Journal of Applied Statistics
Vol. 00, No. 00, Month 200x, 1–101

RESEARCH ARTICLE

The Use of a Reference Variety for Comparisons in Incomplete Series of Crop Variety Trials

Johannes Forkman*

Swedish University of Agricultural Sciences
Department of Crop Production Ecology, Box 7043, SE-750 07 Uppsala, Sweden
(*Received 00 Month 200x; in final form 00 Month 200x*)

*Email: johannes.forkman@slu.se

ISSN: 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/0266476YYxxxxxxxx
http://www.tandfonline.com

In series of crop variety trials, 'test varieties' are compared with one other and with a 'reference' variety that is included in all trials. The series is typically analyzed with a linear mixed model and the method of generalized least squares. Usually, the estimates of the expected differences between the test varieties and the reference variety are presented. When the series is incomplete, i.e. when all test varieties were not included in all trials, the method of generalized least squares may give estimates of expected differences to the reference variety that do not appear to accord with observed differences. The present paper draws attention to this phenomenon and explores the recurrent idea of comparing test varieties indirectly through the use of the reference. A new 'reference treatment method' was specified and compared with the method of generalized least squares when applied to a five-year series of 85 spring wheat trials. The reference treatment method provided estimates of differences to the reference variety that agreed with observed differences, but was considerably less efficient than the method of generalized least squares.

Key words: Agricultural field trials, Crop variety evaluation, Generalized least squares, Series of experiments, Value for cultivation and use

1. Introduction

In VCU trials, Values for Cultivation and Use of new crop varieties are tested for admission to national lists. The crop varieties are tested in series of field experiments (i.e. trials), which may comprise several locations and years or seasons. The series are often analyzed statistically in two stages. At the first stage, each trial is carefully analyzed separately; the expected values of the treatments (i.e. the varieties) are estimated within each trial. At the second stage, the series is analyzed using the estimates from the first stage as observations. This two-stage analysis of series of experiments is associated with fewer computational problems than the one-stage analysis [26] and is convenient when the experiments have varying experimental designs. In addition, in an automated system for routine analysis some of the computer algorithms from the first stage may be re-used at the second stage. The present paper discusses the second stage of the two-stage analysis of series of variety trials.

The dataset for the second stage of the analysis is an $I \times J$-table of observations from I varieties and J trials. Often the table has missing cells, especially when the series span several years, because every year new varieties enter the trials and older, less performing, varieties are phased out. A 'reference variety' is regularly included in all trials. This makes the series connected and facilitates construction of summary tables showing differences between the $I - 1$ 'test varieties', as appropriate for comparative experiments. The expected values of the varieties are usually estimated by use of linear mixed models and generalized least squares [28]. In the simplest case, variety is modeled as a fixed factor and trial as a random factor; the variance components are estimated using the residual maximum likelihood (REML) method [21, 22]. This procedure for analysis of incomplete series of variety trials, which is closely related to the method of fitting constants [32], is efficient [19].

In routine analysis the method of generalized least squares frequently gives apparently strange estimates. The following phenomena may occur:

- In a series of variety trials, the estimate of the difference between a test variety and the reference variety may be outside the range of the differences observed within the trials.
- In a series of variety trials comprising many years, the estimate of the difference between a test variety and the reference variety may not accord with previously reported yearly estimates of differences.

These cases are problematic, because they are hard to understand and easily suspected to be results of incorrect calculation. Stakeholders with genuine interest in trials have difficulty in accepting these phenomena. The present paper highlights and discusses the above mentioned problems in routine analysis of incomplete series of variety trials.

The present study was also motivated by a request for a straightforward method feasible for a new international database application. In this application, a table that summarizes all observed results within selected regions, which may comprise locations from different countries, should be constructed. It should be possible for users to search for results from trials that include a variety of specific interest. Through such selection, the variety of interest would be included in all trials and could therefore be used as a reference variety. Average differences between other varieties and the specific variety of interest could be presented. The other varieties would be compared between themselves through the reported differences to the reference variety. Statistics would be needed to assess the approximate precision in

these comparisons.

The generalized least squares method produces a ranking of the treatments: If Treatment A is better than Treatment B and Treatment B is better than Treatment C, then Treatment A is better than Treatment C. A ranking could also be accomplished by measuring performance of test varieties by differences to a reference variety. This procedure is known to be less efficient than the generalized least squares method, since it does not use all available information [20]. The present paper re-explores the old idea of comparing test varieties through a reference variety [25], but uses a novel method for calculation and compares this 'reference treatment method' with the method of generalized least squares. In Section 2, an example shows that the reference treatment method provides estimates that would easily have been accepted, although much less efficient than the generalized least squares method.

In short, the reference treatment method is carried out in the following steps:

(1) Estimate the variance components of the statistical model using the REML method.
(2) Within each experiment and for each test treatment, calculate the difference between the test treatment and the reference treatment.
(3) Calculate mean differences. When appropriate, use estimates of variance components for weighting.
(4) Calculate standard errors in mean differences and in differences between mean differences.

Through the method of generalized least squares, means are adjusted so that they can be compared with each other, although some observations are missing [19]. This is efficient, but also the source of problematic discrepancies between observed and estimated differences. With the reference treatment method this problem is, on the cost of efficiency, avoided by ignoring covariances. Specifically, in the variance-covariance matrix $\mathbf{V_d}$ of the observed differences to the reference variety, covariances between differences pertaining to different test varieties are zeroed. In this way, the observed mean difference between a test variety and the reference variety is never adjusted due to correlations with test varieties that were experimented in other trials.

Section 2 presents analyses of a five-year series of 85 spring wheat trials with 13 varieties. Analyses were made for each year as well as for the whole series comprising all years, using the reference treatment method as well as the method of generalized least squares. This example shows that the method of generalized least squares produces estimates that appear to be incorrect, since the estimates are not weighted averages, as often expected. The efficiency of the reference treatment method as compared with the method of generalized least squares is quantified. Section 3 specifies the reference treatment method, and Section 4 discusses the methods and the example.

Yates and Cochran [32] and Cochran [5] discussed methods for summarizing estimates from several experiments and proposed weights closely related to those used in Section 3. Nabugoomu, Kempton and Talbot [18] employed the same basic models as the present paper, with random experimental effects and possible grouping of experiments, when considering series of variety trials with heterogeneous variety-by-location variance. The above mentioned problems with apparently strange estimates have not been much noted in the literature [9], although many authors have proposed various extensions of basic linear mixed models ([6], [10], [12], [27], [29]). Bathke et al. [1] derived asymptotic methods for analyzing large series of randomized block experiments with non-normal data.

Caliński et al. [2–4] studied single stage analyses of heteroscedastic observations from complete $I \times J$-tables without grouping of the experiments, whereas the present paper considers second stage analyses of homoscedastic observations from incomplete $I \times J$-tables with possible grouping of environments. At the first stage of the two-stage analysis, experiments are analysed separately according to their specific experimental designs. Weights can be used at the second stage of the two-stage analysis for separating effects of treatment-by-experiment interactions from effects of intra-experimental plot errors [26]. Welham et al. [30] compared the two stage analysis of series of crop variety trials with the one-stage analysis, assuming different models, and concluded, based on a simulation study, that the unweighted two-stage approach performed poorly. On the other hand, Möhring and Piepho [17] compared two-stage analyses using different weighting methods with the one-stage analysis and found that the unweighted two-stage analysis gave reasonable estimates, although weighted two-stage analyses performed better.

2. Example: A series of spring wheat variety trials

As an example, a series 85 VCU trials in spring wheat, carried out during the five-year period 2002–2006, is analyzed. The locations for the trials were not fixed during the years, because the trials were conducted on farmers' fields and not on permanent experimental stations. Section 2.1 illustrates the apparent problems mentioned in the Introduction, and Section 2.2 demonstrates the reference treatment method.

2.1 *Results with the method of generalized least squares*

After each season, the model

$$y_{ij} = \alpha_i + t_j + e_{ij} \tag{1}$$

was fitted to the observed yields, i.e. to the estimated averages from the single trials performed that year. In Equation (1), y_{ij} denote the yield of variety i, $i = 1, 2, \ldots, I$, in trial j, $j = 1, 2, \ldots, J$, and α_i denotes the expected yield of the ith variety, t_j is a random effect of the jth trial, and e_{ij} is a random error. The random effects and the random errors were assumed to be independent and normally distributed with expected value 0 and variances σ_t^2 and σ_e^2, respectively. These variances were estimated using the REML method and the mixed procedure of the SAS System [23].The expected values α_i, $i = 1, 2, \ldots, I$, with $I = 6$ in 2002, $I = 7$ in 2003 and 2004 and $I = 13$ in 2005 and 2006, were estimated using generalized least squares, that is by

$$\breve{\boldsymbol{\alpha}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y} \ , \tag{2}$$

where $\mathbf{y}$ is the vector of observed yields, $\mathbf{X}$ is the design matrix indicating the varieties, and $\hat{\mathbf{V}}$ is the estimated covariance matrix of $\mathbf{y}$.

