The Efficiency of Incomplete Block Designs in On-Farm Trials

by

Robert Peter Ndugwa

Dissertation

Submitted in Fulfillment

of the Academic Requirements

for the Degree of

Master of Science in Biometry

in the

School of Mathematics, Statistics and

Information Technology

University of Natal , Pietermaritzburg

December, 2002

Dedication

To my mother Edith Nakatudde and my entire family.

Declaration

The work described in this dissertation was carried out in the School of Mathematics, Statistics and Information Technology, University of Natal, Pietermaritzburg, from February 2001 to December 2002, under the supervision of Dr. Peter Njuho.

The studies represent original work by the author and have not otherwise been submitted in any form for any other degree or diploma to any University. Where use has been made of the work of others it is duly acknowledged in the text.

Date: December. 2002

Student: ---------- 01-04-2003

Robert Peter Ndugwa

Supervisor: ---------- 01-04-2003

Dr. Peter M Njuho

Acknowledgements

I would like to thank my supervisor Dr. Peter Njuho for his advice, encouragement and guidance in the preparation of this thesis. Thank you for your untiring support and seemingly unlimited belief in me.

I would also like to thank Prof. Linda Haines, Mr. Harvey Dicks and Dr. Henry Mwambi for their guidance while undertaking this study. I would also like to thank the entire staff and postgraduate students of the School of Mathematics, Statistics and Information Technology.

There are special mentors that I must acknowledge due to their importance in my work. Dr. M. Nabasirye, thank you! Prof D. S. Osiru and Prof. E. Adipala, thank you!

There are many friends from behind the scenes that have encouraged and supported my work, and I wish to thank them. Thank you Chris, Symon, Thomas , Andrew, Nosisa and Maureen.

I also thank the Rockefeller Foundation and Makerere University for availing the funds for my studies.

Finally, I would like to thank my family for all the social, moral and logistical support throughout my life.

Robert Peter Ndugwa.

Abstract

Incomplete block designs have been widely applied in several fields of statistical research. However, there has been little work done on applying these designs in on-farm trials. On-farm trials have gained considerable attention in the area of agricultural research. Despite this, little attention has been given to the efficient design of experiments for improvement of the precision of results from these trials. Much of the work conducted in these trials is done using the randomized complete block design. This is a standard design with proper orthogonality properties. However, the size of experiments, variability in farms, missing observations etc., are a few of the problems encountered in on-farm trials. These problems always render the randomized complete block design less efficient in studies in on-farm trials.

This study reviews the incomplete block designs with the aim of analyzing the efficiency of these designs when applied to on-farm trials. Several techniques are employed to classify these designs as compared to the randomized complete block designs. Optimization of these designs is then considered both under constrained and unconstrained design scenarios. Missing observation analysis is discussed. In this study, computations of the efficiencies in constrained and unconstrained cases of on-farm scenarios after incorporating appropriate covariance structures are compared to evaluate designs.

Worst cases of missing observations are identified alongside their effect on the design efficiencies. Overall, we show that the gain in the use of incomplete block designs in on-farm trials is of key importance. However, the maximization of information and minimizing cost cannot always be achieved simultaneously. As a consequence, compromise designs should always be considered, which satisfy requirements as far as possible and are practical at the same time.

Table of Contents

List of Tables

List of Figures

Abbreviations

ANOVA	: Analysis of variance
BIBD	: Balanced incomplete block design
BLUE	: Best linear unbiased estimators
IBD	: Incomplete block design
MINQUE	: Minimum norm quadratic unbiased estimation
MIVQUE	: Minimum variance quadratic unbiased estimation
ML	: Maximum Likelihood
MVN	: Multivariate Normal
OVS	: Orthogonal variation structure
PBIB	: Partially balanced incomplete block
RCBD	: Randomized complete block design
REML	: Residual Maximum Likelihood
SLSE	: Simple least squares estimates
UBIBD	: Unbalanced incomplete block design

Chapter 1

Introduction

On-farm trials are beginning to receive a wide acceptance and credibility where proper statistical procedures have been applied(Ashby *et al*, 1990; Lockeretz, 1987; Stroup *et al*, 1993). Therefore, it is of great importance to pay attention to the efficient design of on-farm trials. The process of making better decisions begins with the proper design of your experiment(Zandastra *et al*,1981). Proper integration of agronomic and socio-economic perspectives into diagnosis, together with an analysis of the existing variability in management practices and yields in farmers' fields, has the potential to improve the efficiency of on-farm trials and at times substitute partially for costly experimentation(Byerlee, Triomphe and Sebillotte, 1991). Efficient designs that lead to extraction of desired information from on-farm trials are required. Therefore, construction of optimal or near-optimal designs to improve on the efficiencies of on-farm trials should be considered.

The randomized complete block design(RCBD) is commonly used in on-farm trials because of the ease and control of site variability by simple blocking, which is preferred over the completely randomized design. But its ability to account for the site variability is limited (Fu, Clarke, Namkoong and Yanchuk, 1998). In many ideal situations, RCBD provides an easier, flexible and less sophisticated analysis as compared to other designs . However, the conditions for on-farm trials are less controlled than those of research stations and this calls for special attention to the design of such experiments. On-farm trials are characterized by several problems such as missing observations, large field layouts ,

high variability etc (Lockeretz, 1987). In on-farm trials, the RCBD provides for bigger blocks and this incorporates a large amount of variability within blocks. This explains why in forest trials, genotype effects obtained from RCBD trials are always inflated or have large standard errors (Magnussen, 1993a).

In on-farm trials, incomplete block designs(IBD) can be of proven value especially in reducing variability through reduction in block sizes. IBD were originally devised to allow for smaller block sizes when equal precision was desired on all treatment comparisons (Yates, 1936). Yates points out two vital situations when small block sizes are desired. One is where the number of treatments is so large that the amount of material needed for a complete block is very heterogeneous. Thus by using smaller blocks the material within the blocks can be more homogeneous, thereby giving more efficient estimates of treatment comparisons. The second situation is when the number of possible units per block is less than the total number of treatments. On-farm trials often fall into one the two situations described above.

In many on-station trials, IBD have shown higher efficiency values than RCBD. Alpha-lattice designs have been shown to be more efficient than RCBD in field trials conducted in the UK (Patterson and Hunter, 1983; Patterson and Williams, 1976). Thus IBD can be a valuable design to replace RCBD in many of these agricultural on-farm trials. Smaller blocks are less heterogenous than bigger blocks and thus larger site variation among blocks is removed from the experimental error so that the contribution of the site variance effects to the error of estimating treatment means is reduced(Fu, Clarke, Namkoong and Yanchuk, 1998).

The procedure of confounding is applied in factorial experiments when the number of treatments is large (Gomez and Gomez,1984). This is aimed at reducing the block size in order to ensure more precise estimation of lower order interactions at the cost of less important higher order interactions. However, in a situation in which all treatment com-

parisons are of interest confounding is not applicable (Federer, 1955). This situation calls for the application of the incomplete block designs.

1.1 Case for on-farm trials

On-farm trials are a vital tool to assess the level of technology transfer and applicability on the actual farmers' fields(Gomez and Gomez, 1984). The gap between the on-station trials where conditions are very well controlled and the farmers' fields where conditions are poorly managed is big in developing countries (Lockeretz, 1987; Lockeretz and Anderson, 1993). Thus on-farm trials can be a solution to closing up such technology transfer problems. There are numerous reasons to justify use of on-farm trials. For instance they can be used to determine the general distribution of response to a new technology and to estimate the frequency of success under a wide range of conditions as well as management and environmental factors that impact on success (St-Pierre and Jones, 1999). They also offer a cost-efficient alternative to researchers when a large number of experimental units are required. The importance of on-farm trials as a component of farming systems research is described fully by Hildebrand and Poey (1985).

On-farm trials can be grouped into multi-site(MS) and single-site(SS) trials. In multi-site trials, several farms are used simultaneously. These trials provide a wide scope of study on a given variable across several locations. However, single-site trials reduce the complexity associated with an experiment involving a multitude of locations. Therefore, this reduction in complexity is done at the expense of the inference range that is lost. Thus the validity of the results from SS trials covers a much narrower range of conditions. Different forms of on-farm trials call for different designs (Mustsaers & Walker, 1990). There are many practical challenges related to the experimental design, data collection procedures and data analysis of on-farm trials (Coe, 1998; Hammerton & Lauckner, 1984; Nokoe, 1999). Some on-farm trials are too large to be manageable as compared to on-station

experiments, which receive a well-planned and controlled setup.

Incomplete block designs are characterized by small block sizes that allow for estimation of treatment effects with a certain degree of precision (John, 1971; Mead, 1994). The randomized complete block design is a "conventional" design used widely in field experiments, especially those carried out on-station. The use of the design assumes well-controlled situations where unbalanced and missing data are never a problem. However, research is increasingly shifting from on-station to on-farm where experimental conditions are difficult to predict. Large variability, unbalanced and missing data are common problems associated with on-farm trials (Hildebrand and Russell, 1996). Such problems make analyzing data from such trials difficult and results are less informative. In many on-farm trials, it is unlikely that an experimenter has any control over the experiment as compared to on-station trials. So it is not possible to use the standard block designs since each situation calls for a specific design in order to accommodate the problems that may arise (Mustsaers and Walker, 1991). If missing observations are likely to occur from some unforeseen events, it would be best to select an appropriate initial design with good characteristics and properties to protect against the whole experiment being ruined.

1.2 Design considerations

An assessment of the efficiency of a design that takes account of unforeseen events that may arise is of great benefit. This would enable an experimenter to consider the risk of using a particular design in practice. An experiment can be designed such that it has a certain degree of controlled non-orthogonality, which will not upset the simplicity of the analysis of variance. In most on-farm trials, the responses for some plots, or in some cases even a full block may be lost for reasons unrelated to the structure of the experiment or the treatments used. This prompts the need to choose a design that will minimize the influence of most on-farm problems affecting the results from the trials. The loss of a full

block has been studied by Gupta and Srivastava (1992) and Bhaumik and Whittinghill (1991). The design of on-farm trials involves the selection of the most appropriate factors to control variability and other related problems with on-farm trials. Typically, traditional experimental designs have been used for such purposes but little has been documented about the success of these designs (Hildebrand and Russell, 1996). Randomized complete block designs have the merit of simplicity but can be inefficient in on-farm trials for a number of reasons.

- It may not be possible to choose homogeneous blocks of sufficiently large size for complete blocks in on-farm trials.
- Blocks/farms may be natural units of a size that does not coincide with the size of a treatment replicate.
- A single set of blocks may not be adequate to account for all sources of variability in an experiment in on-farm trials.
- It may not be apparent how to choose an appropriate set of complete randomized blocks prior to an experiment.

All the above reasons reveal an efficiency deficit that can arise in the use of randomized complete block designs. We should compare the loss in efficiency that could be incurred by using these designs with the loss of efficiency that could be incurred by incomplete block designs. Based on these gains and losses in efficiency, we are able to recommend the right designs for on-farm trials. This study aims to assess the efficiency of the incomplete block designs in general in overcoming the limitations of on-farm trials. Particular emphasis is devoted to the way in which most efficiency factors would behave in varying situations for different forms of incomplete block designs.

A review of incomplete block designs is given in chapter 2. In chapter 3 we restrict attention to the efficiency estimations in designs and the design optimality criterions. Chapter 4 gives an in depth knowledge on the optimal incomplete block designs alongside

the robustness of these designs in the case of missing observations. Chapter 5 considers the different scenarios in on-farm trials and how design selection tools can be used to obtain the most appropriate design. A case study of on-farm trials set up in Uganda is analyzed in this chapter. The conclusions and suggestions are given in chapter 6.

1.3 Overall objectives

To determine the extent to which incomplete block designs could be used in on-farm trials.

Specific objectives

1. To assess the possibilities of using incomplete block designs in on-farm trials.

2. To determine the different levels of efficiency of incomplete block designs with respect to on-farm trials by designing flexible experimental designs.

3. To determine the precision of treatment means and the difference of treatment means using power analysis.

Chapter 2

Literature Review

2.1 Introduction

Major factors for consideration when designing on-farm experiments are the limited and varying farm space, and the inherent variability within a given farm (Riley and Alexander, 1997). In on-farm trials, challenges and conditions are experiment dependent and therefore no set of rules provide the "correct" design. The idea that not every farm requires an identical set of treatments is more novel (Steiner, 1987; Stroud, 1993). It is preferable to limit the number of treatments in order to increase the precision of the experiment. The design strategy described in Gomez and Gomez (1984), which requires equal numbers of plots and replicates on each farm, is not feasible. An incomplete block design provides an alternative to commonly used randomized complete block designs when farms cannot accommodate a full set of treatments(John, 1980). A proper consideration of statistical tools in experimental designs is capable of providing an answer to the problems of on-farm trials. Although statistics is commonly viewed as primarily dealing with post-experimental data analysis, statistical experimental design is of proven value at the planning stages. Fisher(1951; p.3) noted that,

> *"statistical procedure and experimental design are only two different aspects of the same whole, and that they both aid to the logical requirements of the complete process of adding to natural knowledge by experimentation".*

While it is a trivial task to find a good incomplete block design, the topic has been under study for decades and there is a body of research to help find efficient experimental plans. The designs that are most suitable for on-farm trials depend on the questions of interest, and the available resources. Most incomplete block designs are capable of fulfilling this task.

2.2 Incomplete block designs

Incomplete block designs where introduced by Yates(1930) at the Rothamsted Experimental Station. These designs normally pose several questions to researchers:

- Given a resource constraint, what is the best way to choose subsets of the treatments to allocate to the blocks?
- How should the data from the experiment be analyzed?
- How efficient and reliable are these designs?

A common problem arises when the blocks available are not large enough to accommodate all the available treatments. In such instances, we consider the incomplete block designs, where blocks do not contain a full set of treatments. In many forms of investigation that involve screening of large numbers of new treatments or subjects, only a small amount of new material is available and few replications are therefore possible. In an experiment to compare new treatments with old treatments, one is unwilling to allocate more than a small number of replications to the new treatments because of costs involved in screening large numbers of new treatments (Federer, 1955). For instance, in the screening of numerous new fungicides, herbicides, or soil fumigants, at various levels it becomes expensive to run all replications of these treatment combinations. Sometimes it may be desirable or necessary to have the new treatments replicated only once or twice in the experiment and to have the standard or old treatments replicated r times.

A common situation is where farmers cannot apply all the available treatments due to

limited farm size or due to other constraints, therefore research scientists who conduct on-farm trials tend to bias the randomization through farmer selection (Skinner and Mwaniki, 1994). To avoid the bias, Pardey, Roseboom and Beintema (1996) suggested farmer classification procedures which would lead to effective results for particular farm groups. The experimental designs which accomplish this objective are Augmented designs and the related Reinforced and Staircase designs (Das, 1958). For Augmented block designs with new treatments appearing once and the standard treatments appearing r times, one simply sets up block designs for the standard treatments, then enlarges the blocks to accommodate the standard treatments as well as the new treatments (Das, 1958). The statistical analysis is performed on standard yields only to obtain solutions for block effects, which are then used to adjust the yields of the new treatments for the blocks in which they appear (Federer, 1955).

Traditionally, randomized complete block designs (RCBD) have been used in most agricultural trials. A randomized complete block design is constructed on principles of homogeneous experimental units per block (Montgomery, 1976). In basic experimental designs, we apply all treatments within each block because experimental units (plots) located in each block are more or less uniform (John, 1971). Sometimes it is difficult, inconvenient or impossible to apply all treatments in each block (Mead, 1994). This calls for the use of incomplete block designs and forms the basis for their introduction (Yates, 1933).

Balanced incomplete block designs (BIBD) are a special case of incomplete block designs that are constructed such that every pair of treatments occur together in the same block a given number of times (Cochran and Cox, 1957; Bose, Clatworthy & Shrikhande, 1954; Bose, 1947; Rao, 1947). All treatment comparisons are of the same accuracy in these designs. For a given block size, replication, and number of treatments, a BIBD is of maximum efficiency if it exists (Conniffe and Stone, 1974). Conniffe and Stone (1974) de-

fined efficiency as the inverse ratio of the average variance of treatment differences to that of a complete block design with the same replication. Designs where residual effects are negligible are available, and such arrangements include Latin square and Youden squares. These are balanced incomplete block arrangements, with the additional restriction that each treatment occurs just once in each position of the blocks (Conniffe and Stone, 1974). The property of balance, more specifically called total balance, possessed by the BIBD is useful when designing experiments, because it is the next most efficient simple form of design after the orthogonal randomized complete block design (Clarke and Kempson, 1997). Incomplete block designs are expected to be from 1.2 to 1.5 times as efficient as randomized complete block designs (Goulden, 1937). The average gain in efficiency from using incomplete block designs in a series of 166 trials was a factor of 1.43 and in one case as high as 3.82 (Patterson and Hunter, 1983).

Another form of incomplete block designs is the partially balanced incomplete block designs (PBIB). These are designs in which not all pairs of treatments appear together in a block. PBIB designs were introduced by Bose and Nair (1939). Under these we have lattice designs, which is a simple case of the PBIB designs with the L_2 association scheme. For the L_2 association scheme, the treatments are arranged in a square array, and two treatments are first associates if they appear in the same row or column (John, 1971). The practical application of other designs does not provide the real solution in large on-farm agricultural experiments. This is because the sizes and conditions of farms vary considerably. Unbalanced incomplete block designs usually provide a better design to fit the true on-farm situation. Augmented designs, reinforced and staircase designs provided by Das (1958) are more practical in relation to the on-farm trials. Searle (1965) provides a major breakthrough in the analysis of augmented randomized complete block designs. However, a key factor in obtaining appropriate treatment effects is orthogonality. Orthogonal designs have a special role in the calculations of efficiency. An orthogonal design is one in

which treatment differences are estimated independently of block differences (Clarke and Kempson , 1997). The principal consequence of orthogonality is that effects of treatments can be interpreted without simultaneously considering inferences about the block effects. It also simplifies the calculations both for the analysis of variance and for the comparison of treatment effects (Mead, 1994). In these designs all the information on the treatments is obtained from the intra-block analysis. Yates (1933) has the classical statistical account of orthogonal comparative experiments. Yates states that,

> *"Orthogonality is that property of the design which ensures that the different classes of effects to which the experimental material is subject shall be capable of direct and separate estimation without any entanglement".*

Pearce (1953) offers the following condition for orthogonality *" Two classifications are mutually orthogonal if various groups of plots formed by one classification are composed of the same proportionate number of plots of the other".*

Where all pairs of classification are mutually orthogonal, the whole design is said to be orthogonal. In a non-orthogonal design information is obtained from within comparisons and the remaining information is recovered from an inter-block analysis, through comparison between blocks (John, 1987). In the analysis of variance, formulations of an orthogonal decomposition of the data vector are possible such that the total variance is partitioned into components attributable to identifiable causes (Kempthorne, 1952). Rao(1959) applied randomization models to the balanced incomplete block design but did not incorporate complete or incomplete sampling considerations. Although it seems obvious on how to allocate factors into block and treatment, the criteria to be followed are not clearly laid out by many writers (Mead and Curnow, 1983; Preece, 1982; Yates, 1975). Furthermore, in agricultural experiments there is little literature on how this should be done.

This can be illustrated by considering the model

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \tag{2.1}$$

where y_{ijk} is the kth observation due to the ith treatment in the jth block, τ_i denotes the ith treatment effect, β_j denotes the jth block effect , $(\tau\beta)_{ij}$ denotes the interaction effect between the ith treatment and the jth block, and μ is the overall mean, $E(\epsilon_{ijk}) = 0$, $Var(\epsilon_{ijk}) = \sigma^2$ and $cov[\epsilon_{ijk}, \epsilon_{i'j'k'}] = 0$ for $(i, j, k) \neq (i', j', k')$.

Such a model is normally not of full rank and constraints are often placed on one or both the parameters and the estimates in order to obtain a solution. Commonly used constraints include;

$$\sum_i \tau_i = \sum_j \beta_j = \sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0$$

If such constraints are not considered then some of the components in the model may not be estimable(Hocking, 1973; Searle, 1971).

However, the principles related to the design and analysis of on-station experiments are well documented hut how they should be applied to on-farm trials is not well defined. The success of these trials lies on the efficient design of experiments if efficient estimates of treatment means are required. Common to most on-farm trials are the problems of missing observations and unbalanced data. These problems have a significant effect on the orthogonality of the design. However, with a well designed experiment, a good amount of information can be recovered from trials using statistical techniques such as mixed models analysis.

2.2.1 Mixed models review

This consists of both the random effects and fixed effects analysis(Scheffe, 1959). Mixed models were first applied by Fisher(1925) in developing the spilt-plot analysis and in Fisher(1935) while analyzing an experiment involving the testing of varieties at several locations. Several authors have looked at Mixed models since then (Eisenhart, 1947; Plackett, 1960). More recent articles on the subject are by Harville(1977), Searle(1971), and Graybill(1976). These models are a form of the general linear model given as below:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.2}$$

where y, X and β are defined as: y is the vector of n observations, X is a known $n \times p$ matrix of rank r ($r \leq p$), β is a vector of p unknown parameters and ε is the error component such that $E[\varepsilon] = 0$ and $cov[\varepsilon] = \Sigma$.

In the mixed models form, (2.2) can be expressed as (Harville, 1977):

$$y = \sum_{i=1}^{p} x_i \beta_i + \sum_{j=1}^{m} z_j \eta_j \tag{2.3}$$

where $E[y] = X\beta$

z_j is the design matrix for the jth random term and of order $n \times m_j$,

m_j being the number of effects in the jth term,

η_j an $m_j \times 1$ vector of random effects and

$\eta_j \sim N(0, \phi_j I)$ and $m_m = n$ and $z_m = I$, so that $Var[y] = V = \sum_{j=1}^{m} \phi_j Z_j Z_j'$.

The ϕ's form the canonical components of excess variation (Nelder, 1977). They can be interpreted as classical variance components or covariances of the observations (Nelder, 1977; Scheffe, 1959). The advantage of the canonical components are that they have the same interpretations in respect of the variance matrix of the observations for all formulations of the model. It is important to obtain estimates that are Best linear unbiased estimators(BLUEs) and the implication of this for designed experiments is that the experiment needs to be orthogonal for the simple least squares estimates(SLSE) to be BLUEs. In the paper by Houtman and Speed(1983), their concern was not with establishing the equality of SLSEs and BLUEs since for many designs e.g IBD, orthogonality is not obtained and so SLSE are never appropriate. The solution to this is having a model with an orthogonal variation structure (OVS), where the analysis can be based on the hypothesized variance matrix V. Most experimental designs have OVS (Bailey 1982a, 1984). In effect all block designs with equal block sizes and the usual dispersion model have proper orthogonality properties and general balance in treatments (Houtman and Speed, 1983). All in all, partially balanced block designs may or may not fulfill Nelder's (1965, 1968) definition of general balance depending on what decomposition of the treatment subspace is specified. The problem normally arises in examination of the possibilities when ϕ's are

unknown; i.e the ϕ's must be estimated from the data.

There are several estimation techniques which include; Analysis of variance (ANOVA), Maximum Likelihood (ML), Residual Maximum Likelihood (REML), Minimum norm quadratic unbiased estimation (MINQUE) and Minimum variance quadratic unbiased estimation (MIVQUE). The ANOVA estimators are equal to REML, MINQUE, MIVQUE for orthogonal analysis, so long as non-negativity constraints on the variance components do not come into play. ANOVA estimators are location invariant, unbiased, have minimum variance amongst all unbalanced quadratic estimators, and under normality, they are minimum variance amongst all unbiased estimators (Searle, 1971). But they may lead to negative parameter estimates which can fall outside the parameter space. Maximum Likelihood (ML) estimators are biased since they do not consider degrees of freedom lost in estimating the fixed effects of the model and require intensive computations. But non-negativity constraints can be properly imposed. REML estimators have an advantage over ML estimators in that they overcome the loss of degrees of freedom problem of ML and are the same as ANOVA estimators provided that the non-negativity constraints on variance components do not come into play (Harville, 1977). However, for non-orthogonal cases, equivalence of ANOVA and other estimators ceases to hold. In this case, the ANOVA estimators are not available for terms totally confounded with fixed effects and may not have minimum variance. Henderson (1953) provides a method of ANOVA to handle such problems. However, Harville (1977) suggests that REML procedures are to be preferred to Henderson estimators. In studies by Corbeil & Searle (1976), and Harville (1978), comparing ML and REML revealed that although ML estimates are biased, they have smaller mean-squared errors than REML estimates even under orthogonal experiments. In comparing REML and ML for incomplete block designs, the results revealed that there is benefit of using REML mostly in estimating standard errors of the means (Nabugoomu and Allen, 1990, 1994). Fisher (1935) also draws attention to the fact that knowledge of the behavior of the experimental material should be incorporated into the design and

analysis in the form of an hypothesized model if at all minimizing mean-squared errors is desired.

The IBD are crucial to on-farm trials and their use needs to be investigated using the principles highlighted by Yates (1930). There is need for rigorous application of appropriate statistical procedures to the design and analysis of on-farm trials that will enable wide acceptance and improvement in credibility of on-farm trials (Njuho, 1998).

2.2.2 General incomplete block design

The general incomplete block design consists of t treatments in b blocks of size k_j. Treatment $i(i = 1, \ldots, t)$ will have r_i replicates implying that they will appear in r_i plots and block $j(j = 1, \ldots, b)$ will contain k_j plots. The matrix $(t \times b)$ forms the incidence matrix N and it has elements n_{ij} equal to the number of times that treatment i appears in block j. In most incomplete block designs, r_i will be equal to r and all k_j will be equal to k. This implies that the incidence matrix will be composed of zeros and ones. Treatment i appearing in the jth block is denoted by 1 and 0 otherwise. Such designs are referred to as binary designs. If $k_j = k$ for all j, then the design is called proper. If $r_i = r$ for all i, the design is called an equireplicate.

The model

Let y_{ijm} denote the mth observation on the ith treatment in the jth block. We obtain the model;

$$y_{ijm} = \mu + \tau_i + \beta_j + \varepsilon_{ijm} \tag{2.4}$$

where $i = 1, 2, ..., t$; $j = 1, 2, ..., b$; $m = 0, 1, 2, ..., n_{ij}$, ε_{ijm} are the error terms which are uncorrelated random variables each with mean zero and variance σ^2, τ_i is the ith fixed treatment effect, β_j is the jth block effect and μ is the overall mean. The Intra-block analysis and recovery of Inter-block information is widely discussed (John, 1971 ; Cochran and Cox, 1957 ; Khatri and Shah, 1975 ; Bose and Shimamoto, 1952 ; Nair, 1954 ; Yates, 1940).

2.2.3 Intra-block analysis

Let G be the grand total of all the observations, T_i be the sum of all the observations on treatment i, and B_j be the sum of the observations in block j. The intra-block estimates of the treatment effects and the analysis of variance table can be obtained by solving the system of equations below. From (2.4) the following normal equations are obtained using the method of least squares (John, 1971)

$$n\hat{\mu} + \sum_j k_j\hat{\beta}_j + \sum_i r_i\hat{\tau}_i = G \tag{2.5}$$

$$k_j\hat{\mu} + k_j\hat{\beta}_j + \sum_i r_i\hat{\tau}_i = B_j \tag{2.6}$$

$$r_i\hat{\mu} + \sum_j n_{ij}\hat{\beta}_j + r_i\hat{\tau}_i = \sum_i y_{ij} = T_i \tag{2.7}$$

where $j = 1, 2, ..., b; i = 1, 2, ..., t$. The above set of normal equations can be expressed in matrix form as follows;

$$\begin{bmatrix} G \\ T \\ B \end{bmatrix} = \begin{bmatrix} n & 1_t'R & 1_b'K \\ R1_t & R & N \\ K1_b & N' & K \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\tau} \\ \hat{\beta} \end{bmatrix} \tag{2.8}$$

Equation (2.8) can also be expressed as $\hat{Y} = X\beta$ where 1 is a column of all ones, R is the $t \times t$ matrix with diagonal elements $r_1, r_2, \ldots, r_t$ and the off-diagonal elements are zero. Thus $R = diag(r_1, r_2, \ldots, r_t)$. K is a $b \times b$ matrix defined as $K = diag(k_1, k_2, \ldots, k_b)$ and N is the incidence matrix of t rows and b columns where $N = (n_{ij})$, $T' = (T_1, T_2, \ldots, T_t)$, $B' = (B_1, B_2, \ldots, B_b)$, $\tau' = (\tau_1, \tau_2, \ldots, \tau_t)$ and $\beta' = (\beta_1, \beta_2, \ldots, \beta_b)$. The solutions to the matrix system (2.8) depend on the rank of X whether it is of full rank or not full rank. In most incomplete block designs, X is of less than full rank, which implies that many solutions are possible. From John (1971), multiplying both sides of the equation (2.8) by

$$\begin{bmatrix} N^{-1} & 0 & 0 \\ 0 & 1_t & -NK^{-1} \\ 0 & -N'R^{-1} & 1_b \end{bmatrix} \tag{2.9}$$

Note that $R1_t = N1_b$ and $K1_b = N'1_t$.

