ON THE DESIGN AND ANALYSIS OF FIELD EXPERIMENTS

M-1565 April 2001

Walter T. Federer
and
Jose Crossa

Keywords: block design, row-column design, augmented design, parsimonious design, model selection, design axioms, software, competition, border effect.

Abstract:

Topics covered are statistical design axioms, plot technique, experiment design selection, block experiment designs, row-column experiment designs, unreplicated or screening experiment designs, exploratory model selection, multi-site/year trials, nd parsimonious experiment designs. In line with the axiom "design for the experiment, do not experiment for the design", two simple methods for constructing block experiment designs are described. In addition, a software toolkit for constructing optimal or near optimal randomized plans of experiment designs for many situations is discussed. The effect of response model selection is illustrated with examples. Several suggestions for increasing the efficiency of a plant breeding program are given. Augmented and parsimonious experiment designs can be utilized to increase the efficiency of plant breeding program. Some remarks on the construction of these designs are given.

M-1565

ON THE DESIGN AND ANALYSIS OF FIELD EXPERIMENTS

by

Walter T. Federer
343 Warren Hall
Cornell University
Ithaca, New York 14853
e-mail: WTF1@cornell.edu

and

Jose Crossa
Biometrics and Statistics Unit
International Maize and Wheat Improvement Center (CIMMYT)
Apdo Postal 6-641
Mexico, D. F., Mexico
e-mail: J.CROSSA@CGIAR.org

Abstract

Topics covered are statistical design axioms, plot technique, experiment design selection, block experiment designs, row-column experiment designs, unreplicated or screening experiment designs, exploratory model selection, multi-site/year trials, and parsimonious experiment designs. In line with the axiom "design for the experiment, do not experiment for the design", two simple methods for constructing block experiment designs are described. In addition, a software toolkit for constructing optimal or near optimal randomized plans of experiment designs for many situations is discussed. The effect of response model selection is illustrated with examples. Several suggestions for increasing the efficiency of a plant breeding program are given. Augmented and parsimonious experiment designs can be utilized to increase the efficiency of plant breeding programs. Some remarks on the construction of these designs are given.

Keywords: block design, row-column design, augmented design, parsimonious design, model selection, design axioms, software, competition, border effect.

BU-DURUM April 2001

In the Technical Report Series of the Department of Biometrics, Cornell University, Ithaca, New York 14853

ON THE DESIGN AND ANALYSIS OF FIELD EXPERIMENTS

By

Walter T. Federer and Jose Crossa

Introduction

Spatial variability in fields is a universal phenomenon that affects the detection of treatment differences in agricultural experiments by inflating the estimated experimental error variance. Researchers desiring to conduct field trials are faced with this dilemma. They meet this problem by using an appropriate design and layout of the experiment and by using improved statistical methodology for statistical analyses. Owing to the large numbers of genotypes involved in plant breeding programs, small plots are the rule. The smallest unit of area allotted to one genotype or treatment is denoted as an *experimental unit* (e.u.) and has been called a plot (also called a plat in older literature). The arrangement of the e.u.s in an experiment is known as the *experiment design* (ED), and the selection of the treatments (genotypes) to be included in the experiment is known as the *treatment design*. Starting with Sir Ronald A. Fisher's three design principles of blocking, randomization, and replication, many types of EDs have evolved and are still evolving to meet the various situations encountered by researchers. Randomization is necessary in order to obtain an unbiased estimate of the error variance. Many types of EDs such as a completely randomized experiment design, randomized complete block experiment design (RCBD), split-plot experiment design, split-block experiment design, incomplete block (lattice) experiment design (IBD), Latin square ED, Youden ED, lattice square and lattice rectangle EDs, etc. have been described and used in published literature. Lattice and lattice square experiment designs are popular designs for plant breeding evaluation trials. When the incomplete block and lattice square experiment designs were introduced in the 1930s and 1940s, experimental data were analyzed on simple desk calculators. Consequently, the emphasis was on obtaining designs easy to construct and simple to analyze. The use of catalogued lattices and lattice square, such as may be found in Cochran and Cox (1957) and Federer (1955), e.g., is limited owing to the wide variety of numbers of genotypes occurring in practice. This limitation has been lifted owing to developments such as Patterson and Williams (1976), e.g., who devised the resolvable (the incomplete blocks form a complete block or replicate for the set of genotypes) incomplete block designs called alpha designs. A main advantage of alpha designs over the traditional lattices is the flexibility to accommodate any number of

genotypes in any number of replicates and to be able to have incomplete blocks of different sizes. When the field layout is in a row-column shape, either for the entire experiment or within each complete block, EDs for any number of genotypes and replicates can be developed (Nguyen and Williams, 1993; John and Williams, 1995; Federer, 1998b) that control variability in two directions. The row-column EDs have two block components, i.e., blocks in rows and blocks in columns. Likewise, several software packages are available for obtaining randomized plans of these designs for any number of genotypes and replicates. When the entire experiment is laid out in a row-column arrangement, it may be desirable to assure that genotypes do not occur more often than once in a row or a column of the experiment. The so-called "latinized designs" accomplish this. Also, it may be desirable to restrict randomization of genotypes in such a way that certain groups of genotypes do not occur together so that genotypic interference can be avoided. Latinized alpha lattice and row-column designs as well as neighbor restricted designs can be generated using the software package Alpha+ (1996).

