RECOVERY OF INTERBLOCK, INTERGRADIENT, AND INTERVARIETY INFORMATION IN INCOMPLETE BLOCK AND LATTICE RECTANGLE DESIGNED EXPERIMENTS

Walter T. Federer

Biometrics Unit, 337 Warren Hall, Cornell University, Ithaca, NY 14853

BU-1315-MB

April 1997

Recovery of Interblock, Intergradient, and Intervariety Information in Incomplete Block and Lattice Rectangle Designed Experiments

Walter T. Federer

Biometrics Unit, 337 Warren Hall, Cornell University, Ithaca, NY 14853, USA

SUMMARY

Spatial analysis and blocking analysis of experimental results are treated separately in the literature. Here we combine these analyses into a single analysis. The information arising from the random nature of different gradients within incomplete blocks is used to adjust treatment means. We extend Cox's (1958) idea of differential gradients within columns of a Latin square to within blocks for incomplete block and row-column designed experiments, and, in addition, treat them as random effects. With this analysis, the restrictions on randomization due to1 blocking are taken into consideration whereas they are often ignored in spatial analysis literature. Some comments on designing experiments and analyzing experimental results to control heterogeneity are presented. A numerical example illustrates the computational procedure and indicates effect of alternative analyses. The class of augmented experiment designs has been found useful for experiments involving comparisons of standard check treatments with a set of new and untried treatments, usually with one replicate. Interreplicate, interblock, interrow, and/or intercolumn information is available to use in obtaining solutions for new treatment effects. Since the new treatment effects are often considered to be random effects, their distributional properties may be used to increase the efficiency of the experiment. We demonstrate the statistical procedures for recovering this information in block and row-column designs using mixed model procedures.

Key words and phrases: Post-blocking; Covariates; Differential trends; Trend analysis; Mixed model; BLUP; Augmented designs; Interregression information.

1. Introduction

Many unplanned events may occur during the course of conducting an experiment. A planned statistical analysis often needs to be altered or changed in order to accommodate the variation caused by unplanned events. One type of this phenomenon is gradients (trends) in responses that occur within the blocks, rows, and/or columns of the design. Much has been published on spatial analyses but little on connecting spatial analyses with standard block design analyses (see Cox, 1958, and Williams, 1986, for exceptions) and on taking account of the random effects nature of the gradients within the blocking categories, i.e., the recovery of intergradient as well as interblock information. In some types of experiments, part or all of the treatments (varieties) may be considered to be random effects. The recovery of intergradient, interblock, interregression, and intervariety information forms the topic of this paper. These analyses follow the general mixed model approach which has appeared in literature since the late thirties (see, e.g., Yates 1939, 1940a,b) up to the present time (e.g., Searle *et al.* 1992).

. Cox (1958) has described an analysis for a Latin square designed experiment for the situation where there are differential gradients in each column, such as might be found for cows in different parts of their lactation periods. Standard textbook analysis would be invalid for this situation. He considers these gradients as fixed effects whereas differential gradients are treated as random effects in this paper.

Appropriate response models and statistical analyses are necessary to obtain correct analyses of data. No single response model and analysis fits all experiments even with the same experiment design. The model and analysis need to fit the actual situation and all the information in an experiment needs to be extracted. Hence, it is inappropriate to ignore interblock, interrow and intercolumn, interregression, and intervariety information when present. Solutions for treatment effects recovering this information are known to have smaller mean square error. Ignoring interblock information is akin to ignoring whole plot information in a split plot design.

In some situations part or all of the variety effects may be considered to be random. Such is the case when new genotypes are to be screened for further testing. The class of augmented designs was devised to include a large number of new genotypes which usually appear once in an experiment with a number of check or standard varieties being included r times. Other designs are possible (see, e.g., Cullis *et al.* 1989). The new genotypes are considered to be random effects and the checks are fixed effects.

In the following, the above situations are formulated using a form for a mixed model (see, e.g., Searle *et al.* 1992),

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}\,, \tag{1}$$

where $\mathbf{Y}$ is a vector of observations, $\mathbf{X}$ is a design matrix for the fixed effects in the vector β, $\mathbf{Z}$ is the design matrix for the random effects in the vector $\mathbf{u}$, and $\mathbf{e}$ is a vector of random error effects.

After briefly discussing the recovery of interblock information and interrow and intercolumn information in Section 2, we present three possible response models for differential gradients in the incomplete blocks or within a row (or column) of the experiment design in Section 3. For experiments in which the treatments are laid out in a rectangular array within a complete block (replicate), an alternate analysis is given to take into account row and column gradients and their interactions. A numerical example is used to demonstrate the effect of these analyses in Section 4. This is followed by a discussion of post-blocking in an experiment and by a discussion on various aspects of design and analysis. In Section 6, augmented designs and random variety effects are discussed. In Sections 7, 8, and 10, it is shown how to recover intervariety information as well as interblock information or interrow and intercolumn information for three classes of augmented designs. A numerical example of the design in Section 8 is used to illustrate the computational procedure. Some comments are presented in the last section.