This gives the following set of equations

$$\hat{\mu} + \left\{\frac{1'R\hat{\tau}}{N}\right\} + \left\{\frac{1'K\hat{\beta}}{N}\right\} = \frac{G}{N} \tag{2.10}$$

$$T - NK^{-1}B = (R - NK^{-1}N')\hat{\tau} = Q \tag{2.11}$$

$$B - N'R^{-1}T = (K - N'R^{-1}N)\hat{\beta} \tag{2.12}$$

The above equations are of primary interest and are referred to as the reduced intra-block equations (John, 1971). From Giesbrecht (1986), one set of solutions to the equations can be obtained as;

$$\hat{\tau} = (R - NK^{-1}N')^{+}(T - NK^{-1}B) \tag{2.13}$$

and

$$\hat{\beta} = (K - N'R^{-1}N)^{+}(B - N'R^{-1}T) \tag{2.14}$$

where $(K - N'R^{-1}N)^{+}$ denotes the Moore-Penrose generalized inverse of $(K - N'R^{-1}N)$. From (2.11) the reduced normal equations for the estimation of the vector of treatment effects can be expressed in the form;

$$C\hat{\tau} = Q \tag{2.15}$$

where $C = (R - NK^{-1}N')$ and Q depends on the n observations, which are assumed to be uncorrelated with variance σ^2 independent of the block size. Note that $C1 = 0$ and $(K - N'R^{-1}N)1 = 0$, so if we consider a case where C has rank $t-1$ and $(K - N'R^{-1}N)$ has rank $b-1$, then this form of design is connected and as such contrasts and effects are estimable. Thus we obtain the $\hat{\beta}$ from the equations $K1\hat{\mu} + N'\hat{\tau} + K\hat{\beta} = B$
as $\hat{\beta} = K^{-1}B - K^{-1}N'\hat{\tau} - 1_b\hat{\mu}$.

From John(1971), the sum of squares for the model is then given as

$$\begin{aligned} G\hat{\mu} + B'\hat{\beta} + T'\hat{\tau} &= G\hat{\mu} + B'K^{-1}B + (T' - B'K^{-1}N')\hat{\tau} - G\hat{\mu} \\ &= B'K^{-1}B + Q'\hat{\tau} \end{aligned} \tag{2.16}$$

The resultant analysis of variance table for the intra-block analysis using G^2/N as correction factor is then given as follows:

Table 2.1: Analysis of variance table for intra-block analysis

Source of variation	Sums of squares	Degrees of freedom
Blocks(ignoring treatments)	$B'K^{-1}B - \frac{G^2}{N}$	b-1
Treatments(adjusted for blocks)	$Q'\hat{\tau}$	$t-1$
Residual	By subtraction	$N-b-t+1$
Total	$\sum\sum\sum y_{ijm}^2 - \frac{G^2}{N}$	$N-1$

Adjusted treatment and block totals

The adjusted treatment totals Q_i are linearly dependent since $Q'1_t = \sum Q_i = 0$ (where this is not the case, the set of equations $C\hat{\tau} = Q$ would be inconsistent). Therefore ,

$$Q'1_t = T'1_t - B'K^{-1}N'1_t$$

$$Q'1_t = G - B'K^{-1}K1_b = G - G = 0$$

Similarly, $(B - N'R^{-1}T)'1_b = 0$

2.2.4 Inter-block analysis

In the case of incomplete block designs, the block totals may also be expected to provide some information on the treatments, since the set of observations occurring in different blocks are different. This technique introduced by Yates(1940) utilizes the block totals to estimate treatment differences. It is when the blocks are considered random that one can utilize the information provided by block totals. Consider (2.4) simplified such that

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \tag{2.17}$$

Assuming a proper design and let the variance of the intra-block errors e_{ij} be σ_e^2. We assume that the β_j are random variables and therefore $\beta_j \sim iidN(0, \sigma_b^2)$. From the b block totals , considering all the above conditions, the block totals can be regarded as a set of observations:

$$B_j = k\mu + \sum_i n_{ij}\tau_i + (k\beta_j + \sum_j \epsilon_{ij})$$
$$= k\mu + \sum_i n_{ij}\tau_i + \epsilon_j^*$$

thus on the basis of the assumptions made, $\epsilon_j^* \sim iidN(0, k^2\sigma_b^2 + k\sigma_e^2)$. Applying least squares analysis, we minimize

$$\sum_j (B_j - k\tilde{\mu} - \sum_h n_{ih}\tilde{\tau}_h)^2 \tag{2.18}$$

where $\tilde{\mu}$ and $\tilde{\tau}$ denote the inter-block estimates; $\hat{\mu}$ and $\hat{\tau}_i$ denote the intra-block estimates. We obtain normal equations by taking partial derivatives with respect to the parameters and equating them to zero, i.e

$$\frac{\delta}{\delta\tilde{\mu}}(\sum_j (B_j - k\tilde{\mu} - \sum_h n_{ih}\tilde{\tau}_h)^2 = 0 \tag{2.19}$$

This yields

$$\sum_j \sum_h n_{jh}\tilde{\mu} + \sum_j \sum_h n_{jh}\tilde{\tau}_h = G$$

while

$$\frac{\delta}{\delta\tilde{\tau}_i}(\sum_j (B_j - k\tilde{\mu} - \sum_h n_i h\tilde{\tau}_h)^2 = 0 \tag{2.20}$$

yields

$$k\sum_j n_{ij}\tilde{\mu} + \sum_j \sum_h n_{ij}n_{jh}\tilde{\tau}_h = \sum_j n_{ij}B_j$$

for $i = 1, 2, \ldots, t$. Solving the two equations and applying the side condition

$$\sum_i \tilde{\tau}_i = 0$$

we obtain

$$\tilde{\tau}_j = \frac{T_i^* - rk\tilde{\mu}}{r - \lambda}$$

for $i = 1, 2, \ldots, t$. These estimates are orthogonal to the intra-block estimates obtained previously. The estimates of the treatment differences are,

$$\tilde{\tau}_i - \tilde{\tau}_h = \frac{T_i^* - T_h^*}{r - \lambda} \tag{2.21}$$

$h \neq i$

where λ is the number of times treatment i and treatment h appear in a common block. Thus the $Var[\tilde{\tau}_i - \tilde{\tau}_h]$ can be obtained. From $T_i^* = \sum_j n_{ij} B_j$, it follows that

$$Var[T_i^*] = \sum_j n_{ij}^2 Var[B_j]$$

$$Var[T_i^*] = \sum_j n_{ij}^2 (k^2 \sigma_b^2 + k\sigma_e^2)$$

$$Var[T_i^*] = var[T_h^*]$$

Also the covariance

$$Cov[T_i^*, T_h^*] = \lambda(k^2 \sigma_b^2 + k\sigma_e^2)$$

The variance of the treatment differences can be obtained as

$$Var[\tilde{\tau}_i - \tilde{\tau}_h] = \{\frac{1}{r - \lambda}\}^2 (2rk^2\sigma_b^2 + 2rk\sigma_e^2 - 2\lambda k^2 \sigma_b^2 - 2\lambda k \sigma_e^2) \tag{2.22}$$

$$= \frac{2(r - \lambda)k}{(r - \lambda)^2} (k\sigma_b^2 + \sigma_e^2)$$

$$= \frac{2k}{(r - \lambda)} (k\sigma_b^2 + \sigma_e^2)$$

For $r \neq \lambda$ consider a BIBD, its normal equations for the inter-block estimation are given as

$$G = N\hat{\mu} + \sum_i r\tilde{\tau}_i$$

$$\sum_j n_{ij} B_j = rk\tilde{\mu} + r\tilde{\tau}_i + \lambda \sum_h \tilde{\tau}_h$$

Denote $\sum n_{ij} B_j$ by T_i'. Considering the side condition $1'\tau = 0$, the first equation gives $\tilde{\mu} = G/N$, and the second equation becomes

$$T_i' = (r - \lambda)\tilde{\tau}_i + rk\tilde{\mu}$$

hence

$$\tilde{\tau}_i = \frac{T_i' - rk\tilde{\mu}}{r - \lambda} \tag{2.23}$$

Note that $\sum T_i' = KG = trk\tilde{\mu}$ so that $\sum \tilde{\tau}_i \simeq 0$. Thus in the general case, we minimize

$$(B - K\tilde{\mu}1_b - N'\tilde{\tau})'(B - K\tilde{\mu}1_b - N'\tilde{\tau})$$

and obtain the normal equations as

$$\begin{bmatrix} G \\ NB \end{bmatrix} = \begin{bmatrix} bk & 1'R \\ R1K & NN' \end{bmatrix} \begin{bmatrix} \tilde{\mu} \\ \tilde{\tau} \end{bmatrix} \tag{2.24}$$

Multiplying both sides by

$$\begin{bmatrix} 1 & 0 \\ -R1/b & I \end{bmatrix}$$

gives the equations

$$\begin{bmatrix} G \\ NB - \frac{GR1}{b} \end{bmatrix} = \begin{bmatrix} bk & 1'R \\ 0 & NN' - \frac{RJR}{b} \end{bmatrix} \begin{bmatrix} \tilde{\mu} \\ \tilde{\tau} \end{bmatrix} \tag{2.25}$$

Recall that $R1_t = N1_b$ thus the above yields

$$N(B - G1/b) = (NN' - RJR/b)\tilde{\tau}$$

and recalling that $1'R\tau = 0$ the equations become $G/N = \tilde{\mu}$ and

$$N(B - G1/b) = NN'\tilde{\tau}$$

so that

$$\tilde{\tau} = (NN')^{-1}N(B - \frac{G1}{b}) = (NN')^{-1}NB - 1(\frac{G}{bk}) \tag{2.26}$$

Therefore, the existence of the inter-block estimates requires that NN' exists, and hence that the incidence matrix N has rank t. So we have two sets of estimates: $\hat{\tau} = B_{11}Q$ and $\tilde{\tau}$ which is a function of B and G. Since for any i and any j , $cov(Q_i, B_j) = 0$, it follows that the two sets of estimates are orthogonal and under normality are independent.

2.3 Balanced incomplete block designs

An arrangement of t treatments in b blocks of k plots each, $(k < t)$, is known as a Balanced Incomplete Block design(BIBD) if every treatment occurs once and only once in r blocks and any two treatments occur together in λ blocks. The constants t, b, r, k, N and λ are used in incomplete block designs.

b = Number of blocks in the experiment.

t = Number of treatments.

k = Number of experimental units per block.

r = Number of replications for a given treatment in the experiment.

N = Total number of experimental units.

λ = Number of times each pair of treatments appear together in a block (number of concurrences).

n_{ij} = The number of times block j contains a given treatment i. $(i = 1, 2, \ldots, t; j = 1, 2, \ldots, b)$.

Thus a balanced incomplete block design can be denoted as BIBD(t, b, r, k, λ).

The following conditions must be fulfilled in order to obtain a balanced incomplete block design.

1.

$$rt = bk = N \tag{2.27}$$

Consider a single treatment occurring r times. For a given treatment t_i, it must appear together with others in a block $r(k-1)$ times in total.

But it also appears λ times with each other $(t-1)$ treatments. So the total number of concurrences of all pairs of treatments is given as follows

2.

$$\lambda(t-1) = r(k-1). \tag{2.28}$$

such that $\lambda = \frac{r(k-1)}{(t-1)}$

But by definition, λ has to be an integer and thus this equation places restrictions on the possible values of k and t.

3.

$$t \leq b \tag{2.29}$$

A design with $b = t$ and $r = k$ is said to be symmetric. This inequality is due to Fisher (1940) and the proof was presented by Bose (1949). Shrikande (1950) shows that the above three conditions are necessary for the existence of a balanced incomplete block design although not sufficient. Given t treatments and blocks of size k, we can construct BIB designs by taking the t treatments, k at a time in all possible ways.

Consider the model(2.17), for BIBD the following normal equations are obtained(John, 1971)

$$rt\hat{\mu} + k\sum_j \hat{\beta}_j + r\sum_i \hat{\tau}_i = \sum_i \sum_j y_{ij} = G \tag{2.30}$$

$$k\hat{\mu} + k\hat{\beta}_j + \sum_i n_{ij}\hat{\tau}_i = \sum_i y_{ij} = B_j \tag{2.31}$$

$$r\hat{\mu} + \sum_j n_{ij}\hat{\beta}_j + r\hat{\tau} = \sum_i y_{ij} = T_i \tag{2.32}$$

$i = 1, 2, ..., t$ and $j = 1, 2, ..., b$. Equation (2.30) is obtained by minimizing for μ, equation (2.31) forms the block equation and equation (2.32) forms the treatment equations.

The treatments effects are obtained by eliminating block effects. Solving the block equations for $\hat{\beta}_j$. We have,

$$\hat{\beta}_j = \frac{1}{k}\left\{B_j - \sum_i n_{ij}\hat{\tau}_i - k\hat{\mu}\right\} \tag{2.33}$$

or

$$\hat{\beta}_j = \frac{1}{k}B_j - \frac{1}{k}\sum_i n_{ij}\hat{\tau}_i - \hat{\mu} \tag{2.34}$$

Substituting (2.34) into (2.32) we obtain

$$r\hat{\mu} + \sum_j n_{ij}\left\{\frac{1}{k}B_j - \frac{1}{k}\sum_i n_{ij}\hat{\tau}_i\right\} + r\hat{\tau}_i = T_i \tag{2.35}$$

Equation(2.35) simplifies to

$$rk\hat{\mu} + \sum_j n_{ij}\left\{B_j - \sum_i n_{ij}\hat{\tau}_i\right\} + rk\hat{\tau}_i = kT_i \tag{2.36}$$

But

$$\sum_j n_{ij}k\hat{\mu} = rk\hat{\mu}$$

thus

$$rk\hat{\tau}_i - \sum_j \sum_h n_{ij}n_{ih}\hat{\tau}_h = kT_i - \sum_j n_{ij}B_j$$

But

$$\sum_i n_{ij}n_{ih} = \begin{cases} \lambda \rightarrow h \neq j \\ r \rightarrow h = j \end{cases}$$

Further simplification leads to

$$rt\hat{\tau}_i - r\hat{\tau}_i - \lambda\sum_{h\neq j}\hat{\tau}_h = kT_i - \sum_j n_{ij}B_j \tag{2.37}$$

$$= kQ_i$$

where

$$Q_i = T_i - \frac{1}{k}\sum_j n_{ij}B_j$$

$i = 1, 2, \ldots, t$. The component Q_i is the sum of the deviations of each observation on treatment j from the mean of all observations in that block. It is also called the adjusted treatment total for the *ith* treatment.

Imposing the condition $\sum\hat{\tau}_h = 0$ would imply that

$$\sum_{h\neq i}\hat{\tau}_h = -\hat{\tau}_i$$

Therefore $rk\hat{\tau}_i - r\hat{\tau}_i + \lambda\hat{\tau}_i = kQ_i$

$$\left\{r(k-1) + \frac{r(k-1)}{(t-1)}\right\}\hat{\tau}_i = kQ_i$$

which simplifies to

$$\lambda t\hat{\tau}_i = kQ_i \tag{2.38}$$

Thus

$$\hat{\tau}_i = \frac{k}{\lambda t}Q_i$$

for $i = 1, 2, ..., t$. The sum

$$\sum_i n_{ij}B_j$$

can be computed, B_j is the total for block j and $n_{ij} = 1$ only if treatment i is in block j. Summing over the block totals, while considering only those containing treatment i, provides the ANOVA table 2.2.

Table 2.2: Analysis of variance table for inter-block analysis

Source of variation	Degrees of freedom	Sums of squares
Mean	1	$\frac{G^2}{rt}$
Blocks ignoring treatments	$b-1$	$\frac{1}{k}\sum {B_j}^2 - \frac{G^2}{rt}$
Treatments eliminating blocks	$T-1$	$\sum_i \hat{\tau}_i Q_i$
Error	$rt-(b-1)-(t-1)$	By subtraction
Total(uncorrected)	rt	$\sum_i \sum_j {y_{ij}}^2$

The error sum of squares is the sum of squares for intra-block error. The mean square for treatments (adjusted for block) is tested against the error mean square.

2.3.1 Estimating the difference between treatment means

From equation (2.37) we note that

$$Q_i = T_i - \frac{1}{k}\sum_j n_{ij}B_j$$

Consider the estimate $\hat{\tau}_h - \hat{\tau}_i$ for $h \neq i$. From (2.38) substituting for Q_i we have

$$\hat{\tau}_i = \frac{kQ_i}{\lambda t} = \frac{k}{\lambda t}\left\{T_i - \frac{1}{k}\sum_v n_{vi}B_v\right\} \tag{2.39}$$

and

$$\hat{\tau}_h = \frac{kQ_h}{\lambda t} = \frac{k}{\lambda t}\left\{T_h - \frac{1}{k}\sum_i n_{ih}B_i\right\} \tag{2.40}$$

Subtracting (2.39) from (2.40) provides the estimate of the treatment means differences.

$$\hat{\tau}_h - \hat{\tau}_i = \frac{k}{\lambda t}\left\{T_h - T_i - \frac{1}{k}\sum n_i h B_i + \frac{1}{k}\sum_v n_{vi}B_v\right\} \tag{2.41}$$

2.3.2 Computing variances of the estimates

In order to estimate variances we use the idea of computing variances for linear functions of treatment means. From John(1971) and Chakrabarti(1962), the variance of the difference between treatment means can be obtained as $Var(\hat{T}_h - \hat{T}_i)$. That is

$$Var(\hat{T}_h - \hat{T}_i) = \frac{k^2}{t^2\lambda^2}\left\{2r\sigma^2 + \frac{2rk\sigma^2}{k^2} - \frac{4r\sigma^2}{k} + \frac{4\lambda\sigma^2}{k} - \frac{2\lambda k\sigma^2}{k^2}\right\} \tag{2.42}$$

This simplifies to $Var(\hat{T}_h - \hat{T}_i) = \frac{1}{t^2\lambda^2}\{2rk^2\sigma^2 - 2rk\sigma^2 + 2\lambda k\sigma^2\}$

$$Var(\hat{T}_h - \hat{T}_i) = \frac{2k\sigma^2}{\lambda^2 t^2}[kr - r + \lambda] = \frac{2k\sigma^2}{\lambda^2 t^2}[r(k-1) + \lambda] \tag{2.43}$$

But from (2.28) we then have $Var(\hat{T}_h - \hat{T}_i) = \frac{2k\sigma^2}{\lambda^2 t^2}[\lambda(t-1) + \lambda]$

Thus

$$Var(\hat{T}_h - \hat{T}_i) = \frac{2k\sigma^2}{\lambda^2 t^2}[\lambda t] = \frac{2k\sigma^2}{\lambda t} \tag{2.44}$$

The estimate of σ^2 is the error mean square obtained from the analysis of variance of the data. For a balanced incomplete block design, the variance of the treatment differences i.e. $Var(\hat{T}_h - \hat{T}_i)$ is a constant value for all $h \neq i$ (John, 1971). This is because from (2.44), all the variables involved are constant terms such that the variance for the treatment differences across all the treatments is a constant. The variance of the difference between treatment means for a BIBD with each treatment replicated r times in blocks of size k can be generalized as

$$Var(\hat{T}_h - \hat{T}_i) = \frac{2k\sigma_p^2}{\lambda t} \tag{2.45}$$

where σ_p^2 is the plot-to-plot variance for the plots in the incomplete block designs. In comparing the efficiency of incomplete block designs in relation to randomized complete block design, we use the variances of treatment differences as a tool for the measure of efficiency (Das and Giri, 1979). For a randomized complete block design with r replications of each treatment, the variance of the treatment differences is given as

$$Var(\hat{T}_h - \hat{T}_i) = \frac{2\sigma_R^2}{r} \tag{2.46}$$

where σ_R^2 is the plot-to-plot variance among the plots in the randomized complete block designs.

2.3.3 The efficiency factor E

The efficiency of an incomplete block design in which each treatment is replicated r times is measured relative to the complete block design with the same number of replicates. From (2.45) and (2.46), the ratio of the variances of the difference between treatment means can be expressed as

$$\frac{\frac{2\sigma_R^2}{r}}{\frac{2k\sigma_p^2}{\lambda t}} = \frac{\lambda t}{rk}\left\{\frac{\sigma_R^2}{\sigma_p^2}\right\} \tag{2.47}$$

If this ratio is greater than 1 then the BIBD is more efficient than the RCBD with which it is compared (Das and Giri, 1979). The quantity $\frac{\sigma_R^2}{\sigma_p^2}$ is expected to be greater than 1 since BIBD has less variable and more homogenous blocks as compared to the RCBD. The efficiency factor is given as

$$E = \frac{\lambda t}{rk} \tag{2.48}$$

E does not provide a complete evaluation of the efficiency of incomplete block designs but it is an important statistic in the evaluation of this efficiency (John, 1987). It is the lower limit to the efficiency of the balanced incomplete block design compared to a complete block design. The efficiency factor E gives the maximum amount of efficiency that could be lost. However one should be more interested in the complete ratio (2.47) since it provides much more information on the gain in efficiency.

2.4 Partially balanced incomplete block designs

Partially balanced incomplete block designs(PBIBD) were introduced by Bose and Nair (1939) and Bose and Shimamoto(1952) as an extension of balanced Incomplete block designs(BIBD). BIBD have an important property that they are the most efficient among all connected incomplete block designs in which each block has the same number of plots and each treatment is replicated the same number of times(Bose, Shrikhande and Bhattacharya, 1953). However BIBD do not exist for all situations and for certain numbers of treatments they can only exist with large numbers of replications(Bailey, 1985). PBIBD were introduced to solve this problem.

2.4.1 Association schemes

Suppose we have $t = mn$ treatments and that we divide them into m groups of n treatments each. We call treatments in the same group first associates, and treatments in different groups second associates in case of 2 associate classes only. In general, an incomplete block design is partially balanced if

1. The experimental material is divided into b blocks of k units each, with different treatments being applied to the units in the same block.

2. There are t treatments each of which occurs in r blocks.

3. There can be established a relation of association between any 2 treatments satisfying the following requirement;

   Two treatments are either $1^{st}, 2^{nd}, \ldots$,or mth associates.

   Each treatment has exactly n_i, ith associates $(i = 1, 2, \ldots, m)$.

   For any two treatments which are ith associates, the number of treatments common to the jth associates of the first, and the kth associates of the 2^{nd} is P^i_{jk} and is independent of the pair of treatments with which we start. Also $P^i_{jk} = P^i_{kj}$.

4. Two treatments which are ith associates occur together in exactly λ_i blocks$(i = 1, 2)$.

Thus the following holds between the parameters of a PBIBD.

1. $bk = tr$. The numbers $t, b, k, \lambda_1, \lambda_2, \ldots, \lambda_m, n_1, n_2, \ldots, n_m$ are the parameters of the 1^{st} kind and the numbers P^i_{jk} $(i, j, k = 1, 2 \ldots, m)$ the parameters of the second kind, belonging to the design. Therefore , there are $2m + 4$ parameters for the first kind and $m^2(m + 1)/2$ for the second (since $P^i_{jk} = P^i_{kj}$).

2. $n_1 + n_2 + \ldots + n_m = t - 1$

3. $n_1\lambda_1 + n_2\lambda_2 + \ldots + n_m\lambda_m = r(k - 1)$

4. $n_i P^i_{jk} = n_j P^k_{ik} = n_k P^k_{ij}$

5. $\sum_{k=1}^{m} P^i_{jk} = \begin{cases} n_j & \text{if } i \neq j \\ n_j - 1 & \text{if } i = j \end{cases}$

A design with b blocks of size k and $rt = bk$ in which every treatment appears λ_1 times with each of its first associates and λ_2 times with each of its second associates is called a group divisible design. The division of the treatments into groups constitutes a group divisible association scheme. PBIBD with only 2 associate classes($m = 2$) are of special interest. These designs depend on 8 parameters of the 1st kind i.e $t, b, r, k, \lambda_1, \lambda_2, n_1, n_2$ connected by the 3 relations (1), (2), (3) and six parameters of the 2nd kind P^i_{jk} $(i, j, k = 1, 2)$ connected by the relations (4) and (5). Therefore the parameters of the 2nd kind can be exhibited as elements of two symmetric matrices i.e

$$P^1_{jk} = \begin{Bmatrix} P^1_{11} & p^1_{12} \\ p^1_{21} & p^1_{22} \end{Bmatrix}$$

and

$$P^2_{jk} = \begin{Bmatrix} P^2_{11} & p^2_{12} \\ p^2_{21} & p^2_{22} \end{Bmatrix}$$

2.4.2 Analysis of PBIBD

Consider t treatments to be compared in b blocks of size k such that each treatment is replicated r times. Therefore the design requires $N = bk$ experimental units. Assuming

all the requirements for having a partially balanced incomplete block design with two associate classes are fulfilled. Consider the following model.

$$y_{ij} = \mu + t_i + b_j + \epsilon_{ij} \tag{2.49}$$

For $j = 1, 2, \ldots, b$; $i = 1, 2, \ldots, t$ and where y_{ij} is the observation resulting from applying the ith treatment to a unit in the jth block, μ is the general mean, t_i is the ith treatment effect, b_j is the jth block effect and ϵ_{ij} is the random effect which is normally distributed $N(0, \sigma^2)$. Denote the total of all observations for the ith treatment by T_i, and let B_j be the sum of all the k observations from the jth block. Let Q_i denote the adjusted yield for the ith treatment, where Q_i is obtained by subtracting from T_i the sum of the block averages for those blocks in which the ith treatment occurs. For example, if the ith treatment occurs in the blocks $1, 2, 3, \ldots, r$ then

$$Q_i = T_i - \frac{1}{k}\{B_{i1} + B_{i2} + \ldots + B_{ir}\}$$

If we let $S_1(Q)$ be the sum of all the adjusted yields for all the first associates of the $i-th$ treatment. i.e. if the first associates of the ith treatment are the treatments numbered $i_1, i_2, \ldots, i_{n1}$, then

$$S_1(Q) = Q_{i1} + Q_{i2} + \ldots + Q_{in1}$$

Let the grand total of all the N observations be denoted by G. According to Bose, Clatworthy and Shrikhande (1953), we define the constants Δ, H, c_1, c_2 by the following relations.

1. $k^2\Delta = (a + \lambda_1)(a + \lambda_2) + (\lambda_1 - \lambda_2)\{a(f - g) + f\lambda_2 - g\lambda_1\}$,
2. $kH = (2a + \lambda_1 + \lambda_2) + (f - g)(\lambda_1 - \lambda_2)$,
3. $k\Delta c_1 = \lambda_1(a + \lambda_2) + (\lambda_1 - \lambda_2)(f\lambda_2 - g\lambda_1)$,
4. $k\Delta c_2 = \lambda_2(a + \lambda_1) + (\lambda_1 - \lambda_2)(f\lambda_2 - g\lambda_1)$.

where $a = r(k - 1)$, $f = P_{12}^1$ and $g = P_{12}^2$

2.4.2.1 Intra-block and Inter-block analysis

The best linear estimate $\hat{t}_i$ of the treatment effect t_i is given as

$$r(k-1)\hat{t}_i = (k-c_2)Q_i + (c_1-c_2)S_1(Q_i)$$

for $i = 1, 2, \ldots, t$. The variance of the estimate of the difference between two treatment effects is given by

$$V(\hat{t}_i - \hat{t}_u) = \frac{2\sigma^2(k-c_j)}{r(k-1)} \tag{2.50}$$

where the treatments i and u are $j-th$ associates $(j = 1, 2)$. With a completely randomized block design, with the same number of replications r, the variance of the difference between two treatment effects is given as

$$\frac{2\sigma^2_{rcbd}}{r}$$

where σ^2_{rcbd} is its error variance. Therefore if a reduction in the block size does not reduce the error variance then setting $\sigma^2_{rcbd} = \sigma^2$ implies that the efficiency factor would then be given as

$$E_j = \frac{k-1}{k-c_j}$$

$(j = 1, 2)$ which is the ratio of the variance of the estimate of $t_i - t_u$ when the randomized block design is used to the variance of the same estimate when the partially balanced block design is used. Thus the quantities E_1 and E_2 are called the efficiency factors of the two kinds of comparison. Therefore, from all possible contrasts, n_1 of these are estimated each with variance

$$\frac{2\sigma^2(k-c_1)}{r(k-1)}$$

and the remaining n_2 are estimated each with variance

$$\frac{2\sigma^2(k-c_2)}{r(k-1)}$$

Therefore the average variance of the treatment comparison is given by

$$\frac{2\sigma^2\{n_1(k-c_1) + n_2(k-c_2)\}}{r(k-1)(t-1)} \tag{2.51}$$

Thus when we take the ratios of the variance of the RCBD with the above variance we obtain the overall efficiency of the partially balanced incomplete block design. Assuming that $\sigma^2_{rcbd} = \sigma^2$ then the overall efficiency is given as

$$E = \frac{(k-1)(t-1)}{n_1(k-c_1) + n_2(k-c_2)} \tag{2.52}$$

2.5 Other incomplete block designs

2.5.1 Lattice designs

One important characteristic of incomplete block designs is whether the blocks can be grouped so that each group of blocks contains a complete replicate of the set of treatments. When this division of an experiment into replicate groups of blocks is possible the design is said to be resolvable (Clatworthy, 1973; Mead, 1994). Non resolvable designs exist but are less valuable for field trials since they don't allow a two stage removal of field trends, where importantly, the first-stage removal is carried out by replicates which are orthogonal to treatments (Williams and Matheson, 1994). The overall structure of an r replicate resolvable design for t treatments with b blocks of size k within each replicate is an example of a generalized lattice design. Cochran and Cox(1957) give special cases of these designs. The common ones are the lattice designs where $k = b$ and the rectangular lattice designs where $k = b - 1$. In these designs, the concurrence matrices contain either zeros or ones in the off-diagonal positions. Consider a BIBD with k^2 treatments arranged in $b = k(k+1)$ blocks with k runs per block and $r = k+1$ replicates. This type of design is what we call a balanced lattice. The blocks can be grouped into sets in such a way that each set contains a complete replicate.

