ agree in the group division and in the ranking between the two groups (d_1 before d_2), SCFT creates a different grouping and ranks both d_1 and d_2 together (and as worse than the other two). When following the recommended ranking approach (ARFT and then SCFT or GMA, then ARFT (or PARFT or even WZ), and then SCFT), a reasonable unique ranking is achieved (ranks 6,3,1,5,2,4 from top to bottom in Table 8).

Table 9: Five minimum A_3 OA(64, $2^4 4^3 8^1$, 2) as projections from a regular OA(64, $2^5 4^{10} 8^4$, 2) (A_{3ij} refers to words of length 3 from projections with i 4-level and j 8-level factors; PFT$_3$: seven ones, 49 zeroes)

	WZ$_3$								ARFT$_3$					PARFT$_3$						
	A_{300}	A_{310}	A_{301}	A_{320}	A_{311}	A_{330}	A_{321}	rank	0	.143	.333	1	rank	0	.270	.333	.492	.556	.714	rank
PARFT weight	*1*	*.778*	*.714*	*.556*	*.492*	*.333*	*.270*													
1	0	0	0	0	4	0	3	2	147	7	10	4	2	49	3	0	4	0	0	2
2	0	0	1	1	2	0	3	5	147	6	10	5	5	49	3	0	2	1	1	5
3	0	0	0	0	3	1	3	1	147	6	12	3	1	49	3	1	3	0	0	1
4	0	0	0	0	4	0	3	2	147	7	10	4	2	49	3	0	4	0	0	2
5	0	0	0	1	3	0	3	4	147	6	11	4	4	49	3	0	3	1	0	4

Example 3: This example illustrates a complex application of the WZ criterion: we consider the five OA(64, $2^4 4^3 8^1$, 2) that can be obtained as projections of the OA(64, $2^5 4^{10} 8^4$, 2) of Kuhfeld (2009) and have the minimum number of length 3 words, which is A_3=7. All five arrays are regular, and all have the same PFT$_3$ with seven ones and 49 zeroes. They also have the same SCFT$_3$ with 21 ones and 399 zeroes. The WZ$_3$, ARFT$_3$ and PARFT$_3$ patterns are different, however (see Table 9). The projection types for the Wu and Zhang assessment have been ordered as (2,2,2), (2,2,4), (2,2,8),

(2,4,4), (2,4,8), (4,4,4), (4,4,8) from most to least serious (a ranking that only considers the overall number of df in a projection would deviate from this order). Note the increased complexity from having more 4-level factors and an additional 8-level factor in the array. According to the pattern of length 3 words of different types, array 3 is best, followed by tied arrays 1 and 4. Array 2 is worst. $ARFT_3$ and $PARFT_3$ arrive at the same ranking as WZ_3 for this example, while PFT_3 and $SCFT_3$ consider all arrays as equally good, as was mentioned before.

Table 10: PFT_3 for the 14 non-isomorphic OA(24, $2^{11}4^{1}6^{1}$, 2)

				0	1/9	1/3	4/9	5/9	2/3	1
			projection types	*(2,2,2)* *(2,2,6)*	*(2,2,2)* *(2,2,4)*	*(2,2,6)*	*(2,2,2)*	*(2,2,6)*	*(2,2,6)*	*(2,4,6)* *(2,2,4)*
ID	Rank	Rank with PFT_4	*PARFT weights*	*1* *0.7333*	*1* *0.7778*	*0.7333*	*1*	*0.7333*	*0.7333*	*0.5111* *0.7778*
1	13	13		140	90	28	0	0	16	12
2	1	1		152	73	28	4	2	16	11
3	1	3		152	73	28	4	2	16	11
4	13	13		140	90	28	0	0	16	12
5	1	3		152	73	28	4	2	16	11
6	1	1		152	73	28	4	2	16	11
7	12	12		136	90	36	0	0	12	12
8	5	6		156	73	20	4	2	20	11
9	5	7		156	73	20	4	2	20	11
10	5	10		156	73	20	4	2	20	11
11	5	7		156	73	20	4	2	20	11
12	5	7		156	73	20	4	2	20	11
13	5	11		156	73	20	4	2	20	11
14	5	5		156	73	20	4	2	20	11