The estimated differences between the expected values of the test varieties and the reference variety 9601 were calculated and reported for each year as in Table 1, columns 2002–2006. Well-informed readers of Table 1 may react to the estimates presented in these columns, and suppose that the dataset or the analysis is incorrect. For example, Variety 20549, which was included in a single trial in 2005 and in three trials in 2006, yielded 336 kg/ha less than the reference variety in the

single trial performed in 2005, but the estimated difference was −108 kg/ha. In the report (not shown) of the analysis of the only experiment performed in 2005 that included Variety 20549, the estimate of the difference between Variety 20549 and the reference variety was reported as −336 kg/ha. No other trials in 2005 studied Variety 20549. Nevertheless, when analysed together with other trials of that year, the difference was estimated to be −108 kg/ha. This estimate does not fully agree with the observed difference in the trial.

Table 1. Generalized least-squares estimates of differences (kg/ha) to the reference variety 9601, by year and for the complete period 2002–2006. Numbers of trials (N), pairwise tests and standard errors of the differences (SED) refer to the complete period. The reference variety mean was estimated to be 6571 (SE 299) kg/ha. Varieties connected by the same letter are not significantly different on level 5%.

Variety	N	2002	2003	2004	2005	2006	2002–2006		SED
20303	19		518	484	666	580	520	*a*	85
20512	9				289	523	415	*ab*	115
20501	11				334	347	367	*ab*	107
20212	34	173	318	304	432	394	293	*b*	72
20122	77	198	306	440	228	32	242	*b*	59
20211	66	306	210	204	336	150	237	*b*	61
20511	23				305	78	224	*b*	87
9602	79	164	412	277	96	25	195	*b*	58
9601	85	0	0	0	0	0	0	*c*	0
20513	16				-83	-37	-6	*c*	97
20125	69	-664	-361	-550	-542	-535	-517	*d*	60
20549	4				-108	-1053	-737	*d*	161
20548	6				-1156	-1393	-1233	*e*	135

Table 2. Reference treatment method mean differences (kg/ha) to the reference variety 9601, by year and for the complete period 2002–2006. Numbers of trials (N), pairwise tests and standard errors of the differences (SED) are given for the the complete period. The reference variety mean was estimated to be 6570 (SE 299) kg/ha. Varieties connected by the same letter are not significantly different on level 5%.

Variety	N	2002	2003	2004	2005	2006	2002–2006		SED
20303	19		474	426	607	657	514	*a*	100
20512	9				236	464	340	*ab*	143
20212	34	105	273	304	373	472	284	*b*	80
20122	77	198	332	440	228	32	242	*b*	59
20211	66	238	235	204	336	150	232	*b*	62
20501	11				216	197	207	*bc*	132
9602	79	193	411	277	96	25	203	*b*	59
20511	23				301	50	165	*bc*	102
9601	85	0	0	0	0	0	0	*cd*	0
20513	16				-135	-118	-124	*d*	117
20125	69	-699	-361	-559	-551	-535	-524	*e*	61
20549	4				-336	-1044	-855	*e*	205
20548	6				-1187	-1384	-1286	*f*	170

At the end of the five-year period the whole series was analyzed using the model

$$y_{ijk} = \alpha_i + a_j + g_{ij} + t_{jk} + e_{ijk} \ . \tag{3}$$

In Equation (3), y_{ijk} denotes the yield of variety i, $i = 1, 2, \ldots, 13$, in the kth trial, $k = 1, 2, \ldots, K_j$, of year j, $j = 1, 2, \ldots, 5$. Furthermore, α_i is the expected value of the ith variety, a_j is a random effect of the jth year, g_{ij} is a random interaction between the ith variety and the jth year, t_{jk} is a random effect of the kth trial in the jth year, and e_{ijk} is a random error. The trials were nested within years, and the interaction between varieties and trials was included in the error term e_{ijk}. The random effects were assumed to be independent and normally distributed with expected value 0 and variances σ_a^2, σ_g^2, σ_t^2 and σ_e^2, respectively.

Using the mixed procedure of the SAS System and the REML method, the variance components σ_a^2, σ_g^2, σ_t^2 and σ_e^2 were estimated to be 341 589, 3835, 1 634 977, and 75 074, respectively. The expected values α_i, $i = 1, 2, \ldots, 13$, were estimated using the method of generalized least squares.

The eighth column of Table 1 gives the estimates of the differences to the reference variety. The estimates of the five-year series do apparently not accord with previously reported yearly estimates. For example, Variety 20513 was estimated to give 83 and 37 kg/ha lower yield than the reference variety in 2005 and 2006, respectively. The other years, this variety was not experimented. When all five years were analyzed together, the expected long-term difference was estimated to be −6 kg/ha. In this case, the sign of the estimate of the five-years series was luckily the same (negative) as the sign of the yearly estimates. When this is not so, or when the discrepancy is larger, careful readers become suspicious of the validity of the report.

It is customary to report the number of trials, as in Table 1. For example, Variety 20513 was included in 16 trials. However, the estimate, −6, of the difference to the reference variety, was not based on 16 trials only, as one might think.

In 2005, Variety 20501 was included in six trials and produced in average 216 kg/ha higher yield than the reference variety. In 2006, Variety 20501 was included in five trials and produced in average 197 kg/ha higher yield then the reference variety (not shown in Table 1). Through the generalized least squares analyses, the differences were estimated to be 334 and 347 kg/ha, for 2005 and 2006, respectively (Table 1). These estimates were noticeably higher than the observed mean differences. Rather than summary statistics, Table 1 comprises efficient estimates of differences in treatment effects. When all years were taken into account, the estimate was adjusted further upwards, to 367 kg/ha, which is significantly different from 0 (Table 1). Thus, a systematic long-term difference between Variety 20501 and the reference variety was detected. This might be surprising considering that Variety 20501 was only tested in two years and differences between varieties can vary much between years, as a consequence of varying weather.

The general F-statistic [16],

$$F = \frac{\breve{\boldsymbol{\alpha}}'\mathbf{L}'(\mathbf{L}\breve{\mathbf{C}}\mathbf{L}')^{-1}\mathbf{L}\breve{\boldsymbol{\alpha}}}{R} \ , \tag{4}$$

for the hypothesis of no differences between the expected values of the varieties was calculated as 34.15, which may be compared with an F distribution with 12 and 29 degrees of freedom ($P < 0.001$). In Equation (4), $\breve{\boldsymbol{\alpha}}$ is defined as in Equation (2), $\breve{\mathbf{C}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$, $R = 12$ and $\mathbf{L}$ is a matrix of 12 orthogonal contrasts. In the numerator of the F-distribution, the number of degrees of freedom is 12, because twelve orthogonal contrasts are tested. In the denominator, the number of degrees of freedom is 29, since the variety-by-year interaction has 29 degrees of freedom. In the balanced case, with no missing cells and with the same number of trials every year, the expected value of the mean sum of squares for varieties equals the expected value of the mean sum of squares for the variety-by-year interaction. This applies provided that the null hypothesis is true, regardless of the number of trials per year. However, in the unbalanced case, the F-test is approximate. Table 1 includes standard errors for the differences estimated in the five-year series, each calculated from variances and covariances in $\breve{\mathbf{C}}$, using $(\mathbf{L}'\breve{\mathbf{C}}\mathbf{L})^{1/2}$, where now $\mathbf{L}$ is the row vector such that $\mathbf{L}\breve{\boldsymbol{\alpha}}$ is the estimated difference. The last but one column of Table 1 shows the results of pairwise testing without adjustments for multiple comparisons. These results were obtained by comparing the values of the general

t-statistics $\mathbf{L}\check{\boldsymbol{\alpha}}(\mathbf{L}\check{\mathbf{C}}\mathbf{L}')^{-1/2}$ with the t distribution with 29 degrees of freedom [23]. The SAS macro pdmix800 [24] was employed for this computation.

In all statistical analyses, residuals were plotted against predicted values and inspected in normal probability plots. No outliers or systematic patterns were found.

2.2 *Results with the reference treatment method*

An analysis using the reference treatment method as detailed in Section 3 was carried out on the spring wheat dataset. Table 2 reports estimated differences and results of pairwise testing. Note that the order of the varieties is different from Table 1, because the methods rank the varieties differently. The yearly results were obtained using variance estimates from a fit of Model (1), whereas the results of the five-years series were obtained using variance estimates from a fit of Model (3). The same variance estimates were used for the reference treatment method as for the generalized least squares method.

The difference estimates obtained using the reference treatment method are averages. Note specifically that, in contrast to the generalized least squares analysis, the expected difference between Variety 20549 and the reference variety was estimated to be −336 kg/ha in 2005. This was expected, because the only observed difference in yield was −336 kg/ha. Thus, the analysis of the complete series of trials performed in 2005 gave exactly the same estimate as the analysis of the single trial including Variety 20549. In the analysis of the five-years series, the difference between Variety 20549 and the reference variety was estimated to be −855 kg/ha, which is a weighted average of the estimates −336 from 2005 and −1044 kg/ha from 2006. The second estimate, which was based on three trials, gained more weight than the first, which was based on a single trial.