Lattice designs are very important where there are a large number of treatment combinations to be analyzed. Sizes of these designs can be reduced by adopting partially balanced lattices. These designs include the Simple lattice, where 2 replicates of a design for k^2 treatments are in $2k$ blocks of k runs each. e.g consider the 3×3 balanced lattice design below in table 2.3.

Table 2.3: A simple 3×3 balanced lattice design

Replicate 1			Replicate 2		
Block1	Block 2	Block 3	Block 1	Block 2	Block 3
1	4	7	1	2	3
2	5	8	4	5	6
3	6	9	7	8	9

Here we achieve partial balance e.g treatment 2 appears in the same block with treatments 1, 3, 5, 8 and does not appear in a block with any of 4, 6, 7 and 9. A triple lattice is a lattice design with k^2 treatments in $3k$ blocks grouped into 3 replicates. A lattice design with k^2 treatments in $4k$ blocks arranged in 4 replicates is called a quadruple lattice. Cochran and Cox (1957) provide details of some of the available lattice designs. These designs are efficient and optimal. In most practical situations, field conditions dictate the use of resolvable designs (Patterson & Williams, 1976). Yates(1940) acknowledges the importance of resolvable designs which is a major characteristic of most lattice designs. This advantage of lattice designs is broadly shared by other resolvable incomplete block designs especially the $\alpha - designs$.

2.5.2 Alpha designs

Alpha designs are a class of generalized lattice designs, which allows for most practical situations encountered in field trials (Patterson and Williams, 1976). A more detailed study on alpha designs is presented by John and Williams (1995). One advantage is that they provide great flexibility in the number of treatments and block sizes. Alpha-lattice designs in studies of field trials in the UK and Poland have shown more efficiency than RCBD (Patterson and Hunter,1983; Pilarczyk, 1991). In the study of the efficiency of alpha-lattice designs in the international variety trials of barley and wheat, these designs resulted in an average efficiency 18% higher than the randomized complete block designs when average variance was used as the comparison criterion(Yau, 1997).

Patterson and Williams(1976) recommend that the efficiency of the design should be

evaluated before the design is recommended for use. But the computation of the efficiency of the alpha designs is not straight forward as seen below. Williams and Hunter(1978) provide this estimate of efficiency of the alpha designs. Let $N_{v\times b}$ be the incidence matrix, then define

$$C = I - \frac{1}{rk}NN' \tag{2.53}$$

And

$$C_D = I - \frac{1}{rk}N'N \tag{2.54}$$

The subscript D shows that this equation refers to the dual of the original design. One eigenvalue of C_D will be zero; let the non-zero eigenvalues of C_D be denoted by $\theta_1, \theta_2, ..., \theta_{b-1}$. The matrix C has v eigenvalues, which include $\theta_1, \theta_2, ..., \theta_{b-1}$ and also $v - b$ ones.

According to John(1987), the efficiency of the design in question relative to the randomized complete block design is given by E where

$$E = \frac{v-1}{(v-1) + \sum_{i=1}^{b-1} \theta_i^{-1}} \tag{2.55}$$

relative to the randomized complete block design. Paterson and Paterson(1983) recognized (2.55) as the harmonic mean of the eigenvalues.

From Patterson and Williams(1976), an upper bound(U_b) for a resolvable design is given as

$$U_b = \frac{(v-1)(r-1)}{(v-1)(r-1)(b-r)} \tag{2.56}$$

The efficiency computations above are in comparison to a standard randomized complete block design with the same number of treatments and replications.

2.5.3 Cyclic designs

Cyclic designs (John,1971) are vital in the designing of incomplete block designs. They provide a simpler methodology in the designing of incomplete block designs. In recent studies of IBD, cyclic methods of construction of designs have been widely used (Patterson and Williams, 1976; Jarrett and Hall, 1978). For a given set of parameters, many

cyclic designs are possible, and many possess good statistical properties.

Definition 1. *Let $D = (V, B)$ be a BIBD. A bijection σ: $V \rightarrow V$ is called an automorphism of D if $\sigma(B) \in$ ß for all $B \in$ ß. A group G of automorphisms of D is called regular if it acts transitively and faithfully on points, i.e., for any $x, y \in V$, there is a unique $\sigma \in G$ such that $\sigma(x) = y$. If a BIBD has a regular cyclic group of automorphisms, it is called cyclic.*

Note that the order of a regular automorphism group of a (v, b, r, k, λ) BIBD is v. It follows immediately that the complement of a cyclic design is cyclic and multiples of a cyclic BIBD are cyclic. A classical example of a cyclic design is a symmetric design generated by a cyclic difference set. Thus, for any prime power q and any positive integer d, there exists a cyclic $design(d, q)$. An extensive list of cyclic BIBDs can be found in the section on difference families in Colbourn and Dinitz(1996).

Jarrett and Hall (1978) listed efficient generalized cyclic designs, in a sense of maximizing the overall average efficiency factors, with the parameters $10 \leq v \leq 60$ and $r < k \leq 10$. They consider a method for constructing a generalized cyclic design with larger values of the parameters from a generalized cyclic design and give a relationship between the efficiency factors of these designs. They present efficient larger generalized cyclic designs with $v \leq 100$ and $r < k \leq 25$, applying the method to the efficient cyclic designs listed by John (1981). Most of the resulting generalized cyclic designs have fairly high efficiency factors, it therefore seems that the method is useful in constructing efficient larger generalized cyclic designs, provided that the original designs are efficient.

2.6 Summary of designs

There is a comprehensive list of incomplete block designs each with special statistical qualities specific to an area of study. The above represent a few of the incomplete block designs available. In on-farm trials, each problem calls for a special design. Therefore, there is no specific standard incomplete block design that can adequately cover the design

problems in on-farm trials. However, a few of the incomplete block designs discussed in this chapter have qualities that are close to solving on-farm experimental design problems.

Chapter 3

Efficiency Estimation

Efficiency evaluation is a decision tool that enables researchers to select a design that optimizes available resources and produces adequate information. The results from on-farm trials are based on the estimated parameters and therefore estimation procedures that produce efficient estimates are of critical importance. In conducting an experiment, there are always many design issues to resolve(Kempthorne, 1952). These may include deciding which treatments to study, what factors to control and what aspects of a design to randomize (Verdinelli & kadane, 1992). All other aspects relating to considerations of how many experimental units are needed, how many observations should be allocated to each treatment, or what levels of the treatments should be used fall within the ambit of the statistical design(Fisher, 1958). The purpose of efficiency is to improve statistical inferences regarding the quantities of interest by the optimal selection of values for design factors under the control of the investigator, and within the constraints of the available resources(Winer, 1971 ; Ostle, 1963). Several criteria are in existence for this purpose. The efficiency issues discussed in this study are given in the context of improving on on-farm trials results.

3.1 Review of efficiency estimation in designs

Consider an incomplete block design with equal block size k, equal treatment replications r , and each treatment appearing with another λ times in blocks, with the restriction

that each treatment appears at most once in any one block. The normal equations for the treatment parameter estimates τ_i are given as in (2.4), (2.5), and (2.6). Conniffe and Stone (1974) defined the relative efficiency of the incomplete block design as $2/r\bar{v}$ which is the inverse ratio of the average variance to that of a randomized complete block design with the same replication r. And $\bar{v}$ is the average variance of the treatment differences $\tau_i - \tau_j$.

Kempthorne(1956) gives $\bar{v}$ as

$$\bar{v} = \frac{2\sigma^2}{n-1}\sum_{i=2}^{n}\frac{1}{\omega_i} \tag{3.1}$$

where ω_i are the nonzero eigenvalues of the $n \times n$ matrix whose off-diagonal elements are $-\omega_{ij}/k$ and with diagonal elements $(r - r/k)$. Kempthorne(1956) considered the average variance of the elementary treatment contrasts to obtain the harmonic mean of the ω's as a definition of the efficiency factor of a design. Kshirsagar(1958) considered the geometric mean of the ω's as a criterion to measure the efficiency of a design. Thus $(\omega_1\omega_2 \ldots \omega_{t-1})^{-1}$ or

$$\sum_{i=1}^{t-1}\frac{1}{\omega_i}$$

can be used as a measure of the efficiency of a design. For

$$C = rI_t - \frac{1}{k}NN' \tag{3.2}$$

where I_t denotes the identity matrix of order t and $N_{t\times b}$ is the incidence matrix of the design. Then

$$\sum_{i=1}^{t-1}\omega_i = trC = tr(rI_t - \frac{1}{k}NN') \tag{3.3}$$

Thus minimizing either

$$(\omega_1\omega_2 \ldots \omega_{t-1})^{-1}$$

or

$$\sum_{i=1}^{t-1}\frac{1}{\omega_i}$$

subject to the condition

$$\sum_{i=1}^{t-1}\omega_i = K$$

where K is a constant, we obtain the design with the highest efficiency(Kshirsagar, 1958; Mote, 1958 and Conniffe and Stone, 1974). Conniffe and Stone (1974) provided an estimation of the upper and lower bound to the efficiency of an incomplete block design. Thus the $n \times n$ matrix whose off-diagonal elements are $-\omega_{ij}/k$ and with diagonal elements $(r - r/k)$ has rank $n-1$ such that $\omega_1 = 0$. But the sum of the eigenvalues equals to the trace of the matrix and thus $\sum \omega_i = n(r - r/k) = nr(k-1)/k$.

Let

$$\frac{nr(k-1)}{k} = A$$

Since the eigenvalues of the square of the matrix are equal to the squares of the eigenvalues of that matrix, it follows that

$$\sum \omega_i^2 = \frac{n^2 r^2 (k-1)^2}{k^2} + \frac{\sum\sum \omega_{ij}^2}{k^2}$$

Conniffe and Stone (1974) fixed $\sum\sum \omega_{ij}^2$ and hence $\omega_2^2 + \ldots + \omega_n^2 = B$ say. Using the Lagrangian maximization and minimization criterion, it is noted that at a stationary point, n_1 of the eigenvalues equal R_1 and n_2 equal R_2. Where $n_1 R_1 + n_2 R_2 = A$, $n_1 R_1^2 + n_2 R_2^2 = B$, $n_1 + n_2 = n - 1$.

Conniffe and Stone(1974) presented the lower and upper bound to the efficiency of an incomplete block design as

$$min\{\sum_{i=2}^{n} \frac{1}{\omega_i}\} = (n-1)\left\{\frac{n-2}{A - (\frac{n-1}{n-2})^{1/2} P} + \frac{1}{A + (n-2)^{1/2}(n-1)^{1/2} P}\right\} \tag{3.4}$$

and

$$max\{\sum_{i=2}^{n} \frac{1}{\omega_i}\} = (n-1)\left\{\frac{n-2}{A + (\frac{n-1}{n-2})^{1/2} P} + \frac{1}{A - (n-2)^{1/2}(n-1)^{1/2} P}\right\} \tag{3.5}$$

where

$$P = \left\{B - \frac{A^2}{n-1}\right\}^{1/2}$$

Therefore

$$min\{\sum_{i=2}^{n} \frac{1}{\omega_i}\} \leq \sum_{i=2}^{n} \frac{1}{\omega_i} \leq max\{\sum_{i=2}^{n} \frac{1}{\omega_i}\}$$

Using these boundaries, we can compare efficiencies of several designs based on the limits of their stationary values i.e *min* and *max*. Given two designs with the same treatment replications but with different ω_{ij}'s, if the largest stationary value for one design is less than the smallest stationary value for the other design, then the first design is more efficient. This implies that a design that maximizes efficiency can be obtained by choosing ω_{ij} to minimize P (Tocher, 1952; Pearce, 1968 & Kempthorne, 1956).

3.2 The optimality criterion

Optimal experimental designs continue to receive a wide and considerable amount of attention in the statistical literature as acknowledged by Chaloner(1984). Several authors have looked at this field (Kiefer, 1958 & 1959; Fedorov, 1972; Silvey, 1980) and more recently Atkinson and Donev (1992). The goal is to obtain the best design that will optimally provide the unbiased estimation of treatment contrasts with maximum efficiency. There are many optimality criteria that have been studied to accomplish this goal (Atkinson and Donev, 1992). Kiefer(1975) noted that if a binary balanced incomplete block design(BIBD) exists, then it is universally optimal for simultaneously estimating all the treatment contrasts. However, once any of the observations in a BIBD becomes unavailable, its optimality properties are lost. Many researchers have studied the robustness of incomplete block designs when some observations are unavailable(Ghosh, 1982a; Baksalary and Tabis, 1987; Whittinghall, 1986 and 1989). This particular problem is very common in on-farm trials. However, the efficient design of IBD is capable of handling such problems. In this case the selection of the most optimal design that suits the on-farm trial is of great benefit.

3.2.1 Optimal design theory

Classical formulation of optimal design theory is based on a model given as:

$$y_i = \sum_{j=1}^{p} f_j(x_i)\beta_j + \epsilon_i \tag{3.6}$$

for $i = 1, \ldots, n$;

where f_j are known functions of design points x_i, and $\beta_1, \ldots, \beta_p$ are unknown coefficients and ϵ_i are uncorrelated errors with mean 0 and common variance σ^2. The equation (3.6) can be written in the form

$$Y_n = F_n\beta + \epsilon \tag{3.7}$$

Notice that in this case the information matrix is directly proportional to $F_n^T F_n$. The information matrix can be written in the form

$$\frac{1}{n}F_n^T F_n = \int_{\chi} f(x)f(x)^T d\xi_n(x)$$

where ξ_n is the discrete measure that places mass $\frac{1}{n}$ at each x_i (Atkinson and Donev, 1992). Define

$$M(\xi) = \int_{\chi} f(x)f(x)^T d\xi(x) \tag{3.8}$$

for any positive measure ξ. Optimal design criteria are of the form:

Choose ξ to minimize $\Psi\{M(\xi)\}$ for some function $\Psi\{.\}$.

The examples include:

- D-optimality: $\Psi = -log|M(\xi)|$.
- A-optimality: $\Psi = tr\{M(\xi)^{-1}\}$.
- E-optimality: Ψ is largest eigenvalue of $M(\xi)^{-1}$.
- G-optimality: $\Psi = max_{x\in\chi} d(x,\xi)$ where $d(x,\xi) = f(x)^T M(\xi)^{-1} f(x)$.

Note that

$$\int_{\chi} d(x,\xi)d\xi(x) = p.$$

Thus if we can find ξ^* for which $max_{x\in\chi}d(x,\xi^*) = p$, ξ^* must be G-optimal.

General equivalence theorem

Suppose δ_x is a unit point mass at x and consider modifying ξ into

$$\xi^{'}_{\alpha,x} = (1-\alpha)\xi + \alpha\delta_x$$

where $0 < \alpha < 1$. Then

$$M(\xi^{'}_{\alpha,x}) = (1-\alpha)M(\xi) + \alpha M(\delta_x)$$

The derivative of Ψ, in the direction δ_x, is

$$\phi(x,\xi) = lim_{\alpha\downarrow 0}\frac{1}{\alpha}\{\Psi\{M(\xi^{'}_{\alpha,x})\} - \Psi\{M(\xi)\}\}$$

The General equivalence theorem asserts that the following are equivalent:

1. ξ^* minimizes $\Psi\{M(\xi)\}$,
2. $\phi(x,\xi^*) \geq 0$ for all x,
3. $\phi(x,\xi^*)$ achieves its minimum at points of the design, i.e at points x which have positive point measure under ξ^*.

Consider t treatments to be arranged in b blocks comprising a total of n experimental units. Any particular arrangement or design, d is associated with a $t \times b$ incidence matrix $N_d = n_{ij}$, where n_{ij} denotes the number of times the i^{th} treatment appears in the j^{th} block. Let the replication of treatment i be denoted by r_i and the number of units in block j be denoted by k_j. Writing R_d=diag$(r_1,\ldots,r_t)$ and K_d=diag$(k_1,\ldots,k_b)$, the least squares analysis of this design leads to the reduced normal equations given as(equation 2.15)

$$C_d\hat{\tau} = O$$

where $\tau = (\tau_1, \ldots, \tau_t)'$ is the vector of unknown treatment effects, Q is the vector of the adjusted treatment totals, and the information matrix C_d is given as

$$C_d = R_d - N_d K_d^{-1} N_d'$$

The matrix C_d is nonnegative definite and for connected designs it has rank $t-1$.

A connected block design d is said to be variance balanced if and only if all $t-1$ non-zero eigenvalues of C_d are equal(Kiefer, 1975). Let $\omega_1 \leq \omega_2 \leq \ldots \leq \omega_{v-1}$ be the nonzero eigenvalues of C_d.

Definition 2. *A design d is said to be Φ_p-optimal over the class of designs D if it minimizes*

$$\phi = \left\{ \sum_{i=1}^{t-1} \omega_i^{-p} \right\}^{1/p}$$

over $d \in D$.

For $p = \infty$ the above definition becomes a maximization of ω_1, and the optimal design is then said to be E-optimal (Dey and Das, 1989). While for $p = 1$ the common term is A-optimality(Kiefer, 1959). Another measure based on this result is the conventional efficiency ξ which is defined as

$$\xi = \frac{(t-1)\phi}{r - \phi} \tag{3.9}$$

where ϕ is defined as above for $p = 1$, r is the number of treatment replications, and t is the number of treatments. The efficiency ξ is the ratio of the average variance of all pairs of treatment differences to the minimum which would be achieved by a randomized block design, if one existed.

Consider the model given in (2.4)

$$y_{iim} = \mu + \tau_i + \beta_i + \varepsilon_{iim} \tag{3.10}$$

Where $i = 1, 2, ..., t; j = 1, 2, ..., b; m = 0, 1, 2, ..., n_{ij}$;and where n_{ij} denotes the number of experimental units in block j assigned to treatment i , μ denotes the general mean, τ_i the effect of the treatment i, β_j the effect of block j, and ϵ_{ijm} the random error.

Let $D=D(t, b, n, k_m, r_p)$ be a class of connected block designs having t treatments, n experimental units, minimum replication of treatments, r_p, and with maximum block size of k_m.

Definition 3. *A design d^* in a class D of competing designs is said to be E-optimal in D if and only if the smallest non-zero eigenvalue of C_{d^*} is at least as large as that of C_d for any other $d \in D$. Thus d^* is E-optimal if and only if it minimizes the maximum variance of the least square estimators of normalized treatment contrasts.*

Model (3.10) can also be expressed in matrix form as follows;

$$Y = X\beta + \varepsilon \tag{3.11}$$

where $X_{n\times p}$ is design matrix, $\beta_{p\times 1}$ is a vector of regression parameters, $Y_{n\times 1}$ vector of observations and $\varepsilon_{n\times 1}$ is a vector of error terms. We assume that ε is *iid* normal with mean zero and $cov(\varepsilon) = \sigma^2 I$. Thus, from the least squares estimate we have that

$$\hat{\beta} = (X'X)^- X'Y \tag{3.12}$$

where $var(\hat{\beta}) = \sigma^2(X'X)^-$, $\hat{Y}_x = x\hat{\beta}$ and the $var(\hat{Y}_x) = \sigma^2 x(X'X)^- x'$ where $(X'X)^-$ is the generalized inverse of $(X'X)$ if $X'X$ is not of full rank, otherwise $(X'X)^-$ is replaced by the unique inverse $(X'X)^{-1}$. Notice that the variance-covariance matrix of the vector of parameter estimates of β in a least squares analysis is proportional to $(X'X)^{-1}$. Therefore an efficient design is one with a small variance matrix, and the eigenvalues of $(X'X)^{-1}$ provide measures of its size. Thus the design problem involves selecting row vectors $x_{1\times p}$ for all y_i, $i = 1, 2, ..., n$ from a given design space φ such that the design defined by these n vectors is in one way or another optimal. Solutions to this problem involve formulating a criterion based on the above model and using it to obtain optimal designs. There are many criteria that have been developed to enable comparison of experimental designs(Kiefer, 1959). These optimality criteria are based on minimizing the variance of the estimates of the fixed effects and the variance components.

3.2.2 G-optimality criterion

Smith (1918) developed a criterion to obtain optimal designs for regression problems, based on minimization of the maximum variance of any predicted value over the experimental space. Kiefer and Wolfowitz (1960) called this the G-optimality criterion from which we derive the G-efficiency design measure. G-efficiency is a common criterion for optimal design. It is based on the variance of prediction of the candidate points, which is proportional to $(X'X)^{-1}$. Thus, this criterion is related to the information matrix $X'X$. Minimizing the average prediction variance leads to the I-optimality, where "I" denotes integration over the candidate space(Kiefer, 1959). In general the G-efficiency is defined as

$$G - eff = \left\{ \sqrt{\frac{p/N_d}{max_{x \in C} x'(X'X)^{-1}x}} \right\} \times 100 \tag{3.13}$$

where p is the number of parameters in the linear model, N_d is the number of design points and C is a set of candidate points

3.2.3 D-optimality criterion

Wald (1943) proposed the D-optimality criterion which puts more emphasis on the quality of the parameter estimates. A design is said to be D-optimal if for the model the joint confidence region for the vector of unknown parameters is minimum. As a consequence, the global maximum of the determinant of $X'X$ is achieved. D-efficiency is a function of the geometric mean of the eigenvalues , which is given by $|(X'X)^{-1}|^{1/p}$ i.e. the determinant $|(X'X)^{-1}|$ is the product of the eigenvalues of $(X'X)^{-1}$, and the pth root of the determinant is the geometric mean. Thus a design d^* is said to be D-optimal if it has the minimum value of the determinant of $(X'X)^{-1}$. In general D-efficiency is defined as

$$D - eff = \left\{ \frac{| X'X |^{1/p}}{N_d} \right\} \times 100 \tag{3.14}$$

where p and N_d are defined as in (3.13).

3.2.4 A-optimality criterion

The A-optimality criterion minimizes the average variance of the parameter estimates (Chernoff, 1953). i.e. min trace$((X'X)^{-1})$. A-efficiency is a function of the arithmetic mean of the variances, which is given by $trace((X'X)^{-1})/p$. Notice that trace is the sum of the diagonal elements of $(X'X)^{-1}$, which is the sum of the variances and also the sum of the eigenvalues of $(X'X)^{-1}$. In this context A-efficiency is the most logical criterion to use in evaluating design goodness. This is because as orthogonality decreases, both the off-diagonal and diagonal elements of $(X'X)^{-1}$ increase. Thus considering the average variances while ignoring the off-diagonal covariances, is reasonable since variances increase as the covariances increase. In general A-efficiency is given as

$$A - eff = \left\{ \frac{p/N_d}{trace(X'X)^{-1}} \right\} \times 100 \tag{3.15}$$

where p and N_d are defined as above in (3.13)

3.2.5 E-optimality criterion

Another criterion is the E-optimality, which finds a design which maximizes the minimum eigenvalues of $(X'X)$. In other wards, a design d^* is said to be E-optimal if it has least value for λ_{max} where λ_{max} is the maximum eigenvalue of $(X'X)^{-1}$; where $(X'X)^{-1}$ is the inverse of $(X'X)$.

The above criteria can be selected on the basis of the objective of the study. The unpredictable nature of on-farm trials would make one wish to recommend a general optimality criterion. But such criteria are not available and therefore we compare across all criteria to obtain an appropriate design. An illustrative example is presented to show the use of the above optimality criteria.

Example 1

Consider the situation of an on-farm experiment where we are interested in BIB optimal design for 6 farms(blocks) for testing 7 treatments. Suppose that there is a resource constraint such that only 24 experimental units can be considered. Using the SAS procedure OPTEX we can generate and compare such designs. SAS code for the above example is provided in the appendix C.1.

Table 3.1: Results of optimality Efficiency criteria for the on-farm trial considered with $b = 6$, $t = 7$, $k = 4$, and $N = 24$

Design Number	D-efficiency	A-efficiency	G-efficiency	Average prediction Standard error
1	92.8850	85.9649	82.8417	0.7626
2	92.8850	85.9649	82.8417	0.7626
3	92.8850	85.9649	82.8417	0.7626
4	92.8850	85.9649	82.8417	0.7626
5	92.8850	85.9649	82.8417	0.7626
6	92.8850	85.9649	82.8417	0.7626
7	92.6860	85.2011	81.4257	0.7661
8	92.6860	85.2011	81.4257	0.7661
9	92.5877	84.8357	81.4627	0.7677
10	92.5668	84.7444	80.0776	0.7681

The order of the designs range from 1 (most efficient design) to 10 (least efficient design). Note that the efficiency values obtained are in comparison to an orthogonal randomized complete block design. The D-efficiency value remained constant from design 1(92.885) to 6 and decreased slightly to 92.686 for designs 7 and 8 and thereafter dropped to 92.5877 for design 9 and finally to the lowest value of 92.5668 for design 10. The reason for the differing values of these measures across designs is due to the different treatment combinations that can result. Each treatment combination has an effect on the quality of the parameter estimates as mirrored in the design matrix. Other design optimality measures present a similar pattern of performance across all 10 designs. Table 3.2 presents results on the best design obtained by the OPTEX SAS procedure.

Table 3.2: Best design obtained by the OPTEX SAS procedure

Blk/farm 1	Blk/farm 2	Blk/farm 3	Blk/farm 4	Blk/farm 5	Blk/farm 6
7	7	7	6	5	6
6	6	5	5	4	3
5	4	3	4	3	2
2	3	1	1	2	1

The above example illustrates the benefit of using IBD which have less restrictions than RCBD. In this example, the best IBD results in a loss of efficiency compared to a RCBD of 7%, 14% and 18% on the D-efficiency, A-efficiency and G-efficiency criteria respectively and has an average prediction standard error of 0.7626.

3.3 Complexity criterion

The choice of a design is an important determinant of the properties in linear estimation. Thus the design itself is an essential aspect of model selection. Maklad and Nichols(1980) noted that complexity can be a good criterion for use in model selection. Emden(1971) shows that complexity maybe used effectively as a tool in design evaluation. The complexity of a design essentially gauges the nonorthogonality of the design matrix X, as mirrored through the inverse of $X'X$. Ramirez(1989) studied in detail two complexity indices.

Consider a random variable $Y = [Y_1, ..., Y_k]'$ with a distribution $\Omega(Y)$ and a dispersion matrix Λ, then the complexity of $\Upsilon(Y)$ was defined by Emden(1971) in terms of the dispersion matrix Λ as

$$\phi(\Lambda) = \left\{ k \ln(\frac{tr(\Lambda)}{k}) - \ln(|\Lambda|) \right\} /2 \tag{3.16}$$

Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k > 0$ denote the ordered non-zero eigenvalues of Λ. Let $\bar{\lambda}$ be the average of all the eigenvalues(arithmetic mean) and $GM(\lambda)$ denote their geometric mean.

Then

$$\phi(\Lambda) = \left\{ ln(\frac{\bar{\lambda}}{\prod_{i=1}^{k} \lambda_i}) \right\} /2 = (k/2) \ln \left\{ \frac{\bar{\lambda}}{GM(\lambda)} \right\} \tag{3.17}$$

Note that $\phi(\Lambda) \geq 0$ by the arithmetic-geometric mean inequality and $\phi(\Lambda) = 0$ iff Λ is a scalar matrix. Thus $\phi(.)$ can be used to gauge whether a design X yields Gauss Markov estimators $\hat{\beta}(X)$ more precisely than another design H. In order to compare designs on the basis of their relative complexity , we drop the logarithmic scale and say that design X is less complex than a design H whenever $C_1(X, H) \leq 1$, where

$$C_1(X, H) = \frac{tr(\Sigma)}{tr(\Omega)} \left\{ \frac{|\Omega|}{|\Sigma|} \right\}^{1/k} \tag{3.18}$$

where $\Sigma = [X'X]^{-1}$ and $\Omega = [H'H]^{-1}$.

