Fachbereich II – Mathematik - Physik - Chemie

BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

02/2013

Ulrike Grömping

Frequency tables for the coding invariant ranking of orthogonal arrays

Häufigkeitstabellen für das kodierungsinvariante Ranking orthogonaler Felder (englischsprachig)

Reports in Mathematics, Physics and Chemistry

Berichte aus der Mathematik, Physik und Chemie

ISSN (print): 2190-3913

Reports in Mathematics, Physics and Chemistry

Berichte aus der Mathematik, Physik und Chemie

The reports are freely available via the Internet:
http://www1.beuth-hochschule.de/FB_II/reports/welcome.htm

02/2013, December 2013

Frequency tables for the coding invariant ranking of orthogonal arrays
Häufigkeitstabellen für das kodierungsinvariante Ranking orthogonaler Felder
(englischsprachig)

Update history

August 18 2015: Inserted "squared" on p.5, line 13 from bottom

Editorial notice / Impressum

Published by / Herausgeber:
Fachbereich II
Beuth Hochschule für Technik Berlin
Luxemburger Str. 10
D-13353 Berlin
Internet: http://public.beuth-hochschule.de/FB_II/
E-Mail: fbiireports@beuth-hochschule.de

Responsibility for the content rests with the author(s) of the reports.
Die inhaltliche Verantwortung liegt bei den Autor/inn/en der Berichte.

ISSN (print): 2190-3913

Frequency tables for the coding invariant ranking of orthogonal arrays

Ulrike Grömping, Beuth University of Applied Sciences Berlin, Germany

Abstract

Ranking of orthogonal arrays, particularly mixed level orthogonal arrays, is a non-trivial task. Existing methods 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 as proposed by Xu, Cheng and Wu for pure three-level arrays and used e.g. by Schoen for mixed level 18-run arrays. Based on recent insights by Grömping and Xu into the interpretation of the projection frequencies, this report proposes three new types of frequency tables for ranking orthogonal arrays. These are coding invariant, which is particularly important for designs with qualitative factors. The proposed tables are used in the same way as the existing projection frequency tables, but 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 report compares the proposed tables to existing ones based on various examples and recommends the use of two of the three proposals.

Key words: Average R^2 Frequency Tables, Squared Canonical Correlation Frequency Tables, Projection Average R^2 Frequency Tables, Generalized Resolution

1. Introduction

Experimental design is an important tool for gaining as much information as possible from a limited number of experimental runs. One way of designing an experiment is the use of an orthogonal array. This report discusses a particular quality aspect of orthogonal arrays and the experimental designs based on them. Before going into the specifics, some terminology is provided: Orthogonal arrays (OAs) 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. The symbols that occur in a column are called levels; OAs with the same number of levels in all columns are called fixed or pure level arrays. If different columns may have different numbers of levels, an OA is called a mixed level array. The orthogonality of an OA is constituted by the fact that each column contains each of its levels the same number of times (orthogonal to the overall mean), and each pair of columns contains each pair of levels the same number of times (pairwise orthogonality). An OA is said to be of strength t, if each k-factor projection, $k \le t$ contains each k-tuple of levels the same number of times. In the statistical literature, strength t is often denoted as resolution $R=t+1$. Obviously, OAs have at least strength 2, equivalent to resolution III per definition.

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. 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. This leads to a need for mixed level arrays. Thus, mixed level OAs are often used in the screening phase of experiments with qualitative factors. As their construction is by no means straightforward for the practitioner, some collections of orthogonal arrays are available in literature, web and software, e.g. Taguchi (1987), Hedayat, Sloane and Stufken (1999), Kuhfeld (2009), Eendebak and Schoen (2013), Grömping (2013). 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, Schoen 2009, Schoen, Eendebak and Nguyen 2010, Evangelaras, Koukouvinos and Lappas 2011). Of course, such catalogues are useful only if there are criteria for choosing designs from them.

Quality criteria 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 (Xu, Cheng and Wu 2004). While both these criteria have been applied to mixed level designs (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). This report aims at providing tractable quality criteria for orthogonal arrays with qualitative factors that fairly treat mixed level designs. The recent work by Grömping and Xu (2013) will be useful in obtaining three new and conceptually convincing ways to replace minimum projection aberration by other criteria that take care of mixed level arrays in a coding-invariant way and such that WZ's concerns are also addressed. One of the criteria will also prove useful for ranking pure level arrays.

Section 2 presents the existing quality criteria (GWLP, GMA, minimum projection aberration based on projection frequency tables, and WZs proposal), points out in more detail why they leave a gap to be bridged, and motivates the new criteria. Section 3 presents the new criteria and their relation to some existing concepts and provides several examples for their performance. Section 4 discusses ranking of designs and non-equivalence detection as applications of the new metrics. The report closes with a discussion.

