Analysis of Half Diallel Mating Designs

I – A Practical Analysis Procedure for ANOVA Approximation

By G. R. JOHNSON[1]) and J. N. KING[2])

(Received 23rd June 1997)

Abstract

Procedures to analyze half-diallel mating designs using the SAS statistical package are presented. The procedure requires two runs of PROC VARCOMP and results in estimates of additive and non-additive genetic variation. The procedures described can be modified to work on most statistical software packages which can compute variance component estimates. The procedure is relatively simple and provides unbiased estimates for balanced designs and gives good approximations for unbalanced data.

Key words: diallel matings, variance estimates, GCA, SCA.

FDC: 165.3; 165.41; 174.7 *Pinus radiata;* (931).

Introduction

Diallel mating designs are widely used in the genetic improvement programs of many tree species (YEH and HEAMAN, 1987; SNYDER and NAMKOONG, 1978; TALBERT, 1979). Besides their practicality as a dual function mating design that provides both a pedigreed breeding population for selection and a progeny test of parents, they are also highly useful designs for estimating genetic parameters. Estimates of genetic variances and other population parameters provide essential information for the development of breeding strategies. The diallel mating design is of interest, in that the analysis of variance uses the concepts of general combining ability (GCA) and specific combining ability (SCA) to distinguish between the average performance of parents in crosses (GCA) and the deviation of individual crosses from the average of the parents (SCA). In the population improvement strategy using recurrent selection for general combining ability (GCA), we would naturally wish to know the relative amount of the genetic variation caused by additive gene effects (GCA) and whether non-additive gene action is important.

A drawback to the diallel mating design is that it is relatively complex to analyze. Because the genetic effects are not readily separated, they cannot easily be analyzed in a single execution of a linear model procedure in standard statistical packages. The computation of the appropriate sum-of-squares and expected mean squares have been derived in the literature for balanced (GRIFFING, 1956) and unbalanced data (e.g., GARRETSEN and KEULS, 1977: KEULS and GARRETSEN, 1978: BARADAT and DESPREZ-LOUSTAU, 1997). Estimation of the GCA and SCA effects are demonstrated by HUBER *et al.* 1992. Because standard statistical packages cannot handle diallel analyses, the breeder must either program the procedures or use special packages, such as the DIALL program of SCHAFFER and USANIS (1969). These specialty programs lack the convenience and ease associated with large data handling packages and can limit one's options in data analysis. Limita-

[1]) USDA, Forest Service, Forestry Sciences Lab., 3200 SW Jefferson Way, Corvallis, OR 97331-4401, USA

[2]) Ministry of Forests, Research Branch, Ministry of Forests, PO Box 9519, STN PROV GOVT, Victoria, BC, V8W 9C2, Canada

tions involve the amount of data the package can handle (i.e. number of sites and/or replications), the ability to designate effects as fixed or random, and the choice of methods used to estimate the variance components.

In this paper we show a simple and practical method of estimating variance components using standard statistical packages, demonstrating the procedure with SAS (SAS Institute, 1990) and illustrating its use with data from a progeny test of imbalanced disconnected diallels from a series of trials of the New Zealand Radiata Pine Breeding Cooperative. The methodology gives unbiased estimates when the data are balanced and reasonable approximations with unbalanced data. A complete report on the results of this series of trials is reported in a following manuscript (KING *et al.*, 1998).

Description of Procedure

The genetic model used for this analysis is:

$$P = G_A + G_D + E \qquad ...[1]$$

Where P is the phenotype, G_A is the additive genetic effect, G_D is the dominance genetic effect, and E is environmental effect. All epistatic effects are ignored in this simplified model.

The data we use to illustrate the procedure is a 7-year height and diameter measurement taken from progeny of a group of 95 selected radiata pine (*Pinus radiata* D. DON) which were mated in a modified disconnected 5-parent half-diallel design involving a total of 15 diallels. The measurements we report here are from one site planted as a randomized complete block design of 46 replicates as single-tree plots. Details of the development of the "875" series are described in SHELBOURNE *et al.* (1986) and further details are presented in KING *et al.*, 1998.