There is a link between the standard design criteria of A- and D-efficiency and the relative complexity of two designs i.e. A-efficiency is related directly, and D-efficiency is inversely related to C_1 complexity. The link between these criteria is further established by the following theorem.

Theorem 1. *Consider designs X and H, with $\Sigma = [X'X]^{-1}$ and $\Omega = [H'H]^{-1}$, pertaining to the Gauss Markov estimators $\hat{\beta}(X)$ and $\hat{\beta}(H)$, respectively.*

1. *Then $\phi(\Sigma)$, $\phi(\Omega)$ and $C_1(X, H)$ are related as*

$$\phi(\Sigma) - \phi(\Omega) = (1/2) \ln \left\{ \frac{(tr(\Sigma))^k |\Omega|}{(tr(\Omega))^k |\Sigma|} \right\} = (k/2) \ln(C_1(X, H)) \tag{3.19}$$

2. *If X and H are D-equivalent, then H has greater A-efficiency than X if and only if H is less complex than X.*

3. *If X and H are A-equivalent , then H has greater D-efficiency than X if and only if H is more complex than X.*

3.3.1 Numerical example

Consider an on-farm trial to test for 3 treatments on 3 farms. The model for such a trial would be

$$Y_{ij} = \mu + \tau_i + f_j + \epsilon_{ij} \tag{3.20}$$

$i = 1, 2, 3$ and $j = 1, 2, 3$; where Y_{ij} is the observation for the ith treatment in the jth farm, μ is the overall mean,τ_i is the ith treatment effect, and f_j is the jth farm effect. Consider two designs a complete block design and an incomplete block design given in table 3.3.

Table 3.3: Complete block design 1 and Incomplete block design 2

Complete block design 1			Incomplete block design 2		
Farm 1	Farm 2	Farm 3	Farm 1	Farm 2	Farm 3
1	1	1	1	-	1
2	2	2	2	2	-
3	3	3	-	3	3

The design matrix X for each of these examples are given in the appendix A.1. Using the X matrix we obtain the eigenvalues of the $X'X$ matrix. We then compute the arithmetic mean, geometric mean and thereafter the complexity index using 3.17. The results of the computations are presented in table 3.4.

Table 3.4: The eigenvalues, arithmetic and geometric means, and complexity index for the two designs using the model 3.20.

Design Number	Eigenvalues for $(X'X)$	Arithmetic Mean	Geometric mean	Complexity Index
1	15,3,3,3,3	5.4	1.069	4.0492
2	10,3,3,1,1	3.6	1.226	2.6929

In this example design two is less complex(2.6929) as compared to design one. This is because in design 2 the blocks are incomplete but connectedness is maintained. While design 1 is a complete block design with full connectedness, its complexity index compared

to an incomplete connected design is greater. This provides an insight that incomplete block designs can be as efficient or even more efficient than randomized complete block designs.

3.4 Power analysis in designs

In many situations in analyzing designs, there is need to compute the power of a design as compared to other designs in answering a specific objective. Very often RCBD are used in experiments where another design would be appropriate. O'Brien and Lohr(1984) developed a method using ordinary least squares linear models to obtain power values for designs. They used SAS PROC GLM to obtain non-centrality parameters of a non-central F under departures from the null hypothesis of no treatment differences which was the aim of the study. Their study did not cover the aspects of spatial variations that occur in fields since the designs were restricted to a single source of identically independent distributed experimental errors. In the on-farm trials, spatial variability is a common occurrence and more specifically with the use of incomplete block designs, where recovery of inter-block information is required, power analysis is a vital tool in assessing and comparing designs.

3.4.1 Basis for power analysis

The power analysis in design is considered under a mixed model obtained by modifying (2.3) such that

$$y = X\beta + Zu + e \tag{3.21}$$

where

y is a response vector,

X is a design matrix for fixed effects,

β is the vector for the fixed effects parameters,

Z is the design matrix for the random effects,

u is the vector of random effects,

e is the vector of residuals.

In mixed models theory $y \sim MVN\{X\beta, V\}$ where $V = ZGZ' + R$ and

where

$$\begin{Bmatrix} u \\ e \end{Bmatrix} \sim MVN \begin{Bmatrix} G & 0 \\ 0 & R \end{Bmatrix}$$

In mixed models analysis the null hypothesis is H_0:$K'\beta = 0$ where $K'\beta$ is estimable and this can be tested using the generalized F-statistic

$$F = \frac{(K'b)'[K'(X'V^{-1}X)^{-1}K](K'b)}{r(K)} \tag{3.22}$$

where b is the estimate of β, $r(K)$ denotes the rank of K and V is replaced by its estimate depending on the variance-covariance matrix form of G and R. F in equation (3.22) is distributed approximately $F_{r(K),v,\lambda}$. The denominator degrees of freedom(d.f) v, are the d.f to estimate $K'(X'V^{-1}X)^{-1}K$. And λ is the non-centrality parameter which is given as

$$\lambda = (K'\beta)'\{K'(X'V^{-1}X)^{-1}K\}^{-1}(K'\beta) \tag{3.23}$$

Under H_0, $\lambda = 0$ and if H_0 is false $\lambda > 0$. Thus the actual value of λ depends on the design and replication associated with X, $K'\beta$ and V the variance-covariance components.

Power is then defined as the $Prob[F_{r(K),v,\lambda}] > F_{crit}$; where $F_c = F_{r(K),v,0,\alpha}$, the value of the central F at the designated α-level. λ- is the non-centrality parameter under the alternative hypothesis of interest. In evaluating designs, interest is in assessing the expected precision of competing designs rather than the expected power(Mead, 1994). The variance of the estimate of an estimable function in mixed models is given as

$$Var(K'\beta) = K'(X'V^{-1}X)^{-1}K \tag{3.24}$$

Therefore, when comparing designs, we use precision with which they can be expected to estimate functions $K'\beta$ deemed to be of major interest. The idea of designing experiments involves identifying the objectives, treatment structure and formulation of the appropriate sets of contrasts to address the problem or objectives. Use of power analysis

can then be employed to select the most appropriate and best design. Example 2 given below illustrates the procedure of power analysis in design selection.

Example 2

Consider an on-farm trial to test 6 treatments(t) on 6 farms(blocks) which can only accommodate a maximum of $k = 4$ treatments each. Thus the maximum number of experimental plots would be $4 \times 6 = 24$ experimental plots. Therefore, this causes a restriction on the number of treatments to allocate to a given farm since they can only accommodate a maximum of 4 treatments.

Probable designs would be in the class of incomplete block designs e.g. BIBD, PBIBD, RCBD with distorted naturally existing structure etc. Tables 3.5, 3.6 and 3.7 represent the PBIB, BIBD and RCBD designs considered respectivcly.

Table 3.5: Design 1: Partially balanced incomplete block design(PBIB) layout with 4 treatments per farm

farm 1	farm 2	farm 3	farm 4	farm 5	farm 6
1	1	4	6	5	3
2	2	5	3	6	6
3	3	6	1	4	5
4	5	1	2	1	4

Table 3.6: Design 2: Balanced incomplete block design(BIBD) layout with 4 treatments per farm

farm 1	farm 2	farm 3	farm 4	farm 5	farm 6
2	1	3	4	6	4
1	2	1	6	4	6
4	6	2	3	5	1
3	5	5	5	2	3

Table 3.7: Design 3: Randomized complete block design(RCBD) layout with 4 treatments per farm

farm 1	farm 2	farm 3	farm 4	farm 5	farm 6
2	1	3	4	6	4
1	2	1	6	4	6
4	6	2	3	5	1
3	5	5	5	2	3
5	3	4	2	3	5
6	4	6	1	1	2

The RCBD design in table 3.7 shows a constraint where complete block designs are not possible. This implies that the treatments in lower portion of the table cannot be accommodated in the design implementation.

Assuming that the variance among farms is given as $\sigma_f^2 = 3.5$ and the variance between plots within a given farm is $\sigma_{pf}^2 = 5$. We can then evaluate these designs using PROC MIXED in SAS. Using simulated data we are able to assess the power of these designs in estimating a given contrast i.e. (1, 0 ,-1, -1, 0 ,1). Results from the simulation study are presented in table 3.8. The SAS procedure used in the case are given in Appendix C.2.

Results from the SAS output

Table 3.8: Results for power estimate for testing the contrast(1, 0 ,-1, -1, 0 ,1)

Design	ndf	ddf	alpha	fcritical	ncparm	F value	$Pr > F$	power
PBIBD	1	13	0.05	4.66719	10.27	10.27	0.0069	0.84157
BIBD	1	13	0.05	4.66719	11.4	11.40	0.005	0.87669
RCBD	1	13	0.05	4.66719	5.53	5.53	0.0268	0.58558

where ndf is the numerator degrees of freedom, ddf is denominator degrees of freedom, alpha is level of significance, fcritical is the F-critical observed value, ncparm is the product of ndf and F.

From table(3.8) we notice the discrepancy between the incomplete block designs and the randomized complete block design as far as power of a design is concerned. In this particular scenario RCBD is less appropriate(Power=0.58558) as compared to BIBD and PBIBD with power of 0.87669 and 0.84157 respectively. In this example power analysis demonstrates how designs can be assessed on their application for a given situation and constraint.

Power analysis gives the prior probability of detecting an expected relevant difference between treatments. This would involve use of previous knowledge about the experiment and the subject matter under experimentation.

Chapter 4

Stability Status of Designs

Optimal incomplete block designs are particularly important for most agricultural experiments since they can provide the maximum information from an experiment. Cheng(1978) provides much information on the designing and analyzing these designs. The actual design of experiments is aimed at optimizing certain characteristics of the statistical procedures to be used which depend on the settings for the experimental conditions experienced(Schwabe, 1996). Designing on-farm trials is more complicated than designing on-station trials (Hilderbrand and Russell, 1996). There are many factors of influence in on-farm trials and these interact in many different ways and this makes designing such experiments even more difficult(Coe, 1998). Possible confounding factors may arise for a well designed on-farm trial due to various reasons. Missing observations and breakdown in communication between the farmer and the researcher are some of the factors which may occur during implementation of the trial. These are some of the factors that affect design stability. The stability status of the design is assessed through consideration of missing plot evaluation, and robustness of the resulting design.

Kiefer(1975) dealt with general criteria in the block design setting. In his paper, the universal optimality of balanced block designs is adequately discussed. There many ways in which a design loses its optimality; but mostly it is through distortion of a design either by loss of some treatments(missing observations). However, it is important to know how

much efficiency is lost when optimality distortions occur. This is particularly important since the information provided by a design largely depends on how efficient a design will he even after loss of some treatments. It is therefore important to study the effects of missing observations on the design optimality of incomplete block designs.

4.1 Analysis of the effect of Missing plots

Unlike on-station trials, a common problem in most on-farm trials is the unavailability of some observations for analysis(Lockeretz, 1987). In some cases an entire block may be lost for reasons not related to the structure of the experiment. Since missing observations are a real possibility in on-farm trials, it is very important to select a design that would minimize the influence of these missing observations on the efficiency of the design selected. Studies concerning the loss of a single treatment, loss of whole blocks and the loss of any number of treatments in a single block have been done(Ghosh,1982a,b; Whittinghill, 1989; Dey, 1993; Prescott and Mansson, 2001; Most, 1975; Das and Kageyama, 1992; Bhaumik and Whittinghill, 1991 and Gupta and Srivastava, 1992). From a practical point of view it is unlikely that a researcher or farmer has any influence over which treatments become missing.

However, if missing observations are a common occurrence due to unforeseen events, then it is best for the researcher to choose an initial design with good robustness properties to guard against the ruination of the experiment (Prescott and Mansson, 2001). Hedayat and John(1974) developed resistant balanced incomplete block designs which remain variance balanced even if all observations of a particular treatment become unavailable before the analysis. Ghosh(1982a) proved that a BIBD with r replications is robust to the unavailability of any $r-1$ observations as well as any $r-1$ blocks. This is because the resulting design remains connected with respect to treatment. Since we are interested in the treatment effects, we look at the C_d matrix which is the treatment information matrix for a general block design d which is a $t \times t$ matrix given as in (3.2). A general

block design d consists of t treatments allocated to plots in b blocks. The number of plots in each block is given by $K = (k_1, \ldots, k_b)'$. Define $r_d = (r_1, \ldots, r_t)$ as the vector of replications of treatments in the design d, and $N_d = n_{ij}$ is the $t \times b$ incidence matrix whose elements n_{ij} equal the number of times treatment i appears in block j, $i = 1, \ldots, t$ and $j = 1, \ldots, b$. The treatment information matrix for a general block design $t \times t$ can be expressed as

$$C_d = r_d^{'} - N_d K^{-1} N_d^{'} \tag{4.1}$$

where $r_d^{'} = diag(r_1, \ldots, r_t)$ and $K^{-'} = diag(1/k_1, \ldots, 1/k_b)$. The matrix C_d is symmetric, nonnegative definite, and bas row sums equal to zero. When the design is connected, C_d has rank $t-1$ and its resultant eigenvalues can be given by the vector $\omega = (\omega_1, \omega_2, \ldots, \omega_t)'$ where ω_i is the *ith* largest eigenvalue and $\omega_t = 0$. C_d can be expressed in canonical form (Prescott and Mansson, 2001) as

$$C_d = \sum_{i=1}^{t-1} \omega_i z_i z_i^{'} \tag{4.2}$$

where z_i is the normalized eigenvector corresponding to ω_i such that $z_i^{'} z_i = 1$ and $z_i^{'} z_j = 0$ for $i \neq j$. To obtain the treatment estimates $\hat{\tau}$ of $\tau = (\tau_1, \ldots, \tau_t)$, we need to have the generalized inverse of C_d say G such that $C_d G C_d = C_d$. Prescott and Mansson(2001) give this inverse as

$$G = \sum_{i=1}^{t-1} \frac{1}{\omega_i} z_i z_i^{'}.$$

This implies that $\hat{\tau} = G(T - N_d K^{-'} B)$ where T and B are the vectors of the treatment and the block totals respectively. Also $var(\hat{\tau}) = G\sigma^2$. In assessing the robustness of designs to missing observations, use is made of the pairwise treatment comparisons based on $var(\hat{\tau_{i1}} - \hat{\tau_{i2}})$ for $i_1, i_2 = 1, \ldots, t$ and $i_1 \neq i_2$. If Υ is a contrast matrix of dimension $(t(t-1)/2) \times t$ identifying these differences, then the diagonal elements of the $\Upsilon G \Upsilon' \sigma^2$ are the variances of these contrasts. Then average variance of the pair-wise treatment differences(Prescott and Mansson, 2001) is given by

$$Ave\left\{var(\hat{\tau_{i1}} - \hat{\tau_{i2}})\right\} = \frac{tr(\Upsilon G \Upsilon')\sigma^2}{t(t-1)/2} = \frac{2tr(\Upsilon G \Upsilon')\sigma^2}{t(t-1)} \tag{4.3}$$

In this context, for a balanced incomplete block design, $k_j = k$ for all j and $r_i = r$ for all i such that the information matrix of the treatment effects for this design reduces to

$$C_d = \frac{\lambda t}{k}\left\{I - \frac{J}{t}\right\}$$

where J is the matrix of all 1's and λ is the number of times that treatment i and j appear together in the same block, and I is the identity matrix. Thus C_d has $t-1$ non-zero eigenvalues all equal to $t\lambda/k$. Since $\left\{I - \frac{J}{t}\right\}$ is idempotent, the generalized inverse of C_d denoted by G is given by

$$G = \frac{k}{\lambda t}\left\{I - \frac{J}{t}\right\}$$

For a BIBD all estimates of the treatment differences have variance $2\sigma^2 k/\lambda t$ and the estimated pair-wise treatment difference for a completely randomized design with t treatments and r replications is given by $2\sigma^2/r$. The efficiency factor of the BIBD is $\lambda t/rk$. In the event of loss of some observations, the new properties of the resulting design largely depend on the specific configuration of the missing observations(Whittinghill, 1986). Some variances of the pair-wise treatment differences will be increased while others may not. This change affects the eigenvalues of the original design and thus the robustness of the whole resultant design. In the evaluation of these designs we use several approaches.

4.1.1 The average variance and the relative efficiency of the resulting designs

For t_i observations missing, the sum of the eigenvalues of the information matrix is given as $\frac{t(t-1)\lambda}{(k-t)}$ and its only tbe t_i eigenvalues that are affected but not ω_t which is always equal to zero. Thus the average variance expressed in terms of the non-zero eigenvalues of the information matrix is given as the sum of the eigenvalues of its generalized inverse. For a given design d the average variance(Whittinghill, 1989) is given as

$$Avevar(\hat{\tau_{i1}} - \hat{\tau_{i2}}) = \frac{2}{t-1}\sum_{i=1}^{t-1}\frac{1}{\omega_i(d)}\sigma^2 \qquad (4.4)$$

Therefore the relative efficiency(R) of two designs d_1 and d_2 is defined as the ratio of their average variances which can be given as

$$R = \frac{\sum_{i=1}^{t-1} \frac{1}{\omega_i(d_2)}}{\sum_{i=1}^{t-1} \frac{1}{\omega_i(d_1)}} \tag{4.5}$$

R is compared to 1 whereby if $R > 1$, then design 1 is better than design 2. While if $R < 1$ then design 2 is better than 1. In the event that $R = 1$ then the two designs are equally good.

4.1.2 Mini-max variability criterion

The min-max variability method given by Whittinghill(1989) compares the variability V_d of the $t-1$ non-zero eigenvalues of the resultant C-matrix of design d introduced in section 4.2 where

$$V_d = \frac{\{\sum_{i=1}^{t-1}(\omega_i - \bar{\omega})^2\}}{(t-2)} \tag{4.6}$$

where $\bar{\omega}$ is the mean of the eigenvalues ω_i for $i = 1, \ldots, t-1$. Therefore, using this measure of variability, a resulting design d_1 is said to be better or more nearly variance balanced than a resulting design d_2 if V_{d_1} is nearer to 0 than V_{d_2}. When d is variance balanced, its V_d=0.

In the on-farm trial situation, where the problem of missing plots is fairly common, the criteria discussed above could be applied in order to compute standard errors and evaluate design efficiency.

Presence of outliers and missing observations or blocks are some of the problems likely to be encountered in on-farm trials especially when data has to be recorded by the farmer. BIBD is known to have robustness. The discussion in the following section establishes a further role of BIBD in the case of such problems.

4.2 Robustness of balanced incomplete block designs

Design problems are often encountered even for well-planned on-farm trials. These range from missing observations, presence of outliers, loss of all observations pertaining to a

given treatment, etc. This results in poor performance of an optimal design and loss in efficiency. The idea of robustness is explained further in the following definition.

Definition 4. *A design d is said to be robust against one or more of the above distortions if it remains insensitive to the presence of one or more of the above distortions in terms of design properties such as connectedness, variance balance, efficiency, optimal properties, etc.*

Connectedness of block designs is a vital property since it ensures that estimability of all paired differences among treatment effects is possible. Balanced incomplete block designs have a special property of complete connectedness. Therefore, for these designs, the loss of some plots may result in the design losing its connectedness with respect to treatments and it will no longer be a BIBD. The following definitions elaborate more on the robust and strongly robust cases of a BIBD.

Definition 5. *A BIBD(t, b, k, r, λ) is said to be robust against the unavailability of any q observations if the block design obtained by omitting any q observations remains connected with respect to treatment.*

Definition 6. *A BIBD(t, b, k, r, λ) is said to be strongly robust against the unavailability of any q observations if the block design obtained by omitting any q observations remains completely connected.*

Ghosh(1982b) shows that a BIBD is robust against the unavailability of any $r - 1$ observations. However, a BIBD is not robust against the unavailability of any r observations. This is because if all r observations corresponding to a particular treatment are lost then the treatment will be disconnected from the other treatments. Therefore, BIBD is robust against the unavailability of any q observations so long as q is less than $r - 1$. The robust and strongly robust properties are summarized in theorem 2 and 3 below.

Theorem 2. . *A BIBD(t, b, k, r, λ) is robust against the unavailability of all observations in any $r - 1$ blocks.*

Theorem 3. . *A BIBD (t, b, k, r, λ) is strongly robust against the unavailability of any $k - 1$ observations.*

Note that a BIBD is not strongly robust against the unavailability of any k observations. The maximum number of observations q that should be unavailable in order for

the BIBD not to be strongly robust is $k-1$.

According to Prescott and Mansson(2001), missing observations affect a design by their effect on the eigenvalues of the BIB design and thereby affect the average variance and the relative efficiency of the design as compared to the complete block design. They note that in a situation where one observation is missing, the properties of the resultant design are the same regardless of the type of treatment lost as well as its position in the initial design. In such a case, the specific variances of the pairwise treatment differences depend on the type of treatment lost for that specific observation, but the average and the maximum variance will be unchanged. For the case of two missing observations, it is rather complicated since it calls for different configurations, each with different eigenvalues, to be considered. Let g be the number of treatments common to the pairs of blocks. There are different possible cases in the loss of two observations that exist (Prescott and Mansson, 2001). Some of these include:

1. Different treatments in different blocks where neither treatment is common to the two blocks.

2. Different treatments in different blocks, with one of the two treatments common to both blocks.

3. Two treatments lost in the same block.

4. Both treatments are different and occur in both blocks.

5. Two replicates of the same treatment missing. etc

For all such situations, two eigenvalues are affected. The two eigenvalues affected take on values given as $\omega_{t-2} = \frac{t\lambda}{k} - 1 + x$ and $\omega_{t-1} = \frac{t\lambda}{k} - 1 - x$. Table 4.1 presents the results on average variance and relative efficiency for a single missing observation and two missing observations.

Table 4.1: Summary of cases involving loss of a treatment(s) in a design

Observations missing	Average Variances	Relative efficiency
One missing Observation $(t-1)$	$\frac{2k\sigma^2}{t\lambda} + \frac{2k^2\sigma^2}{t(t-1)\lambda(t\lambda-k)}$	$\{1+\frac{k}{(t-1)(t\lambda-k)}\}^{-1}$
Two missing observations $(t-2)$ $\omega_{t-1}=\frac{t\lambda}{k}-1+x$	$\frac{2k\sigma^2}{t\lambda} + \frac{4k^2(t\lambda-k+kx^2)}{t(t-1)\lambda\{(t\lambda-k)^2-k^2x^2\}}$	$\{1+\frac{2k(t\lambda-k+kx^2)}{(t-1)\{(t\lambda-k)^2-k^2x^2\}}\}^{-1}$

For the case $\omega_{t-1} = \frac{t\lambda}{k} - 1 - x$ the x values and the number of configurations can take on diffcrent values.

Table 4.2: Summary of value of x and number of configurations

Case	Value of x	Number of configurations
1	$x = g/\{k(k-1)\}$,for $g = 0, \ldots, k-1$	$(k-g)^2$
2	$x = (k-g)/\{k(k-1)\}$,for $g = 1, \ldots, k-1$	$2g(k-g)$
3	$x =$ ●	$bk(k-1)/2$
4	$x = (2k-g)/\{k(k-1)\}$,for $g = 2, \ldots, k$	$g(g-1)$
5	$x = (k^2-2k+g)/\{k(k-1)\}$, for $g = 1, \ldots, k$	g

In general the smallest loss in efficiency occurs in cases 3 and a section of case 1 when the $g = 0$. The worst scenario is where two replicates of the same treatment are lost from different blocks(case 5). In general the loss in efficiency will largely depend on the experimental design used and how the treatments are laid out in this design. The incorporation of prior knowledge at the design and implementation stage can be a vital tool in improving the design and minimizing efficiency losses.

Treatments tested in on-farm trials can take different forms depending on the researchers interest. Treatments of factorial structure may be used by researchers especially soil fertility scientists. Maintaining such a structure may lead to large set of treatments which could cause application problems in on-farm trials. The optimality of factorial treatment structure in incomplete block design is discussed in the next section.

4.3 Factorial treatment structure in incomplete blocks

Many factorial and fractional factorial designs have been applied in agricultural experiments. These designs are very common in soil fertility, agronomic and animal production trials. Because not all interaction levels can be run in these experiments, use of fractional factorial designs has gained considerable attention. The work by Yates(1935) on factorial trials addresses the necessity of proper blocking in experimentation. Proper blocking greatly increases the precision of experimental designs(Kempthorne, 1947). Farms or farms in regions are a common blocking factor in most on-farm trials. Therefore, in on-farm trials, farm*treatment factor interactions need due attention. Yates(1935) address this problem on differential responses in different blocks in agricultural experiments. The interpretation of most factorial designs with a high degree of confounding or fractional replication depends on the assumption that block-treatment interactions are negligible. This is however not the case in on-farm trials since the regions are expected to contribute an environmental effect and the different management practices in different farms are expected to have an effect as well. In fact the investigation of farm-treatment factor interactions is often one of the objectives of an on-farm trial.

4.3.1 Choosing an IBD for a factorial treatment structure

Selecting an incomplete block design depends on

- The number of factors and treatment combinations involved.
- The size of the blocks
- The resources available to determine number of replications needed.

There are several lists of designs where one can easily obtain an incomplete block design. Montgomery(1997), Cochran and Cox (1957) are a few such references. The designs provided in these references do not cover the confounding structure of the block- treatment factor interactions. In on-farm trials, analysis of block-treatment factor interactions is an

important component. Thus using incomplete block designs for on-farm trials requires obtaining such information concerning these interactions. Therefore, the guiding principles in selecting incomplete block designs with a factorial treatment structure for on-farm trials, is the ability for the design to provide estimates for the farm-treatment factor interactions. Secondly, its important to select designs with blocks that are resolution III and above (Cochran and Cox ,1957; Mead, 1994). In such a situation, one will be able to obtain estimates of main effects for the farms/blocks. Such recommended designs are given in table 4.3.

Table 4.3: Factorial treatment designs with resolution III and above for different numbers of factors

Number of Factors	Farm size	Number of farms	Design resolution
3	4	4	III
4	8	2	IV
5	8	4	III
6	16	4	IV
6	8	8	III
7	16	4	IV

An on-farm trial involving 4 farms, each of size 4 accommodating 3 factors qualifies to be a design of resolution III. In such a trial its possible to estimate main effect and two factor interactions. Similarly a design with 4 farms of size 16, with 7 factors is of resolution IV. Researcher choice of design resolution depends on the higher interaction factors to be confounded. The choice of which depends on the important component to be investigated.

Analysis of factorial treatment structure in IBD

Steps for analyzing an incomplete block factorial trial.

- Factors should enter the model in the order of main effects, all two factor interactions and finally all block-treatment factor interactions. It is important to use the sequential sums of squares to analyze the data. This is because it allows for the interaction effect to be investigated free of the main effect.

- We examine the mean squares for the block-treatment factor interactions for any abnormal values. If any of the block-treatment effects are abnormal, then a reason must be established for the lack of consistence across the blocks.

- If all the block-factor interactions are small, then we can drop the effects from the model. Therefore analyzing such a trial would proceed as a typical factorial experiment.

Notice that a large mean square would imply that the effect of this factor is not consistent across the experiment. And therefore a plot of the main effects by block can help in understanding the changing nature of the main effects across blocks. The aforementioned idea is the result of the fundamental differences between expectation and variation models in respect to the behavior of marginal terms(Nelder, 1977). A more elaborate study on the design and analysis of factorial designs is given by Yates(1933), Plackett and Burman(1946) , Fisher(1942), Bose(1947) and Rao(1947).

In general, several methods have been suggested for the construction and selection of optimal incomplete block designs(Nguyen & Williams, 1993; Nguyen, 1994). Most of these methods are based on optimizing various criteria such as the A-,E-,D-optimality criteria. Huber and Zwerina(1996) identify four principles underlying the optimality of designs. These include level balance, orthogonality, minimal overlap and utility balance. These principles adequately define what makes a design efficient in relation to other designs. In this context, level balance refers to equal replications of each treatment in the experiment, minimal overlap is attained when the alternatives within each treatment combination choice set have non-overlapping treatment levels, and utility balance is attained when the utilities of alternatives within treatment combination choice sets are the same. Orthogonality is as defined by Yates(1933), i.e where the different classes of effects to which the experimental material is subject are capable of direct and separate estimation without any entanglement. Most of these methods are based on algorithms that optimize

design efficiency parameters in order to obtain an optimal incomplete block design.