2. Standard Statistical Analyses with Recovery of Interblock Information

The textbook response model for a resolvable row-column (lattice square or lattice rectangle) designed experiment for v = bk treatments in k rows and b columns within r complete blocks (replicates) is for g = 1,..., r, h = 1,..., k, i = 1,..., b. and j = 1,..., v:

$$Y_{ghij} = \mu + \beta_g + \rho_{gh} + \gamma_{gi} + \tau_j + \epsilon_{ghij}\,; \tag{2}$$

where μ is a general mean effect, β_g is the gth replicate effect, ρ_{gh} is the hth row effect in replicate g, γ_{gi} is the ith column effect in replicate g, τ_j is the jth treatment effect, ϵ_{ghij} is a random error effect distributed with mean zero and variance σ_ϵ^2. ρ_{gh} and γ_{gh} are random effects distributed with zero means and variances σ_ρ^2 and σ_γ^2, respectively. For v=bk treatments in b incomplete blocks of size k, the standard response model is obtained by dropping the subscript h and the term ρ_{gh} from (2).

The adjusted treatment effects recovering row and column information are

$$\hat{\tau}^* = \left[r\mathbf{I}_v - (\mathbf{NR}' \;\; \mathbf{NC}') \begin{bmatrix} b^*\mathbf{I}_{rk} & \mathbf{RC}-\mathbf{J} \\ \mathbf{RC}' & k^*\mathbf{I}_{rb} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{NR} \\ \mathbf{NC} \end{bmatrix} + \mathbf{J}/k \right]^{-1}$$

$$\times \left[\mathbf{YT} - (\mathbf{NR}' \;\; \mathbf{NC}') \begin{bmatrix} b^*\mathbf{I}_{rk} & \mathbf{RC}-\mathbf{J} \\ \mathbf{RC}' & k^*\mathbf{I}_{rb} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{YR} \\ \mathbf{YC} \end{bmatrix} \right], \tag{3}$$

where $b^* = b + \hat{\sigma}_\epsilon^2 / \hat{\sigma}_\rho^2$ and $k^* = k + \hat{\sigma}_\epsilon^2 / \hat{\sigma}_\gamma^2$, **YR** = row totals minus b times the replicate mean, **YC** = column totals minus k times the replicate mean, **YT** = treatment totals minus r times the overall mean, $\mathbf{I}_x$ is an identity matrix of side x, **RC** is the row-column incidence matrix, **NR** is the row-treatment incidence matrix, **NC** is the column-treatment incidence matrix, and **J** is a matrix of ones to add the restriction that the sum of the effects is zero. These effects plus $\hat{\mu}$ are the adjusted means presented in textbooks when ANOVA solutions for the variance components are used. The variance-covariance matrix for treatment effects is obtained as $\hat{\sigma}_\epsilon^2$ times the first factor on the right side of (3). An approximate average variance of a difference, which is less than or equal to the correct one, may be obtained as $2\hat{\sigma}_\epsilon^2$ times the $1/v^{\text{th}}$ root of the determinant of the first factor in (3). The approximation is useful for large v and unequal standard errors. If the treatment (eliminating row and column effects) sum of squares is desired, obtain intrablock solutions $\hat{\tau}$ from (3) by using b for b^* and k for k^*. Then the sum of squares is $\hat{\tau}'$ times the last term in (3) with b and k replacing b^* and k^*.

When missing values occur, the diagonal matrix of $k\boldsymbol{I}_{rb}$ is replaced by a diagonal matrix, say $\boldsymbol{K}$, with number of elements in a column $k_{gi} \leq k$ on the diagonal; the diagonal matrix $b\mathbf{I}_{rk}$ is replaced by **B**, say, with the number of elements $b_{gh} \leq b$ on the diagonal. Also, $r\boldsymbol{I}_v$ is replaced by a diagonal matrix, say $\boldsymbol{R}$, which has replicate numbers $r_j \leq r$ for treatment j on the diagonal. Otherwise, the analysis proceeds as described above with the degrees of freedom appropriate for the number of observations in the experiment.

3. Statistical Analyses with Recovery of Intergradient and Interblock Information

The following statistical analysis applies equally well to incomplete block designs as to row-column designs. The linear model used in place of model (2) is either

$$Y_{ghi} = \mu + \beta_g + \rho_{gh} + \pi_{gh}\,a_{ghi} + \tau_i + \epsilon_{ghi}\,, \tag{4}$$

$$Y_{ghi} = \mu + \beta_g + \rho_{gh} + \pi_{gh}\,a_{ghi} + \pi_g a_{ghi} + \tau_i + \epsilon_{ghi}\,, \tag{5}$$

or

$$Y_{ghi} = \mu + \beta_g + \rho_{gh} + \pi_{gh}\, a_{ghi} + \bar{\pi} a_{ghi} + \tau_i + \epsilon_{ghi} \;, \tag{6}$$

where the a_{ghi} are the centered linear regression values of position within block (or row) gh (e.g., for k = 3, the values are –1, 0, and 1 and for k = 4, the a_{ghi} values are –3, –1, 1, and 3 for any block or row), π_{gh} is the linear regression coefficient for block gh and is a random effect distributed with mean zero (4), mean π_g (5), or mean $\bar{\pi}$ (6), and variance σ_π^2, and the other effects are defined as for (2). Note that the orthogonal polynomial values b_{ghi} for quadratic (or higher) regressions δ_{gh} could be added to equations (4) – (6) if the situation warranted differential curvilinear regressions within blocks. Other regression forms are possible in this framework.