The following notation will be used: the letter s stands for the number of factor levels. An orthogonal array of resolution R = strength $R-1$ in N runs with n factors will be denoted as OA(N, $s_1,\ldots,s_n$, $R-1$), with s_1, …, s_n, possibly but not necessarily distinct, or as OA(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). The unsquared letter R always refers to the resolution of a design, while R^2 denotes the coefficient of determination. k-factor projections are often denoted by an index set $\{u_1,…,u_k\}$; R factor projections are of particular interest for this report.

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

In brief, the GWLP with entries A_3, A_4,… quantifies the amount of confounding among sets of three, four, … factors, and the resolution of the design is the number R for which $A_R > 0$, but $A_k = 0$ for all $k < R$. The GMA criterion ranks designs by minimizing the GWLP entries from left to right, which automatically maximizes the resolution. More formally, the generalized word length pattern (GWLP) of an OA(N, $s_1,…,s_n$, R–1) (Xu and Wu 2001) is most easily defined using the model matrix $\mathbf{M} = (\mathbf{M}_0, \mathbf{M}_1, \mathbf{M}_2, …, \mathbf{M}_n)$ 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. $\mathbf{M}_k = (\mathbf{X}_{1…k}, …, \mathbf{X}_{n-k+1…n})$, with $\mathbf{X}_{u1…uk}$ the $N\times((s_{u1}-1)\bullet…\bullet(s_{uk}-1))$ model matrix of the interaction among factors $\{u_1,…,u_k\}$ obtained as element-wise products of one column from each of the k main effects model matrices. The elements $A_0,A_1,A_2,A_3,…$ of the GWLP can be calculated as the sums of squared column averages of the respective portions of $\mathbf{M}$, i.e. $A_k=\mathbf{1}^\mathrm{T}\mathbf{M}_k\mathbf{M}_k^\mathrm{T}\mathbf{1}/N^2$. That sum can be split into contributions from the separate k-factor projections, i.e.

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

the $a_k(u_1,…,u_k)$ are called projection frequencies and are the basis of minimum projection aberration, as presented in the next section. The coding of matrix $\mathbf{M}_1$ (orthogonal, squared column length normalized to N) will be called "normalized orthogonal coding" in the following.

2.2. Projection frequencies

The following two lemmata state the two results from Grömping and Xu (2013) that are most important for this paper.

Lemma 1 (Grömping and Xu 2013).

Consider an OA(N, $s_1 \ldots s_n$, $R-1$), $c \in \{u_1,\ldots,u_R\} \subseteq \{1,\ldots,n\}$, and C = $\{u_1,\ldots,u_R\} \setminus \{c\}$. Denote by $\mathbf{X}_c$ the N x (s_c-1) main effects model matrix in orthogonal coding for the factor c. Then, the projection frequency $a_R(u_1,\ldots,u_R)$ 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.

Note that orthogonal coding encompasses orthogonality to the intercept column, i.e. the classical dummy coding, although yielding main effects model columns that are orthogonal to each other, is not considered to be orthogonal coding. Furthermore, note that the individual R^2 values depend on the particular choice of orthogonal coding, while the sum of the R^2 values is independent of that choice.

Lemma 2 (Grömping and Xu 2013).

Consider an OA(N, $s_1 \ldots s_n$, $R-1$), $c \in \{u_1,\ldots,u_R\} \subseteq \{1,\ldots,n\}$, and C = $\{u_1,\ldots,u_R\} \setminus \{c\}$. Denote by $\mathbf{X}_c$ the N x (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, up to the interaction of degree $R-1$, except for the intercept column. Then, the projection frequency $a_R(u_1,\ldots,u_R)$ is the sum of the squared canonical correlations between $\mathbf{X}_c$ and $\mathbf{F}_C$.

Remark: Under normalized orthogonal coding, all columns of the full model matrix in Lemma 2 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 (Hotelling 1936). Details on canonical correlation analysis can e.g. be found in Härdle and Simar (2003). In brief, canonical correlation analysis partitions the linear relation between an Nxp matrix $\mathbf{X}$ and an Nxq matrix $\mathbf{Y}$ into uncorrelated pairs of linear combinations ($\mathbf{u}_i$=$\mathbf{X}\mathbf{a}_i$, $\mathbf{v}_i$=$\mathbf{X}\mathbf{b}_i$), i=1,…,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 $\mathbf{X}$ and $\mathbf{Y}$, which implies that the squared canonical correlations provide a coding invariant way of partitioning the projection frequencies. 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.

2.3. Minimum projection aberration

For resolution R designs, the minimum projection aberration criterion primarily looks at the projection frequencies $a_R(u_1,\ldots,u_R)$ of the different R factor projections and ranks design d_2 as better than design d_1, if d_2 has fewer projections with high projection frequencies. The tool for assessing this criterion are the so-called projection frequency tables (PFTs), as defined below:

<u>Definition 1</u>

(i) For an OA(N, $s_1,\ldots,s_n$, $R-1$), the projection frequency table (PFT$_k$, $k \geq R$, Xu, Cheng and Wu 2004 for k=R) is defined as the frequency table of the $\binom{n}{k}$ values $a_R(u_1,\ldots,u_k)$, $\{u_1,u_2,\ldots,u_k\} \subseteq \{1,\ldots,n\}$.