4.4 Standard versus control treatments

In many on-farm trials it is not possible for experimental material to meet the requirements of a classical design. Designing an experiment is sometimes viewed as picking from a library of designs the design recipe which comes closest to fitting the particular situations of the experiment, and then making compromises in the objectives of the experiment and the structure of the experimental material in order to force the experiment into the recipe's requirements(Mead, 1994). Standard designs are vital, but how often their requirements are or can be naturally met in on-farm trials remains the question(Kuehl, 2000). Consider an on-farm situation of comparing t test treatments, denoted by $1, 2, ..., t$, with a standard treatment, denoted by O , in b blocks of size k . Assuming an additive model, the fixed effect model for a one- way elimination of heterogeneity is given as

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \tag{4.7}$$

where μ denotes the general mean , τ_i the effect of the ith treatment and β_j the effect of the jth block. ε_{ij} are uncorrelated random errors with mean 0 and variance σ^2. The quantity n_{ij} is design dependent and denotes the number of times that treatment i is administered to a plot in block j (Thus for some i and j , $n_{ij} = 0$). This experiment/scenario aims at inference about the treatment differences $\tau_1 - \tau_o, \tau_2 - \tau_o, ..., \tau_t - \tau_o$ the comparisons between test treatments and the standard treatment. Contrasts among the new treatments are of equal interest, but contrasts between the new treatments and the standard are required to be more accurately estimated. This design will be a proper block design so long as

$$\sum_{j=1}^{b} n_{0j} n_{ij} \tag{4.8}$$

$i = 1, 2, ..., t$ and

$$\sum_{j=1}^{b} n_{ij} n_{i'j} \tag{4.9}$$

$i, i' = 1, 2, ..., t$ $i \neq i'$ do not depend on i and i'.

This implies that this form of balanced incomplete block design(BIB) is one for which the $var(\widehat{\tau_i - \tau_o})$ and the $cov(\widehat{\tau_i - \tau_o}, \widehat{\tau_{i'} - \tau_o})$ are independent of i and i' , $i \neq i'$, where the notation $\widehat{\tau_i - \tau_o}$ denotes the best linear unbiased estimator of $\tau_i - \tau_o$. Notice also that provided $\sum_j n_{oj} n_{ij} \geq 0$ then the BIB will be connected. Martin and Eccleston (1993) note that the best criterion for selecting optimal designs in such situations considers only the sum of the variances of the pairwise contrasts between the new treatments and the standard treatment. In such a case the standard treatment gets more replicates than the new treatments. Consider a design in the class of designs $d(t, b, k)$; if we reinforce a control treatment then we have a class of designs $d^*(t+1, b, k)$. Constantinc(1983) defines a design $d^*(t+1, b, k)$ as being trace optimal for control over a collection D of designs in $d^*(t+1, b, k)$ if

$$\frac{1}{t}\sum_{i=1}^{t} var_{d^*}(\tau_i - \tau_o) \leq \frac{1}{t}\sum_{i=1}^{t} var_d(\tau_i - \tau_o) \tag{4.10}$$

for all $d \in D$

Trace optimality for control considers paired comparisons with the control. A design is trace efficient for the control if it is trace-better for control than the great majority of designs. Consider a situation where k divides b and r, and the control treatment appears bk^{-1} times in each row of d^*, while the other treatment occurs rk^{-1} times in each row of d^*. If each distinct pair of treatments from the set$[1, 2, ..., t]$ occurs in λ blocks of d, then according to Bellman(1970), the information matrix of d^* for treatment effects can be written as

$$kC_{d^*} = \begin{bmatrix} b(k-1) & -r1' \\ -r1 & (r(k-1)+\lambda)I - \lambda J \end{bmatrix} \tag{4.11}$$

where 1 is the column vector with all its entries 1, I is the identity matrix and J is the matrix with all entries 1. C_d is a non-negative definite matrix with zero row sums such that, $kC_d = diag(kr_{d^0}, ..., kr_{d^t}) - N_d N_d'$ where $N_d = (n_{dij})$ with n_{dij} indicating the number of times treatment i appears in the jth block of d ; r_{di} is the replication number

of treatment i in d. The rank of C_{d^*} is t thus the generalized inverse of kC_{d^*} is given as

$$k^{-1}C_{d^*}^{-} = \begin{bmatrix} 0 & 0' \\ 0 & (r(k-1)+\lambda)^{-1}r^{-1}(rI+\lambda J) \end{bmatrix} \tag{4.12}$$

Notice that C_{d^*} and $C_{d^*}^{-}$ are of dimension $t+1$. Ghosh(1982b) noted that a BIB design is D-,A-, and E-optimal and as such designs d^* are more efficient for control than any other design in D . Constantine(1983) proved that the idea of reinforcing is in fact compatible to a satisfactory extent especially when it comes to trace efficiency for control. Part of the main objective for on-farm trials is to compare new treatments against the farmer's practice (control). The findings of Ghosh(1982b) on the efficiency of BIB emphasizes the need to use BIBDs in on-farm trials.

Chapter 5

Possible Scenarios of On-farm Trials

In chapter 3 and 4, we looked at the various forms of measuring efficiency in designs and how optimal incomplete block designs can be obtained through use of several criteria. We highlighted a number of issues in incomplete block designs that need to be dealt with if a satisfactory strategy for on-farm trials designs is to be obtained. In this chapter we apply the methods and criteria presented in this thesis to various scenarios involved in on-farm trials to assess the performance of these different designs. This will provide a paradigm for analysis of on-farm trials and clarify several important issues.

5.1 Introduction to the problem

As in all agricultural experiments, the proper planning of on-farm trials is an important step to ensure that valid results and conclusions are obtained (Gomez and Gomez, 1984). The identification of the most appropriate experimental design is very crucial in the planning stage (Fisher, 1953). The statistical validity of a design, and its ability to adequately estimate the required parameters as well as its flexibility in case of any eventualities are important components(Stroud, 1993). It is always difficult, and at times impossible to recover enough information from a poorly designed experiment. On-farm trials, whether managed by the farmer or researcher require simplicity in design that takes into account the practical situation. Most constraints are due to limited resources such as land and treatment materials. Complexity in design makes the work of the management and data

collection cumbersome, especially for the farmer. There are several constraints and conditions that occur in on-farm trials as far as plot and block treatments are concerned. In many on-farm trials we are faced with a problem where the whole experiment, with or without replications cannot be carried out on one smallholder's farm. This would then necessitate the use of several farms, either as replicates or single plots. In the case where there are no limitations on the plot sizes, resources(treatments) and availability of the farms, then a complete trial is performed on all farms. Such a situation calls for the use of a standard design depending on the nature and treatments being tested. In situations where the farms are large enough to accommodate all treatments, but not large enough to allow for replications, we take farms as replicates. Here use of RCBD may be feasible where a farm acts as a block receiving all treatments.

Possible scenarios in on-farm trials are discussed in the following section. These scenarios are based on possible problems encountered in on-farm trials. The performance of these scenarios is analysed using the optimality and efficiency criteria discussed in the previous chapters.

5.2 The scenarios in on-farm trials

Very often, we encounter situations where farmers are not capable of accommodating all the available treatments. This is because the land available to farmers is of varying sizes, where not all the available treatments can be accommodated. In such a case some farmers select a few treatments whereas others take a full set of treatments for experimenting.

5.2.1 Scenario I: Farms unable to accommodate all treatments

Consider a situation were the test materials are in abundance but the available land sizes do not allow for the allocation of all treatments on the same farm. This would result in some farms not receiving certain treatments. Such situations call for the use of incomplete block designs. Balancing treatments in this scenario is very important in order to obtain

valid and meaningful results. The following example illustrates such a scenario. Consider an on-farm trial where farms can only accommodate a maximum of 4 treatments. In order to achieve balance,

- For 5 treatments, we would need 5 farms with each pair of treatments appearing together on a given farm 3 times.
- For 6 treatments, we would need 15 farms with each pair of treatments occurring together on a given farm 6 times.
- For 7 treatments, we need 35 farms with each pair of treatments occurring together on the same farm 10 times.

These combinations are derived as noted in section (2.27 and 2.28) for the BIBD. Each of the above cases highlights a variation in order of the different levels of precision achieved and calls for different size of experimental material which depends on its availability.

5.2.2 Scenario II: Use of farms as experimental units

In some situations farms can be used as experimental units if all treatments cannot be applied on a single farm. We would then have farms as experimental plots/units. Normally there are two possibilities, i.e Farms are either identical or variable in nature. In experimental design theory, interest is in observing the effects of treatments when applied to identical experimental material(Fisher, 1930). In case of identical farms, we apply treatments to the farms randomly and in such a case, the completely randomized design can be an appropriate design. However, for variable farms, this calls for a blocking of identical farms. Therefore we would then group the similar farms together and then apply the treatments accordingly.

Consider an on-farm trial with 4 treatments to be tested A,B,C,D,and 12 farms($farm = 1, 2, \ldots, 12$) which are located in 3 different soil zones. In this situation, we can group the farms in similar soil zones together and then apply the treatments to the respective

farms falling in a given common soil zone . Suppose farms 4,7,11,2 fall in zone 1, while farms 1,9,6,10 fall in zone 2, and farms 8, 5, 12, 3 fall in zone 3. In this case each farm is considered as an experimental unit. The design layout of the groupings of farms in zones is presented in table 5.1.

Table 5.1: Design layout of the groupings of farms in zones for the above example

Soil Zone 1		Soil Zone 2		Soil Zone 3	
Farm	**Treatment**	**Farm**	**Treatment**	**Farm**	**Treatment**
4	A	1	A	8	B
7	B	9	D	5	A
11	C	6	C	12	D
2	D	10	B	3	C

This illustrates an example of a design under strict assumption on farms falling in each zone to be homogenous. This is a possible scenario in on-farm trials.

5.2.3 Scenario III: Naturally occurring imbalances

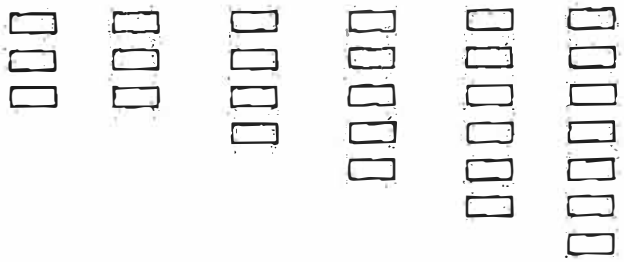
The example considered here describes a situation where balance in treatment allocation within a block is not possible due to naturally existing blocks. This would imply that we have farms taking on varying numbers of treatments. We can still obtain efficient and sensible designs when the block size is not equal to the number of treatments(Mead, 1994).

Consider an on-farm trial testing 7 treatments (A, B, C, D,E,F, and G). Suppose that the available farms are of unequal sizes. This means that each farm is capable of accommodating different numbers of treatments.

Case I

Consider 6 farms where farms are considered as blocks. If the farms contain 3, 3, 4, 5, 6 and 7 plots respectively then a suitable design can be obtained. The design layout is as given in the figure 5.1.

Figure 5.1: The columns represent the farms and within columns we have the plots for each farm.



Assuming equal importance of all treatment comparisons, we can allocate the 7 treatments to the plots within the various farms such that efficient and meaningful results can be obtained. Any 3 of the 7 available treatments could be accommodated in farm 1 or farm 2, whereas farm 6 accommodates all treatments. A possible layout of treatments in each of the six farms is presented in the figure 5.2. The treatments within a farm are not randomized.

Figure 5.2: Allocation of treatments into the plots within farms for Example 5. Experimental plan for comparing 7 treatments in 6 blocks with three, three, four, five, six and seven plots.

Blocks					
FarmI	*FarmII*	*FarmIII*	*FarmIV*	*FarmV*	*FarmVI*
A	*D*	*B*	*A*	*A*	*A*
B	*E*	*C*	*B*	*C*	*B*
C	*F*	*D*	*E*	*D*	*C*
–	–	*G*	*F*	*E*	*D*
–	–	–	*G*	*F*	*E*
–	–	–	–	*G*	*F*
–	–	–	–	–	*G*

Case II

Consider 5 farms where farms are considered as blocks. If the farms contain 3, 4, 5, 7 and 9 plots respectively then a suitable design layout can be obtained as shown in the figure 5.3. Treatments within a farm are non-randomized.

Figure 5.3: Allocation of treatments into the plots within farms for Example 6. Experimental plan for comparing 7 treatments in 5 blocks with three, four, five, seven, and nine plots

Blocks				
FarmI	*FarmII*	*FarmIII*	*FarmIV*	*FarmV*
A	*D*	*A*	*A*	*A*
B	*E*	*B*	*B*	*B*
C	*F*	*C*	*C*	*C*
–	*G*	*D*	*D*	*D*
–	–	*E*	*E*	*E*
–	–	–	*F*	*F*
–	–	–	*G*	*G*
–	–	–	–	*F*
–	–	–	–	*G*

This example demonstrates a situation where Farm IV can accommodate all 7 treatments and Farm V can accommodate more than 7 treatments.

The two possible on-farm trials layout discussed in scenario III are compared to a RCBD on the criterion of the resulting standard errors of estimating treatment differences. This evaluation is presented in the following section.

5.2.4 Evaluation of designs on the basis of standard errors

Consider a randomized complete block design with 7 treatments, 7 blocks and with each treatment replicated 7 times. This design requires 49 experimental units.

Consider also a balanced incomplete block design with parameters $BIBD(t, r, b, k, \lambda) = (7, 4, 7, 4, 2)$. This design would require 28 experimental units. The two sample designs are given in appendix A.2. We use a dummy analysis in Genstat to compare these designs for efficiency by examining the resulting standard errors(s.e). Table 5.2 presents the results of the standard errors.

Table 5.2: standard errors of the 2 unbalanced incomplete block design with a BIBD structure and RCBD in estimating treatment differences

Estimated difference	BIB s.e's	RCBD s.e's	UBIB1 s.e's	UBIBD2 s.e's
$B - A$	0.756σ	0.535σ	0.727σ	0.707σ
$C - A$	0.756σ	0.535σ	0.730σ	0.707σ
$D - A$	0.756σ	0.535σ	0.765σ	0.741σ
$E - A$	0.756σ	0.535σ	0.746σ	0.741σ
$F - A$	0.756σ	0.535σ	0.746σ	0.761σ
$G - A$	0.756σ	0.535σ	0.736σ	0.761σ
$F - G$	0.756σ	0.535σ	0.738σ	0.707σ

Unbalanced incomplete block design is denoted by UBIB where UBIB1 and UBIB2 refer to the design layout in figures 5.2 and 5.3 respectively.

From table 5.2, assuming equal variance σ^2 for all four designs, we notice that the precision of the different comparisons for the two unequal block designs differs slightly from the balanced incomplete block design while for the RCBD, the difference is greater. However, both designs are efficient and sensible for the conditions under which they were set. In some cases these designs perform better in estimating certain treatment differences e.g. in comparing (F-G). However, RCBD with a standard error value of 0.535σ, seems to be the best design in terms of figures if at all a complete blocks design can be set . But under this on-farm scenario, the argument is that complete blocks are impossible. In addition

to the above, RCBD would require 49 experimental plots for this experiment in contrast to 28 experimental units required by the Incomplete block designs. Therefore, there is a big gain in terms of resource savings. In this study we also assumed that treatments are of equal importance. But sometimes this may not be the case. Mead(1994) clarifies that in such a situation, an experimental procedure is to consider the desired relative precision of different comparisons and thereby choose the appropriate replication r for each treatment such that the resulting variances of the treatment differences for the experiment incorporates the relative importance of each treatment. By emphasizing use of BIBD in on-farm trials, it should be noted that more than one treatment may be accommodated in situations where experimental material is available. This would lead to an increase in precision as demonstrated by the difference in standard errors for UBIB1 and UBIB2. The fact that more treatments were accommodated by some farms in UBIB2 led to a smaller standard error.

5.2.5 Scenario IV: Control versus standard treatments

Researchers are particularly interested in comparing standard treatments and a control treatment with a specified number of treatment replications. Such a trial can be set up in smaller incomplete blocks and a control treatment augmented onto the design. An incomplete block set up balanced in such a way that the objectives of the experiment are capable of being tested can be the augmented incomplete block designs. Consider 7 standard treatments and a control treatment O. Thus in total we have eight treatments which are denoted as A,B,C,D,E,F,G and O. Suppose it is decided that the control treatment O has to appear in every block and that each of the other standard treatments has to be replicated 4 times.

An appropriate experimental plan for this study would require seven blocks of size 5 plots each. A suitable design for this study is shown in figure 5.4.

Figure 5.4: Experimental plan for comparing 7 standard treatments plus 1 control treatment in seven blocks of size five plots.

Blocks						
FarmI	*FarmII*	*FarmIII*	*FarmIV*	*FarmV*	*FarmVI*	*FarmVII*
A	*B*	*A*	*B*	*A*	*C*	*A*
B	*C*	*C*	*E*	*B*	*D*	*D*
D	*D*	*F*	*F*	*C*	*E*	*E*
G	*F*	*G*	*G*	*E*	*G*	*F*
O	*O*	*O*	*O*	*O*	*O*	*O*

We examine performance of this design using a Genstat analysis on standard errors of treatment comparisons. A procedure exists in Genstat(Genstat 4.2, 2000) that allows one to evaluate various designs using only dummy variables. The procedure used to produce results given in table 5.3 is given in Appendix A.3 The Genstat procedure is based on regression analysis and therefore the results are presented as regression coefficients and standard errors associated with them. Using the covariance matrix in Appendix A.3, we can obtain other s.e's of treatment comparisons.

Table 5.3: Estimates of the regression coefficients and standard errors of the treatment comparisons

Estimated difference	Estimate	Standard error
B-A	0.000	0.745
C-A	0.000	0.745
D-A	0.000	0.745
E-A	0.000	0.745
F-A	0.000	0.745
G-A	0.000	0.745
O-A	0.000	0.645

It should be noted that all the estiamtes are zero because no data has been used. Our interest in this case is in the standard errors used in treatment versus control comparisons. Suppose we wish to compare treatments B and E, noting that the available information is

on B-A and E-A. The standard error of the comparison (B-E) can be obtained as follows: Using the property of connectedness,

$$B - E = (B - A) - (E - A)$$

The variance of (B-E) denoted as Var(B-E) is computed as

$$Var(B - E) = Var(B - A) + Var(E - A) - 2Cov(B - A, E - A)$$

From Appendix A.3, $Var(B-A) = Var(E-A) = 0.5556$ and $Cov(B-A, E-A) = 0.2778$

$$Var(B - E) = 0.5556 + 0.5556 - 2(0.2778)$$

$$Var(B - E) = 0.5556$$

The standard error

$$S.e(B - E) = \sqrt{Var(B - E)} = 0.7453$$

All standard treatment comparisons have the same standard error 0.7453. The standard error of the comparison of standard treatments with the control treatment O is given as 0.645. This is because a BIBD design was considered.

In this scenario, the control treatment O is replicated more times and it provides for a comparison of each standard treatment with the control treatment within each block/farm. In effect the control treatment is estimated with greater precision than the standard treatments. This scenario represents an on-farm case where the farmers own treatment is compared with the other standard treatments. The design is balanced in the standard treatments, efficient and adequately provides all the required information. Choice of a suitable set of treatments is obviously essential for a successful experiment and the appropriate treatment design is a vital part of any good experimental design. However, on-farm experiments often have large non-treatment sources of variability and the proper allocation of treatments to units using an appropriate block design is also essential for effective experimentation.

5.3 Constrained and unconstrained scenarios

A few of the different scenarios that occur in on-farm trials were discussed in section 5.2. We classify such scenarios as the constrained and unconstrained. Constrained scenarios occur in situations where the standard design is distorted either by the researcher to fit a given situation or due to unavoidable circumstances. Naturally occurring events may force a design to be distorted due to loss of some treatments. The unconstrained scenarios occur when a standard design is set and analyzed as planned with no distortions. A general on-farm mathematical model for both constrained and unconstrained scenarios is of the form:

$$y_{ijk} = \mu + f_i + b(f)_{ij} + \tau_k + f\tau_{ik} + e_{ijk} \tag{5.1}$$

where

y_{ijk} is the observation on the j^{th} block at the i^{th} farm for the k^{th} treatment.

f_i is the i^{th} farm effect; $i = 1, 2, \ldots, f$

$b(f)_{ij}$ is the ij^{th} block within farm effect; $j = 1, 2, \ldots, b$

τ_k is the effect of the k^{th} treatment; $k = 1, 2, \ldots, k$

$f\tau_{ik}$ is the ik^{th} farm-by-treatment interaction effect.

μ denotes the general mean.

e_{ijk} denotes the random error term.

The assumptions for the random elements are

the f_i are i.i.d. $N(0, \sigma_f^2)$

the $b(f)_{ij}$ are i.i.d. $N(0, \sigma_{bf}^2)$

the $f\tau_{ik}$ are i.i.d. $N(0, \sigma_{tf}^2)$

the e_{ijk} are i.i.d. $N(0, \sigma^2)$

The assumption in on-farm trials is that a large number of farms are available from which a sample of farms can be selected. Thus, farms are a random sample from a population of farms and they form a random effect. Therefore all interaction effects involving a farm will also be assumed to be random(Hocking, 1973).

Gomez and Gomez (1984) noted that large variation among farms and between fields in a farm is very common, and this forms a range of sites of various environmental variations. This problem can be handled by spatial modelling of the environmental variation(Legendre and Fortin, 1989). In this context we will introduce a spatial model of environmental variation to cater for the effects from the different locations of these farms. The spatial variation can occur both in constrained and unconstrained cases. Thus, incorporation of a spatial model in on-farm trials is considered in the following section.

5.3.1 Incorporation of a spatial model

It should be noted that spatial variations in fields used in on-farm trials exist. These variations are complicated and difficult to describe. Several authors have investigated the use of spatial analysis techniques for improving the precision of estimating treatment contrasts for data from large field trials(Zimmerman and Harville, 1991; Cullis and Gleeson, 1991). This is because in field trials variation results from many factors such as soils, land gradients/steeps, plant nutrient intake etc. Spatial modelling techniques are able to use covariance between sample points by modelling it as a function of distance between these points. This accounts for both correlation between clustered farms and potential correlation between plots that are close to each other. Several spatial models exist for fitting these variations(Magnussen, 1990; Cressie, 1991). We employ a widely used spherical spatial model by Cressie(1991) because of its flexibility and easy convergence. Here every experimental plot is represented by coordinates (i, j) and therefore a yield from such a plot receiving treatment k is given as

$$y_{ij(k)} = \tau_k + e_{ij} \tag{5.2}$$

We assume that the error component e_{ij} may have nonzero means and covariances which are functions of their distances apart. We assume that e_{ij} has mean $\bar{e}$ which has a component of the drift parameter γ and variance given as V, where $V = v_{ij}$. The spherical

covariance function is given as(Martérn, 1986)

$$v_{ij} = \begin{cases} \sigma_e^2 \left\{ 1 - \frac{3}{2}(\frac{d_{ij}}{\rho}) + \frac{1}{2}(\frac{d_{ij}}{\rho})^3 \right\} & \text{for } 0 < d_{ij} \leq \rho \\ \sigma_0 + \sigma_e^2 & \text{for } d_{ij} = 0 \\ 0 & \text{for } d_{ij} > \rho \end{cases} \quad (5.3)$$

The range of the covariance function is the distance after which observations become uncorrelated. Therefore d_{ij} is the unit distance between plots i and j, ρ is the maximum distance apart beyond which plot yields are uncorrelated which is called the *range parameter* (Magnussen, 1990). We define e as

$$\boldsymbol{e} = \boldsymbol{V}^{\frac{1}{2}}\boldsymbol{n} \quad (5.4)$$

where $V^{\frac{1}{2}}$ is the cholesky factor/square root of V and n is the normally distributed vector with mean of zero and variance 1. Therefore by specifying drift and range parameters γ and ρ, we are capable of obtaining several spatial error distributions across the fields.

The following process illustrates how data with a V matrix that fulfills cholesky factor conditions are generated. These data are later used in both constrained and unconstrained scenarios. Six V matrices are generated using the model (5.3) with different levels of the range parameter. This is done using a SAS program for each of the designs and the respective range parameters. Vectors of sizes of the total numbers of observations T for each design for each farm and treatment applications are generated. i.e

$$\boldsymbol{T}_{ij} = \boldsymbol{t}_i + \boldsymbol{f}_j \quad (5.5)$$

$i = 1, 2, \ldots, t$ (treatment), $j = 1, 2, \ldots, f$ (farm)

A vector n equal to the size of the total number of observation for that specific design is generated with a mean 0 and variance 1. Then a final vector of the data W given as

$$\boldsymbol{W} = \boldsymbol{T} + \boldsymbol{e} \quad (5.6)$$

is generated. Data are finally analyzed using both SAS Proc Mixed procedures and Residual Maximum likelihood(REML)procedures in Genstat after assembling the appropriate

input files.

From the designs, we obtain the design matrix, the incidence matrix N of the design, the matrix of replications r_d and the matrix k of the block sizes. We use these to obtain the required C_d matrix which is given as

$$C_d = r_d - N_d k^{-1} N_d' \tag{5.7}$$

We use the C_d matrix corresponding to the design to obtain the appropriate eigenvalues which are used in calculating the various criteria.

5.3.2 The constrained scenarios

Consider a study to compare $t = 6$ standard treatments with a control treatment O in b blocks of size $k + 1$, where 1 is the additional unit in each block representing the control treatment(farmers treatment). In total we consider 7 treatments for this scenario. The design layouts considered in this section are given in Appendix B.1. The fact that an additional treatment was applied led to constrained condition both for BIBD and RCBD. The performance of these designs under the stated condition is evaluated to determine their performance using the previously discussed optimality and efficiency measure criteria. The design parameters for the 7 design scenarios evaluated are presented in Table 5.4.

Table 5.4: Summary of design parameters for the constrained scenarios

Design scenario	Number of blocks	replications	block size	Lambda λ	Total number of treatments
1	10	5	4	2	40
2	6	1	2	-	12
3	15	5	3	1	45
4	20	10	4	4	80
5	15	10	5	6	75
6	6	5	6	4	36
7	6	6	7	6	42

It should be noted that arrangement of treatments A to F meet BIBD conditions. How-

ever , the condition does not hold anymore with the addition of the control treatment O. Let the treatments be A, B, C, D, E, F and O. In analyzing these trials we require the matrix C_d(John,1971) which is given as $C_d = r_d - N_d k^{-1} N_d'$ as in (5.7)

Design scenario 1. The randomized BIBD with the additional control treatment(reinforced) given as scenario 1 in Appendix B.1 for constrained designs is considered. Each of the treatments A, B, C, D, E,and F is replicated $r = 5$ times while treatment O is replicated $r_o = 10$ times since it appears on each block. Thus the 7×7 diagonal r_d matrix of the above design is given as

$$\begin{pmatrix} 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 10 \end{pmatrix}$$

The 10×10 diagonal matrix k is also given as

$$\begin{pmatrix} 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix}$$

The 7×10 incidence matrix N_d for this design is given as

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Thus the 7×7 treatment information matrix C_d given as

$$C_d = r_d - N_d k^{-1} N_d'$$

is

$$C_d = \begin{pmatrix} 3.75 & -0.5 & -0.5 & -0.5 & -0.5 & -1.25 & -1.25 \\ -0.5 & 3.75 & -0.5 & -0.5 & -0.5 & -0.5 & -1.25 \\ -0.5 & -0.5 & 3.75 & -0.5 & -0.5 & -0.5 & -1.25 \\ -0.5 & -0.5 & -0.5 & 3.75 & -0.5 & -0.5 & -1.25 \\ -0.5 & -0.5 & -0.5 & -0.5 & 3.75 & -0.5 & -1.25 \\ -0.5 & -0.5 & -0.5 & -0.5 & -0.5 & 3.75 & -1.25 \\ -1.25 & -1.25 & -1.25 & -1.25 & -1.25 & -1.25 & 7.5 \end{pmatrix} \tag{5.8}$$

The eigenvalues of C_d are given as ω_1=(8.750, 4.250, 4.250, 4.250, 4.250, 4.250, 0.000). The average variance expressed in terms of the non-zero eigenvalues of the information matrix is given as;

$$AveVar(\tau_i - \tau_0) = \frac{2}{t-1} \sum_{i=1}^{t-1} \frac{1}{\omega_i} \sigma^2 \tag{5.9}$$

This gives

$$= \frac{1}{3} \times 1.290756\sigma^2 = 0.4303\sigma^2$$

or more correctly as the sum of the eigenvalues of its generalized inverse is given as

$$AveVar(\tau_i - \tau_0) = \frac{2}{t-1} \sum_{i=1}^{t-1} \omega_{i(gi)} \sigma^2 \tag{5.10}$$

where $\omega_{i(gi)}$ are the eigenvalues of the generalized inverse of the information matrix C_d and σ^2 is the common variance. From the above, the eigenvalues of the generalized inverse

are given as; (0.2353, 0.2353, 0.2353, 0.2353, 0.2353, 0.1143, 0.0000). Similarly ,

$$AveVar(\tau_i - \tau_0) = \frac{1}{3} \times 1.2908\sigma^2 = 0.4303\sigma^2$$

Analyzing all other design scenarios using a similar approach we have the results in Table 5.5.