Example 4. This example investigates the 14 non-isomorphic OA(24, $2^{11}4^{1}6^{1}$, 2) from Eendebak and Schoen (2013). All 14 arrays have the same GWLP and WZ_3 patterns: $(A_3,\ldots,A_{13})$=(42, 103, 245.33,

333.33, 484, 436.33, 218.67, 141.33, 34, 9, 0), (A_{300}, A_{310}, A_{301}, A_{311}) = (4,7,20,11), where A_{3ij} refers to words of length 3 from projections with i 4-level and j 6-level factors. PFT_3 (see Table 10) shows a much more spread-out pattern for these arrays than for the ones considered in the previous examples, and $ARFT_3$ and $PARFT_3$ (not shown) are even more spread out, because particular values of PFT_3 imply different average R^2 for different factors in a projection ($ARFT_3$) or have different weights according to different projection types ($PARFT_3$). However, neither ARFT nor PARFT nor SCFT refine the ranking or distinguish more arrays for this example. Some additional refinement can be achieved by including PFT_4 (see Table 10); there remain some tied arrays, however.

Table 11: Number of classes distinguished for OA(32,4^a,2), a=3,4,6,7,8,9

a	No. of arrays	GWLP	PFT_3	$SCFT_3$	PFT_3 & $SCFT_3$	PFT_3 & PFT_4	PFT_3, $SCFT_3$ & PFT_4	PFT_3 & GWLP	PFT_3, PFT_4 & GWLP	$SCFT_3$ & GWLP	$SCFT_3$, PFT_3 & GWLP	all four criteria
3	44	12	12	40	40	12	40	12	12	40	40	40
4	32983	51	211	11324	11339	287	11748	287	287	11733	11748	11748
5	108339											
6	31779	171	2705	26038	26482	12242	28017	4037	12242	26546	26875	28017
7	6564	114	869	5332	5400	3056	5747	1218	3056	5375	5429	5747
8	283	7	76	211	212	192	241	76	270	211	212	241
9	20	1	12	15	15	15	15	12	15	15	15	15

Example 5: The final example illustrates the use of SCFT for ruling out design equivalence, even in large sets of fixed level arrays. PFTs have been previously used for this purpose. Table 11 shows the numbers of classes distinguished by various metric combinations for the tractable series of non-isomorphic OA(32,4^a,2). For these symmetric arrays, ARFT and PARFT are equivalent to PFT. Clearly, $SCFT_3$ is the key contributor to discriminating the arrays of Table 11.

6. Discussion

This paper introduced three new metrics for assessing the quality of BAs. It has been argued that the projection frequency tables introduced previously (Xu, Cheng and Wu 2004) should be refined for ranking mixed level arrays, in order to better reflect statistical implications from confounding in terms of bias and imprecision of estimates. The primary metric for comparing R-factor projections of resolution R arrays should be ARFT_R, the average R^2 table over all factors in R-factor projections. This table has an entry for each factor in each of its R-factor projections and thus avoids the need to obtain an overall assessment of a mixed-level projection. The SCC frequency table, SCFT_R, considers the distribution of R^2 values from regressing individual main effect df on full models in $R-1$ other factors, given the factor is coded in the worst possible orthogonal way in terms of concentrating all the confounding on a few df. For regular arrays, SCFT_R has the values "0" and "1" only, implying complete or no confounding for all main effects df in worst case coding. For R=2, strict non-regularity (no "1" SCCs) is crucial for estimability of main effect coefficients in at least all pairs of factors. The third metric that was introduced here, PARFT_R, averages the R ARFT_R contributions for each projection before tabulation; as PARFT_R may seem a natural alternative or even preferable to some readers, this metric has been included here, although the author prefers ARFT_R because of its greater conceptual clarity. It has been recommended to rank arrays by ARFT_R and resolve ties by SCFT_R, possibly preceding the entire process by a ranking in terms of A_R or even GMA.