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

Table 1: The five GMA OA(16, $2^3 4^2$, 2) and design 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

<u>Example 1</u>: We consider the five minimum aberration OA(16, $2^3 4^2$, 2) plus a sixth OA(16, $2^3 4^2$, 2) that was investigated by Wu and Zhang 1993 together with two of the minimum aberration designs. The designs themselves are given in Table 1. The designs marked by d_1, d_2 and d_3 are isomorphic to the ones that have been investigated by Wu and Zhang 1993.

Table 2: GWLP and PFT3 for the designs of Table 1

	GWLP				PFT$_3$			
	Rank	A_3	A_4	A_5	Rank	0	0.5	1
1 (d_3)	5	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

Table 2 shows the GWLP and PFT_3 for the designs from Table 1. According to GMA, design 1 (d_3) of Table 1 is worst, the other five designs are equivalent. PFT_3 further distinguishes design 3 from the other GMA designs: it has only 3 instead of 4 triples with 1 word of length 3 and is therefore considered better. Among the three Wu and Zhang designs, d_1 and d_2 are equivalent, while d_3 is worse.

2.4. The case for modifications to PFTs for mixed level designs

Xu, Cheng and Wu (2004) introduced PFT for pure 3-level designs only. For mixed level designs, the projection frequencies correspond to R factor projections of different patterns of numbers of factor levels. This issue was raised by WZ for mixed 2-level and 4-level designs, and by Grömping (2011) in general. A look at Lemma 1 clarifies the reason behind these concerns. The degree of confounding that the factor c faces in the projection $\{u_1,\ldots,u_R\}$ is reasonably measured by the average of the R^2 values for the main effects columns of the factor, not by their sum: the average R^2 value is $a_R(u_1,\ldots,u_R)/(s_c - 1)$. The larger the average R^2 for a factor in an R factor projection, the higher the percentage of the variation in the factor's main effects model matrix that is explained by the other factors, and the higher the confounding risk for this factor's main effect within the projection. For mixed level designs, the same $a_R(u_1,\ldots,u_R)$ value from different projections may imply a different degree of confounding; likewise, within the same projection, main effects from factors with different numbers of levels have different confounding severities. For example, in a resolution III 3-factor projection 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 projection, whereas a sum of "1" for a 4-level factor is an average of 1/3 and means much less severe aliasing for that factor. Based on this insight, this report will provide three different ways to tabulate the confounding risk from an orthogonal array: average R^2 frequency tables (ARFT) tabulate the average R^2 values for each factor within each projection, i.e. a total of $R\binom{n}{R}$ average R^2 values are tabulated. Projection wise average R^2 frequency tables (PARFT) average the R R^2 values tabulated in ARFT within each projection (PARFT). Furthermore, a finer degree of detail is introduced by tabulating a value for each individual degree of freedom. While one might think of tabulating individual df R^2 values, this is generally not a good idea, because these are coding dependent, which is inadequate for qualitative factors. A coding invariant individual df approach can be obtained by tabulating the squared canonical correlations instead (see Lemma 2), which leads to squared canonical correlation frequency tables (SCFTs). All three types of tables and the quality criteria based on them are defined and exemplified in Section 3, after investigating WZ's approach in the following section.

2.5. Wu and Zhang (1993)

WZ proposed to treat different types of projections in mixed level designs differently. Their solution was very specific to the designs they studied: designs 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 designs or even designs 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 projections of only 2-level factors (i.e. zero 4-level factors), k factor projections with one 4-level factor, and k factor projections 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 projection. 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, 2^34^2, 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. The ranking approach proposed here will rectify this undesired ranking.

Table 3: The Wu and Zhang patterns for the designs 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 designs, 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 design; all their designs are therefore regular. Design 3 of Example 1 is not of that nature, but nevertheless shows the same WZ pattern as the best WZ design. For the 16 run cases, it has been investigated whether there are better designs according to the WZ criterion within the larger set of all the non-isomorphic designs with the respective numbers of factors and levels. (Attention was restricted to the 16 run designs, because there is an easily manageable number of them, whereas the 32 runs designs have not even been enumerated by Schoen and Eendebak 2010.) It was found that the WZ Type 0 MA designs remain best in the overall set of designs 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 designs. This is not surprising because regular designs have repeatedly been found to have undesirable projection properties, e.g. by Xu and Deng (2005).

An obvious generalization of WZ's method for designs 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 projection. 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 designs 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 designs into consideration or to take weighted sums of the A_{ij}. The proposals of Section 3 take up the idea of weighted sums, however not on the overall GWLP level but for projections instead. The weights arise very naturally from the perspective of averaging R^2 values that became available through the work of Grömping and Xu (2013). This approach also works for designs with more than two different numbers of levels, for which generalization of the WZ approach would be 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 3.4 contains two examples 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. These underline the complexity involved in WZ's approach.

3. The new criteria

The following definitions will be demonstrated with a worked example of an 8 run design with two 2-level factors and one 4-level factor (see Table 4), before applying them all to the designs from Table 1.