The statistical analysis uses individual tree measurements and follows the model:

$$Y_{ijk} = \mu + R_i + G_j + G_k + S_{jk} + E_{ijk} \qquad ...[2]$$

where Y_{ijk} is the individual tree observation, μ is the overall mean and; R_i = effect of the ith replication; G_j and G_k are the general combining ability (GCA) effects of the *j*th and *k*th parents; S_{jk} is the specific combining ability (SCA) effect of the *j*th and *k*th parents; and E_{ijk} is the error associated with the *ijk*th tree. All effects were considered random (Model II, GRIFFING, 1956). The expected mean squares and genetic interpretations of a half-diallel design, excluding selfs, where parents are randomly assigned to diallels and the data is balanced (complete) are shown in *Table 1*. Calculation of the additive and dominance genetic variance is straightforward after determining the GCA variance (σ^2_{GCA}) and SCA variance (σ^2_{SCA}). The relationships are as follows:

$$\sigma^2_a = 4\sigma^2_{GCA}$$
$$\sigma^2_d = 4\sigma^2_{SCA}$$

The difficulty in analyzing diallels is that the main genetic treatment, the full-sib cross, is not readily dividable into its component parts; those being the contribution of each parent (parental GCAs) and the deviation of the full-sib value from the mid-parent mean (SCA).

This procedure requires two runs of PROC VARCOMP (SAS Institute, 1990) to complete the analysis. For a single diallel, in the first run, the model includes the categorical variables for replications (REP) and crosses (CROSS). (If row plots or non-contiguous multi-tree plots are used, the replication-by-cross interaction would be included in the model). The SAS code for our single-tree-plot data was:

```
PROC VARCOMP METHOD = TYPE1;
 CLASS REP CROSS;
 MODEL HEIGHT = REP CROSS;
RUN;
```

For this example the TYPE1 method was used so that results would be comparable with the DIALL package which uses ANOVA-based estimators of variance components (SAS also provides maximum-likelihood procedures). The CROSS variable is a unique identifier for each full-sib cross and can be derived from parental variables (MA and PA) to ignore reciprocal effects when parental variables are ordered for the number of parents in the diallel (e.g. for 5-parent diallels MA and PA are 1 to 5) by:

```
IF PA < MA THEN CROSS = 10*PA + MA;
IF MA < PA THEN CROSS = 10*MA + PA;
```

In this example the error term for CROSS is the error mean square (same as REP-by-CROSS mean square). If the CROSS mean square is found to be insignificant then the analysis

Table 1. – Expected mean squares and genetic interpretations of a half diallel (from GRIFFING, 1956).

Source	df	Expected Mean Square
Replication	r-1	
Cross	[n(n-1)/2]-1	$\sigma^2_e + r\sigma^2_{Cross}$
GCA	(n-1)	$\sigma^2_e + r\sigma^2_{sca} + r(n-2)\sigma^2_{gca}$
SCA	n(n-3)/2	$\sigma^2_e + r\sigma^2_{sca}$
Error	(r-1){[n(n-1)/2]-1}	σ^2_e
Total	rn-1	

where: r and n refer to number of replications and parents per diallel respectively.
Genetic interpretations
σ^2_{gca} = Covariance of half-sib families = $^1/_4\ \sigma^2_a$
σ^2_a = additive genetic variance
σ^2_{sca} = Covariance of full-sib families – (2 * (Covariance of half-sib families) = $^1/_4\ \sigma^2_d$
σ^2_d = dominance genetic variance
σ^2_e = error variance
$\sigma^2_{CROSS} = [1/2 - 1/(n+1)]\ \sigma^2_a + {}^1/_4\ \sigma^2_d$

would be complete since genetic variation would not have been observed.

The VARCOMP procedure produces an unbiased estimate of the CROSS variance component (σ^2_{Cross}). The expectation of σ^2_{Cross} is the covariance of individuals within a full-sib family minus the covariance of individuals from different full-sib families. For our simplified model, the covariance of individuals within a full-sib family is one-half the additive variance plus one-quarter the dominance variance ($^1/_2\sigma^2_a + {}^1/_4\sigma^2_d$). The covariance of individuals from different full-sib families is zero if all the full-sib families are unrelated. In a diallel, however, many full-sib families are related to one another. For a five-parent half-diallel, any full-sib family is related to six out of the nine ($^2/_3$) remaining families as half-sibs. Therefore, the value of covariance of individuals from different full-sib families for the five-parent-half-diallel is $^2/_3\ (^1/_4\sigma^2_a) = 1/6\ \sigma^2_a$. Thus $\sigma^2_{Cross} = {}^1/_3\sigma^2_a + {}^1/_4\sigma^2_d$ for our example. This value changes as the number of parents/crosses in the diallel changes. The value for covariance of individuals from different full-sib families for 3, 4, 5, and 6-parent half-diallels are 1/4, 1/5, 1/6, and 1/7 σ^2_a respectively. Note that this covariance = 1/(number of parents +1) σ^2_a. When the data are unbalanced, the following equation can be used to determine the covariance among individuals from different full-sib families for an unbalanced half-diallel without selfs displayed in the upper right diagonal of the crossing matrix:

$$\frac{\left(\sum_{i=1}^{n}\sum_{j=i+1}^{n} X_{ij}\left(\frac{\left(\sum X_{i.}+\sum X_{.i}+\sum X_{j.}+\sum X_{.j}-2X_{ij}\right)}{\sum_{i=1}^{n}\sum_{j=i+1}^{n} X_{ij}}\right)\right)}{\sum_{i=1}^{n}\sum_{j=i+1}^{n} X_{ij}}\ (^1/_4\sigma^2_a)$$

where x_{ij} represents the number of progeny for the family of parent *i* crossed with parent *j*.

Imbalance affects the covariance of individuals from different full-sib families; hence, when the data are not balanced, biased estimates will result if one uses the simple approach, i.e. covariance = 1/(number of parents +1) σ^2_a. However, it requires a considerable amount of imbalance to change this to a great degree. For example, if one cross is missing from a diallel with more than three parents there is no effect at all. The effect of imbalance is demonstrated later with the sample data.

The CROSS sum-of-squares can be further partitioned into general combining ability (GCA) and specific combining ability (SCA) sum-of-squares using a modification of the procedure outlined by SANDERS (1987). Dummy variables, representing the GCA effect of each parent, must be produced for each observation which, when used in the regression model, produces the GCA regression sum-of-squares. For our 5-parent half-diallel example, five dummy variables were formed to represent parents within diallels. Parents within diallels were designated parent 1 through parent 5 (P1 to P5). In any cross where parent 1 was used the P1 value was set to 1, otherwise it was given the value 0. This was done for each parent, so that each cross had two dummy variables which were 1 and three which were 0. The dummy variables whose value was 1 represented the two parents of the cross (SAS code is shown in Appendix 1). PROC VARCOMP was run a second time with P1 through P4 and placed in the model before CROSS. The variable P5 was not necessary since the first four variables are of full rank (only 4 degrees of freedom for parents in a 5-parent diallel). The SAS code to obtain the proper sum-of-squares for a single five-parent half-diallel is as follows:

```
PROC VARCOMP METHOD=TYPE1;
      CLASS REP CROSS P1 P2 P3 P4;
      MODEL HEIGHT = REP P1 P2 P3 P4 CROSS;
RUN;
```

The sum of the Type I sum-of-squares for the variables P1 through P4 represent the GCA sum-of-squares. The CROSS sum-of-squares in this analysis is what remains of the first CROSS sum-of-squares after the GCA sum-of-squares has been removed and represents the SCA sum-of-squares. The cross variance component (σ^2_{Cross}) is an unbiased estimate of σ^2_{sca}. Dominance genetic variation can be estimated as $\sigma^2_d = 4\ \sigma^2_{Cross}$. Unfortunately there is no clean estimate of σ^2_{gca} in the SAS output since the GCA line in the ANOVA is represented by multiple lines in the SAS output, i.e. the GCA effect is the sum of the effects of the four parents. As a result, SAS does not compute an appropriate expected mean square for the GCA mean square or give an accurate estimate of σ^2_{gca}.

An alternative to calculating σ^2_{gca} through the expected mean squares, and the approach we used, is to solve for σ^2_{gca} using information which is available from the two runs of the VARCOMP procedure. In our example, the σ^2_{Cross} estimate from the first VARCOMP procedure estimates of $^1/_3\sigma^2_a + {}^1/_4\sigma^2_d$, and the σ^2_{sca} estimate from the second VARCOMP procedure estimates $^1/_4\sigma^2_d$. The additive variance (σ^2_a) can be estimated from the available information since we have two equations and two unknowns.

