1988

Relative accuracy of a neighbour method for field trials

W J. Lill
NSW Department of Agriculture Research Institute

A C. Gleeson
NSW Department of Agriculture Research Centre

Brian R. Cullis
NSW Department of Agriculture Research Institute, bcullis@uow.edu.au

Relative accuracy of a neighbour method for field trials

Abstract

Two scries of simulation experiments were used to investigate the accuracy of treatment and variance estimation with a neighbour analysis of field trials proposed by Gleeson & Cullis (1987). The first series examined the accuracy of residual maximum likelihood (REML) estimation of seven theoretical error models applicable to field trials. REML estimation provided accurate estimates of the variance parameters, but the Ftest of treatments was slightly biased upward (to +2·4%) for first differences models and slightly biased downwards (to –1·4%) for second differences models. The second series of simulations, based on 19 uniformity data sets, illustrated that treatment effects were consistently estimated more accurately by the REML neighbour (RN) analysis of Gleeson & Cullis (1987) than by incomplete block (IB) analysis with recovery of interblock information. The relative gain in accuracy of RN over IB depends on the amount of systematic variation or 'trend' in the trial, and ranged from 6 to 18% with an average of 12% for a range of trend and error variances commonly encountered in field trials. The predicted average standard errors of pairwise treatment differences from the RN analysis were in close agreement with their empirical estimates, indicating that the predicted average S.E.D. is approximately valid.

Keywords

accuracy, neighbour, trials, field, relative, model

Disciplines

Engineering | Science and Technology Studies

Publication Details

Lill, W. J., Gleeson, A. C. & Cullis, B. R. (1988). Relative accuracy of a neighbour method for field trials. The Journal of Agricultural Science, 111 (2), 339-346.

J. agric. Sci., Camb. (1988), **111**, 339–346. *Printed in Great Britain*

Relative accuracy of a neighbour method for field trials

BY W. J. LILL, A. C. GLEESON* AND B. R. CULLIS

NSW Department of Agriculture Research Institute, Wagga Wagga 2650, Australia
** NSW Department of Agriculture Research Centre, Tamworth 2340, Australia*

(*Revised MS. received 22 March 1988*)

SUMMARY

Two series of simulation experiments were used to investigate the accuracy of treatment and variance estimation with a neighbour analysis of field trials proposed by Gleeson & Cullis (1987). The first series examined the accuracy of residual maximum likelihood (REML) estimation of seven theoretical error models applicable to field trials. REML estimation provided accurate estimates of the variance parameters, but the F test of treatments was slightly biased upward (to +2·4%) for first differences models and slightly biased downwards (to −1·4%) for second differences models. The second series of simulations, based on 19 uniformity data sets, illustrated that treatment effects were consistently estimated more accurately by the REML neighbour (RN) analysis of Gleeson & Cullis (1987) than by incomplete block (IB) analysis with recovery of interblock information. The relative gain in accuracy of RN over IB depends on the amount of systematic variation or 'trend' in the trial, and ranged from 6 to 18% with an average of 12% for a range of trend and error variances commonly encountered in field trials. The predicted average standard errors of pairwise treatment differences from the RN analysis were in close agreement with their empirical estimates, indicating that the predicted average S.E.D. is approximately valid.

INTRODUCTION

The primary aim of most field experiments is the unbiased and efficient estimation of treatment contrasts. One of the obstacles to achieving this aim with randomization based analyses such as randomized complete blocks is the likely correlation among neighbouring plots. Incomplete block designs, first introduced by Yates (1936) in the form of lattice designs have been very popular and were made more applicable by the development of generalized lattice designs (Patterson, Williams & Hunter, 1978). These minimize the effects of correlation among plots by using small blocks. Alternative approaches include the use of neighbouring plot values as covariates (Papadakis, 1937) and modelling the error structure of the plots (Kiefer & Wynn, 1981; Martin, 1986).

Patterson & Hunter (1983) showed, in a study of 244 cereal variety trials grown in the United Kingdom since 1975, that the use of generalized lattice or alpha designs instead of complete block designs resulted in an average reduction of 30% in variance of varietal yield differences and indicated potential for further improvement using neighbour methods. Kempton & Howes (1981) estimated that Papadakis' analysis increased the precision of varietal differences among varieties of wheat by 8·7% compared with incomplete block analysis.

