ARTICLE IN PRESS

Journal of Statistical Planning and Inference ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

# D-optimal two-level parallel-flats block designs with partial replication

Shin-Fu Tsai *, Chen-Tuo Liao

*Division of Biometry, Institute of Agronomy, National Taiwan University, Taipei 10617, Taiwan*

ARTICLE INFO

*Article history:*
Received 24 January 2017
Received in revised form 29 May 2017
Accepted 16 August 2017
Available online xxxx

*Keywords:*
Optimal design
Orthogonal blocking
Parallel-flats design
Screening experiment
Variance component

ABSTRACT

Under the assumption of random block effects, a new class of two-level factorial block designs with partial replication is proposed for estimating the user-specified requirement sets and variance components. A noteworthy feature of the proposed designs is that the within-block and between-block replicates are both conducted, such that the components of variance can be unbiasedly estimated. Under the framework of parallel-flats block designs, a set of sufficient conditions is presented for design characterization, and an algorithm is developed for systematically constructing the proposed designs. Using the proposed algorithm, a design catalogue is generated as a reference for experimentation. Some examples are given to demonstrate that the proposed designs are promising alternatives for practical applications.



## 1. Introduction

When conducting a physical experiment, a natural phenomenon is that the experimental outcomes might be contaminated with systematic and random errors. Fisher (1926) proposed several principles for quantifying these errors, including the principles of randomization, blocking and replication, such that agricultural field trials can be implemented and analyzed in a more efficient manner. These principles have been widely adopted in many industrial and scientific studies. Random arrangement of treatments or treatment combinations can avoid systematic errors, which have not been recognized in advance, among the experimental units. When the experimental units are drastically heterogeneous, these units are divided into several groups called blocks, such that the intra-block variation is smaller than the inter-block variation. If blocking is done well, variability over units would be quantified, and the parameters of interest can be estimated in a more precise manner than an unblocked design. Typically, block effects can be classified into two types: fixed and random block effects. When the levels of a blocking factor are randomly chosen and statistical inference is made for the entire level population, the block effects are reasonably treated as random effects. Otherwise, fixed block effects. For example, if an industrial experiment is performed by some randomly chosen operators, the operator effects are then regarded as random block effects in the data analysis. Similarly, if a food product is manufactured by raw materials, which are randomly sampled from different batches, then the batch effects are often assumed to be random. Under the assumption of random block effects, an extra component of variance is introduced to the experimental response, and a more comprehensive analysis is required for inferencing the factorial effects and variance components. For an excellent introduction to the analysis of experiments with fixed and random block effects, the reader is referred to Wu and Hamada (2009).

Replication of treatments or treatment combinations can provide an unbiased estimate for the error variance and improve the estimation efficiency. A well-designed experiment with sufficient replicates offers a solid basis for statistical inference. In

* Corresponding author.
*E-mail address:* shinfu@ntu.edu.tw (S.-F. Tsai).

http://dx.doi.org/10.1016/j.jspi.2017.08.004
0378-3758/

practice, an unreplicated factorial design is often used for screening active effects at the preliminary stage of a multifactorial study, because of its run-size economy. Based on the principle of effect sparsity that only few effects are vital and many of them are trivial, several methods have been developed for identifying truly active effects from unreplicated factorial experiments. For a comprehensive review and comparison among these proposals, the reader can consult Hamada and Balakrishnan (1998). When the effect sparsity principle is not fully satisfied, most of the existing methods would not work out reliable findings, primarily due to the lack of a replication-based estimate for the error variance. Fully replicating all the treatment combinations is a straightforward way to get the pure replicates. However, this can rapidly outgrow the cost of an experiment. A cost-effective compromise is to add a fraction of repeated runs to an unreplicated design, such that the error variance can be unbiasedly and efficiently estimated through these pure replicates. Recently, partially replicated factorial designs have received much attention by several authors, including Butler and Ramos (2007), Lupinacci and Pigeon (2008), Dasgupta et al. (2010), Chatzopoulos et al. (2011), Ou et al. (2013), Bird and Street (2016), and Li and Qin (2017), among others. Based on the domain knowledge or prior information, potentially active effects can be explicitly enumerated by researchers in some experimental studies. The collection of these factorial effects is called the requirement set, which typically consists of all the main effects and certain two-factor interactions. Liao and Chai (2004) first introduced two-level parallel-flats designs with some identical flats for estimating the user-specified requirement sets. Subsequently, Liao and Chai (2009) proposed a set of sufficient conditions for a two-level parallel-flats design with exactly two identical flats to be D-optimal over all competing designs. Tsai and Liao (2011) further extended these results to mixed two- and three-level parallel-flats designs with respect to the A-, D- and E-optimality criteria. However, none of these works takes the variability over units into account. This motivates us to develop systematic approaches for constructing and analyzing two-level factorial block designs with partial replication, such that a broader collection of partially replicated designs can be employed for addressing real-world problems.