Wu and Zhang (1993) previously proposed separate consideration of numbers of words from different types of projections and introduced the criterion "type 0 MA". They treated arrays with two- and four-level factors, for at most two four-level factors, and they provided a few optimal arrays under their criterion. Example 3 (Table 9) showed that a generalized version of WZ's approach gets intricate with more different levels and more factors per level. Thus, it is not surprising that the WZ type 0 MA approach has not entered statistical practice in any breadth. Average R^2 frequency tables (ARFTs) share the advantages of the Wu and Zhang method without carrying its burden of complexity, and ties from ranking by ARFT_R can potentially be resolved by SCFT_R.

SCFT_R also yields an additional possibility for assessing array equivalence: equivalent arrays must have the same SCFT_R; Table 11 shows that SCFT_R is able to discriminate large sets of non-isomorphic arrays into very many equivalence classes, which substantially reduces the burden of isomorphism checking. However, there are also cases for which all SCFTs are the same, while other criteria discriminate between arrays; this particularly happens in sets for which many or all arrays are regular.

Section 3.3 briefly touched upon the work of Tsai and Gilmour (2010), who discussed a Q_B criterion that they considered a bridge between alphabetic optimality and aberration criteria; they showed that aberration criteria can be obtained as the limits of Q_B for certain sequences of prior probabilities assigned to effects being active in an assumed maximal model. It might be of interest to investigate, whether this idea can be brought to bear for the aberration-related frequency tables of this paper.

Acknowledgements

Eric Schoen contributed orthogonal arrays. Eric Schoen and Hongquan Xu were competent discussion partners regarding the appropriate way of quantifying confounding for mixed level OAs. Hongquan Xu gave valuable comments on the manuscript. Rosemary Bailey drew my attention to the work by James and Wilkinson. Two anonymous referees prompted substantial improvements to the manuscript's flow and instigated inclusion of supersaturated designs. Deutsche Forschungsgemeinschaft Grant GR 3843/1-1 supported the initial phase of this work.

References

Ai, M., Fang, K.-T. and He, S. (2007). E(χ^2)-optimal mixed-level supersaturated designs. *Journal of Statistical Planning and Inference*, 137, 306-316.

Booth, K.H.V. and Cox, D.R. (1962). Some systematic supersaturated designs. *Technometrics*, 4, 489-495.

Cheng, S.W. and Wu, C.F.J. (2001). Factor Screening and Response Surface Exploration (with discussion), *Statistica Sinica*, 11, 553–604.

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.

Eendebak, P.T. and Schoen, E.D. (2013). Complete series of non-isomorphic orthogonal arrays. URL: http://pietereendebak.nl/oapage/ Accessed May 24, 2013.

Evangelaras, H., Koukouvinos, C. and Lappas, E. (2007). 18-run nonisomorphic three level orthogonal arrays. *Metrika*, 66, 31-37.

Evangelaras, H., Koukouvinos, C. and Lappas, E. (2011). 27-run nonisomorphic three level orthogonal arrays: Identification, evaluation and projection properties. *Utilitas Mathematica*, 84, 75-87.

Fang, K.-T., Lin, D.K.J. and Liu, M.Q. (2003). Optimal mixed-level supersaturated design. *Metrika*, 58, 279-291.

Georgiou, S. (2014). Supersaturated designs: A review of their construction and analysis. *Journal of Statistical Planning and Inference*, 144, 92-109.

Grömping, U. (2011). Relative projection frequency tables for orthogonal arrays. Report 1/2011, Reports in Mathematics, Physics and Chemistry, Department II, Beuth University of Applied Sciences Berlin.

Grömping, U. (2013). Frequency tables for the coding invariant ranking of orthogonal arrays. Report 2/2013, Reports in Mathematics, Physics and Chemistry, Department II, Beuth University of Applied Sciences Berlin.

