J. R. Statist. Soc. B (1983),
45, *No.* 2, *pp.* 151–211

Nearest Neighbour (NN) Analysis of Field Experiments

By G. N. WILKINSON,† S. R. ECKERT, T. W. HANCOCK and O. MAYO

Waite Agricultural Research Institute, University of Adelaide, Adelaide, Australia

[*Read before the* Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, 8th December, 1982, Professor* R. N. Curnow *in the Chair*]

SUMMARY

The paper is in two parts. Part I presents results of a Monte Carlo randomization study of Papadakis's covariance-method of NN analysis which show that (i) a non-iterated Papadakis analysis tends to be conservatively biassed; (ii) iteration of the analysis as suggested by Bartlett (1978) leads to substantial positive bias in the treatment F ratio; (iii) the method is very inefficient when there are substantial trend effects in the data. A theoretical explanation of these results is given.

Part II describes a new method of NN analysis discovered by the first author and developed in collaboration with the co-authors. The method is essentially a "moving-block" analogue of classical forms of analysis for "fixed" blocks (or rows, columns). It avoids the defects of Papadakis's method and leads to approximately unbiassed analyses. It is nearly always and often substantially more efficient on average than classical analyses of complete or incomplete block experiments, and also more efficient than standard analyses of Latin or lattice square designs if there are appreciable row × column interactions in the data. New criteria of design for NN balance are described. Validity of the new method under randomization is demonstrated empirically with Monte Carlo studies.

Keywords: ANALYSIS OF FIELD EXPERIMENTS; NEAREST NEIGHBOUR METHODS; NN ANALYSIS; INTRA-N ANALYSIS; PAPADAKIS'S METHOD; MONTE CARLO; RANDOMIZATION DISTRIBUTIONS; BIAS IN F RATIO; EFFICIENCY; SPATIAL MODELS; NN-BALANCED DESIGNS

0. INTRODUCTION

The idea of adjusting yields of field experiments for local trend effects by a covariance analysis with respect to treatment-corrected yields of adjacent plots was suggested 45 years ago by Papadakis (1937) and subsequently discussed by Bartlett (1938). Regrettably this important idea then suffered a long period of neglect until interest in it was revived by Atkinson (1969), see also Yates (1970, p. 148). Pearce and Moore (1976) drew attention to the substantial gains in accuracy of treatment estimation that were possible from practical application of Papadakis's method, and Bartlett (1978), in a paper read to the Royal Statistical Society, re-examined its theoretical properties with reference to nearest-neighbour models of a symmetrical autoregressive form. Bartlett also suggested iteration of the method, using treatment estimates from the previous iteration to redefine the nearest-neighbour covariate for the current. Pearce (1980) and Kempton and Howes (1981) present additional empirical evidence on the practical value of the method.

† The first author is currently the Reserve Bank of Australia Senior Research Fellow in Agriculture (1981–82), and was previously an Australian Research Grants Research Fellow at the University of Adelaide. Grateful acknowledgement is made to both funding bodies. The co-authors thank the Wheat Industry Research Council for grants for computing equipment.

Present address: Dr G. N. Wilkinson, Biometry Section, Waite Agricultural Institute, Glen Osmond, South Australia 5064.

0035–9246/83/45151 $2.00

Much of the theoretical discussion of Papadakis's method has centred on stationary models. We believe this is a mistake in the context of field experiments, where non-stationary trend effects are typically a dominant consideration. Indeed the primary role of nearest-neighbour adjustment in what we term nearest-neighbour (or NN) analysis of field experiments can be described as that of providing a continuous form of local detrending, in contrast to the stepwise block-detrending of classical methods of analysis. The present paper reconsiders the Papadakis method from this point of view, and also describes a more efficient alternative method.

The paper is organized in two parts corresponding to these methods. Part I describes first the results of extensive Monte Carlo randomization studies of the Papadakis method with uniformity data. The results show that while a non-iterated Papadakis analysis is reasonably valid under randomization (with a slight tendency towards conservative bias) iteration of the analysis as suggested by Bartlett (1978) leads to serious upward bias in the treatment F-ratio. The most serious defect of the method, however, was found to be its inherent inefficiency when trend effects are appreciable. A theoretical explanation of the inefficiency is derived in Section I.3 on the basis of a *smooth trend + independent error* model. A simple explanation of the bias from iteration is also given.

The source of inefficiency in the Papadakis method was found to be the prior correction of yields for treatment effects when forming the nearest-neighbour covariate. This led the first author to discover an alternative, more efficient method of NN analysis which can be properly described as the "moving-block" analogue of the classical "fixed-block" analysis methods developed at the Rothamsted Experimental Station. The new method will be described in Part II, and we defer further discussion of it to the introductory section II.1.

Part I. Papadakis's Method

I.1. MONTE CARLO RANDOMIZATION STUDY

Data sets

Three sets of uniformity data were used in the study, from Wiebe (1935), Mercer and Hall (1911) and Kempton and Howes (1981). They are described below.

The Wiebe data comprise 1500 wheat yields from 1500 plant rows 15 ft long and 1 ft apart in a 125 row x 12 column field arrangement. We split this into two sets designated W1 and W2, comprising columns 1–6 and 7–12 respectively. W1 is more variable than W2 and includes a patch with very high yield. As the experiment was sown with an 8-row grain drill which produced a recurring pattern of row effects on yield, the data were further reduced by summation to 8-row plot totals. Elsewhere in the paper we also consider 4-row totals, and distinguish the two cases with subscripts 8 or 4. A contour diagram based on the 8-row totals in shown in Fig. 1.

The Mercer and Hall data, designated MH, are wheat yields from a 20 x 25 array of plots 10.8 ft long and 11 rows wide, with 9 inch row spacing (imputed from the specified plot area, 0.002 acres). We condensed these data by summation to a 20 x 6 array of 4-column totals (omitting column 25) to obtain yields for relatively long plots, for which 1-D NN adjustment would be appropriate.

The third data set (KH) from Kempton and Howes (1981, p. 64) comprises barley yields from a 28 x 7 array of 5 ft x 14 ft plots.

Experimental design and analysis

The study involved a (3 x 2 x 2 x 2) factorial set of 24 runs and 100 randomized designs for 2 or 3 replicates of 30 treatments in 5 x 6 arrays superimposed on the data sets W1, W2 and MH, plus an additional 2 x 2 factorial set of 4 runs on KH with 4 replicates of 49 treatments in 7 x 7 arrays. A 10-cycle iterated Papadakis analysis was then done in each case, making a total of 28 000 individual analyses. The additional factors varied in the study were

Randomization: Unrestricted, restricted;
Method of analysis: 1-D, 2-D

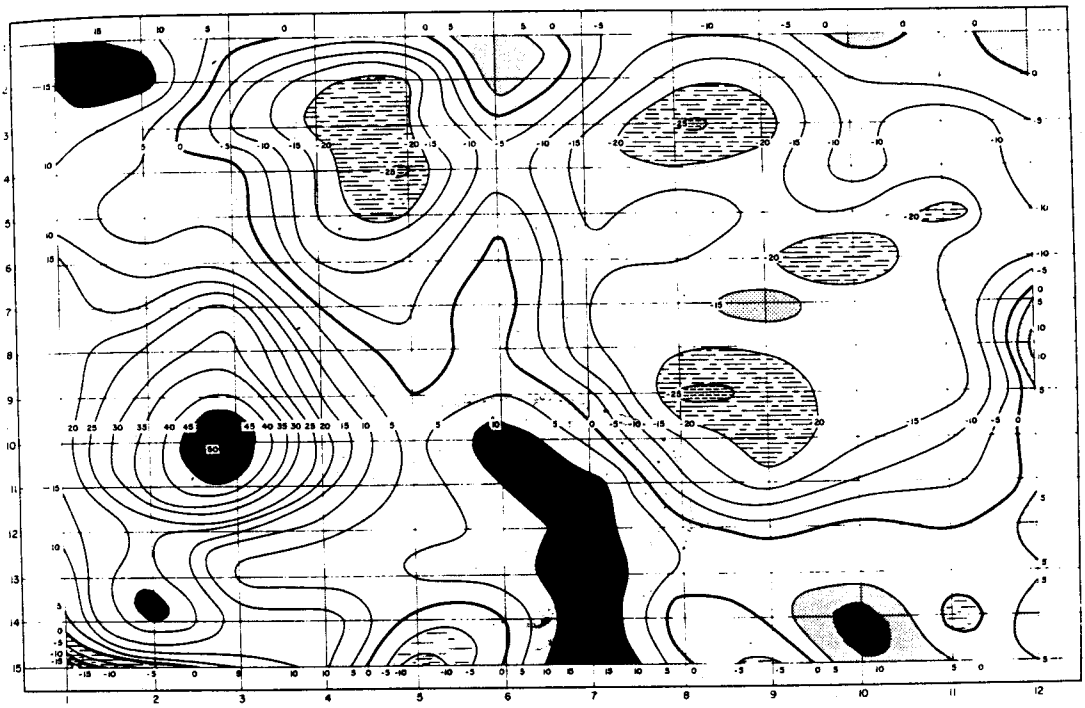


Fig. 1. Contours of 8-row grain yields of the wheat uniformity data of Wiebe (1935), at 5 per cent intervals of deviation from mean yield. Grid lines indicate centre lines of the 8-row plots (8 ft × 15 ft). Contours terminate at the centre lines of boundary plots.

Unrestricted randomization was that for a randomized block design with the specified numbers of replicates and treatments. Otherwise randomization was restricted to give second-level partial NN balance (see Section II.6), except for KH where the restricted randomization was that for a balanced lattice square design, for comparison with a run of 40 randomizations on KH reported in Kempton and Howes (1981). The Papadakis analyses were performed with either a single NN covariate for column neighbours (the 1-D case) or else two covariates for row and column neighbours respectively (2-D). Covariate values for boundary plots were defined using just the neighbours available, as border plots were not included in the designs.

Our original study, completed in December 1980, included only 6 runs on W1, W2 and MH, with unrestricted randomization and 1-D analysis. When Kempton and Howes' paper appeared in 1981, we added a run on KH to resolve an apparent discrepancy between our conclusions and theirs. Kempton (personal communication) suggested that the discrepancy might be due to the restricted randomization in their study (they also used a 2-D analysis) so we extended our study as described above. (Kempton has since found an error in their original calculations, and their revised Monte Carlo results are now concordant with ours.)

Results

The numerical output from our study is very extensive, and only an abbreviated summary is presented here in Tables 1 and 2, with results for only the first and tenth cycles of iteration. Convergence behaviour will be reported elsewhere but we note here that 4 iterations would have been sufficient for statistical purposes.

If a Papadakis analysis were performing properly, under randomization theory, the Monte Carlo means of the treatment mean square (*TMS*) and residual mean square (*RMS*) should not differ

TABLE 1

(1a) *Factorial means of Monte Carlo means*

	TMS_1/V_0			TMS_{10}/V_0			RMS_1/V_0			RMS_{10}/V_0		
GM (excluding KH)	**1.58**			**1.25**			**1.67**			**0.84**		
Data	2.14	1.73	1.09	1.32	1.36	1.09	2.27	1.89	1.10	0.89	0.87	0.78
Rep	1.84	1.37	**1.40**	1.38	1.14	**1.04**	2.07	1.51	**1.54**	0.79	0.87	**0.84**
Rand	1.70	1.49		1.32	1.18				*Anal.*	0.82	0.86	
Max. *SE*	0.9%, 1.2% (KH)											
	$F_1/E(F)$			$F_{10}/E(F)$								
GM	**0.90**			**1.53**								
Data	0.83	0.88	1.00	1.49	1.59	1.50						
Rep	0.89	0.91	**0.90**	1.90	1.32	1.24						
Rand	0.96	0.84		1.61	1.45							
Max. *SE*	1.3% 1.8% (KH)			1.5%, 2.1% (KH)								

(1b) *Factorial means of ratios of Monte Carlo CV^2 to theoretical χ^2 or F equivalents*

	RMS_1			RMS_{10}			F_1			F_{10}		
GM	**0.69**			**0.98**			**1.13**			**1.57**		
Data	0.77	0.74	0.58	0.78	0.96	1.26	1.01	1.35	1.06	1.39	1.81	1.52
Rep	0.70	0.68	**0.90**	1.38	0.70	**0.65**	1.18	1.09	1.18	2.04	1.20	**1.10**
Max. *SE*	3.9%, 5.5% (KH)			5.5%, 7.8% (KH)			4.1%, 5.8% (KH)			5.9%, 8.3% (KH)		

Factorial means (geometric) from statistical analyses of (a) the means and (b) relative variances (CV^2) of 28 Monte Carlo distributions for TMS/V_0, RMS/V_0 and treatment *F*-ratio *F* from cycles 1 and 10 (indicated by subscripts) of iterated Papadakis analyses. Means for non-significant main effects are omitted. Experimental factors:

Data: $W1_8$, $W2_8$, MH, KH (mean for KH in bold type under *Replication*).

Replication: 2, 3, 4 (KH); *Randomization*: Unrestricted, Restricted; *Analysis*: 1-D, 2-D.

significantly from the residual mean square ignoring treatments (V_0), nor the Monte Carlo means of treatment *F*-ratio (F) from the expected values $E(F)$ of the relevant central *F* distributions. Table 1(a) presents geometric grand means and marginal means of the Monte Carlo means of TMS/V_0, RMS/V_0 and $F/E(F)$ in cycles 1 and 10, for statistically significant factors. Interactions are not included but are discussed later. Note that the method of analysis affected only RMS_{10}/V_0.

The summary means were derived from weighted factorial analyses of the logarithms of Monte Carlo means, with inverse weights the CV^2 for the corresponding central χ^2 or *F* distributions. Table 1(b) presents the corresponding entries for ratios of the CV^2 of the Monte Carlo distributions to their theoretical χ^2 or *F* equivalents, but only for the two important factors *data set* and *replication*. The entries were derived from unweighted factorial analyses of logarithms.

The main purpose of Table 2 is to show that the theory developed later in Section I.3 gives predicted values in close agreement with the empirical results in the table. The table also gives the divisors V_0 for the calculations in Table 1, the Monte Carlo means of Papadakis regression coefficients, and other particulars for 1-D analyses to be described now. For clarity the residual mean squares in Table 2 have been normalized to the dimension-free form (CV per cent)^2 with respect to the grand mean of the data.

The *intra-N RMS* in Table 2 is the within-replicate mean square of deviations of observations from their NN means. It provides, in combination with the within-replicate mean square of the

TABLE 2

Residual mean squares (RMS) ignoring treatments, NN regression coefficients and efficiency values for iterated Papadakis analyses on the data sets shown

Number of replicates:		2			3			4
Data sets:		$W1_8$	$W2_8$	*MH*	$W1_8$	$W2_8$	*MH*	*KH*
RMS's ignoring treatments								
Within reps		176.1	90.9	33.5	160.3	77.9	39.0	126.5
Intra-N		23.5	25.4	33.5	40.0	19.9	30.5	35.2
NN covariance analysis (V_0):	1-D	23.9	25.8	28.2	40.2	20.1	28.0	35.0
	2-D	21.6	23.5	28.5	33.5	19.5	27.6	35.0
Variance ratio ω		10.2	4.4	0.5	5.0	4.9	0.9	4.4
Papadakis regression coefficients (1-D)								
Ignoring treatments		0.99	0.97	0.50	0.95	0.97	0.67	0.94
Cycle 1	MC	1.15	1.09	0.54	1.06	1.04	0.70	1.04
	P	*1.18*	*1.09*	*0.56*	*1.04*	*1.04*	*0.71*	*1.00*
Cycle 10	MC	1.00	1.00	0.80	0.99	1.00	0.84	0.98
Efficiency loss % Cycle 1	MC	73	55	7	44	42	10	35
	P	*71*	*51*	*9*	*42*	*41*	*11*	*30*

MC indicates Monte Carlo means and P the corresponding values predicted from the *trend + error* model (Section I.3), and the variance ratio ω for trend effects.

unadjusted data, an estimate of a variance ratio ω for trend effects in the data given by the formula

$$1 + \hat{\omega} = RMS\,(\text{Within reps})/RMS\,(\text{intra-N})\ 1.5. \qquad (1)$$

The predicted values in Table 2 were determined by the corresponding $\hat{\omega}$ values. The Monte Carlo *efficiency loss* (per cent) for cycle 1 has been calculated as $100(1 - V_0/RMS_1)$, where V_0/RMS_1 is the inverse of the relevant mean ratio RMS_1/V_0. It is discussed further below.

Discussion of results

The F_1 ratios in Table 1(a) indicate a degree of conservative bias which is closely correlated with the variance ratio ω for trend effects. It is negligible only for the MH data, in which there is very little trend. The bias is greater with restricted randomization. Iteration, however, see F_{10}, produces a substantial positive bias which would be unacceptable in practical use. The bias is inversely correlated with the number of replicates and is less with restricted randomization, though in terms of *increase* in bias from iteration, the randomization effect is small.

The TMS/V_0 values show that there is a substantial degree of inefficiency in treatment estimation by the Papadakis method. The extent is again closely correlated with the variance ratio ω for trend effects. Iteration, see TMS_{10}/V_0, reduces the degree of inefficiency, by about a half in the worst cases, but does not eliminate it.*

The inefficiency in treatment estimation is roughly matched in cycle 1 by a corresponding upward bias in error-variance estimation, see RMS_1/V_0, which is also closely correlated with the ω ratios, so that the conservative bias in F_1 is small. Iteration, however, leads to a *negative* bias in the error-variance estimates of about 16 per cent. It is this negative bias from iteration coupled with the inefficiency of treatment estimation which produces the serious bias in the F-ratio.

* The apparent inefficiency, however, is very largely an artifact of the Papadakis analysis of covariance. See Discussion and Reply.

The efficiency of an analysis may be represented for comparative purposes as the product of two factors, namely an efficiency factor for size of error variance and an average efficiency factor for treatment estimation relative to the specified error variance. The loss of efficiency entered in Table 2 refers to the first of these factors, and is most severe just when high efficiency is needed, namely when trend effects in the data are substantial.

It may be noted from Table 2 that the Papadakis regression coefficient shifts towards unity under iteration and in the case of MH there is a substantial upward bias from iteration. One explanation is that the "autoregressive" character of the regression has been greatly intensified by iteration, but there may be others.

With regard to the relative variances of the Monte Carlo distributions, see Table 1(b). There are clearly substantial deviations from the corresponding theoretical values for χ^2 or F, with a strong dependence on the actual data of the analysis (determined by both *data set* and *replication*). Iteration exacerbates the deviations.

A final remark concerns interactions. Though not included in Table 1(a), there were significant interactions of *data set* with *replication* and *randomization*. The general effect was the same in each case, namely, that the relative effects of the *replication* and *randomization* factors were small in the case of MH and greatest with W1. The largest effects were those of *replication* on the TMS_1 and RMS_1 means for W1, these being approximately doubled by reduction of the number of replicates from 3 to 2. There were no interactions in the case of F_{10}.

I.2. SERIAL CORRELATIONS

The theory in the next Section is based on a *smooth trend + independent error* model which will be physically appropriate for field experiments if there are no interplot competition effects. Here we present some empirical evidence regarding this model.

The serial correlations presented in Fig. 2 are correlations within, and averaged over the columns of the data sets $W1_8$, $W2_8$, MH and also $W1_4$, $W2_4$. They have been calculated both before and after local detrending by subtracting means of column neighbours, which is equivalent to second differencing each column; and also, in the case of $W1_4$ and $W2_4$, after elimination of an alternating row effect attributable to the 8-row grain drill used in the experiment.

Considering first the correlations of the undetrended data, there are two noteworthy features. The first, in the case of the Wiebe data, the fairly rapid swing of the correlations from positive to negative with increasing lag, with a swing back to positive in the case of W1. This behaviour is incompatible with a stationary first-order autoregressive model as assumed by Bartlett (1978) and many of his discussants. The MH and KH data show similar behaviour, but to a weaker extent. The second feature is the extreme smoothness of trend in the correlations, which is highly symptomatic of the presence of non-stationary trend effects in the data.

By contrast the correlations of the data after detrending by NN adjustment show the kind of random sampling fluctuation one would expect, and appear to be fluctuating randomly about zero from lag 3 on; except for the W_4 data if the alternating row effect is not first removed, for then the correlations oscillate in a manner symptomatic of this row effect. The serial correlations of the detrended data are thus compatible with either an independent-error model or a finite correlation structure of order at most 2, depending on the magnitude of the first and second correlations, whose expected values are $-\frac{2}{3}$ and $\frac{1}{6}$ under the independent-error model. The independent-error model appears to be adequate for MH, KH and nearly so for $W1_4$ and $W2_4$. In the case of the 8-row Wiebe data, however, curvature effects in trend are evidently appreciable, giving rise to first serial correlations of -0.26, -0.44 respectively in the detrended data. The local linear detrending is more effective for the 4-row Wiebe data, producing first correlations of -0.41, -0.54 after elimination of the alternating row effect. Our overall conclusion is that a finite correlation structure needs to be invoked only to represent residual trend effects if the level of detrending is insufficient. A more efficient alternative is to increase the level of detrending (see Section II.7, Table 6).

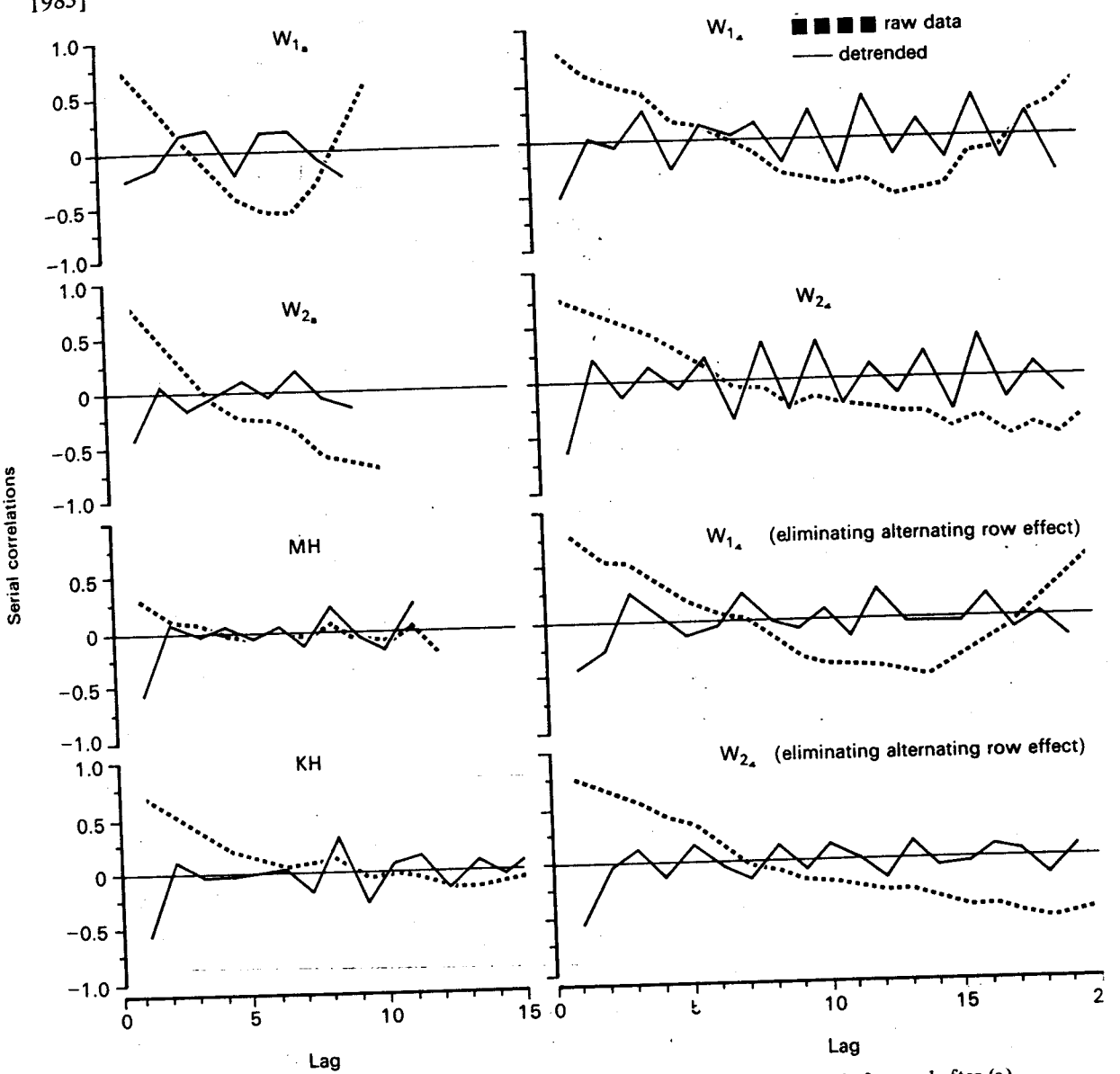


Fig. 2. Average serial correlations within columns of the W_8, MH, KH and W_4 data before and after (a) detrending by linear NN adjustment and (b) elimination of an alternating row effect in the W_4 data.

I.3. THEORY

Here we derive theoretical explanations for some of the properties of a Papadakis analysis observed in the Monte Carlo study. A *smooth trend + independent error* model (with additive treatment effects) is assumed for deriving results on efficiency. We restrict attention to 1-D analysis with respect to column neighbours, and to avoid irrelevant mathematical complications we assume that the experimental design is a randomized block design with r replicates of p treatments, plus a row of border plots at each end, and with randomization restricted to ensure first-level partial NN balance as defined in Section II.6, namely that no treatment is (i) adjacent to itself and (ii) adjacent to any other treatment more than once. Plots other than border plots are described as internal plots.

Notation

Papadakis analysis depends in a complex way on NN relations in the design, for which we need an appropriate notation. We first define some set operators $\mathcal{N}$, $\mathcal{T}$, $\mathcal{U}$ as follows, with experimental units described as plots:

$\mathcal{N}i$: the set of column neighbours of plot i, namely $\{i-1, i+1\}$ if plots are serially indexed within columns. Similarly for a set of plots S, $\mathcal{N}$ S denotes the set of all column neighbours of the plots in S.

$\mathcal{T}i$: the treatment on plot i. Similarly $\mathcal{T}$ S denotes the set of treatments on the plots S.

$\mathcal{U}j$: the set of internal plots with treatment j. Similarly for $\mathcal{U}$ S, where S is a set of treatments.

These operators may be concatenated. In particular, $\mathcal{U}\mathcal{T}i$ is the set of internal plots with the treatment on plot i. Repetitions in a set are to be distinguished. Thus $\mathcal{N}\mathcal{N}i$ or $\mathcal{N}^2 i$ denotes the neighbours of the neighbours of i and is the set $\{i-2, i, i, i+2\}$ or $\{i-2, i(2), i+2\}$, in which i occurs twice. Non-existent plots in a set, such as possibly $i-2$ or $i+2$ in the preceding example, are to be treated as null elements with no further influence on subsequent operations.

Set expressions as above will be used as subscripts to indicate means of the data over the corresponding sets. Thus $\bar{y}_{\mathcal{N}i}$ denotes the NN mean for y_i, namely $(y_{i-1}+y_{i+1})/2$, and $\bar{y}_{\mathcal{U}\mathcal{T}i}$ denotes the mean of the r internal plots with treatment $\mathcal{T}i$.

NN adjustments

The two forms of NN adjustment with which we are concerned are, for data values y_i,

$$y_i'(b) = y_i - b\bar{y}_{\mathcal{N}i}, \tag{2}$$

$$y_i^*(b) = y_i - b(\bar{y}_{\mathcal{N}i} - \bar{t}_{\mathcal{T}\mathcal{N}i}), \tag{3}$$

where b is a coefficient which in practice will be chosen to minimize the residual variance of the adjusted values, and t in (3) denotes an estimate of a treatment parameter τ, hence $\bar{t}_{\mathcal{T}\mathcal{N}i}$ the mean of such estimates for the neighbours of i.

The first is the form of adjustment used in the new method of NN analysis described in Part II. If $b = 1$ the adjustment (2) effects a complete local linear detrending by symmetric differencing, which is equivalent to second differencing with $y_i'(1) = -\frac{1}{2}\Delta^2 y_i$. If $b \neq 1$, only partial linear detrending takes place.

The second form (3) is the general form of Papadakis adjustment, with prior correction of the NN covariate for treatment effects. In the first cycle of an iterated analysis the treatment means $\bar{y}_{\mathcal{U}j}$ are used as the initial estimates of treatment parameters τ_j and (3) can be rewritten with a subscript 1 for the first cycle as

$$y_{1i}^*(b) = y_i - b_1(\bar{y}_{\mathcal{N}i} - \bar{y}_{\mathcal{U}\mathcal{T}\mathcal{N}i}). \tag{4}$$

Note that the $2r$ elements in $\mathcal{U}\mathcal{T}\mathcal{N}i$ are all distinct and do not include i, under the assumption of first-level partial NN balance.