The resulting normal equations for equation (4) for $Y_{ghi} - \bar{y}_g = Y_{ghi} - \hat{\mu} - \hat{\beta}_g$ values are

$$\begin{bmatrix} k\mathbf{I}_{rb} & \mathbf{0}_{rb\times rb} & \mathbf{NB}_{rb\times v} \\ \mathbf{0} & \Sigma\, a_{ghi}^2\, \mathbf{I}_{rb} & \mathbf{NG}_{rb\times v} \\ \mathbf{NB}' & \mathbf{NG}' & r\mathbf{I}_v \end{bmatrix} \begin{bmatrix} \rho_{rb\times 1} \\ \pi_{rb\times 1} \\ \tau_{v\times 1} \end{bmatrix} = \begin{bmatrix} \mathbf{YB}_{rb\times 1} \\ \mathbf{YG}_{rb\times 1} \\ \mathbf{YT}_{v\times 1} \end{bmatrix} , \tag{7}$$

where $\mathbf{0}$ is a matrix of zeros since the sum of the a_{ghi} in each block gh is zero, ***NB*** is the block-by-treatment incidence matrix, ***NG*** is a matrix of a_{ghi} values for treatment i in block gh, ***YB*** is a vector of block totals for $Y_{ghi} - \bar{y}_{g..}$ values, ***YG*** is a vector of sums of products of a_{ghi} and Y_{ghi} values for each block gh, and the other terms are as defined previously. The restrictions that the sums of the block and treatment effects are zero is added using a **J** matrix (matrix of ones). Intrarow and intragradient (fixed effect) solutions result. Replacing k by k + $\sigma_\epsilon^2/\sigma_\rho^2$ and $\sum a_{ghi}^2$ = C by C + $\sigma_\epsilon^2/\sigma_\pi^2$ results in treatment effects adjusted for interrow and intergradient information.

The expected value (ANOVA) of the error mean square is taken to be σ_ϵ^2. The expected value of the gradient (eliminating block and treatment effects) mean square has the form

$$\sigma_\epsilon^2 + g_0\sigma_\pi^2 = \sigma_\epsilon^2 + \sigma_\pi^2 C(rk - k - 1)/rb \;, \tag{8}$$

as found using *Mathematica*. The expected value of the block (eliminating gradient and treatment effects) mean square has the form $\sigma_\epsilon^2 + k_0\sigma_\rho^2$, where k_0 using *Mathematica* or SAS Proc GLM is 2.8149 for the example. For v = 16, k = 4, and r = 5 in a balanced lattice square design the expected value of the columns (or rows) mean square after eliminating rows (or columns) and treatment effects is $\sigma_\epsilon^2 + 3\sigma_\gamma^2$. The coefficient of 3 is larger than $k_0 = 2.8149$. For g_0, the value is 15 for this example as compared to $C = 20$. When r becomes large, k_0

approaches k and g_0 approaches Σa^2_{ghi}. Solutions for the various variance components allow for treatment effects recovering interblock (interrow) and intergradient information as follows:

$$\hat{\tau}^* = \left[r\mathbf{I}_v - \mathbf{NB}'\ \mathbf{NB}\Big/\left(k + \hat{\sigma}^2_\epsilon/\hat{\sigma}^2_\rho\right) - \mathbf{NG}'\ \mathbf{NG}\Big/\left(C + \hat{\sigma}^2_\epsilon/\hat{\sigma}^2_\pi\right) + \mathbf{J}\big/k\right]^{-1}$$

$$\times\left[\mathbf{YT} - \mathbf{NB}'\ \mathbf{YB}\Big/\left(k + \hat{\sigma}^2_\epsilon/\hat{\sigma}^2_\rho\right) - \mathbf{NG}'\ \mathbf{YG}\Big/\left(C + \hat{\sigma}^2_\epsilon/\hat{\sigma}^2_\pi\right)\right]. \tag{9}$$

The variance-covariance matrix for $\hat{\tau}^*$ is $\hat{\sigma}^2_\epsilon$ times the first term on the right hand side of (9).

A row-column classification does not account for gradients which are in directions different from the rows or columns. An alternate analysis to that presented above is to fit polynomial terms to the rows and to the columns in each complete block of an experiment designed as a lattice rectangle (or an incomplete block in a rectangular layout) design; then, include differential row-column interaction terms such as linear $\times$ linear, linear $\times$ quadratic, quadratic $\times$ linear, and/or quadratic $\times$ quadratic for each complete block to obtain an ANOVA and adjusted treatment effects. The number of regression terms included determines the number of degrees of freedom in each complete block. For random regression effects within complete blocks, interregression information may be recovered in the same manner as for interblock information. The expected value of regression mean squares may be evaluated using some such program as MATHEMATICA or MAPLE. The computational procedures described in this Section may be programmed in GAUSS, SAS, or other packages (Federer and Wolfinger, 1996, 1997; Barnard and Federer, 1997).

4. A Numerical Example

The numerical example used to illustrate the statistical analysis with recovery of interblock and intergradient information and interregression (intertrend) information is the one presented by Wadley (1946) and given in Table 12.5 of Cochran and Cox (1957) for an experiment designed as a balanced lattice square. We consider differential gradients within rows as an alternative analysis for columns. Since a rather high coefficient of variation, 44%, was obtained from the analysis on count for the standard lattice square analysis (Cochran and Cox, 1957; Federer, 1955; Kempthorne, 1952), some alternative analysis and/or transformation of data may be required. In addition, the intrarow-column error mean square was slightly larger than the treatment (eliminating row and column effects) mean square. This appears illogical since it is unlikely that the null hypothesis would be true for 16 different chemical spray treatments involving a check treatment. The lattice square analyses for means of three counts, square roots of count, and the arcsine transformation of count were computed with ratios of treatment to residual mean squares being $\frac{21.30}{22.67}$, $\frac{0.4716}{0.5759}$, and $\frac{17.49}{21.19}$, respectively, all ratios being less than one.