For Variety 20513, the estimated difference in the five-year series is −124 kg/ha, which is a weighted average of the estimates −135 and −118 kg/ha that were obtained for 2005 and 2006, respectively. Unlike the generalized least squares method, the reference treatment method never gives estimates of differences to the reference variety that are not simple or weighted averages of differences observed within trials. The estimate, −124 kg/ha, of the difference to the reference variety, was indeed based on 16 trials, as reported.

Variety 20501 was the third best performing variety according to the generalized least squares analysis (Table 1), but only the sixth best according to the reference treatment method . The estimates 216 and 197 kg/ha, reported in Table 2, are simple averages of within-trial differences between Variety 20501 and the reference variety. In the analysis of the five-years series, the difference was estimated to be 207 kg/ha, which is a weighted average of 216 and 197 kg/ha. This estimate, 207 kg/ha, was not significantly different from 0 (Table 2).

Expected differences were estimated as Equations (6) and (14), and the standard errors of the differences as the square roots of Equations (10) and (18), for yearly results and the five-years series, respectively. Equivalently, expected differences may be estimated as Equation (28), and standard errors as the square roots of diagonal elements of $\mathbf{S}$ in Equation (31); this gives the same result. The F-statistic (33) was 26.4 on 12 and 29 degrees of freedom ($P < 0.001$), indicating differences between varieties.

2.3 *Comparison of efficiency*

The reference treatment method was less efficient than the generalized least squares method, as seen by comparing approximate standard errors of differences (SED)

Table 3. Ratio of SED with the method of generalized least squares to SED with the reference treatment method, for all pairs of varieties

Variety	20125	20211	20212	20303	20501	20511	20512	20513	20548	20549	9601	9602
20122	0.98	0.98	0.90	0.85	0.81	0.85	0.80	0.83	0.80	0.78	0.99	0.98
20125		0.99	0.91	0.86	0.82	0.86	0.81	0.83	0.80	0.79	0.98	0.97
20211			0.92	0.86	0.82	0.86	0.81	0.84	0.80	0.79	0.98	0.94
20212				0.92	0.81	0.81	0.81	0.81	0.76	0.76	0.90	0.87
20303					0.82	0.82	0.84	0.83	0.76	0.76	0.85	0.85
20501						0.90	0.87	0.89	0.82	0.79	0.81	0.81
20511							0.89	0.95	0.79	0.79	0.85	0.85
20512								0.92	0.75	0.75	0.80	0.80
20513									0.80	0.80	0.83	0.83
20548										0.93	0.80	0.80
20549											0.78	0.78
9601												0.99

of Table 2 with those of Table 1. The SED are approximate since variances in estimates of variance components were ignored. As a consequence of larger SED, the reference treatment method produced significant differences less often than the generalized least squares method. The reference treatment method was particularly inefficient for comparisons between test varieties that were not tested all years. This is shown by Table 3, which presents ratios between SED using the generalized least squares method to SED using the reference treatment method. These ratios vary from 0.75 to 0.99. Efficiency is usually measured as ratio in variance (i.e. by squared SED ratios). In this example, the efficiency of the reference treatment method to the generalized least squares method varied from 0.56 to 0.99. Efficiency measured as the ratio of the average variances (averages over all comparisons) was 0.67. For pairs of varieties that were tested in all five years, the reference treatment method was almost as efficient as the generalized least squares method. For pairs of varieties that were missing some years, the efficiency of the reference treatment method was considerably lower.

The reference treatment method as specified in Section 3 takes into account weights, w_{ij} and u_j, which are inverses of variances. These are defined in Equations (13) and (15), respectively. To investigate the importance of these weights in the present example, a calculation was made using equal weights, i.e. $w_{ij} = 1$ and $u_j = 1$ for all i and j. Applying Equation (C1) in Appendix C, the efficiency of this equal-weights method to the reference treatment method was 0.90, measured as the ratio of average variances. The efficiency of the equal-weights method to the generalized least squares method was 0.60.

3. Specification of the reference treatment method

Section 3.1 specifies the reference treatment method for series with independent experiments, Section 3.2 considers series of experiments that can be classified according to for example points of time or locations, and Section 3.3 derives the method for general linear mixed models. Within each section, equations for variances and covariances of weighted mean differences are derived, making possible assignments of standard errors of differences, confidence intervals and probability values.

3.1 *Series with independent experiments*

Let y_{ij} denote an observation from treatment i, $i = 1, 2, \ldots, I$, in experiment j, $j = 1, 2, \ldots, J$. Let

$$y_{ij} = \alpha_i + t_j + e_{ij} \ , \tag{5}$$

where α_i denotes the expected value of the ith treatment, t_j is a random effect of the jth experiment, and e_{ij} is a random error. The random effects and the random errors are assumed to be independent and normally distributed with expected value 0 and variances σ_t^2 and σ_e^2, respectively.

The series comprises I treatments and J experiments, but y_{ij} may not have been observed for all pairs of i and j. Let subscript $i = c$ if the treatment is the reference treatment. For all i and j such that y_{ij} has been observed, let $d_{ij} = y_{ij} - y_{cj}$. Let $\mathcal{J}(i) = \{i : d_{ij} \text{ exists}\}$, and let J_i be the number of elements in $\mathcal{J}(i)$. Thus, the number of experiments including treatment i is J_i. The difference δ_i between the expected value α_i of the ith treatment and the expected value α_c of the reference treatment can be estimated by

$$\delta_i^\star = \frac{1}{J_i} \sum_{j \in \mathcal{J}(i)} d_{ij} \ , \tag{6}$$

which is distributed as $\mathrm{N}(\alpha_i - \alpha_c, 2\sigma_e^2/J_i)$ when $i \neq c$. Using Model (5) and the reference treatment method, the expected value α_c of the reference treatment is estimated by

$$\alpha_c^\star = \frac{1}{J} \sum_{j=1}^{J} y_{cj} \ , \tag{7}$$

and the expected value α_i of treatment i, $i \neq c$, is estimated by

$$\alpha_i^\star = \alpha_c^\star + \delta_i^\star \ . \tag{8}$$

PROPOSITION 3.1 *Let J_{pq} be the number of experiments that include both treatments p and q. The covariance between $\alpha_p^\star$ and $\alpha_q^\star$, as defined by Equations* (7) *and* (8)*, is*

$$cov(\alpha_p^\star, \alpha_q^\star) = \begin{cases} \dfrac{\sigma_t^2 + \sigma_e^2}{J} \ , & \text{if } p = q = c \\ \dfrac{\sigma_t^2}{J} + \Big(\dfrac{2}{J_p} - \dfrac{1}{J}\Big)\sigma_e^2 \ , & \text{if } p = q \neq c \\ \dfrac{\sigma_t^2}{J} - \Big(\dfrac{1}{J} - \dfrac{J_{pq}}{J_p J_q}\Big)\sigma_e^2 \ , & \text{if } \begin{cases} p \neq q \\ p \neq c \\ q \neq c \end{cases} \\ \dfrac{\sigma_t^2}{J} \ , & \text{if } \begin{cases} p \neq q \\ q = c \ . \end{cases} \end{cases}$$

Proof See Appendix A. ■

COROLLARY 3.2 *The variance in the estimated difference between two test treatments is*

$$var(\alpha_p^\star - \alpha_q^\star) = \Big(\frac{1}{J_p} + \frac{1}{J_q} - \frac{J_{pq}}{J_p J_q}\Big)\, 2\,\sigma_e^2 \ , \tag{9}$$

where $p \neq q$, $p \neq c$ and $q \neq c$, and the variance in the estimated difference between a test treatment and the reference treatment is

$$var(\delta_i^\star) = \frac{2\sigma_e^2}{J_i} \ . \tag{10}$$

In applications, the variance components σ_t^2 and σ_e^2 are unknown. These may be estimated using the REML method, which, in balanced cases, gives the same estimates, when positive, as the traditional ANOVA method [22]. Let $\hat{\sigma}_e^2$ be the REML estimate of σ_e^2. The ratio between the estimate, Equation(6), of the difference $\alpha_i - \alpha_c$, $i \neq c$, and the standard error $2\hat{\sigma}_e^2/J_i$ may be compared with a t distribution with $N - I - J + 1$ degrees of freedom, where N is the total number of observations. Similarly, the ratio between the difference $\alpha_p^\star - \alpha_q^\star$, where $p \neq q$, $p \neq c$ and $q \neq c$, and the square root of Equation (9), with $\hat{\sigma}_e^2$ substituted for σ_e^2, may be compared with a t distribution with $N - I - J + 1$ degrees of freedom.

3.2 *Series with experiments in groups*

When the experiments can be classified into groups, for example groups of experiments conducted at about the same point of time or at the same location, the following model is useful. Let y_{ijk} denote an observation from treatment i, $i = 1, 2, \ldots, I$, in the kth experiment, $k = 1, 2, \ldots, K_j$, of group j, $j = 1, 2, \ldots, J$. Let

$$y_{ijk} = \alpha_i + a_j + g_{ij} + t_{jk} + e_{ijk} \ . \tag{11}$$

In Equation (11), α_i denotes the expected value with the ith treatment, a_j is a random effect of the jth group, g_{ij} is a random interaction between the ith treatment and the jth group, t_{jk} is a random effect of the kth experiment in the jth group, and e_{ijk} is a random error. The random effects and interactions are assumed to be independent and normally distributed with expected value 0 and variances σ_a^2, σ_g^2, σ_t^2 and σ_e^2, respectively.