From (4) we obtain treatment estimates

$$t_{1j} = \bar{y}_{1\mathcal{U}j}^* = \bar{y}_{\mathcal{U}j} - b_1(\bar{y}_{\mathcal{N}\mathcal{U}j} - \bar{y}_{\mathcal{U}\mathcal{T}\mathcal{N}\mathcal{U}j}). \tag{5}$$

Substitution of (5) in (3) gives the form of the Papadakis adjustment in the second cycle,

$$y_{2i}^* = y_i - b_2(\bar{y}_{\mathcal{N}i} - \bar{y}_{\mathcal{U}\mathcal{T}\mathcal{N}i}) + b_1 b_2 \underset{4r}{(\bar{y}_{\mathcal{N}\mathcal{U}\mathcal{T}\mathcal{N}i}} - \underset{4r^2}{\bar{y}_{(\mathcal{U}\mathcal{T}\mathcal{N})^2 i}}), \tag{6}$$

$$\underset{2}{} \quad \underset{2r}{}$$

with the relevant divisors shown under each mean. Now the main points to note are (i) the set $\mathcal{N}\mathcal{U}\mathcal{T}\mathcal{N}i$ includes y_i twice and (ii) $(\mathcal{U}\mathcal{T}\mathcal{N})^2 i$ includes $\mathcal{U}\mathcal{T}i$ twice, so that (6) can be re-expressed

with $c = b_1 b_2/(2r)$ as

$$y_{2i}^* = (1-c)y_i + c\bar{y}_{\mathcal{N}\mathcal{T}i} + \text{terms in other } y_k,\ k \notin \mathcal{U}\mathcal{T}i. \tag{7}$$

This kind of structure in the y_{2i}^* persists also for subsequent iterations but with the coefficient c becoming more complex. As we have observed in our Monte Carlo studies, there tends to be an alternation in convergence behaviour, which we think is due to the fact that the set $\mathcal{N}_i^k$ does not include i if k is odd, but does so if k is even.

Heuristic explanations

We can now provide heuristic explanations for the three main properties of a Papadakis analysis noted in Section I.1. However, for space reasons more detailed argument will be presented subsequently for only the most important property, the loss of efficiency.

(i) *Loss of efficiency* (as defined in Section I.1). As our argument hinges on the necessity for prior correction of the Papadakis covariate for treatment effects, note that such prior correction is not needed for Williams' Type II designs (Williams, 1952), which have exact NN balance.

Our remarks therefore do not apply to this special case. Otherwise consider the situation where there are substantial trend variations in the field so that almost complete linear detrending ($b \approx 1$) is necessary to minimize the spatial variance of the y_i' in (2), which corresponds to V_0 in Section I.1. No such complete detrending is possible with the Papadakis adjustment (3) for any value of b, because the treatment correction of the NN covariate adds in trend effects from other replicates of the experiment.† The minimum spatial variance of the $y_i^*(b)$ cannot therefore be as low as that for the $y_i'(b)$ and a loss of efficiency ensues. This loss is avoided by the new method of analysis in Part II which works directly on the $y_i'(b)$ in (2).

(ii) *Conservative bias in the first cycle*. The treatment estimate t_{1j} in (5) involves 3 means, the last of a large number, $2r^2$, of y-values. In any difference of two such estimates there is bound to be, on average, some cancellation of terms, so that differences are estimated more accurately than will be indicated by the residual mean square of the analysis. The treatment F-ratio will therefore be conservatively biassed.

(iii) *Upward bias from iteration*. We reason here by an analogy. Consider a sample of $n \geqslant 2$ independent observations $y_1, y_2, \ldots, y_n$ with common expected value μ and variance σ^2. An unweighted least squares analysis of the data will result in an unbiassed analysis of variance with the expected values of $n\bar{y}^2$ and the sample variance s^2 both equal to σ^2 if $\mu = 0$. Now consider the transformed data

$$y_i' = (1-c)y_i + c\bar{y}, \quad 0 < c < 1, \quad i = 1, 2, \ldots, n. \tag{8}$$

The expected value of y_i' is still μ, as for y_i, and $\bar{y}' = \bar{y}$. The expected value of the sample variance of the y_i', however, may be shown to be $(1-c)^2\sigma^2$. Conventional unweighted least squares analysis of the y_i' thus leads to a positively biassed analysis. Now examining the form (7) of a Papadakis-adjusted value y_{2i}^* in the second cycle of analysis, a precise parallel may be seen between (7) and (8), with treatment estimates derived from (7) unaffected by the reduced coefficients of the y_i, which will however affect the residual variance. We can therefore expect iteration to produce an upward bias in the treatment F-ratio, as we observed.

Note on randomization theory

Our development of the foregoing arguments will invoke randomization theory. The conventional approach to this treats the numerical data as fixed, with the inferential properties of the analysis determined entirely by the randomization population of possible treatment assignments. The first author's view is that this is not entirely in accord with R. A. Fisher's ideas.

† This does not imply that the trend is nearly linear. Consider for instance a sine curve for underlying trend with amplitude say double the local error standard deviation, σ_η, and verify that second differences of 30 or 40 ordinates do indeed have a small spatial variance relative to σ_η^2.

Fisher certainly considered that some conceptualization of the statistical properties of the data would be necessary, to avoid "over-conditioning" on inferentially irrelevant irregularities in the numerical data. Accordingly in the present context we shall formulate our analysis with reference to a "*bivariate*" reference set comprising the population of all possible yield realizations for the experiment apart from additive treatment effects, under a *smooth trend + independent error* model, in combination with the randomization population of possible treatment assignments. This approach also avoids some of the complexities of conventional randomization theory. (Note that the often-used term "permutation" is not applicable for describing random treatment assignment to border plots.)

From this point of view a symbol such as y_i is to be regarded below as a bivariate random variable, with index i assigned to a random distribution over the plots of the design. For the yield of a given plot we use the notation $y_i \mid i$ when necessary.

Smooth trend + independent error model

We specify the model for yield realizations on given plots i as

$$y_i \mid i = \tau_{\mathcal{F}i} + \xi_i + \eta_i(0, \sigma^2), \tag{9}$$

where ξ_i is the (fixed) trend component on plot i and the η_i are independent local errors with variance σ_η^2. We shall be concerned only with variation within replicates of the design. Assign now a uniform distribution to i over the internal plots of the experiment, and let σ_ξ^2 denote the average spatial variance of the trend components ξ_i within replicates. The within-replicate variance of y_i over the "bivariate" reference set is then

$$V(y_i) = \sigma_\xi^2 + \sigma_\eta^2. \tag{10}$$

The smoothness-of-trend assumption here is that the residual trend components after local linear detrending, namely

$$\xi_i' = \xi_i - \bar{\xi}_{\mathcal{N}i} \tag{11}$$

are relatively small relative to the standard deviation of local errors; or more precisely that the spatial variance $V(\xi_i')$ is small compared with σ_η^2.† We formally give effect to this assumption in our subsequent calculations by putting $\xi_i' = 0$, as is customary after detrending in time-series analysis; equivalently by putting $\bar{\xi}_{\mathcal{N}i} = \xi_i$.

Since we are only concerned with variance properties here, we may assume without loss of generality that the treatment constants τ_j are zero. Then the NN adjustments (2) and (4) can be re-expressed formally, using the smoothness approximation, and with an obvious extension of notation, as

$$y_i'(b) \approx (1-b)\xi_i + \eta_i'(b), \tag{12}$$

$$y_i^*(b) \approx (1-b)\xi_i + b\bar{\xi}_{\mathcal{N}\mathcal{F}\mathcal{N}i} + \eta_i'(b) + b\bar{\eta}_{\mathcal{N}\mathcal{F}\mathcal{N}i}, \tag{13}$$

omitting the subscript 1 for the first cycle.

Loss of efficiency (cycle 1)

Let $\omega = \sigma_\xi^2/\sigma_\eta^2$ and now formally put $\sigma_\eta^2 = 1$ so that variances calculated below will be multiples of σ_η^2. From (12) the variance of y_i' is

$$V(y_i') = (1-b)^2\omega + 1 + \gamma b^2/2, \tag{14}$$

where $\gamma = 1$. If there were no borders γ would be $1 + 2/d$, where d is the number of rows or columns of the design in the direction of NN adjustment. The minimum variance is therefore, together the corresponding optimal value of b,

$$V(y_i')_{\min} = 1 + (1-b)\omega = 1 + b\gamma/2; \quad b = \omega/(\omega + \gamma/2). \tag{15}$$

For the Papadakis form (13) the variance calculations are more complex since first we must distinguish between boundary and non-boundary internal plots, and in the former case invoke an additional approximation of the form $\xi_{i\pm1} \approx \xi_i$ to simplify (13) with respect to the one internal plot neighbour of a boundary plot i in $\mathcal{U}\mathcal{T}\mathcal{N}i$. All correlations between ξ, η terms in the expansion of (13) are zero except for pairs $(\xi_i, \xi_j; i \neq j)$ for plots i, j in the same replicate. This correlation is $-1/(p-1)$ where p is the number of treatments. (The correlation would be less in magnitude if the condition $i \neq j$ arising from the first-level NN balance assumption were relaxed.) For the sake of simplicity we shall omit terms of order $1/(rp)$ or $1/(rd)$ and hence obtain the variance of y_i^* as

$$V(y_i^*) = \{(1-b^*)^2 + bb^*/(2r)\}\omega + \{1 + \gamma bb^*/2\}, \tag{16}$$

where $b^* = b(r-1)/r$ and γ is 1 or $1 + 2/d$ according to the presence or absence of border plots. The minimum variance and corresponding optimal b for the Papadakis adjustments are therefore

$$V(y_i^*)_{\min} = 1 + (1 - b_p^*)\omega = 1' + b_p(\gamma + \omega/r)/2;$$
$$b_p = \omega/\{\omega(2r-1)/(2r) + \gamma/2\}, b_p^* = b_p(r-1)/r. \tag{17}$$

Comparison of (17) and (15) shows first that the optimal b for Papadakis analysis is greater than that in (15) and hence that $V(y_i^*)_{\min} > V(y_i)_{\min}$. There is thus an inherent loss of efficiency in the Papadakis method which is attributable to the prior correction of the Papadakis covariate for treatment effects. Note also that b_p^* in (17) is less than b in (15). The relative efficiency factor for variance estimation is given by the ratio

$$E = \frac{1 + (1-b)\omega}{1 + (1 - b_p^*)\omega}, \tag{18}$$

where b and b_p^* are as specified in (15) and (17) respectively. This is tabulated for a range of ω values in Table 3 with $\gamma = 1$.

TABLE 3

Papadakis relative efficiency factor (%) for variance estimation, calculated from (15), (17) and (18) with $\gamma = 1$

	ω	0	$\frac{1}{2}$	1	2	3	4	5	6	8	10	12	∞
r =	2	100	92	83	70	60	53	48	43	36	31	27	0
	3	100	95	89	79	71	65	60	55	48	43	38	0
	4	100	96	92	84	78	72	68	63	57	51	47	0

The foregoing formulae with $\gamma = 1 + 2/d$ for the no-border case have been used to calculate the predicted values for the Papadakis regression coefficients and efficiency losses in Table 2 (for cycle 1). The agreement with the empirical values is clearly good.

I.4. DISCUSSION

The problem of bias in an iterated Papadakis analysis could be rectified in principle by applying either analytically or empirically derived correction factors to the residual variance. The real disability of the method is its inherent inefficiency when there are appreciable trend effects in the data. The new method of NN analysis in Part II is largely free of this defect.

The Monte Carlo results and conclusions drawn in this paper are in conflict to some extent with those of Bartlett (1978), chiefly with regard to the hitherto unsuspected bias from iteration. The main reasons for the differences would appear to be (i) the omission in Bartlett's models of a trend component, which we have shown to be critically important, and (ii) his use of first-order autonormal models for the errors, which do not always fit the experimental data examined here, as the correlograms in Fig. 2 show. There is also a mathematical slip in his Section 2, noted

previously by Martin (1978), which affects some of his conclusions. Due to division by a total rather than residual sum of squares for the covariate, his β' is too small by a factor of $(r-1)/r$, which explains the difference he found between his β' and β''.

Yields from field experiments nearly always include, in addition to trend effects, local and independent error components attributable to plant variation within plots, local soil irregularities, etc. Thus the efficiency results of Atkinson (1969) and Ripley (1978) for a Papadakis analysis under a first-order autonormal model, see also Bartlett (1981), appear to be only of academic interest in this context, as they do not extend to what Besag (1977) has termed *errors-in-variables* formulations, or to the *trend + error* model considered here. In the absence of interplot competition an independent-error or finite-order correlation model would appear to be the most appropriate, see Wynn (1978), provided that the essential trend component is also included. Under randomization theory the Papadakis method becomes fully efficient only when it has nothing to do ($b = 0$). However it has some continuing usefulness for adjustment of standard analyses when local trend effects are relatively small.

Part II. A More Efficient Alternative

II.1. PRELIMINARY REMARKS

The alternative method of NN analysis to be described here originated from our theoretical studies of Papadakis's method, which had shown that the inherent inefficiency of the latter was due to the prior correction of the NN covariate for treatment effects. It thus occurred to the first author that treatment effects might be estimated more efficiently by equating treatment totals of the simpler NN-adjusted values $y_i - b\bar{y}_{\mathcal{N}i}$ in (2) to their expected values, initially with $b = 1$ and then with a value of b for optimal efficiency determined not by covariance analysis but from a variance ratio for trend effects in the data. A parallel with classical Rothamsted methodology for incomplete block designs immediately came to mind, where first an intra-block analysis is performed in which treatment effects are estimated from treatment totals of intra-block deviations $y_i - \bar{y}_{\mathcal{B}i}$, where $\mathcal{B}i$ denotes the block containing plot i, and hence $\bar{y}_{\mathcal{B}i}$ the corresponding mean, and then (if necessary) a recovery of inter-block information with an analysis equivalent to analysing partially block-detrended values $y_i - b\bar{y}_{\mathcal{B}i}$ with a coefficient b for optimal efficiency determined from a variance ratio for interblock variation. We envisaged, therefore, intra-N and optimal NN analyses as the corresponding phases of analysis in a *moving block* extension of the classical *fixed block* methodology, with NN adjustment providing the continuous analogue of the step-wise block-detrending of classical analysis. (See Section II.7, however, for comments regarding 2-way trend elimination in classical analysis.)

The foregoing analogy breaks down in one important respect. For whereas in classical analysis the elimination of block effects is complete in that the transformation to intra-block deviations is irreversible, the same is not true of NN detrending, since the corresponding intra-N transformation is largely reversible. In the case of column NN adjustment for an appropriately bordered design, for instance, the reverse transformation would recover the original data modulo linear trend within columns so that, if we represent the intra-N transformation by $\mathbf{y} \to \mathbf{N}\mathbf{y}$, a weighted least squares analysis of $\mathbf{N}\mathbf{y}$ with a generalized inverse of $\mathbf{N}\mathbf{N}^{\mathrm{T}}$ as weight matrix would be equivalent to an unweighted analysis of the original data with elimination only of linear trend within columns.

Thus the hoped-for gain in efficiency of treatment estimation from continuous local detrending by NN adjustment will not materialize unless an estimation process is applied which effectively "breaks" this kind of reversible connection with the original data. Basing estimation on the unweighted treatment totals of the NN-adjusted data (and unweighted sums of products for covariates, if any) achieves this effect and leads to highly efficient estimates, as our Monte Carlo studies in Section II.7 confirm. The device also seems peculiarly apposite in relation to randomization theory, which conventionally is concerned only with the numerical values of the adjusted data and their modified expectations, the statistical nature of the adjustments (elimination of trend, etc.) being otherwise irrelevant.

Exact randomization theory for the new method will not be given here, but a joint study of it has been initiated with A. W. Davis. We have conjectured that when adjustment is with respect to nearest neighbours, as considered here, and certain restrictions on randomization are applied to ensure a degree of NN balance in the design, the statistical properties of the estimates under randomization will depend primarily on only the spatial variance and the relevant first and second serial correlations of the adjusted data excluding treatment effects (in comparison with classical theory for standard designs, where spatial variance alone is of primary relevance). Dr Davis's preliminary work provides some support for this.

In the following two Sections we first describe the method in detail for linear 1-D NN analysis for a single treatment factor and illustrate it with an analysis of a Rothamsted experiment with exact NN balance in the sense of Williams (1952). In the next two Sections we deal with extensions of the method for covariates and for 2-D NN analysis. *We restrict attention throughout to randomized complete-block designs with appropriately treated border plots for NN adjustment.* Some restriction on randomization is necessary to secure a degree of NN balance for optimal efficiency of treatment estimation. Appropriate balance conditions are described in Section II.6. Results of Monte Carlo studies and related data which provide empirical evidence on the approximate validity of the method under randomization and its efficiency are presented in Section II.7. Implications of the new methodology are discussed in Section II.8.

Computer implementation of the methodology for both design generation and data analysis will be only briefly mentioned. Fuller details of our computer programs will be given elsewhere.

II.2. LINEAR 1-D NN ANALYSIS

This simplest form of NN analysis is effective with long narrow plots as commonly used in cereal breeding experiments. The relevant form of NN adjustment is as specified in (2), Section I.3, with respect to nearest-neighbour plots across the narrower plot dimension. An alternative representation of (2) is as a mixture of y_i and the intra-N deviation $y_i'(1) = y_i - \bar{y}_{\cdot i}$,

$$y_i'(b) = y_i - b\bar{y}_{\cdot i} = (1-b)y_i + by_i'(1). \tag{19}$$

In the first, intra-N analysis phase, b is taken to be 1, so that NN adjustment effects complete local linear detrending, as noted in Section I.3. From this analysis a value of b is determined for a subsequent, optimized analysis.

Assuming physical stratification of the experiment into r replicates of p treatments in complete blocks of p plots, with appropriate border plots for NN adjustment, the relevant data for treatment estimation are the within-block deviations of the $y_i'(b)$, namely

$$y_i''(b) = y_i'(b) - \overline{y'_{\cdot i}}(b) = (1-b)(y_i - \bar{y}_{\cdot i}) + b(y_i'(1) - \overline{y'_{\cdot i}}(1)). \tag{20}$$

Note that reducing the intra-N deviations by their relevant block means $\overline{y'_{\cdot i}}$ has the important statistical effect of eliminating additionally an average curvature effect of trend over each block.

Estimation equations

For either an intra-N analysis ($b = 1$) or a subsequent optimized analysis (see below) the estimation equations for treatment parameters are simply

$$E(T_j'') = T_j'', \quad j = 1, 2, \ldots, p, \tag{21}$$

where the T_j'' are the totals of the $y_i''(b)$ over the internal plots with the respective treatments j. In practice we first form the simpler equations

$$E(T_j') = T_j', \quad j = 1, 2, \ldots, p, \tag{22}$$

where T_j' are treatment totals of the $y_i'(b)$. An explicit matrix representation of the equations is

$$\mathbf{B}'\boldsymbol{\tau} = (r\mathbf{I}_p - \tfrac{1}{2}b\mathbf{A})\boldsymbol{\tau} = \mathbf{T}', \tag{23}$$

where $\boldsymbol{\tau}$ is the vector of treatment parameters and $\mathbf{A}$ is a $p \times p$ "adjacency" matrix with entries a_{jk}

equal to the number of times treatment k is a neighbour of a treatment j on an internal plot. (Treatments on border plots render **A** non-symmetric.) The equations are then converted to the required form (21), in matrix notation $\mathbf{B}''\boldsymbol{\tau} = \mathbf{T}''$, by premultiplying both sides by the projection matrix $\mathbf{Q} = \mathbf{I}_p - \mathbf{1}_p\mathbf{1}_p^{\mathrm{T}}/p$, since $\mathbf{T}'' = \mathbf{QT}'$ and hence $\mathbf{B}'' = \mathbf{QB}'$. Computationally the **Q** operation on (23) performs "column sweeps" on $\mathbf{B}'$, $\mathbf{T}'$, that is, reduces each column by its mean to sum to zero.

The equations $\mathbf{B}''\boldsymbol{\tau} = \mathbf{T}''$ are consistent and although singular are readily solved by say the Jordan inversion algorithm (Hemmerle, 1967) if the reciprocals of "zero" pivots (only one in this case) are set to zero as they occur. Their solution is unique modulo an arbitrary additive constant. The Jordan algorithm returns the particular solution with $\hat{\boldsymbol{\tau}}_p = 0$, and to eliminate arbitrariness in subsequent calculations below it is necessary to adjust the solution $\hat{\boldsymbol{\tau}}$ by a mean sweep $\hat{\boldsymbol{\tau}} \rightarrow \mathbf{Q}\hat{\boldsymbol{\tau}}$ to sum to zero. We also need to know the adjusted solution in the explicit form $\hat{\boldsymbol{\tau}} = \mathbf{MT}'$, as the multiplier matrix **M** is needed subsequently. If **M** is initially the generalized inverse produced by the Jordan algorithm, the transformation $\mathbf{M} \rightarrow \mathbf{QMQ}$ reduces it by row and column sweeps to the required matrix **M**.

Residuals, analysis of variance

Note first that the estimation equations above are completely determined by the initial totalling operation without reference to correlational properties of the data. We can also form at this stage a vector of residuals

$$\mathbf{z}'' = \mathbf{y}'' - \hat{\mathbf{E}}(\mathbf{y}''; \hat{\boldsymbol{\tau}}) \tag{24}$$

and hence a residual sum of squares $\mathbf{z}''^{\mathrm{T}}\mathbf{z}''$. Computationally we first form y-residuals $\mathbf{z}$ with $z_i = y_i - \hat{\tau}_{\mathcal{F}i}$, then transform $\mathbf{z}$ by NN adjustment to $\mathbf{z}'$ (with the same b as before), and finally reduce $\mathbf{z}'$ to $\mathbf{z}''$ with a sweep of its block means.

The spatial correlation structure of the NN adjusted values $y_i'(b)$ now enters into the calculations, first to determine a variance matrix for the treatment-estimates and hence a mean square for treatments, and second to determine a residual degrees-of-freedom divisor for the residual sum of squares, to complete the analysis of variance. According to our conjecture in Section II.1, only the first two serial correlations (ρ_1', ρ_2') of the $y_i'(b)$ in the direction of NN adjustment are of primary relevance in these calculations, assuming the necessary degree of NN balance in the design. We therefore proceed formally by specifying all other correlations of the $y_i'(b)$ to be zero.

Two methods of determining values for ρ_1', ρ_2' will be discussed below. For the present we take their assignment for granted. A variance matrix $\mathbf{V}_T$ for $\mathbf{T}'$ as a multiple of the within-block variance (v') of the $y_i'(b)$ is then readily determined by an incremental algorithm scanning for nearest neighbours. An explicit representation of its (j, k)th element (see Section I.3 for notation) is

$$V_T(j,k) = r\delta_{jk} + \sum_{i \in \#j}^{r} \sum_{\pm}^{2} \{\rho_1' \, \delta_{\mathcal{F}(i\pm1),k} + \rho_2' \, \delta_{\mathcal{F}(i\pm2),k}\} \tag{25}$$

where δ is the Kronecker delta and $\pm$ indicates summation over pairs of terms with alternatively plus and minus signs in their subscripts, omitting terms referring to border plots.

The variance matrix $\mathbf{V}_\tau$ of the treatment estimates as a multiple of v' and hence a mean square *TMS* for treatment effects in terms of a generalized inverse $\mathbf{V}_\tau^{-1}$ of $\mathbf{V}_\tau$ are evaluated as

$$\mathbf{V}_\tau = \mathbf{M}\mathbf{V}_T\mathbf{M}^{\mathrm{T}}, \quad TMS = (\hat{\boldsymbol{\tau}}^{\mathrm{T}}\mathbf{V}_\tau^{-}\hat{\boldsymbol{\tau}})/(p-1). \tag{26}$$

Now consider each z_i' in the explicit form $z_i' = \mathbf{r}_i^{\mathrm{T}}\mathbf{y}'$ with $\mathbf{r}_i$ derived via the matrix **M**. If $\mathbf{C}'$ denotes the assigned correlation matrix for $\mathbf{y}'$, the expected value of the sum of squares $\mathbf{z}'^{\mathrm{T}}\mathbf{z}'$ is

$$E(\mathbf{z}'^{\mathrm{T}}\mathbf{z}') = \sum_i E(z_i'^2) = \sum_i \mathbf{r}_i^{\mathrm{T}}\mathbf{C}'\mathbf{r}_i = \sum_i \mathrm{trace}\,(\mathbf{C}'\mathbf{r}_i\mathbf{r}_i^{\mathrm{T}}), \tag{27}$$

each term of which is computed progressively with an incremental scanning algorithm. Subtraction of a correction factor for block means then gives the expected value of the residual sum of squares as

$$E(\mathbf{z}''^{\mathrm{T}}\mathbf{z}'') = E(\mathbf{z}'^{\mathrm{T}}\mathbf{z}') - r(1 + 2\alpha_1\rho_1' + 2\alpha_2\rho_2'), \tag{28}$$

where, if each block has d rows or columns in the direction of NN adjustment, $\alpha_k = (d-k)/d$. This is the equivalent degrees-of-freedom divisor (*EDF*) for computing the residual mean square *RMS* of the analysis of variance.

Assignment of correlations

We discuss two methods here, which assume a *smooth trend + error* model of the form (9) in Section I.3 in terms of which $y_i'(b)$ may be re-expressed approximately assuming sufficient smoothness of trend as

$$y_i'(b) \approx (1-b)\xi_i + \eta_i'(b), \tag{29}$$

omitting treatment parameters.

In method 1, which we use in practice, we assume the errors η_i in (9) are *independent* (uncorrelated would be mathematically sufficient), which leads to the values $\rho_1' = -\frac{2}{3}$, $\rho_2' = \frac{1}{6}$ for the intra-N analysis with $b = 1$. Then from a preliminary randomized block analysis, the *RMS* of which estimates $(\omega + 1)\sigma_\eta^2$, and the intra-N analysis *RMS*, which estimates $v' = 1.5\sigma_\eta^2$, we obtain an estimate of the ratio $\omega = \sigma_\xi^2/\sigma_\eta^2$ for trend effects (see Section I.3), which is used to determine an optimal b value (see below). For the optimized NN analysis the correlations ρ_1' and ρ_2' are redefined in terms of ω and b from the relations (omitting a common multiplier σ_η^2)

$$v' = (1-b)^2\omega + (1 + b^2/2), \quad v'\rho_1' = (1-b)^2\omega - b, \quad v'\rho_2' = (1-b)^2\omega + b^2/4, \tag{30}$$

in which the common term $(1-b)^2\omega$ arises from assuming the serial correlations of the trend components ξ_i to be approximately 1 by virtue of smoothness of trend. Note that some loss of efficiency is to be expected if the level of detrending is inadequate or if the independence assumption breaks down in some other way. Evidence on this will be presented in Section II.7.

To check the effects on NN analysis of assuming an independent-error model we have also analysed our uniformity data in Section II.7 with ρ_1' and ρ_2' determined directly as the serial correlations of the adjusted data y_i'', ignoring treatments. This is method 2. However, we do not have a workable method of estimating ρ_1' and ρ_2' in the presence of real treatment effects. We cannot use the serial correlations of the treatment-corrected residuals z_i'' for ρ_1' and ρ_2' without some adjustment, as our Monte Carlo studies show that doing so leads to downward bias in the estimates of error variance. A conventional approach would be to form linear equations to transform the serial variance array of the residuals to what it would have been theoretically if no adjustment for treatments had been needed, but our attempt at this approach led to highly unstable estimates which repeatedly gave rise to variance matrices for treatment effects which violated the requirement of non-negative definiteness. A basic statistical difficulty is that the imposition of treatment effects on the data according to a randomized plan partially obliterates the serial correlation structure we are seeking to estimate.

With method 2 it is necessary to transform the variance array $[1, \rho_1', \rho_2']\, v'$ estimated from the y_i'' with $b = 1$ to an equivalent array $[1, \rho_1, \rho_2]\, \sigma_\eta^2$ for the errors η_i, assuming $\rho_l(\eta) = 0$ for $l > 2$. A computational method of doing this which lends itself to generalization for 2-D NN analysis (Section II.5) is as follows. Let $\mathbf{x}_0$ denote the 5-element vector, with $b \neq 1$ in general,

$$\mathbf{x}_0 = [-\tfrac{1}{2}b, 1, -\tfrac{1}{2}b, 0, 0] \tag{31}$$

and $\mathbf{x}_l$ the cyclically right-shifted version of $\mathbf{x}_0$ with shift $l \geqslant 0$. Let $\mathbf{u}_l$ denote the 5-element vector comprising sums $u_{l,i}$ of neighbouring elements $x_{0,i\pm l}$ (if they exist) of $x_{0,i}$ in $\mathbf{x}_0$. Then with $\mathbf{u}_0 = \mathbf{x}_0$, the required relation connecting the ρ_l and ρ_l' is

$$\begin{bmatrix} 1 \\ \rho_1' \\ \rho_2' \end{bmatrix} v' = \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \otimes [\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2] \cdot \begin{bmatrix} 1 \\ \rho_1 \\ \rho_2 \end{bmatrix} \sigma_\eta^2, \qquad (32)$$

where the multiplication operation for the direct product (a 3 × 3 matrix) is the calculation of inner products of the vector pairs $(\mathbf{x}_l, \mathbf{u}_m)$, $0 \leqslant l, m \leqslant 2$. In the intra-N analysis the equations (32) are solved for σ_η^2, ρ_1 and ρ_2 and hence an estimate of ω obtained from the randomized block *RMS* which estimates $(\omega + 1)\sigma_\eta^2$.