One possible alternative model is that there are differential linear trends in each row (or column) of the lattice square designed experiment. The polynomial linear regression coefficients are –3, –1, 1, and 3 with a sum of squares of $C = 20$. The regressions are within rows. A fixed effects analysis results in a residual mean square for count of 18.97, a 16% reduction over the standard lattice square analysis. The Type III mean squares ratios of treatment to residual are 23.15/18.97 and 0.5453/0.5445 for count and square root of count, respectively. Although the F-ratios are greater than one, the residual mean square appears large in light of distribution theory. If counts have a Poisson distribution, the estimated theoretical variance is 10.905/3 or approximately 4 since the experiment mean is 10.905. The theoretical variance of square roots is 1/4. The obtained residual mean squares are much larger, indicating extraneous variation. Even if the differential gradient analysis is more appropriate than the standard lattice square analysis, it appears another form of spatial analysis is required

Response model (4) accounted for more of the experimental variation than did the standard lattice square analysis but there still appears to be considerable variation remaining. Therefore, we shall use the regression (trend) analysis described in Section 3. In order to have the same number of degrees of freedom for controlling within complete block variation, i.e., 15 + 15 = 30, row linear (RL) and row quadratic (RQ), column linear (CL), and interactions LL = RL $\times$ CL, LQ = RL $\times$ CQ and QQ = RQ $\times$ CQ were included. Interaction QL = RQ $\times$ CL and column quadratic (CQ) mean squares were less than the residual mean square, and were not included as they would be relegated to the residual mean square using the Bozovich *et al.* (1956) procedure. As may be noted, there is a dramatic reduction in the error mean square from 22.67 to 11.91, or a reduction of 1 – 11.91/22.67 = 47.5%. Also, treatments now show a significant difference, F = 28.97/11.91 = 2.43, at the 2% level. For square root of count, F = 0.7087/0.3579 = 1.98. Note that significant differences among the treatments are also indicated if a RCBD ANOVA is used, i.e., F = 82.95/38.88 = 2.13 ($F_{.025} = 2.06$). A residual mean square of 11.91 is closer to the theoretical variance than before. Using square root of count produced an error mean square of 0.36 (REML solution = 0.33) which is much closer to the theoretical variance of 0.25. It appears that extra-Poisson variation is present.

Both types of analyses are more reasonable than the standard ANOVA for a balanced lattice square. Regression analysis controls experimental variation much better than differential gradient analysis, i.e., a residual mean square of 11.91 vs. 18.97. Evidently the trends were not in the directions of the row-column orientation. When CQ (column quadratic) and QL (row quadratic by column linear) terms were included in the analysis, the residual mean square was 12.85, still considerably less than 22.67 from the lattice square design or 18.97 for the differential gradient analysis.

5. Discussion

During the course of conducting an experiment, events occur which are not controlled by the original blocking for the experiment. For example, a field experiment on alfalfa may exhibit a patch of yellowing in a part of the experiment which may be caused by excessive rain during the previous year. There are two ways of handling this problem. First and probably best is to obtain a measure of amount of yellowing on each experimental unit and then use this measurement as a covariate. Second, a new block for the yellowed area of the experiment may be designated and this can then be taken care of in the analysis as an additional block. Note that this is equivalent to using covariance with a 0 or 1 independent variate to signify the presence or absence of yellowing. The same procedure may be used to handle other situations such as water standing in part of the experiment and insect, disease, or animal damage to a part of the experiment. In marketing experiments, a part of the experiment may be damaged by fire, water, or wind and the part affected may be handled as described above. The following axiom is useful in determining whether or not to use post-blocking:

> **Axiom:** Any event occurring during the course of an experiment which is not caused by or is a response of the treatments in the experiment is a candidate for removal by post-blocking or covariance.

Likewise, patchy or spotty occurrences can be accounted for in a similar manner. One word of caution here is that some of the events which occur cause a treatment by event interaction which is a treatment effect. One such event is winter heaving (or kill) for a group of cultivars such as alfalfa, winter wheat, winter rye, etc. It may not be possible to estimate these interactions owing to complete confounding of some effects.

Control of within-complete-block heterogeneity is best accomplished with a row-column arrangement within each complete block. Such designs have been denoted as lattice rectangle or resolvable row-column experiment designs. These designs have the desirable properties of the Latin square design. Hence, as a measure of insurance, resolvable row-column designs should be used whenever the experimenter even suspects that there may be removable variation in two directions with possibly differential trends or gradients in one direction. If an incomplete block design has been used and then trends occur within some or all of the incomplete blocks, the procedure of Section 3 will be useful in removing this type of experimental variation. Differential trends in two directions can be handled by appropriate statistical analyses. Recovery of interrow and intercolumn information, of interblock and intergradient information, or of interregression information should always be done when analyzing data. Ignoring this type of information is inefficient use of resources and information.