Define d_{ijk} as the difference between the observations of the treatments i and c in the kth experiment in group j, whenever these observations exist. Thus $d_{ijk} = y_{ijk} - y_{cjk}$. Let $\mathcal{K}(i,j) = \{k : d_{ijk} \text{ exists}\}$ and K_{ij} be the number of experiments with treatments i (and c) in group j (i.e. the number of elements in $\mathcal{K}(i,j)$). The mean difference over experiments of group j is normally distributed:

$$\frac{1}{K_{ij}} \sum_{k \in \mathcal{K}(i,j)} d_{ijk} \sim \mathrm{N}\Big(\alpha_i - \alpha_c \ , \ 2\big(\sigma_g^2 + \frac{\sigma_e^2}{K_{ij}}\big)\Big) \ . \tag{12}$$

Define weights $\{w_{ij}\}$ as

$$w_{ij} = \begin{cases} \dfrac{1}{2(\sigma_g^2 + \sigma_e^2/K_{ij}))} & \text{if} \quad K_{ij} > 0 \\ 0 & \text{if} \quad K_{ij} = 0 \ . \end{cases} \tag{13}$$

Let $\mathcal{J}(i) = \{j : d_{ijk} \text{ exists for some } k\}$. Thus $\mathcal{J}(i)$ is the set of all subscripts for groups with experiments including treatment i. Then

$$\delta_i^\star = \frac{1}{\sum_{j=1}^J w_{ij}} \sum_{j \in \mathcal{J}(i)} \frac{w_{ij}}{K_{ij}} \sum_{k \in \mathcal{K}(i,j)} d_{ijk} \tag{14}$$

is a pooled estimate of $\delta_i = \alpha_i - \alpha_c$. The expected value α_c of the reference treatment can be estimated by

$$\alpha_c^\star = \frac{1}{\sum_{j=1}^J u_j} \sum_{j=1}^J \frac{u_j}{K_j} \sum_{k=1}^{K_j} y_{cjk} \ , \tag{15}$$

where $u_j = (\sigma_a^2 + \sigma_g^2 + \sigma_t^2/K_j + \sigma_e^2/K_j)^{-1}$, and the expected value of α_i, $i = 1, 2, \ldots, I$, may be estimated by

$$\alpha_i^\star = \alpha_c^\star + \delta_i^\star \ . \tag{16}$$

Proposition 2 gives the variance of $\alpha_i^\star$, $i = 1, 2, \ldots, I$, and the covariances between $\alpha_p^\star$ and $\alpha_q^\star$ when $p \neq q$.

PROPOSITION 3.3 *Let K_{pqj} denote the number of experiments, in the jth group, that include the treatments p and q. The covariance $cov(\alpha_p^\star, \alpha_q^\star)$ between $\alpha_p^\star$ and $\alpha_q^\star$, as defined by Equations* (15) *and* (16) *is*

$$
\begin{cases}
\dfrac{1}{\sum_{j=1}^{J} u_j}\ , & \text{if } p=q=c \\
\dfrac{1}{\sum_{j=1}^{J} u_j}+\dfrac{1}{\sum_{j=1}^{J} w_{pj}} & \\
\quad +\dfrac{2}{\sum_{j=1}^{J} w_{pj}\sum_{j=1}^{J} u_j}\sum_{j=1}^{J} w_{pj}u_j\big(-\sigma_g^2-\dfrac{\sigma_e^2}{K_j}\big)\ , & \text{if } p=q\neq c \\
\dfrac{1}{\sum_{j=1}^{J} u_j}+\dfrac{1}{\sum_{j=1}^{J} w_{pj}\sum_{j=1}^{J} u_j}\sum_{j=1}^{J} w_{pj}u_j\big(-\sigma_g^2-\dfrac{\sigma_e^2}{K_j}\big) & \\
\quad +\dfrac{1}{\sum_{j=1}^{J} w_{qj}\sum_{j=1}^{J} u_j}\sum_{j=1}^{J} w_{qj}u_j\big(-\sigma_g^2-\dfrac{\sigma_e^2}{K_j}\big) & \\
+\dfrac{1}{\sum_{j=1}^{J} w_{pj}\sum_{j=1}^{J} w_{qj}}\sum_{j\in\mathcal{J}(p,q)} w_{pj}w_{qj}(\sigma_g^2+\dfrac{K_{pqj}\sigma_e^2}{K_{pj}K_{qj}})\ , & \text{if } \begin{cases} p\neq q \\ p\neq c \\ q\neq c \end{cases} \\
\dfrac{1}{\sum_{j=1}^{J} u_j}+\dfrac{1}{\sum_{j=1}^{J} w_{pj}\sum_{j=1}^{J} u_j}\sum_{j=1}^{J} w_{pj}u_j\big(-\sigma_g^2-\dfrac{\sigma_e^2}{K_j}\big)\ , & \text{if } \begin{cases} p\neq q \\ q=c \end{cases},
\end{cases}
\tag{17}
$$

where the sum over $\mathcal{J}(p,q)=\mathcal{J}(p)\cap\mathcal{J}(q)$ *is 0 if* $\mathcal{J}(p,q)$ *is empty.*

Proof See Appendix B. ■

COROLLARY 3.4 *The variance in the estimated difference between two test treatments is*

$$
\begin{aligned}
var(\alpha_p^\star-\alpha_q^\star) = &\frac{1}{\sum_{j=1}^{J} w_{pj}} \ + \ \frac{1}{\sum_{j=1}^{J} w_{qj}} \\
&- \frac{2}{\sum_{j=1}^{J} w_{pj}\sum_{j=1}^{J} w_{qj}}\sum_{j\in\mathcal{J}(p,q)} w_{pj}w_{qj}(\sigma_g^2+\frac{K_{pqj}\sigma_e^2}{K_{pj}K_{qj}})\ ,
\end{aligned}
$$

where $p\neq q$, $p\neq c$, $q\neq c$ *and the last sum is 0 if* $\mathcal{J}(p,q)$ *is empty, and the variance in the estimated difference between a test treatment and the reference treatment is*

$$
var(\delta_i^\star)=\frac{1}{\sum_{j=1}^{J} w_{ij}}\ . \tag{18}
$$

An alternative expression for the variance in the estimated difference between two treatments is provided in Appendix C.

Let $\hat{\sigma}_a^2$, $\hat{\sigma}_g^2$, $\hat{\sigma}_t^2$ and $\hat{\sigma}_e^2$ be the REML estimates, restricted to non-negative values, of σ_a^2, σ_g^2, $\sigma_t^2$and σ_e^2, respectively. Let $\hat{w}_{ij}=(2(\hat{\sigma}_g^2+\hat{\sigma}_e^2/K_{ij}))^{-1}$ and $\hat{w}_{pqj}=(2(\hat{\sigma}_g^2+\hat{\sigma}_e^2/K_{pqj}))^{-1}$, with K_{pqj} defined as in Proposition 2. Then,

$$
\hat{\delta}_i=\frac{1}{\sum_{j=1}^{J}\hat{w}_{ij}}\sum_{j\in\mathcal{J}(i)}\frac{\hat{w}_{ij}}{K_{ij}}\sum_{k\in\mathcal{K}(i,j)} d_{ijk}
$$

is an estimator of $\alpha_i - \alpha_c$, and $\hat{\delta}_p - \hat{\delta}_q$ is a reference treatment method estimator of $\alpha_p - \alpha_q$. The ratio between $\hat{\delta}_i$ and its standard error, $(\sum_{j=1}^{J} \hat{w}_{ij})^{-1/2}$, may be used as a test statistic for the difference between the ith treatment and the reference treatment. Similarly, the ratio between $\hat{\delta}_p - \hat{\delta}_q$ and the standard error

$$\Big(\frac{1}{\sum_{j\in\mathcal{J}(p)} \hat{w}_{pj}} + \frac{1}{\sum_{j\in\mathcal{J}(q)} \hat{w}_{qj}} - \frac{2}{\sum_{j=1}^{J} \hat{w}_{pj} \sum_{j=1}^{J} \hat{w}_{qj}} \sum_{j\in\mathcal{J}(p,q)} \hat{w}_{pj}\hat{w}_{qj}(\hat{\sigma}_g^2 + \frac{K_{pqj}\hat{\sigma}_e^2}{K_{pj}K_{qj}}) \Big)^{1/2}$$

may be used as a test statistic for the difference between the pth and the qth treatment, $p \neq c$, $q \neq c$. These statistics can be approximately compared with a t distribution with $\sum_{i=1}^{J} J_i - I - J + 1$ degrees of freedom.