For optimal NN analysis the equations (32), with $\mathbf{x}_0$ defined by the optimal b value, produce a variance array $[1, \rho_1', \rho_2']\, v'$ as a multiple of σ_η^2, the elements of which must be augmented by $(1-b)^2\omega$ as in (30), to allow for the ξ_i term in (29). Rescaling this array then gives the appropriate values of ρ_1', ρ_2'.

Optimal NN analysis

Choosing an adjustment coefficient b in the NN-adjusted values $y_i''(b)$ to optimize the efficiency of treatment estimation reduces approximately to choosing its value to minimize the variance v' of the adjusted values, because the average efficiency factor e for treatment estimation varies very slowly with b, compared with the *RMS*. (The factor e is defined to be the nominal variance $2v'/r$ divided by the average variance for estimated treatment differences.) Since $y_i'(b) = y_i - b\bar{y}_{\cdot vi}$, the optimization problem is formally analogous to that of linear covariance analysis within blocks, and the optimal b is given in terms of within-block variance and covariance by

$$b = \frac{\mathrm{cov}\,(y_i, \bar{y}_{\cdot vi})}{V(\bar{y}_{\cdot vi})} \approx \frac{\omega + \rho_1(\eta)}{\omega + (1 + \rho_2(\eta))/2}, \qquad (33)$$

noting that $\bar{y}_{\cdot vi} \approx \xi_i + \bar{\eta}_{\cdot vi}$. Under the independent-error model, (33) gives $b = \omega/(\omega + \frac{1}{2})$, though this value may be slightly too low if the level of NN adjustment is inadequate, since local trend curvature tends to make $\rho_1(\eta) > 0$, and likewise ρ_2, to a lesser degree.

Note that the optimization problem is not solved in practice by a covariance analysis of treatment-corrected values z_i on their NN means $\bar{z}_{\cdot vi}$. Though the latter would minimize approximately the residual sum of squares of the z_i (not exactly because the estimates of treatment effects which effect both the z_i and the $\bar{z}_{\cdot vi}$ would be revised in the subsequent analysis), the degrees-of-freedom divisor for converting the sum of squares to an *RMS* estimating v' is not constant as in a standard covariance analysis, but in fact increases as b decreases.

II.3. ANALYSIS OF A ROTHAMSTED EXPERIMENT

Detailed Monte Carlo studies of the new method will be presented in Section II.7. Here we demonstrate the effectiveness of the method with the results of NN analysis of a Rothamsted experiment on control of mildew in barley (Jenkyn *et al.*, 1979) using an NN-balanced design (Dyke and Shelley, 1976).

The data analysed are the 1975 grain yields which may be found in Draper and Guttman (1980), who also analysed them by a different method. There were 4 spray treatments, *nil*, *early*, *late* and *repeated*, labelled $(0, 1, 2, R)$. The experiment comprised a single row of 38 plots subdivided into 9 blocks of 4 with an additional treated border plot at each end. The design was exactly balanced in the sense that every treatment occurred in combination with every other treatment as a neighbouring treatment, on both left- and right-hand sides.

The results are summarized in Table 4. As the Rothamsted analysis was extended to estimate neighbouring as well as direct treatment effects, we did likewise with our NN analysis. The 16 degrees of freedom for NN treatments comprise 2 each for left- and right-hand treatment comparisons for each of the 4 direct treatment levels. However, our analysis showed no significant

TABLE 4

Intra-N analysis summary of Rothamsted experiment with comparisons of residual variances. Grain yields in 0.01 tons/hectare units, grand mean 580

	DF	*F*	*Efficiency factors (%)*	
			NN analysis	*Rothamsted*
Treatment F-ratios:				
Direct treatments	3	68.1	122	98
NN treatments	16	1.7	92	73

	DF	*RMS*	*RMS/e* (direct treatments)	
Residual variances:				
NN analysis ignoring NN treatments	28.3	175	144	Efficiency gain 49%
cf. Rothamsted analysis eliminating NN treatments	12	209	214	
Rothamsted analysis ignoring NN treatments	28	422		
Randomized Block analysis	24	362		
Draper and Guttman analysis	23	273		

Treatment estimates:	*0*: −49	*1*: 6	*2*: 21	*R*: 22	*SED*: 5.7
NN analysis					
RB analysis	−49	7	29	14	9.0

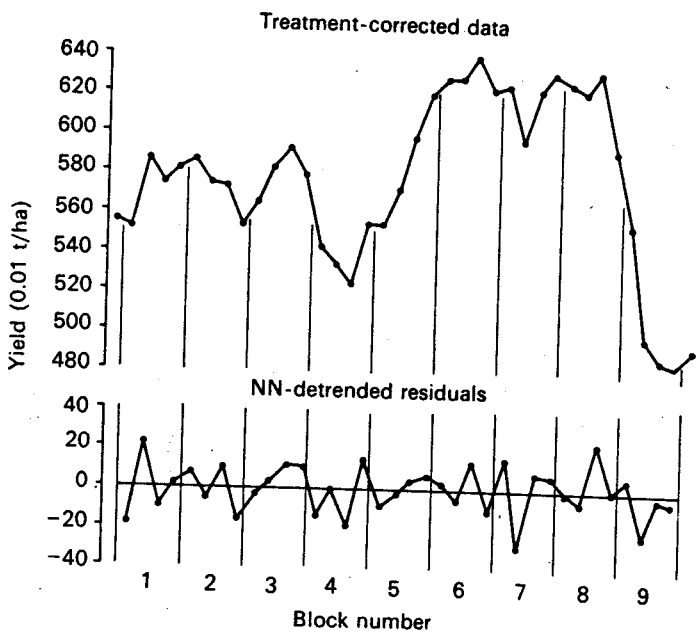


Fig. 3. Treatment corrected data from Rothamsted experiment, showing the trend effects and the NN detrended residuals.

evidence of neighbour treatment effects and we have therefore presented only the *RMS* ignoring NN treatments, from an intra-N analysis which turned out to be close to optimal for the data. Residual serial correlations were −0.59, +0.24 (cf. $-\frac{2}{3}, \frac{1}{6}$). Fig. 3 shows the effectiveness of NN detrending for the Rothamsted data.

Because of the different bases of calculation of the *RMS*, the Rothamsted efficiency factors differ from ours (though in the same proportion). It is therefore more appropriate to compare effective variances *RMS/e*, which show an improvement of nearly 50 per cent in nett efficiency by NN analysis, in comparison with the Rothamsted analysis. The gain in efficiency relative to a randomized block analysis for direct treatments only (and note there are only 4 plots per block) is 151 per cent, and more than 90 per cent relative to the Draper and Guttman analysis (since their efficiency factor would have been less than 1). The latter analysis included one additional constant which essentially did no more than "round the corners" off the stepwise block-detrending of a randomized block analysis.

In the Rothamsted analysis the block structure was ignored and a 4-term Fourier series fitted for detrending. With this method there is a large reduction in *RMS* from 422 to 209 when the additional constants for NN treatments are included in the analysis. This would have led Jenkyn *et al.* to claim significant effects for NN treatments ($F = 2.8$), whereas the intra-N analysis results suggest in fact that the additional NN treatment constants in the Rothamsted analysis have merely acted as additional local detrending constants, though not as efficiently as NN adjustment in the intra-N analysis. Note that the cost of NN adjustment is represented by a loss of only 3.7 degrees of freedom, in comparison with 8 d.f. for stepwise block detrending. For further discussion see Section II.8.

Note finally that the anomalous drop in treatment effect for R from a randomised block analysis is eliminated by the NN analysis. This illustrates the kind of artifact than an inefficient analysis can produce.

II.4. COVARIANCE ANALYSIS

This extension is needed in particular to adjust an NN analysis for missing data, using indicator variables for the missing values as covariates. If there are q covariates $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_q$, the estimation (24) are augmented with q equations equating sums of products of NN-adjusted covariates $\mathbf{x}_k''$ with $\mathbf{y}''$ to their expected values, that is, with $SP_k = \mathbf{x}_k''^{\mathrm{T}}\mathbf{y}''$.

$$E(SP_k) = SP_k, \quad k = 1, \ldots, q. \tag{34}$$

It is important to note that SP_k is also equal to $\mathbf{x}_k''^{\mathrm{T}}\mathbf{y}'$, as this simplifies both forming the estimation equations and determining an augmented variance matrix $\mathbf{V}_{C+T}$ for the right-hand sides from ρ_1' and ρ_2', as a multiple of v'. If $\mathbf{V}_{CC}$, $\mathbf{V}_{TC}$ and $\mathbf{V}_{TC}^{\mathrm{T}}$ denote the additional submatrices in the augmented variance matrix, and $\mathbf{x}_{k,l}''$ denotes a lagged version of $\mathbf{x}_k''$ with lag l and with undefined elements set to zero, the additional elements of $\mathbf{V}_{C+T}$ are

$$V_{CC}(j,k) = \mathbf{x}_j''^{\mathrm{T}}\mathbf{x}_k'' + \sum_{i=1}^{2} \rho_i' \{\mathbf{x}_j''^{\mathrm{T}}\mathbf{x}_{k,-i}'' + \mathbf{x}_{j,-i}''^{\mathrm{T}}\mathbf{x}_k''\}, \tag{35}$$

$$V_{TC}(j,k) = T_j(\mathbf{x}_k'') + \sum_{i=1}^{2} \rho_i' \{T_j(\mathbf{x}_{k,-i}'') + T_j(\mathbf{x}_{k,+i}'')\}, \tag{36}$$

where T_j indicates a total for treatment j. In the estimation equations let γ_k, $k = 1, \ldots, q$, denote the covariance parameter for $\mathbf{x}_k$, and $\mathbf{E}_\tau(\mathbf{y}')$ the treatment component of the expectation of $\mathbf{y}'$. Then noting that $\mathbf{x}_k''^{\mathrm{T}}\mathbf{x}_l' = \mathbf{x}_k''^{\mathrm{T}}\mathbf{x}_l''$,

$$E(SP_k) = \mathbf{x}_k''^{\mathrm{T}}\mathbf{E}_\tau(\mathbf{y}') + \sum_{l=1}^{q} \gamma_l \mathbf{x}_k''^{\mathrm{T}}\mathbf{x}_l'', \quad k = 1, \ldots, q \tag{37}$$

and

$$\mathbf{E}(\mathbf{T}(\mathbf{y}')) = \mathbf{B}'\tau + \sum_{l=1}^{q} \gamma_l \mathbf{T}(\mathbf{x}_l'), \tag{38}$$

which is adjusted to $\mathbf{E}(\mathbf{T}(\mathbf{y}''))$ as in Section II.2.

The augmented estimation equations may be solved by the Jordan algorithm as in Section II.2. If $\mathbf{M}_{C+T}$ is the inverse matrix produced by this algorithm, with submatrices $\mathbf{M}_{CC}, \mathbf{M}_{TC}, \mathbf{M}_{CT}, \mathbf{M}_{TT}$, the following row and column sweeps with $\mathbf{Q}$ convert it to the required form for expressing the solutions in terms of the SP_k and T_j' and hence the y_i':

$$\mathbf{M}_{TC} \to \mathbf{Q}\mathbf{M}_{TC}, \quad \mathbf{M}_{CT} \to \mathbf{M}_{CT}\mathbf{Q}, \quad \mathbf{M}_{TT} \to \mathbf{Q}\mathbf{M}_{TT}\mathbf{Q}. \tag{39}$$

Completion of the NN analysis proceeds as in Section II.2 but with separately weighted sums of squares for covariate and treatment effects, and the additional subtraction of estimated covariate effects when forming the residual vector $\mathbf{z}''$. A corresponding extension of the algorithm for computing the expected value of the residual sum of squares automatically adjusts the residual degrees of freedom for the covariate adjustments.

II.5. 2-D NN ANALYSIS

Even with long narrow plots an increase in efficiency may result if a linear 1-D analysis is extended to give second degree detrending using second-nearest neighbours as well as first, and with plots whose dimensions differ by only a factor of 3 or less, additional efficiency may be gained alternatively with 2-D NN adjustments with respect to row, column and diagonal NNs. The complications in the analysis are essentially the same for both situations assuming appropriate border plots, so we restrict attention here to 2-D analysis.

The general form of NN adjustment is specified by

$$y_i'(\mathbf{b}) = y_i - \sum_{m=1}^{3} b_m \bar{y}_{\backslash m i} \tag{40}$$

with respect to column (i.e. row-lagged), row (column-lagged) and 4 diagonal NNs respectively. Under the smooth-trend assumption the NN-adjusted values will be approximately free of the trend components ξ_i if the sum of the b_m is unity and we use the term *intra-N analysis* for this case. Otherwise $y_i'(\mathbf{b})$ may be re-expressed approximately in terms of the model (9), omitting treatment constants, as

$$y_i'(\mathbf{b}) \approx (1 - b_S)\xi_i + \eta_i'(\mathbf{b}); \quad b_S = \sum_{m=1}^{3} b_m. \tag{41}$$

If $b_1 = b_2 = 1$, $b_3 = -1$, NN adjustment provides full second degree detrending in two dimensions, with $y_i' = \frac{1}{4}\Delta_R^2\Delta_C^2 y_i$, where Δ_R^2 and Δ_C^2 are second difference operators for the row and column directions. However, we often find that in an optimal NN analysis b_1 say is close to 1 and b_2 and b_3 are much smaller, with $b_2 \approx -b_3$, if the plots are narrower in the column direction. In effect one coefficient (row or column) provides nearly all the linear detrending effect and the other two chiefly a partial adjustment for curvature. See the examples in Section II.7.

In the estimation equations the matrix $\mathbf{B}'$ now becomes

$$\mathbf{B}' = r\mathbf{I} - \sum_{m=1}^{3} \frac{b_m}{d_m} \mathbf{A}_m; \quad d_m = 2, 2, 4 \text{ respectively}, \tag{42}$$

where $\mathbf{A}_m$ is the adjacency matrix for the mth class of neighbour. The correlation structure for variance calculations comprises 8 correlations ρ_{kl}' with row and column lags $k, l \leqslant 2$, and there are 8 corresponding classes of neighbour with lags $(\pm k, \pm l)$, with 4 neighbours in each class if both lags are non-zero. The variance formulae (25), (35) and (36) generalize in a straightforward manner for the additional classes of neighbour.

Under the *independent-error* model we assign zero values to the correlations ρ_{kl} for the η_i and

map to corresponding values ρ'_{kl} for the NN-adjusted values (method 1). In method 2 we estimate the ρ'_{kl} from the $y''_i(\mathbf{b})$ ignoring treatments and map back to the ρ_{kl}. The generalization of the mapping (32) is as follows. Let $\mathbf{x}_{00}$ denote the 5 × 5 matrix

$$\begin{bmatrix} -b_3/4 & -b_1/2 & -b_3/4 & 0 & 0 \\ -b_2/2 & 1 & -b_2/2 & 0 & 0 \\ -b_3/4 & -b_1/2 & -b_3/4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{43}$$

and let $\mathbf{x}_{kl}$ denote the cyclically right- and down-shifted version of $\mathbf{x}_{00}$ with shifts k, l respectively. Define 5 × 5 matrices $\mathbf{u}_{kl}$ comprising sums of neighbours in $\mathbf{x}_{00}$ (if they exist) with lags $(\pm k, \pm l)$, let $\mathbf{X}$, $\mathbf{U}$ be 9 × 1 symbolic vectors with elements $\mathbf{x}_{kl}$, $\mathbf{u}_{kl}$ in standard order, and let $\boldsymbol{\rho}$, $\boldsymbol{\rho}'$ be the corresponding 9 × 1 vectors of the correlations ρ_{kl}, ρ'_{kl}. Then the required mapping relation analogous to (32) is

$$\boldsymbol{\rho}' v' = \mathbf{X} \otimes \mathbf{U} \,.\, \boldsymbol{\rho}\, \sigma_\eta^2, \tag{44}$$

the direct product being a 9 × 9 matrix comprising sums of the 25 products of corresponding elements in the matrix pairs $\{\mathbf{x}_{kl}, \mathbf{u}_{mn}\}$, $0 \leqslant k, l, m, n \leqslant 2$. In (28) the correction factor for block means in the residual degrees of freedom becomes

$$-r\{1 + 2 \sum_{k,\, l} \alpha_{kl} \rho'_{kl}\}; \quad \alpha_{kl} = (d-k)\,(c-l)/(dc), \tag{45}$$

where d, c are the numbers of rows and columns respectively in a block.

For an optimized analysis the variance array $\boldsymbol{\rho}'v'$ inferred from $\boldsymbol{\rho}(\eta)\sigma_\eta^2$ via (44) is augmented by $(1 - b_S)^2 \omega \mathbf{1}$ and rescaled to adjust $\boldsymbol{\rho}'$ for the ξ_i term in (41).

With the above modifications of the relevant formulae, each analysis phase proceeds as described in Section II.2, except for the following modifications to determine optimal b values.

For method 1 a first intra-N analysis is performed with provisional input b values summing to 1 such as (1, 0, 0), solely to determine provisional estimates of treatment effects to enable treatment corrected values $z_i = y_i - \hat{\tau}_{\mathcal{T}i}$ to be calculated. The z_i are also likewise adjusted for covariate effects (if any). A within-block NN covariance analysis of the z_i with respect to their NN means $\bar{z}_{\mathcal{N}mi}$, $m = 1, 2, 3$, is then performed to determine an optimal *balance* of NN adjustment with respect to the 3 classes of neighbour, with a Lagrangian adjustment of the equations for $\mathbf{b}$ to ensure that $b_S = 1$. (The form of the adjustment is as specified below for method 2.) While relative b values determined in this way appear to be reasonably satisfactory for an intra-N analysis, covariance analysis of the z_i does not give a satisfactory value for b_S, for the reason noted at the end of Section II.2. A second intra-N analysis with the optimally balanced NN adjustment is then performed to determine an *RMS* for estimating the variance ratio ω for trend effects, noting that $E(RMS) = v' = (1 + \nu)\sigma_\eta^2$, where $\nu = \Sigma_{m=1}^{3} b_m^2/d_m$ under the assumption of independent errors. An optimal value of b_S is given by $b_S = \omega/(\omega + \nu)$, and $\mathbf{b}$ is rescaled with this value for a final, optimized analysis.

It must be stressed that the *smooth trend + independent error* model is only being invoked to characterize the variance structure of the $y'_i(\mathbf{b})$ for optimally balanced NN adjustment. If 2-D adjustment is really needed the model will not provide an adequate characterization for values adjusted with respect to one class of neighbour alone and it cannot therefore be used to determine an optimal balance of adjustment (it would lead in fact only to the answer $b_1 = b_2 = \frac{1}{2}b_3$, corresponding to a mean of all 8 NN values).

In method 2 we also do two intra-N analyses, using the correlations ρ_{kl} calculated from the corresponding ρ'_{kl} in the first to determine optimally balanced NN adjustments for the second by a theoretical NN covariance analysis of the y_i. The theoretical regression equations, in a

slightly modified form with a Lagrangian correction $\alpha\mathbf{1}$ of the right-hand side to ensure that $b_S = 1$, are

$$\frac{1}{2}\begin{bmatrix} 1+\rho_{20} & \rho_{11} & \rho_{01}+\rho_{21} \\ \rho_{11} & 1+\rho_{02} & \rho_{10}+\rho_{12} \\ \rho_{01}+\rho_{21} & \rho_{10}+\rho_{12} & \frac{1}{2}(1+\rho_{20}+\rho_{02}+\rho_{22}) \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \rho_{10} \\ \rho_{01} \\ \rho_{11} \end{bmatrix} + \alpha\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad .(46)$$

With the equations expressed in matrix notation as $\mathbf{\Lambda b} = \boldsymbol{\lambda} + \alpha\mathbf{1}$, the required value of α is

$$\alpha = (1 - \mathbf{1}^T \mathbf{\Lambda}^{-1} \boldsymbol{\lambda})/(\mathbf{1}^T \mathbf{\Lambda}^{-1} \mathbf{1}). \quad (47)$$

The correlations are recalculated in the second intra-N analysis and hence an estimate of ω is obtained as in Section II.2. For the final optimized analysis $\mathbf{b}$ is recalculated from the equation

$$(\mathbf{\Lambda} + \omega\mathbf{1}\mathbf{1}^T)\,\mathbf{b} = \boldsymbol{\lambda} + \omega\mathbf{1}. \quad (48)$$

II.6. DESIGNS WITH NN BALANCE

We have found that some restriction on randomization to ensure a degree of NN balance in the experimental design is necessary, both to increase the average efficiency of treatment estimation and to reduce, concomitantly, the amount of variation between standard errors of estimated treatment differences (*SED*s). It is of interest that these two properties appear to go together. A similar effect has been noted with incomplete block designs (W. B. Hall, personal communication).

A condition of exact NN balance has been described by Williams (1952), namely that all pairs of treatments should occur as nearest neighbours equally often (usually once), including or excluding self-pairs. We term this a condition of exact first-level NN balance, as it involves only a first level of NN relationship. Second-nearest neighbour relations (level 2) and so on are also involved below. Note that exact first-level balance does not determine a single *SED* for all treatment differences. For the Rothamsted experiment in Section II.3 the 6 calculated *SED*s ranged from 0.884 to 0.919 of nominal, being affected also by higher-level NN relations.

Exact NN balance is too stringent a condition in practice for reasonably large numbers of treatments, say 15 or more, as too many replicates are needed. Here we propose conditions of *partial* NN balance. For a particular class of neighbour (row, column etc.), the condition of rth-level partial NN balance is that all the neighbouring treatments of any treatment up to the rth level should be different from that treatment and from each other; or equivalently, that any pair of treatments should occur in NN relationship at most once, including up to rth-level neighbours, with no treatment in such relationship to itself. A composite criterion extending NN relationship to include row, column and diagonal neighbours is similarly defined, with a triplet of integers such as (2, 2, 0) to specify the maximum levels of NN relationship to be considered for each class of neighbour.

A parallel between the above and classical concepts of partial balance for fixed-block methodology is readily seen, with *level* of NN relationship the counterpart of block size (e.g. level 2 ~ block size 5). There are differences in their effect on analysis, however, and the average efficiency of an NN analysis and the range of *SED*s are both also affected by the degree of departure of the data from a *smooth trend + independent error* model. This is illustrated in the next section. For 1-D NN analysis of designs with 3 or more reps we have found that second-level partial NN balance is sufficient to produce high efficiency and an acceptably narrow range of *SED*s if this model holds, likewise (2, 2, 0)-level balance fo 2-D analysis. The kind of *SED* distribution obtained with second-level balance and independent errors is illustrated in Table 5, with *SED*s classified according to NN relationship between treatments. NN-related treatments tend to have smaller *SED*s and the lower tail of the distribution for first NNs is associated with the comparisons involving a border plot treatment.

Our computer design-generator can readily produce randomized design plans with up to (2, 2, 0)-level partial NN balance including appropriate border treatments, for 2-4 reps of a

Second NN

TABLE 5

Frequency distribution of 276 SEDs (relative to nominal) for estimated treatment differences from a 1-D intra-N analysis of a 3 rep × 24 treatment experiment with second-level partial NN balance. The SEDs are classified according to the NN relationship between the treatments

Range:	0.89	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	1.00	1.01	1.02	1.03	*Total*
First NNs:	1	0	3	9	9	13	26	16	1							78
Second NNs:						1	6	17	24	15	7	2				72
Other:									1	11	33	51	24	6		126

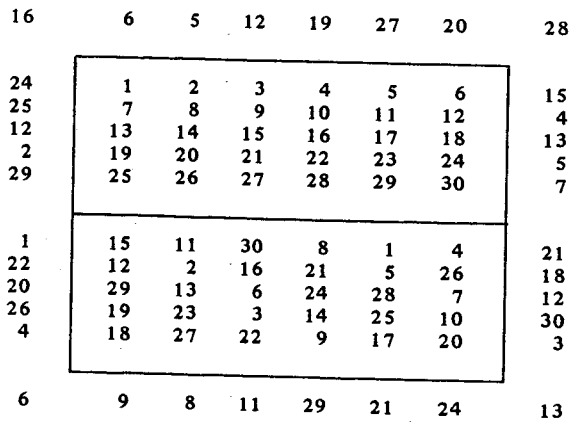


Fig. 4. Bordered design with (2, 2, 0)-level partial NN balance for 2 reps × 30 treatments.

reasonably large number of treatments. An example is given in Fig. 4 with treatments in the first rep numbered in standard order to simplify study of its properties. This level of NN balance also seems to ensure that "undesirable" randomizations do not occur. We should also note here that we have not succeeded so far in producing any design with a non-zero level of diagonal NN balance.

Border plots are essential as otherwise the accuracy of treatment comparisons involving treatments on boundary plots is greatly reduced. It is readily verified for instance that if a boundary plot is linearly detrended using its next two neighbours on the inward side, the detrended value is simply double that of the next plot inward. In other words, in the absence of border plots, the boundary plots automatically assume that role for linear 1-D detrending, and thereby are lost as replicates of the treatments concerned. For 2-D analysis, however, border plots on two of the four sides could be omitted without too much sacrifice in quality of the analysis if it were known (from shape of plot, etc.) that the total NN adjustment would be dominated by linear adjustment in one direction.

A final note concerns a secondary condition of NN balance, namely that for any pair of treatments their lists of neighbours (to specified level) should have at most one element in common. We have found that this enhances efficiency, etc. for 2-replicate designs but it is not at present included in our design generator.

II.7. VALIDITY UNDER RANDOMIZATION, EFFICIENCY

Many Monte-Carlo trials of the new analysis have been done using randomly generated sets of 100 partially balanced NN designs in conjunction with the data sets described in Part I. Table 6 is a

TABLE 6

Summary statistics for NN analysis by two methods (M) from 28 Monte Carlo randomization trials

Design Data	V_0 ($CV\%^2$)	M	EDF	$\frac{F}{E(F)}$	$\frac{RMS}{V_0}$	e (%)	$\frac{eV_0}{RMS}$	SED min	SED max
2 X 24:									
MH	27.28	1, 2*†	21.7	1.15	0.93	100.2	107.8	0.95	1.03
KH	22.94	1	15.1	0.91	1.29	100.6	77.8	0.85	1.09
		2	19.7	0.99	0.99	76.2	77.3	0.94	1.28
$W1_8$	11.23	1	15.1	0.98	1.15	100.6	87.7	0.85	1.09
		2	19.0	1.09	0.91	79.3	87.3	0.93	1.24
$W2_8$	20.39	1	15.1	0.97	1.28	100.6	78.8	0.85	1.09
		2	19.9	1.03	0.96	73.6	76.2	0.94	1.32
3 X 24:									
MH	27.89	1, 2*†	44.1	1.11	0.95	102.2	107.2	0.96	1.01
KH	29.46	1	35.7	0.93	1.10	104.0	94.2	0.90	1.03
		2	39.4	0.96	1.00	95.6	95.5	0.94	1.08
$W1_8$	16.75	1	35.7	0.90	1.36	104.0	76.7	0.90	1.03
		2	49.3	0.98	0.97	77.8	79.2	0.95	1.27
$W2_8$	17.91	1	35.7	1.02	1.14	104.0	91.0	0.90	1.03
		2	42.5	1.06	0.96	89.5	93.3	0.95	1.14
$W1_4$ + cov.	17.88	1	33.9	1.03	1.12	101.2	90.6	0.90	1.10
		2	38.2	1.04	0.99	89.1	90.0	0.95	1.20
$W2_4$ + cov.	24.63	1†	36.6	1.04	1.09	101.0	92.8	0.92	1.09
		2†	39.6	1.05	0.99	91.3	92.4	0.95	1.17
3 X 40:									
KH 1-D	30.55	1	59.3	0.92	1.14	100.3	87.9	0.91	1.04
		2	66.0	0.94	1.02	89.6	88.3	0.96	1.10
2-D	27.65	1	56.4	0.94	1.20	101.7	85.0	0.90	1.04
		2	68.1	0.92	1.12	94.1	83.8	0.95	1.08
W_8 1-D	17.85	1	59.0	0.89	1.34	99.5	74.2	0.91	1.05
		2	80.3	0.98	0.99	75.1	76.2	0.97	1.26
2-D	13.86	1	55.0	0.92	1.50	103.1	68.9	0.89	1.05
		2	71.5	1.36	0.94	66.7	70.7	0.99	1.49

* Both methods gave essentially the same results.
† Optimal b or $b_S < 0.95$.

summary of results from 28 of these tests, classified according to method of analysis (M) and *rep* X *treatment* design parameters. Method 1 assumes a *smooth trend + independent error* model. Method 2 uses correlations estimated from the data ignoring treatments, as described in Sections II.2 and II.5, for comparison with method 1. The first two groups of trials are of 1-D NN analysis with respect to column neighbours, and correspond to those of the Papadakis method in Part I except for the reduced number (24) of treatments to allow for border rows. Group 2 includes trials of covariance analysis on $W1_4$ and $W2_4$ with an indicator variable as covariate to eliminate an alternating row effect arising from the use of an 8-row seed drill; see Section I.2. The third group includes tests of 2-D analysis and for this purpose border plots were included all around in the designs. One set of 100 designs was used for all trials within each group, except for the last group where two separate sets were necessary for KH and W_8, with 8 x 5 and 4 x 10 replicate configurations respectively. The levels of NN balance were (0, 2, 0) for the 24-treatment designs and (2, 2, 0), (1, 2, 0) for the 40-treatment designs for KH and W_8.