Table 4: An OA(8, $4^1 2^2$, 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 model matrix up to 2-factor interactions for the design of Table 4

1	2	3	4	5	6	7	8	9	10	11	12	13
	A	B	C_1	C_2	C_3	AB	AC_1	AC_2	AC_3	BC_1	BC_2	BC_3
1	−1	−1	$-\sqrt{2}$	$-\sqrt{2/3}$	$-\sqrt{1/3}$	1	$\sqrt{2}$	$\sqrt{2/3}$	$\sqrt{1/3}$	$\sqrt{2}$	$\sqrt{2/3}$	$\sqrt{1/3}$
1	−1	−1	0	$\sqrt{8/3}$	$-\sqrt{1/3}$	1	0	$-\sqrt{8/3}$	$\sqrt{1/3}$	0	$-\sqrt{8/3}$	$\sqrt{1/3}$
1	−1	1	$\sqrt{2}$	$-\sqrt{2/3}$	$-\sqrt{1/3}$	−1	$-\sqrt{2}$	$\sqrt{2/3}$	$\sqrt{1/3}$	$\sqrt{2}$	$-\sqrt{2/3}$	$-\sqrt{1/3}$
1	−1	1	0	0	$\sqrt{3}$	−1	0	0	$-\sqrt{3}$	0	0	$\sqrt{3}$
1	1	−1	0	0	$\sqrt{3}$	−1	0	0	$\sqrt{3}$	0	0	$-\sqrt{3}$
1	1	−1	$\sqrt{2}$	$-\sqrt{2/3}$	$-\sqrt{1/3}$	−1	$\sqrt{2}$	$-\sqrt{2/3}$	$-\sqrt{1/3}$	$-\sqrt{2}$	$\sqrt{2/3}$	$\sqrt{1/3}$
1	1	1	0	$\sqrt{8/3}$	$-\sqrt{1/3}$	1	0	$\sqrt{8/3}$	$-\sqrt{1/3}$	0	$\sqrt{8/3}$	$-\sqrt{1/3}$
1	1	1	$-\sqrt{2}$	$-\sqrt{2/3}$	$-\sqrt{1/3}$	1	$-\sqrt{2}$	$-\sqrt{2/3}$	$-\sqrt{1/3}$	$-\sqrt{2}$	$-\sqrt{2/3}$	$-\sqrt{1/3}$

Table 5 shows the model matrix with interactions up to degree 2 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 effects in the absence of 2-factor interactions. $A_3 = a_3(1,2,3) = 1$ can then be obtained

- from the matrix of three-factor interactions (not shown) by the definition,
- as the single R^2 value from regressing any of the 2-level columns on the full model matrix in the other two factors (e.g. column 2 on columns 1, 3 to 6, 11 to 13); with the coding of Table 5, the right-hand side can omit all main effects columns, i.e. columns 11 to 13 suffice for the right-hand side.
- as the sum of the three R^2 values from regressing the main effects columns of the 4-level factor (columns 4 to 6) on the full model in the two-level factors (columns 1 to 3 and 7); these are 0.5, 1/6 and 1/3, respectively. Again, the coding of Table 5 allows to restrict the right-hand side to column 7.
- as the only squared canonical correlation from column 2 vs the full model matrix in factor B and C without intercept (columns 3 to 6 and 11 to 13; restrictable to columns 11 to 13)
- as the only squared canonical correlation from column 3 vs the full model matrix in factor A and C without intercept (columns 2, 4 to 6 and 8 to 10, restrictable to columns 8 to 10)
- as the sum of the three squared canonical correlations from columns 4 to 6 vs the full model matrix in factors A and B without intercept (columns 2, 3, 7, restrictable to column 7). The squared canonical correlations are 1, 0, 0 with the three columns 2, 3 and 7 on the right-hand side, and 1 with only column 7 on the right-hand side. This latter case illustrates that it is possible to have fewer canonical correlations than the left-hand side factor has df. If that happens, the squared canonical correlations will be supplemented with the appropriate number of zeroes, in order to have always a squared canonical correlation for each df.

3.1. Average R^2 frequency tables (ARFTs)

According to Lemma 1, $a_R(u_1,\ldots,u_R)$ is a sum of R^2 values when regressing the main effects model matrix $\mathbf{X}_c$ on the full model matrix of the other $R-1$ factors in the projection indexed by $\{u_1,\ldots,u_R\}$, and it was discussed before, that the average R^2 value $a_R(u_1,\ldots,u_R)/(s_c - 1)$ is a reasonable measure for the degree of aliasing for factor c. In the above example, the average R^2 value is 100% for the two 2-level factors, but 33.3% only for the 4-level factor (of course, both 2-level factors and generally all factors with the same number of levels in the same projection have the same average R^2 value). The idea of the average R^2 frequency tables is to simply tabulate these individual averages for all factor projection combinations from R factor projections, so that the sum of the frequencies is $R\binom{n}{R}$:

Definition 2:

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

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

For the worked example, $n=R=3$, $a_3(1,2,2)=1$, $s_1=s_2=2$, $s_3=4$, 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 respective R-factor projections are completely aliased; for one factor-projection combination, the average R^2 is 1/3, i.e. the respective factor is partially aliased in the respective R-factor projection.