In neighbour analysis (Wilkinson *et al.* 1983) the plot variation is assumed to comprise two components, an underlying 'trend' (representing local conditions such as fertility, soil moisture, etc.) which changes smoothly from plot to plot, and an independent 'measurement error'. Presence of a trend will cause an association among plots. Wilkinson *et al.* (1983) noted that taking second differences of the data, namely,

$$y_i - \frac{y_{i-1} + y_{y+1}}{2},$$

where y_i is the value for the ith plot, approximately removed or at least greatly reduced the trend component.

All neighbour analyses proposed for field experiments are based on a 'trend + error' model, and each employs, at least implicitly, some level of differencing of the data to reduce the trend component to a more statistically manageable form. Examples of such analyses are given in Wilkinson *et al.* (1983), Patterson & Hunter (1983), Green, Jennison & Seheult (1985), Williams (1986), Besag & Kempton (1986) and Gleeson & Cullis (1987). These neighbour

analyses differ in the assumptions about trend, and their method of estimation.

Gleeson & Cullis (1987) pointed out that the specific models assumed for the trend in many of the proposed analyses belong to a general class of autoregressive-integrated moving average (ARIMA) models. Hence they proposed that, starting from the lowest order, a series of ARIMA models be fitted to the trend of each data set until an appropriate diagnostic check indicates that the model adequately fits the data. Gleeson & Cullis (1987) based their estimation on the residual maximum likelihood (REML) technique first proposed by Patterson & Thompson (1971). REML was shown by Cooper & Thompson (1977) to provide efficient estimates of parameters of low-order autoregressive-moving average models like those in Gleeson & Cullis (1987).

Before any neighbour method can be confidently recommended for the routine analysis of field trials, the accuracy of the treatment estimation and the unbiasedness of the variance estimation should be thoroughly examined. Unless the variance estimation is shown to be unbiased, comparison of a neighbour method with standard alternatives such as incomplete block analysis on the basis of average standard errors of pairwise treatment differences is a pointless exercise. To make such a comparison valid, there should be close agreement between the predicted average standard error of pairwise treatment differencees from the neighbour method and the empirical estimate (Besag & Kempton, 1986).

The first aim of this paper is to report on a simulation study of the performance of REML estimation of variance components of theoretical neighbour models which have been found in practice to be appropriate for a large number of field trials. The F test for treatments and the probability of falsely rejecting the null hypothesis of no treatment differences (i.e. the type I error rate) are also examined. The second aim is to use the trends of 19 sets of uniformity data as a basis for a simulation study to compare the relative accuracy of treatment estimation of the neighbour method proposed by Gleeson & Cullis (1987) and incomplete block analysis, and to examine the empirical and predicted standard errors of pairwise treatment differences for the neighbour method.

NEIGHBOUR ANALYSIS

In this section we present a brief description of the neighbour analysis proposed by Gleeson & Cullis (1987). The analysis applies to the usual two-dimensional layout of several rows of long narrow plots adjacent on their longer side. However, for notational convenience in describing the analysis we consider the case of a single row of plots.

Assuming the plots are indexed in field order, we write a neighbour model in matrix and vector notation as

$$y = D\tau + \xi + \eta,$$

where y is an n-vector of plot yields, τ is a t-vector of treatment effects, D is the corresponding $n \times t$ design matrix, ξ is an n-vector of trend effects, and η is an n-vector of measurement errors, assumed to be independent $N(0, \sigma^2)$ deviates.

The elements of ξ are assumed to follow an ARIMA (p, d, q). ARIMA (p, d, q) denotes a model for a random process whose dth differences follow an autoregressive (order p)-moving average (order q) process. Analyses of field trials suggest that models with $p = 0$ or 1, $d = 1$ or 2 and $q = 0$ will generally provide an adequate description of the random trend. Besag & Kempton (1986) assumed an ARIMA $(0, 1, 0)$ for the trend, and Green *et al.* (1985) used an ARIMA $(0, 2, 0)$ model for the trend.

After differencing the model is

$$\Delta y = \Delta D\tau + \Delta\xi + \Delta\eta$$

where Δ is an $(n-\delta) \times n$ matrix that takes differences to order d. The differenced trend has expectation 0 and variance $\sigma_t^2 V(\theta)$, say where σ_t^2 is the variance of $\Delta\xi$ and V is a matrix of correlation coefficients. For example, if $-1 < \theta < 1$, V could be

$$V(\theta) = \begin{pmatrix} 1 & \theta & \cdots & \theta^{n-d-1} \\ \theta & 1 & \cdots & \theta^{n-d-2} \\ \vdots & \vdots & \ddots & \vdots \\ \theta^{n-d-1} & \theta^{n-d-2} & \cdots & 1 \end{pmatrix}$$

This is the matrix of correlation coefficients from a Markov (first-order autoregressive) process. If the elements of $\Delta\xi$ were assumed to be independent and identically distributed as $N(0, \sigma_t^2)$, $\theta = 0$, and $V = I_{n-d}$, the identity matrix of order $n-d$.