Table 5.5: Summary of results for the constrained cases

Scenario	Eigenvalues of C_d matrix	$AveVar(\tau_i - \tau_0)$
1	(8.750, 4.250, 4.250, 4.250, 4.250, 4.250, 0.000)	$0.4303\sigma^2$
2	(3.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0)	$3.4286\sigma^2$
3	(11.667, 3.667, 3.667, 3.667, 3.667, 3.667, 0)	$0.4830\sigma^2$
4	(17.5, 8.5, 8.5, 8.5, 8.5, 8.5, 0)	$0.2151\sigma^2$
5	(14.0, 9.2, 9.2, 9.2, 9.2, 9.2, 0)	$0.2050\sigma^2$
6	(5.833, 4.833, 4.833, 4.833, 4.833, 4.833, 0)	$0.4020\sigma^2$
7	(6.0, 6.0, 6.0, 6.0, 6.0, 6.0, −0.002)	$0.3333\sigma^2$

The high value of average variance observed for scenario 2 ($3.4286\sigma^2$) is due to the fact that only one replication was used whereas the low value of $0.2050\sigma^2$ was as a result of maximum number of replications(10) and $\lambda(6)$.

Evaluation of scenarios

The evaluation of scenarios 1-7 whose average variance of $(\tau_i - \tau_0)$ are presented in table 5.5 based on conventional efficiency, average variance of differences and minimum variability criteria is considered. The results of the efficiency measures are given in the table 5.6.

Table 5.6: Results of Conventional efficiency, Average variance and mini-max variability criterion

Scenario	Conventional efficiency ξ	Mini-max variability criterion	Average variance of differences
1	91.7	3.3750	0.3440
2	53.8	1.5000	0.1890
3	80.0	10.667	0.0846
4	91.8	13.500	0.0490
5	97.1	3.8400	0.0798
6	99.4	0.1670	0.0614
7	100.0	0.0000	0.0314

Conventional efficiency criterion

The conventional efficiency criterion(3.9) results are given in table 5.6. Scenarios 1, 4, 5, and 6 provide reasonable efficiency values as compared to scenario 7(RCBD) which has a value of 100. Scenario 2 and 3 have very low conventional efficiency values. Scenario 2 with conventional efficiency value of 53.8 is the worst under this criterion.

The choice of the design to use in on-farm trial would not only be based on a given criterion but also on availability of resources. Under the constrained situation one needs to be aware of amount of information achievable.

Mini-max variability criterion

According to this measure of variability, a resulting design is said to be better or nearly variance balanced than another design if its variability value is nearer to 0 than the latter's. The mini-max variability criterion values given in Table 5.6 were computed using (4.6). Scenario 6 provides the best design with a variability value= 0.167. The worst

scenario is 4 with a variability value of 13.50. Scenario 7 has a $value = 0$ implying that this design(RCBD) is variance balanced.

Average variance of differences criterion

The average variance expressed in terms of the non-zero eigenvalues of the information matrix is given as the sum of the eigenvalues of its generalized inverse. Based on this criterion we notice that scenario 7(RCBD) has the minimum value of the average variance of differences(0.0314) followed by scenario 4 with a value 0.049.

Evaluation of scenarios based on A-,E-,D-optimality and Efficiency factor

The results of the efficiency measures are given in Table 5.7

Table 5.7: Results for A-,E-,D-optimality and efficiency factor

Scenario	A-optimality	E-optimality	D-optimality	efficiency factor
1	1.29075	4.250	12132.57	0.80
2	10.2860	0.500	0.109375	0.00
3	1.44900	3.667	7735.950	0.60
4	0.64540	8.500	776484.297	0.80
5	0.61500*	9.200*	922714.133*	0.90
6	1.20600	4.833	15380.690	0.96
7	1.00000	6.000	46656.000	1.00

Based on the A-optimality criterion(3.2.1) , scenario 5 is the best since it has the minimum value for A-optimal. Scenario 7(RCBD) has a value of 1.0 which is higher than that obtained by both 4 and 5. The worst scenario for this measure is 2 with an A-optimality value of 10.286. E-optimality also agrees with A-optimality criterion. Here the 2 scenarios 5 and 4 are the best respectively and they are both better than scenario 7(6.0). The worst scenario is 2(0.5). The D-optimality criterion also gives similar results as seen in Table 5.7 . We investigate these designs further using the SAS OPTEX Procedure.

The SAS OPTEX(SAS System version 8) Procedure produces the efficiency measures for defined designs. Using this procedure the efficiency measures for the scenarios were

obtained and presented in table 5.8. The SAS OPTEX Procedure produces the best design for a given scenario. Each scenario is characterized by design parameters given in table 5.4

Table 5.8: Efficiency measures of the respective cases using the SAS OPTEX procedure

Scenario	D-efficiency	A-efficiency	G-efficiency	Aver. Pred. S.E†
1	96.4131	92.7265	88.1594	0.6359
2	73.7596	45.2055	56.0612	1.4240
3	94.4730	88.5766	87.9794	0.7084
4	97.8945	95.6855	93.4326	0.5715
5	98.9599	97.8991	93.6262	0.5219
6	98.9729	97.9802	95.3087	0.5170
7	100.00	100.00	100.00	0.5528

†Aver.Pred.S.E means Average Prediction Standard Error.

From table 5.8 we note that among the incomplete block designs, case 6 is identified as the best design by all the criteria in comparison to the orthogonal design case 7. It is closely followed by case 5 which is marginally different from case 6. The worst design is case number 2 which on average loses about 40% of its efficiency when compared to the orthogonal design. Therefore, for the constrained cases, we would then select either case 6 or case 5 in the family of incomplete block designs depending on the resources available. The good performance of scenario 6 could be attributed to the fact that only one treatment was missing per block. Therefore, it has qualities close to those of RCBD. In addition, it is less complex as compared to other incomplete block scenarios.

Evaluation using the Complexity criterion

The evaluation of the scenarios using the complexity criterion(3.16 and 3.17) was considered. The summary of computations and results are presented in table 5.9.

Table 5.9: Complexity criterion computations for the different design cases

Scenario	$\bar{\lambda}$	$\prod_{i=1}^{\hat{}} \lambda_i$	$\phi(\Lambda)$
1	5	12132.57	0.1265
2	1	0.109375	1.1065
3	5.0003	7735.952	0.3533
4	10	776484.297	0.1265
5	10	922714.133	0.04022
6	4.9997	15380.6947	0.0076995
7	6	46656	0.0

From table 5.9, based on the complexity design criterion, we notice that the worst design is case 2($\phi(\Lambda) = 1.1065$) and the best design is case 6($\phi(\Lambda) = 0.0076995$) followed by design 5(0.04022) in the class of the constrained incomplete block designs. Design 7 provides a value of 0 since it is variance balanced.

The results discussed under the constrained cases illustrate the possibility of conducting on-farm trials under limited resources and restrictions. The researchers become aware of gains/ losses of information under these constraints. This would enable researchers to select a design knowing well in advance how much is expected from such a design.

5.3.3 The unconstrained scenario

Four BIB designs, one cyclic design and one RCBD were generated using the design procedures in the statistical package Genstat for 7 treatments(A, B, C, D, E, F, and ●). The design matrix was obtained for each design. We denote the designs as BIB1, BIB2, BIB3, BIB4, CYCLIC and RCBD. Several criteria were applied in analyzing these designs which include complexity criterion, optimality criteria (A, D,and E), efficiency factor measures, conventional efficiency, mini-max variability criterion, etc. The design parameters are given in table 5.10.

Table 5.10: Summary of design parameters for the unconstrained case

Design	blocks b	Replications r	block size k	Lambda λ	Total number of expt* units	Number of treatments
BIB1	7	3	3	1	21	7
BIB2	7	4	4	2	28	7
BIB3	14	6	3	2	42	7
BIB4	14	8	4	4	56	7
CYCLIC	7	3	3	1	21	7
RCBD	7	7	7	7	49	7

* expt stands for experimental.

The data for these designs were analyzed using the statistical packages Genstat and SAS and the results based on the criteria of interest presented. As in constrained cases we also develop the C_d matrix and we use it to obtain the necessary eigenvalues and other information pertaining to the design which formulates the basis to compare these designs. The results from the various criteria on the designs are given in Tables 5.11 and 5.12.

Table 5.11: Standard errors(s.e) of differences of means for each design and the relative efficiencies of the 5 designs to RCBD

Case	S.E Reml	S.E Anova	Efficiency factor	conv.eff
BIB1	0.663	0.715	0.778	0.8935
BIB2	0.6571	0.6831	0.875	1.024
BIB3	0.3789	0.4279	0.778	0.894
BIB4	0.2681	0.2828	0.875	1.0244
CYCLIC	0.9965	1.189	0.778	0.8939
RCBD	0.3909	0.4	1	1.0

Table 5.12: Average variance, Optimality values and Complexity values for the designs

Case	Ave.var	A-optimal value	D-optimal	E-optimal	Complexity criterion
BIB1	0.8573	2.5718	161.246	2.333	0
BIB2	0.5714	1.7143	1838.27	3.50	0
BIB3	0.4285	1.2854	10346.3	16.66	0
BIB4	0.2857	0.8571	117649	16.67	0
CYCLIC	0.8569	2.5707	161.6	2.334	0
RCBD	0.2381	0.7143	44471.32	2.646	0

Discussion of results

From tables 5.11 and 5.12 we note the following: from the standard error of the differences of means using both Reml and Anova, BIB4 has the minimum value followed by BIB3 and RCBD respectively. The maximum value is obtained in the cyclic design which shows that this design had the highest variance in estimating treatment means differences. On the basis of the conventional efficiency factor, the two designs BIB2 and BIB4 have values greater than 1 implying that they are more efficient than the RCBD based on this criterion. The other 3 IBD seem to lose about 11% of the efficiency in comparison to the RCBD. On the criterion of the average variance of the estimates of treatment means differences according to Kshirsagar(1958), RCBD is the best design followed by BIB4 and then BIB3. Using the A-optimality criterion, RCBD is the A-optimal design overall and in the class of the IBD the A-optimal design is BIB4 i.e. the design which minimizes the A-optimal value is the A-optimal design. The D-optimal design is BIB4. RCBD ranks second under this criterion and the worst is BIB1. The E-optimal design is BIB4. It is followed by BIB3, BIB2, RCBD, CYCLIC and the worst is BIB1. Since all these designs are variance balanced, the complexity criterion turns out to be 0 for all.

The results from the unconstrained scenarios indicate that a particular set of BIB designs can be used in on-farm trials where resource constraints do not exist.

Power analysis in unconstrained designs

Using power analysis in the SAS PROC MIXED procedure, we assessed the performance of designs given in Table 5.10. For illustration purposes, suppose a variance 2.5 among blocks(farms) and a variance of 4 between plots within a given block(farm) are assumed. Under this condition the power in estimating the treatment contrast $(1, 1, 1, -1, -1, -1, 0)$ was computed for each of the designs. This contrast represents a comparison between treatments A, B, C versus D, E, and F. The results are presented in Table 5.13 where obs denotes number of observed contrasts, ndf denotes numerator degrees of freedom, ddf denotes denominator degrees of freedom, alpha, denotes the significance level of the test, ncparm is the product of ndf and the observed F-value and fcrit denotes the critical value of F. The results show that design BIB2, BIB4 and RCBD have power values above 0.5.

Table 5.13: Power analysis of the unconstrained designs in estimating a treatment contrast

Design	Obs	ndf	ddf	ncparm	alpha	fcrit	Power
BIB1	1	1	8	3.04	0.05	5.31766	0.33627
BIB2	1	1	15	4.57	0.05	4.54308	0.51607
BIB3	1	1	22	3.84	0.05	4.30095	0.46576
BIB4	1	1	36	5.12	0.05	4.11317	0.59568
CYCLIC	1	1	8	1.80	0.05	5.31766	0.21990
RCBD	1	1	36	5.01	0.05	4.11317	0.58645

However, in estimating this particular contrast, the incomplete block design (BIB4) is more powerful than the randomized complete block design(RCBD). The cyclic design is not suitable for the estimation of this contrast and it has the least power value of 0.21990. Therefore, depending on the interest of the researcher, i.e the question to be answered, we can use power analysis to obtain the appropriate design in order to obtain the best information in answering the research questions. The procedure is to express the research question in from of a contrast and we analyze the power of the designs in estimating that contrast. For this particular contrast, BIB4 would be the most appropriate design since it has a power value greater than that for the RCBD.

5.4 Missing plot evaluation in BIB designs

In this section we consider two scenarios 1 and 4 under the constrained scenarios given in Table 5.4. We evaluate these designs on the missing plots criteria given in section 4.1.

Scenario 1

From Table 5.4 $t = 6$ treatments, $k = 3$ treatments per block, $b = 10$ blocks, $r = 5$ replications of treatments and $\lambda = 2$. The average variance of a pair-wise treatment difference for the complete design is

$$(2k/\lambda t)\sigma^2 = 2 \times 3/2 \times 6)\sigma^2 = \sigma^2/2$$

The efficiency of this design relative to a completely randomized design with the same number of treatments t and replications r is

$$\lambda t/rk = 2 \times 6/5 \times 3 = 0.8$$

For one missing observation we have Average Variance

$$AveVar(\hat{t}_{i1} - \hat{t}_{i2}) = \frac{2k\sigma^2}{t\lambda} + \frac{2k^2\sigma^2}{t(t-1)\lambda(t\lambda - k)} = \frac{32\sigma^2}{60} \tag{5.11}$$

with Relative efficiency to the complete design

$$Rel.Eff = \{1 + \frac{k}{(t-1)(t\lambda - k)}\}^{-1} = 0.9375. \tag{5.12}$$

With two missing observations, there are 5 different configurations, each with different eigenvalues to be considered. However, four of these configurations have sub-cases that arise depending on g, the number of treatments common to the pairs of blocks.
These cases are:

- I. Two observations in same block.
- II. Different treatments in separate blocks(no treatment common to the two blocks).

- III. Different treatments in different blocks(and one of the two treatments common to both blocks.
- IV. Both treatments are different and occur in both blocks.
- V .Two replicates of the same treatment lost.

The results of average variances and their corresponding relative efficiencies are presented in table 5.14.

Table 5.14: The average variances and their corresponding relative efficiencies

Case	g value	x	x^2	Average Variance	Relative efficiency	Number of configurations
I	g=0	0	0	0.566700	0.88240	30
II	g=1	1/6	1/36	0.567500	0.88110	9
	g=2	2/6	1/9	0.569998	0.87720	4
III	g=1	2/6	1/9	0.569998	0.87720	4
	g=2	1/6	1/36	0.567500	0.88110	4
IV	g=2	4/6	4/9	0.580520	0.86100	2
	g=3	3/6	1/4	0.574300	0.87063	6
V	g=1	4/6	4/9	0.580520	0.86130	1
	g=2	5/6	25/36	0.588960	0.84895	2
	g=3	1	1	0.600000	0.83300	3

From the table(5.14) we notice that the best case is where two observations are lost from the same block and the blocks are arranged in such a way that, the number of treatments common to the pairs of blocks is zero. This is case I with $g = 0$ and it provides a relative efficiency of 0.8824 with an average variance of 0.5667. The minimum loss of efficiency is about 11%. However, the worst case occurs when two replicates of the same treatment are lost in a situation where the blocks are arranged in such a way that the number of treatments common to the pairs of blocks is 3 i.e $g = 3$. In this case, such a situation leads to a minimum loss of efficiency of about 16% with an average variance of about 0.6. This is case V with a g-value= 3. From Figure 5.5 we notice that on average there is a drastic fall in efficiency as we move down the different cases(g-values). The graph displays

a sharp fall/rise in efficiency for the g values between 3g2(case III with g=2) and 4g3(case IV with g=3) respectively. With case III and g-value=2 we lose 2 treatments in the same block with 2 treatments common to the pairs of the blocks. While for case IV and g-value=2, we lose 2 different treatments which occur in both blocks, and the blocks are such that there are 2 treatments common to the pairs of blocks. This is because for case IV, g-value=2, there is a direct impact on the connectivity of the design and this makes the design less efficient thus a sharp fall in the resulting efficiency. From 4g3(case IV, g=4) to 5g3(case V, g=3) we experience a steady fall in the efficiency. With two treatments lost and both are different and occur in both blocks, g the number of treatments common to the pairs of blocks has a significant effect on the resulting efficiency. In this example we notice that for g=3 the efficiency registered is 0.87063 for case IV as compared to g=2 which is 0.8610. Therefore the greater the number of treatments common to the pairs of blocks, the greater the efficiency that will be obtained in case of losing two different observations if they both occurred in both blocks in the original design. This can also be noticed from figure 5.6 of the average variance against cases(g values). This figure also portrays a similar picture in terms of the average variances obtained.

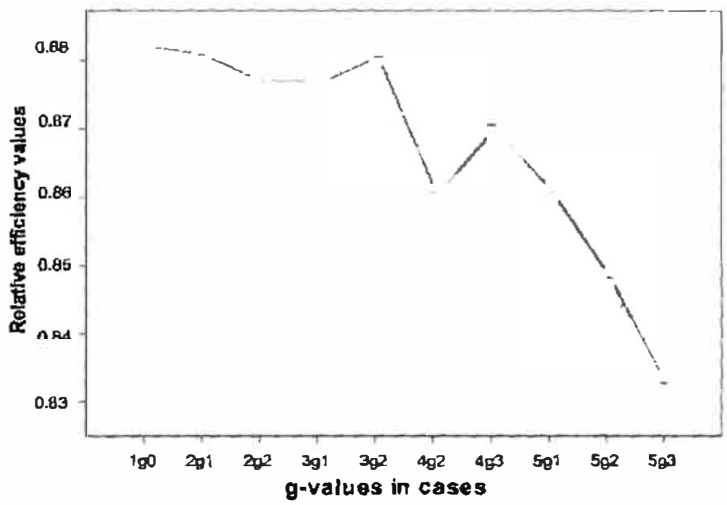


Figure 5.5: The relative efficiencies against the g-values in the various cases for scenario 1.

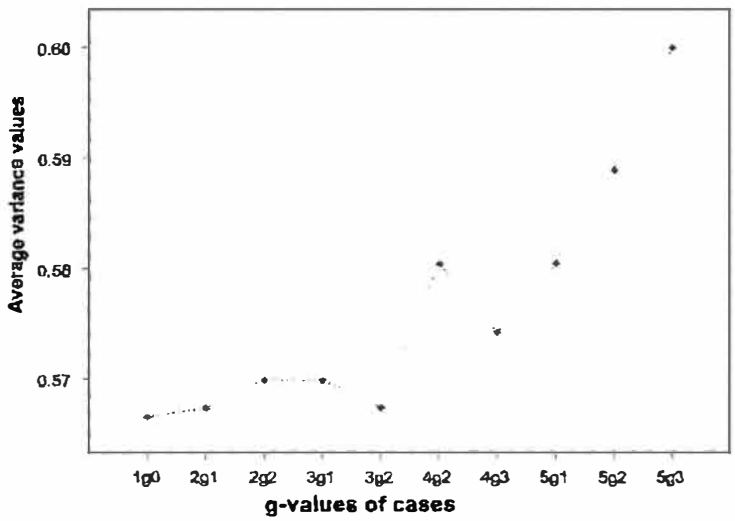


Figure 5.6: Average variance of the different cases verses the g-values of the cases for the loss of an observation for scenario 1

Scenario 4

In this study we have $t = 6$ treatments, $k = 3$ treatments per block, $b = 20$ blocks, $r = 10$ replications of treatments and $\lambda = 4$. The average variance of a pairwise treatment difference for the complete design is $(2k/\lambda t)\sigma^2 = \sigma^2/4$. Its efficiency relative to the completely randomized design with t treatments and r replications is $\lambda t/rk = 0.8$.

For one missing observation we have Average Variance $=0.257143\sigma^2$ with Relative efficiency to the complete design as 0.97222.

For two missing observations, the results are presented in table 5.15.

Table 5.15: Average variances, relative efficiencies and Number of configurations for Scenario 4 for 2 missing observations

Case	g value	x	x^2	Average variance	Relative efficiency	Number of configurations
I	g=0	0	0	0.264300	0.945895	60
II	g=1	1/6	1/36	0.264350	0.945720	9
	g=2	2/6	1/9	0.264545	0.945020	4
III	g=1	2/6	1/9	0.264545	0.945020	4
	g=2	1/6	1/36	0.264350	0.945720	4
IV	g=2	4/6	4/9	0.265332	0.942216	2
	g=3	3/6	1/4	0.264870	0.943860	6
V	g=1	4/6	4/9	0.265332	0.942216	1
	g=2	5/6	25/36	0.265930	0.940100	2
	g=3	1	1	0.266700	0.937400	3

From the table 5.15, we notice that this situation is in agreement with the first scenario, i.e. the best case is where two observations are lost from the same block and the blocks are arranged in such a way that the number of treatments common to the pairs of blocks is zero.

However, in this design the number of blocks b and λ is increased and we notice a great improvement in the relative efficiencies as compared to the original design. The minimum loss of efficiency in this design is about 6% for the worst case(V) with $g = 3$ as compared to the worst case in scenario 1 which has a minimum loss of efficiency of about 16% un-

der a similar case. On the basis of the average variances, scenario 4 still has all values less than 0.27 as compared to the previous which has values ranging between 0.5 to 0.6. Therefore, on average scenario 4 is a better design than scenario 1 as far as missing plots are concerned. Figure 5.7 is a graph of the relative efficiencies of this scenario against the cases and it is not much different from the graph for scenario 1 in behavior. Similarly, Figure 5.8 also reveals a similar pattern as the average variance graph for scenario 1. The only difference is in the magnitude of the efficiencies and the average variances.

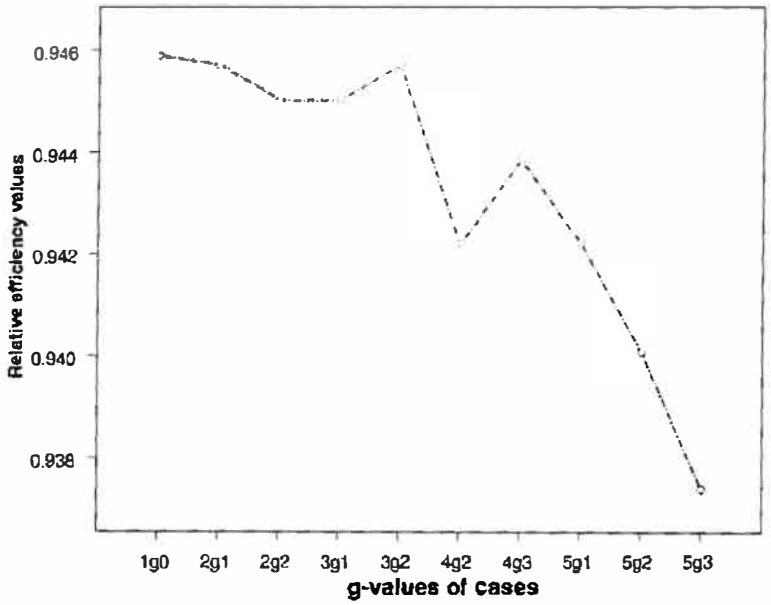


Figure 5.7: Relative efficiencies of the different cases versus the g-values of the cases for the loss of 2 observations in Scenario 4

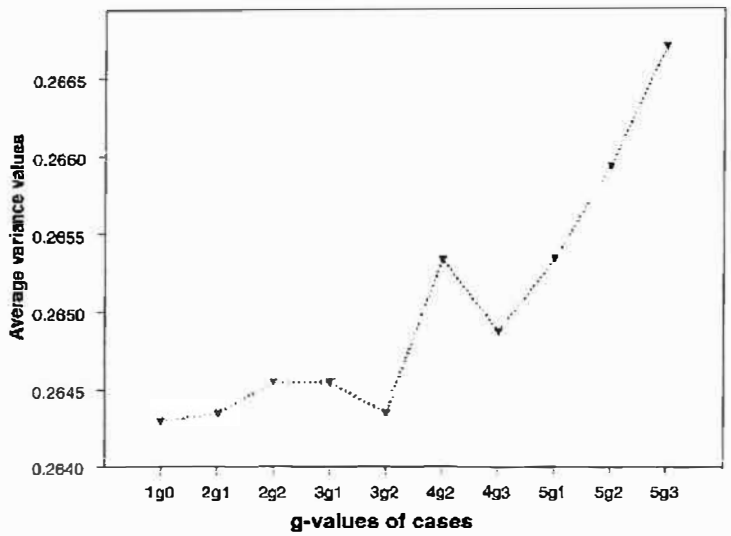


Figure 5.8: Average variances of the different cases versus the g-values of the cases for the loss of 2 observations in scenario 4

5.4.1 Discussion of the results

In many on-farm trials there is a high chance of not having all the observations available at the end of the trial and hence many missing observations. Therefore, in designing on-farm trials particularly in incomplete block designs we need to safe guard against case V which provides lower efficiency values i.e losing 2 replicates of the same treatment. The results reveal that the greater the number of treatments common to the pairs of blocks(g), the greater the loss in efficiency. Therefore, for a better design we would need to minimize the number of treatments common to the pairs of blocks in cases where two replicates of the same treatment are most likely to be missing. In cases where both treatments are different and occur in both blocks(case IV), we would maximize the number of treatments common to pairs of blocks in order to obtain maximum efficiency in case of losing 2 observations. The best situation is case I where we lose different treatments in different blocks where neither treatment is common to the two blocks and the number of treatments common to the pairs of blocks(g) is zero.

This case provides for proper connectivity within the design and therefore treatment comparisons/estimates can still be obtained. Designs that consist of blocks with many common treatments tend to average to very low efficiency values.

This is why the worst designs would be where the missing observations are for the same treatment in similar blocks. In general, to obtain minimal loss in efficiency of a design as a result of loss of treatments, we require an incomplete block design with as few treatments common to blocks as possible.

5.5 Case study: Cowpea trials

A cowpea trial was conducted in two major cowpea growing districts in Eastern Uganda to investigate the performance of two promising cultivars with different pest management technologies. Pest infestation data were collected weekly throughout the study and yield data were collected at the end of the trials.

The treatment combinations were;

1. Variety Ebelat; at $60 \times 20cm$; no spraying(T1)
2. Variety Ebelat; at $60 \times 20cm$; sprayed once at budding,flowering and podding(3 sprays a season)(T2)
3. Variety Ebelat; at $60 \times 20cm$; sprayed weekly(8 sprays a season)(T3)
4. Variety Makerere; at $60 \times 20cm$; no spraying(T4)
5. Variety Makerere;at $60 \times 20cm$; sprayed once at budding,flowering and podding(3 sprays a season)(T5)
6. Variety Makerere;at $60 \times 20cm$; sprayed weekly(8 sprays a season)(T6)
7. Farmers variety and method of management(T7)

A control treatment which is the farmers own management practice was included in the trial. Two incomplete randomized block trials and one Randomized complete block design trial were considered. Two in Pallisa district and one in Kumi district. The designs for the trials are presented in the Appendix B.3.

Analysis of designs based on the efficiency measure criteria

Results for the performance of designs on the efficiency measures are presented in tables 5.16 and 5.17.

Table 5.16: Results for A-optimal, E-optimal and D-optimal criteria

Trial	A-optimal	E-optimal	D-optimal
PalisaIBD	1.29075	4.2500	12132.57
KumiIBD	30.1880	0.1667	0.000665
RCBD	1.00000	6.0000	46656

Table 5.17: Results for the complexity criterion, average variance, conventional efficiency and efficiency factor

Trial	Complexity criterion	Average Variance	Conventional efficiency	efficiency factor
PalisaIBD	0.1265	0.4303	0.917	0.8
KumiIBD	3.6580	10.063	0.040	0.8
RCBD	0.0000	0.0314	1.000	-

The optimality criteria(A-,E- and D-) reveal that RCBD is better followed by the IBD for Pallisa and Kumi is the worst. As seen in the table above, RCBD has the minimum value of the A-optimality(1.0), it has the maximum value of E-optimality(6.0) and has maximum value of D-optimality(46656). It is consistently followed by the Pallisa IBD. Similarly, with the complexity, Average variance and conventional efficiency measures,they all identify RCBD as the best design in these trials.