Three types of spatial analyses are described above. Other forms such as nearest neighbor, smoothing and kriging may be used in the context of the experiment design and blocking structure of the experiment (Federer *et al.*, 1997). The spatial analysis parameters may be considered in the mixed model context and internearest-neighbor, intersmoothing, and interkriging information recovered in the manner described. The appropriate statistical analysis and model are necessary in order to obtain the best solutions for treatment effects.

6. Augmented Designs and Random Variety Effects

A design used for screening genotypes in plant breeding trials is a layout wherein a single check variety is systematically spaced throughout the experimental area at the rate of one check to $n-1$ genotypes. Many procedures for obtaining adjusted genotype means have been suggested since the 1930s. A recent method of adjustment and analysis has been proposed by Cullis *et al.* (1989). The class of augmented experiment designs was constructed as an alternative to the above one. These designs pose several advantages over the systematic check design.

Various augmented experiment designs have been presented in the literature (Federer, 1955a, 1961; Federer and Raghavarao, 1975; Federer *et al.*, 1975; Federer and Wright, 1988). The purpose of the following sections is to present a statistical analysis for these experiment designs making use of the information obtained from the random blocking effects and from the distributional effects of the augmented (or new) treatments in the experiment. Since augmented designs are used to screen a set of new treatments (varieties) for which there is limited information and often material, these treatment or varietal effects are often considered to be distributed around some mean and with a common variance σ_T^2 (see Cullis *et al.*, 1989). Herein we consider that each new treatment is included once in the experiment but this need not be the case as the procedure is easily extendible to take additional replication into account. There are c check or standard treatments which are used to obtain the experiment design prior to adding the augmented treatments. The check treatment yields are used to obtain solutions for blocking and check treatment effects. The replicate and block effects are used to adjust the new treatment effects. From the mean square for the new treatments, an estimate of the variance component σ_T^2 is obtained and used for adjusting new treatment means for their distributional effects. Adjustment for distributional properties of the random effects makes use of all information from an experiment, producing the well-known BLUP-like or empirical best linear unbiased predictors.

In Section 7, augmented block experiment designs are considered. In Section 8, augmented row-column experiment designs such as those described by Federer and Raghavarao (1975) and Federer *et al.* (1975) are the subject of discussion. An example illustrating the design of Section 8 is presented in Section 9. In Section 10, we present statistical analyses for augmented resolvable row-column designs such as those presented by Federer and Wright (1988) and those which could be obtained from lattice rectangle or resolvable row-column designs by a "variety cutting" procedure.

7. Augmented Block Experiment Designs

Among the experiment designs in this class are the augmented randomized complete block designs (ARCBD), augmented balanced incomplete block designs (ABIBD), and augmented partially balanced incomplete block designs (APBIBD). With respect to the augmented or new treatments all these designs are incomplete in that all new treatments do not appear together in the replicates or blocks. Recovery of interreplicate and interblock information is needed for a more efficient analysis. First consider an ARCBD with c checks and n new treatments for a total of $v = c + n$ treatments in r blocks. Let the c check treatments appear once in each of the blocks [note that the c treatments could appear in the proportions $n_1 : n_2 : \cdots : n_c$ in each of the r blocks and the design would still be an orthogonal one (see Federer, 1991, Ch. 7)]. Since the n new treatments each occur once in the experiment the observation can only contribute to the new treatment estimate and nothing to block, overall mean or error estimation (Federer and Raghavarao, 1975). From an analysis on check treatment results only, the expected value of the block mean square is $\sigma_\epsilon^2 + c\sigma_\beta^2$, and an ANOVA solution of σ_β^2 is the difference between the block mean square and the residual mean square divided by c. This variance component is used to obtain adjusted new treatment means recovering interblock information. For our analysis, we use the following linear model:

$$\mathrm{Y}_{ij} = (\mu + \beta_i + \tau_j + \epsilon_{ij})n_{ij}\,, \tag{10}$$

where μ is a general mean effect, β_i is the ith block effect distributed with mean zero and variance σ_β^2, τ_j is the j th treatment effect, $i = 1, \cdots, r, j = 1, \cdots, v$, n_{ij} is one if the jth treatment occurs in ith block and zero otherwise, and ϵ_{ij} is a random error effect distributed with mean zero and variance σ_ϵ^2.

The new treatment sums of squares is

$$\sum_{i=1}^{r}\sum_{j=1}^{n}\hat{\tau}_j(\mathrm{Y}_{ij} - \bar{\mathrm{y}}_i)n_{ij} - \left(\sum_{i=1}^{r}\sum_{j=1}^{n}\hat{\tau}_j n_{ij}\right)^2 \Big/ \mathrm{n}\,, \tag{11}$$

where Y_{ij} is the yield of new treatment j in block i, and the other sums of squares are from standard procedures. The expected value of (11) is $\left(\sigma_\epsilon^2 + \sigma_\tau^2\right)(\mathrm{n}-1)$. Substituting $\mathbf{NB} - \mathbf{J}$ for $\mathbf{NB}$ in the normal equations results in solutions for the τ_j effects. The treatment effects adjusted for interblock and intervarietal information are:

$$\hat{\tau} = \left[\begin{pmatrix} \mathrm{r}\mathbf{I}_c & \emptyset \\ \emptyset' & \mathbf{I}_n(1 + \hat{\sigma}_\epsilon^2/\hat{\sigma}_\tau^2) \end{pmatrix} - \mathbf{NB}'\left(\mathbf{K} + \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_\beta^2}\,\mathbf{I}_r\right)^{-1}(\mathbf{NB} - \mathbf{J}\emptyset)\right]^{-1}$$