Optimal plot size and shape, border effects, competition between e.u.s, and experimental techniques may be factors contributing to the variability in experimental results. These topics are discussed in the following section. Sections on selecting an ED, blocked EDs, row-column EDs, unreplicated or screening EDs, exploratory model selection, multi-site/year trials, and parsimonious EDs are presented.

The following eight axioms should be followed whenever a field experiment is contemplated. Discussion of the first five axioms may be found in Federer (1984, 1993).

Axiom I: *A complete, precise, and rigorous description of the population to which inferences are to be made is essential if inferences are to have any meaning.*

Axiom II: *Design for the experiment, do not experiment for the design.*

Axiom III: *Use the minimum amount of blocking possible to control heterogeneity among the experimental units (e.u's).*

Axiom IV: *Treatments with different numbers of randomizations will have different numbers of replication, will have different e.u.'s, and will have different error variances.*

Axiom V: *A valid error variance for the difference between two treatment effects must contain all sources of variation in the e.u.'s except that due to the treatments themselves.*

The following three axioms need to be considered when selecting a response model and statistical analysis for the experiment:

Axiom VI: *The experimental layout and the conduct of the experiment are part of the design as far as spatial variation and experimental variation are concerned.*

Almost universally, statisticians and data analysts only consider the experiment design (plan) selected for an experiment, but the actual layout of the experiment in the

field is mostly ignored. For example, suppose eight complete blocks of a randomized complete block design (RCBD) with seven treatments are selected. Then, the actual field layout is eight rows (the blocks) and seven columns. When selecting a response model, the data analyst should consider this as a row-column design (RCD) and not a RCBD in order to account for the spatial variation present in the experiment. During the course of conducting an experiment certain events may occur that introduce heterogeneity into an experiment. For example, an experimenter may be forced by weather to stop planting in the middle of a block, insects or disease may invade from one side of an experiment, a cultivator may have one shoe going one inch deeper than the others, water may stand on part of the experiment, sprinkle irrigation may not be uniformly distributed, non-uniformity among note-takers may occur, or any one of a variety of other causes may occur to introduce variation into the experimental responses. All such items need consideration when selecting a response model and statistical analysis.

Axiom VII: *In order to extract the maximum amount of information from an experiment, an appropriate response model must be selected. An analysis of experimental data must be one that accounts for the heterogeneity present in the experiment.*

As stated above, all sources of variation in experimental results need to be considered in model selection. More often a data analyst will not know what types of variation are present and will need to do exploratory model selection such as described by Federer, Crossa, and Franco (1998), for example. Available computer software simplifies the task of model selection to a great extent. Several forms of spatial analyses for designed experiments are available (Cullis and Gleeson, 1991; Gilmour *et al.*, 1997; Federer, Newton and Altman, 1997, e.g.).

Axiom VIII: *The treatment design (the selection of entities or treatments to be included in an experiment) for an experiment is a vital and crucial component for a successful experiment. The treatments should be selected to maximize information in an experiment and this leads to efficient experimentation.*

The inclusion of points of reference, controls, or standards in an experiment is vital for the success of many experiments. For example, in comparing yields for new genotypes, one could select the top 10%, say. Without the inclusion of a standard genotype with which to compare the new genotypes, selecting the top 10% may be meaningless as all could be far below the yield of the standard genotype.

Plot Technique

Variability is an omnipresent feature of field experiments. The experimenter needs to utilize procedures that control heterogeneity present in an experiment. Heterogeneity can arise from variation present in the experimental sites and from effects occurring during the conduct of the experiment. Heterogeneity in the error variance can also arise from selecting an inappropriate statistical model for data analysis. Some ways of controlling heterogeneity are (Federer, 1955, 1984):

(i) Refining experimental techniques,
(ii) selecting uniform material and/or a uniform environment,
(iii) grouping or blocking material into uniform subgroups, and
(iv) measuring related variables and using covariance techniques.

Inappropriate model selection can have considerable effect on the size of the error mean square. As an example, consider the data in Table 12.3 of Cochran and Cox (1957). Their lattice square analysis results in an error mean square of 9.57. Alternatively, one may replace the column variable with the variable differential linear gradients within rows in the response model to obtain an error mean square of 4.06 (Federer, Crossa, and Franco, 1998). Their model would require 9.57/4.06 = 2.4 times more replication in order to achieve the same standard error of a difference between two means. To demonstrate that the previous example is not an isolated case, a trend analysis using the model

Count = replicate + treatment + linear row gradient with replicate (R1) + quadratic row gradient within replicate (R2) + linear column gradient within replicate (C1) + R1*C1 + R1*C2 + R2*C2 + R2*C3 + error

for the data in Table 12.5 of Cochran and Cox (1957), results in an error mean square of 8.95 as compared to their value of 22.67. Ri is the ith orthogonal polynomial row regression coefficient of yield on position and Cj is the jth orthogonal polynomial column regression coefficient of yield on position. Again, their model would require 22.67/8.95 = 2.5 times more replication to achieve the same standard error of a difference between two treatments. Other examples are easily found.