Relative efficiency measures

Three standard errors of the treatment differences are computed, i.e maximum standard error of differences, the average standard error of differences, and the minimum standard error of differences. The relative efficiency of the incomplete block design to the

randomized complete block design is then computed as

$$\boldsymbol{Rel.Effi.} = \frac{\text{Standard Error of differences for RCBD}}{\text{Standard Error of differences for IBD}}$$

The summary of results is given in the table (5.18) below.

Table 5.18: Average variances, Standard errors(S.E) and Relative efficiency values for the respective trials for estimating differences between means.

	Kumi IBD	Pallisa IBD	Pallisa RCBD	Relative efficiency values	
				Kumi IBD	Pallisa IBD
Maximum s.e of differences	0.4137	0.3325			
Average s.e of differences	0.4078	0.2843	0.2562	0.628	0.90116
Minimum s.e of differences	0.3929	0.2147			
Average variance of differences	0.1664	0.08176	0.06575	0.3951	0.8042

The relative efficiency values are less than 1 for the two case studies which implies that RCBD performed better than the two IBD. The disparity can be explained in the high level of missing observations that occurred in the two incomplete block designs. This was due to the long drought during the months of September to January 2001 in these regions. Other trials were destroyed by wild or domestic animals that roam around uncontrolled during the night. This exactly describes a true scenario in on-farm trials.

In general, the randomized complete block designs performed better in these trials. The efficiency measure criteria considered here show a greater difference in the two sets of designs. However, it should be noted that the incomplete block designs for these trials were not selected as the best from a set of all possible designs. This study considered an incomplete block design without assessing several designs that would even be more efficient for this experiment. This thesis covers several approaches that can be used for this design selection problem. In light of the costs involved in setting up the on-farm

trials, there is a significant difference in costs incurred in setting up a RCBD and an IBD. Therefore, on this basis, we can adopt an incomplete block design with highest efficiency values, and good parameters that can achieve high precision and yet restrict ourselves to the available resources.

Chapter 6

Conclusions

This study gives some insight into the key benefits and weaknesses in the use of incomplete block designs in on-farm trials. Several techniques are used in the evaluation of designs in order to obtain a proper incomplete block design solution for a given situation in on-farm trials. The proper use of these techniques is of key importance. This study illustrates a tradeoff between using incomplete block designs in place of randomized complete block designs. It shows that adequate information can be obtained without unduly large experiments.

The results reveal that designs that have a balance between control and standard treatments are not always the best designs. This is because power estimations are clearly related to the design parameters. Thus with hypothesized parameters, one is capable of assessing and obtaining the best design decisions for the many different scenarios that exist in on-farm trials. Although the approaches discussed here are theoretically complex, they depend on clear and simplified assumptions to enable one to make the best decisions regarding experimental designs for on-farm trials.

6.1 Specific contributions

The thesis adds to the state of knowledge about on-farm trials in the following ways:

- A review of the literature about incomplete block designs reveals a clear benefit in the use of these designs(Kempson & Clarke, 1997; Goulden, 1937; Harrington, 1948;

Patterson & Hunter, 1983). The review shows a clear gain in efficiency attained by the use of incomplete block designs.

- We review the different forms of incomplete block designs which adequately fit the true on-farm situation, especially in the variability of farms which calls for unequal block sizes and augmented designs(Das, 1958), which accommodate the trials of standard treatments versus control treatment.

- Analytical techniques exist for constrained and unconstrained on-farm trials. Handling of missing data scenarios which is a major characteristic of on-farm trials has been considered. Mixed models analysis today provides proper analysis of incomplete block design trials. And therefore rigid orthogonality in designs remains less of a limitation as far as obtaining the results from a trial is concerned.

- The methods demonstrated allow for flexibility of designs and provide researchers with information on the gain / loses of efficiency depending on the design used which is mainly controlled by the available resources and conditions. Incorporation of the appropriate covariance model is another important area of designing on-farm trials. This enables a researcher to fully account for the variability in the results that are obtained. Consideration of several efficiency criteria and measures for constrained and unconstrained designs can lead to slightly different conclusions. This confirms that the use of augmented designs and all other constrained designs can be of great help to on-farm trials especially since they fit the on-farm setting perfectly.

- In comparing incomplete block designs to the randomized complete block design, the results reveal in general that the RCBD is better. However the IBD have efficiency values close to that of the RCBD and thus could be useful in on-farm trials given the compensation in the costs involved in having full blocks in relation to incomplete blocks. The power calculations for the performance of designs provides a concise approach to comparisons of all forms of designs.

The key question here was how to evaluate the efficiency of incomplete block designs in relation to randomized complete block designs. The methods considered reveal that in several cases the IBD are close to being as efficient as RCBD. In addition to reducing the time spent on complete blocks, there is a big saving in the costs involved in setting up complete blocks. Therefore, on the basis of this argument, incomplete block designs are capable of providing design solutions in on-farm trials particularly where randomized complete block is hard to implement, either by nature of land, resources, limitations in complete blocking schemes, etc.

Researchers have been particularly reluctant about the application of IBD to on-farm trials due to the complexity involved in designing, laying out and analyzing these designs. However, with the emergence of computing services, these problems can fully be addressed today. Adoption of randomized complete block design, results in an increase in error variance due to large block sizes in on-farm trials (Das and Giri, 1979). In environments where considerable variation occurs over the testing ground, smaller blocks are most suitable and this renders the incomplete block designs most suitable. Although on-farm tests in incomplete blocks vary in their effectiveness in answering research questions, trials have clearly indicated that incomplete blocks are a necessary design for on-farm trials.

6.2 Further research

We have looked at the different incomplete designs, and considered criteria that can be used to select an optimal and efficient design from a class of all possible designs. However, these are standard criteria and each of them uses an independent method of selection depending on what the aim of the experiment maybe or the interest of a researcher. Therefore, more research will be required in the area of how to integrate all the different approaches in order to have a standard method catering for all the general selection methods. This would reduce the amount of work involved in the selection of the best design. On-farm trials are very unpredictable due to the nature and lack of control of several key factors in these trials. Factors such as natural catastrophes, destruction by wild animals and farmer negligence in case of farmer-managed trials. Therefore, in order to achieve genuine and acceptable results, both the farmer and researcher need to give all their possible input in order for these trials to succeed. Even if proper or efficient designs are selected for these trials, the final results and efficiency will depend on what type or kind of data is obtained from these trials. In summary, there is a considerable gain in the use of incomplete block designs for on-farm trials.

In some cases the difference between the RCBD and IBD is very minimal, thereby prompting the use of IBD. At times the nature of the field of trial will in no way call for the use of RCBD. In such cases therefore we need the incomplete block designs. In on-farm trials this is normally the case as illustrated in the different scenarios.

Bibliography

Ashby, J.A. 1990. Evaluating technology with farmers: A handbook. *CIAT Publication no.187.* Apartado Aereo 6713, Cali, Columbia.95 pp.

Atkinson, G.L. and Donev, A.N. 1992. *Optimum Experimental Designs,* Clarendon Press. Oxford, 327.

Bailey, R.A. 1984. Quasi-complete Latin squares: construction and randomization. *J. Roy. Statist. Soc.* B 46, 323-334.

Bailey, R.A. 1985. Factorial design and Abelian groups. *Linear Algebra Appl.* 70, 349-368.

Baksalary, J.K. and Tabis, Z. 1987. Conditions for robustness of block designs against the unavailability of data. *J. Statist. plann. and Inf.* 16, 49-54.

Bellman, R. 1970. *Introduction to Matrix Analysis* 2nd ed. New York: McGraw-Hill.

Bhaumik. D.K., Whittinghill, D.C., 1991. Optimality and robustness to the unavailability of blocks in block designs. *J. Roy. Statist. Soc. Ser.* B 53, 399-407.

Bose, R.C. 1939. On the construction of balanced incomplete block designs. *Ann. Eugen.* 9, 353-99.

Bose, R.C. and Nair, K.R. 1939. Partially balanced block designs. *Sankhya Vol.4* , pp. 337-373.

Bose, R.C. 1947. Mathematical theory of the symmetrical factorial design. *Sankhya 8,* 107-166.

Bose, R.C. and Shimamoto,T. 1952. Classification and analysis of partially balanced designs with two associate classes. *J.Amer.Stat. Assn., Vol.47,*pp. 151-190.

Bose, R.C., Clatworthy, W.H. and Shrikhande, S.S. 1954. Tables of partially balanced designs with two associate classes. *Technical bulletin no* 107, North Carolina Agricultural Experiment station.

Bose, R.C., Shrikhande, S.S. and Bhattacharya, K.N, 1953. On the construction of group divisible incomplete block designs. *Ann. Math. Statist.* 24, 167-195.

Byerlee,D., Triomphe, B. and Sebillotte, M. 1991. *Experimental Agricultural Journal.* Vol. 27

Cheng, C.S. 1978. Optimal designs for the elimination of multi-way heterogeneity *Ann.Statist.*, 6 1262-1272.

Cheng, C.S. 1983. Construction of optimal balanced incomplete block designs for correlated observations. *Ann.Statist.,* 11, 240-246.

Chernoff, H. 1953. Locally optimal designs for estimating parameters. *Ann. Math. Statist. 24,* 586-602.

Chakrabarti, M.C. 1962. *Mathematics of design and analysis of experiments* Asia Publishing House.

Clarke,G.M. Kempson,R.E., 1997. *Introduction to the Design and Analysis of Experiments,* Arnold.

Clatworthy, W.H. 1973. *Tables of two-associates-class partially balanced designs.* Appl. Math. Ser. 63. National Bureau of Standards, Washington.

Coe, R. 1998. *Participatory on-farm experimentation in Agroforestry: experiences and the role of biometrics.* The XIXth International Biometric Conference. IBC 98 Invited Papers pp 1-10.

Colbourn, C.J and J.H.Dinitz. 1996. *The CRC Handbook of combinatorial Designs,* CRC Press

Conniffe, D and Stone, J. 1974. The efficiency factor of a class of incomplete block designs. *Biometrika,* 61, 633-636.

Constantine, G.M. 1983. On the trace efficiency for control of reinforced balanced incomplete block designs. *J.Roy. Statist. Soc. B* 45, 31-36.

Cochran, W.G. and Cox, G.M., 1957. *Experimental Design.* Wiley, New York.

Corbeil, W.G. and Searle, S.R. 1976. Restricted maximum likelihood(REML) estimation of variance components in the mixed model. *Technometrics 18,* 31-38

Cressie, N. 1991. *Statistics for Spatial Data.* New York: John Wiley Sons.

Cullis, B.R., and Gleeson, A.C. 1991. Spatial analysis of field experiments- an extension to two dimensions. *Biometrics, 47,* 1449-1460.

Das, M.N. 1958. On reinforced incomplete block designs. *J. Ind. Soc. Agric. Statist. 10,* 73-77.

Das, M.N. and Giri, N.C. 1979. *Design and Analysis of Experiments.* New Delhi: Wiley Eastern.

Das. A., Kageyama, S., 1992. Robustness of BIB and extended BIB designs against the unavailability of any number of observations in a block. *Comput. Statist. Data Anal. 14.* 343-358.

Dey.A., 1993. Robustness of block designs against missing data. *Statist. Sinica 3.* 219-231.

Dey, A. and Das,A. 1989. On some E-optimal block designs. *Metrika 36,* 269-278.

Dunnet, C.W.1955, A multiple Comparison procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association,* 50, 1096-1121.

Eisenhart, C. 1947. The assumptions underlying the analysis of variance. *Biometrics 3:* 121.

Emden, M.H. Van 1971. An analysis of complexity. *Mathematical centre Tracts,* Volume 35. Mathematical Centrum , Amsterdam.

Federer, W.T. 1955. *Experimental Design.* The MacMillan Co., New York, NY.

Fedorov, V.V., 1972. *Theory of Optimal experiments.* Academic Press.

Fisher, R.A. 1953. *The design of experiments.* Oliver and Boyd, Ltd., Edinburgh and London.

Fisher, R.A. 1940. An examination of the different possible solutions of a problem in incomplete blocks. *Annals of Eugenics, 10,* 52-75.

Fisher, R.A. 1958. *Statistical methods for research workers.* Oliver and Boyd, Edinburgh, 10th ed.,chapters VII and VII.

Fisher, R.A. 1942. The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eugen. 11,* 341-353.

Fu, Y.B., Clarke, G.P., Namkoong, G., and Yanchuk, A.D. 1998. Incomplete block designs for genetic testing: statistical efficiencies of estimating family means. *Can. J. For. Res.* 28: 977-986.

Fu, Y.B., Yanchuk, A.D., Namkoong, G., and Clarke, G.P. 1999. Incomplete block designs for genetic testing Statistical efficiencies with missing observations. *For. Sci. 45(3):* 374-380.

Ghosh, S. 1982a.Information in an observation in robust designs. *Commun. Statist. Part A- Theory Methods 11,* 1173-1184.

Ghosh, S. 1982b. Robustness of BIBD againsst the unavailability of data. *J.Statist. Plann. and Infer.*6.pp 29-32.

Giesbrecht, F.G., 1986. Analysis of data from incomplete block designs. *Biometrics 42*, 437-448.

Graybill, F.A., 1976. *Theory and applieation of the linear model.* Wadsworth, Belmont CA.

Gomez and Gomez 1984. *Statistical procedures for Agricultural Research, with emphasis on rice.* Los Banos, The Philippines: International Rice Research Institute.

Goulden, C.H. 1937. Efficiency in field trials of Pseudo-factorial and incomplete randomized block methods. *Can. J. Res.* C. 15: 231

Gupta,V.K., Srivastava, R., 1992. Investigations on robustness of block designs against missing observations. *Sankhya-Indian J.Statist. Ser.B* 54, 100-105.

Hammerton, J.L. and Lauckner,F.B. 1984. *On-farm Experimentaion: A Manual of Suggested Experimental procedures.* Trinidad: CARDI(Caribbean Agricultural Research and Development Institute).

Harville, D.A., 1977. Maximum Likelihood approaches to variance component estimation. *Journal of the American Statistical Association,* 72, 320-340.

Heydayat. A. and John. P.W.M., 1974. Resistant and susceptible BIB designs. *Ann.Statist. 2* 148-158.

Hildebrand, P.E., and F. Poey. 1985. *On-farm Agronomic Trials in Farming Systems Research and Extension.* Lynn Rienner Publ., Boulder, Colorado.

Hildebrand, P.E., and J.T. Russell. 1996. *Adaptability analysis. A method for the design, analysis and interpretation of on-farm research-extension.* Iowa State University Press.

Hocking, R.R. 1973. A discussion of the two-way mixed model. *Amer. Stat. 27(4):* 148152.

Houtman, A.M. and Speed, T.P., 1983. Balance in designed experiments with orthogonal block structure. *Ann. Statist.11,* 1069-1085.

Huber, J. and Zwerina Klaus, 1996. The importance of Utility balance in efficient choice designs. *Journal or Marketing Research,* 33(August), 307-317.

Jacroux,M. 1984. On thc Optimality and Usagc of Rcinforced Block Designs for Comparing Test Treatments with a Standard Treatment. *J.R. Statist. Soc.B 46.* pp.316-322.

Jarrett, R.G. 1989. A review of bounds for the efficiency factor of block designs. *Austral. J. Statist.* 31,118-129.

Jarrett,R.G. and Hall, W.B. 1978. Generalized cyclic incomplete block design. *Biometrika, 65,2*, pp. 397-401.

John,J.A and Mitchell,T.J 1977. Optimal incomplete block designs. *J. R. Statist. Soc.,Ser.B*, 39, 39-43.

John, J.A. 1981 Efficient cyclic designs. *J. Roy. Statist. Soc. B* 43 76-80.

John, J.A. 1987. *Cyclic Designs*. NY: Chapman & Hall.

John, J.A. Williams E.R. 1995. *Cyclic designs and computer-generated designs*. NY: Chapman Hall.

John, P.W.M 1971. *Statistical Design and Analysis of Experiments*. New York: Macmillan.

John, P.W.M. 1973. A balanced design for eighteen varieties. *Technometrics*, 15, 641-642.

John, P.W.M. 1980. *Incomplete block designs*. NY: Marcel Dekker, Inc.

Kempthorne, O., 1947 A Note on Differential Responses in Blocks, *Jour. Agr. Sci.*, 37, 245-248.

Kempthorne, O. 1952. *The Design and Analysis of Experiments*. Wiley, New York.

Kempthorne, O. 1956. The efficiency factor of an incomplete block design. *Ann. Math. Statist.27*,846-9

Khatri, C.G., and Shah, K.R. 1975. Exact variance of combined inter- and intra-block estimates in incomplete block designs. *J. Am. Stat. Assoc. 70*: 402-406.

Kiefer,J. 1958. On the Nonrandomized Optimality and Randomized Non-optimality of Symmetrical Designs. *Ann. Math. Statist. 29*, 675-699.

Kiefer, J. 1959. Optimal Experimental Design *J. Roy. Stat. Soc.,Ser. B*, 21, 272-319.

Kiefer, J. and Wolfowitz, J. 1960. Optimum designs in regression problems. *Ann. Math. Statist. 30*, 271-294.

Kiefer, J. 1961. Optimal design in regression problems, II. *Ann. Math. Statist.* 32 298-325.

Kiefer, J. 1975. Construction and Optimality of geberalized Youden dsigns. In: J.N.Srivastava, Ed. , A survey of statistical Design and Linear models.North-Holland,Amsterdam , 333-353.

Kshirsagar,A.M. 1958. A note on incomplete block designs. *Ann. Math. Statist. 29*, 207-10

Kuehl,R.O. 2000. *Design of Experiments:Research design and Analysis*, 2nd edition, Cambridge university press.

Legendre, P., and Fortin, M.J. 1989. Spatial pattern and ecological analysis. *Vegetatio, 80:* 107-138.

Littell, R.C.,G.A.Millikan, W.W.Straup, and R.D.Wolfinger. 1996. SAS System for Mixed Models. SAS Inst., Inc., Cary, NC.

Lockeretz, W. 1987. Establishing the proper role for on-farm research. *Amer. J. Altern. Agric.* 2:132-136.

Lockeretz, W., and M.D. Anderson. 1993. *On-farm research. Ch. 8 in: Agricultural Research Alternatives*. U. Nebraska Press, Lincoln, Nebraska. p. 99-115.

Maklad, M. S. and Nichols, S. T. 1980. *A new approach to model structure discrimination.* IEEE Transactions on System, Man, and Cybernetics 10 78-84.

Magnussen, S. 1990. Application and comparison of spatial models in analyzing tree-genetics field trials. *Can. J. For. Res. 20*: 536-546.

Magnussen, S. 1993. Design and analysis of tree genetic trials.*Can. J. For. Res.23* :1144-1149.

Martérn, B. 1986. *Spatial Variation.* Springer-Verlag, New York.

Martin, R.J and Eccleston, J.A, 1993. Incomplete block designs with spatial layouts when observations are dependent. *J. statist. Plann. and Inf. 3577-92*

Mead. R. and Curnow, R.N. 1983. *Statistical methods in agriculture and experimental biology.* Chapman and Hall, London, UK.

Mead, R. 1994. *The design of Experiments.* Cambridge University Press

McCullagh, P. and Nelder, J.A. 1989. *Generalized Linear models.* 2nd ed. Chapman and Hall: London.

Montgomery, 1997. *Design and Analysis of Experiments,* 4th ed., John Wiley and Sons, New York.

Most, B.M. 1975. Resistance of balanced incomplete block designs. *Ann. Statist. 3* , 1149-1162.

Mote, V.L. 1958. On a min-max property of a balanced incomplete block design. *Ann. Math. Statist. 29*, 910-4

Mutsaers, H.J.W and P.Walker, 1991. *On-farm Research in Theory and Practice.* IITA, Ibadan, Nigeria, pp,39-45.

Mutsaers, H.J.W and P.Walker, 1990. Farmers maize yields in S.W. Nigeria and the effect of variety and fertilizer: an analysis of variability in on-farm trials. *Field Crops Research 23*, 265-278.

Nabugoomu, F. and Allen, O.B., 1990. Estimation of a ratio of variance components using likelihood-based procedure. *In Biometry for Development: Proceedings of the first East and Central Africa Network meeting of the Biometrics Society.* Eds: Patel, M.S. and Nokoe, S. Nairobi: Kenya.

Nelder, J.N.1965. The analysis of experiment with orthogonal block structure. *Proc. R. Statist. Soc.* Lond. A. 283, 147-178.

Nelder, J.N. 1977. A reformulation of the linear models (with discussion). *J. Roy. Statist. Soc.* A 140, 48-78.

Nguyen, N-K. 1993. An algorithm for constructing optimal resolvable block designs. *Commun. Statist. B* 22, 911-923.

Nguyen, N-K. 1994. Construction of optimal incomplete block designs by computer. *Technometrics* 36, 300-307.

Njuho, P.M. 1998. Statistical issues related to on-farm trials. *AFSRE 15th International Symposium Proceedings, pp* . 837-842.

Nokoe, S. 1999. On-farm trials: surgical or preventive approach . *Journal of Forest Resources*, 16(2)(In press).

O'Brien,R.G. and Lohr, V.I. 1984. Power analysis for univariate linear models: the SAS system makes it easy. *SAS Users' Group International: Proceedings of the Ninth Annual Conference. Cary, NC*: SAS Institute, Inc.

Ostle, B. 1963. *Statistics in Research* 2nd Edition. Iowa State Press, Ames, Iowa.

Pardey, P.G., Roseboom, J. and Beintema, N. 1996. Agricultural research in Africa: Three decades of development. ISNAR briefing paper No. 19. The Hague.

Patterson, H.D. and Hunter, E.A. 1983. The efficiency of incomplete block designs in National List and Recommended List cereal variety trials. *Journal of Agricultural Science 101*, 427-433.

Patterson, H.D., and Patterson, L.J. 1976. An algorithmn for constructing α-lattice designs. *Ars Combinatoria, 16 A*: 87-98.

Patterson, L.J., and Williams, E.R. 1976. A new class of resolvable incomplete block designs. *Biometrika, 63*: 83-92.

Pearce, S.C. 1953. *Field experimentation with fruit trees and other perennial plants.* Tech. Commun. Commonwealth Bur. Hort. Plantation Crops, 23.

Pearce, S.C. 1968. The mean efficiency of equi-replicate designs. *Biometrika 55*, 251-253.

Pilarczyk, W. 1991. The efficiency of α-designs in Polish variety testing field trials. *Plant varieties and seeds 4*, 37-41.

Plackett, R.L., and Burman, J.P. 1946. The design of optimum multifactorial experiments. *Biometrika 33*, 305-325.

Preece, D.A., 1982. The design and analysis of Expriments: What has gone wrong? *Utilitas Mathematica,* Vol. 21A, pp. 201-244.

Prescott, P. and Mansson, R. 2001. Robustness of balanced incomplete block designs to randomly missing observations. *J. Statist. Plann. and Inf. 92*, 283-296.

Pukelsheim, F. 1993. *Optimal Design of Experiments.* John Wiley and Sons, Inc., New York.

Ramirez, D.E. 1989. On two noncquivalent measures of complexity. *IEEE Transactions on System, Man and Cybernetics* 19: 1073-1077.

Rao, C.R, 1947. General methods of analysis for incomplete block designs. *J. Amer. Stat. Assoc 42*:115-132

Rao, C.R. 1959. On the construction of partially balanced incomplete block designs with more that three associate classes *Bull. Calcutta Statist. Ass.* 9, 87-92.

Riley, J and Alexander,C. 1997. Statistical literature for participatory on-farm research. *Experimental Agriculture 33*:73-82

SAS Institute Inc. 1989. SAS Technical Report P-188, SAS/QC software examples, Version 6.

Shah, K.R. 1960. Optimality criteria for incomplete block designs. *Ann. Math. Statist. 31*, 791-4.

Scheffe, H. 1959. *The analysis of Variance,* New York: John Wiley & Sons.

Shrikhande, S.S. 1950. The Impossibility of certain Symmetrical balanced incomplete block designs. *Ann. Math. Statist.,* 21, 106-111.

Searle, S. R., 1971. *Linear Model.* New York, NY: John Wiley & Sons, Inc.

Searle, S. 1971a. Topics in variance component estimation. *Biometrics 27*: 176.

Searle, S.R., 1987 *Linear Models for Unbalanced Data.* John Wiley

Schwabe, R. 1996. *Optimum designs for Multi-factor models. Lecture notes in statistics.* Springer, New York.

Silvey, S.D. 1980. *Optimal Designs.* Chapman & Hall, New York.

Skinner H and Mwaniki B M, 1994. DAREP/RTDU Tools and tillage on-farm research: Sept. 1993-August 1994. Dryland Agricultural research and Extension Project, Box 27, Embu, Kenya.

Steiner, K.G. 1987. *On-farm Experimentaion Handbook for Rural Development Projects.*Eschborn,Germany; Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ) GmbH.

Stroud, Ann. 1993. *Conducting on farm experiemnts.*CIAT, Apartado Aereo 6713. Cali, Columbia

St-Pierre, N.R and Jones, L.R. 1999. Interpretation and design of Nonregulatory On-farm feeding Trials. *J.Anim. Sci.* Vol.77, Suppl. 2/J.

Tocher, K.D. 1952. The design and analysis of block experiments(with discussion). *J. Roy .Statist. Soc. B,* 14,45-100.

Verdinelli, I., and Kadane, J.B., 1992. Bayesian designs for maximising information and outcome. *J. Amer. Statist. Ass. 87,* 510-515.

Wald, A. 1943. On the efficient design of statistical investigations. *Ann. Math. Stat,* Vol. 14, pp 134-140.

Williams, E.R., and Matheson, A.C. 1994. *Experimental design and analysis for use in tree improvement.* CSIRO, Victoria, Australia.

Williams, E.R. and Patterson H.D. 1977. Upper bound for efficiency factors in block designs. *Austral. J. Statist.* 19, 194-201.

Winer, B.J. 1971. *Statistical principles in experimental design.* 2nd ed. McGraw-Hill, New York.

Whittinghill, D.C., 1986. *The effect of the loss of observations in a design originally balanced.* Preliminary Report 2. Department of Statistics, Temple University, Philadelphia.

Whittinghill, D.C., 1989. Balanced block designs robust against the loss of a single observation. *J. Statist. Plann. and Inf.* 22,71-80.

Whittinghill, D.C., 1995. A note on the optimality of block designs resulting from the unavailability of scattered observations. *Utilitas Math.* 47, 21-31.

Yates, F. 1933. The analysis of replicated experiments when the field results are incomplete. *Emp. Jour. Exp. Agr.* 1, 129-142.

Yates, F. 1935. Complex experiments. *J.Roy. Statist.Soc(Supp.)* 2 181-247.

Yates, F. 1936. A new method of arranging variety trials involving a large number of varieties. *J.Agric.Sci. 26*, 424-55.

Yates, F.1940. The recovery of inter-block information in balanced incomplete block designs. *Ann.Eugen.* 10, 317-325.

Yates, F. 1975. The early history of experimental design. In a survey of statistical design and linear models, edited by J.N.Srivastava, Amsterdam: North-Holland , pp 581-592.

Yau, S.K., 1997. Efficiency of alpha-lattice designs in international variety yield trials of barley and wheat. *J. of Agric. Science,* Cambridge Univ. Press 128, 5-9.

Zandstra, E.C.Price, J.A. Litsinger and R.A Morris. *A methodology for on-farm cropping systems research,* IRRI, Los Banos,Philippines,pp.104-113.

Zimmerman, D.L., and Harville, D.A. 1991. A random field approach to the analysis of field plot experiments. *Biometrics,* 47, 223-239.

Appendix A

Designs considered

A.1 Design matrix for the complexity criterion examples

For design 1

$$\begin{pmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{12} \\ y_{22} \\ y_{32} \\ y_{13} \\ y_{23} \\ y_{33} \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \mathbf{1} & 0 & 0 & 1 & 0 & 0 \\ \mathbf{1} & 0 & 1 & 0 & 1 & 0 & 0 \\ \mathbf{1} & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & \mathbf{1} & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & \mathbf{1} & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & \mathbf{0} & \mathbf{1} & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ f_1 \\ f_2 \\ f_3 \end{pmatrix} + \varepsilon_{ij} \qquad \text{(A.1)}$$

For design 2.

$$\begin{pmatrix} y_{11} \\ y_{21} \\ y_{22} \\ y_{32} \\ y_{13} \\ y_{33} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & \mathbf{1} & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & \mathbf{1} \\ 1 & 0 & 0 & 1 & 0 & 0 & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ f_1 \\ f_2 \\ f_3 \end{pmatrix} + \varepsilon_{ij} \qquad \text{(A.2)}$$

A.2 RCBD and IBD trial design layout

Treatment are listed as 1,2,3,4,5,6,7 denoting A,B,C,D,E,F,G respectively in the designs.