The first statistic in Table 6 is the residual mean square V_0 from an NN covariance analysis ignoring treatments, with 1 or 3 NN covariates respectively for the 1-D or 2-D cases. Since there are no real treatment effects in the data, V_0 estimates the theoretical minimum variance of the data with respect to the specified level of NN adjustment. We therefore use it for assessment of efficiency. All other statistics in Table 6 are Monte Carlo means of 100 values, or functions thereof. The mean treatment F-ratio has been adjusted for the varying numbers of error degrees of freedom by dividing it by the expected value $EDF/(EDF-2)$ of a corresponding central F distribution. Note that the mean EDF is affected by the correlations assigned under methods 1 and 2, and hence also the mean RMS in almost the same proportion. We have therefore multiplied the computed mean efficiency factor e by V_0/RMS to give a "nett" efficiency value for treatment estimation relative to V_0. Note finally that the entries for SED (min, max) are expressed relative to the nominal SED, $(2v'/r)^{\frac{1}{2}}$.

In all cases except MH and $W2_4$ the results presented are for intra-N analysis, as we found that in these cases the optimal b or b_S value was close to 1. Table 7 gives mean optimal b values for MH and $W2_4$ and for the 2-D analyses of KH and W_8. Average standard errors for the three groups of entries in Table 6 are 4.5, 3.8 and 3.0 per cent respectively for $F/E(F)$, and 2.7, 1.5 and 1.2 per cent for RMS/V_0 and eV_0/RMS. For all other entries the SEs are of the order 0.1 per cent or less except for MH and W_{2_4} where variation in the optimal b values induced additional variation, with a maximum SE of 0.8 per cent for EDF in the case of MH(2 × 24). In Table 7 the SEs are 0.024, 0.015 for MH and 0.01 or less for all other entries.

TABLE 7

Monte Carlo means of optimal b values for method 1 *in comparison with values from NN covariance analysis ignoring treatments*

Data set	MH 2 × 24	MH 3 × 24	$W2_4$ 3 × 24	KH 3 × 40			W_8 3 × 40		
				b_1	b_2	b_3	b_1	b_2	b_3
Method 1	0.44	0.49	0.93	0.99	0.19	−0.18	0.96	0.34	−0.30
NN covariance analysis ignoring treatments	0.46	0.47	0.86	0.94	0.40	−0.38	0.97	0.66	−0.61

In discussing the results we shall restrict attention to two major points of interest.

With regard first to validity under randomization, it is clear that exact unbiassedness is not attained by either method of analysis. Nevertheless the degree of bias is usually very small. The slightly larger positive biasses (15 per cent, 11 per cent) in the F ratios for MH are reduced to 4 and 6 per cent respectively if EDF is reduced by 2 to allow for the optimization as suggested by Bartlett (1938) for the Papadakis method. The intra-N analyses of MH gave slightly lower values 0.97, 1.04 for $F/E(F)$. There is one isolated instance of appreciable bias (36 per cent), from 2-D analysis of W_8 by method 2. We have no explanation for this at present. Note that no such bias appeared when method 1 was used for 2-D analysis of W_8.

The second main point of interest is *efficiency*. Here there are some surprising effects. With method 1 the efficiency factor e for treatment estimation is typically 100 per cent or more. However except in the case of MH, for which the independent-error model holds very well, the mean RMS is larger than V_0 so that the nett efficiency relative to V_0 is somewhat less, to a varying extent. If inadequacy of the *smooth trend + independent error* model were the primary cause of the reduction in nett efficiency, one would expect greater efficiency from method 2, but although a reduction in RMS is obtained by this method, there is also a corresponding reduction in the efficiency factor e, so that nett efficiency is almost the same as for method 1. Extension to 2-D

detrending reduces the *RMS* but relative to the corresponding, reduced value V_0, nett efficiency is in fact lower than for 1-D NN adjustment. We conjecture therefore that heterogeneity in trend variation is the main factor limiting the efficiency of the analysis. Note that method 1 tends to underestimate the range of *SED*s for treatment comparisons, assuming those from method 2 to be more appropriate.

The nett efficiencies are nevertheless reasonably high, and comparable with the efficiency factors for say lattice or lattice square designs with similar-sized replicates. One can therefore expect substantial improvements in overall efficiency relative to classical methods because of reductions in *RMS* effected by the continuous local detrending in NN analysis. The ratios RMS/V_0 in Table 8 give an indication of the gains in efficiency from reduction in error variance that could be expected from NN analysis relative to various forms of classical analysis of the specified data sets. The only ratios less than one relate to the very uniform data set MH.

TABLE 8

Ratios to V_0 of residual variances within rep, row and column classifications

Design:	3 X (4 X 6)				3 X (8 X 5)	3 X (4 X 10)
Data:	MH	KH	$W1_s$	$W2_s$	KH	W_s
Classification:						
Rep	1.16	3.60	9.33	4.79	3.51	15.57
Rep/row	1.23	3.90	8.95	4.34	3.66	16.36
Rep/column	0.70	1.10	3.75	2.56	2.01	4.47
Rep/(row X column)	0.70	1.01	2.13	1.51	1.81	4.29

A simple theoretical answer can be given in terms of relative error variances for the efficiency of NN analysis relative to a standard randomized block analysis, assuming a *smooth trend + independent error* model. For 1-D analysis with border plots present the relative efficiency is given approximately by

$$E = \frac{1+\omega}{1+(1-b)\omega}\,; \quad b = \frac{\omega}{\omega+\frac{1}{2}}\,, \tag{49}$$

from the minimum variance formula (15) in Section I.3. The same formula for E holds for 2-D analysis with b_S substituted for b and with $b_S = \omega/(\omega+\nu)$ as given in Section II.5. The value of E is about 150 per cent if $\omega = 1$, rises to about 400 per cent if $\omega = 5$, and to more than 700 per cent if ω is as high as 10, as in Table 2. The gains would be much less relative to incomplete block designs with small blocks, but substantial nevertheless if trend variation were appreciable. The most competitive classical designs would be those of the Latin and lattice square type, where elimination of both row and column effects within squares does provide a continuous form of detrending. However, detrending by eliminating row and column effects alone will not be satisfactory if there is appreciable row x column interaction in the data, as when the variation is patchy for instance, and one is rarely in the position of knowing that row x column interaction is negligible.

Kempton and Howes (1981) have used data from 118 cereal trials with 6-rep balanced 5 x 5 lattice square designs, to compare Papadakis's method of NN analysis with 3 forms of standard analysis eliminating progressively rep, row and column effects. They report substantial gains in efficiency from NN analysis except in comparison with the full lattice square analysis where the relative efficiencies ranged from 64 to 170 per cent, with an average of little over 100 per cent. Other data in Fig. 1(b) of their paper show that the gain in efficiency from NN analysis is greatest

RP 100

when it is most needed, that is when variability in the data is high. Their Fig. 1(b) shows an average gain in efficiency of about 150 per cent relative to randomized block analysis when the initial *CV* is in the range 12.5–17.5 per cent.

II.8. DISCUSSION

We have presented in Part II a new and efficient method of analysing field experiments. We see it as a logical development in classical Rothamsted methodology, an evolution from step-wise block detrending or marginal row and column detrending to continuous local detrending in 1 or 2 dimensions. We also suggest that this extension might have resolved the controversy between "Student" and Fisher on semi-systematic versus randomized designs. For an essay on this topic and relevant references see Wilkinson and Mayo (1982). Whereas Papadakis's method has been considered merely as an adjunct to classical methods (Bartlett, 1978), we are here envisaging the new NN methodology as a replacement for classical methods in many areas of application, particularly for plant breeding experiments such as varietal trials.

Our justification of the method is largely pragmatic – it works, in the empirical studies done so far. Much remains to be done in developing more theoretical knowledge of its properties and possible extensions of it. As we have mentioned, a joint study of the randomization theory for the method is being undertaken with A. W. Davis. We know that the method is approximately, but not exactly unbiassed under randomization, and therefore further empirical tests would be needed to verify approximate unbiassedness in different areas of application.

The method assumes a *smooth trend + independent error* model. We do not have a satisfactory way of estimating correlation structure in the presence of treatment effects, though our method 2 is available as a check method for studies with uniformity data.

Our work suggests that substantial gains in accuracy of treatment estimation are to be expected with the new method. Practical trials of the methodology have been carried out at the Waite Agricultural Research Institute by Dr D. H. B. Sparrow, the Senior Barley Breeder, with 12 experiments involving 3 sets of 90 lines at 4 sites each and sown with partially NN-balanced designs. The reductions in residual variance from NN analysis, relative to randomized block analysis ranged up to a maximum of 50.0 per cent, the highest reductions being associated with the most variable sites. Consequently nearly all multiple-replicate cereal breeding trials at the Waite Institute have been sown this year with partially balanced NN designs, and further trials with the new method are being conducted by the New South Wales Department of Agriculture.

Improvement in accuracy of treatment estimation has a critical effect on the selection of the top-yielding lines in plant breeding experiments. Using one of the above barley trials as an illustration, we found from comparison of yield estimates from randomized block and NN analyses that the estimated yields of the top 4 lines were reduced from 29 to 38 per cent above grand mean to 21–24 per cent, and more importantly, there was a 30 per cent misclassification of the top 10 by the randomized block analysis, with 3 of the top 10 estimated yields being reduced in rank to eighteenth, twenty-second and thirty-sixth under NN analysis. The continuous local detrending has the effect of preventing any treatment from achieving a spuriously high rank by virtue of a lucky assignment to high-yielding patches in the field.

One of the important advantages of the new methodology is flexibility in design for reasonably large numbers of treatments. Our computer design generator can readily achieve second-level partial NN balance for almost any reasonable rectangular configuration of replicates in the field. This flexibility is of particular importance in experiments where the use of multi-plot seed drills may largely dictate the dimensional parameters of the design. So far we have not turned our attention to NN-balanced designs for small numbers of treatments but other research is being done in this area. See, for example, Freeman (1979).

With regard to factorial experiments with a reasonably large total number of treatments the approach we intend examining is to analyse the data with respect to a single composite treatment factor and then subject the estimated treatment effects to a factorial analysis. If the range of *SE*s

of treatments estimates is small, the factorial analysis will be very nearly orthogonal. With stratified experiments of the split-plot type we imagine including higher-stratum factors such as main plots provisionally as additional treatment factors, and proceeding as above for factorial experiments, with appropriate higher-stratum residual variances determined from the factorial analysis for calculating F tests and standard errors for the treatment effects estimated in those strata.

Perhaps the most fundamental issue that will be raised in discussion of the paper is the role of randomization. Some might argue that sufficient refinement of the NN methodology would render randomization unnecessary but we would not support this view. Randomization provides a protection against spurious effects in our inferences.

But certainly the protection is more limited than is commonly realized. A fundamental principle of uncertain inference due to Fisher (see Wilkinson, 1977) may be expressed in inverted form as follows: one can arrive at spurious or misleading inferences by failing to take into proper account *all* the relevant evidence. Thus approximate unbiassedness under randomization is not a sufficient requirement of validity, the analysis must also be as efficient as possible. It was a defective form of detrending in the Rothamsted analysis described in Section II.3, for instance, that led to spuriously significant estimates of neighbour treatment effects, and it could be argued that many past analyses of field experiments by classical methods are scientifically invalid because of inadequate detrending. Certainly the supposed equality of standard errors for many orthogonal or balanced designs may be only a mathematical illusion, for under NN analysis very different standard errors arise unless the experimental design has a high degree of NN balance.

ACKNOWLEDGEMENTS

The authors are very grateful to M. S. Bartlett, A. G. Constantine, A. W. Davis, G. V. Dyke, W. B. Hall, A. T. James, R. J. Martin, J. A. Nelder, H. D. Patterson and S. C. Pearce for valuable discussions and correspondence regarding the paper, to Helen Simpson and Karen Kain for meticulous typing of the drafts, and to Ruth Hughes for the excellent artwork in Figs 1, 2 and 3. Dr A. J. Lawrance is particularly thanked for his patient handling of the paper during the refereeing, and likewise the referees, whose comments led to significant improvements in presentation. We are particularly indebted to H. D. Patterson for his special interest in the work during a visit to Australia, and to A. W. Davis for his current work on the randomization theory.

REFERENCES

Atkinson, A. C. (1969) The use of residuals as a concomitant variable. *Biometrika*, **56**, 33–41.

Bartlett, M. S. (1938) The approximate recovery of information from field experiments with large blocks. *J. agric. Sci.*, **28**, 418–427.

——— (1978) Nearest neighbour models in the analysis of field experiments (with Discussion). *J. R. Statist. Soc.* B, **40**, 147–174.

——— (1981) A further note on the use of neighbouring plot values in the analysis of field experiments. *J. R. Statist. Soc.* B, **43**, 100–102.

Besag, J. (1977) Errors-in-variables estimation for Gaussian lattice schemes. *J. R. Statist. Soc.*, B, **39**, 73–78.

——— (1978) Contribution to Discussion of Bartlett (1978). *J. R. Statist. Soc.*, B, **40**, 165–166.

Draper, N. R. and Guttman, I. (1980) Incorporating overlap effects from neighbouring units into response surface models. *Appl. Statist.*, **29**, 128–134.

Dyke, G. V. and Shelley, C. F. (1976) Serial designs balanced for effects of neighbours on both sides. *J. agric. Sci.*, **87**, 303–305.

Freeman, G. H. (1979) Some two-dimensional designs balanced for nearest neighours. *J. R. Statist. Soc.* B, **41**, 88–95.

Hemmerle, W. J. (1967) *Statistical Computations on a Digital Computer*. Waltham, Massachusetts: Blaisdell.

Jenkyn, J. F., Bainbridge, A., Dyke, G. V. and Todd, A. D. (1979) An investigation into inter-plot interactions, in experiments with mildew on barley, using balanced designs. *Ann. appl. Biol.*, **92**, 11–28.

Kempton, R. A. and Howes, C. W. (1981) The use of neighbouring plot values in the analysis of variety trials. *Appl. Statist.*, **30**, 59–70.

Martin, R. J. (1978) Contribution to Discussion of Bartlett (1978). *J. R. Statist. Soc.* B, **40**, 163–165.

Mercer, W. B. and Hall, A. D. (1911) The experimental error of field trials. *J. agric. Sci.*, **4**, 107–132.

Papadakis, J. S. (1937) Méthode statistique pour des expériences sur champ. *Bull. Inst. Amél. Plantes á Salonique*, **23**.
Pearce, S. C. (1980) Randomized blocks and some alternatives: a study in tropical conditions. *Trop. Agric.*, **57**, 1–10.
Pearce, S. C. and Moore, C. S. (1976) Reduction of experimental error in perennial crops, using adjustment by neighbouring plots. *Exp. Agric.*, **12**, 267–272.
Ripley, B.D. (1978) Contribution to Discussion of Bartlett (1978). *J. R. Statist. Soc.*, B, **40**, 162–163.
Wiebe, G. A. (1935) Variation and correlation in grain yield among 1500 wheat nursery plots. *J. agric. Res.*, **50**, 331–357.
Wilkinson, G. N. (1977) On resolving the controversy in statistical inference. *J. R. Statist. Soc.*, B, **39**, 119–171.
Wilkinson, G. N. and Mayo, O. (1982) Control of variability in field trials: an essay on the controversy between "Student" and Fisher, and a resolution of it. *Utilitas. Math.*, **21B**, 169–188.
Williams, R. M. (1952) Experimental designs for serially correlated observations. *Biometrika*, **39**, 151–167.
Wynn, H. P. (1978) Contribution to Discussion of Bartlett (1978). *J. R. Statist. Soc.* B, **40**, 161.
Yates, F. (1970) *Experimental Design*. London: Griffin.

DISCUSSION OF THE PAPER BY WILKINSON *ET AL.*

Dr H. D. Patterson* (ARC Unit of Statistics, Edinburgh): Today's authors are to be congratulated on their bold, innovative and important paper. I have no doubt that the subject is important. Neighbour analysis has more potential to improve the accuracy of field plot experiments than, for example, the theory of optimal block design. I speak as a practitioner with some responsibility for designing trials of new varieties of agricultural crops and a healthy respect for Dr Yates's lattice and other incomplete block designs. Most of the trials I deal with nowadays have long narrow plots so I restrict my remarks to one-dimensional design and analysis.

My first comment about the paper itself concerns the standards against which the new methods are to be judged. Following tradition the authors calculate efficiency relative to complete blocks. Would it not be fairer, particularly if the number of varieties is large, to make the comparison against the best available incomplete block designs? Thus, in Section II.8 the authors refer to 12 barley experiments each with 90 varieties, and quote a reduction in residual variance from NN analysis, relative to randomized block analysis, of up to 50 per cent. What would the reduction have been relative to an analysis of incomplete blocks of say 9 or 10 plots? In the UK small blocks can be very effective. My colleague Mr E. A. Hunter has calculated, for example, that in recent years generalized lattices in National List and Recommended List trials of new cereal varieties have reduced variances of varietal differences by an average of 30 per cent. This result is based on 244 trials.

Nevertheless, conventional block analysis does ignore some information on positional effects even when blocks are small and neighbour analysis will usually, if not always, effect an improvement. That brings me to my second question. Does the NN method put forward this evening use the additional information efficiently? Here I regret to say I have some reservations. It seems to me that the authors have allowed inefficiency to creep into their analysis.

Firstly, border plots are used solely to allow second differences to be calculated on end plots; the direct information they provide on varietal differences is ignored. In my view this is wasteful and will result in net loss of information per plot rather than the intended gain.

Secondly, but more seriously, the use of unweighted variety totals of detrended yields in equation (21) of the paper leads to loss of precision when errors in the original values contain random components. One can appreciate the authors' anxiety not to lose or appear to lose the advantages of detrending. But the simple fact is that in using equation (21) they have departed from well-established least squares principles and should not be surprised if some inefficiency ensues.

So I come to my third point. I agree with the authors' objective of improving on the Papadakis method by including a white noise component in the model, if only to allow for errors in measurement of plot dimensions, amount of seed, yield and so on. But why is it not possible to achieve the objective by conventional least squares?

With this question in mind Dr E. R. Williams of CSIRO Canberra and I have been looking at least squares analysis of several autocovariance models. The simplest model has a white noise

* As presented to the meeting in proposing the vote of thanks.

component with variance σ_1^2 and a component with variance σ_2^2 and correlation ρ^i between plots in the same replicate and distance i plots apart. The overall correlations are $\lambda\rho, \lambda\rho^2, \ldots$, where $\lambda = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$. We call this the exponential variance or EV model. It fits the UK cereal data well. The average value of λ in 167 recent trials is 0.71, i.e. about 30 per cent of the total plot error variance is associated with the white noise component. The average value of ρ is 0.94.

Given empirical values of λ and ρ variety means can be estimated by generalized least squares. Thus, if $\mathbf{y}_i$ is the vector of yields and $\mathbf{X}_i$ the design submatrix for replicate i, the generalized least squares estimate of τ, the vector of variety means, is given by $\hat{\tau}$, where

$$\sum_i (\mathbf{X}_i^T \hat{\mathbf{V}}^{-1} \mathbf{y}_i) = \sum_i (\mathbf{X}_i^T \hat{\mathbf{V}}^{-1} \mathbf{X}_i)\, \hat{\tau} \tag{1}$$

and $\hat{\mathbf{V}}$ estimates the autocovariance matrix. The latter can be written in the form

$$\mathbf{V} = \sigma_1^2 \mathbf{I} + \sigma_2^2 (1 - \rho^2)\, \Delta_{(\rho)}^{-1}/(1 + \rho^2), \tag{2}$$

where

$$\Delta_{(\rho)} = \frac{1}{2} \begin{pmatrix} \frac{b}{\rho} & -b & 0 & 0 & \ldots & 0 \\ -b & 2 & -b & 0 & \ldots & 0 \\ 0 & -b & 2 & -b & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \frac{b}{\rho} \end{pmatrix}. \tag{3}$$

and $b = 2\rho/(1 + \rho^2)$.

Except at the end plots the elements of $\Delta_{(\rho)}\, \mathbf{y}_i$ are the partially detrended yields used by Dr Wilkinson and his colleagues and equation (21) of the paper can be written

$$\sum_i (\mathbf{X}_i^T\, \Delta_{(\rho)}\, \mathbf{y}_i) = \sum_i (\mathbf{X}_i^T\, \Sigma_{(\rho)}\, \mathbf{X}_i)\, \hat{\tau}_{NN}, \tag{4}$$

where $\hat{\tau}_{NN}$ is the NN estimate of τ.

Comparison of (1) and (4) shows that the authors' NN analysis is a special case of the generalized least squares analysis with σ_1^2 in the EV model set to zero or, equivalently, λ given unit value. Thus, for given ρ and ignoring differences in the treatment of end plots, NN analysis give exactly the same results as an analysis based on a first-order autoregressive error model. The benefits the authors expected to obtain by including a white noise component have not materialized.

By contrast, equation (1) gives improved accuracy when σ_1^2 is not zero. Those who insist on including a deterministic trend component in the error model can obtain exactly the same answers by least squares analysis of second differences $\Delta_{(1)}\, \mathbf{y}_i$ instead of yields $\mathbf{y}_i$. Over a range of values of ρ from 0.7 upwards $\Delta_{(\rho)}$ is well approximated by $\Delta_{(1)}$ so that, on the EV model, we can write

$$\mathrm{Var}\,(\mathbf{y}_i) = \sigma_1^2 \mathbf{I} + \sigma_2^2 (1 - \rho^2)\, \Delta_{(1)}^{-1}/(1 + \rho^2) \tag{5}$$

and so avoid the complication of partial detrending.

Whatever method is used we must be sure that valid estimates of error are available. It is reassuring to hear that estimates of error are nearly unbiased in the Papadakis and NN methods but, as Dr Wilkinson and his colleagues show in their results for the iterated Papadakis method, things can go badly wrong. A randomization theory for NN methods as elegant as that for block designs is unlikely to emerge but I for one will be content with "approximate" validity. In real life that is all we have anyway.

Finally, to try to put the new methods into perspective, I have calculated theoretical variances of varietal differences under the empirical exponential variance rule for two hypothetical three-

replicate trials. Trial A has 16 varieties and trial B has 48. The average variances $(t/h)^2$ are as follows:

	Trial A	*Trial B*
Randomized block analysis	0.067	0.094
Incomplete block analysis	0.056	0.061
EV neighbour analysis	0.053	0.055

The results show that the main benefits come from incomplete blocks with approximately optimal block size (4 plots per block for trial A and 6 plots per block for trial B). Nevertheless, the additional improvement resulting from neighbour analysis by generalized least squares is worth having and not otherwise available. The fact that it is now theoretically possible to obtain almost as accurate results in a trial with 48 varieties as in a trial with 16 varieties is noteworthy. Ideally, we would like the accuracy with which varietal differences are estimated to be independent of the number of varieties in trial. In line with the authors' observations the range of variances of individual differences narrows with decreasing average variance.

I believe that much of the improvement of which neighbour analysis is capable has already been realized by Dr Wilkinson and his colleagues. Work remains to be done but they have shown the way. Less adventurous spirits can now follow. I have much pleasure in proposing the vote of thanks.

Mr J. E. Besag (University of Durham): The first version of tonight's paper appeared some two years ago; since then it has survived and evolved through various stages of development and has even gained an additional author. The eventual result is an important contribution to the design and analysis of field experiments, especially varietal trials.

In practice, one-dimensional adjustment will often be adequate and, like the authors, I shall concentrate primarily on that case. First I want to consider the basic estimation equations (21) and (23). These seem rather to be plucked from thin air—even more so in the original version of the paper. In any case, what befalls the ξ_i's? When $b \neq 1$, they do not seem to be eliminated, unless an appeal to randomization is made, in which case the adjacency matrix **A** suffers the same fate; and, so far as I can see, randomization conditional on **A** removes the ξ_i's only for rather special designs.

As an alternative approach, one can consider a simple conditional expectation formulation, of which the authors will surely disapprove but which also aims at local detrending. I omit complications such as the inclusion of errors in variables and conventional trend terms which might be incorporated in practice; see the Discussion of Professor Bartlett's 1978 paper. However, in common with the authors, I include border plots used solely for NN adjustment: this is appealing in principle but would normally be considered wasteful. Thus, assume an additive model,

$$\mathbf{Y} = \mathbf{D}\tau + \mathbf{Z},$$

where $\mathbf{Y}$ denotes the random vector of all plot yields, τ contains the p treatment effects, $\mathbf{D}$ is the corresponding design matrix and $\mathbf{Z}$ represents plot effects, measured about zero without any loss of generality. Further assume that, at each of the n internal plots,

$$E(Z_i \mid all\ z_j, j \neq i) = b\bar{z}_{Ni}, \quad \text{var}\,(Z_i \mid \ldots) = \kappa;$$

at the border plots, assume merely that the marginal expectation of Z_i is zero. For the moment, treat b as fixed.

Unfortunately, the dispersion matrix $\mathbf{V}$ of $\mathbf{Y}$ cannot be identified because of the lack of border assumptions but it is easily shown that $\mathbf{GVG} = \kappa\mathbf{G}$, where $\mathbf{G}$ is the matrix whose (i, j) element is

$$G_{ij} = \begin{cases} 1 & j = i, \quad i \text{ internal}, \\ -\frac{1}{2}\,b & j \in Ni, \quad i \text{ internal}, \\ 0 & \text{otherwise}. \end{cases}$$

Thus, $\kappa^{-1}\mathbf{G}$ is a generalized inverse of $\mathbf{V}$, and in practice will be effectively close to the full inverse,

though not even symmetric. Adopting a generalized least squares approach then suggests that the estimate $\hat{\tau}$ of τ should satisfy

$$\mathbf{D}^{\mathrm{T}}\mathbf{G}\mathbf{D}\,\hat{\tau} = \mathbf{D}^{\mathrm{T}}\mathbf{G}\mathbf{y}$$

or, in the authors' notation,

$$(r\mathbf{I} - \tfrac{1}{2}\,b\mathbf{A})\,\hat{\tau} = \mathbf{T}',$$

which is exactly their equation (23)! This equation is soluble for $b < 1$ and the inconsistency when $b = 1$ does not affect estimates of treatment contrasts—formally $b \uparrow 1$ is again equivalent to equation (21). Thus, for fixed b, the estimates coincide with those of the authors and indeed that is equally true of the iterated Papadakis method with border-plot adjustment; in brief, there is nothing new in equations (21) and (23). These remarks generalize immediately to more complex adjustment in one or two dimensions. Note that, with the conditional expectation assumptions, $\hat{\tau}$ is unbiased for τ and that an unbiased estimate of κ is available, since if we define residuals $\hat{\mathbf{Z}} = \mathbf{Y} - \mathbf{D}\,\hat{\tau}$ then $E(\hat{\mathbf{Z}}^{\mathrm{T}}\mathbf{G}\hat{\mathbf{Z}}) = (n - p)\,\kappa$.

The real differences between the iterated Papadakis, the conditional expectation and the authors' methods arise in the choice or estimate of b, in the assignment of standard errors for treatment contrasts and in the rôle of design, including, in practice, the exploitation of border plots. In the iterated Papadakis method, where b is estimated by repeated covariance analysis, the formation of standard errors seems unsatisfactory, as emphasized by the authors tonight. In the conditional expectation formulation, the natural step is to invoke a Gaussian assumption, which leads to an auto-Normal model, and to estimate b by maximum likelihood. One then has the problem of model-based standard errors, when one knows that at best the fertility model is crude. In tonight's paper, there is a new approach in which one either takes $b = 1$ or chooses b optimally according to a "*smooth trend + independent errors*" model, a model which I for one find appealing.

As regards the yields of border plots, the authors' use of them merely to adjust for edge effects in equations (21) and (23) seems inefficient. It is also unclear how such an approach could be adapted to cope with border complications: for example, a recent practice at Rothamsted has been to leave every third plot untreated and to use it exclusively as a spraying run. Note also that for full two-dimensional adjustment, borders can easily increase the size of the experiment by 50 per cent, say. For the Papadakis and auto-Normal methods, in contrast to the mathematical development earlier in this discussion, it is usual to make convenient and not unreasonable assumptions at the boundary and to include the corresponding yields fully in the equations for treatment estimates. This should lead to an increase in accuracy, though standard errors may require special attention.