If it is further assumed that for each plot the differenced trend is independent of the measurement error, i.e. $\mathrm{cov}(\Delta\xi, \eta) = 0$, the expectation of the differenced data is $\Delta D\tau$ and their variance is $\sigma_t^2 V + \sigma^2 \Delta\Delta' = \sigma^2 H$, say, where $H = \psi V + \Delta\Delta'$, $\psi = \sigma_t^2/\sigma^2$ and Δ' denotes the transpose of the matrix Δ.

The variance parameters to be estimated are then σ^2, σ_t^2 and θ or alternatively σ^2, ψ and θ. These parameters are estimated by REML, with one small but important modification to the original procedure described in Patterson & Thompson (1971). The degree of freedom for estimating σ^2 are taken to be $n-d-k-t+1$, where k is the number of variance parameters in H. In the models here, for example, $k = 2$ if estimating both ψ and θ, of $k = 1$ if $\theta = 0$. Subtraction of k from the degrees of freedom of the original REML procedure, $n-d-t+1$, is suggested in unpublished work by C. A. McGilchrist of the University of New South Wales, and the simulation results of this paper support his findings that the

estimation of σ^2 is more accurate with this modification.

The vector of treatment effects τ is then estimated by weighted least squares, using $\hat{H}^{-1}$ as weights, where $\hat{H}$ is the REML estimate of H. Because $\hat{H}$, and not H, is used to calculate the estimated treatment effects and hence the treatment sums of squares, the F test for treatments is approximate. This also applies to incomplete block analysis with recovery of inter-block information. Hence it is of interest to examine the closeness of the estimate of F to its expected value of $\nu/(\nu-2)$ where ν denotes the residual degrees of freedom, and also the realized probability of falsely rejecting the null hypothesis (type *I* error) and its proximity to 0·05.

SIMULATION EXPERIMENTS

Neighbour analyses of over 500 cereal variety trials conducted by the New South Wales Department of Agriculture (NSWDA) in the past 3 years suggest that trend can be adequately represented by one of three ARIMA models (0, 1, 0), (0, 2, 0) or (1, 1, 0). A test of model adequacy (Cullis, 1987) based on the first two residual autocorrelations, indicated that the simplest model for trend (0, 1, 0), was adequate for most trials.

Two series of simulation experiments were conducted, the first based on simulated ARIMA (0, 1, 0), (0, 2, 0) or (1, 1, 0) trends and the second based on trends calculated from 19 uniformity data sets.

Series 1 methods.

Four hundred simulations were generated of 4 rows × 25 columns for the seven models in Table 1. The ratios of the variance of the differenced trend, σ_t^2, to the variance of the measurement error, σ^2, were chosen to represent the range found in the analyses of the 500 field trials.

The measurement error variance, σ^2, was set to one and elements of η were generated as $N(0, 1)$ deviates using the NAG library subroutine G05DDF. The variance of the differenced trend, σ_t^2, was calculated as the product of σ^2 and ψ. To generate an ARIMA (0, 1, 0) for each row of the 4 × 25 array, a series of 25 $N(0, \sigma_t^2)$ deviates was generated and the cumulative sums calculated. For example, if we denote by ϵ_1, ϵ_2, $\epsilon_3, \ldots$, e_{25} a series of independent $N(0, \sigma_t^2)$ deviates, then the plot values of an ARIMA (0, 1, 0) trend are taken as ϵ_1, $\epsilon_1+\epsilon_2$, $\epsilon_1+\epsilon_2+\epsilon_3$, ..., $\epsilon_1+\epsilon_2+\ldots+\epsilon_{25}$. The cumulative summing is repeated to generate an ARIMA (0, 2, 0), i.e. the plot values of an ARIMA (0, 2, 0) trend are taken as ϵ_1, $2\epsilon_1+\epsilon_2$, $3\epsilon_1+2\epsilon_2+\epsilon_3$, etc. To generate an ARIMA (1, 1, 0) trend, NAG library subroutine G05EWF was used to generate a series of first order autoregressive deviates ν_1, $\nu_2, \ldots$, ν_{25}. The plot values of an ARIMA (1, 1, 0) trend were then taken as ν_1, $\nu_1+\nu_2, \ldots$, $\nu_1+\nu_2+\ldots+\nu_{25}$.