A.2.1 Randomized complete block design

Treatment combinations on each unit of the design for RCBD

Blocks						
1	5	5	5	5	1	3
3	6	4	4	6	6	1
7	2	7	3	7	3	7
4	1	2	6	3	5	4
2	4	1	2	2	4	2
5	7	6	1	4	7	5
6	3	3	7	1	2	6

A.2.2 Incomplete block design

Treatment combinations on each unit of the design for IBD

Blocks						
1	2	6	6	7	3	5
7	7	2	5	4	4	3
3	4	7	1	3	6	2
6	1	5	4	5	2	1

A.3 Genstat analyses based on standard errors

```
GenStat Release 4.21  (PC/Windows XP)     04 october 2002 17:57:14

 Randomized complete Blocks example

 Treatment factors are listed in the order: Treat1

model[dispersion=1]yield3 fit farm3 + Treat3

 141  model[dispersion=1]yield3
 142  fit farm3 + Treat3

 Regression Analysis
```

```
 Response variate: yield3
     Fitted terms: Constant + farm3 + Treat3

*** Summary of analysis ***

             d.f.          s.s.          m.s.       v.r.
Regression     12            0.            0.       0.00
Residual       36            0.            0.
Total          48            0.            0.
* MESSAGE: ratios are based on dispersion parameter with value 1

Standard error of observations is fixed at 1.00

 Estimates of parameters

                              estimate         s.e.       t(*)
Constant                         0.000        0.515       0.00
farm3 2                          0.000        0.535       0.00
farm3 3                          0.000        0.535       0.00
farm3 4                          0.000        0.535       0.00
farm3 5                          0.000        0.535       0.00
farm3 6                          0.000        0.535       0.00
farm3 7                          0.000        0.535       0.00
Treat3 2                         0.000        0.535       0.00
Treat3 3                         0.000        0.535       0.00
Treat3 4                         0.000        0.535       0.00
Treat3 5                         0.000        0.535       0.00
Treat3 6                         0.000        0.535       0.00
Treat3 7                         0.000        0.535       0.00
* MESSAGE: s.e.s are based on dispersion parameter with value 1

Parameters for factors are differences compared with the reference
level:
             Factor  Reference level
              farm3  1
             Treat3  1


Balanced incomplete block design example

 model[dispersion=1]yield1 fit blk + treat4

 164  model[dispersion=1]yield1
 165  fit blk + treat4

 Regression Analysis

 Response variate: yield1
     Fitted terms: Constant + blk + treat4

*** Summary of analysis ***

             d.f.          s.s.          m.s.       v.r.
Regression     12            0.            0.       0.00
```

```
Residual        15          0.          0.
Total           27          0.          0.
* MESSAGE: ratios are based on dispersion parameter with value 1

Standard error of observations is fixed at 1.00

 Estimates of parameters
                            estimate        s.e.      t(*)
Constant                       0.000       0.681      0.00
blk 2                          0.000       0.756      0.00
blk 3                          0.000       0.756      0.00
blk 4                          0.000       0.756      0.00
blk 5                          0.000       0.756      0.00
blk 6                          0.000       0.756      0.00
blk 7                          0.000       0.756      0.00
treat4 2                       0.000       0.756      0.00
treat4 3                       0.000       0.756      0.00
treat4 4                       0.000       0.756      0.00
treat4 5                       0.000       0.756      0.00
treat4 6                       0.000       0.756      0.00
treat4 7                       0.000       0.756      0.00
* MESSAGE: s.e.s are based on dispersion parameter with value 1

Parameters for factors are differences compared with the reference
level:
             Factor  Reference level
                blk  1
             treat4  1

Example for Unbalanced incomplete block/farm design 1

 171  model[dispersion=1]yield1
 172  fit farm + treatment1

 Regression Analysis

 Response variate: yield1
     Fitted terms: Constant + farm + treatment1

*** Summary of analysis ***

             d.f.        s.s.        m.s.      v.r.
Regression     11          0.          0.      0.00
Residual       16          0.          0.
Total          27          0.          0.
* MESSAGE: ratios are based on dispersion parameter with value 1

Standard error of observations is fixed at 1.00

*** Estimates of parameters ***
                            estimate        s.e.      t(*)
Constant                       0.000       0.715      0.00
farm 2                         0.000       0.946      0.00
farm 3                         0.000       0.810      0.00
```

farm 4	0.000	0.786	0.00
farm 5	0.000	0.769	0.00
farm 6	0.000	0.734	0.00
treatment1 2	0.000	0.727	0.00
treatment1 3	0.000	0.730	0.00
treatment1 4	0.000	0.765	0.00
treatment1 5	0.000	0.746	0.00
treatment1 6	0.000	0.746	0.00
treatment1 7	0.000	0.736	0.00

* MESSAGE: s.e.s are based on dispersion parameter with value 1

Parameters for factors are differences compared with the reference level:

Factor	Reference level
farm	1
treatment1	1

```
rkeep v = covv
print covv
173  rkeep v= covv
174  print covv
```

covv

Constant	0.5110				
farm 2	-0.3404	0.8947			
farm 3	-0.2877	0.4149	0.6565		
farm 4	-0.3563	0.4480	0.3753	0.6182	
farm 5	-0.3515	0.4668	0.3904	0.4077	0.5914
farm 6	-0.3363	0.4469	0.3852	0.3992	0.4071
treatment1 2	-0.2659	0.0130	-0.0665	0.0081	0.0502
treatment1 3	-0.2671	0.0083	-0.0703	0.0607	0.0043
treatment1 4	-0.1820	-0.2154	-0.1740	-0.0513	-0.1124
treatment1 5	-0.1649	-0.2237	-0.1037	-0.1120	-0.1167
treatment1 6	-0.1649	-0.2237	-0.1037	-0.1120	-0.1167
treatment1 7	-0.1781	-0.1037	-0.1641	-0.0938	-0.0976

	Constant	farm 2	farm 3	farm 4	farm 5
farm 6	0.5390				
treatment1 2	0.0055	0.5292			
treatment1 3	0.0035	0.2687	0.5327		
treatment1 4	-0.1082	0.2654	0.2807	0.5845	
treatment1 5	-0.1117	0.2468	0.2479	0.3038	0.5559
treatment1 6	-0.1117	0.2468	0.2479	0.3038	0.3059
treatment1 7	-0.0963	0.2666	0.2676	0.2935	0.2759

	farm 6	treatment1 2	treatment1 3	treatment1 4	treatment1 5
treatment1 6	0.5559				
treatment1 7	0.2759	0.5410			

	treatment1 6	treatment1 7

Example for Unbalanced incomplete block design 2

```
 167  model[dispersion=1]yield2
 168  fit farm2 + treat2

 Regression Analysis

 Response variate: yield2
     Fitted terms: Constant + farm2 + treat2

 Summary of analysis

             d.f.         s.s.         m.s.      v.r.
Regression     10           0.           0.      0.00
Residual       17           0.           0.
Total          27           0.           0.
* MESSAGE: ratios are based on dispersion parameter with value 1

Standard error of observations is fixed at 1.00

*** Estimates of parameters ***
                            estimate         s.e.      t(*)
Constant                       0.000        0.707      0.00
farm2 2                        0.000        0.887      0.00
farm2 3                        0.000        0.758      0.00
farm2 4                        0.000        0.737      0.00
farm2 5                        0.000        0.735      0.00
treat2 2                       0.000        0.707      0.00
treat2 3                       0.000        0.707      0.00
treat2 4                       0.000        0.741      0.00
treat2 5                       0.000        0.741      0.00
treat2 6                       0.000        0.761      0.00
treat2 7                       0.000        0.761      0.00
* MESSAGE: s.e.s are based on dispersion parameter with value 1

Parameters for factors are differences compared with the reference
level:
             Factor  Reference level
              farm2  1
             treat2  1
 rkeep v=cova1
print cova1
 169  rkeep v=cova1
 170  print cova1

                  cova1

     Constant      0.5000
      farm2 2     -0.3333      0.7876
      farm2 3     -0.3333      0.4121      0.5745
      farm2 4     -0.3333      0.4501      0.3783      0.5429
      farm2 5     -0.3333      0.4712      0.3805      0.4121      0.5408
     treat2 2     -0.2500      0.0000      0.0000      0.0000      0.0000
     treat2 3     -0.2500      0.0000      0.0000      0.0000      0.0000
     treat2 4     -0.1667     -0.1969     -0.1030     -0.1125     -0.1178
```

treat2 5	-0.1667	-0.1969	-0.1030	-0.1125	-0.1178
treat2 6	-0.1667	-0.2117	-0.0545	-0.1210	-0.1579
treat2 7	-0.1667	-0.2117	-0.0545	-0.1210	-0.1579

	Constant	farm2 2	farm2 3	farm2 4	farm2 5
treat2 2	0.5000				
treat2 3	0.2500	0.5000			
treat2 4	0.2500	0.2500	0.5492		
treat2 5	0.2500	0.2500	0.2992	0.5492	
treat2 6	0.2500	0.2500	0.3029	0.3029	0.5788
treat2 7	0.2500	0.2500	0.3029	0.3029	0.3288

	treat2 2	treat2 3	treat2 4	treat2 5	treat2 6
treat2 7	0.5788				

	treat2 7

```
Standard versus control treatment example output
FACTOR [modify=yes;nvalues=35;levels=7] block FACTOR
[modify=yes;nvalues=35;levels=8] treat variate[values=35(0)]yield
VARIATE [nvalues=35] yield

model[dispersion=1]yield fit block + treat

 Regression Analysis
 Response variate: yield
     Fitted terms: Constant + block + treat

 Summary of analysis

             d.f.          s.s.          m.s.       v.r.
Regression     13            0.            0.       0.00
 Residual      21             0             0
 Total         34            0.            0.

*** Estimates of parameters ***
                              estimate          s.e.      t(*)
Constant                         0.000         0.645      0.00
block 2                          0.000         0.667      0.00
block 3                          0.000         0.667      0.00
block 4                          0.000         0.667      0.00
block 5                          0.000         0.667      0.00
block 6                          0.000         0.667      0.00
block 7                          0.000         0.667      0.00
treat 2                          0.000         0.745      0.00
treat 3                          0.000         0.745      0.00
treat 4                          0.000         0.745      0.00
treat 5                          0.000         0.745      0.00
treat 6                          0.000         0.745      0.00
treat 7                          0.000         0.745      0.00
treat 8                          0.000         0.645      0.00
* MESSAGE: s.e.s are based on dispersion parameter with value 1
```

Parameters for factors are differences compared with the reference level:

Factor	Reference level
block	1
treat	1

NB: The standard errors(s.e.) for treat 2 to treat 8 are respectively the s.e.'s for the treatment comparison B-A, C-A, D-A, E-A, F-A, and G-A.

rkeep v=cova print cova

```
 41  rkeep v=cova
 42  print cova
```

cova

Constant	0.4167				
block 2	-0.1667	0.4444			
block 3	-0.2222	0.2222	0.4444		
block 4	-0.1667	0.2222	0.2222	0.4444	
block 5	-0.2222	0.2222	0.2222	0.2222	0.4444
block 6	-0.1667	0.2222	0.2222	0.2222	0.2222
block 7	-0.2222	0.2222	0.2222	0.2222	0.2222
treat 2	-0.2778	-0.0556	0.0556	-0.0556	0.0000
treat 3	-0.2222	-0.1111	-0.0556	-0.0556	-0.0556
treat 4	-0.2778	-0.0556	0.0556	0.0000	0.0556
treat 5	-0.2222	-0.0556	0.0000	-0.1111	-0.0556
treat 6	-0.2222	-0.1111	-0.0556	-0.1111	0.0000
treat 7	-0.2778	0.0000	0.0000	-0.0556	0.0556
treat 8	-0.2500	-0.0556	0.0000	-0.0556	0.0000

	Constant	block 2	block 3	block 4	block 5
block 6	0.4444				
block 7	0.2222	0.4444			
treat 2	0.0000	0.0556	0.5556		
treat 3	-0.1111	0.0000	0.2778	0.5556	
treat 4	-0.0556	0.0000	0.2778	0.2778	0.5556
treat 5	-0.1111	-0.0556	0.2778	0.2778	0.2778
treat 6	-0.0556	-0.0556	0.2778	0.2778	0.2778
treat 7	-0.0556	0.0556	0.2778	0.2778	0.2778
treat 8	-0.0556	0.0000	0.2778	0.2778	0.2778

	block 6	block 7	treat 2	treat 3	treat 4
treat 5	0.5556				
treat 6	0.2778	0.5556			
treat 7	0.2778	0.2778	0.5556		
treat 8	0.2778	0.2778	0.2778	0.4167	

	treat 5	treat 6	treat 7	treat 8

The covariance matrix cova is given above and we can use the results from the matrix to obtain the standard errors of other

comparisons.

Appendix B

Examples of designs considered

B.1 Constrained cases

Design scenario 1

Blocks	1	2	3	4	5	6	7	8	9	10
Plot 1	C	A	A	C	E	C	B	A	E	E
plot 2	D	B	F	D	D	E	F	B	B	A
plot 3	A	F	D	B	F	F	C	E	D	C
plot 4	0	0	0	0	0	0	0	0	0	0

Design scenario 2

Block	1	2	3	4	5	6
plot 1	0	0	0	0	0	0
plot 2	A	B	C	D	E	F

Design scenario 3

Plot 1	A	A	A	A	A	B	B	B	B	C	C	C	D	D	E
Plot 2	B	C	D	E	F	C	D	E	F	D	E	F	E	F	F
Plot 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Design scenario 4

Blk	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
plt1	A	A	A	A	A	A	A	A	A	A	B	B	B	B	B	B	C	C	C	D
plt2	B	B	B	B	C	C	C	D	D	E	C	C	C	D	D	E	D	D	E	E
plt3	C	D	E	F	D	E	F	E	F	F	D	E	F	E	F	F	E	F	F	F
plt4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Design scenario 5

Block	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
plt 1	A	A	A	B	B	C	B	A	A	A	A	A	A	B	A
plt 2	B	B	B	C	C	D	C	B	C	C	D	B	B	D	C
plt 3	C	C	C	D	D	E	E	E	D	D	E	D	D	E	E
plt 4	D	E	F	E	F	F	F	F	E	F	F	E	F	F	F
plt 5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Design scenario 6

block	1	2	3	4	5	6
plt 1	A	A	A	A	B	A
plt 2	B	B	B	B	C	C
plt 3	C	C	C	D	D	D
plt 4	D	D	E	E	E	E
plt 5	E	F	F	F	F	F
plt 6	0	0	0	0	0	0

Design scenario 7 RCBD

Block	1	2	3	4	5	6
PLT 1	A	A	A	A	A	A
PLT 2	B	B	B	B	B	B
PLT 3	C	C	C	C	C	C
PLT 4	D	D	D	D	D	D
PLT 5	E	E	E	E	E	E
PLT 6	F	F	F	F	F	F
PLT 7	0	0	0	0	0	0

B.2 Unconstrained designs

Treatment combinations on each unit of the design

BIBD1

Blocks	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Units														
1	4	6	2	2	4	4	1	1	3	4	3	1	2	6
2	7	8	3	7	3	2	6	8	7	8	5	7	7	2
3	2	7	4	5	7	5	5	3	1	6	6	4	6	1
4	8	5	1	3	6	8	4	5	8	3	2	5	1	8

Treatment factors are listed in the order: Treatments

BIBD2

Blocks	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Units														
1	1	1	2	1	2	3	4	2	3	1	2	1	1	1
2	5	2	3	3	4	5	6	3	4	4	5	3	2	2
3	7	6	7	4	5	6	7	4	5	5	6	6	4	3
4	8	8	8	8	8	8	8	6	7	6	7	7	7	5

Treatment factors are listed in the order: Treatments

CYCLIC DESIGN

Blocks	1	2	3	4	5	6	7	8
Units								
1	2	5	2	5	1	3	5	3
2	8	8	4	7	7	2	3	8
3	6	4	8	3	5	6	2	1
4	7	6	1	4	6	4	1	7

RCBD

Blocks	1	2	3	4	5	6	7	8	9	10
Units										
1	3	8	8	2	6	8	7	5	2	1
2	4	4	1	5	7	3	4	2	8	6
3	1	3	3	8	5	1	1	8	6	8
4	2	5	7	1	4	7	2	7	3	7
5	5	1	6	7	1	6	5	3	7	5
6	6	6	5	3	8	4	3	4	5	2
7	8	2	2	6	2	2	6	1	1	4
8	7	7	4	4	3	5	8	6	4	3

B.3 Design layouts for the Numerical example

Plot	Blk1	Blk2	Blk3	Blk4	Blk5	Blk6
1	T1	T1	T1	T1	T1	T2
2	T2	T2	T2	T2	T3	T3
3	T3	T3	T3	T4	T4	T4
4	T4	T4	T5	T5	T5	T5
5	T5	T6	T6	T6	T6	T6
6	T7	T7	T7	T7	T7	T7

Plts	Blk1	Blk2	Blk3	Blk4	Blk5	Blk6	Blk7	Blk8	Blk9	Blk10
1	T1	T1	T1	T1	T1	T2	T2	T2	T3	T3
2	T2	T2	T3	T4	T5	T3	T4	T5	T4	T4
3	T3	T4	T5	T6	T6	T6	T5	T6	T5	T6
4	T7	T7	T7	T7	T7	T7	T7	T7	T7	T7

The Letter T stands for treatment and the treatments numbers 1, 2, 3, 4, 5, 6, 7 correspond to the various alphabetical letters A, B, C, D, E, F, O in that order.

Layout of the On-farm trials in Uganda(Pallisa district)RCB design

Plots	Blk1	Blk2	Blk3	Blk4	Blk5	Blk6
1	T3	T4	T5	T6	T1	T2
2	T2	T7	T4	T2	T5	T7
3	T4	T2	T3	T7	T6	T1
4	T6	T3	T1	T3	T2	T5
5	T1	T5	T7	T4	T3	T6
6	T5	T6	T2	T5	T7	T4
7	T7	T1	T6	T1	T4	T3

Appendix C

SAS programmes

C.1 The SAS code for example on efficiency measures

```
data can;
do treat=1 to 7;
do blk=1 to 6;
output;
end;
end;
proc optex data=can seed = 23568 coding=orth;
class treat blk;
model treat blk;
generate n=24;
run;
output out=bibd blockname=blk;
proc print data=bibd;
run;
```

C.2 The SAS codes for examples on Power analysis

The designs data and how it was set up is presented in Table C.1.

BIBD1

```
data bibd17;
input blk plot trt yield1;
datalines;
   (DESIGN SEATS HERE)
;
```

```
run;
proc mixed noprofile;
class blk plot trt;
model
yield1=blk plot trt;
parms(4)/noiter;
model yield1=trt;
random blk;
parms(2.5)(4)/noiter;
contrast 'trt x lin' trt 1 1 1 -1 -1 -1 0;
ods ouput contrast=nc;
data pwr; ndf=1;
ddf=8;
f=3.04;
alpha=0.05;
ncparm=ndf*f;
fcrit=finv(1-alpha,ndf,ddf,0);
power=1-probf(fcrit,ndf,ddf,ncparm);
proc print;
var ndf ddf
ncparm alpha fcrit power;
run;
```

BIBD2

```
data bibd27;
input blk2 plot2 trt2 yield2;
datalines;

   (DESIGN SEATS HERE)
;

run;
proc mixed noprofile;
class blk2 plot2 trt2;
model yield2=blk2 plot2 trt2;
parms(4)/noiter;
model yield2=trt2;
random blk2;
parms(2.5)(4)/noiter;
contrast 'trt x lin' trt2 1 1 1 -1 -1 -1 0;
ods output contrast=nc;
run;
data pwrd;
```

```
ndf=1;
ddf=15;
f=4.57;
alpha=0.05;
ncparm=ndf*f;
fcrit=finv(1-alpha,ndf,ddf,0);
power=1-probf(fcrit,ndf,ddf,ncparm);
proc print;
var ndf ddf ncparm alpha fcrit power;
run;
```

BIBD3

```
data bibd37;
 input bllk2 pllot2 trrt2 yield3;
 datalines;
 (DESIGN SEATS HERE)
  ;
 run;
 proc mixed noprofile;
 class bllk2 pllot2 trrt2;
 model yield3=bllk2 pllot2 trrt2;
 parms(4)/noiter;
 model yield3=trrt2;
 random bllk2;
 parms(2.5)(4)/noiter;
 contrast 'trt x lin' trrt2 1 1 1 -1 -1 -1 0;
 ods output contrast=nc;
 run;
 data pwrd;
 ndf=1;
 ddf=22;
  f=3.84;
 alpha=0.05;
 ncparm=ndf*f;
 fcrit=finv(1-alpha,ndf,ddf,0);
 power=1-probf(fcrit,ndf,ddf,ncparm);
 proc print;\\
 var ndf ddf ncparm alpha fcrit power;
 run;
```

BIBD47

```
data bibd47;
```

```
input bllk4 pllot4 trrt4 yield4;
datalines;
(DESIGN SEATS HERE)
;
run;
proc mixed noprofile;
class bllk4 pllot4 trrt4; model yield4=bllk4 pllot4 trrt4;
parms(4)/noiter;
model yield4=trrt4;
random bllk4;
parms(2.5)(4)/noiter;
contrast 'trt x lin' trrt4 1 1 1 -1 -1 -1 0;
ods output contrast=nc;
run;
data pwrd;
ndf=1;
ddf=36;
f=5.12;
alpha=0.05;
ncparm=ndf*f; fcrit=finv(1-alpha,ndf,ddf,0);
power=1-probf(fcrit,ndf,ddf,ncparm);
proc print;
var ndf ddf ncparm alpha fcrit power;
run;
```

Cyclic

```
data bibdcyc;
input blcy plcy trcy yieldcy;
datalines;
(DESIGN SEATS HERE
;
run;
proc mixed noprofile;
class blcy plcy trcy;
model yieldcy=blcy plcy trcy; parms(4)/noiter;
model yieldcy=trcy;
random blcy;
parms(2.5)(4)/noiter;
contrast 'trt x lin' trcy 1 1 1 -1 -1 -1 0;
ods output contrast=nc;
run;
data pwrd;
ndf=1;
ddf=8;
```

```
f=1.8;
alpha=0.05;
ncparm=ndf*f;
fcrit=finv(1-alpha,ndf,ddf,0);
power=1-probf(fcrit,ndf,ddf,ncparm);
proc print;
var ndf ddf ncparm alpha fcrit power;
run;
```

RCBD

```
data rcbdc; input blrcbd plrcbd trrcbd yieldrcbd;
datalines;
(DESIGN SEATS HERE)
run;
proc mixed noprofile;
class blrcbd plrcbd trrcbd;
model yieldrcbd=blrcbd plrcbd trrcbd;
parms(4)/noiter;
model yieldrcbd=trrcbd; random blrcbd;
parms(2.5)(4)/noiter;
contrast 'trt x lin' trrcbd 1 1 1 -1 -1 -1 0;
ods output contrast=nc;
run;
data pwrd; ndf=1;
ddf=36;
f=5.01;
alpha=0.05;
ncparm=ndf*f;
fcrit=finv(1-alpha,ndf,ddf,0);
power=1-probf(fcrit,ndf,ddf,ncparm);
proc print;
var ndf ddf ncparm alpha fcrit power;
run;
```

Appendix D

Data for unconstrained scenario designs

Table D.1: Data table for the Unconstrained scenario designs. Designs are listed in the order of block, plot, treatment, and yield.

BIBD17	BIBD27	BIBD37	BIBD47	Cyclic	RCBD
1 1 2 5.2	1 1 5 9.2	1 1 6 6.64	1 1 2 5.97	1 1 6 6.4	2 1 3 4.2
2 1 4 4.1	2 1 1 3.8	2 1 1 4.5	2 1 2 6.1	2 1 5 7.5	3 1 7 3.7
3 1 7 5.2	3 1 6 6.4	3 1 7 4.8	3 1 1 4.63	3 1 5 6.4	4 1 5 6.5
4 1 1 4.9	4 1 5 8.1	4 1 5 6.65	4 1 1 4.8	4 1 5 9.8	5 1 4 4.1
5 1 7 4.3	5 1 5 7.9	5 1 3 3.9	5 1 4 5.9	5 1 4 5.1	6 1 6 5.5
6 1 5 6.5	6 1 3 4.2	6 1 4 4.2	6 1 5 8.4	6 1 7 4.6	7 1 7 4.1
7 1 3 5.2	7 1 4 6.8	7 1 2 6.2	7 1 5 8.1	7 1 2 5.9	1 2 3 5.7
1 2 6 6.9	1 2 7 3.4	8 1 4 5.2	8 1 6 6.7	1 2 2 5.9	2 2 4 4.5
2 2 5 9.1	2 2 7 5.5	9 1 2 6.8	9 1 6 7.1	2 2 7 3.6	3 2 3 4.9
3 2 6 7.5	3 2 4 5.6	10 1 4 5.1	10 1 4 5.6	3 2 6 5.4	4 2 2 5.1
4 2 5 7.9	4 2 6 6.8	11 1 5 7.7	11 1 2 4.8	4 2 2 6.3	5 2 6 5.8
5 2 2 4.8	5 2 1 5.2	12 1 1 4.7	12 1 4 5.2	5 2 6 8.1	6 2 7 4.5
6 2 6 6.2	6 2 6 7.2	13 1 5 7.5	13 1 3 4.7	6 2 3 4.9	7 2 3 4.2
7 2 4 6.6	7 2 6 6.6	14 1 5 8.1	14 1 6 6.4	7 2 4 3.9	1 3 5 6.9
1 3 1 4.5	1 3 4 5.5	1 2 1 5.6	1 2 1 4.9	1 3 7 6.3	2 3 1 4.9
2 3 2 5.4	2 3 4 4.1	2 2 7 3.9	2 2 4 4.6	2 3 4 7.3	3 3 1 4.8
3 3 4 5.4	3 3 5 6.6	3 2 2 4.1	3 2 4 5.2	3 3 1 5.1	4 3 7 4.2
4 3 7 4.9	4 3 7 5.3	4 2 6 7.1	4 2 3 5.3	4 3 3 4.9	5 3 5 7.5
5 3 3 3.7	5 3 2 6.1	5 2 2 6.4	5 2 3 4.2	5 3 3 5.3	6 3 4 5.1
6 3 3 5.1	6 3 1 3.3	6 2 5 7.6	6 2 2 5.5	6 3 1 3.8	7 3 6 6.3
7 3 1 4.7;	7 3 3 6.6	7 2 4 5.6	7 2 2 4.1	7 3 1 5.2;	1 4 4 4.7
	1 4 3 4.3	8 2 6 5.4	8 2 7 4.8		2 4 2 6.7
	2 4 2 4.6	9 2 1 4.3	9 2 7 5.5		3 4 5 7.1
	3 4 1 4.5	10 2 3 4.2	10 2 1 5.4		4 4 3 3.9
	4 4 2 6.2	11 2 3 3.9	11 2 5 7.6		5 4 2 5.9
	5 4 3 3.9	12 2 5 7.9	12 2 3 4.6		6 4 2 5.1
	6 4 7 4.8	13 2 3 5.1	13 2 5 7.1		7 4 2 4.7
	7 4 2 6.1;	14 2 6 6.3	14 2 2 6.1		1 5 6 6.1
		1 3 4 5.1	1 3 4 4.8		2 5 7 4.8
		2 3 3 3.3	2 3 5 7.6		3 5 4 5.5
		3 3 1 5.5	3 3 7 4.3		4 5 4 6.6
		4 3 7 4.6	4 3 4 4.3		5 5 7 5.8
		5 3 6 6.7	5 3 5 7.3		6 5 5 7.9
		6 3 2 4.9	6 3 7 4.9		7 5 4 5.5
		7 3 7 5.1	7 3 1 5.3		1 6 7 4.9
		8 3 3 4.6	8 3 1 4.1		2 6 5 9.1
		9 3 6 6.4	9 3 4 5.8		3 6 6 6.6
		10 3 7 5.1	10 3 7 5.1		4 6 6 7.6
		11 3 1 4.4	11 3 7 4.8		5 6 1 4.1
		12 3 4 5.5	12 3 7 4.2		6 6 1 4.2
		13 3 2 6.5	13 3 1 4.3		7 6 1 4.6
		14 3 7 3.8;	14 3 3 4.7		1 7 2 6.3
			1 4 3 5.1		2 7 6 6.7
			2 4 6 6.6		3 7 2 4.9
			3 4 6 5.8		4 7 1 5.1
			4 4 2 5.7		5 7 3 5.4
			5 4 6 6.4		6 7 3 4.1
			6 4 1 5		7 7 5 7.9;
			7 4 6 6.7		
			8 4 3 5.3		
			9 4 2 5.4		
			10 4 5 6.8		
			11 4 3 4.9		
			12 4 5 8.3		
			13 4 6 5.3		
			14 4 7 5.1;		

Table D.1. Data table for the Unconstrained scenario designs. Designs are listed in the order of block, Plot, Treatment, and Yield.