As regards design, the authors stipulate "*partial* NN *balance*". However, if there is freedom to design the experiment with spatial analysis in mind, then all sorts of possibilities are opened up. For example, one might monitor the pattern of fertility by the inclusion of occasional untreated check plots. And one must also make comparisons with the best classical designs involving the same total number of plots, not with randomized blocks or to the exclusion of available border plots. Tonight's paper is not in the same spirit as Professor Bartlett's, where nearest-neighbour adjustment was viewed merely as an "*ancillary exploratory device*", without particular demands on design.

I should like next to summarize some comparisons of my own, based on the same data sets as used by the authors; see Table D1. Twenty randomized layouts were produced for each set of yields; unfortunately, in the absence of a sophisticated design generator, the authors' balance requirement had to be relaxed to one in which no treatment appeared next to itself nor next to any other more than *twice*, rather than once. The methods of analysis were randomized blocks (*RB*), uniterated Papadakis (*Pap*), iterated Papadakis (*I Pap*), Wilkinson *et al.* with $b = 1$ (*Intra*) and optimized (*Extra*), auto-Normal (*AN*) and a simple version of errors in variables (*E in V*) with interaction parameter $b = 1$. In the Papadakis and Gaussian analyses, border plots were fully utilized in the equations for treatment estimates; note that, for the *New MH* data, the border plots provided a complete fifth replicate. The programs were written in APL.

TABLE D1
Details of data sets used

Data	*Overall dimensions*	*Replicates*	*Replicate dimensions*	*Border plots*
$W1_s$	14 × 6	3	4 × 6	2 × 6
$W2_s$	14 × 6	3	4 × 6	2 × 6
New MH	10 × 12	4	4 × 6	2 × 12
New KH	23 × 7	3	7 × 7	2 × 7

For each data set and method of analysis, the averages over layouts of two quantities *Emp* (for Empirical) and *Pre* (for Predicted) are shown in Table D2, where

$$Emp = \frac{1}{p(p-1)} \sum_{1}^{p} \sum_{1}^{p} (\hat{\tau}_k - \hat{\tau}_l)^2,$$

$$Pre = \frac{1}{p(p-1)} \sum_{1}^{p} \sum_{1}^{p} \text{estimated var } (\hat{\tau}_k - \hat{\tau}_l).$$

Clearly, since the treatment effects are null, small values of *Emp* reflect high accuracy in the estimation of pairwise treatment comparisons; *Pre* is the mean squared standard error of such a contrast, as predicted by the particular type of analysis. In a randomization framework, the average values of *Emp* and *Pre* are required to be approximately equal. For each data set, the figures have been scaled so that the average *Emp* for randomized blocks is 100. I know that some will bemoan the continued use of uniformity data with dummy treatments but at least this is a rigorous approach under the assumption of treatment additivity, since all the methods of analysis also act additively.

TABLE D2
Average values of Emp and Pre over 20 layouts

		R B	*Pap*	*I Pap*	*Intra*	*Extra*	*A N*	*E in V*
$W1_s$	*Emp*	100	26	16	14	15	13	13
	Pre	107	30	14	15	16	21	21
$W2_s$	*Emp*	100	34	22	23	26	20	21
	Pre	88	36	18	24	25	23	28
New MH	*Emp*	100	50*	48*	67	66	47*	44*
	Pre	100	51*	43*	67	63	41*	43*
New KH	*Emp*	100	33	22	23	23	22	22
	Pre	102	34	15	21	22	22	21

The general pattern of Papadakis results in Table D2 accords with the authors' conclusions concerning bias; also there is impressive agreement between the average *Emp* and *Pre* values for *Intra* and *Extra*. On the other hand, it is clear, in particular from the *New MH* results, that, when border plots provide a substantial proportion of the layout, *Intra* and *Extra* can be relatively inefficient in treatment estimation, especially since the starred values in fact refer to completely randomized internal blocks. This inefficiency is accentuated by the *Emp*'s for *AN* and *E in V* but equally it is clear that here the randomization means of *Emp* and *Pre* may disagree substantially: the fact that likelihood methods make no claim to address randomization distributions is not entirely consoling. As a whole, Table D2 suggests that spatial analysis is potentially very useful but that more work needs to be done. In some earlier studies, partly with Dr Richard Martin, several methods of spatial analysis, not including the authors', were examined on uniformity data

with dummy treatments in six replicates, each 5 × 5. Even with randomized blocks layouts, the likelihood-based methods slightly out-performed balanced lattice square design and analysis. Incidentally, it is instructive to reformulate the balanced lattice square model in terms of conditional expectations.

In conclusion, I see an increasingly important role for spatial methods in the analysis of field experiments, especially when sophisticated classical designs are unavailable or have not been applied. Despite tonight's claims, the eventual choice of methods still remains open and no doubt we shall hear of further contenders—even this evening. It gives me great pleasure to second the vote of thanks to the authors for their innovative and thought-provoking paper.

The vote of thanks was passed by acclamation.

Professor S. C. Pearce (University of Kent at Canterbury): I do not propose to make any contribution to the theoretical aspects of this paper, my interests being more empirical. Some 40 years ago I received a copy of Papadakis's original paper and I have been using his method on and off ever since. Also, I have recently made a study of some 89 bodies of data with a view to finding which method of local control is most generally effective in reducing the error variance (Pearce, 1978, 1980).

My general experience is as follows: If there is a clear fertility pattern and the blocks are well chosen, the error variance will be reduced, but Papadakis's method will be as good. If the blocks are badly chosen they will be ineffective, but neighbouring plots will still work. (Of course, there are experiments with so much random variation that nothing works, but they do not affect the argument.) Clearly the nearest neighbour approach is to be preferred if someone can give us a form that is soundly based. Here I suggest that there is no need immediately to study the most complicated cases. For example, why examine blocks and adjustments by neighbours in conjunction? If the blocks are effective the adjustments are not going to remove the same variation twice. (Also, it would be disastrous if they did.) On the other hand, if the blocks are badly chosen, the trends will be encapsulated within them and it is the trends that are the source of much of the difficulty. Eventually we are going to have to choose between the two approaches, not use them together.

Also, the one-dimensional case is not just a step to a desired two-dimensional solution but is for practical purposes the more important. For example, one unsolved problem concerns the designing of an experiment on a terrace that snakes its way round the contour of a hill. A one-dimensional nearest neighbour method could be invaluable. Again, it appears to be agreed that these methods will mostly be used with long, narrow plots placed side by side. The two-dimensional case will present problems of its own. In the meantime the world is waiting for a solution of the one-dimensional.

Professor M. S. Bartlett: May I first add my congratulations to the authors for a very stimulating paper. As this raises many important issues, the discussion could become over-fragmented, so I will begin by referring to some of the *general* criteria which are relevant to field experimental design and analysis. These general criteria include: (1) Validity. (2) Efficiency. (3) Robustness. (4) Relevance. I will also add a fifth, which should precede the others, for without it we could not even begin, viz. (0) Intelligibility.

If we list next the various broad categories of design and analysis under consideration, these would include the following:

(a) Systematic designs with no randomization;
(b) Standard designs with randomization ((i) simple designs, (ii) more complex designs)
(c) Standard designs, plus covariance on position (or on control plots);
(d) Nearest-neighbour adjustments using treated plots ((i) Papadakis, (ii) iterated Papadakis, (iii) analysis of local fertility variation, with some treatment balance—Wilkinson *et al.*)

In my own notes on this paper I could not resist assigning scores to the various categories for all the different criteria, and some of the audience may like to do likewise. My point here, however, is that *all* the criteria must be considered; some categories score better on some of the criteria, but worse on others. Ultimately it must be a matter of judgement and experience which procedure is optimal, depending on the relative importance given to each criterion.

Note that (1) and (2) can only be appraised precisely for a specific model, (3) for two or more alternative models, and (4) for the appropriateness for the data in hand of the various models

consistent with the analysis. Fisher was prepared (especially with his randomization principle) to sacrifice some efficiency for validity and robustness; but, as Dr Wilkinson has reminded us, he thereby came into conflict with "Student".

In the paper it is claimed that (d) (i) (Papadakis) is approximately valid but not very efficient, and that (d) (ii) (iterated Papadakis) is more efficient but less valid. It is further claimed that the new nearest-neighbour methods (d) (iii) are approximately valid and even more efficient. I must admit that the criticisms of (d) (ii) on the grounds of invalidity (presence of bias) and lack of robustness were rather unexpected, as underlying my view of the Papadakis method as a useful ancillary device was the tacit assumption that randomization rendered it reasonably foolproof. This appears to be not necessarily so, though I would note that Dr Wilkinson's criticism seems to apply in principle to his own recommended methods (cf. his remark on the projected study with A. W. Davis). I am indeed not always clear on the extent of the difference between his procedures and iterated Papadakis, which should, when convergent, be equivalent to maximum likelihood estimation on the nearest-neighbour models I considered. (Non-linear estimation of many parameters can cause problems, but these should be largely avoidable in the present context, as the non-linearity only arises from the covariance coefficients when unknown.) Dr Wilkinson advocates a different model, of fertility trend plus independent error; but, if the analysis is sensitive even with randomization to the model assumed, it is not satisfying the criterion of robustness, and its value would be that much diminished. Moreover, for his model the procedure (c) with covariance or position (i.e. fitting a trend) would seem quite a valid and logical method, but somewhat dismissed (contrast the remarks in Wilkinson and Mayo, 1982, with the findings of Pearce, 1980).

To sum up, the authors have unearthed some snags with the use of iterated Papadakis, but the degree of validity and robustness of their own methods still seems to need further investigation, as the authors themselves accept, before their status in comparison with standard methods can be comprehensively assessed.

Finally, on one or two points of detail:

(i) The recommendation for "balanced designs" is sensible, if any kind of nearest-neighbour adjustments are to be made.

(ii) I am indebted to Dr Wilkinson (and Dr R. J. Martin) for pointing out the theoretical slip in my 1978 paper.

(iii) I am not clear how far the bias with iterated Papadakis would still be present if the nearest-neighbour models I assumed were correct.

(iv) In Bartlett (1981) I pointed out that with "two-sided" nearest-neighbour adjustments (even with no treatment adjustments) the variance of a sum of squares was greater than classical theory would suggest. Wilkinson *et al.* have emphasized that with their methods the effective number of degrees of freedom for calculating the "error mean square" is less than the nominal number, but they imply that, after correction for this, classical sampling theory is approximately valid. This would seem to need further checking.

Dr G. H. Freeman (National Vegetable Research Station): If this contribution is a little incoherent it is because I was just sitting down to write it when I was consulted by a colleague who wished to analyse some data showing a trend across his experimental plots. Thus, as far as I am concerned, this is a problem of considerable practical importance and I therefore welcome the present paper. The general principle of checking the validity of the Papadakis approach to NN analysis is, in my view, extremely useful since I have used this approach for some time not because I want to but because it is vitally important to reduce extraneous sources of error in field experiments.

There are basically three uses for NN analysis: (1) the elimination of fertility trends, (2) consideration of unequal competition, (3) patchy occurrence of pests and diseases. The first and third may reasonably have smooth trend + independent error models, but the second is more doubtful. Most analyses have concentrated on fertility trends, but the patchy occurrence is also important—indeed the example I referred to above is of a pest whose appearance in the field is far from uniform. Papadakis' method used in situations of this sort with small plots has considerably reduced the variability. Two-dimensional analysis often shows b_1 near 1 and b_2 near 0.

Restriction to randomized complete block designs may be desirable to simplify the theory, but the same sort of approach may work with incomplete block designs or row and column

designs. A possible design which combines a lattice square with (1, 1, 0)-level balance, in the notation of the present paper, is as follows:

9	3	17	1	16	22	4	6	5	15	10	1	21	16	14	1	3
18	1	2	3	4	5	20	24	1	8	12	19	25	1	7	13	17
8	6	7	8	9	10	22	11	10	19	3	23	11	9	20	2	22
13	11	12	13	14	15	6	18	14	2	25	6	17	15	3	24	7
12	16	17	18	19	20	13	5	17	21	9	12	4	18	21	10	6
25	21	22	23	24	25	4	7	23	15	16	5	8	22	14	16	1
7	1	2	21	4	8	19	24	1	11	13	19	25	1	7	4	25
10	19	20	23	1	22	16	23	3	8	20	2	12	9	6	21	11
23	1	4	2	5	3	18	9	22	15	1	17	10	23	14	1	22
15	11	14	12	15	13	10	2	16	24	13	9	3	16	25	12	16
23	21	24	22	25	23	4	11	8	17	25	5	11	7	18	24	5
24	6	9	7	10	8	21	20	14	3	7	21	19	15	2	8	6
18	16	19	17	20	18	12	23	5	6	19	13	22	4	6	20	1
14	20	5	13	12	1	20	2	24	14	4	16	5	24	10	16	15

This design is exact, in the sense that each variety has *all* other varieties as neighbours in rows or columns once and once only. This is the only case where such a scheme would be likely to work, since with r replicates of v variates it needs $4r = n(v-1)$, if there are n concurrences. Thus $r = 6$, $v = 25$ leads to the required design which is, further, two bordered lattice squares with 3 replicates each.

Most work with small numbers of treatments in two dimensions uses complete or quasi-complete Latin squares, so also has (1, 1, 0) balance. It is possible, at least for a 5 × 5 quasi-complete square to obtain overall second-order balance by a careful choice of square, appropriately bordered. Thus, using the sequence 3 (0 1 4 2 3) 0 for rows and 1 (0 2 3 4 1) 0 for columns one obtains the following design:

	1	2	0	3	4	
3	0	1	4	2	3	0
0	2	3	1	4	0	2
1	3	4	2	0	1	3
2	4	0	3	1	2	4
4	1	2	0	3	4	1
	0	1	4	2	3	

In rows and columns combined each treatment thus has each other as first neighbour five times and as second neighbour four times.

I should like to end by thanking the authors for their contribution to this problem and by hoping that they will make their computer program for their method of analysis widely available – preferably in Genstat. I need it, if possible by the end of this week.

Dr A. M. Herzberg and Dr H. P. Wynn (Imperial College of Science and Technology): In a recent paper (Herzberg and Wynn, 1982), we have given a general approach to Yates' inter-block and intra-block analysis. Consider a partitioned regression model of the form

$$E(\mathbf{y}) = \mathbf{X}_1 \boldsymbol{\theta}_1 + \mathbf{X}_2 \boldsymbol{\theta}_2, \qquad (1)$$

where $\mathbf{Y}$ is an $n \times 1$ vector of observations, $\mathbf{X}_1$ and $\mathbf{X}_2$ are matrices of explanatory variables of

size $n \times p_1$ and $n \times p_2$, respectively, and $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are vectors of unknown parameters of sizes $p_1 \times 1$ and $p_2 \times 1$, respectively. A special case is when $\boldsymbol{\theta}_1$ is the vector of treatment parameters and $\boldsymbol{\theta}_2$ is the vector of plot parameters, $\mathbf{X}_1$ is the incidence matrix of plots times treatments and $\mathbf{X}_2$ is the incidence matrix of the nearest neighbour plots.

Herzberg and Wynn (1982) have shown that a necessary and sufficient condition that an inter-block and intra-block analysis can be obtained for the treatment balanced case is that

$$\mathbf{X}_1^{\mathrm{T}}(\mathbf{X}_2\mathbf{X}_2^{\mathrm{T}})^r\,\mathbf{X}_1 = a_r\mathbf{I} + b_r\mathbf{J}\ (r = 1, 2, \ldots). \tag{2}$$

The authors of the present paper have encouraged the reader to consider the nearest neighbour analysis as a moving average problem over plots and the use of an inter-block and intra-block analysis. This problem can be put into the framework just described.

Let the blocks of any plot be its north, south, east and west nearest-neighbours. Then $\mathbf{X}_2$ will be an $n \times n$ matrix with 0's and 1's, n being the number of plots. For example, consider the square array of 16 plots in Fig. 1 with the plots numbered as shown.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Fig.1. A square array of 16 plots

Then

$\mathbf{X}_2 =$

0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0
1	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0
0	1	0	1	0	0	1	0	0	0	0	0	0	0	1	0
1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1
1	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0
0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0
0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0
0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0
0	0	0	0	1	0	0	0	0	1	0	1	1	0	0	0
0	0	0	0	0	1	0	0	1	0	1	0	0	1	0	0
0	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0
0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1
1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1
0	1	0	0	0	0	0	0	0	1	0	0	1	0	1	0
0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	1
0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	0

where wrap-around of the plots is considered. A design with four treatments, 1, 2, 3, 4, for which (2) is true, is given in Fig. 2.

1	2	3	4
3	2	1	4
3	4	1	2
1	4	3	2

Fig. 2. A design with four treatments, 1, 2, 3, 4, for the array in Fig. 1.

It can easily be seen that for all arrays a design with two equally replicated treatments will always satisfy (2).

Further details will appear in a forthcoming paper.

Mr R. Mead (University of Reading): There are many aspects of this paper I would like to discuss. Specific questions on which I would welcome the authors' comments are:

(i) How can the adequacy of their NN model be tested?

(ii) How reasonable is it to assume that b is invariant over the whole experiment?

(iii) How can the possible bias of treatment effect estimates be assessed?

Incidentally, I do deprecate the authors' assumption that treatment effect estimates from their model are correct and their discussion of results from other methods as "misclassifications" (Section II.8). If the method is justified on the empirical basis that it works, there is no justification for assuming that one set of results is correct rather than another. I shall discuss two particular aspects of the NN method from the experimenter's view, and shall ask simple questions hoping for simple answers. (Are not all important statistical ideas explainable in simple terms?)

If detrending gives a marked improvement on a randomized block model, then this implies heterogeneity within blocks and this, in turn, implies that smaller blocks are needed. This happens to coincide with my belief that we allow or encourage experimenters to use overlarge blocks. We have allowed the computational simplicity of the RB design to replace the statistical concept of small homogeneous blocks, even though the computational simplicity is now largely irrelevant. The use of smaller blocks should improve the estimation of effects. To test this hypothesis, consider the analysis of the 38 plots in Section II.3. If we superimpose blocks of 2 plots or of 3 plots, instead of the original blocks of 4 plots, we obtain residual mean squares of 0.0151 and 0.0324 for blocks of 2, and 0.0346, 0.0247 and 0.0390 for blocks of 3, compared with 0.0363 for blocks of 4. These figures seem to support the hypothesis. To what extent does the use of smaller, properly homogeneous, blocks reduce the need for NN analysis?

I suggest that comparison of different models is difficult and that, in particular, we must be cautious in our discussion of degrees of freedom where these do not derive explicitly from chi-squared distributions. Consider the set of 37 differences, $z_i = y_i - y_{i-1}$ for the 38 plot yields and the least squares analysis of (i) $z_1, z_3, \ldots, z_{37}$, (ii) $z_2, z_4, \ldots, z_{36}$, (iii) all 37 z values. The first two analyses are, of course, identical to those for the models with blocks of 2 units discussed earlier, and the residual mean squares are based on 15 or 16 d.f. with 18 or 19 d.f. implicitly allocated to block effect estimation. The third analysis, allowing for correlations of the z_i, gives 34 d.f. for the residual mean square and none for block effects. Thus, if we analyse either half set of differences, each of which involves all the original observations, then we are allowing for block effects, but if we analyse both, then we do not allow for block effects. The treatment effect estimates for all three analyses are very similar, and similar also to the authors' NN results in Table 4. In all analyses, the set of residuals for z_i, shown below for (iii) are similar and show little autocorrelation pattern ($r = 0.19$), and I believe that we should accept that the models give very similar fits. But the assumptions apparently made in the models appear to offer differing interpretations, and to some extent the adequacy of the fit depends not on the sum of squares of residuals but on the degrees of freedom and other correction factors. So my question is, how do we compare different models and what do the degrees of freedom of different models represent?

Residuals of z_i for model (iii)

−0.11	+0.42	−0.20	+0.07	+0.04	−0.10	+0.07	−0.30	+0.13	+0.17
+0.13	−0.06	−0.45	−0.12	+0.01	+0.20	+0.10	+0.08	+0.29	+0.18
+0.10	−0.01	+0.10	−0.07	−0.07	−0.19	+0.18	+0.07	−0.03	−0.06
+0.20	+0.02	−0.35	−0.60	−0.10	+0.06	0.00			

Dr B. D. Ripley (Imperial College): The Waite team are to be congratulated on a difficult but stimulating paper. I am grateful to them and Professor Terry Speed for earlier discussions.

Their NN analysis is not as novel as the authors believe. An iterated Papadakis analysis with fixed regression coefficient, or generalized least squares or maximum likelihood for a conditional autoregression yields

$$(\mathbf{D}^{\mathrm{T}}\mathbf{D})\,\hat{\boldsymbol{\tau}} = \mathbf{D}^{\mathrm{T}}\mathbf{Y}^{*} - a\mathbf{D}^{\mathrm{T}}\mathbf{N}(\mathbf{Y}^{*} - \mathbf{D}\,\hat{\boldsymbol{\tau}}),\ \mathbf{Y}^{*} = \mathbf{Y} - \mathbf{B}\,\hat{\boldsymbol{\beta}}$$

(Ripley, 1978, 1981; Martin, 1978, 1982) where $\mathbf{Y}$ is the data vector, $\mathbf{D}$, $\mathbf{B}$ and $\mathbf{N}$ the design, block and neighbour incidence matrices, $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$ the treatment and block parameters and a a parameter. This can be recast as

$$\mathbf{D}^{\mathrm{T}}(\mathbf{I} - a\mathbf{N})\,\mathbf{Y}^{*} = \mathbf{D}^{\mathrm{T}}(\mathbf{I} - a\mathbf{N})\,\mathbf{D}\,\hat{\boldsymbol{\tau}}$$

which differs from (21) only in (i) summing over all plots, border as well as internal, and (ii) how the block parameters are estimated. Equation (21) can be expressed as

$$\mathbf{P}(\mathbf{D}^{\mathrm{T}}\mathbf{C}\mathbf{D})\,\hat{\boldsymbol{\tau}} = \mathbf{P}[\mathbf{D}^{\mathrm{T}}\mathbf{C}\mathbf{Y} - a\mathbf{D}^{\mathrm{T}}\mathbf{C}\mathbf{N}(\mathbf{Y} - \mathbf{D}\,\hat{\boldsymbol{\tau}})], \quad \mathbf{1}^{\mathrm{T}}\hat{\boldsymbol{\tau}} = 0,$$

where $\mathbf{P} = (\mathbf{I} - p^{-1}\mathbf{1}\mathbf{1}^{\mathrm{T}})$, $a = b/$(number of neighbours) and $\mathbf{C}$ is the diagonal matrix of entries 1 or 0 for internal and border plots respectively. This form provides a convenient iterative computation for $\boldsymbol{\tau}$ in $p - 1$ variables. In removing block *means* the authors make a false assumption of orthogonality which must slightly impair the efficiency of their analysis. The differences between these analyses (and true iterated Papadakis) yield small changes in the treatment estimates, since appreciable neighbour correlations will force b near one.

Why then does the paper rate NN analysis much above iterated Papadakis? The answer is that the authors do not compare treatment estimates directly, nor do they compare like with like. In Part II the "treatment mean square" is a generalized mean square. More seriously, the residuals differ between Parts I and II. Part II uses $(\mathbf{I} - a\mathbf{N})(\mathbf{Y} - \mathbf{D}\,\hat{\boldsymbol{\tau}})$ minus its block mean for both *RMS* and V_0 (with $\tau \equiv 0$). Part I apparently uses $\mathbf{Y}$-block mean-$\mathbf{D}\,\hat{\boldsymbol{\tau}}$-covariate = $(\mathbf{I} - \hat{a}\mathbf{N})(\mathbf{Y} - \mathbf{D}\,\hat{\boldsymbol{\tau}} - \mathbf{B}\hat{\boldsymbol{\beta}})$ for the Papadakis *RMS but* the ordinary *RMS* for V_0. Thus all residual mean squares except V_0 in Part I are on differenced scale, concerning $\mathbf{Y}'$ rather than $\mathbf{Y}$. The difference is analogous to that between estimating the variance of an AR (1) process and that of its innovations process. If my interpretation is correct then the apparent biases found are to be expected.

Section I.2 suggests that Fig. 2 is inconsistent with an autoregressive model. The "smooth trend plus independent error" and "error in variables" models are essentially equivalent, being different ways to explain high power at low spectral frequencies. The theoretical autocorrelations of neither explain Fig. 2. However, the series involved are very short and the sample autocorrelations are seriously biased. The "U-shaped" behaviour of Fig. 2 is consistent with a low-order autoregressive model. The "detrended" data show some evidence of over-differencing.

I am unhappy about the choice of the data used. The aggregation is artificial. Do the plots produced adequately reflect the variation found in genuine long narrow plots? The Mercer and Hall data are known to have an atypical periodicity (Mercer and Hall, 1911; McBratney and Webster, 1981; Ripley, 1981). (Dr P. A. Burrough has found a similar effect elsewhere from earlier drainage operations.) Fig. 1 may be an artistic impression, but if accurate shows the Wiebe data to be smoother than any uniformity trial I have examined.

Clearly the Waite team and neo-Papadakists have much to learn from each other. In particular I would like to see more work done on the validity of the standard errors ascribed by the various methods.

Mr R. A. Kempton (Plant Breeding Institute, Cambridge): As previous discussants have indicated, for b known, equation (23) gives the iterated Papadakis solution, and hence the novelty of the analytic method lies in the derivation of an estimator for b and an unbiased estimate of error for treatment comparison.

The authors propose a preliminary analysis with $b = 1$. This seems a sensible starting point since my experience of the nearest neighbour method suggests that, for the one-dimensional analysis, there is no appreciable gain in precision if $b < 0.6$, and $b > 0.8$ is required for any substantial gain.

In the authors' optimal analysis, b is estimated by minimizing the residual mean squares (RMS), obtained by dividing the residual sum of squares (RSS) by its equivalent degrees of freedom (EDF) which itself depends on b (equation (28)). The iterative Papadakis algorithm (Bartlett, 1978) gives a solution close but not identical to the minimum RSS solution. For the Rothamsted example I find the Papadakis estimator gives $\hat{b}_p = 0.93$ compared to 0.94 from minimizing the RSS, while according the authors' criterion the "optimal" estimator gives $\hat{b} = 0.96$. However, the authors' approximate formula (equation (33) with $\omega = 2.07$) gives $\hat{b} = 0.81$: perhaps they would explain this discrepancy and why they have not proceeded to the optimal solution.

Comparison of the efficiencies for the refined NN method (Table 6) with those given for Papadakis, using NN balanced designs without border plots, in Table 1(a) (column headed TMS_{10}/V_0) suggests that the new method is no more efficient. The bias in treatment F-ratio for the Papadakis analysis, which does not seem to have been wholly removed by the authors' analysis when $b \neq 1$, is substantially reduced when treatments are moderately well replicated and balanced for nearest neighbours. Since the Papadakis method estimates treatment effects with moderately high efficiency and the computing algorithm is relatively straightforward, it still provides, for the present, a useful *ad hoc* method for analysing selection trials where an exact estimate of treatment error is not required. However, the authors' proposals for unbiassed error estimation and particularly for design are important and should substantially improve the effectiveness of the nearest neighbour method.

My own interest in nearest neighbour methods currently lies in compensating for the effect of interplot interference in small plot trials (Kempton, 1982). Interference may be due, for example, to competition between varieties ($\beta < 0$) or to the spread of disease from untreated plots ($\beta > 0$). Here there is a direct relationship between the responses of neighbouring plots, rather than an underlying relationship between treatment adjusted responses arising from local trends in soil fertility. The plot values may thus be defined by the model

$$y_i = \tau_{\mathcal{T}_i} + \beta \bar{y}_{\mathcal{N}_i} + \eta_i, \qquad i = 1, 2, \ldots, n$$

with possibly added block terms to remove large scale spatial effects. If the η_i are iid $N(0, \sigma^2)$ random variables and β is known, the treatment means are estimated by

$$\hat{\tau}_j = \bar{y}_{\mathcal{U}_j} - \beta \bar{y}_{\mathcal{N}_{\mathcal{U}_j}}$$

and the variance of treatment contrasts is independent of the design. (Note, that Herzberg discussed this model in Bartlett (1978) but wrongly identified it with the Papadakis method.) The estimation of β gives rise to an analogous problem to Dr Wilkinson's method, in that the function to be minimized is not simply the RSS, $\Sigma(y_i - \tau_{\mathcal{T}_i} - \beta \bar{y}_{\mathcal{N}_i})^2$, but $|\mathbf{I} - \beta \mathbf{W}|^{-2/n}$ RSS, where $\mathbf{W}$ is a matrix specifying the neighbours $\mathcal{N}_i$ (including border plots where available) for each plot i.

Dr R. J. Martin (University of Sheffield): I would like to congratulate the authors on a most interesting paper, which I hope will stimulate much further work in the use of spatial methods for analysing field experiments.