Another variable that considerably affects variation between e.u.'s is competition. Designs for measuring competition have been presented by Federer and Basford (1992), for example. The proper choice of an e.u. and the spacing between e.u.'s can eliminate this variable. Consider a maize trial laid out in two rows of an e.u. with one meter (m) between all maize rows. Instead of using this arrangement, the two (or more) rows of an e.u. may be spaced 0.25 m apart and the e.u.'s are placed 1.75 m apart. This arrangement preserves the same density per hectare as the one-meter apart arrangement but eliminates the effect of competition between e.u.'s. A rule of thumb for grass species is that the distance between e.u.'s should equal the height of the plants. For wheat varietal trials, e.u's one m apart should suffice. Such an arrangement has the advantage that cultivation for weed control can be continued longer than for equally spaced rows. In the early stages of a breeding program involving large numbers of new untried genotypes, the experimenter may wish to use single-row e.u.'s. The density within a row could be increased to have the usual density per hectare and the rows could be the height of the plants apart to effectively eliminate competition between e.u.s..

Border effect can also affect heterogeneity in the experimental site. The choice of border material can be used to diminish this effect. Federer and Basford (1992) suggest using a mixture of all treatments in the experiment as the border material in order to equalize competition effects.

Selecting an Experiment Design

Following Axioms I, II, and III above, a plan should be selected to fit the experiment under consideration. The experimenter should not have to change the plan to fit a plan from a catalogue of plans such as given in Cochran and Cox (1957), for example. Software packages and methods are available to construct plans for almost any situation. The experiment design plans may be resolvable, that is all v treatments (genotypes, lines, varieties) occur in one complete block. A non-resolvable experiment design is one in which the v treatments are not grouped into complete blocks. To accommodate any number v of genotypes in an experiment, unequal block sizes, say k and k + 1, may be used. Such arrangements have been discussed by Patterson and Williams (1976) and Khare and Federer (1981). Their results eliminate the need to add or delete genotypes in a proposed experiment in order to fit a catalogued ED.

Block Experiment Designs

Block designs may be complete (all v treatments are included in each block) or incomplete (a subset of the v treatments appears in a block). A complete block experiment design (CBD) is used in situations where it is presumed that all the v experimental units (e.u.s) in a block are relatively homogeneous and variability cannot be further controlled by subdividing into smaller blocks. When this is not the case, an incomplete block design (IBD) is indicated. The incomplete block size k should be one that groups the experimental area into homogeneous sub-groups. For example, suppose v = 228 genotypes. It would usually be very difficult to select uniform blocks of size 228 e.u.s. Therefore, incomplete blocks are to be considered. For v = 228, k = 2, 3, 4, 6, 8, or 12 are possible incomplete block sizes. Such IBDs may be easily constructed as shown by Patterson *et al.* (1976 and 1985), Khare and Federer (1981), and Federer (1995). Also, these authors show that the incomplete block size may vary, say k and k + 1, to accommodate various values of v. To illustrate a method of construction, let v = 45, k = 5, and r = 3. For the first replicate of s = 9 incomplete blocks, the numbers 1-45 are written as

Replicate or complete block 1

1	10	19	28	37
2	11	20	29	38
3	12	21	30	39
4	13	22	31	40
5	14	23	32	41
6	15	24	33	42
7	16	25	34	43
8	17	26	35	44
9	18	27	36	45

The numbers in the rows form the incomplete block arrangement for replicate 1. The incomplete block arrangements (groupings) for replicate 2 are formed by taking the main

right diagonal of replicate 1 which is 1, 11, 21, 31, and 41. These numbers form the first incomplete block of replicate 2. The numbers within columns are cyclically permuted to form the incomplete blocks of replicate 2. Replicate 3 is formed by taking the main right diagonal of replicate 2 and cyclically permuting the numbers within columns. The two resulting arrangements are

Replicate 2					Replicate 3				
1	11	21	31	41	1	12	23	34	45
2	12	22	32	42	2	13	24	35	37
3	13	23	33	43	3	14	25	36	38
4	14	24	34	44	4	15	26	28	39
5	15	25	35	45	5	16	27	29	40
6	16	26	36	37	6	17	19	30	41
7	17	27	28	38	7	18	20	31	42
8	18	19	29	39	8	10	21	32	43
9	10	29	30	40	9	11	22	33	44

Pairs of numbers either occur together or they do not to form a two associate, 0 and 1, class design and is an optimal IBD. There are $v(v-1)/2 = 45(44)/2 = 990$ pairs of treatments. 270 of the pairs occur together in incomplete blocks to form first associates. 720 pairs do not occur together in incomplete blocks to form zeroth associates. Additional replicates may be obtained by continuing the above procedure. The incomplete blocks above are randomly allotted to the incomplete blocks in the field and then the numbers within each incomplete block are randomly allotted to the $k = 5$ e.u.s within an incomplete block.

Another simple method (Federer, 1995) to form incomplete blocks of size $k = 2$ or 3 is given below. For $k = 2$, $v = 2s$ treatments and $s = v/2$ incomplete blocks per complete block. For the first replicate, arrange the v treatments as

1	2	3	4	5	...	v/2-1	v/2
v/2+1	v/2+2	v/2+3	v/2+4	v/2+5	...	v-1	v

to form v/2 incomplete blocks of size $k = 2$. The second replicate is formed by moving the items in row 2 above one place to the right and cyclically permuting the items. Row one remains as is.

1	2	3	4	5	...	v/2-1	v/2
v	v/2+1	v/2+2	v/2+3	v/2+4	...	v-2	v-1

The third replicate is formed from the second replicate in the same manner and is

1	2	3	4	5	...	v/2-1	v/2
v-1	v	v/2+1	v/2+2	v/2+3	...	v-3	v-2

If this process is continued to obtain $r = v/2$ replicates (complete blocks), each treatment in row 1 will appear once with each of the treatments in row 2, that is, each treatment has

v/2 first associates. Since none of the treatments in row 1 appear with any of the other treatments in row 1 in an incomplete block, each treatment will have v/2 – 1 zeroth associates. This 0,1 association scheme cannot be improved upon.