To warrant fitting a covariance structure to the differenced trend, for example fitting an ARIMA (1, 1, 0) model for trend, the variance of the differenced trend must be quite large compared to the measurement error variance, and in the example presented here $\sigma_t^2 = 10\sigma^2$.

Four replicates of a 5 × 5 balanced lattice design were imposed on the 4 × 25 array. The treatment allocation remained the same for all 400 repetitions of each of the seven models. The treatment effects added to the model were all zero, i.e. $\tau_i = 0$ for $i = 1, \ldots, 25$, so that the calculated F-statistics should average 1·029 for first differences models and 1·031 for second differences models. Under the null hypothesis of no treatment differences, critical values of $F_{24,71}$ for first differences ($\theta = 0$), $F_{24,70}$ for first differences ($\theta \neq 0$), and $F_{24,67}$ for second differences ($P = 0{\cdot}05$) were used to calculate type I error rates.

Series 1 results (Table 1)

Measurement error variance σ^2 was well estimated by REML, with a maximum bias of −3%. The variance

Table 1. *Mean REML estimates of σ^2, σ_t^2 and θ, mean estimated F values, expected F values and estimated type I error rates ($P = 0{\cdot}05$) for 400 simulations of seven neighbour models*

Trend model	Variance true σ_t^2	Variance true σ^2	Parameters estimated $\hat{\sigma}^2$	Parameters estimated $\hat{\sigma}_t^2$	Estimated F value $\hat{F}$	Estimated F value $E(\hat{F})$	Type I error rate ($P = 0{\cdot}05$) $\hat{P}$
(0, 1, 0)	1·0	1	0·97	1·0	1·054	1·029	0·083
(0, 1, 0)	0·1	1	0·99	0·10	1·039	1·029	0·060
(0, 1, 0)	0·01	1	0·97	0·018	1·036	1·029	0·050
(0, 2, 0)	0·1	1	1·00	0·10	1·017	1·031	0·055
(0, 2, 0)	0·01	1	0·99	0·010	1·020	1·031	0·038
(0, 2, 0)	0·001	1	0·99	0·0012	1·019	1·031	0·050
(1, 1, 0)	10	1	1·00	9·9	1·038	1·029	0·065
	($\theta = 0{\cdot}5$)		($\hat{\theta} = 0{\cdot}51$)				

of the differenced trend, σ_t^2, was generally well estimated by REML except for the (0, 1, 0) trend with the largest measurement error ($\sigma^2 = 100\sigma_t^2$).

Estimated F values showed a slight positive bias (up to +2·4%) for first differences models, whereas for second differences models there was a slight negative bias (down to −1·4%).

The estimated type I error rates were reasonably close to the expected value of 0·05. The average type I error rate over the four first differences models, 0·064, suggests that in practice a few too many significant differences may be indicated by neighbour analysis. For second differences models there may be slightly too few significant differences.

Series 2 methods

The second series of simulations was based on the 19 sets of uniformity data described in Table 2. Data for sets 1–10 have been collected by plant breeders in New South Wales Department of Agriculture since 1984 and data sets 11 and 12 were collected by 1978, so they should all reflect modern practices in cultural management of field trials. The others came from published uniformity data. Sets 11 and 12 are, respectively, the first four and last three rows of U.K. spring barley data in Kempton & Howes (1981). Wiebe (1935) published a uniformity data set comprising 12 rows of 31 plots of wheat in the U.S.A. To enable use of an efficient incomplete block design the last plot of each row was omitted in our study. To keep the size of each data set manageable and of a size to that generally used in later stage cereal variety trials, Weibe's data were divided into four groups of three rows of 30 plots. Similarly the wheat uniformity data of Garber, McIlvaine & Hoover (1926) were divided into three sets, each of two rows of 45 plots.

The type of incomplete block design imposed on each set (Table 2) depended on the number of plots per row. The order of priority for selecting an incomplete block design was a balanced lattice, a rectangular lattice and an alpha design (Patterson *et al.* 1978). The treatment allocation of the selected design for each set remained the same for all simulations based on that set.