$$\times \left[\begin{pmatrix} \mathbf{YC} \\ \mathbf{YN} \end{pmatrix} - \mathbf{NB}'\left(\mathbf{K} + \frac{\hat{\sigma}_\epsilon^2}{\hat{\sigma}_\beta^2}\,\mathbf{I}_r\right)^{-1}\mathbf{YB}\right], \tag{12}$$

where $\mathbf{K}$ is a diagonal matrix, $\emptyset$ is a matrix of zeros, $\mathbf{NB}$ is the block-treatment incidence matrix, and $\mathbf{YB}$ and $\mathbf{YT}$ are block and treatment totals corrected for the replicate means. The checks have ones in every block and the restriction used to solve for treatment effects is that the sum of the check effects is zero. Variances of differences among treatment effects are obtained from the variance-covariance matrix, $\hat{\sigma}_\epsilon^2$ times the first term on the right hand side of (12).

The statistical analysis outlined above is directly applicable to an analysis for ABIBD, APBIBD, and other augmented incomplete block designs. The $\mathbf{K}$ and $\mathbf{NB}$ matrices will need to be adjusted to take into account the experiment design for the check treatments. When the check treatments are in an incomplete block experiment design, their effects will need to be adjusted for recovery of interblock information. The remainder of the analysis proceeds as described above.

8. Augmented Row-Column Experiment Designs

A number of augmented row-column designs (ARRCD) have been presented by Federer and Raghavarao (1975) and Federer *et al.* (1975). One particular row-column design for few checks and many new treatments is to have the checks repeated more than once in each row and/or column. José Crossa, CIMMYT, personal correspondence, used a row-column design of 12 columns and 15 rows with two checks which appear either two or three times in each column and twice in each row with 60 check plots and 120 new treatments for an experiment conducted by Mathew Reynolds, CIMMYT (see Table 1). The design was used over several sites. The linear model used here is

$$Y_{ghi} = (\mu + \rho_g + \gamma_h + \tau_i + \epsilon_{ghi})n_{ghi} \,, \tag{13}$$

where μ is a general mean effect, ρ_g is the effect of the gth row distributed with mean zero and variance σ^2_ρ, γ is the effect of the hth column distributed with mean zero and variance σ^2_γ, τ_i is the effect of the ith treatment where the check treatments are fixed effects and the new treatments are random effects distributed with mean τ and variance σ^2_τ, and n_{ghi} is one if treatment i occurs in row g and column h and zero otherwise.

The intrarow-column solutions for new treatments are $\hat{\tau}_i = Y_{ghi} - \hat{\mu} - \hat{\rho}_g - \hat{\gamma}_h$ where $\hat{\mu}$, $\hat{\rho}_g$, and $\hat{\gamma}_h$ are the mean, intrarow, and intracolumn solutions. The sum of squares among new treatments is $\sum_{i=1}^{n}\hat{\tau}_i^2$ - $(\sum_{i=1}^{n}\hat{\tau}_i)^2/n$. The expected values for Type III row and column mean squares may be evaluated using various computer software. Then ANOVA or REML solutions for the variance components σ^2_ρ, σ^2_γ, and σ^2_ϵ may be obtained. These solutions and $\hat{\sigma}^2_\tau$ are then substituted in the normal equations and the matrices are expanded to include the new treatments. With these changes the new treatment effects with recovery of interrow, intercolumn, and intervariety information are obtained. The variances of difference may be obtained as described in Section 7. The following numerical example illustrates the procedure for row-column augmented designs.

9. ARCD Example

An augmented row-column design involving two check varieties each appearing twice in each of the 15 rows and two or three times in each of the 12 columns for a total of 30 replicates for each check, was conducted by Dr. Mathew Reynolds, CIMMYT. Of the 180 experimental units, 60 were allocated to checks and 120 to the 120 new genotypes. The field layout and grain weights for the 122 cultivars are presented in Table 1. Owing to the fact that the SAS Proc GLM constraint sets the highest numbered effect equal to zero, it is desirable to give checks the highest numbers, 121 and 122 here. Since the design is not row-column connected, we use polynomial regression values of row and column positions and their interactions. For this particular analysis, row polynomials up to 12th degree and column polynomials up to 10th degree were computed using the computer program given by Wolfinger *et al.* (1997). Following the Bozovich *et al.* (1956) procedure, the polynomials R1, R2, R4, R8, R10, C1, C2, C3, C4, C6, and C8 were retained, where Ri and Ci are ith degree polynomial regressions. Then from the 16 interaction terms Ri*Ci, i = 1,..., 4, R1*C1, R1*C2, and R1*C3 were added to the regression model. These 14 regressions appeared to explain satisfactorily the spatial variation present. The 120 new cultivars and the regressions are considered to be random effects. A SAS Proc Mixed program for recovering interregression and intervariety information is given by Wolfinger *et al.* (1997). Since the number of new is often large, arranging the new effects adjusted for interrow, intercolumn, and intervariety information, from largest to smallest, is desirable for the experimenter and is a feature included in the code. Selected material from SAS Proc GLM and Proc Mixed outputs is given in Table 2. Only the top 15 REML means and the associated intraregression least squares means are presented because of space. The ranks of the fixed effect means are given to demonstrate the change in ranks from a random effects model. With respect to the fixed effect means, 52 new genotypes were above the 121 check mean of 907 and 96 above the 122 check mean of 827. 36 of the REML means were larger than the 121 check mean of 910 and 114 were above the 122 check mean of 827.