For k = 3, v = 3s treatments and s = v/3 incomplete blocks. The first replicate is

1	2	3	4	...	v/3-1	v/3
v/3+1	v/3+2	v/3+3	v/3+4	...	2v/3-1	2v/3
2v/3+1	2v/3+2	2v/3+3	2v/3+4	...	v-1	v

The second replicate is formed from replicate 1 by retaining row 1, moving the treatments in row 2 one place to the right, and moving the treatments in row 2 two places to the right and is

1	2	3	4	...	v/3-1	v/3
2v/3	2v/3+1	v/3+2	v/3+3	...	2v/3-1	2v/3-2
v-1	v	2v/3+1	2v/3+2	...	v-3	v-2

This process may be continued to obtain r = v/3 replicates of an IBD with only zero and first associates. Such an IBD is efficient.

However, an even easier method for constructing an IBD is to use a software package such as GENDEX (Nguyen, 2001). This toolkit can be used to obtain randomized plans for r replicates for resolvable IBDs for any v divisible by k. For non-resolvable IBDs, the requirement is that vr = bk, where b is the number of incomplete blocks of size k. For the above example with v = 228, a few simple commands will give the printed output for r = 6 replicates, say, and incomplete blocks of k = 4, say. Such a toolkit is a great time and labor saving device and produces optimal or near optimal IBDs.

Row-Column Experiment Designs

A set of v treatments replicated r times may be placed in a row-column design such as the Latin square, Youden, or other design. A set of vr = bk e.u.s may be placed in k rows and b columns. Randomized plans for such designs may be obtained using a software toolkit such as GENDEX (Nguyen, 2001). For example, v = 25 treatments with r = 4 replicates may be placed in a 10 ×10 RCD, in a 5 × 20 RCD, or in a 2 × 50 RCD.

Since plant breeders have experiments with large v, resolvable row-column or lattice rectangle designs (RRCD) will be of greater interest to them. In a RRCD, the v treatments are arranged in k rows and s columns within each complete block. Thus ks = v. The more well-known RRCDs are the balanced lattice square and the semi-balanced lattice square experiment designs where k = s (Cochran and Cox, 1957; Federer, 1955, e.g.). The grouping of treatments needs to vary from complete block to complete block for at least two of the complete blocks in order to obtain solutions for treatment effects. In constructing RRCDs, an attempt is made to optimize the plan by grouping treatments in such a manner as to minimize the variance of a difference between two treatment means. RRCDs may be obtained using the GENDEX toolkit. For example, a RRCD for v = 45, r = 4, k = 5, and s = 9 in randomized form is easily obtained and is given below.

38	35	18	41	28	05	26	11	40		26	41	19	31	42	01	34	25	16
20	15	25	27	24	14	32	02	29		45	29	03	43	08	22	20	07	05
31	21	04	44	07	06	36	16	09		11	04	06	09	27	33	12	35	24
19	13	22	39	45	12	37	34	42		14	10	32	39	17	21	40	38	13
33	08	30	03	01	17	43	10	23		23	37	28	18	36	02	15	44	30
41	17	32	07	05	15	04	43	23		06	08	37	07	40	11	25	19	15
33	16	42	13	38	36	45	25	06		01	32	35	18	14	29	36	16	03
14	22	31	27	01	10	40	28	35		41	38	31	12	43	13	02	23	22
18	44	11	20	37	24	03	21	34		20	04	17	42	24	28	39	21	26
08	29	30	26	09	19	02	12	39		30	34	45	10	44	33	05	27	09

An iterative procedure is used to obtain an experiment design. The above design was obtained at try 68 and had an efficiency rating of 97.25% relative to the best row-column design theoretically possible. Another attempt or more tries may result in a design with a higher efficiency rating. However, 97.25% is quite close to 100% and therefore is near-optimal. The efficiency measure referred to here considers only intra-block, not inter-block, efficiency and is for the plan and not the conducted experiment.

For the example in the previous section with v = 228, an experimenter may wish to use a RRCD with v = 228, r = 6, k = 12, and s = 19 (Federer, 1998b). A reason for this choice is that row-column designs are efficient in allowing for response models that allow many types of experimental variation to be taken into account (Culllis and Gleeson, 1991; Federer, Crossa, and Franco, 1998, e.g.).

Unreplicated or Screening Experiment Designs

In the early stages of a breeding program, a plant breeder is faced with evaluating the performance of large numbers of genotypes. Frequently the seed supply is limited but even if it is not, the large number of genotypes necessitates using a single e.u. for a genotype. One of the early procedures was to plant one row for each line or genotype, one after the other in long, continuous rows. This was called the line-to-row method. At this stage many lines were discarded based mostly on characteristics other than yield. Then, the survivors from the first screening would be placed in yield trials replicated either at a single site or with one replicate at each of several sites, or they might be screened further using one of the following methods.

A second, and popular, procedure is the one known as systematically spaced checks. In this procedure, a standard check genotype is systematically spaced every kth e.u. Several statistical procedures have been devised over the years to compare the yield of a new genotype with the standard variety. Some experimenters have used the standard check in every other e.u., some in every third e.u., others in every tenth e.u., and so forth. This procedure can require an inordinate amount of space, labor, and other resources devoted to check plots of the single standard genotype.