The trend for each set was taken to be the values predicted from a cubic spline fitted to the data of each row of the set. A cubic spline fit to uniformity data is illustrated in Fig. 1 using the barley data of NSWDA4. These values were differenced and the variance of the differences, σ_t^2, was calculated. Three values of σ^2 were selected, equal to σ_t^2, $2\sigma_t^2$ and $10\sigma_t^2$, and three data sets generated for each of the 19 sets of Table 1 by adding independent $N(0, \sigma^2)$ measurement errors to the trend. Each of the 57 sets was simulated 200 times.

Lin, Poushinsky & Jui (1983) define a measure of within site variability as 'the percentage of environmental variability due to soil factors (R_p^2)'. For the 57 sets generated here, R_p^2 ranged from 8 to 98% with a mean of 75%.

Each set of simulated data was analysed as a randomized complete block (RCB) design, as an

Table 2. *Details of uniformity data used in Series 2 simulations*

Set	Name	Crop*	Year	Plot size (m)	Array size	IB design†
1	NSWDA1	B	1984	20 × 2	2 × 25	5 × 5BL
2	NSWDA2	B	1985	20 × 2	4 × 12	4 × 3BL
3	NSWDA3	B	1986	10 × 2	3 × 25	5 × 5BL
4	NSWDA4	B	1986	10 × 2	3 × 25	5 × 5BL
5	NSWDA5	B	1986	10 × 2	3 × 25	5 × 5BL
6	NSWDA6	W	1984	17 × 1	4 × 24	6 × 4α
7	NSWDA7	W	1985	17 × 2	4 × 24	6 × 4α
8	NSWDA8	W	1984	20 × 2	3 × 28	7 × 4α
9	NSWDA9	W	1985	20 × 2	4 × 24	6 × 4α
10	NSWDA10	W	1985	20 × 2	4 × 24	6 × 4α
11	KH1	B	1978	4·6 × 1·2	4 × 28	7 × 4α
12	KH2	B	1978	4·6 × 1·2	3 × 28	7 × 4α
13	WIEBE1	B	1927	4·6 × 1·2	3 × 30	6 × 5RL
14	WIEBE2	B	1927	4·6 × 1·2	3 × 30	6 × 5RL
15	WIEBE3	B	1927	4·6 × 1·2	3 × 30	6 × 5RL
16	WIEBE4	B	1927	4·6 × 1·2	3 × 30	6 × 5RL
17	GARBER1	W	1924	18·6 × 4·3	2 × 45	9 × 5α
18	GARBER2	W	1924	18·6 × 4·3	2 × 45	9 × 5α
19	GARBER3	W	1924	18·6 × 4·3	2 × 45	9 × 5α

* B, barley; W, wheat.

† BL, balanced lattice; RL, rectangular lattice; α, alpha design.