The rest of this article is organized as follows. Section 2 introduces some notation and terms. When the block effects are regarded as random effects and the requirement set is specified, Section 3 gives a set of sufficient conditions for a two-level parallel-flats block design to be D-optimal over all competing designs. Based on these sufficient conditions, an algorithm is developed for systematically constructing the proposed designs, and a design catalogue is then generated as a reference for experimentation. Section 4 discusses the analysis of partially replicated factorial block designs. In addition, a simulated experiment is analyzed by using the replication-based inference procedure to demonstrate that the proposed designs are feasible for practical applications. Concluding remarks are given in the final section.

## 2. Notation and definitions

Some notation and definitions to be used throughout this article are introduced as follows.

### 2.1. Parallel-flats block designs

Let $d_i$ be a two-level design, which consists of all the solutions $\boldsymbol{t}$ satisfying the linear equations $\boldsymbol{At} = \boldsymbol{c}_i$ over $GF(2)$, where $\boldsymbol{A}$ stands for a $p \times n$ alias matrix of rank $p$; $\boldsymbol{c}_i$ stands for a $p \times 1$ coset indicator vector; and $GF(2)$ denotes the Galois field of order two. Typically, $d_i$ is called a single-flat design. The juxtaposition of $f$ single-flat designs $d_1, d_2, \ldots, d_f$ is further defined as an $f$ parallel-flats design, which is abbreviated as $f$-PFD and is denoted by $d$, namely, $d = \{d_1, d_2, \ldots, d_f\}$. It is clear that an $f$-PFD is constructed by collecting all the solutions of linear equations determined by the matrix pair $(\boldsymbol{A}, \boldsymbol{C})$, where $\boldsymbol{C} = [\boldsymbol{c}_1\ \boldsymbol{c}_2 \cdots \boldsymbol{c}_f]$ is a $p \times f$ coset indicator matrix. Note that the columns of $\boldsymbol{C}$ are not necessarily all distinct. Specifically, an $f$-PFD with exactly two identical flats is called an $f$-PFDR by Liao and Chai (2004). For an introduction to the general theory of parallel-flats designs, the reader is referred to Section 15.12 of Cheng (2014).

Suppose that the treatment combinations determined by $(\boldsymbol{A}, \boldsymbol{C})$ are allocated to $b$ blocks of size $k$. Under the framework of parallel-flats designs, a blocking scheme can be characterized through a partition on the columns of coset indicator matrix $\boldsymbol{C}$ given by

$$\boldsymbol{C} = [\boldsymbol{C}_1 \quad \boldsymbol{C}_2 \quad \cdots \quad \boldsymbol{C}_b],$$

where $\boldsymbol{C}_j = [\boldsymbol{c}_{j1}\ \boldsymbol{c}_{j2} \cdots \boldsymbol{c}_{jg}]$; and $g$ denotes the number of flats in each block. Note that the block size $k = g \times 2^{n-p}$, and the total run size $N = f \times 2^{n-p} = b \times g \times 2^{n-p} = b \times k$. The treatment combinations determined by $(\boldsymbol{A}, \boldsymbol{C}_j)$ are then randomly allocated to the experimental units in the $j$th block. This two-level factorial block design is called an $f$ parallel-flats block design abbreviated as $f$-PFBD. A special class of $f$-PFBDs called the $f$-PFBDRs is defined by $(\boldsymbol{A}, \boldsymbol{C})$, where each submatrix $\boldsymbol{C}_j$ of $\boldsymbol{C}$ can be further expressed as

$$\boldsymbol{C}_j = [\boldsymbol{c}_{j1} \quad \boldsymbol{C}_{j0}].$$

Note that the columns of $\boldsymbol{C}_{10}, \boldsymbol{C}_{20}, \ldots, \boldsymbol{C}_{b0}$ are all distinct, and $\boldsymbol{c}_{11}, \boldsymbol{c}_{21}, \ldots, \boldsymbol{c}_{b1}$ are the same as the first column of $\boldsymbol{C}_{10}$. Furthermore, let $\boldsymbol{C}_0$ be the $p \times (f-b)$ matrix derived by juxtaposing $\boldsymbol{C}_{10}, \boldsymbol{C}_{20}, \ldots, \boldsymbol{C}_{b0}$, that is,

$$\boldsymbol{C}_0 = [\boldsymbol{C}_{10} \quad \boldsymbol{C}_{20} \quad \cdots \quad \boldsymbol{C}_{b0}].$$

The $(f-b)$-PFBD determined by $(\boldsymbol{A}, \boldsymbol{C}_0)$ is obtained by removing the $b$ identical flats from an $f$-PFBDR. Equivalently, an $f$-PFBDR can be constructed by adding the first flat of its first block to itself and to the other $b-1$ blocks of an $(f-b)$-PFBD.