3.3 *General series*

Often the experiments can be classified with regard to more than a single factor. For example, the experiments may have been conducted on several points of time, e.g. years, and at several locations. In this case, let y_{ijkl} denote an observation from treatment i, $i = 1, 2, \ldots, I$, in the lth experiment, $l = 1, 2, \ldots, L_{jk}$ at point of time j, $j = 1, 2, \ldots, J$ and location k, $k = 1, 2, \ldots, K_j$. Let

$$y_{ijkl} = \alpha_i + a_j + b_k + g_{ij} + h_{ik} + t_{jkl} + e_{ijkl} \ . \tag{19}$$

In Equation (19), α_i is the expected value using the ith treatment, a_j is a random effect of the jth point of time, b_k is a random effect of the kth location, g_{ij} is a random interaction between the ith treatment and the jth point of time, h_{ik} is a random interaction between the ith treatment and the kth location, t_{jkl} is a random effect of the lth experiment at the jth point of time and the kth location, and e_{ijkl} is a random error. The random effects are assumed to be independent and normally distributed with expected value 0 and variances σ_a^2, σ_b^2, σ_g^2, σ_h^2, σ_t^2 and σ_e^2, respectively. In series of crop variety trials, points of time are often years or seasons. In these cases, it is often reasonable to assume that effects of points of time, which are mainly due to changes in weather, are uncorrelated, especially if new plots are used every year. Model (19) can be augmented with the interaction between points of time and locations, provided that the effects of this interaction are not confounded in the design with the effects of points of time, locations or trials. If this is fulfilled, let

$$y_{ijkl} = \alpha_i + a_j + b_k + g_{ij} + h_{ik} + s_{jk} + t_{jkl} + e_{ijkl} \ . \tag{20}$$

In Equation (20), h_{ik} is a random interaction between the ith treatment and the kth location, and s_{jk} is a random interaction between the jth point of time and the kth location. The random effects are assumed to be independent and normally distributed with expected value 0 and variances σ_a^2, σ_b^2, σ_g^2, σ_h^2, σ_s^2, σ_t^2 and σ_e^2, respectively.

Let N denote the total number of observations, and M the total number of experiments. Let $\mathbf{I}(k) \equiv \mathbf{I}_k$ denote a $k \times k$-identity matrix and $\mathbf{1}_k$ a k-vector of ones. Models (5), (11), (19) and (20) can be written

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}_t\boldsymbol{\gamma}_t + \mathbf{e} \ , \tag{21}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}_a\boldsymbol{\gamma}_a + \mathbf{Z}_g\boldsymbol{\gamma}_g + \mathbf{Z}_t\boldsymbol{\gamma}_t + \mathbf{e} \ , \tag{22}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}_a\boldsymbol{\gamma}_a + \mathbf{Z}_b\boldsymbol{\gamma}_b + \mathbf{Z}_g\boldsymbol{\gamma}_g + \mathbf{Z}_h\boldsymbol{\gamma}_h + \mathbf{Z}_t\boldsymbol{\gamma}_t + \mathbf{e} \ , \tag{23}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}_a\boldsymbol{\gamma}_a + \mathbf{Z}_b\boldsymbol{\gamma}_b + \mathbf{Z}_g\boldsymbol{\gamma}_g + \mathbf{Z}_h\boldsymbol{\gamma}_h + \mathbf{Z}_s\boldsymbol{\gamma}_s + \mathbf{Z}_t\boldsymbol{\gamma}_t + \mathbf{e} \ , \tag{24}$$

respectively. In Equations (21)–(24), $\mathbf{X}$, $\mathbf{Z}_a$, $\mathbf{Z}_b$, $\mathbf{Z}_g$, $\mathbf{Z}_h$, $\mathbf{Z}_s$ and $\mathbf{Z}_t$ denote full column-rank design matrices for treatments, points of time, locations, treatment-by-time interactions, treatment-by-location interactions, time-by-location interactions, and experiments, respectively, whereas $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}_a$, $\boldsymbol{\gamma}_b$, $\boldsymbol{\gamma}_g$, $\boldsymbol{\gamma}_h$, $\boldsymbol{\gamma}_s$, $\boldsymbol{\gamma}_t$ and $\mathbf{e}$ denote vectors of fixed treatment effects, random point of time effects, random location effects, random treatment-by-time effects, random treatment-by-location effects, random time-by-location effects, random experiment effects and random model errors, respectively.

Let $\mathbf{G}_a = \sigma_a^2\mathbf{I}(\text{rank}(Z_a))$, $\mathbf{G}_b = \sigma_b^2\mathbf{I}(\text{rank}(Z_b))$, $\mathbf{G}_g = \sigma_g^2\mathbf{I}(\text{rank}(Z_g))$, $\mathbf{G}_h = \sigma_h^2\mathbf{I}(\text{rank}(Z_h))$, $\mathbf{G}_s = \sigma_s^2\mathbf{I}(\text{rank}(Z_s))$, $\mathbf{G}_t = \sigma_t^2\mathbf{I}(\text{rank}(Z_t))$ and $\mathbf{R} = \sigma_e^2\,\mathbf{I}_N$. Let $\mathbf{V}$ denote the covariance matrix of $\mathbf{y}$. For Models (5), (11), (19) and (20), respectively,

$$\mathbf{V} = \mathbf{Z}_t\mathbf{G}_t\mathbf{Z}_t' + \mathbf{R} \ ,$$

$$\mathbf{V} = \mathbf{Z}_a\mathbf{G}_a\mathbf{Z}_a' + \mathbf{Z}_g\mathbf{G}_g\mathbf{Z}_g' + \mathbf{Z}_t\mathbf{G}_t\mathbf{Z}_t' + \mathbf{R} \ ,$$

$$\mathbf{V} = \mathbf{Z}_a\mathbf{G}_a\mathbf{Z}_a' + \mathbf{Z}_b\mathbf{G}_b\mathbf{Z}_b' + \mathbf{Z}_g\mathbf{G}_g\mathbf{Z}_g' + \mathbf{Z}_h\mathbf{G}_h\mathbf{Z}_h' + \mathbf{Z}_t\mathbf{G}_t\mathbf{Z}_t' + \mathbf{R} \ ,$$

$$\mathbf{V} = \mathbf{Z}_a\mathbf{G}_a\mathbf{Z}_a' + \mathbf{Z}_b\mathbf{G}_b\mathbf{Z}_b' + \mathbf{Z}_g\mathbf{G}_g\mathbf{Z}_g' + \mathbf{Z}_h\mathbf{G}_h\mathbf{Z}_h' + \mathbf{Z}_s\mathbf{G}_s\mathbf{Z}_s' + \mathbf{Z}_t\mathbf{G}_t\mathbf{Z}_t' + \mathbf{R} \ .$$

Generally, linear mixed models can be written $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}$, where $\mathbf{X}$ and $\mathbf{Z}$ denote design matrices for treatments and random effects, respectively, whereas $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\mathbf{e}$ denote vectors of fixed treatment effects, random effects and random model errors, respectively. The covariance matrix of $\mathbf{y}$ is $\mathbf{V} = \mathbf{Z}'\mathbf{G}\mathbf{Z} + \mathbf{R}$, where $\mathbf{G}$ is the covariance matrix of $\boldsymbol{\gamma}$ and $\mathbf{R}$ is the covariance matrix of $\mathbf{e}$. Given $\mathbf{V}$, the generalized least squares estimator of $\boldsymbol{\alpha}$ is $\boldsymbol{\alpha}^\circ = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, and the best linear unbiased predictor of $\boldsymbol{\gamma}$ is $\boldsymbol{\gamma}^\circ = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}^\circ)$. In complete series, that is in tables without missing cells, treatments are orthogonal to experiments, which results in uncorrelated generalized least squares estimates of treatment differences and experimental effects: $\text{cov}(\mathbf{L}\boldsymbol{\alpha}^\circ, \boldsymbol{\gamma}^\circ) = \mathbf{L}\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} = \mathbf{0}$, when $\mathbf{L}$ is a row vector of contrast constants.

At the second stage of the two-stage analysis, the residual variance is a sum of two components: the variance due to treatment-by-experiment interaction, and the variance due to intra-experimental plot errors. When $\mathbf{R}$ is modeled as $\sigma_e^2\,\mathbf{I}_N$, the residual variance is assumed to be homogeneous. In general, if the elements of $\mathbf{y}$ are sorted by experiments, $\mathbf{R}$ is a block-diagonal matrix. Estimated covariance matrices from the analyses of the first stage can be used as blocks. This makes it possible to estimate the effects of the treatment-by-experiment interaction. Equations (5), (11), (19) and (20) can be augmented by a term u_{ij}, u_{ijk}, u_{ijkl} and u_{ijkl}, respectively, for random treatment-by-experiment interaction effects. However, from a

data processing point of view it is difficult to store all estimated covariance matrices for future use at second stages [26]. For this reason, $\mathbf{R}$ is often approximated by $\sigma_e^2\mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix [17].