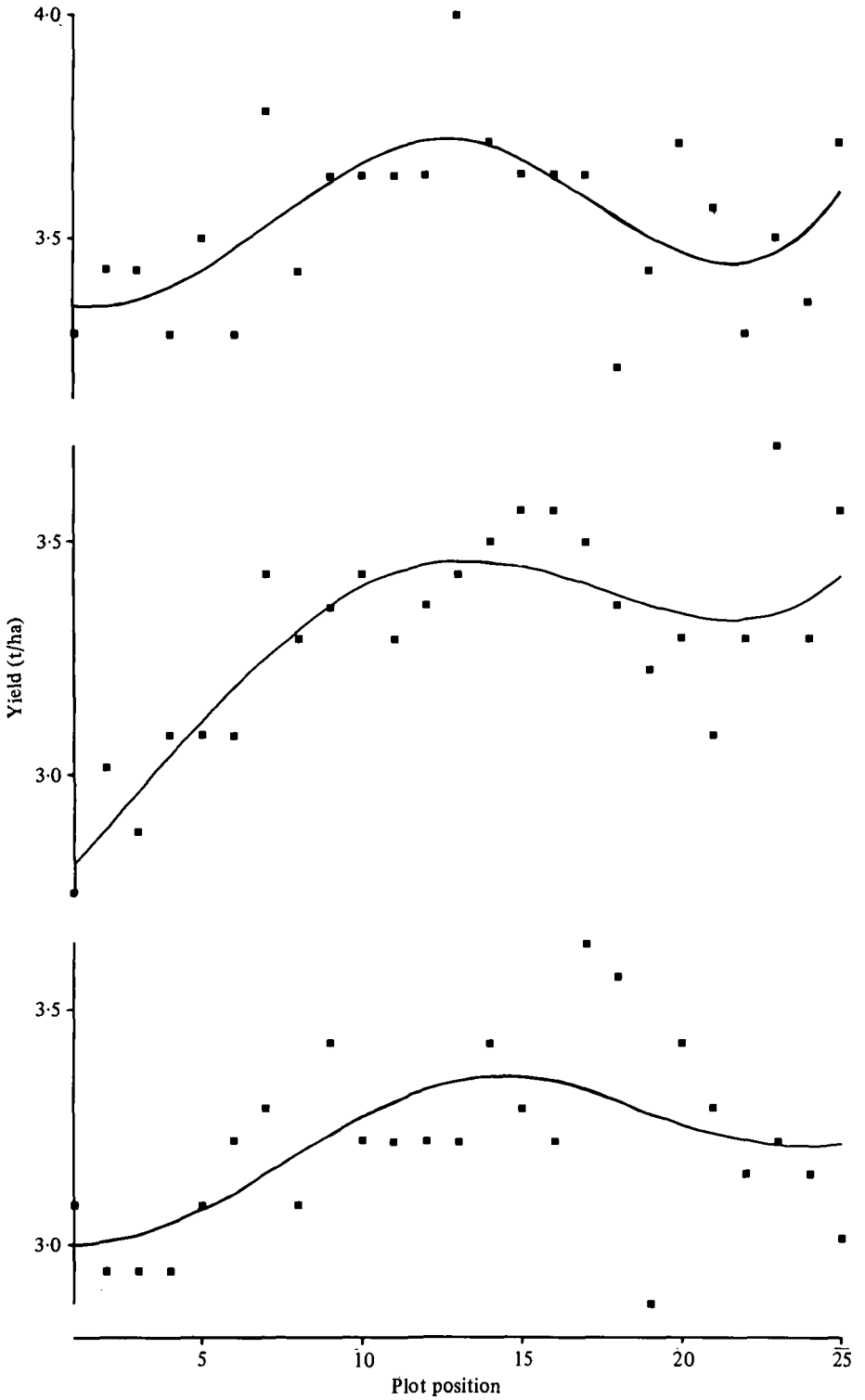

Fig. 1. Cubic splines. (—) fitted to yields (■) from a barley uniformity trial.

Table 3. *Mean inaccuracy of treatment estimation for three levels of measurement error, over 200 simulations expressed as a percentage of that for RCB analysis with the highest level of measurement error*

Data set	$\sigma^2 = 10\sigma_t^2$			$\sigma^2 = 2\sigma_t^2$			$\sigma^2 = \sigma_t^2$		
	RCB	IB	RN	RCB	IB	RN	RCB	IB	RN
1	100	92	83	74	56	41	71	50	30
2	100	99	88	67	56	44	62	48	33
3	100	55	48	93	35	23	92	32	17
4	100	72	63	85	41	30	83	36	23
5	100	63	56	91	36	28	89	31	21
6	100	99	90	65	58	44	59	49	33
7	100	91	83	72	48	38	69	41	29
8	100	91	84	74	53	42	69	47	32
9	100	91	84	75	53	41	72	47	32
10	100	77	70	82	46	35	79	40	26
11	100	76	67	83	43	32	81	36	24
12	100	70	61	89	43	29	88	37	21
13	100	76	68	85	46	33	83	41	23
14	100	82	73	82	50	36	80	44	27
15	100	75	68	85	42	33	83	36	24
16	100	71	64	88	41	30	87	36	23
17	100	89	82	78	53	41	74	46	31
18	100	96	90	71	58	45	67	49	33
19	100	76	66	87	48	33	85	42	24
Mean	100·0	81·1	73·0	80·3	47·6	35·7	77·5	41·4	26·6

incomplete block (IB) design with recovery of inter-block information and by the REML neighbour (RN) analysis of Gleeson & Cullis (1987). For each RN analysis, ARIMA trend models were fitted in the order (0, 1, 0), (1, 1, 0), (0, 2, 0), (1, 2, 0) until an adequate model was found. An adequate model was found from this set of four in every analysis. The test of model adequacy was based on the first two autocorrelations of the recursive residuals following Ansley & Newbold (1979) but using distributions derived by Cullis (1987).

The mean absolute treatment effect $\frac{1}{t}\Sigma|\hat{\tau}_i|$ was used to measure inaccuracy. Since this can only be used as a comparative measure, each is given as a percentage of the inaccuracy for an RCB analysis when $\sigma^2 = 10\sigma_t^2$.