Insert Tables 1 and 2

10. Augmented Resolvable Row-Column Experiment Designs

Lattice square experiment designs are resolvable row-column experiment designs and with "variety-cutting" may be used to construct augmented resolvable row-column designs, ARRCD (Federer and Wright, 1988). For $v = k^2$ there are $k - 1$ suitable arrangements from a balanced lattice square design to construct an

ARRCD. Nam-Ky Nguyen (1994, personal communication) has prepared software for constructing lattice rectangle designs for $v = kb$. These designs may be used to construct ARRCDs by "variety cutting" in the manner described by Federer and Wright (1988). An ANOVA for an ARRCD is obtained using the following linear model

$$Y_{ghij} = (\mu + \beta_g + \rho_{gh} + \gamma_{gi} + \tau_j + \epsilon_{ghij})\eta_{ghij} \,, \tag{14}$$

where μ is a general mean effect, β_g is a gth replicate effect and is distributed with mean zero and variance σ^2_β, ρ_{gh} is the kth row effect in replicate g and is distributed with mean zero and variance σ^2_ρ, γ_{gh} is the ith column effect in replicate g and is distributed with mean zero and variance σ^2_γ, τ_j is the jth treatment effect, ϵ_{ghij} is a random error effect distributed with mean zero and variance σ^2_ϵ, and η_{ghij} is an indicator variable which is one when treatment j occurs in row h and column i in replicate g and zero otherwise.

The various sums of squares are computed in the manner described previously, taking into account the nature of the various matrices involved. The intrarow-column solution for a new treatment effect is

$$\hat{\tau}_j = Y_{ghij} - \hat{\mu} - \hat{\beta}_g - \hat{\rho}_{gh} - \hat{\gamma}_{gi} = Y_{ghij} - y_g - \hat{\rho}_{gh} - \hat{\gamma}_{gi} \,, \tag{15}$$

when $Y_{....}$ the grand total for the checks, contains an equal number of each of the effects ρ_{gh}, γ_{gi}, and τ_j. That is, when the restrictions on the solutions are that

$$\sum_{h=1}^{a}\hat{\rho}_{gh} = \sum_{i=1}^{b}\hat{\gamma}_{gi} = \sum_{j=1}^{c}\hat{\tau}_g = 0 \,, \tag{16}$$

these effects disappear in the total $Y_{....}$ from the check yields. A sum of squares for new treatments is $\sum_{j=1}^{n}\hat{\tau}_j^2 - \left(\sum_{j=1}^{n}\hat{\tau}\right)^2 / \mathrm{n}$. The expected value is $(n-1)(\sigma^2_\epsilon + \sigma^2_\tau)$. With estimates of σ^2_ϵ, σ^2_ρ, σ^2_γ, σ^2_β, and σ^2_τ, new treatment effects recovering interreplicate, interrow, intercolumn, and intervariety information are obtained as described previously. Likewise, variances for differences of these adjusted effects may be obtained as explained before.

11. Comments

Computation of sums of squares and adjusted effects for the proposed analyses pose little difficulty owing to the availability of computer software packages and codes (Barnard and Federer, 1997; Federer and Wolfinger, 1996; Wolfinger *et al.*, 1997). Standard statistical packages such as GAUSS, SAS, and GENSTAT can be used to recover interblock and intervariety information. Some packages such as MAPLE, MATHEMATICA, and SAS may be used to obtain expected values for mean squares. Then, solutions for treatment effects recovering interreplicate, interblock, interrow, intercolumn, and intervariety information are possible. These computations pose little difficulty using GAUSS and concatenating submatrices or multiples thereof. Also, SAS Proc Mixed produces REML solutions for these variance components and would be the preferred procedure according to Searle *et al.* (1992). Using SAS Proc Mixed, it is irrelevant whether or not the design is connected. For the example in Section 9, the program runs even when row, column, and row by column interaction are listed as random. The estimated standard errors are much larger than for connected analyses owing to the excessive over-parameterization.

Acknowledgment

The constructive and clarifying comments and suggestions of an Associate Editor and referees were helpful in arriving at this version of the paper. Their efforts and patience are greatly appreciated.

References

Barnard, J. and W. T. Federer (1997). GENSTAT and SAS programs for recovering interblock, interrow-column, and intergradient information. BU-1383-M, Technical Report Series, Biometrics Unit, Cornell University, Ithaca, NY 14853 (cost $5).

Bozovich, H., T. A. Bancroft, and H. O. Hartley (1956). Power of analysis of variance test procedures for certain incompletely specified models. I. Annals Mathematical Statistics **27**:1017-1043.

Cochran, W. G. and G. M. Cox (1957). *Experimental Designs*, 2nd ed. John Wiley & Sons, Inc., New York.

Cox, C. P. (1958). The analysis of Latin square designs with individual curvatures in one direction. *Journal of the Royal Statistical Society, Series B* **20**, 193-204.

Cullis, B. R., J. L. Warwick, J. A. Fisher, and B. J. Read (1989). A new procedure for the analysis of early generation variety trials. *Applied Statistics* **38**, 361-375.