Recent research has concentrated on stationary models—the authors' use of a "smooth trend plus error" model is a welcome innovation. Their evidence for the fit of this model, and of the predictions made from it, is impressive, but I am not convinced that stationary models might not also have performed well. I suspect many time-series analysts looking at Fig. 2, and noting that sample autocorrelations are correlated and for high lags have large standard errors, would diagnose a low-order autoregression (see Anderson, 1976, pp. 21, 25). For the Rothamsted experiment a second-order autoregression appears to fit reasonably well, and, interestingly, both the first-order and second-order exact maximum likelihood fits give estimated treatment contrasts essentially the same as in Table 4, using respectively 5 and 6 degrees of freedom.

The correlations in Fig. 2 decline fairly rapidly initially, yet it is often stated (see Besag, 1981) that for agricultural data there is a very slow decay in the correlations, which possibly could be modelled by a first-order auto-normal scheme. Are the authors' data sets atypical? Note that realizations from this scheme with moderately high correlations might prove difficult to diagnose as stationary.

In the classical analysis randomization is used to justify an analysis that would otherwise be incorrect, and a method of estimation that would otherwise be inefficient. It tends to be forgotten that the theoretical justification for randomization with fixed errors is not strong (Kempthorne, 1952, Chapters 7, 8, 10), and that acceptance of the method in this country was controversial (it has never been fully accepted in some European countries). I also believe that Fisher implicitly assumed some underlying error distribution for which randomization led to sums of squares being a mixture of distributions approximating to the relevant χ^2 distribution. With information on the underlying distribution it is quite logical to restrict the randomization to a set large enough to ensure some robustness, but small enough to ensure an approximately constant high efficiency. In the classical approach this leads to row and column blocking, etc. and restricted randomizations (Grundy and Healy, 1950). Some related discussion for stationary models is in a forthcoming paper of mine. Another justification of the classical approach was that everyone would get the same answers. Nowadays, however, there is more emphasis on the assumptions for an analysis being reasonable.

To return to stationary models, which when fitted by maximum likelihood need not suffer the defects of the Papadakis method, there is some evidence (see also above) that the analysis is reasonably robust to the choice of model, and can be at least as efficient as the best classical design. Border plots are not needed, or can be used as full observations if they are present. Theoretical progress is possible on the choice of an efficient set of designs when treatment effects are estimated either by maximum likelihood or by ordinary least-squares (see Martin, 1982, and a forthcoming paper of mine). For example, for the first-order auto-normal scheme nearest-neighbour balanced designs, for which variances of treatment contrasts are approximately constant, are highly efficient, but ordinary least-squares is almost as efficient as maximum likelihood. If correlations decay quicker, very different results follow.

Finally, Fig. 1 shows quite strongly correlations in both directions. Can the authors explain why in Part I "Analysis" is not important?

Miss A. Laurence (University of Kent at Canterbury): I shall confine my comments to Part I of this paper.

The authors note that the iterated analysis oscillates towards the final solution. In a study I have conducted on simulated data using the Papadakis method of analysis this effect has been utilized to increase the rate of convergence. Two cycles are performed, the average of the estimated treatment means obtained and used as the basis for a further cycle. This is repeated until the estimated treatment means converge. In the variety of situations I have considered the iterative scheme converged within 10 cycles, although cases can be envisaged where the RMS remains approximately constant but the estimated treatment means do not converge. I would welcome some information on the effect of iteration on the estimated treatment means in this study.

A further problem encountered is that, forming covariates from neighbours within the trial only, iteration may cause a drift in the estimated treatment means. This can be eliminated by centralising on each cycle so that the weighted mean of the estimated treatment means is the grand mean. This is essentially a reparametrization of the model.

Under the smooth trend + independent error model of this paper the smooth trend will typically consist of removable and non-removable trend effects. The residuals are then a combination of these non-removable trend effects and the independent error terms. Therefore V_0 is a positively biased estimate of σ_η^2. In my study using simulated data the RMS was compared to the known value of σ_η^2. It was found that when the random variation is due to independent error terms only the RMS is not seriously biased. In the presence of correlated error terms the RMS is a positively biased estimate of σ_η^2 although the bias does decrease as the analysis is iterated. The conflicting results for the RMS depend on the definition of the true error variance, $E(V_0)$ or σ_η^2; see also Dr Ripley's remarks.

In a conventional analysis of an orthogonally designed trial, the standard error of a treatment mean is, of course, $(r^{-1}\ \mathrm{RMS})^{\frac{1}{2}}$. I have found that an analogous quantity often provides a good estimate of this standard error in a Papadakis analysis. An exception occurs, under a completely randomized design, when there is large between row variation, small within row variation and adjustment is by row neighbours. The standard error can then become negatively biased in an iterated analysis. The differing conclusions between the studies indicate that the *TMS* may incorporate trend effect which, relative to the RMS, is not reduced by iteration.

Mr G. V. Dyke (Rothamsted Experimental Station): Dr Jenkyn and I have considered the authors' application of intra-NN analysis to our 1975 barley experiment (Jenkyn *et al.*, 1979) and to certain sets of constructed yields showing specified "interference" effects.

In the real experiment the authors' method renders the effects of interference non-significant; we have found in this and other experiments that, with practicable numbers of plots, they are very near the limit of detection. They are, however, corroborated by assessments of disease and accord with biologically-based expectation. Our method of fitting Fourier curves of relatively long wavelength is intended to describe smoothly a profile of potential yield crudely indicated by "block" means. The authors' method, we suggest, may "overfit" short-wave variation that is more reasonably ascribed to inter-plot (treatment) interference. We therefore doubt whether it is fair to dismiss interplot interference on the grounds that its effects are non-significant when the authors' Intra-NN analysis is applied. There is abundant biological evidence that these effects do occur; an example from a recent experiment involving triadimefon, a vapour-active material, is:

Mildew: mean logits (and % by back transformation)

Plot treatment	*Treatment to western neighbour*			
	None	*Triadimefon*	*Tridemorph*	*Mean*
None	–	−1.02 (11.5)	−0.52 (26.1)	−0.77 (17.6)
Triadimefon	−3.47 (0.1)	–	−3.36 (0.1)	−3.41 (0.1)
Tridemorph	−2.07 (1.6)	−2.54 (0.6)	–	−2.30 (1.0)
SED 0.137 (14 d.f.)				

Our interest in the effect of interference on the results of field experiments led us to do a series of simulations. In two of the sets of constructed yields interference is the only source of variation in addition to direct treatment-effects:

Treatment *A* yields 5 on all its plots and lessens yield of every adjacent plot by 1 (by 2 if *A* is present on both sides).

Treatments *B*, *C*, *D* yield 10 in the absence of interference.

Design I is serially balanced as was the barley experiment; Design II is chosen to be "unbalanced" thus:

BACD BADC BACD . . . . (36 plots)

		Treatment estimate			
		A	*B*	*C*	*D*
Design I	(Theoretical	5	10	10	10)
	Randomized block	5	9.3	9.3	9.3
	Intra-NN	5	8.7	8.7	8.7
Design II	(Theoretical	5	10	10	10)
	Randomized block	5	9.0	9.4	9.6
	Intra-NN	5	9.0	9.2	9.3

In both simulations intra-NN analysis gives estimates further removed from the theoretical yields in the absence of interference than the conventional treatment-means; in Design II the difference is small, in Design I large.

Added after the meeting: These sets of data were constructed originally for investigating the behaviour of the Papadakis method in the presence of interference; I noted that in a serially-balanced experiment with "simple" interference as defined above the Papadakis regression coefficient is identically zero. It seems also that, for such designs, the authors' matrices $\mathbf{B}'$, $\mathbf{T}'$ (p. 000) have zero column-totals without premultiplication by $\mathbf{Q}$.

Mr C. R. Muirhead (Imperial College, London): I would like to make a few remarks concerning the error structure in the one-dimensional case. Firstly, a swing in the serial correlations from positive to negative values with increasing lag, as seen in Fig. 2, would be expected under an AR (1) model whose auto-regressive parameter is close to 1; further details are given by Anderson (1979).

Secondly, both the model used in this paper and the AR (1) model could be generalized by a one-dimensional version of Besag's (1977) errors-in-variables model, i.e. with reference to equation (9),

$$y_i \mid i = \tau_{\mathcal{F}i} + \xi_i + \eta_i,$$

$$(1 - \alpha B)(1 - \alpha B^{-1})\,\xi_i = (1 - \alpha B^{-1})\,\epsilon_i, \quad |\alpha| \leqslant 1.$$

Here $Bx_i = x_{i-1}$ and $\{\eta_i\}$, $\{\epsilon_i\}$ are uncorrelated, both within and between series, with $E(\eta_i) = E(\epsilon_i) = 0$, var $(\eta_i) = \sigma_\eta^2$, var $(\epsilon_i) = \sigma_\epsilon^2$. The stationary AR (1) model arises when $\sigma_\eta^2 = 0$ and $|\alpha| < 1$, whereas the smooth trend and uncorrelated error model arises as $\alpha \to 1$, $\sigma_\epsilon^2 \to 0$ and $\sigma_\xi^2 = \sigma_\epsilon^2/(1 - \alpha^2) \to \omega\sigma_\eta^2$. Adapting method 1, α, σ_ξ^2 and σ_η^2 could be estimated by carrying out analyses based on $\{y_i'(0)\}$, $\{y_i'(b^*)\}$ and $\{y_i'(1)\}$ respectively, where $0 < b^* < 1$, and then equating the observed residual sums of squares (rather than RMS's) to their expectations. To obtain the optimal NN analysis note that, for α and ω given, var $(y_i'(b))$ is minimized by taking

$$b = \frac{2\omega\alpha}{1 + \omega(1 + \alpha^2)}.$$

Dr F Yates: It is now over 40 years since I introduced various types of incomplete-block and lattice-square designs, having particularly in mind the need of plant breeders for comparing large numbers of varieties. These were all based on Guassian least squares; randomization ensured that the assumption of independent errors is satisfied; mathematically, therefore, they are extremely simple. During and after the war I became involved in other problems and have not followed in detail the development of Papadakis and nearest-neighbour analysis, and therefore will not waste your time by discussing the finer points of Dr Wilkinson's paper.

The point I chiefly want to make is that if we are using simulation—and it is quite a good weapon now that we have computers—we should ask ourselves rather carefully what questions are at issue, and see whether it is possible to design realistic simulations to answer them.

In nearest-neighbour (NN) analysis, and in Papadakis analysis, we are adjusting the treatment means in a rather involved way. I was not at all clear, from a hasty and very belated examination of Dr Wilkinson's paper, whether the treatment effects that emerge from the analysis will, in fact, be a true representation of the treatment differences. Dr Wilkinson was pinning his faith in quite a lot of the paper on these simulations, but all were on uniformity trials. It would be a perfectly simple matter to extend these by adding sets of imaginary treatment effects and to see how well they agreed with the final estimates. Good agreement, relative to estimated experimental error, would provide practical confirmation of the soundness of the rather involved theoretical arguments given in the paper.

An even simpler simulation procedure, using data from actual variety trials with partial NN balance such as those discussed in Section II.8, would be to add sets of 90 arbitrary increments to the yields of the 90 lines in an experiment. In a randomized block analysis the changes in the estimates of the line yields will correspond exactly to the added increments. How closely does this hold in an NN analysis? This procedure has the important advantage that information which

is relevant to the fertility variations on the type of land actually used for the trials can be accumulated.

This is in no sense intended as a criticism of Dr Wilkinson's approach. Practical experimenters have long been aware of the need to make allowances for local fertility variations, which are sometimes glaringly obvious from examination of the yields or residuals when they are set out in plan form. Before computers were available little could be done, apart from close attention to design and plot shape, other than rejection of defective areas. Computers greatly expand the possibilities, but at the cost of the logical simplicity of the Fisherian methods.

Attempts to improve the estimates by fitting periodic or autoregressive functions are to my mind completely unrealistic, and have shown themselves to be so when attempted in practice. Dr Wilkinson mentioned in Section II.3 the failure of a four-term Fourier series fitted to a Rothamsted experiment. A two-dimensional spectral analysis of the Mercer and Hall data, in a paper which contains some impressive but very misleading computer-drawn diagrams (McBratney and Webster, 1981), claimed to have revealed a wavelength of about 40 ft across the drill rows, which the authors attributed to an earlier ridge and furrow system. But a graph in the Mercer and Hall paper clearly shows this is solely due to three peaks on the west half of the field; no similar peaks occur on the east half.

Dr P. J. Green, Dr C. Jennison and Dr A. H. Seheult (University of Durham): We should like to thank the authors for a genuinely stimulating paper—one which has stimulated us to seek an alternative analysis of their very sensible and plausible model. Their key assumption is that the plot yields $\mathbf{y}$ can be additively decomposed as $\mathbf{y} = \boldsymbol{\xi} + \mathbf{D}\boldsymbol{\tau} + \boldsymbol{\eta}$, where the fertility effects $\boldsymbol{\xi}$ are spatially smooth, $\mathbf{D}$ is the design matrix for treatment effects $\boldsymbol{\tau}$ over *all* plots and $\boldsymbol{\eta}$ is the residual error.

We consider first a purely data-analytic approach, in which we will make explicit use of the smoothness assumption. The simplest description of smooth fertility in one dimension is to require that in some sense the second differences $\xi_{i-1} - 2\xi_i + \xi_{i+1}$ along columns should be small. Adding the usual requirement that the residuals $\boldsymbol{\eta}$ should also be small, a least squares approach leads us to estimates of $\boldsymbol{\tau}$ and $\boldsymbol{\xi}$ which minimize the *penalty function* $\lambda \boldsymbol{\xi}^{\mathrm{T}} \boldsymbol{\Delta}^{\mathrm{T}} \boldsymbol{\xi} \boldsymbol{\Delta} + \boldsymbol{\eta}^{\mathrm{T}} \boldsymbol{\eta}$, where the matrix $\boldsymbol{\Delta}$ takes second differences along columns. Here, λ is a tuning constant that can be varied to control the degree of smoothness in the estimate of $\boldsymbol{\xi}$—the larger λ the smoother $\boldsymbol{\xi}$; cf. Huber, 1978. We shall call this method *least squares smoothing* and suppose for the present that λ is fixed.

The resulting estimates of $\boldsymbol{\tau}$ and $\boldsymbol{\xi}$ are readily computable solutions of the simultaneous equations

$$\boldsymbol{\tau} = (\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}\,\mathbf{D}^{\mathrm{T}}(\mathbf{y} - \boldsymbol{\xi}),$$

$$\boldsymbol{\xi} = (\mathbf{I} + \lambda\,\boldsymbol{\Delta}^{\mathrm{T}}\boldsymbol{\Delta})^{-1}\,(\mathbf{y} - \mathbf{D}\boldsymbol{\tau}).$$

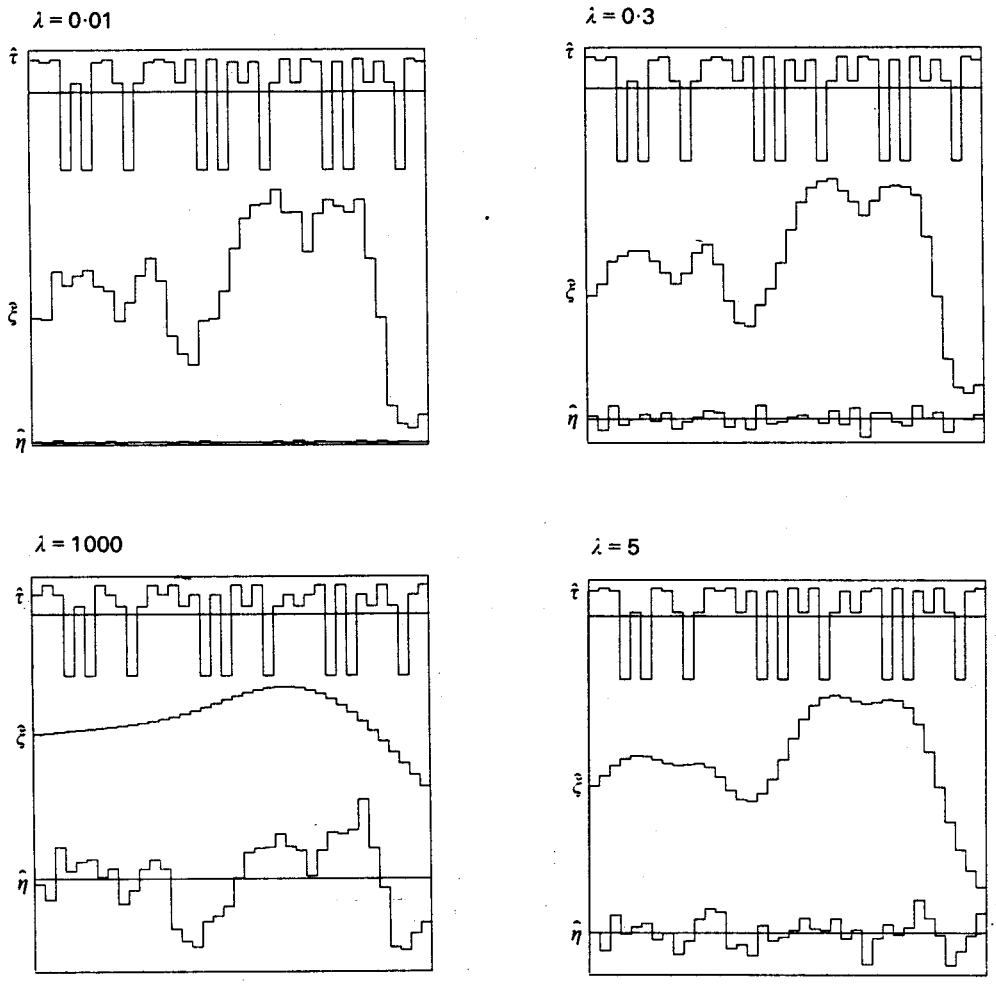
However, it is worth noting that for given $\boldsymbol{\xi}$, the first equation gives the usual estimate of $\boldsymbol{\tau}$. On the other hand, for given $\boldsymbol{\tau}$, the second equation tells us to estimate $\boldsymbol{\xi}$ by applying the *smoothing matrix* $(\mathbf{I} + \lambda\,\boldsymbol{\Delta}^{\mathrm{T}}\boldsymbol{\Delta})^{-1}$ to $\mathbf{y} - \mathbf{D}\boldsymbol{\tau}$.

The equations also suggest the usable iterative scheme whereby we alternately fix one of $\boldsymbol{\tau}$ and $\boldsymbol{\xi}$ and solve for the other.

We now suppose that $\mathbf{D}$ is given and that $\boldsymbol{\xi}$ and $\boldsymbol{\tau}$ are fixed but unknown, and introduce the usual assumptions about the residuals: $E(\boldsymbol{\eta}) = \mathbf{0}$ and $\mathrm{var}(\boldsymbol{\eta}) = \sigma^2\mathbf{I}$. This enables us to evaluate the precision of least squares smoothing estimates for various values of λ, and to compare our method with that of the authors with $b = 1$.

All these estimates are of the form $\hat{\boldsymbol{\tau}} = \mathbf{A}\mathbf{y}$ for some matrix $\mathbf{A}$, with $E(\hat{\boldsymbol{\tau}}) = \boldsymbol{\tau} + \mathbf{B}\boldsymbol{\Delta}\boldsymbol{\xi}$ and $\mathrm{var}(\hat{\boldsymbol{\tau}}) = \sigma^2\mathbf{A}\mathbf{A}^{\mathrm{T}}$ for an appropriate matrix $\mathbf{B}$.

If $\boldsymbol{\Delta}\boldsymbol{\xi} \doteq \mathbf{0}$ the estimates are therefore approximately unbiased with variances $\sigma^2\,\mathbf{A}\mathbf{A}^{\mathrm{T}}$. In addition, least squares smoothing produces an estimate $\hat{\boldsymbol{\eta}}$ and if $\boldsymbol{\Delta}\boldsymbol{\xi} \doteq \mathbf{0}$, $E(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta})$ is approximately equal to a known multiple of σ^2, allowing σ^2 to be estimated in the usual way. Computations show that the entries in $\mathbf{A}\mathbf{A}^{\mathrm{T}}$ are smallest for least squares smoothing with large values of λ. This is to be expected since as $\lambda \to \infty$ this method tends to a covariance analysis fitting linear fertility trends in each column, and when $\boldsymbol{\Delta}\boldsymbol{\xi} = \mathbf{0}$ this is the most efficient analysis.



Figs D1–4. Different decompositions $(\hat{\tau}, \hat{\xi}, \hat{\eta})$ of the 38 barley yields from the Rothamsted mildew control experiment, for $\lambda = 0.01, 0.3, 5$ and 1000.

The authors' method is less efficient in this case since it fits correlated $\mathbf{y}''$ by *unweighted* least squares.

In order to apply the least squares smoothing method it is necessary to choose a value of λ. A large value of λ is desirable as this gives relatively small entries in $\mathbf{A}\mathbf{A}^{\mathrm{T}}$; on the other hand, λ must be small enough to allow an accurate fit of a smooth but non-linear fertility pattern. The latter point is illustrated in Figs D1–4 for the mildew control data. Least squares smoothing with $\lambda = 1000$ produces a smooth estimate of fertility but the residuals exhibit a definite pattern: if $\lambda = 5$ is used this pattern is taken up by the fertility and little or no structure is apparent in the residuals.

One method of choosing a value for λ, therefore, is to subjectively assess the degree of roughness of $\hat{\boldsymbol{\xi}}$ and the amount of structure in $\hat{\boldsymbol{\eta}}$ from graphs of the decomposition for various values of λ. Another possibility is to augment our assumptions so that λ becomes a parameter. It transpires that our estimate of $\boldsymbol{\tau}$ is also the generalized least squares estimate in a model for which, in addition to the assumptions above about $\boldsymbol{\eta}$, $\boldsymbol{\xi}$ is random with $E(\Delta\boldsymbol{\xi}) = 0$ and $\mathrm{var}(\Delta\boldsymbol{\xi}) = \lambda^{-1}\sigma^2\mathbf{I}$. This leads to an alternative estimate of σ^2, and if in addition we assume

normality for (ξ, η) we can estimate λ by maximum likelihood. For the mildew control experiment, the estimate of λ is about 0.3, yielding a decomposition of the data in which $\hat{\xi}$ exhibits four distinct smooth peaks, and there is no structure apparent in $\hat{\eta}$; see Figs D1–4.

It is instructive to compare our estimating equations for τ with those of NN adjustment and ordinary covariance analysis. For each of these methods these equations are of the form $\mathbf{D}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})(\mathbf{y}-\mathbf{D}\tau)=\mathbf{0}$, where $\mathbf{S}$ is $(\mathbf{I}+\lambda\Delta^{\mathrm{T}}\Delta)^{-1}$ for least squares smoothing, is $\mathbf{C}(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}$ for covariates represented as the columns of $\mathbf{C}$, and for the authors' NN analysis is an arbitrary member of the family defined by

$$(\mathbf{Sy})_i = \begin{cases} y_i, & \text{for border plots} \\ \frac{1}{2} rby_{i-1} + (1-r)\, y_i + \frac{1}{2} rby_{i+1} & \\ \quad + r(1-b)\, \bar{y}_{\mathcal{A}\, i} + rb\overline{y'_{\mathcal{A}\, i}}(1), & \text{for internal plots} \end{cases}$$

for $0 < r < 1$. Thus both NN analysis and least squares smoothing allow for spatial structure in a way which is similar to adjusting for covariates.

Two further points follow from this representation of the NN method. Firstly, our iterative method described above can be used to compute the NN adjusted estimates of τ . Secondly, examination of the form of $\mathbf{S}$ for NN adjustment shows that it combines moving average smoothing and block adjustment. This compromise between the use of blocks and real spatial information is avoided in our least squares smoothing method.

Finally, we should like to thank the authors for giving us the opportunity of seeing an advance copy of tonight's paper.

Dr R. L. Smith (Imperial College, London): The conflict between randomization and balance in the design of experiments, to which the authors refer as the Student–Fisher controversy, is one which arises in other areas of statistical application, for example, clinical trials. The conflict is present even in the simplest case of a one-way layout with two treatments. I do not see how the concept of NN-balance applies in this case. It seems to me that the controversy is still unresolved.

Kiefer and Wynn (1981) have considered optimal design in the presence of correlated observations. A natural question, raised by the authors' discussion, is whether a similar theory can be developed to handle "smooth trend". I offer here some very tentative suggestions of the form such a theory might take. The discussion is restricted to the one-way layout with two treatments, and is based on the development of "accidental bias" given by Efron (1971) in the context of biased coin designs.

Define δ_n to be $+1$ or -1 according as the nth treatment unit is assigned the first or second treatment. Our model is

$$y_n = \mu + \alpha\delta_n + \xi_n + \eta_n, \quad 1 \leqslant n \leqslant N,$$

where α is the treatment effect, ξ_n the trend component and η_n the error. Assuming $\Sigma\, \xi_n = 0$, the effect of the trend on the difference of treatment means is proportional to $\Sigma\, \xi_n\delta_n$, so we are led to consider the variance of this quantity.

If the sequence $\{\delta_n, 1 \leqslant n \leqslant N\}$ is asymptotically stationary as $N \to \infty$, then we may define covariances

$$\gamma_k = E\{\delta_n \delta_{n+k}\}$$

and an associated spectral density

$$f_\delta(\omega) = (2\pi)^{-1} \sum_{-\infty}^{\infty} \gamma_k e^{-i\omega k}$$

with respect to the stationary distribution. For fixed $\xi_1, \ldots, \xi_N$, we then have

$$E\{(\Sigma\, \delta_n \xi_n)^2\} = 2\pi N \int_{-\pi}^{\pi} R(\omega)\, f_\delta(\omega)\, d\omega \qquad (1)$$

where $R(\omega) = (2\pi N)^{-1} \mid \Sigma\ e^{i\omega n}\xi_n|^2$ is the sample periodogram of the ξs.

If we impose the normalizing condition $2\pi N \int R = 1$ then the maximum value of (1) is just the maximum value of $f_\delta(\omega)$. Efron (1971) therefore suggested that $\max_\omega f_\delta(\omega)$ be taken as a measure of the vulnerability of a design to "accidental bias". However, if no further restriction is placed on R, the "best"design from this point of view is the completely randomized design, for which f_δ is constant – a trivial conclusion.

Suppose now the additional condition is given that the ξs are "smooth". One way to make this concept precise is to specify

$$\Sigma(\xi_n - \xi_{n-1})^2 \leqslant \eta\ \Sigma\ \xi_n^2 \tag{2}$$

for some small η. In terms of $R(\omega)$, this condition becomes

$$\int_{-\pi}^{\pi} R(\omega) \cos \omega d\omega \geqslant (1 - \eta/2) \int_{-\pi}^{\pi} R(\omega)\, d\omega. \tag{3}$$

If we now maximize (1), subject to the normalizing condition and to (3), we get

$$\max \{f_\delta(\omega): 0 \leqslant \omega \leqslant \cos^{-1}(1 - \eta/2)\} \tag{4}$$

as a new measure of the vulnerability of the design to bias caused by smooth trend.

For example, for a randomized blocks design with block length G (an even integer) we have

$$\begin{aligned} 2\pi f_\delta(\omega) = 1 - [1 - \sin((2G + 1)\ \omega/2)/\sin(\omega/2) \\ + \cos(\omega/2)\{\cos(\omega/2) - \cos((2G + 1)\ \omega/2)\}/ \\ \sin^2(\omega/2) - 2G]/(2G\ (G - 1)) \end{aligned}$$

which is zero at $\omega = 0$ and has a maximum $G/(G - 1)$ attained at $\omega = 2\pi/G, 4\pi/G, \ldots, \pi$. For sufficiently small η, the maximum in (4) will be attained at $\omega = \cos^{-1}(1 - \eta/2)$.

I congratulate the authors for a stimulating paper which has opened up many questions for further investigation.

Mr Robin Thompson (ARC Unit of Statistics University of Edinburgh): I would like to comment on the link between the author's Nearest Neighbour (NN) analysis, the Papadakis procedure (P) and Incomplete block analysis.

Firstly, P and NN give the same treatment estimates for a given value of b. For a model $\mathbf{y} = \mathbf{X}\tau$ then the P estimate satisfies $\mathbf{X}^T\mathbf{X}\,\tau = \mathbf{X}^T\mathbf{y} - b\mathbf{X}^T\mathbf{L}\,(\mathbf{y} - \frac{1}{2}\mathbf{X}\,\tau)$, where $\mathbf{L}$ is a matrix indicating if plots are neighbours, i.e. $r \times$ treatment effect = treatment total − adjustment for neighbours (adjusted for treatments). Alternatively $\mathbf{X}^T(\mathbf{I} - \frac{1}{2}b\mathbf{L})\,\mathbf{X}\,\tau = \mathbf{X}^T(\mathbf{I} - \frac{1}{2}b\mathbf{L})\,\mathbf{y}$ which is the authors' equation (23).