For the RN analyses the average empirical standard error (*Emp*) and the average predicted standard error (*Pre*) of the pairwise treatment differences were calculated. For t treatments,

$$Emp = \sqrt{\left(\sum_{i=1}^{t}\sum_{j=1}^{t}\frac{(\hat{\tau}_i-\hat{\tau}_j)^2}{t(t-1)}\right)} \quad (i \neq j)$$

and

$$Pre = \sqrt{\left(\sum_{i=1}^{t}\sum_{j=1}^{t}\frac{\text{estimated variance } (\hat{\tau}_i-\hat{\tau}_j)}{t(t-1)}\right)} \quad (i \neq j).$$

Because *Emp* and *Pre* are used only as comparative measures within each of the 19 sets in Table 1, each is expressed as a percentage of *Emp* when $\sigma^2 = 10\sigma_t^2$.

Table 4. *Empirical and predicted average standard errors of pairwise treatment differences from a neighbour analysis at three levels of measurement error, over 200 simulations, expressed as a percentage of* Emp *for the highest level of measurement error*

Data set	$\sigma^2 = 10\sigma_t^2$		$\sigma^2 = 2\sigma_t^2$		$\sigma^2 = \sigma_t^2$	
	Emp	*Pre*	*Emp*	*Pre*	*Emp*	*Pre*
1	100	100	49	51	36	38
2	100	100	50	48	37	36
3	100	100	49	52	36	37
4	100	99	48	48	36	39
5	100	101	50	50	37	39
6	100	102	49	49	37	37
7	100	100	46	48	35	36
8	100	99	50	49	38	37
9	100	93	50	45	39	34
10	100	99	51	47	39	35
11	100	97	48	47	35	35
12	100	100	49	49	34	35
13	100	96	48	48	35	35
14	100	102	49	49	37	38
15	100	101	49	50	36	38
16	100	97	48	48	36	37
17	100	102	50	51	38	42
18	100	105	51	53	38	41
19	100	101	50	50	37	37
Mean	100·0	97·7	49·2	49·1	36·6	37·3

Series 2 results

The outstanding feature of Table 3 is the consistent reduction in inaccuracy or, alternatively, the consistent increase in accuracy, going from an RCB analysis to an IB analysis to an RN analysis, and the dependence of the size of the increase in accuracy on the value of σ^2. The averages for the 19 trends studied here indicate that an RN analysis increases the accuracy of treatment estimation compared with an RCB analysis by about the same amount as reducing the measurement error for the RCB analysis by a factor of 10. The RN analysis was consistently more accurate than the IB analysis. As the measurement error decreased, there was a greater advantage, in treatment accuracy, of RN over IB. The increase in accuracy ranged from an average of 8·1% for $\sigma^2 = 10\sigma_t^2$ to 14·8% for $\sigma^2 = \sigma_t^2$.

The empirical and predicted average standard errors of differences (S.E.D.S) for the RN analysis are listed, as percentages of the empirical average S.E.D. for $\sigma^2 = 10\sigma_t^2$, in Table 4. There is close agreement between the empirical and predicted average S.E.D.S with only one of the 57 pairs differing by more than 5%. A 5-fold and 10-fold decrease in measurement error resulted in a 51 and 63% reduction, respectively, in the average S.E.D. over the 19 sets. The agreement between the empirical and predicted average S.E.D.S was not affected by the changes in σ^2.

DISCUSSION

For a range of neighbour models applicable to field trials, accurate estimates of the measurement error variance, σ^2, were obtained using a slightly modified REML estimation. The modification was to reduce degrees of freedom for estimating σ^2 by a number of parameters in the correlation matrix of the (differenced) data. This reduction in the degrees of freedom improved the accuracy of the REML estimation of σ^2 and reduced the slight upward bias of the F test for first differences models.

Accurate estimates of the variance of the differenced trend, σ_t^2, were also generally obtained, the largest discrepancy occurring when σ_t^2 was much smaller than σ^2. Swallow & Monahan (1984) obtained a similar upward bias in the estimate of the block variance in incomplete block analysis, when block variance was less than $\frac{1}{5}$ of the measurement error.

For first differences models, the average upwards bias (+1·2%) of the F test and the average type I error rate of 0·064 indicate that slightly too many significant results will be obtained with RN analysis. For second differences models, there may be slightly too few significant differences.