Grömping and Xu (2013) 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. Consequently, GR can be obtained from ARFT$_R$ by subtracting the square root of the largest table header from R+1. As the largest header for the worked example is "1", the example design has GR=3.

3.2. Squared canonical correlation frequency tables (SCFTs)

ARFT$_R$ does 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 2.4, individual main effect df should be considered based on the coding-invariant squared canonical correlations from considering the main effects model matrix of a particular factor on the Y side, the full model matrix of $R-1$ other factors on the X side of a canonical correlation analysis. The squared canonical correlations provide the R^2 values for individual df that one obtains with the worst case coding, where "worst case" means that the sum $a_R(u_1,\dots,u_R)$ of individual R^2 values is distributed over individual main effects df of the Y side factor as unequally as possible. SCFT$_R$ tabulates a value for each main effects df within each factor projection combination:

Definition 3:

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

(ii) Minimum squared canonical correlation aberration ranks designs according to their SCFT$_R$, in complete analogy to minimum projection aberration.

For the worked example, the single squared canonical correlation 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:

Squared canonical correlation	0	1
frequency	2	3

The benefit of a good ranking in terms of minimum squared canonical correlation aberration is less obvious than that of minimum average R^2 aberration. Therefore, we consider an additional example:

Example 2: There are 44 non-isomorphic OA(32, 4^3, 2), 10 of which are GMA (1 word of length 3). The best and worst designs and their SCFT$_3$ tables are given in Table 6. In the worst case, the one word of length 3 can be concentrated on a single df for all three factors in the only 3-factor projection. This implies the pattern of three ones and six zeroes that is observed for the worst design of Table 6. In the best case, the one word of length 3 is as evenly distributed over the df as possible for an OA (there is no design with 9 squared canonical correlations of 1/3 each).

Table 6: The best and worst GMA OA(32, 4^3, 2) and their SCFT$_3$

SCFT$_3$

	0	0.25	0.375	1
best	0	3	6	0
worst	6	0	0	3

Best design (number 9 among the GMA designs, transposed)

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

Worst design (number 1 among the GMA designs, transposed)

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

Table 7 shows the frequency distribution of the best and the worst design; besides the fact that the worst design has only 16 distinct runs, while the best design has 26 distinct runs, the critical characteristic to which $SCFT_3$ has responded is the very systematic pattern of factor level combinations attained in the worst design: each (1,2) vs. (3,4) contrast is completely aliased with the 2-factor interaction of the other (1,2) vs. (3,4) contrasts. This is most easily seen for the relation of factor C to the AB interaction: C in (1,2) only occurs with either both A and B in (1,2) or with both A and B in (3,4), while C in (3,4) only occurs for the other two cases. The worst design has such a completely confounded df for each of its factors, while the best design does not display any such pattern.

Table 7: Combination frequencies for the best and worst design of Table 6

			Best					Worst			
		C	1	2	3	4		1	2	3	4
A	B										
1	1		2	0	0	0		2	0	0	0
	2		0	2	0	0		0	2	0	0
	3		0	0	2	0		0	0	2	0
	4		0	0	0	2		0	0	0	2
2	1		0	2	0	0		0	2	0	0
	2		0	0	1	1		2	0	0	0
	3		1	0	0	1		0	0	0	2
	4		1	0	1	0		0	0	2	0
3	1		0	0	1	1		0	0	2	0
	2		2	0	0	0		0	0	0	2
	3		0	1	0	1		2	0	0	0
	4		0	1	1	0		0	2	0	0
4	1		0	0	1	1		0	0	0	2
	2		0	0	1	1		0	0	2	0
	3		1	1	0	0		0	2	0	0
	4		1	1	0	0		2	0	0	0

In the previous section, we saw that $ARFT_R$ is related to the generalized resolution GR. Straightforward considerations imply that $GR=R$ comes with at least one squared canonical correlation of "1". The reverse is not true, though. $SCFT_R$ is related to a different type of generalized resolution, GR_{ind}: Grömping and Xu (2013) introduced GR_{ind} as the stricter version of generalized resolution that reacts to the most severe aliasing in individual main effects df. Thus, $GR_{\text{ind}} = R+1-\max r_{1;R}$ with $r_{1;R}$ 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 table header from R+1, i.e. in the same way in which GR can be obtained from $ARFT_R$.

$SCFT_R$ has a relation to another concept: design regularity. The worst design of Tables 6 and 7 is a regular design, as defined e.g. in Wu and Hamada (2009): all its df are either completely aliased or independent. It can be shown that all regular designs in this sense have $SCFT_R$ with headers "0" and

"1" only (completely aliased df are reflected by squared canonical correlations of "1", independent effects by squared canonical correlations of "0"). As was previously observed, regular designs often have undesirable projection properties (cf. e.g. Xu and Deng 2005). This is also seen here, as – for a given number of words of length R – the "0"-"1" type SCFT$_R$ are necessarily worst. Unfortunately, for setups where regular designs are possible, catalogued designs are often regular (e.g. many of the designs in the Kuhfeld 2009 catalogue).