Define the I-vector $\boldsymbol{\delta} = (\delta_1, \delta_2 \ldots, \delta_I)$ of differences in expected values as

$$\boldsymbol{\delta} = \boldsymbol{\alpha} - \mathbf{1}_I\alpha_c \ , \tag{25}$$

where α_c is the expected value of the reference treatment, so that $\delta_c = 0$. Denote by $\mathbf{Z}_t$ the subset of columns of $\mathbf{Z}$ that form the design matrix of the experiments. Let $\mathbf{x}_i$ be the ith column of $\mathbf{X}$, and $\mathbf{A} = \mathbf{Z}_t'\text{diag}(\mathbf{x}_c)$. Then $\mathbf{Ay}$ is the M-vector of reference treatment observations, with variance $\mathbf{AVA}'$. The N-vector of within-experiment differences to the reference is

$$\mathbf{d} = (\mathbf{I}_N - \mathbf{Z}_t\mathbf{A})\mathbf{y} \ . \tag{26}$$

The expected value of $\mathbf{d}$ is $E(\mathbf{d}) = E(\mathbf{y} - \mathbf{Z}_tE(\mathbf{Ay})) = E(\mathbf{y} - \mathbf{Z}_t\mathbf{1}_M\alpha_c) = \mathbf{X}(\boldsymbol{\alpha} - \mathbf{1}_I\alpha_c) = \mathbf{X}\boldsymbol{\delta}$, because $E(y) = \mathbf{X}\boldsymbol{\alpha}$ and $\mathbf{Z}_t\mathbf{1}_m\alpha_c = \mathbf{X}\mathbf{1}_I\alpha_c$. The covariance matrix of $\mathbf{d}$ is

$$\mathbf{V_d} = (\mathbf{I}_N - \mathbf{Z}_t\mathbf{A})\mathbf{V}(\mathbf{I}_N - \mathbf{Z}_t\mathbf{A})' \ . \tag{27}$$

Through the reference treatment method each treatment is compared with the reference treatment directly. As discussed in the introduction, covariances in $\mathbf{V_d}$ between treatments are not considered for the estimation of the differences between the test treatments and the reference, i.e.

$$\tilde{\mathbf{V}}_\mathbf{d} = \sum_{i=1}^{I} \text{diag}(\mathbf{x}_i)\mathbf{V_d}\text{diag}(\mathbf{x}_i)$$

is used for the estimation of $\boldsymbol{\delta}$ as defined in Equation (25). Thus,

$$\boldsymbol{\delta}^\star = (\mathbf{X}'\tilde{\mathbf{V}}_\mathbf{d}^-\mathbf{X})^-\mathbf{X}'\tilde{\mathbf{V}}_\mathbf{d}^-\mathbf{d} \ , \tag{28}$$

is the I-vector of estimates $\delta_i^\star$, with $\delta_c^\star = 0$, $i = 1, 2, \ldots, I$. In Equation (28) generalized inverses are used, because $\tilde{\mathbf{V}}_\mathbf{d}$ and $\mathbf{X}'\tilde{\mathbf{V}}_\mathbf{d}^-\mathbf{X}$ are singular. The reference treatment method estimator $\alpha_c^\star$ of the expected value α_c of the reference treatment is found by regressing $\mathbf{Ay}$ on $\mathbf{1}_M$, using generalized least squares:

$$\alpha_c^\star = (\mathbf{1}_M'(\mathbf{AVA}')^{-1}\mathbf{1}_M)^{-1}\mathbf{1}_M'(\mathbf{AVA}')^{-1}\mathbf{Ay} \ . \tag{29}$$

The off-diagonal values of $\mathbf{AVA}'$ involve variances between groups (locations or points of time) and variances for variety-by-group interaction. When these between-groups variances are large, the estimate $\alpha_c^\star$ will be close to an unweighted average of group means. When they are small, $\alpha_c^\star$ will be close to a weighted average of group means, using numbers of experiments as weights. The vector $\boldsymbol{\alpha}^\star$of reference treatment method estimates of expected values is

$$\boldsymbol{\alpha}^\star = \mathbf{1}_I\alpha_c^\star + \boldsymbol{\delta}^\star \ . \tag{30}$$

PROPOSITION 3.5 *The covariance matrix var($\boldsymbol{\alpha}^\star$), of $\boldsymbol{\alpha}^\star$ as defined in Equation*

(30), *is*

$$\mathbf{C} = \mathbf{1}_I(\mathbf{1}'_M(\mathbf{AVA}')^{-1}\mathbf{1}_M)^{-1}\mathbf{1}'_I + \mathbf{S} + \mathbf{T}' + \mathbf{T} \ ,$$

where

$$\mathbf{S} = (\mathbf{X}'\tilde{\mathbf{V}}_{\mathbf{d}}^{-}\mathbf{X})^{-}\mathbf{X}'\tilde{\mathbf{V}}_{\mathbf{d}}^{-}\mathbf{V}_{\mathbf{d}}((\mathbf{X}'\tilde{\mathbf{V}}_{\mathbf{d}}^{-}\mathbf{X})^{-}\mathbf{X}'\tilde{\mathbf{V}}_{\mathbf{d}}^{-})' \ , \tag{31}$$

$$\mathbf{T} = (\mathbf{X}'\tilde{\mathbf{V}}_{\mathbf{d}}^{-}\mathbf{X})^{-}\mathbf{X}'\tilde{\mathbf{V}}_{\mathbf{d}}^{-}(\mathbf{I}_N - \mathbf{Z}_t\mathbf{A})\mathbf{V}\mathbf{B}' \ .$$

Proof Let $\mathbf{B} = \mathbf{1}_I(\mathbf{1}'_M(\mathbf{AVA}')^{-1}\mathbf{1}_M)^{-1}\mathbf{1}'_M(\mathbf{AVA}')^{-1}\mathbf{A}$. From Equations (26), (27), (28), (29) and $\text{var}(\boldsymbol{\alpha}^\star) = \text{var}(\mathbf{1}_I\alpha_c^\star) + \text{var}(\boldsymbol{\delta}^\star) + \text{cov}(\mathbf{1}_I\alpha_c^\star\ , \boldsymbol{\delta}^\star) + \text{cov}(\boldsymbol{\delta}^\star, \mathbf{1}_I\alpha_c^\star)$, it follows that $\mathbf{C} = \mathbf{BVB}' + \mathbf{S} + \mathbf{T}' + \mathbf{T}$. Finally note that $\mathbf{BVB}' = \mathbf{1}_I(\mathbf{1}'_M(\mathbf{AVA}')^{-1}\mathbf{1}_M)^{-1}\mathbf{1}'_I$. ■

In applications, the variance components are unknown. Let $\hat{\mathbf{V}}$ denote the estimate of $\mathbf{V}$, obtained by substituting the REML estimates for the unknown variance components, and let $\hat{\mathbf{C}}$ and $\hat{\boldsymbol{\alpha}}$ denote the covariance matrix $\mathbf{C}$ and the estimates $\boldsymbol{\alpha}^\star$, respectively, calculated using $\hat{\mathbf{V}}$ in place of $\mathbf{V}$. Approximate tests of linear combinations of the expected values α_i, $i = 1, 2, \ldots, I$, can be carried out using general t- and F-statistics. Consider the statistical hypothesis $\text{H}_0 : \mathbf{L}\boldsymbol{\alpha}$, where $\mathbf{L}$ is a matrix of full row rank. When $\mathbf{L}$ is a single row, the general t-statistic is applicable:

$$t = \frac{\mathbf{L}\hat{\boldsymbol{\alpha}}}{\sqrt{\mathbf{L}\hat{\mathbf{C}}\mathbf{L}'}} \ . \tag{32}$$

When $\mathbf{L}$ comprises $R \geq 1$ rows, the general F-statistic given by

$$F = \frac{\hat{\boldsymbol{\alpha}}'\mathbf{L}'(\mathbf{L}\hat{\mathbf{C}}\mathbf{L}')^{-1}\mathbf{L}\hat{\boldsymbol{\alpha}}}{R} \tag{33}$$

may be used for testing the hypothesis H_0. Using Models (19) and (20), the contrasts in the treatments are compared with variances for treatment-by-time and treatment-by-location interactions estimated with $\sum_{i=1}^{J} J_i - I - J + 1$ and $\sum_{i=1}^{K} K_i - I - K + 1$ degrees of freedom, respectively, where J_i is the number of points of time comprising the ith treatment, and K_i is the number of locations comprising the ith treatment. For this reason it is reasonable to compare Equations (32) and (33) with a t distribution with $\min\{\sum_{i=1}^{J} J_i - I - J + 1\ ,\ \sum_{i=1}^{K} K_i - I - K + 1\}$ degrees of freedom and an F distribution with R and $\min\{\sum_{i=1}^{J} J_i - I - J + 1\ ,\ \sum_{i=1}^{K} K_i - I - K + 1\}$ degrees of freedom, respectively. Alternatively, the generalized Satterthwaite method [8, 23] can be used. When $\mathbf{L}$ is a single row, it is assumed that $\mathbf{LCL}'$ is approximately chi-square distributed with $2E((\mathbf{LCL}))^2/\text{var}(\mathbf{LCL}')$ degrees of freedom. Let $\mathbf{M}$ denote the observed information matrix, obtained at maximization of the residual likelihood, and let $\mathbf{g}$ denote the vector of derivatives of $\mathbf{LCL}'$ with respect to the variance components. Then $\mathbf{gMg}'$ approximates $\text{var}(\mathbf{LCL}')$, and the Satterthwaite approximation of the degrees of freedom is $2E((\mathbf{LCL}'))^2/\mathbf{gMg}'$. Kenward and Roger [13, 14] proposed further approximations for general t- and F-tests.