A third procedure used in the screening of genotypes for yield and other characteristics is augmented experiment designs. An augmented experiment design

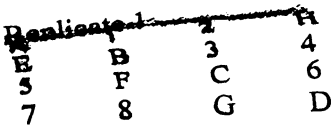
(AED) is constructed by selecting the c check or standard genotypes to be included and then selecting an appropriate experiment design for these check genotypes. Then, the block sizes or the number of rows and columns are increased to accommodate n new genotypes. To illustrate, let c = 15 checks be arranged in r = 5 replicates and b = rs = 25 incomplete blocks of size k = 3. Let n = 300 new genotypes. By enlarging the 25 incomplete blocks from k = 3 to k = 15 to accommodate 3 + 12 = 15 e.u.s, the 300 new genotypes can be put into these 25 incomplete blocks. The 12 new genotypes and the three checks are randomly allotted to the 15 e.u.s in each of the 25 incomplete blocks. The 15 check genotypes may be two standard genotypes and 13 promising and surviving new genotypes from previous screening cycles. These designs allow screening early generation genotypes at the same time as evaluating promising new genotypes. Combining the evaluation of different cycles of selection in a plant breeding program leads to efficient experimentation. Since the estimation of block effects and error mean squares do not depend on the yields of the unreplicated new treatments, the experimenter may decide not to harvest some of the new genotypes owing to unfavorable characteristics such as lodging, disease, etc. This will not affect the resulting statistical analyses that are based on check responses.

The above example used an IBD for the check genotypes. Any experiment design may be used to obtain an augmented ED. A row-column ED with r rows and c columns with k < r or c check genotypes and n new genotypes may also be used. For example, let r = 6, c = 9, k = 3 check genotypes (A, B, C), and n = 36 new genotypes. A plan for this augmented row-column ED is

A	1	2	B	3	4	C	5	6
7	8	B	9	10	C	11	12	A
13	B	14	15	C	16	17	A	18
B	19	20	C	21	22	A	23	24
25	26	C	27	28	A	29	30	B
31	C	32	33	A	34	35	B	36

Not all row and column effects have solutions in the above design but functions of row and column effects can be used, e.g., linear trend in rows, quadratic trend in rows, cubic trend in rows, etc. and the same for column effects (see Federer, 1998a, e.g.). The numbers 1 to 36 are randomly allotted to the 36 new genotypes and the rows and columns are randomized.

Federer (2000, 2001) shows how to construct augmented RRCDs and presents a statistical analysis with computer software code for these designs. These experiment designs require 2s or 3s check genotypes where s is the number of rows of a RRCD. The largest n can be accommodated when k = s to obtain augmented lattice square experiment designs. It should be pointed out that a large proportion of the 2k or 3k checks should be promising new genotypes requiring further testing. In this manner, the number of plots allocated to standard genotypes is minimized and the evaluation procedure made efficient. To illustrate, let k = s = 4, 2k = 8 check genotypes (A, B, C, D, E, F, G, H), n = rk(k – 2) = 32 new genotypes, and k = r = 4 complete blocks or replicates of the check genotypes. An unrandomized plan for the design is



Replicate 2			
A	9	10	G
H	B	11	12
13	E	C	14
15	16	F	D

Replicate 3			
A	17	18	F
G	B	19	20
21	H	C	22
23	24	E	D

Replicate 4			
A	25	26	E
F	B	27	28
29	G	C	30
31	32	H	D

The check genotypes, which could be one or two standards and 7 or 6 new promising genotypes requiring further testing, may be placed on *any* two of the right diagonals. Again, not all row and column effects will have solutions which necessitates use of some function of these effects such as row-linear, column linear, and perhaps the interaction of these regressions.

Of these three procedures, augmented designs have several advantages over the other two procedures. These are:

(i) More than one check genotype can be included in the experiment.
(ii) Standard errors of a difference between two new genotypes are available.
(iii) Standards errors of a difference between two checks are available.
(iv) Standard errors of a difference between a check and a new genotype is available.
(v) Fewer cycles of selection are possible.

Patterson and Silvey (1980) presented a plan for introducing a new genotype into production. It would appear that the use of augmented experiment designs could be utilized to improve the efficiency of the breeding and selection program. It is possible that the number of years suggested by the authors could be reduced through the use of such screening designs as augmented EDs. In the second or third cycle of evaluation, AEDs could be used at each of several sites (Federer, Reynolds, and Crossa, 2001). This could decrease the number of cycles in evaluating a set of new genotypes. AEDs allow comparative evaluations throughout all cycles of a program thus allowing fewer cycles.

Moreau *et al.* (2000) studied the efficiency of marker-assisted-selection (MAS) with phenotypic selection under different circumstances including traits sensitive to genotype × environment interaction. They concluded that:

(i) When genotype × environment interactions are included in the model, it is always optimal to perform one replication per trial.
(ii) When investment is high enough, it appears optimal to do only a small number of trials even when genotype × environment interaction is important.

(iii) It may be useful to use checks and/or replicate a small subset of the population sample within each trial.
(iv) MAS uses fewer trials than phenotypic selection.