3.3. 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 average of the R factor wise average R^2 for each R factor projection. PARFT$_R$ tabulates these averages:

Definition 4:

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

(ii) The respective minimum projection average R^2 aberration ranks designs 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 dfs $1/(s_{ui}-1)$ (i.e., (1+1+1/3)/3=7/9). Thus, the design in the worked example has a PARFT$_3$ with the only entry "1" for the header 7/9. For only 2-level factors, PARFT would use unmodified projection frequencies, for only 4-level factors, PARFT would divide the projection frequency 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 $a_R(u_1,\ldots,u_R)$ with $R/(s_{u_1}+\ldots+s_{u_R})$; 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. Therefore, the alternative weighting has not been pursued.

After defining the new criteria, the next section applies them to some example situations, starting with Example 1 that was considered above.

3.4. Examples

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 designs 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 design 1, 4 for the other designs). For PFT_3, the sum of the products of frequencies with headers is A_3, the same is true for a third of the sum of products of frequencies with headers 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). For most designs 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 design 3, there are two projections with ½ words of length 3. 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 $PARFT_3$ header "7/18" with frequency 2 for design 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 headers "1" and "1/3" for most designs, and the additional "1/2" and "1/6" for design 3. Let us conclude this technical portion with the detailed derivation of $ARFT_3$ for design 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 half a word each and translates into the entry "4" for the header "1/2" from the two 2-level factors and the entry "2" for the header "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 design is design 3, like it was with PFT_3 and WZ (for WZ tied with designs 2 and 5): the design has the fewest factor-projection combinations with complete aliasing (average R^2 of 1), it has only 5 df-projection combinations with squared canonical correlation 1, and the worst projection average R^2 for this design is smallest. All criteria also single out design 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 designs 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).

<u>Example 2 revisited</u>: The two OA(32, 4^3, 2) of Table 6 are fixed level designs with s=4 levels for each factor. Consequently, PFT_3, $PARFT_3$ and $ARFT_3$ are equivalent, and the Wu and Zhang approach is equivalent to GMA. Table 9 shows the quality criteria for these designs. PARFT headers and ARFT headers are $1/(s-1)$=1/3 times the PFT headers, and PARFT entries coincide with PFT entries, while ARFT entries are R times the respective PFT entries. As was discussed previously, $SCFT_3$ is the only criterion that adds independent information.

Table 9: The quality criteria for the designs from Table 6

	WZ_3	PFT_3	$ARFT_3$	$SCFT_3$				$PARFT_3$
	A_{33}	1	1/3	0	1/4	3/8	1	1/3
PARFT weight	*1/3*							
Best	1	1	3	0	3	6	0	1
Worst	1	1	3	6	0	0	3	1

<u>Example 3:</u> This example uses 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 designs 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 10). 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 design. According to the pattern of length 3 words of different types, design 3 is best, followed by tied designs 1 and 4. Design 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 designs as equally good, as was mentioned before.

Table 10: 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							rank	$ARFT_3$				rank	$PARFT_3$						rank
	A_{300}	A_{310}	A_{301}	A_{320}	A_{311}	A_{330}	A_{321}		0	.143	.333	1		0	.270	.333	.492	.556	.714	
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 4. This example considers the eleven non-isomorphic GMA OA(64, $4^3 2^2$, 3) that were obtained from Eendebak and Schoen (2013).

Table 11: $SCFT_4$ for the 11 GMA OA(64, $2^2 4^3$, 3) from Eendebak and Schoen (2013)

	WZ_4		PFT_4		$ARFT_4$			$PARFT_4$			$SCFT_4$				
	A_{42}	A_{43}	0	1	0	1/3	1	0	1/2		0	0.25	0.5	0.75	1
PARFT weight	*2/3*	*1/2*								Rank					
1	0	2	3	2	12	6	2	3	2	11	36	0	0	0	8
2	0	2	3	2	12	6	2	3	2	9	33	3	0	3	5
3	0	2	3	2	12	6	2	3	2	10	34	0	4	0	6
4	0	2	3	2	12	6	2	3	2	7	33	0	6	0	5
5	0	2	3	2	12	6	2	3	2	7	33	0	6	0	5
6	0	2	3	2	12	6	2	3	2	4	30	6	0	6	2
7	0	2	3	2	12	6	2	3	2	6	32	0	8	0	4
8	0	2	3	2	12	6	2	3	2	5	29	4	8	0	3
9	0	2	3	2	12	6	2	3	2	2	30	0	12	0	2
10	0	2	3	2	12	6	2	3	2	1	26	8	8	0	2
11	0	2	3	2	12	6	2	3	2	2	30	0	12	0	2