Probability values can be adjusted for multiple comparisons by the Tukey-Kramer method [15]. The Dunnett method [7] is appropriate if the only comparisons of interest are those between the test treatments and the reference treatment.

4. Discussion

This paper highlighted practical problems when communicating generalized least squares estimates in series of variety trials. Performance is often measured as difference in yield to a reference variety. As shown, generalized least squares estimates of differences can appear strange when compared with observed differences.

Sometimes it is suggested that test varieties could be compared indirectly through the use of a reference variety. This recurrent idea was examined by specification of a 'reference treatment method'. The reference treatment method gives estimates of differences to the reference treatment that are averages or weighted averages of observed differences, thereby avoiding the striking discrepancies produced by the method of generalized least squares. However, for indirect comparisons of test varieties, through comparison with the reference variety, discrepancies between estimates and observations can occur, i.e. with the reference treatment method, an estimate of the expected difference between two test varieties can be outside the range of observed differences.

It might seem restrictive that the reference treatment method requires that a reference variety must be included in all trials. In practice this is seldom a problem, since series of VCU trials are planned, and a reference variety is always included to facilitate comparison. However, if the series is not designed for a joint analysis, this requirement can be problematic. Through the reference treatment method, estimates of differences can depend on the choice of reference variety. The reference treatment method is intended for the situation that the reference variety is of particular interest. The reference variety may be a variety mix or a standard variety that is well known.

The main problem discussed in this paper is that point estimates do not always compare well with what is truly observed. This problem can be diminished if confidence intervals are provided. When estimated differences deviate from observed differences, standard errors in differences are usually large and confidence intervals wide. By presenting confidence intervals for differences it is pedagogically made clear that estimates are uncertain.

The efficiency in the example, of the reference treatment method as compared with the method of generalized least squares, was evaluated. In the five-years series, the standard errors in the differences were consistently higher with the reference treatment method than with the generalized least squares method. This was not surprising, because the comparison was made under the assumption that the model was correct, and the reference treatment method ignores some correlations in the data. When the model is incorrect, the reference treatment method could possibly in some cases perform better than the generalized least squares method. Further studies are needed to investigate if this might be the case. Otherwise, the reference treatment method cannot be recommended from a statistical point of view. It should be more important that the statistical tests of the differences between the varieties are made as efficiently as possible, than that estimated and observed differences agree.

The discrepancies in results between the methods of generalized least squares and the reference treatment method indicates that the model does perhaps not describe the observations sufficiently well. In a two-way table of yields from I varieties and J trials, there may be variety-by-trial interaction that is not included

in the additive models of the present paper. Instead of investigating other ways to make calculations one could, in future research, consider other models.

As shown in Section 3.3, in series that are complete, generalized least squares estimates of treatment differences are uncorrelated with experimental effects. If possible, a series of trials should be designed as a complete series, because this eliminates the problems of separating effects of treatments from effects of trials. Series of complete variety trials can be analyzed using GGE or AMMI analyses [11, 31], which provides insight into variety-by-trial interactions. Series of VCU trials can, for reasons mentioned in the Introduction, rarely be de designed as complete series. However, if the series is made 'as complete as possible', such apparent problems that were discussed in the present paper may be partially avoided.

5. References

References

[1] A.C. Bathke, S.W. Harrar, H. Wang, K. Zhang, and H.-P. Piepho, *Series of Randomized Complete Block Experiments With Non-Normal Data*, Comput. Stat. Data Anal. 54 (2010), pp. 1840–1857.

[2] T. Caliński, S. Czajka, Z. Kaczmarek, P. Krajewski, and W. Pilarczyk, *Analyzing Multi-Environment Variety Trials Using Randomization-Derived Mixed Models*, Biometrics 61 (2005), pp. 448–455.

[3] T. Caliński, S. Czajka, Z. Kaczmarek, P. Krajewski, and W. Pilarczyk, *A Mixed Model Analysis of Variance for Multi-Environmental Variety Trials*, Stat. Pap. 50 (2009), pp. 735–759.

[4] T. Caliński, S. Czajka, Z. Kaczmarek, P. Krajewski, and W. Pilarczyk, *Analysing the Genotype-by-Environment Interactions Using a Randomization-Derived Mixed Model*, J. Agric. Biol. Environ. Stat. 14 (2009), pp. 224–241.

[5] W.G. Cochran, *The combination of estimates from different experiments*, Biometrics 10 (1954), pp. 101–129.

[6] J.-B. Denis, H.-P. Piepho, and F.A. van Eeuwijk, *Modelling Expectation and Variance for Genotype by Environment Data* Heredity 79 (1997), pp. 162–171.

[7] C.W. Dunnett, *New tables for multiple comparisons with a control.* Biometrics 20 (1964), pp. 482-491.

[8] A.H.T. Fai, and P.L. Cornelius, *Approximate F-tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-plot Experiments*, J. Stat. Comput. Simul. 54 (1996), pp. 363-378.

[9] J. Forkman, J. (2011), *Linear mixed models for series of variety trials*, Biuletyn Oceny Odmian 33 (2011), pp. 63–72.

[10] A. Frensham, B. Cullis, and A. Verbyla, *Genotype by Environment Variance in a Two-Stage Analysis*, Biometrics 53 (1997), pp. 1373–1883.

[11] H.G. Gauch, *Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs*, Elsevier, Amsterdam, 1992.

[12] X. Hu, and J. Spilke, *Variance-Covariance Structure and its Influence on Variety Assessment in Regional Crop Trials*, Field Crop. Res. 120 (2011), pp. 1–8.

[13] M.G. Kenward, and J.H. Roger, *Small sample inference for fixed effects from restricted maximum likelihood*, Biometrics 53 (1997), pp. 983–997.

[14] M.G. Kenward, and J.H. Roger, *An improved approximation to the precision of fixed effects from restricted maximum likelihood*, Comput. Stat. Data Anal. 53 (2009), pp. 2583–2595.

[15] C.Y. Kramer, *Extension of multiple range tests to group means with unequal numbers of replications*, Biometrics 12 (1956), pp. 307–310.
[16] R.C. Littell, G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Shabenberger O., *SAS for Mixed Models*, 2nd ed., SAS Institute, Cary, 2006.
[17] J. Möhring, and H.-P. Piepho, *Comparison of Weighting in Two-Stage Analysis of Plant Breeding Trials*, Crop Sci. 49 (2009), pp. 1977–1988.
[18] F. Nabugoomu, R.A. Kempton, and M. Talbot, *Analysis of Series of Trials where Varieties Differ in Sensitivity to Locations*, J. Agric. Biol. Environ. Stat. 4 (1999), pp. 310–325.
[19] H.D. Patterson, *Routine least squares estimation of variety means in incomplete tables*, J. natn. Inst. agric. Bot. 14 (1978), pp. 401–412.
[20] H.D. Patterson, *Analysis of Series of Variety Trials*, in *Statistical Methods for Plant Variety Evaluation*, R.A. Kempton, and P.N. Fox, eds., Chapman & Hall, London, 1997, pp. 139–161.
[21] Patterson, H. D., and Thompson, R. (1971), "Recovery of Inter-Block Information when Block Sizes are Unequal," *Biometrika*, 58, 545–554.
[22] D.L. Robinson, *Estimation and Use of Variance Components*, The Statistician 36 (1987), pp. 3–14.
[23] SAS Institute, *SAS/STAT 9.2 user's guide: the mixed procedure*, SAS Institute, Cary, 2008.
[24] A.M. Saxton, *A macro for converting mean separation output to letter groupings in proc mixed*, in *Proceedings of the 23rd SAS Users' Group International*, SAS Institute, Cary, 1998, pp. 1243–1246.
[25] V. Silvey, *Methods of analysing NIAB variety trial data over many sites and several seasons*, J. natn. Inst. agric. Bot. 14, pp. 385–400.
[26] A. Smith, B. Cullis, and A. Gilmour, *The Analysis of Crop Variety Evaluation Data in Australia*, Aust. N. Z. J. Stat. 43 (2001), pp. 129–145.
[27] A. Smith, B. Cullis, and R. Thompson., *Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend*, Biometrics 57 (2001), pp. 1138–1147.
[28] A.B. Smith, B.R. Cullis, and R. Thompson, *The Analysis of Crop Cultivar Breeding and Evaluation Trials: an Overview of Current Mixed Model Approaches*, J. Agric. Sci. 143 (2005), pp. 449–462.
[29] C.M. Theobald, M. Talbot, and F. Nabugoomu, *A Bayesian Approach to Regional and Local-Area Prediction from Crop Variety Trials*, J. Agric. Biol. Environ. Stat. 7 (2002), pp. 403–419.
[30] S.J. Welham, B.J. Gogel, A.B. Smith, R. Thompson, and B.R. Cullis, *A Comparison of Analysis Methods for Late-Stage Variety Evaluation Trials*, Aust. N. Z. J. Stat. 52 (2010), pp. 125–149.
[31] W. Yan, and N.A. Tinker, *Biplot Analysis of Multi-Environment Trial Data: Principles and Applications*, Can. J. Plant Sci. 86 (2006), pp. 623–645.
[32] F. Yates, and W.G. Cochran, *The Analysis of Groups of Experiments*, J. Agric. Sci., Camb. 28 (1938), pp. 556–580.