It would appear that the AED admirably fits their conclusions and that the procedure of Sprague and Federer (1951) for optimum allocation of resources to maximize genetic advance would be useful here. AEDs may be used for mass screening on the basis of phenotypic selection and then MAS used on the survivors from the initial screening. A combination of AEDS, PEDS (discussed later), phenotypic selection, and MAS will result in a reduction of the number trials (site, year), costs, and cycles of selection.

For mapping specific genomic segments affecting quantitative trait loci (QTL) and for studying QTL × environment interaction (QTL × E) with the aid of molecular markers, a set of families from a suitable population such as F_2, backcross, recombinant inbred, or double haploid are grown in field trials in different environments. The precision by which the different regions of the chromosomes and the magnitude of their effects are estimated depends, among other factors, on the number of families included in the field evaluation. Usually no more than 200 families are evaluated but with 500 or more families QTL estimates will be more precise. The lack of sufficient seeds as well as limited resources for managing large replicated trials in several environments precludes testing very large numbers of families in several environments. One possible solution to these problems could be the use of unreplicated field designs as discussed above in a variety of environments. AEDs may help to increase the precision of estimating QTL effects and QTL × E effects. The *a priori* control of local variability by using a suitable number of replicated checks in an AED and the *a posteriori* exploratory model selection analysis will help to increase the precision of QTL estimation.

Exploratory Model Selection

Probably statistics courses and textbooks in the past have been responsible for the notion that a response model must be selected at the time of planning the experiment and that there is one and only one response model for a given experiment design. These ideas have persisted even though experimenters have used several transformations for a data set and selected one for the statistical analysis. The use of transformations is just one form of exploratory data analysis. The idea of exploratory data analysis has been in the literature since the late 1940s and the 1950s. The idea of exploratory model selection appeared in papers in the mid-1950s. Spatial analyses for designed experiments have been around for many years. A classical paper on exploratory model selection is the one by Box and Cox (1964). Despite these results, there are still individuals who criticize exploratory model selection. True, procedures for picking a model from a class of models could be improved and could be made less subjective. However, this is no reason for not considering a class of plausible models for a data set and then selecting an appropriate model for the statistical analysis. Such procedures do have an effect on the Type I error but the effect is usually minimal. Any model selection procedure may be investigated via simulations as was done by Federer, Crossa, and Franco (1998) and Federer, Wolfinger, and Crossa (2001), e.g. Exploratory model selection is made easy when computer codes

are available such as the seven codes found in the papers by Federer (2001), Federer and Wolfinger (2001a, 2001b, 2001c, 2001d), Federer, Singh, and Wolfinger (2001), and Federer, Wolfinger, and Crossa (2001).

An example will demonstrate how effective exploratory model selection is under a complex spatial variation pattern. The example is for seven treatments on tobacco plants designed as a RCBD but laid out as an eight row by seven column design. Some of the models in the class of models investigated by Federer, Crossa, and Franco (1998) for Y = plant height are:

$$Y = \text{row} + \text{treatment} + \text{error} \quad (1)$$

$$Y = \text{row} + \text{column} + \text{treatment} + \text{error} \quad (2)$$

$$Y = R1 + R2 + R3 + R5 + R6 + R7 + C1 + C2 + C3 + C5 + C1*R1 + C2*R1 + C2*R3 + C3*R2 + C4*R1 + C4*R2 + \text{treatment} + \text{error} \quad (3)$$

$$Y = \text{row} + \text{column} + C1*R1 + C2*R1 + C2*R3 + C3*R2 + C4*R1 + C4*R2 + \text{treatment} + \text{error} \quad (4)$$

$$Y = \text{row} + C2(\text{row}) + C3(\text{row}) + C4(\text{row}) + \text{treatment} + \text{error} \quad (5)$$

Ri is the ith degree orthogonal polynomial regression for row positions on responses and Cj is the jth orthogonal polynomial regression for column positions on responses. The residual (error) mean squares were 30,228 for model (1), 7,352 for model (2), 4,204 for model (3), 4,418 for model (4), and 11,310 for model (5). Either model (3) or (4) appears to be the appropriate model to account for the spatial variation present in this data set. There were dramatic differences in the residual mean squares for the different models. The F-values for treatments versus residual were also quite different.

The recovery of inter-random effect information should always be performed as this leads to treatment means with smaller standard errors. Many computer packages have software for accomplishing the recovery of random-effect information using mixed model procedures.

When deciding to perform exploratory model selection, one needs to know if the standard textbook response model for a particular ED suffices. An experimenter usually has some idea of what constitutes a well-controlled experiment. In cereal trials, if the coefficient of variation is 5%, say, little is to be gained from exploratory model selection. However, if the coefficient of variation is15%, say, it would appear that a response model is not appropriate and a search should be made for a more appropriate one. In other words, if the residual error mean square is relatively small, further reduction would appear to be unlikely.

Multi-Site and/or Multi-Year Experiments

When new genotypes are released, they are for a particular region. In order to test the adaptation of released genotypes, they need to go through multi-site testing (see Patterson and Silvey, 1980, e.g.). These trials are to determine the general adaptability of a genotype over a region. It is often presumed that a random selection of sites is made but this is never the case. Sites for testing are selected for a variety of reasons, e.g, the willingness of a farmer to allow a test on his farm. Test sites may also be selected to

represent a variety of conditions found in the region. Prior to release, a genotype is often tested over several years. This testing is essential to determine how a genotype interacts with the environments it is expected to encounter. A genotype should perform relatively well over all environments, i.e., it should have good general adaptability.