Grömping, U. (2016). The DoE.base Package (Full factorials, orthogonal arrays and base utilities for DoE packages). R package version 0.28. In: R Core Team (2016).

Grömping, U. and Bailey, R.A. (2016). Regular fractions of factorial arrays. In *mODa 11 – Advances in Model-Oriented Design and Analysis*, Kunert, J., Müller, C. and Atkinson, A. (eds.) Springer International Publishing Switzerland, pp. 143-151.

Grömping, U. and Xu, H. (2014). Generalized resolution for orthogonal arrays. *The Annals of Statistics*, 42, 918-939.

Gupta, V.K., Parsad, R., Bhar, L.M. and Kole, B. (2008). Supersaturated Designs for Asymmetrical Factorial Experiments. *Journal of Statistical Theory and Practice*, 2, 95-108.

Härdle, W. and Simar, L. (2003). *Applied multivariate statistical analysis*. Springer, New York, NY.

Hedayat, S., Sloane, N. and Stufken, J. (1999). *Orthogonal Arrays: Theory and Applications*. Springer, New York, NY.

Hotelling, H. (1936). Relations Between Two Sets of Variates. *Biometrika*, 28, 321–377.

James, A.T. and Wilkinson, G.N. (1971). Factorization of the residual operator and canonical decomposition of nonorthogonal factors in the analysis of variance. *Biometrika*, 58, 279-294.

Kuhfeld, W. (2009). Orthogonal Arrays. Website courtesy of SAS Institute http://support.sas.com/techsup/technote/ts723.html.

R Core Team (2016). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

Schoen, E.D. (2009). All orthogonal arrays with 18 runs. *Quality and Reliability Engineering International*, 25, 467-480.

Schoen, E.D., Eendebak, P.T. and Nguyen, M.V.M. (2010). Complete enumeration of pure-level and mixed-level orthogonal arrays. *Journal of Combinatorial Designs*, 18, 123-140.

Stufken, J. and Tang, B. (2007). Complete enumeration of two-level orthogonal arrays of strength d with $d+2$ constraints. *Annals of Statistics*, 35, 793-814.

Taguchi, G. (1987). *System of Experimental Design, Vol. 1 and 2*. Unipub/Kraus, White Plains, New York, NY.

Tsai, P.-W. and Gilmour, S.G. (2010). A General Criterion for Factorial Designs under Model Uncertainty. *Technometrics*, 52, 231-242.

Vasilev, N., Schmitz, Ch., Grömping, U., Fischer, R. and Schillberg, S. (2014). Assessment of Cultivation Factors that Affect Biomass and Geraniol Production in Transgenic Tobacco Cell Suspension Cultures. *PLoS ONE*, 9(8): e104620 (7 pages).

Wu, C.F.J. and Zhang, R. (1993). Minimum aberration designs with 2-level and 4-level factors. *Biometrika*, 80, 203-209.

Xu, H. (2015). Nonregular factorial and supersaturated designs. In *Handbook of Design and Analysis of Experiments*, Dean, A., Morris, M., Stufken, J. and Bingham, D. (eds) Chapman and Hall/CRC, 339-370.

Xu, H., Cheng, S.-W. and Wu, C.F.J. (2004). Optimal Projective Three-Level Designs for Factor Screening and Interaction Detection. *Technometrics*, 46, 280-292.

Xu, H. and Deng, L.Y. (2005). Moment Aberration Projection for Nonregular Fractional Factorial Designs. *Technometrics*, 47, 121-131.

Xu, H. and Wu, C.F.J. (2001). Generalized Minimum Aberration for Asymmetrical Fractional Factorial Designs. *Annals of Statistics*, 29, 1066-1077.

Yamada, S. and Matsui, T. (2002). Optimality of Mixed-Level Supersaturated Designs. *Journal of Statistical Planning and Inference*, 104, 459-468.