Appendix A. Proof of Proposition 1.

Proof By Model (5), $\mathrm{var}(\alpha_c^\star) = (\sigma_t^2 + \sigma_e^2)/J$ and

$$\mathrm{cov}(\alpha_c^\star, \delta_i^\star) = \mathrm{cov}\Big(\frac{1}{J}\sum_{j=1}^{J} e_{cj}, \frac{1}{J_i}\sum_{j\in\mathcal{J}(i)}(e_{ij} - e_{cj})\Big) = \sum_{j\in\mathcal{J}(i)}\frac{-\sigma_e^2}{J_i} = \frac{-\sigma_e^2}{J}$$

when $i \neq c$. Moreover, $\text{var}(\delta_i^\star) = 2\sigma_e^2/J_i$ when $i \neq c$, and

$$\text{cov}(\delta_p^\star, \delta_q^\star) = \frac{1}{J_p J_q} \text{cov}\big(\sum_{j \in \mathcal{J}(p)} (y_{pj} - y_{cj}), \sum_{j \in \mathcal{J}(q)} (y_{qj} - y_{cj})\big) = \frac{J_{pq}\sigma_e^2}{J_p J_q} \ ,$$

when $p \neq q$, $p \neq c$ and $q \neq c$. The second equation of the proposition follows from $\text{var}(\alpha_p^\star) = \text{var}(\alpha_c^\star) + \text{var}(\delta_p^\star) + 2\,\text{cov}(\alpha_c^\star, \delta_p^\star)$. The third equation follows from $\text{cov}(\alpha_p^\star, \alpha_q^\star) = \text{var}(\alpha_c^\star) + \text{cov}(\alpha_c^\star, \delta_p^\star) + \text{cov}(\alpha_c^\star, \delta_q^\star) + \text{cov}(\delta_p^\star, \delta_q^\star)$, where $\text{cov}(\delta_p^\star, \delta_q^\star) = \sigma_e^2 J_{pq}/(J_p J_q)$. The fourth equation follows from $\text{cov}(\alpha_p^\star, \alpha_c^\star) = \text{var}(\alpha_c^\star) + \text{cov}(\alpha_c^\star, \delta_p^\star)$. ■

Appendix B. Proof of Proposition 2.

Proof Write $\delta_i^\star$ as $\delta_i^\star = \sum_{j \in \mathcal{J}(i)} v_{ij} \bar{d}_{ij\cdot}$, where $v_{ij} = w_{ij}/\sum_{j \in \mathcal{J}(i)} w_{ij}$ and $\bar{d}_{ij\cdot} = \sum_{k \in \mathcal{K}(i,j)} d_{ijk}/K_{ij}$. By Equation (12), $\text{var}(\bar{d}_{ij\cdot}) = 1/w_{ij}$. Consequently

$$\text{var}(\delta_i^\star) = \text{var} \sum_{j \in \mathcal{J}(i)} v_{ij} \bar{d}_{ij\cdot} = \frac{1}{\sum_{j=1}^J w_{ij}} \ . \tag{B1}$$

Since $\text{cov}(\bar{d}_{pj\cdot}, \bar{d}_{qj\cdot}) = \text{cov}(\sum_{k \in \mathcal{K}(p,j)} (y_{pjk} - y_{cjk}) \ , \sum_{k \in \mathcal{K}(q,j)} (y_{qjk} - y_{cjk}))/(K_{pj}K_{qj})$ $= (K_{pj}K_{qj}\sigma_g^2 + K_{pqj}\sigma_e^2)/(K_{pj}K_{qj}) = \sigma_g^2 + K_{pqj}\sigma_e^2/(K_{pj}K_{qj})$, the covariance between $\delta_p^\star$ and $\delta_q^\star$ is

$$\begin{aligned} \text{cov}\big(\sum_{j \in \mathcal{J}(p)} v_{pj} \bar{d}_{pj\cdot} \ , \sum_{j \in \mathcal{J}(q)} v_{qj} \bar{d}_{qj\cdot}\big) &= \sum_{j \in \mathcal{J}(p,q)} v_{pj} v_{qj} \ \text{cov}(\bar{d}_{pj\cdot} \ , \bar{d}_{qj\cdot}) \\ &= \frac{1}{\sum_{j=1}^J w_{pj} \sum_{j=1}^J w_{qj}} \sum_{j \in \mathcal{J}(p,q)} w_{pj} w_{qj} (\sigma_g^2 + \frac{K_{pqj}\sigma_e^2}{K_{pj}K_{qj}}) \ . \end{aligned} \tag{B2}$$

Write $\alpha_c^\star$ as $\alpha_c^\star = \sum_{j=1}^J x_j \bar{y}_{cj\cdot}$, where $x_j = u_j/\sum_{j=1}^J u_j$ and $\bar{y}_{cj\cdot} = \sum_{k=1}^{K_j} y_{cjk}/K_j$. Since $\text{cov}(\bar{d}_{ij\cdot} \ , \bar{y}_{cj\cdot}) = \text{cov}(\sum_{k \in \mathcal{K}(i,j)} (y_{ijk} - y_{cjk}) \ , \sum_{k=1}^{K_j} y_{cjk})/(K_{ij}K_j) = (-K_{ij}K_j\sigma_g^2 - K_{ij}\sigma_e^2)/(K_{ij}K_j) = -\sigma_g^2 - \sigma_e^2/K_j$, the covariance between $\delta_i^\star$ and $\alpha_c^\star$ is

$$\begin{aligned} \text{cov}\big(\sum_{j \in \mathcal{J}(i)} v_{ij} \bar{d}_{ij\cdot} \ , \sum_{j=1}^J x_j \bar{y}_{cj\cdot}\big) &= \sum_{j \in \mathcal{J}(i)} v_{ij} x_j \ \text{cov}(\bar{d}_{ij\cdot} \ , \bar{y}_{cj\cdot}) \\ &= \frac{1}{\sum_{j=1}^J w_{ij} \sum_{j=1}^J u_j} \sum_{j \in \mathcal{J}(i)} w_{ij} u_j (-\sigma_g^2 - \frac{\sigma_e^2}{K_j}) \ . \end{aligned} \tag{B3}$$

The variance $\text{var}(\alpha_c^\star)$ in the estimator of the expected value of the reference treatment is $1/\sum_{j=1}^J u_j$, because $\text{var}(\sum_{k=1}^{K_j} y_{cjk}/K_j) = 1/u_j$. The second equation of the proposition follows from $\text{var}(\alpha_p^\star) = \text{var}(\alpha_c^\star + \delta_p^\star) = \text{var}(\alpha_c^\star) + \text{var}(\delta_p^\star) + 2\,\text{cov}(\alpha_c^\star \ , \delta_p^\star)$ using Equations (B1) and (B3). The fourth equation of the proposition follows from $\text{cov}(\alpha_p^\star \ , \alpha_c^\star) = \text{cov}(\alpha_c^\star + \delta_p^\star \ , \alpha_c^\star) = \text{var}(\alpha_c^\star) + \text{cov}(\alpha_c^\star \ , \delta_p^\star)$ using Equation (B3). The third equation of the proposition follows from $\text{cov}(\alpha_p^\star, \alpha_q^\star) = \text{cov}(\alpha_c^\star + \delta_p^\star \ , \alpha_c^\star + \delta_q^\star)$ $= \text{var}(\alpha_c^\star) + \text{cov}(\alpha_c^\star \ , \delta_p^\star) + \text{cov}(\alpha_c^\star \ , \delta_q^\star) + \text{cov}(\delta_p^\star \ , \delta_q^\star)$ using Equations (B2) and (B3). ■

Appendix C. Another expression for the variance in the estimated difference between two treatments.

With the reference treatment method as applied to Model (11), the variance in the estimated difference between two treatments is

$$\begin{aligned}\mathrm{var}(\alpha_p^\star - \alpha_q^\star) = & \frac{2}{(\sum_{j\in\mathcal{J}(p)} w_{pj})^2} \sum_{j\in\mathcal{J}(p)} w_{pj}^2(\sigma_g^2 + \frac{\sigma_e^2}{K_{pj}}) \\ & + \frac{2}{(\sum_{j\in\mathcal{J}(q)} w_{qj})^2} \sum_{j\in\mathcal{J}(q)} w_{qj}^2(\sigma_g^2 + \frac{\sigma_e^2}{K_{qj}}) \\ & - \frac{2}{\sum_{j=1}^J w_{pj} \sum_{j=1}^J w_{qj}} \sum_{j\in\mathcal{J}(p,q)} w_{pj}w_{qj}(\sigma_g^2 + \frac{K_{pqj}\sigma_e^2}{K_{pj}K_{qj}}) \ . \end{aligned} \tag{C1}$$