These mixed level strength 3 plans are of equal quality based on all criteria except for $SCFT_4$ (see Table 11): they have (A_4, A_5) = (2, 1), and one word of length 4 each comes from the two quadruples with one 2-level and three 4-level factors. Thus, each PFT_4 has two ones and three zeroes, $PARFT_4$ replaces each one with a 0.5 and keeps the zeroes unchanged, and each $ARFT_4$ has frequency "2" for the entry "1" and frequency "6" for the entry "1/3" from these quadruples, with 12 zeroes from the other three quadruples. $SCFT_4$ contains additional information: apart from two tied pairs, it uniquely ranks the 11 designs. For the first=worst design, it is possible to code the factors such that the one word of length 4 in the respective quadruple relates to a single df for all four factors in both quadruples. This is not the case for any other design. For the best four designs, only the 2-level factor within each quadruple has a completely confounded df (which cannot be avoided, of course).

Example 5. This example considers the four non-isomorphic GMA OA(64, $2^4 4^3$, 3) that were also obtained from Eendebak and Schoen (2013). Table 12 shows all criteria for these designs. WZ_4 considers all designs equally suitable, PFT_4, $ARFT_4$ and $PARFT_4$ distinguish between the first and the other three designs (first=worst, the only regular design among the four), and $SCFT_4$ can uniquely rank all four designs. The $SCFT_4$ entry "32" for the first design implies that each of the 8 4-factor projections with one word of length 4 can be coded such that there is a completely aliased df for each of the four factors in the projection. For the other three designs, the worst possible coding in terms of concentrating all the confounding on a particular df turns up 20, 18 or 16 completely aliased df only.

Table 12: Quality criteria for the four non-isomorphic GMA OA(64, $4^3 2^4$, 3)

		WZ_4				rank	PFT_4			rank	$ARFT_4$					rank	$PARFT_4$				rank	$SCFT_4$				rank
		A_{40}	A_{41}	A_{42}	A_{43}		0	1/2	1		1	1/6	1/3	1/2	1		0	1/3	1/2	2/3		0	1/4	1/2	1	
	$PARFT_4$ weight	*1*	*5/6*	*2/3*	*1/2*																					
1		0	0	4	4	1	27	0	8	4	108	0	20	0	12	4	27	0	4	4	4	228	0	0	32	4
2		0	0	4	4	1	25	4	6	1	100	8	16	8	8	1	25	4	4	2	1	216	0	24	20	3
3		0	0	4	4	1	25	4	6	1	100	8	16	8	8	1	25	4	4	2	1	214	0	28	18	2
4		0	0	4	4	1	25	4	6	1	100	8	16	8	8	1	25	4	4	2	1	204	16	24	16	1

Example 6. The final example investigates the 14 non-isomorphic OA(24, $2^{11}4^1 6^1$, 2) from Eendebak and Schoen (2013). All 14 designs 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 13) shows a much more diverse pattern for these designs than for the ones considered in the previous examples, and $ARFT_3$ and $PARFT_3$ (not shown) diversify even more, because particular headers 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, the ranking cannot be refined by including ARFT, PARFT or SCFT for this example.

Table 13: PFT_3 for the 14 non-isomorphic OA(24, $2^{11}4^16^1$, 2)

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

4. Applications

4.1. Ranking recommendations

Generally, it is most desirable to rank designs according to their behavior w.r.t. the most severe confounding, i.e. w.r.t. confounding from R factor projections in resolution R designs. That was the rationale of the proposal to look at PFTs; however, PFTs must be considered as problematic for mixed level designs. The other existing criterion, Wu and Zhang's approach, becomes quite complicated for complex design structures (see e.g. Example 3) and 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). The three new criteria from the previous section are possible alternatives: $ARFT_R$ tabulates an average R^2 value per factor-projection combination, $PARFT_R$ tabulates an average of average R^2 values for each projection, $SCFT_R$ tabulates a squared canonical correlation per combination of main effects degree of freedom with projection. As mentioned before, the approach taken by $SCFT_R$ tabulates the df wise R^2 values, provided the maximally concentrated factor parametrization has been used.

As $SCFT_R$ concentrates on the detail within each factor and assumes a worst case factor parametrization, it is considered as a secondary criterion, after using one of the other criteria as the primary one. The previous section has already illustrated the differences between the criteria. This section discusses which primary criterion should be used in the mixed level case (for fixed level designs, we use PFT_R, which is equivalent to the other two).