Secondly, there are at least five different models used for the variance matrix (V). (1) The most explicit is (30) used in the expectation of the residual mean square and testing of treatment effects. (2) It is not clear what model is used in the expectation of the randomized block mean square. (3) For (23) to be efficient, V should be inversely proportional to $(\mathbf{I} - \frac{1}{2}b\mathbf{L})$, i.e. an autoregressive scheme. (4) For the residual mean square to be an efficient statistic, then V should be inversely proportional to $(\mathbf{I} - \frac{1}{2}b\mathbf{L})^2$. (5) It is not clear if b is a parameter chosen to give simple structure to V or just a weighting factor used to optimize treatment estimation.

P uses the same residual mean square and the implicit use of the models in (3) and (4) partly explains the problems in estimating β iteratively (E. A. Hunter, personal communication).

Finally suppose we have a linear plan and take two copies of the plan and data. In the first copy (I) suppose plots $2j + 1$, $2j + 2$ are in the same block and in the second copy (II) suppose plots $2j$, $2j + 1$ are in the same block. Then NN and P treatment estimation corresponds to recovery of interblock information using the two copies of the data and these blocks of size two.

The analogy can be taken further. (1) To give an interblock analysis and to compare treatment estimates in the two strata. (2) To suggest efficient designs, if we have an efficient $2r$ replicate design in blocks of 2 then this suggests using r replicates to give the blocks of two in the first copy and using the other r replicates to orientate and order the blocks. (3) Efficiency considerations

carry over. The Rothamsted design has efficiencies $\frac{2}{3}$ and $\frac{4}{9}$ considering first and second neighbours and 120 per cent = $100 \times (\frac{3}{2}) \times (2 \times \frac{2}{3}) / (4 \times \frac{2}{3} - \frac{4}{9})$ per cent. Efficiency factors can also be used in the P procedure in conjunction with Dr Wilkinson's algorithm implemented in Genstat to speed the convergence of treatment estimates and also to interpret the treatment sum of squares. (4) Natural interpretation of border strata and maximum likelihood estimation of variance parameters. (5) To allow the two copies of the same plot to be at the same level of some other blocking factor. Recovery of this information corresponds to using the error in variables model mentioned by Dr Patterson and Dr Besag.

Dr A. C. Atkinson (Imperial College, London): I have three points which have not been totally covered by other contributors to the discussion.

First, several speakers have shown how very close the new estimator is to various forms of the Papadakis estimator. Why did this not emerge from Dr Wilkinson's simulations? The answer lies in the design of the simulation experiment. It is a basic principle of designing such experiments that we should seek to remove variation between samples from our comparisons. To compare two estimators, this means that we should use a matched pair analysis rather than treating them as two separate samples. Had the results been collected in such a way, I feel sure the close relationship between the two estimators would have emerged well before the reading of the paper to our Society.

Secondly, the estimates of the treatment contrasts obtained from the Papadakis and new methods are presumably highly similar. The variances, and therefore the efficiency factors, that Dr Wilkinson gets seem to be very different. Therefore at least one of the methods of estimating variances of treatment differences must be inappropriate or wrong.

Thirdly, just before the beginning of Part II, the authors refer to "the efficiency results of Atkinson and Ripley". Both of us (Dr Ripley in greater detail than I in my much earlier paper) showed the close relationship between the Papadakis method and maximum likelihood estimation for AR1 processes. There seems to be a logical flaw in what is said about this in the paper. The implication is that because our results tie the Papadakis method to AR1 processes, the method will be no use for other models. This does not logically follow any more than it does for least squares in the absence of normality. Further, we have heard in this discussion that it is sometimes very difficult to discriminate between AR1 processes and moving average processes. Also, of course, since the new method is similar to the Papadakis method, if results for one method are irrelevant, or as the authors put it "only of academic interest", the same strictures must apply to results for the other method. It may of course be true that the results are irrelevant, but the paper contains only an unsubstantiated dismissal. I do feel some proof is required.

Mr C. W. Howes (Plant Breeding Institute, Cambridge): I want to speak in favour of 2-D models for use in cereal breeding trials, which is the area in which the method described tonight has been tried. These trials are necessarily smaller than those in statutory variety trials. Kempton and Howes (1981) showed that in 118 wheat breeding trials, with plots of size 5 × 14 ft, correlations of plot yields within columns of trials were equally as important as correlations within rows even when the plots were rectangular. Mechanical and human variation, induced by many field operations being carried out along the length of the plots, may well account for this. There may have been a dominant trend in one direction, but this direction could not be predicted in advance.

Thus, the NN analysis described tonight would need border plots on all four sides, even if a 1-D analysis was subsequently found to be adequate. This should be taken into account when comparing efficiencies of the NN analysis with those designs which do not require borders. For example, a fairer version of Table 8 would compare classical designs with NN designs with reduced replication. Given, say, 150 plots to test 25 treatments, the choice could be between 6 replicates of a balanced 5 × 5 lattice square design or only 4 replicates (144 plots) of a bordered NN design.

The following contributions were received in writing, after the meeting.

T. Caliński (Academy of Agriculture, Poznań, Poland): It is a privilege to be invited to contribute to the discussion on a paper read before the Royal Statistical Society. In this particular case it offers an early opportunity to study a paper which brings new results aimed at solving one

of the most important problems in the statistical analysis of agricultural field experiments. In fact, the idea of adjusting yields of field experiments for local trend effects has a long history, going back to the first decades of this century, or perhaps even earlier. In Poland, for example, the idea was already discussed in a classical book by Zaleski (1927) and quite early considered statistically by Neyman (1929).

In the present paper a thorough investigation of the Papadakis (1937) method, together with its iterated version suggested by Bartlett (1978), is presented and a new, more efficient method, is suggested. The results, as they are shown, support the new approach considerably and may open a new direction of research. But I must confess that I have found the paper not completely satisfactory in the presentation of the underlying theory. The authors have made a promising attempt to devise a general analysis of field experiments by extending the classical "fixed" block analysis to its "moving block" analogue. To convince the reader that this has been achieved one would need to show the extension by modelling the new method in a way reflecting the common model and its least-squares solution used for general block designs. The model (9) suggests that the usual model for the n yields of a block experiment is extended by an $n \times 1$ vector ξ, which in an intra-block analysis is considered as fixed, subject to some constraints, and is a random vector in an inter-block analysis. But this is not clearly and consistently exposed in the paper and, therefore, it is difficult to follow the various derivations of the method. It seems that the authors are aware of the incompleteness of the theory when they write (in Section II.8) that the justification of the method is largely pragmatic. It seems that when they succeed in satisfactory presentation of the model they will also be able to develop more theoretical knowledge of the properties of the method. Anyway, the empirical studies presented in the paper are encouraging and the authors can be congratulated for the results they have obtained till now.

Dr A. W. Davis (CSIRO, Adelaide, Australia): The authors suggest that their method may be viewed as a resolution of the classical controversy on semi-systematic versus randomized designs, and indeed it appears essential to clarify the position of their work with regard to the contemporary debate (referred to by several discussants of Bartlett, 1978) concerning:

(a) "Model-based" methods, in which the yield distribution is specified by a parametric model, and the analysis is carried out conditional on the particular treatment layout, and

(b) "Design-based" methods, in which the stochastic element is provided solely by the randomization of the treatments, conditional upon the underlying yields.

The need for this clarification is highlighted by the fact that the simulations, heavily relied on by the authors as empirical justification of their approach, are based entirely on treatment randomization over uniformity data. On the other hand, the estimation equations and variance calculations in the text, which are used in the numerical presentation based on the simulation, appear to be conditional upon a particular treatment layout. One is therefore led to look for a rapprochement between the randomization and model-based approaches of the kind established, for example, by Nelder (1965) for a large class of designs, but which is far from obvious when neighbour relationships are to be taken into account. Of course, the authors' model is not specified parametrically in such a way that maximum likelihood, or even Gauss–Markov, is directly applicable. Indeed, the fact that the trend ξ is unspecified, apart from its smoothness, makes it almost inevitable that randomization must be a component of the model itself, as the authors have recognized in their "bivariate" reference set. However, this involves difficulties. If we write $\zeta_i = \xi_i + \eta_i$ for the ith plot, then the jth treatment total can be expressed as

$$T_j' = \left(r\tau_j - \tfrac{1}{2} b \sum_{k=1}^{p} a_{jk}\tau_k \right) + \Sigma_{\mathcal{N}j} (\zeta_i - b\psi_i), \qquad (1)$$

where $\psi_i = \bar{\zeta}_{\mathcal{N}i}$ is the average ζ on the neighbours of plot i. Apart from the randomization, no mechanism is provided for the ξ's to vanish when the expectation is taken in (22), or to contribute to the variance. But the problem then arises that the adjacency matrix $\mathbf{A} = (a_{jk})$ does not remain invariant under the authors' restricted randomization, which is completely "symmetric" in the treatments. Indeed, under the randomization the T_j' will have a spherical covariance matrix, rather than the form $\mathbf{V}_T$ given in (25), whose derivation purports to condition upon a particular layout.

Under the usual randomized blocks randomization, the ζ-term in (1) will have variance

$$p^{-1}\Sigma_i[(\zeta_i - \bar{\zeta}_{\mathcal{A}i}) - b(\psi_i - \bar{\psi}_{\mathcal{A}i})]^2 \qquad (2)$$

for each T_j', where the sum extends over all plots. The optimal b in this case is thus the regression coefficient of the ζ_i's on the neighbour means ψ_i (eliminating blocks), and the minimum value of (2) is proportional to the variance of the ζ's conditional on the ψ's. Obviously this result would generalize to several neighbour classes. Thus, if we bypass (21) and (22) and simply adopt (23) as the estimation equation, then, at least at the level of a randomization analysis, the procedure is seen to make very efficient use of the information contained in the ψ's about the ζ's, lending support to the authors' claims to near-optimality. My own simulations, based only on the randomized blocks randomization, confirm that the resulting estimates of treatment effects may have considerably smaller variances than the Papadakis estimators, provided that good estimates of b are available. The symmetry of the authors' restricted randomization implies moreover that the estimates are unbiased. Nevertheless, the problem remains that since **A** is variable equation (26) is inapplicable, and so we have no valid formula for the (spherical) covariance matrix $\mathbf{V}_\tau$ of the treatment estimates under the stated randomization. The sphericity implies that it is inappropriate to use (26) to assign different efficiencies to treatment contrasts in the simulation on the basis of neighbour relationships in particular layouts (Table 5). The use of (26) in calculating the "treatment mean square", and the derivation of the EDF should also be noted.

It is quite possible, of course, that the authors' results provide a useful approximation. However, a rigorous treatment of their model and overall approach would appear to require a highly restricted randomization under which the adjacency matrices $\mathbf{A}_m$ remain invariant. Indeed, more than this would be required for a useful estimate of $\mathbf{V}_\tau$, based on (2) for example. It is not difficult to show that the main lines of the authors' analysis could be recovered if it were possible to construct randomizations having the following properties: (i) The adjacency matrices $\mathbf{A}_m$ are constant for all admissible randomizations. (ii) Any treatment j has an equal probability of falling on any plot within a block, independently for different blocks. (iii) In those blocks in which treatment k occurs on a class m neighbour of treatment j, k has an equal probability of falling on *any* class m neighbour of j (independently for different blocks). (iv) In those blocks in which k is not a neighbour of j, k has an equal probability of falling on any non-neighbour of j (independently for different blocks). For the Type IIa design of Williams (1952), all treatment relabellings provide a large class of randomizations with constant **A** matrix, but it is not obvious how independence between blocks may be achieved.

Professor Oscar Kempthorne (Iowa State Univ): The ideas addressed by the authors are very interesting and worthy of study. I favour expressing the situation in the following terms for the case of n experimental units on a line. If we assume, though we should be quite hesitant in doing so, that treatments are additive to unit plot effects with some entirely random noise, we can write for our data vector: $\mathbf{y} = \mathbf{u} + \mathbf{X}\tau + \mathbf{e}$, with $\mathbf{y}$ and $\mathbf{u}$ the vectors of data and unknown unit effects, respectively, and $\mathbf{X}$ the incidence matrix of the treatments. Obviously, with this formulation, treatment differences are not estimable. A next step comes to mind easily. Express the unit values as a linear model, e.g. a polynomial. We are then in a standard linear model situation – the linear covariance model, say, $\mathbf{y} = \mathbf{X}\tau + \mathbf{Z}\theta + \varepsilon$. This presents no problems if we assume $\mathbf{e} \sim N_n(\theta, \sigma^2\mathbf{I})$, and if we should wish to specify a variance matrix, $\mathbf{V}$, with parameters including correlations, to be determined, we know how to proceed. So we could, for instance, assume that $\mathbf{V}$ is determined by a simple autoregressive process. To do this would be to assume trend as given by $\mathbf{Z}\theta$. We would certainly be able to specify the likelihood function and proceed in the traditional way. We would want to reach, in the end, a "confidence" region for the differences of (assumed) additive treatment effects.

This approach may seem unduly assumptive. So one considers how one can use neighbouring units. One can decide that individual elements u_i are predicted as well by neighbouring values of u_i, or, in other terms, that for some matrix $\mathbf{N}$ (which one might take to be stochastic) it is the case that $\mathbf{u}$ is very close to $\mathbf{Nu}$; one might hypothesize that $\mathbf{u} - \mathbf{Nu} \sim N_n(\rho, \sigma_1^2\mathbf{I})$. Write $\mathbf{Z} = \mathbf{u} - \mathbf{Nu}$, and then we have the model: $\mathbf{y} = \mathbf{Nu} + \mathbf{X}\tau + \mathbf{e} + \mathbf{Z}$, or, perhaps preferably,

$$\mathbf{y} = (\mathbf{I} - \mathbf{N})^{-}(\mathbf{I} - \mathbf{N})\gamma + \mathbf{X}\tau + \mathbf{e} + \mathbf{Z}.$$

Now where should one go? I am not clear and cannot explore the question on this occasion.

The Papadakis method is to have N of special form involving only nearest neighbours. Then we are to consider Ny, and we are to consider $\mathbf{y}^* = (\mathbf{I} - \beta\mathbf{N})\,\mathbf{y}$, where β is a parameter to be determined. I am inclined to suggest that we are in very deep water. We certainly have a non-linear model for $\mathbf{y}^*$, and I would be interested in various aspects of identifiability of τ (subject to $\Sigma\,\tau_i = 0$). Perhaps I have missed seeing this matter discussed in the literature. I am not surprised by the difficulties the authors mention. Does convergence occur? I note that if we can assume $\mathbf{u} \doteq \mathbf{Nu}$, then $(\mathbf{I} - \mathbf{N})\,\mathbf{y} \doteq (\mathbf{I} - \mathbf{N})\,\mathbf{X}\,\tau$, a simple linear model. If, however, $\mathbf{u} \doteq \beta\mathbf{Nu}$, then to get rid of $\mathbf{u}$, we are to form $(\mathbf{I} - \beta\mathbf{N})\,\mathbf{y} \doteq (\mathbf{I} - \beta\mathbf{N})\,\mathbf{X}\,\tau$, and we have a non-linear model in β and τ which I do not understand.

Let us suppose that a "neighbour" model presents no identifiability problems. We can fit it, and presumably, obtain purported confidence regions for true treatment differences. Can we judge whether the conclusions are valid (a heavy word and notion). The authors refer to "conventional" randomization theory. I wonder what this randomization theory is. I see that it, whatever it is, is "not entirely in accord with R. A. Fisher's ideas". I wonder what "over-conditioning" is, and what "inferentially irrelevant irregularities" are. Also, Fisher's writings are quite obscure, in that he appeared, in 1956, to reject much of his own 1935 book, and also to reject non-parametric ideas entirely. I see that the authors would like to have the unbiasedness in some Yatesian analysis of variance sense. Why should one want this? Perhaps there are good reasons. As I mentioned in my 1952 book, this does not happen in the ordinary designs and their ordinary analyses when covariates are introduced. However, this does not cause irremediable problems. One can make tests of any hypothesized treatment effects by adjusting them out, then performing a randomization test with the developed test criterion. This is the way some workers, including myself, advocate for very important experiments, and, as far as I can see, Yatesian unbiasedness, which is a curious mixture of randomization and Gauss–Markov linear model theory, does not arise. One was interested in Yatesian unbiasedness when one had no alternative to ordinary analysis of variance, which I have long taken to be, merely, an approximation to a full randomization test approach (or a re-randomization approach if one accepts the Tukey renaming). One often hears, as a supposed profoundly wise aphorism, that the scientist needs estimates not tests. Unless, perhaps, one is a fiducialist (but perhaps even then) estimation with bounds of uncertainty and testing are, essentially, two sides of a single coin; and only a foolish person fails to see both sides of a coin. Of course, in many, but by no means all, cases, the test of a single null hypothesis is not of interest, and tests of "wider hypotheses", leading to intervals are needed.

I see the efforts using neighbour data as a valid and useful attempt to increase precision of comparative experiments, and I am glad to see the present paper. There are many aspects of the effort that merit deep discussion and evaluation and further work.

I am concerned, for instance, that one may obtain apparent very high precision by model search, and one cannot, except, perhaps, by randomization tests, attach reasonable measures of uncertainty to conclusions that are obtained. Also, in general, the internal errors of a single experiment are only a part of the total error that is involved in a scientific effort.

Professor C. McGilchrist (University of New South Wales, Kensington): Although not the main thrust of this very useful paper, Wilkinson *et al.* point out the need to use a non-stationary component in modelling such data as the Mercer Hall data. Contour plots of the data, supported by the pattern of computed spatial correlations, suggest a slowly varying component which gives rise to substantial variation over the whole field. The authors regard this component as a slowly varying trend but it would be more consistent with Box–Jenkins modelling to regard the "trend" as an integrated error process. In either case differencing is appropriate to obtain a simple error covariance structure. For the ungrouped Mercer Hall data, row and column first differencing is adequate.

The Papadakis style analysis presented is a welcome addition to the analysis of field trials. It appears robust to different types of interplot correlations which it does not seriously attempt to model.

Dr J. A. Nelder (Rothamsted Experimental Station): I have interpreted the authors' "smooth trend + independent error" model as implying

$$\mathbf{z} = \delta^2\mathbf{y} = \delta^2\mathbf{X}\,\boldsymbol{\beta} + \delta^2\boldsymbol{\eta}\ , \qquad (1)$$

where y are the yields, δ^2 is the central second-difference operator with coefficients $(-\frac{1}{2}, 1, -\frac{1}{2})$, **X** is the design matrix and $\boldsymbol{\eta}$ is a vector of i.i.d. variables. If the model were correct we could write

$$\left.\begin{aligned} E(\mathbf{z}) &= \mathbf{DX}\boldsymbol{\beta}, \\ \text{var}(\mathbf{z}) &= \sigma^2 \mathbf{DD}^{\mathrm{T}}, \end{aligned}\right\} \tag{2}$$

and use generalized least-squares to estimate $\boldsymbol{\beta}$. (**D** is the matrix which takes second differences.) However, the authors point out, correctly, that this approach will yield essentially the same result as analysing the original ys with equal weights and design matrix **X**. (Note that I am not here including block effects in the model.) Now I draw a different conclusion from the authors about this equivalence. They appear to be denying the optimality of generalized least squares in this context and propose to use instead a consistent estimator given in their equation (21). By contrast I conclude that *if* (2) *is true then detrending by differencing will not have the effect desired.* However, experiments with data, particularly the Rothamsted linear experiment discussed in the paper, convince me that such differencing works, hence I conclude that (1) is wrong as a description of the data in question. Evidence that (1) is unsatisfactory may be obtained from the autocorrelations from the uniformity data given in their Fig. 2. If (1) is true then we expect $\rho_1 = -\frac{2}{3}$, $\rho_2 = \frac{1}{6}$, $\rho_j = 0$, $j > 2$ for the autocorrelations of **z**. All the estimates of ρ_1 in Fig. 2 appear to be less than $-\frac{2}{3}$ in absolute value, some markedly less. I have therefore attempted to amend (1) in the simplest way possible by adding a white-noise term on the z-scale giving

$$\mathbf{z} = \delta^2 \mathbf{y} = \delta^2 \mathbf{X}\boldsymbol{\beta} + \delta^2 \boldsymbol{\eta} + \mathbf{e}. \tag{3}$$

If we now write var $(\boldsymbol{\eta}) = \sigma'^2 \mathbf{I}$, var $(\mathbf{e}) = \sigma^2 \mathbf{I}$ and $k = \sigma'^2/\sigma^2$ we have

$$\left.\begin{aligned} E(\mathbf{z}) &= \mathbf{DX}\boldsymbol{\beta}, \\ \text{var}(\mathbf{z}) &= \sigma^2(k\mathbf{DD}' + \mathbf{I}). \end{aligned}\right\} \tag{4}$$

Now the first serial correlation is given by

$$\rho_1 = \frac{-k}{1 + 3k/2}$$

and $\rho_2 = -\frac{1}{4}\rho_1$, $\rho_j = 0$, $j > 2$ as before. Thus $\rho_1 = -\frac{1}{2}$ for $k = 2$, and generally ρ_1 goes from zero to $-\frac{2}{3}$ as k goes from 0 to ∞. We now consider the model (4) in relation to the Rothamsted linear experiment.

For fixed k we can estimate σ^2 by

$$\hat{\sigma}^2 = (\mathbf{z} - \hat{\boldsymbol{\mu}})\,\mathbf{V}^{-1}(\mathbf{z} - \hat{\boldsymbol{\mu}})/(n - p),$$

where n is the number of observations (here 36), p the number of parameters fitted and $\mathbf{V} = k\mathbf{DD}' + \mathbf{I}$.

Fitting only treatment effects, and allowing k to vary from 0 to ∞ we obtain the results of the following table:

k	S.e. $(t_0 - t_3)$	*Efficiency b'*
0	4.04	26.7
1	4.96	60.4
3	5.84	76.9
20	8.02	86.1
∞	20.34	82.0

This shows how the s.e. of $t_0 - t_3$ (taken as representative) varies with k, and also the efficiency of the estimator b' proposed by the authors in their equation (22), i.e. without block adjustment. This shows that the s.e. is sensitive to the choice of k; while the estimates of the treatment effects also change, the differences are much less striking. The efficiency of the authors' estimator can be quite low if k is small.

The estimation of k poses problems; if Normality is assumed the deviance can be minimized, it taking the form $(n-p)\,\hat{\sigma}^2 + \ln|\mathbf{V}|$. Alternatively we might try to make the actual first autocorrelation of the residuals equal that predicted by the model. The first procedure gives a value of k close to 1, while the second makes k about 2. Whichever value is chosen the standard error obtained is less than the authors', their value being presumably inflated by the relative inefficiency of their estimator. Note that in this analysis only 6 parameters are fitted, 2 being implied by the second differencing, 3 for the treatment effects, and 1 for the variance ratio k. No intercept is necessary on the z-scale and the treatments contain an aliasing in the form $\Sigma\, t_j = 0$. Thus model (2) appears to have been successful for this experiment, with only 6 parameters used and a low standard error for estimated treatment differences. However, it may not be the whole story in that the autocorrelation structure of the residuals is not specially well matched by those of the model. For $k = 2$ the values are

Order	*Auto-correlation of residuals*	*Auto-correlation predicted*
1	−0.46	−0.45
2	0.23	0.08
3	−0.26	0.03
4	0.04	0.01

The second and third autocorrelations are somewhat too large in absolute value compared with the predictions from the model.

The question of block effects needs further investigation. For $k = 0$ the introduction of block effects on the z-scale proves to be quite unnecessary, as indeed one would hope, the second differencing being all that is required. For $k = \infty$, block effects on the z-scale can be converted to trends on the y-scale and their inclusion gives a s.e. for treatment differences of about 8.0 compared with 9.0 for the randomized block analysis. For k small one would expect block effects to be unnecessary. Note that for the randomized block analysis the inclusion of blocks (on the y-scale) does not affect the variance matrix of the estimates, except by reducing σ^2. If differencing is applied to data it is not obvious what block effects should be included. If differencing is successful one would expect them to be unnecessary.

I very much hope that the ideas of this paper will be taken up and developed vigorously. We clearly need a randomization theory, as the authors themselves state, which connects NN-designs with an infinite model, by matching first and second-order moments in a way similar to that of classical randomization theory. We need, too, further models like (4) for which differencing is an effective trend-removing technique, coupled with extensions to two dimensions. I look forward with interest to these developments.

Dr D. A. Preece (Rothamsted Experimental Station): The degree of precision to be chosen for recording a variate (i.e. whether the variate values should be recorded to, for example, the nearest 5 kg/plot, the nearest 1 kg, or the nearest $\frac{1}{2}$ kg, etc.) was not mentioned at the meeting but deserves attention, this topic having been much neglected in writings on field experimentation. I have in recent years come across many examples of yield variates recorded to an insufficient degree of precision, the grouping errors being too large to be regarded as contributing just another component to experimental error (Preece, 1982, p. 212). Indeed I have often found experimenters to be unaware of the "Divide-by-Four Rule", which considers as acceptable a grouping interval that is no greater than the expected standard error per plot divided by 4 (Cochran and Cox, 1957, p. 60). (Dyke, 1974, pp. 163–164, is less stringent in being happy with a "Divide-by-Two Rule"). Neither a nearest-neighbour analysis nor any other sophisticated form of analysis will help an experimenter who has used an insufficient degree of precision of recording. But does the work reported in the authors' paper suggest to them any new thoughts on what *is* a sufficient degree of precision of recording?

Incidentally, some columns of the Wiebe data have a different degree of precision of recording from the others. I discussed this in detail elsewhere (Preece, 1981, p. 38).

The authors replied later, in writing, as follows:

We should like first to thank the many discussants for their contributions to a varied and valuable discussion of the paper. Twenty-one discussants spoke at the meeting, a record for the Society we believe, and indicative of the importance of the topic. Our reply will concentrate on several major points of interest that have emerged, and we trust that discussants will forgive us for not otherwise covering in our reply every point made in the discussion.

R.1. *The Papadakis method*

Three discussants (Besag, Ripley and Thompson) pointed out that iteration of the Papadakis estimation equations (for treatments) with a *fixed* value of the regression coefficient determines the same estimates as the equations (21) or (23) of our NN analysis, depending on whether block effects are eliminated or not. We did not ourselves notice this important equivalence because we had established that the uniterated Papadakis method was inherently inefficient with regard to minimizing error variance and that iteration led to bias, so we took no further interest in the form of the estimation equations. In the Monte Carlo studies, of course, the regression coefficient was not fixed but re-estimated in the usual way in each iteration of the covariance analysis.

The equivalence is important as evidence that a Papadakis covariance analysis is basically incorrect as an *error* analysis, the apparent inefficiency of treatment estimation being in some measure a consequential artifact. This fact, coupled with the demonstrated validity of the new form of NN analysis, answers Dr Atkinson's second point. Dr Atkinson also suggested we should have made a paired-comparison of the two methods in our Monte Carlo trials, so we need to point out that our studies of Papadakis's method were substantially complete before the new NN method was even discovered, so to speak.

R.2. *Interpretation of the sample correlograms in Fig.* 2

Ripley and Muirhead queried our interpretation of Fig. 2, drawing attention to biases in the sample autocorrelations arising from the shortness of the series involved. See also Martin's discussion. We should have mentioned that our correlations were calculated as *simple* correlations of the corresponding portions of lagged series. This definition appears to have better small-sample properties than the alternative definition (cf. Anderson, 1979) in which the sum of products in the numerator is of deviations from the mean of the whole series, rather than from the means of the two relevant subseries in which, for lag k, the first and last k values respectively are omitted, as in Kendall (1954). In particular, for an AR (1) process with autocorrelation parameter ϕ near unity, there is no swing in expectation with increasing lag to large negative correlations with the Kendall definition (K), as there is with Anderson's, (A). For instance, with $n = 15$, the expected values of the first 10 autocovariances, when normalized with respect to the expected value of the sample variance, have limiting values according to their definition, as $\phi \to 1$,

	1	2	3	4	5	6	7	8	9	10
A:	0.69	0.43	0.22	0.05	−0.08	−0.17	−0.23	−0.27	−0.28	−0.26
K:	0.75	0.52	0.34	0.19	0.08	0.02	0	0	0	0

(In the case of K the limiting expectation is zero for all lags $k \geqslant (n-1)/2$.) Thus at least in the case of the Wiebe data it is clear that an AR(1) process cannot account for the swings in Fig. 2 to large negative correlations.

R.3. *Stationary versus non-stationary models*

Several discussants (Patterson, Besag, Ripley, Martin and Muirhead) considered NN analysis with reference to stationary models, so we need to clarify our reasons for regarding a non-stationary model as inferentially necessary. The argument hinges on what is understood by the term *trend*. We take it that trend in a field experiment is an observable attribute of the actual data which, if judged highly significant with an appropriate test, is recognized as a definite source of bias in treatment comparisons if not removed. If the trend is to be removed by an agreed detrending process such as subtraction of NN means from the data, a model explaining how the removed trend might have arisen is of no further inferential relevance. Only a valid model for the detrended data is required.