The outstanding result of this study was the consistent increase in accuracy of treatment estimates from RN analysis compared with IB. For a fixed value of σ_t^2 the relative gain in accuracy of RN compared with IB increased as σ^2 decreased. Agricultural researchers, having chosen a trial site, have no control over the inherent within site variability which determines σ_t^2. It is possible, however, to reduce the measurement error σ^2 (by, say, improved experimental procedures). The results of this study indicate that the more experimental error can be reduced, the greater the increase in accuracy from RN compared with IB.

The close agreement between the empirical and predicted average S.E.D.s for RN analysis over the 19 data sets at three levels of measurement error indicates that the predicted average S.E.D. is approximately valid. This will enable future studies, with real experiments, of the efficiency of RN analysis compared with randomization based analyses such as RCB and IB, on the basis of predicted average S.E.D.S.

The often substantial increase in accuracy of treatment estimates suggests that agricultural researchers should use RN analysis, particularly if there is any evidence or suspicion that there may be some systematic variation across the trial site. Provided that interpretation is made realizing that, as with recovery of interblock information, slightly more than the true number of significant differences may be indicated by the F test for treatments, we believe there is sufficient evidence presented here to encourage the use of RN analysis.

Further research on neighbour methods, which will include examining the influence of outliers and finding optimal experimental designs for neighbour methods, should enhance the already substantial benefits offered by neighbour methods.

REFERENCES

ANSLEY, C. F. & NEWBOLD, P. (1979). On the finite sample distribution of residual autocorrelations in auto-regressive-moving average time series models. *Biometrika* **66**, 547–553.

BESAG, J. & KEMPTON, R. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics* **42**, 231–251.

COOPER, D. M. & THOMPSON, R. (1977). A note on the estimation of the parameters of the autoregressive-moving average process. *Biometrika* **64**, 625–628.

CULLIS, B. R. (1987). Analysis of repeated measures data from designed experiments. Ph.D. thesis, University of New South Wales.

GARBER, R. J., MCILVAINE, R. C. & HOOVER, M. M. (1926). A study of soil heterogeneity in experimental plots. *Journal of Agricultural Research* **33**, 255–268.

GLEESON, A. C. & CULLIS, B. R. (1987). Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics* **43**, 277–288.

GREEN, P. J., JENNISON, C. & SEHEULT, A. H. (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society, Series B* **47**, 299–315.

KEMPTON, R. A. & HOWES, C. W. (1981). The use of neighbouring plot values in the analysis of variety trials. *Applied Statistics* **30**, 59–70.

KIEFER, J. & WYNN, H. P. (1981). Optimum balanced block and latin square designs for correlated observations. *Annals of Statistics* **9**, 737–757.

LIN, C. S., POUSHINSKY, G. & JUI, P. Y. (1983). Simulation study of three adjustment methods for the modified augmented design and comparison with the balanced lattice square design. *Journal of Agricultural Science, Cambridge* **100**, 527–534.

MARTIN, R. J. (1986). On the design of experiments under spatial correlation. *Biometrika* **73**, 247–277.

PAPADAKIS, J. S. (1937). Methode statistique pour des expériences sur champ. *Bulletin de l'Institut d'Amélioration des Plantes à Salonique*, No. 23.

PATTERSON, H. D. & HUNTER, E. A. (1983). The efficiency of incomplete block designs in National List and Recommended List cereal variety trials. *Journal of Agricultural Science, Cambridge* **101**, 427–433.

PATTERSON, H. D. & THOMPSON, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58**, 545–554.

PATTERSON, H. D., WILLIAMS, E. R. & HUNTER, E. A. (1978). Block designs for variety trials. *Journal of Agricultural Science, Cambridge* **90**, 395–400.

SWALLOW, W. H. & MONAHAN, J. F. (1984). Monte Carlo comparison of ANOVA, MIVQUE, REML and ML estimators of variance components. *Technometrics* **26**, 47–57.

WEIBE, G. A. (1935). Variation and correlation in grain yield among 1500 wheat nursery plots. *Journal of Agricultural Research* **50**, 331–357.

WILKINSON, G. N., ECKERT, S. R., HANCOCK, T. W. & MAYO, O. (1983). Nearest neighbour (NN) analysis of field experiments (with discussion). *Journal of the Royal Statistical Society, Series B* **45**, 152–212.

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

YATES, F. (1936). A new method of arranging variety trials involving large numbers of varieties. *Journal of Agricultural Science, Cambridge* **26**, 424–455.