In performing statistical analyses of multi-site and/or multi-year trials, several methods are available, each with their advantages and drawbacks. Following results of Federer, Reynolds, and Crossa (2001), it is recommended that

(i) the most appropriate model for each experiment be obtained,
(ii) the best estimate of the treatment means be obtained,
(iii) the means are standardized using the transformation of mean/standard error of a mean,
(iv) the analysis for environments and treatments be obtained, and
(v) the fact that the expected error mean square of the standardized means is the parameter one be utilized.

This method allows for different response models and different experiment designs at each site and year. A second method for combining results over sites is discussed by Federer, Reynolds, and Crossa (2001). It is a modification of the one presented in Cochran and Cox (1957), Chapter 14.

Parsimonious Experiment Designs

Site to site and year to year variation can often be identified. Such factors as fertility level, date of planting, date of harvesting, moisture level, disease level, insect level, etc. contribute to site to site and year to year variation. Can the effect of these factors on genotype performance be evaluated at a single site? The answer is that they can be. Since it is easier to add levels rather than subtract amounts of these factors, a site that is limiting in these factors could be utilized to assess their effects on genotype performance. Such an experiment would diminish if not eliminate the need for most multi-site testing. For such an experiment, the experimenter could use a factorial treatment design but this would require a very large experiment. The class of designs described by Federer and Scully (1993) and Federer (1993), Chapter 10, can be used effectively to reduce such multi-factor experiments in an efficient manner. They denoted these designs as parsimonious experiment designs (PEDS). For a PED, the levels of one or more factors are varied *within an e.u.* and a *response function* of yield is used rather than a single value such as weight per e.u. PEDs allow a wide coverage of levels of a factor and efficient utilization of material and space. The experimenter knows the levels of these factors when PEDs are used whereas they are usually unknown in multi-site trials. Two or more factors may be varied within a single e.u. of a PED. For example, fertility may be varied in one direction of an e.u. and date of planting in another direction. Various other procedures have been discussed by these authors. Harvesting costs are increased for PEDs but travel and other off site expenses are eliminated. Further discussion of PEDs may be found in Federer and Scully (1993) and Federer (1993).

Each set of genotypes and environments requires individual attention rather than resorting to generalizations. However, from some experimental results on maize, it was found that most of the site by genotype, year by genotype, and site by year by genotype interactions could be accounted for by date of planting. Biological date of planting rather than calendar date of planting is the important date to keep in mind. An optimal biological date of planting will vary from a calendar date from site to site and year to year. April 15 may be the optimum biological date in one year at one site and May 1 in another year at the same site.

Discussion and Summary

The versatility and availability of computer software and the developments in statistical design and analysis over the past twenty years need to be incorporated into all plant breeding programs. Computer constructed plans for experiments and exploratory model selection procedures are available to optimize plant breeding procedures and to obtain optimal or near optimal EDs and statistical analyses for experimental data. Using these more efficient procedures allows more research information to be obtained with less personnel, material, and finances.

In selecting an ED for an experiment, the experimenter should use his/her knowledge of spatial variation that is likely to be encountered for the planned experiment. Using this knowledge, the most appropriate ED that accounts for the presumed spatial variation should be selected. Row-column and resolvable row-column EDs are more capable of controlling spatial variation than block designs, usually fit the experimental lay-out better (Axiom VI), and more complex statistical response models using interactions of row and column effects are available.

Augmented experiment designs and parsimonious experiment designs will be useful for increasing the number of families that need to be tested when mapping QTLs and doing marker assisted selection. This will allow the testing of more families in a larger number of environments in an efficient manner.

Literature Cited

Alpha+ (1996). Experimental designs for variety trials and many-treatment experiments. CSIRO Forestry and Forest Products, P. O. Box 4008, Queen Victoria Terrace, Canberra ACT 2600, Australia, and Biomathematics and Statistics Scotland, The University of Edinburgh, James Clerk Maxwell Building, The King's Building, Edinburgh EH9 3JZ, Scotland.

Box, G. E. P. and D. R. Cox (1964). An analysis of transformations. Journal of the Royal Statistical Society, Series B 26:211-252.

Cochran, W. G. and G. M. Cox (1957). *Experimental Designs, second edition.* John Wiley & Sons, Inc., New York.

Cullis, B. R. and A. C. Gleeson (1991). Spatial analysis of field experiments – An extension to two directions. Biometrics 47:1449-1460.

Federer, W. T. (1955). *Experimental Design: Theory and Application.* The Macmillan Co., New York First printing 1955; second printing 1962. Oxford and IBH Publishing Co., New Delhi, Bombay, and Calcutta. Indian Edition, first printing 1967; second printing 1974.

Federer, W. T. (1984). Principles of statistical design with special reference to experiment and treatment designs. In *Statistics: An Appraisal- Proceedings of the 50th Anniversary Conference*, Iowa State Statistical Laboratory (H. A. David and H. T. David, Editors), pp. 77-104.

Federer, W. T. (1993). *Statistical Design and Analysis for Intercropping Experiments. Volume I: Two Crops.* Springer-Verlag, Berlin, Heidelburg, New York, Chapter 10.

Federer, W. T. (1995). A simple procedure for constructing experiment designs with incomplete block sizes 2 and 3. Biometrical Journal 37(8):899-907.