Table 8 above showed the new quality criteria for all designs from Table 1. In this example, $ARFT_3$ ranked the Wu and Zhang designs d_1, d_2 and d_3 the way Wu and Zhang (1993) would have liked their method to work, while $PARFT_3$ ranked design d_3 before d_2 like the WZ method does. What is the reason for this behavior? Put simply, $ARFT_R$ considers complete confounding for a particular 2-level factor as equally severe, regardless if the factor is confounded from an R-tuple with only 2-level factors in which the other 2-level factors are also completely confounded or if the factor is confounded from an R-tuple with e.g. $R-1$ 4-level factors for which the 4-level factors are partially confounded only. On the contrary, $PARFT_R$ downweights average R^2 values from projections with more df. Design d_2 has a completely confounded triple of 2-level factors, which accounts for three $ARFT_3$ entries of “1” each, and there are three further average R^2 values of “1” that originate from the triples of the three 2-level factors with both 4-level factors. Design d_3 has 7 R^2 values of “1”, but none of them comes from the triple with all 2-level factors; instead, two triples with one 4-level factor and two 2-level factors are affected. As a consequence from considering the A_{30} first for WZ_3 or from considering averages within a projection for $PARFT_3$, design d_2 takes on the worst case in WZ_3 and $PARFT_3$, while design d_3 does not. It is now a matter of judgment, which ranking behavior appears more appropriate. While one might argue that two-factor interactions are more likely to be strong if only 2-level factors are involved, there is no real evidence to back this up. Therefore, this report recommends $ARFT_R$ as the most suitable criterion for the primary ranking of designs. In most cases, $PARFT_R$ will not further distinguish ties from $ARFT_R$, so that $PARFT_R$ should not be considered. $SCFT_R$ can be quite helpful as a secondary criterion, as discussed above. If $ARFT_R$ and $SCFT_R$ cannot distinguish between designs, higher dimensions can be considered. As the new criteria are not defined for higher dimensions, GWLP and PFT have to be used.

4.2. Detection of non-equivalence

PFTs have been previously used for distinguishing more designs than possible with other criteria (e.g. Schoen 2009). PFTs, ARFTs and PARFTs are all based on the projection frequencies $a_R(u_1,\ldots,u_R)$, thus substantial additional discriminatory power cannot be expected from ARFTs and PARFTs. SCFTs, however, use an additional source of information, the canonical correlations. This implies that they can contribute an independent source for the assessment of non-isomorphism. The following example shows that SCFTs can add substantial discriminatory power: Table 14 shows the numbers of classes distinguished for the tractable series of non-isomorphic $OA(32,4^a,2)$. Clearly, $SCFT_3$ is the key contributor to discriminating among these.

Table 14: Number of classes distinguished for OA(32,4^a,2), a=3,4,6,7,8,9

a	No. designs	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

5. Discussion

This report introduced three new metrics for assessing the quality of orthogonal arrays. It has been argued that the projection frequency tables introduced previously (Xu, Cheng and Wu 2004) and used for both fixed and mixed level designs e.g. by Schoen (2009) should not be used for ranking mixed level designs, because they yield unfair comparisons of projections with different patterns of numbers of factor levels. The primary metric for comparing R factor projections of resolution R designs 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 new squared canonical correlation 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 designs, $SCFT_R$ has the headers "0" and "1" only, implying complete or no confounding for all main effects df in worst case coding. 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 does not recommend its use.

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 designs with two- and four-level factors, for at most two four-level factors, and they provided a few optimal designs under their criterion. A few other authors followed up on their proposal: Mukerjee and Wu (2001) generalized their approach to designs in n factors at s levels with one factor at s^r levels or one factor each at s^{r1} and s^{r2} levels. Ankenman (1999) studied two- and four-level designs, but disagreed with Wu and Zhang's distinction of different types of words and used the overall GWLP later defined in general by Xu and Wu (2001); he provided some minimum aberration designs according to his perspective. Wu and Hamada (2009) also provided some minimum aberration designs in two and four levels according to the Wu and Zhang approach. In this report, a generalized version of the Wu and Zhang approach was applied to mixed level designs for several of the examples, not restricting attention to designs with at most two factors at a larger number of levels, and neither to designs with all numbers of levels a power of the same s. However, Example 3 (Table 10) showed that 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 projection frequency tables (ARFTs) share the advantages of the Wu and Zhang method without carrying its burden of complexity. Ties from ranking by ARFT_R should be resolved by SCFT_R.

SCFT_R also yields an additional possibility for assessing design equivalence: equivalent designs must have the same SCFT_R; Table 14 shows that SCFT_R is able to discriminate large sets of non-isomorphic designs 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 designs.

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.

References

Ankenman, B.E. (1999). Design of Experiments with Two- and Four-Level Factors. *Journal of Quality Technology* **31**, 363-375.

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.

Grömping, U. (2013). The DoE.base Package (Full factorials, orthogonal arrays and base utilities for DoE packages). R package version 0.25-2. In: R Core Team (2013). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

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. and Xu, H. (2013). Generalized resolution for orthogonal arrays. Report 1/2013, Reports in Mathematics, Physics and Chemistry, Department II, Beuth University of Applied Sciences Berlin.

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

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

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

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

Mukerjee, R. and Wu, C.F.J. (2001). Minimum Aberration designs for mixed factorials in terms of complementary sets. *Statistica Sinica* **11**, 225-239.

R Core Team (2013). 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.
Wu, C.F.J. and Hamada, M. (2009, 2nd ed.). *Experiments. Planning, Analysis and Optimization*. Wiley, New York.
Wu, C.F.J. and Zhang, R. (1993). Minimum aberration designs with 2-level and 4-level factors. *Biometrika* **80**, 203-209.
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.