Federer, W. T. (1955). *Experimental Design-Theory and Application.* The Macmillan Company, New York (republished by Oxford & IBH Publishing Company, Calcutta, 1967 and 1974).

Federer, W. T. (1955a). Augmented (or hoonuiaku) designs. *Hawaiian Planter's Record* **55**, 191-208.

Federer, W. T. (1961). Augmented designs with one-way elimination of heterogeneity. *Biometrics* **17**, 447-473.

Federer, W. T. (1991). *Statistics and Society – Data Collection and Interpretation.* Marcel Dekker, Inc., New York. Chapter 7.

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

Federer, W. T., Nair, R. C. and Raghavarao, D. (1975). Some augmented row-column designs. *Biometrics* **31**, 361-374.

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

Federer, W. T. and R. D. Wolfinger (1996). GAUSS and SAS for recovery of interblock and intervariety information. BU-1384-M, Technical Report Series, Biometrics Unit, Cornell University, Ithaca, NY 14853 (cost $5).

Federer, W. T. and J. Wright (1988). Construction of lattice square designs. *Biometrical Journal* **30**, 77-85.

Kempthorne, O. (1952). *The Design and Analysis of Experiments.* John Wiley & Sons, Inc., New York.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components.* John Wiley and Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore.

Wadley, F. M. (1946). Incomplete block designs in insect population problems. *Journal of Economic Entomology* **38**, 651-654.

Williams, E. R. (1986). A neighbour model for field experiments. *Biometrika* **73**, 279-287.

Wolfinger, R. D., Federer, W. T., and Cordero-Brana, O. (1997). SAS PROC GLM and PROC MIXED codes for recovering information in augmented designs. **Agronomy J.** (accepted).

Yates, F. (1939). The recovery of interblock information in variety trials arranged in three dimensional lattices. *Annals of Eugenics* **9**, 136-156.

Yates, F. (1940a). The recovery of interblock information in balanced incomplete block designs. *Annals of Eugenics* **10**, 317-325.

Yates, F. (1940b). Lattice squares. *Journal of Agricultural Science* **30**, 672-687.

Table 1.
Layout and grain weights for a augmented row-column design
of two checks (121 and 122) and 120 new genotypes.

Column

Row	1	2	3	4	5	6	7	8	9	10	11	12
1	121	32	42	122	62	72	121	92	102	122	2	12
	1040	902	839	966	796	936	878	894	819	744	907	923
2	23	121	43	53	122	73	83	121	103	113	122	13
	843	916	731	942	862	815	894	942	782	800	683	907
3	24	34	121	54	63	122	84	93	121	114	3	122
	916	769	882	816	807	808	894	827	776	861	783	718
4	122	33	44	121	64	74	122	94	104	121	4	14
	880	922	718	936	826	840	888	788	767	902	786	870
5	25	122	45	55	121	75	85	122	105	115	121	15
	893	845	1011	788	864	894	929	708	793	837	716	857
6	26	35	122	56	65	121	86	95	122	116	5	121
	845	963	832	788	812	967	958	914	721	888	712	896
7	121	36	46	122	66	76	121	96	106	122	6	16
	938	922	1046	796	810	963	1006	854	878	848	688	832
8	27	121	47	57	122	77	87	121	107	117	122	17
	851	824	924	948	807	922	941	850	780	919	630	782
9	28	37	121	58	67	122	88	97	121	118	7	122
	741	882	995	1036	967	973	942	858	810	1043	814	866
10	122	38	48	121	68	78	122	98	108	121	8	18
	692	917	1017	981	971	931	971	937	977	985	852	999
11	29	122	49	59	121	79	89	122	109	119	121	19
	755	650	975	976	940	911	983	840	968	959	829	895
12	30	39	122	60	69	121	90	99	122	120	9	121
	791	892	912	1133	931	1031	1063	1060	855	1050	898	975
13	121	40	50	122	70	80	121	100	110	122	19	20
	738	727	666	926	837	881	798	869	898	853	821	894
14	21	122	41	51	121	71	81	122	101	111	121	11
	943	775	929	793	1104	919	786	816	1023	1090	878	1109
15	22	31	122	52	61	121	82	91	122	112	1	121
	742	767	937	764	1084	1096	1121	1057	927	916	833	1013

Table 2.
Type III mean squares, REML solutions for means
of the top 15 new genotypes with fixed effect means.

Source of variation	Degrees of freedom	Type III (GLM) mean square	F-ratio
R1	1	28,678	8.31
R2	1	12,832	3.72
R4	1	4,993	1.45
R8	1	20,170	5.85
R10	1	15,069	4.37
C1	1	12,953	3.76
C2	1	48,712	14.12
C3	1	42,867	12.43
C4	1	22,613	6.56
C6	1	31,220	9.05
C8	1	77,300	22.41
R1*C1	1	52,885	15.33
R1*C2	1	24,977	7.24
R1*C3	1	7,998	2.32
Entry	121	8,411	2.44
Residual	48	3,449	

Cultivar	REML mean	Fixed effect mean	rank
60	974	1062	2
21	951	1038	8
11	947	968	25
99	944	1049	4
2	942	1087	1
35	937	1056	3
118	937	1038	7
58	937	1001	12
111	934	975	20
46	932	966	27
120	932	998	14
61	931	957	36
38	927	1041	5
82	927	959	35
90	926	962	30
121	910	907	53
122	826	827	97