Federer, W. T. (1998a). Recovery of interblock, intergradient, and intervariety information in incomplete block and lattice rectangle designed experiments. Biometrics 54:471-481.

Federer, W. T. (1998b). A simple procedure for constructing resolvable row-column experiment designs. BU-1438-M in the Technical Report Series of the Department of Biometrics, Cornell University, Ithaca, NY.

Federer, W. T. (2000). Construction and analysis of an augmented lattice square design. BU-1485-M in the Technical Report Series of the Department of Biometrics, Cornell University, Ithaca, New York.

Federer, W. T. (2001). Analysis for an experiment designed as an augmented lattice square design. In *Formulas, Software, and Techniques for Geneticists and Breeders* (Manjit S. Kang, editor), Food Products Press, New York, to appear.

Federer, W. T. and K. E. Basford (1991). Competing effects designs and models for two dimensional field arrangements. Biometrics 47:1461-1472.

Federer, W. T., J. Crossa, and J. Franco (1998). Forms of spatial analyses with mixed model effects and exploratory model selection. BU-1406-M in the Technical Report Series of the Department of Biometrics, Cornell University, Ithaca, NY.

Federer, W, T., R. C. Nair, and D. Raghavarao (1975). Some augmented row-column designs. Biometrics 31:361-373.

Federer, W. T., E. A. Newton, and N. S. Altman (1997). Combining standard block analyses with spatial analyses under a random effects model. In *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions* (T. G. Gregoire *et al.*, Editors), Springer, New York, Heidelburg, pp. 373-386.

Federer, W. T. and D. Raghavarao (1975). On augmented designs. Biometrics 31: 29-35.

Federer, W. T., M. Reynolds, and J. Crossa (2001). Combining results from augmented designs over sites. Agronomy Journal 93:389-395.

Federer, W. T, and B. T. Scully (1993). A parsimonious statistical and breeding procedure for evaluating and selecting desirable characteristics over environments. Theoretical and Applied Genetics 86:612-620.

Federer, W. T. and C. S. Schlottfeldt (1954). The use of covariance to control gradients in experiments. Biometrics 10:282-290.

Federer, W. T., M. Singh, and R. D. Wolfinger (2001). SAS/GLM and SAS/MIXED for trend analyses using Fourier and polynomial regression for centered and non-centered variables. In *Formulas, Software, and Techniques for Geneticists and Breeders* (Manjit S. Kang, editor), Food Products Press, New York, to appear.

Federer, W. T. and R. D. Wolfinger (2001a). Augmented row-column design and trend analyses. In *Formulas, Software, and Techniques for Geneticists and Breeders* (Manjit S. Kang, editor), Food Products Press, New York, to appear.

Federer, W. T. and R. D. Wolfinger (2001b). Code for simulating degrees of freedom for items in principal components analysis of variance. In *Formulas, Software, and Techniques for Geneticists and Breeders* (Manjit S. Kang, editor), Food Products Press, New York, to appear.

Federer, W. T. and R. D. Wolfinger (2001c). PROC GLM and PROC MIXED codes for trend analyses for row-column designed experiments. In *Formulas, Software, and Techniques for Geneticists and Breeders* (Manjit S. Kang, editor), Food Products Press, New York, to appear.

Federer, W. T. and R. D. Wolfinger (2001d). PROC GLM and PROC MIXED for trend analyses of incomplete block and lattice rectangle designed experiments. In *Formulas, Software, and Techniques for Geneticists and Breeders* (Manjit S. Kang, editor), Food Products Press, New York, to appear.

Federer, W. T., R. D. Wolfinger, and J. Crossa (2001). Principal components (PC) and additive main effects and multiplicative interaction (AMMI) trend analyses for incomplete block and lattice rectangle designed experiments. In *Formulas, Software, and*

Techniques for Geneticists and Breeders (Manjit S. Kang, editor), Food Products Press, New York, to appear.

Gilmour, A. C., B. R. Cullis, and A. P. Verbyla (1997). Accounting for natural and extraneous variation in the analysis of field experiments. Journal of Agricultural, Biological, and Environmental Sciences 2(3):269-293.

John, J. A. and E. R. Williams (1995). *Cyclic and Computer Generated Designs.* Chapman and Hall, London.

Khare, M. and W. T. Federer (1981). A simple construction procedure for resolvable incomplete block designs for any number of treatments. Biometrical Journal 23:121-132.

Moreau, L., S. Lemarie, A. Charcosset, and A. Gallais (2000). Economic efficiency of one cycle of marker-assisted selection. Crop Science 40:329-337.

Nguyen, N-K (2001). A toolkit for generating designs of experiments. http://designcomputing.hypermart.net/gendex.

Nguyen, N-K and E. R. Williams (1993). An algorithm for constructing optimal resolvable row-column designs. Australian Journal of Statistics 35:363-370.

Patterson, H. D. and V. Silvey (1980). Statutory and List Recommended Trials of crop varieties in the United Kingdom. Journal of the Royal Statistical Society, Series A 143(3):219-252.

Patterson, H. D. and E. R. Williams (1976). A new class of resolvable incomplete block designs. Biometrika 63:83-92.

Patterson, H. D., E. R. Williams, and L. Patterson (1985). A note on resolvable incomplete block designs. Biometrical Journal 27:75-79.

Sprague, G. F. and W. T. Federer (1951). A comparison of variance components in corn yield trials: II. Error, year × variety, location × variety, and variety components. Agronomy Journal 43:535-541.