However unlike a theoretically known bias whose removal presents no complications, trend in data represents a bias which has to be estimated, and the closer the data are to apparent stationarity the more the estimates of bias become relatively insignificant in magnitude, compared

with errors of estimation, so that in the limit the *observable* status of the presumed bias changes to that of a random error. In the latter situation detrending is not merely unnecessary but would lead to a loss of information. Thus in general some compromise is necessary between detrending and not detrending, and we have proposed basing the analysis on a weighted mixture of undetrended and detrended values as specified by the formula $y_i'(b) = (1-b)\,y_i + b y_i'(1)$. In the paper we choose the mixing proportions $(1-b), b$ to minimize the spatial variance of the $y_i'(b)$, but further reflection leads us to believe that it might be more appropriate to assign the value p to $(1-b)$, where p is the calculated significance probability of an appropriate test for trend. We shall examine the implications of this more fully in the future.

Given the special status of *trend* in the reasoning as outlined above, we are understandably sceptical about basing statistical inference from field experiments on stationary models. However, it may be that a stationary model, while perhaps failing to characterize trend in the field adequately, nevertheless indicates that almost complete detrending is necessary, and its relevance can then be judged according to the variance properties assigned to the detrended data.

Consider for instance the AR (1) model with autoregressive parameter ϕ. Patterson, Besag and Ripley pointed out that this process leads to essentially the same estimation equations for treatment effects as we derived. Here $b = 2\phi/(1+\phi^2)$ and it may be noted that if $\phi > 0.8$, then $b > 0.9756$, which is very close to unity. For the detrended values $-\frac{1}{2}\Delta^2 y_i$ the AR (1) process implies the following autocorrelations,

$$\rho_1' = \frac{-(1-\phi^2)+(1-\phi)^4}{4\phi\, v(\phi)}, \qquad \rho_k' = \frac{\phi^{k-2}(1-\phi)^4}{4\, v(\phi)} \quad \text{for } k \geqslant 2,$$

where $v(\phi) = 1.5 - 2\phi + 0.5\phi^2$. With $\phi = \frac{1}{2}$, the values are $-11/20$, $1/40$, $1/80$, $1/160$ . . . and the limiting values as $\phi \to 1$ are $-\frac{1}{2}, 0, 0, 0, \ldots$. Thus, for $\phi > \frac{1}{2}$, nearly all the persistent correlation structure of the AR (1) process is absorbed in generating "trend" that is removable by second differencing, the differenced data having negligible autocorrelations after the first, and indeed the proportion of total variance that is removable by second differencing tends to unity as $\phi \to 1$. It should be noted that adding an independent error term to the AR (1) model, as in Patterson's EV model, renders both the first and second correlations non-negligible in the detrended data. We may expect similar properties for detrended data to arise from almost any stationary model with initially positive and monotone-decreasing autocorrelations and thus we are led to infer that (in the absence of interplot competition) a finite correlation struction with only the first two autocorrelations nonzero would usually be adequate in practice for characterizing detrended data in the 1-D case. However, when it comes to determining a coefficient b for partial detrending, only the relative proportion of total variance attributable to trend is relevant and thus it would seem unwise to adopt a stationary model in which this proportion was artificially tied to the correlational properties of the detrended data, as in the AR (1) model considered by Besag, or Patterson's EV model.

R.4. *Weighted least squares analysis*

We noted in Section II.1 that a weighted least squares analysis of NN-detrended data $\mathbf{y}' = \mathbf{N}\mathbf{y}$ with weight matrix $(\mathbf{N}\mathbf{N}^{\mathrm{T}})^-$ is equivalent by back-transformation to an unweighted least squares analysis of the raw data $\mathbf{y}$ eliminating only an exact linear trend component within each column, so that the analysis fails to achieve the anticipated gain in efficiency from local detrending. The dilemma arises because NN-detrending is only an approximate elimination process so that the detrended values still retain in numerically codified form the original trend values, modulo exact linear trend within columns. We resolved the dilemma in one way as described in Section II.1, but Dr Patterson asks "why is it not possible to achieve the objective [an optimal analysis] by conventional least squares?" The same question is implicit in Nelder's contribution. For the reasons discussed in the previous section we limit the scope of our reply here to analysis of NN-detrended data, and for convenience to the linear 1-D case.

Inverting Patterson's question to "is it possible . . .", the answer is "yes" in a qualified sense. For if there were some perturbation of the assumed variance matrix $\mathbf{V}'$ for $\mathbf{y}'$ from the factorized form $\mathbf{V}' \propto \mathbf{N}\mathbf{N}^{\mathrm{T}}$, the equivalence of a least squares analysis of $\mathbf{y}'$ with weight matrix $(\mathbf{V}')^-$ to an unweighted analysis of $\mathbf{y}$ would be broken, and broken completely if the perturbation from the factorized form affected the variance matrix for every subset of detrended values. Presumably in these circumstances, a weighted least squares analysis of $\mathbf{y}'$ would fully retain

the benefits of local detrending. The full requirement would be that there exist no variance matrix $\mathbf{V}$ such that $\mathbf{V}' \propto \mathbf{NVN}^{\mathrm{T}}$, in relation either to $\mathbf{y}'$ or to any subset of $\mathbf{y}'$.

That such a perturbation is inferentially necessary is implied by the approximate nature of NN detrending, for in relation to the *smooth trend + independent error* model, the effects of residual trend effects in $\mathbf{y}'$ will need to be represented in the analysis by a perturbation in $\mathbf{V}'$ from the factorized form $\mathbf{NN}^{\mathrm{T}}$. If $\mathbf{V}'$ is specified by assigning non-zero values to only the first and second autocorrelations, it may only be necessary to keep these values away from $-\frac{2}{3}, +\frac{1}{6}$ respectively, in order to obtain an efficient least squares analysis of $\mathbf{y}'$.

How much more efficient such a least squares intra-N analysis would be than the form we give (assuming the same $\mathbf{V}'$) remains to be seen. We suspect that the gain would be usually only slight.

One important point is that basing estimation on unweighted treatment totals $\mathbf{T}'$ as we do in Section II.2 greatly simplifies the form of the corresponding randomization analysis derived by Davis; see Section R.6 below.

R.5. *Reduction of detrended values to deviations from replicate means*

While this reduction is not *exactly* orthogonal, as Ripley notes, the associated reduction in the average efficiency factor for treatment estimation is typically only 1 per cent or less, and usually more than offset by a corresponding reduction in residual variance of typically a few per cent. Thus we do not agree with Nelder that the adjustment is unnecessary. From a mathematical point of view it arises naturally as a consequence of the fixed replicate structure of the experimental design as an additional requirement to NN balance.

R.6. *Randomization analysis*

Davis's contribution raises many questions relating to randomization, a topic considered also or touched on by Besag, Bartlett, Martin and Kempthorne. The views expressed below are primarily those of the first author, who is indebted to Dr Davis for valuable further discussions regarding them.

Fisher viewed randomization of experiments as an essential requirement for imparting a degree of objective validity to the statistical inferences from them, in the face of uncertainties of various kinds. That he also considered a randomization analysis to be an *adjunct* to and not a substitute for model-based reasoning is clear from his first use of a randomization test on paired-comparison data, in Section 21 of *Design of Experiments* (Fisher, 1935), for here the randomization analysis is applied to the *arithmetic mean* as the relevant statistic, to demonstrate the wider validity under randomization of inferences based on the Normal distribution. Fisher appears never to have supported a purely non-parametric randomization approach to statistical analysis, of the kind to which Professor Kempthorne refers. Indeed he specifically disparages it in Section 21.1 which he added to the seventh edition (1960) of *Design of Experiments*.

However, Professor Kempthorne's opinion that Fisher may have rejected randomization analysis completely by 1956 is not correct, and probably stems from misinterpreting a sentence in Chapter 4, Section 7 *Statistical Methods and Scientific Inference* (Fisher, 1956), where the "it" in "it has no useful role to play in the formation of opinion" is intended to refer to randomization in the active sense of a deliberate mathematical act of random choice in the derivation of inferences *from given data*, not to the prior physical randomization of the associated experiment. The context, if not the sentence itself, makes this clear.

Robustness under randomization is not in itself a sufficient criterion for the inferential validity of an analysis. Professor Bartlett listed the further requirements of efficiency and relevance. In Fisherian theory, efficiency is a concomitant attribute of a fully relevant analysis in which all discernibly relevant features of the data have been taken into account (Wilkinson, 1977). In particular such an analysis would be conditioned with respect to relevant ancillary statistics. Since some of the latter may be properties of the experimental design, it is clear that model-based reasoning must play an essential role in determining not only the statistics on which a randomization analysis will be based, but also the relevant reference set of randomized designs. This is an important point which seems not to have arisen in previous work on randomization theory, so we amplify it now.

Davis correctly emphasized that an exact randomization analysis must refer to a highly restricted set of randomizations, or, in equivalent terms here, to the *subset* of all possible randomizations which agree with the realized design in respect of features deemed essential to

the analysis. In past work the relevant features have been regarded as determined by the structure of the design itself. For example, for a partially balanced square lattice design, the relevant subset from this point of view would be those randomizations in which the block relations (i.e. whether in the same block or not) of every pair of treatments are the same as in the realized design. But the more significant point is that the block relations between pairs of treatments are ancillary statistics determining the accuracy of the corresponding treatment comparisons, and they are the only such "design" ancillaries under the usual assumption of independent errors within blocks, so that the same subset of randomizations is arrived at as from design considerations alone. Thus the appropriateness of the square lattice design, as with other classical designs, is very closely tied to the *independent errors within blocks* model.

Consider now a linear 1-D intra-N analysis under a stationary model for the detrended values in which only the first two 1-D autocorrelations are allowed non-zero values. The relevant design ancillaries will then be the first- and second-level NN relations between pairs of treatments and the relevant subset of possible randomizations comprises those which have the same NN relations between treatments as in the realized design. Clearly the experimental design will need some corresponding level of NN balance if the analysis is to have any degree of robustness under randomization. Davis indicated some conditions of symmetry under randomization which he regards as necessary, and his unpublished work shows that if these conditions hold, the analysis will be robust against departures from zero of the higher-order autocorrelations in the model, as we conjectured. In practice the necessary conditions may hold only approximately, because of border effects in particular.

In our Monte Carlo studies of NN analysis we generated designs at random from all possible designs with a specified level of NN balance and, as Davis points out, a randomization analysis for this reference set would produce a spherical error distribution for treatment estimates. This result, however, is inferentially irrelevant, as the only properties one may legitimately examine over the set of all possible randomizations are those which firstly are invariant under permutation of treatment labels. In this category are the *RMS* and *F*-ratio for treatment effects from the intra-N analysis, the *EDF* and the average efficiency factor *e* for treatment estimation. However, a second invariance requirement does not hold exactly, since, in the Monte Carlo studies, *EDF* and *e* were not exactly invariant over the designs. Thus, all possible relevant subsets of randomizations as described above are not exactly equivalent under permutation of treatment labels. However, the coefficient of variation of *EDF* and *e* over designs was typically 1 per cent or less for 2×24 designs and 0.5 per cent or less for 3×24 designs, and a similar degree of approximate invariance also held for the histograms of *SED*'s (relative to nominal). We would therefore claim that our Monte Carlo studies are valid for the limited conclusions we draw from them regarding approximate unbiasedness and average efficiency. A final point here concerns the status of the trend components ξ_i when not eliminated as in intra-N analysis, a point of concern to Besag, Caliński and Davis. From a model-based point of view one would have to conceptualize each trend curve as a realization from a population of smooth trend curves. When the variance of trend at each point of the experiment becomes small relative to the independent error variance σ_η^2, the effect on the spatial correlation structure of the data would likewise be small so that in the limit one arrives at the ordinary randomized block form of analysis. On the other hand, when the trend variance is very large, the conceptual population of trends becomes irrelevant, since the realized trend itself would be removed by detrending. The essential point about randomization here is that the less that can be discerned about trend behaviour from the data themselves the more dependent the inferences become on their justification under randomization.

R.7. *Comparisons with classical analysis*

Patterson and Besag suggested that a comparison of NN analyses with classical analyses of the best available incomplete-block designs would be desirable, and we would agree. However, such a proposal raises a number of fundamental issues which we discuss now.

The first point about such comparisons is that 1-D NN analyses involving only first NN adjustments should be compared with classical analyses of 3-plot block experiments, NN analyses involving 1st and second NN adjustments compared with classical analyses of 5-plot block experiments, and so forth. Likewise in the case of 2-D NN analysis, the effect of NN adjustments with respect to first NN's in all three of the row, column and diagonal categories should be compared with the effects of adjustments for row, column and square effects in classical analyses of designs

laid out in 3 × 3 squares, and so forth. And now of course it is immediately apparent that NN analysis, with sufficiently high-order NN balance in the associated designs, has a flexibility in making optimal use of several levels and categories of NN adjustment that is not available in classical analyses which are tied to the fixed block or row–column structure of the corresponding classical forms of design. This is an important point from a practical point of view.

The next point is that for comparison, residual mean squares need to be expressed on a common basis as estimates of true residual variance. We should have done this in connection with our Table 8, since *RMS*'s from NN analysis as defined in the paper include a multiplicative variance adjustment factor such as 1.5 in a linear 1-D intra-N analysis. But if compared on this basis it is clear without empirical evidence that an optimized NN analysis will almost always determine a smaller residual variance than classical analysis because the latter does not adjust for continuous trend effects as does NN analysis, a point which Dr Patterson mentions himself.

However, the efficiency factors for treatment estimation relative to true residual variance may also differ with the two forms of analysis, and now it becomes apparent that the *nominal* efficiency factors cannot be compared, for the classical form of analysis is in fact to some degree invalid because, by virtue of the presence of continuous trend variation, the assumption of stationarity and statistical independence of errors within blocks or within rows and columns does not pertain, and in particular a classical analysis on this assumption must in fact be less efficient than the nominal efficiency factors indicate. More than this, the classical analysis must in reality be less efficient than an optimized NN analysis applied to the same data, and the degree to which the nominal efficiency factors of the classical analysis overestimate the intrinsic efficiencies will be exacerbated by the fact that the classical design of the experiment will usually have a poor degree of moving NN balance, which increases the disparity in the real accuracy of treatment comparisons, so that a concomitant lowering of real average efficiency results.

Thus the crux of the matter is that the relative accuracy of treatment estimates from NN and classical analyses must be assessed on the basis of a common model, and one which for field experiments incorporates a continuous trend component or equivalent correlation structure. From this point of view comparative studies like that in Kempton and Howes (1981), are to some extent deficient, leading in that instance to underestimation of the comparative value of NN analysis.

R.8. *Border plots*

Three discussants (Patterson, Besag, Howes) suggested that the need for border plots with NN analysis should be taken into account in comparing relative efficiencies of the NN and classical methods. However, in continuation of our remarks of the preceding section, we stress that the apparent absence of a requirement for border plots with classical designs is simply a consequence of assuming stationarity of errors within blocks. Were the classical analyses to be extended to eliminate continuously varying trend, it would be found that the accuracy of treatment comparisons involving boundary-plot treatments was substantially reduced through the absence of border plots.

On another aspect, we agree with Julian Besag that the efficiency of an NN analysis may be increased by including the information in border-plot values in the estimation equations. Of course only certain comparisons would gain in accuracy, and the gain would usually be small since the border-plot values would enter the analysis with relative variance $1 + \omega$ compared with $1 + (1 - b_s)\omega$ for the partially detrended internal-plot values. The latter variance factor tends to $1 + v$ (see Section II.5) as $\omega \to \infty$, whereas the former tends to infinity.

As Mr Howes suggests, some experiments will need border plots all around for 2-D analysis. However, one would like to be sure that the gain in accuracy from 2-D analysis more than compensated for the additional border plots. We suggest that, for the 118 wheat trials discussed in Kempton and Howes (1981), 1-D analysis with optimized adjustment for both first and second NN's might have been just as effective as their 2-D analyses. Note that there is one special form of 2-D analysis for which no border plots at all are necessary, namely when the adjustment coefficients b_R, b_C, b_D for row, column and diagonal NN means are constrained to be in the proportions $1 : 1 : -1$. In an intra-N analysis these proportions give full second degree 2-D detrending for not only the centre plot but all except the four corner plots of a 3 × 3 square (see Wilkinson and Mayo, 1982). Thus in a 2-D analysis, in which adjustment is based on a moving 3 × 3 square of neighbours, adjustment with respect to off-centre squares can be used

for all boundary values and then only the four corner values of the whole design will be left with some residual row × column interaction effect after NN adjustment. As this form of 2-D analysis will be optimal only for highly variable trend in the field, practical tests of the analysis would be necessary before a decision to omit border plots in the design on a regular basis was made.

R.9. *The Rothamsted mildew control experiment*

We shall comment here on two alternative forms of analysis that were applied to this experiment in the discussion, namely the least squares smoothing analysis of Green, Jennison and Seheult, and a modified NN analysis by Nelder. But first some comments about our own analysis.

We should have mentioned that in practice we do not rely on the theoretically determined optimal analysis alone. Our computer program does a second optimal analysis using the b value which minimizes a quadratic function determined by the three *RMS*'s of, respectively, a randomized block analysis ($b = 0$), the intra-N analysis ($b = 1$) and the first optimized analysis. Occasionally we find by this means that the first theoretically determined b value is substantially too low, as was the case with the Rothamsted experiment, where the first theoretical value was 0.81, as found also by Kempton. The second optimal value calculated by our program was 1.07. However, the corresponding analysis produced an *RMS* slightly larger than that of the intra-*N* analysis, so we settled on the latter for presentation in the paper. We also extended our analysis of the Rothamsted data to include NN adjustment with respect to second neighbours as well, but found no improvement on the simpler analysis. Quadratic detrending lost more through an increase in variance due to the adjustments than it gained from a slight reduction in $\hat{\sigma}^2_\eta$.

We greatly admire the originality of approach in the method proposed by the trio from Durham University and believe it may be useful in studies of trend behaviour. However, their criterion of optimization is not compatible with ours and appears to lead to a less satisfactory analysis of the Rothamsted experiment for which we found complete detrending by second differencing to be necessary in comparison with their smoothing compromise with $\lambda = 0.3$.

Nelder's suggestion of adding an independent component to the model for detrended values certainly has the effect of perturbing their variance matrix from the factorized form involving $\mathbf{NN}^{\mathrm{T}}$ (or $\mathbf{DD}^{\mathrm{T}}$ in Nelder's notation), and so may lead to an efficient weighted least squares analysis as noted in Section R.4. However, there appears to be no scientific basis for such a model and it also has the undesirable effect of decreasing the magnitude of both the first and second autocorrelations in direct proportion as k decreases, whereas residual trend effects would reduce the first in magnitude from $-\frac{2}{3}$ but increase the second above $\frac{1}{6}$. A check on unbiasedness with the method would be desirable, as difficulties may arise in particular with the residual degrees-of-freedom divisor.

Mr Dyke suspects that our method of detrending the Rothamsted data may "overfit short-wave variation that is more reasonably ascribed to interplot interference". That there is some partial confounding of local trend effects with those of neighbouring treatments is not denied; this accounts for the perturbation in the estimates of treatment effects from the analysis of his artificial data. However, the loss of efficiency from this source in estimating the effects of neighbouring treatments is only about 25 per cent, and thus does not account for the lack of significance we found for these effects.

R.9. *Interplot competition*

Kempton's work on allowing for interplot competition effects in NN analysis is of considerable interest and is related to unpublished work of A. G. Constantine and R. L. Correll at the CSIRO's Division of Mathematics and Statistics in Adelaide. We put forward here a simple idea for detrending data for competition studies, which we arrived at as follows for the 1-D case, with subscripts 1 and 2 for first and second NN's; First form values $y'_i = y_i + \bar{y}_{\mathcal{N}_1 i}$ to cancel the immediate competition effects and then detrend by forming $y''_i = y'_i - \bar{y}'_{\mathcal{N}_1 i}$. This simplifies to $y''_i = y_i - \bar{y}_{\mathcal{N}_2 i}$, that is, one can use second NN's for detrending and thus preserve most of the information on competitive effects in the detrended values.

R.10. *Experimental designs for NN analysis*

We hope Dr Freeman will continue his studies of NN balance, as theoretical design generators for completely and partially NN-balanced designs are needed for further mathematical work on randomization analysis. (At the Waite Institute the design problem is being studied by Dr Deborah

Street.) On practical aspects, it would certainly be possible to modify an NN analysis to allow for a more complex block structure but for maximum efficiency (not merely simplicity) it is desirable to eliminate sources of discontinuity in trend in the experiment. Unavoidable or accidental discontinuities can be course be allowed for with appropriate covariates. For 1-D analysis it is an advantage if NN adjustments do not cross the boundaries between replicates, for then the latter can be physically separated (by pathways, etc.) without affecting the NN analysis.

With reference to a comment by Besag we see no practical need for checkplots within replicates – indeed their presence would complicate the analysis unnecessarily. However, the treatments may of course include certain controls, such as standard varieties, in randomized locations.

R.11. *Comments on individual points*

Having dealt at length with several major aspects of the discussion we limit our comments here to a selection of individual points on which our views may be relevant.

Professor Pearce should note that NN designs can readily be generated for a single line of plots such as along a narrow terrace on a hillside, and if he carefully examines the empirical evidence presented in the paper he ought to feel confident that the new NN method can provide valid and efficient analyses for such experiments.

Mr Mead raised the question whether the partial-detrending coefficient b should be varied over the experiment. It should not be varied within replicates, since that degree of spatial homogeneity of the analysis is necessary to ensure robustness under randomization. Varying b from replicate to replicate could be considered, but the effect on efficiency is likely to be very small.

Mr Mead's "deprecation" needs some rebuttal. It is a fact of statistical inference that inefficient analyses can produce wrong conclusions, and in the case of the NN barley trial discussed in Section II.8, the misclassification of 3 of the top 10 lines by a randomized block analysis was quite clearly a consequence of local fertility effects which the analysis ignored. We also checked the stability of the NN analysis results by extending the analysis to 2-D adjustment. On the matter of degrees of freedom of a residual sum of squares from NN detrending, the interpretation of them that is inferentially relevant is as an indicator of the relative accuracy of estimated residual variance. In our Monte Carlo studies we found that the calculated *EDF* was in reasonably close accord with the relative variance of the *RMS*.

Various aspects of Dr Ripley's discussion have been dealt with in preceding sections, but some additional points need comment or rebuttal. On the charge of not comparing "like with like", we note that we did show, on the basis of a common, non-stationary model that the Papadakis method, uniterated, could not reduce the spatial variance of the adjusted values to the same extent as the new NN analysis; and our explanation of how bias from iteration arises is quite clear and simple. Dr Ripley is under some misapprehension about V_0, which was derived in exactly the same way in Part I as in Part II. On the data used in the Monte Carlo studies the aggregation is quite valid and not misleading. It is an inherent property of NN analysis that doubling the length of all plots, or their width, more than doubles the statistical accuracy of the results (in terms of variance) because the additional averaging of yield inherent with larger plots increases the effectiveness of smooth local detrending. Fig. 1 for the Wiebe data is based on yield contours produced by GENSTAT, which of necessity invokes a local-smoothing algorithm.

We would be interested to see Dr Martin's analyses of the Rothamsted data, particularly with regard to the standard errors assigned to treatment comparisons, as different methods can often produce similar estimates of treatment effects yet differ greatly in their assessments of accuracy. With reference to his discussion of Fig. 2, the impression "that for agricultural data there is a very slow decay in the correlations" probably derives from looking at autocorrelations calculated marginally over a whole two-dimensional array rather than conditionally as correlations within rows or columns, as would be relevant for 1-D NN analysis. That there can be very great differences between the marginal and conditional correlations is evident from comparing the following marginal autocorrelations of the W_8 data with the conditional correlations in Fig. 2:

Lag	1	2	3	4	5	6	7	8	9	10
	0.88	0.74	0.61	0.47	0.37	0.33	0.28	0.23	0.27	0.31

The strong persistence in these values is due to the substantial between-column variance in the data. Of course the distinction between marginal and conditional correlations vanishes for fully NN-detrended data.

Dr Yates referred to possible bias in treatment estimation under randomization. There is a possibility of slight bias in one sense because, in particular, boundary or border plots would retain the same treatments (with some permutation) in the subset of inferentially relevant randomizations (see Section R.6). However, it is clear mathematically that exact unbiasedness would pertain over the set of all possible randomizations. We checked on this empirically nevertheless, as Dr Yates suggested, and found no sign of any bias within the narrow limits of accuracy determined by 100 randomizations. It was not necessary to add artificial treatment effects to the uniformity data, as it is mathematically clear that the analysis would itself react additively to such changes.

In response to Dr Smith we note that NN design and analysis resolves the Student–Fisher dispute in the sense of reconciling the need for continuous local detrending with Fisher's randomization requirement. However, if there are only 2 treatments, there appears to be no more effective balanced solution with adequate randomization than the classical one of a set of randomized 2 × 2 Latin squares, in which row × column interaction variance determines the accuracy of the treatment comparisons. If the units are all in one line the design is a string of randomized sandwiches of the form ABBA. Dr Smith's interesting mathematical formulation may not have enough structure yet to recognize this particular "optimal" solution.

We are impressed with Mr Thompson's ingenuity in representing 1-D NN analysis as a combining analysis involving two passes over the data with respect to two alternative 2-plot block structures. He has used this kind of trick before to make GENSTAT analyse diallel tables and certain partially balanced designs. We look forward to seeing a numerical demonstration that the idea really works for NN analysis.

We are grateful to Dr Caliński for his remarks and particularly for the early Polish references. The difficulties he found regarding the underlying theory will, we hope, be resolved by the clarification of the theory given in Sections R.3–R.6, and a forthcoming paper on randomization aspects by A. W. Davis.

Finally we should like to reiterate thanks to all the contributors for a discussion which will clearly widen and deepen readers' understanding of nearest neighbour methods, and which has provided the authors, in particular, with valuable insights for further developments.

REFERENCES IN THE DISCUSSION

Anderson, O. D. (1976) *Time Series Analysis and Forecasting*. London: Butterworths.

——— (1979) Formulae for the expected values of the sampled variance and covariances from series generated by general autoregressive integrated moving average processes of order (p,d,q). *Sankhya* B, **41**, 177–195.

Besag, J. (1981) On a system of two-dimensional recurrence equations. *J. R. Statist. Soc.* B, **43**, 302–309.

Cochran, W. G. and Cox, G. M. (1957) *Experimental Designs*, 2nd edn. New York: Wiley.

Dyke, G. V. (1974) *Comparative Experiments with Field Crops*. London: Butterworths.

Efron, B. (1971) Forcing a sequential experiment to be balanced. *Biometrika*, **58**, 403–417.

Fisher, R. A. (1935) *Design of Experiments* (8th ed., 1966). Edinburgh: Oliver and Boyd.

——— (1956) *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.

Grundy, P. M. and Healy, M. J. R. (1950) Restricted randomization and quasi-Latin squares. *J. R. Statist. Soc.* B, **12**, 286–295.

Herzberg, A. M. and Wynn, H. P. (1982) A general approach to Yates's inter-block and intra-block analysis. *Utilitas Mathematica*, **21B**, 189–203.

Huber, P. J. (1978) Robust smoothing. In *Robustness in Statistics* (R. L. Launer and G. N. Wilkinson, eds), New York: Academic Press.

Kempthorne, O. (1952) *Design and Analysis of Experiments*. New York: Wiley.

Kempton, R. A. (1982) Adjustment for competition between varieties in plant breeding trials. *J. Agric. Sci. Camb.*, **98**, 599–611.

Kendall, M. G. (1954) Note on bias in the estimation of autocorrelation. *Biometrika*, **41**, 403–404.

Kiefer, J. and Wynn, H. (1981) Optimum balanced block and Latin square designs for correlated observations. *Ann. Statist.*, **9**, 737–757.

McBratney, A. R. and Webster, R. (1981) Detection of ridge and furrow pattern by spectral analysis of crop yield. *Int. Statist. Rev.*, **49**, 45–52.

Martin, R. J. (1982) Some aspects of experimental design and analysis when errors are correlated. *Biometrika*, **69**, 597–612.

Neyman, J. (1929) The theoretical basis of different methods of testing cereals, Part II. The method of parabolic curves. In *Pedigree Seed Cultures*, No. 4, Warsaw: K. Buszczynski & Sons.

Nelder, J. A. (1965) The analysis of randomized experiments with orthogonal block structure. I and II. *Proc. Roy. Soc.* A, **283**, 147–162 and 163–178.

Pearce, S. C. (1978) The control of environmental variation in some West Indian maize experiments. *Trop. Agric. (Trinidad)*, **55**, 97–106.

Preece, D. A. (1981) Distributions of final digits in data. *Statistician*, **30**, 31–60.

——— (1982) The design and analysis of experiments: What has gone wrong? *Utilitas Mathematica*, **21A**, 201–244.

Ripley, B. D. (1981) *Spatial Statistics*. New York: Wiley.

Załęski, E. (1927) *Metodyka Doświadczeń Rolniczych* (Methodology of Agricultural Experimentation). Lwow: Piller–Neumann.