$\sigma^2_{Cross} = {}^1/_3\sigma^2_a + {}^1/_4\sigma^2_d$ ...σ^2_{Cross} from the first VARCOMP

$\sigma^2_{sca} = {}^1/_4\sigma^2_d$ ...σ^2_{Cross} from the second VARCOMP

therefore:

$$\sigma^2_{Cross} - \sigma^2_{sca} = {}^1/_3\sigma^2_a + {}^1/_4\sigma^2_d - {}^1/_4\sigma^2_d = {}^1/_3\sigma^2_a.$$

$$\sigma^2_a = 3 \cdot (\sigma^2_{Cross} - \sigma^2_{sca}) \qquad ...[3]$$

Alternatively, the information from the second run of VARCOMP can also be used to generate the mean square for the GCA effect and this used to estimate σ^2_{gca} by equating mean squares to their expected mean squares. This requires additional, and sometimes more complicated steps. For balanced data the GCA expected mean square (EMS) coefficients can be estimated using the formulae from KEMTHORNE and CURNOW (1961) for partial diallels:

$$\text{GCA EMS} = \sigma^2 + r\ \sigma^2_{sca} + [rs(n-2)/(n-1)]\ \sigma^2_{gca} \qquad ...[4]$$

Where r, n, and s refer to the number of replications, parents, and crosses per parent respectively. To obtain σ^2_{gca} estimates in those few cases where the data are balanced (complete) one can use equation [4] to obtain the appropriate coefficient for the σ^2_{gca} component and estimate σ^2_{gca} as:

$$\sigma^2_{gca} = (MS_{gca} - MS_{sca}) / [rs(n-2)/(n-1)] \qquad ...[5]$$

Estimation of σ^2_{gca} using expected mean squares for unbalanced data is more complex for two reasons. First, the coefficients for σ^2_{sca} is usually not the same for the GCA and SCA lines of the ANOVA, thus simply subtracting the SCA mean square from the GCA mean square does not isolate σ^2_{gca}. However, this can be accomplished through some relatively simply matrix algebra (see BARADAT and DESPREZ-LOUSTAU, 1997, for an example). The second complication is that the coefficients for both σ^2_{sca} and σ^2_{gca} are no longer simple formulas. KEULS and GARRETSEN (1977) provide calculations for expected mean squares for an unbalanced diallel, but the calculations do not take into account incomplete replications. While the coefficient for σ^2_{gca} is a formulae; the σ^2_{sca} components require computing the trace of a matrix used in the

formation of the sum-of-squares. The programming for the proper coefficients can be found in SCHAFFER and USANIS (1969). SEARLE (1971) provides an excellent general discussion on variance component estimation with unbalanced data.

Use of the mean or harmonic mean for r, n or s in equation [5] was inadequate for our example data set. Estimates of the GCA variance component coefficient were obtained using the mean number of crosses per parent and the mean number of replications per cross. Both arithmetic and harmonic means were inserted into equation [3]. The estimates were compared with the correct coefficients which were produced by the DIALL program (*Table 2*). The estimates produced with means were very similar to the correct estimates, but the arithmetic mean coefficient estimates were always larger than the correct coefficients (*Table 2*). The correlation of coefficients estimates using the simple and harmonic means with the DIALL (correct) coefficients were 0.9943 and 0.9824 respectively. While the correlations were high, the difference in the coefficients ranged from –12% to +15%, a range we considered too wide for our use.

Table 2. – Estimates of the coefficient for σ^2_{gca} term in the gca Ems of 15 half diallels from a New Zealand progeny test, comparing the correct coefficient computed by DIALL and estimates using means, and harmonic means.

Diallel	Parents	Crosses Made	σ^2_{gca} Coefficient Estimate		
			DIALL *	Means	Harmonic Means
2	5	10	64.94	67.80	66.91
3	5	10	66.91	69.60	68.77
4	5	8	39.24	45.60	40.19
5	4	6	38.11	42.00	41.38
6	5	9	56.31	60.00	58.61
7	5	7	39.50	43.50	41.22
8	5	9	59.42	63.00	61.18
9	5	7	34.02	38.10	34.02
11	5	8	43.03	47.70	44.13
12	5	8	45.39	48.90	45.90
13	5	6	34.17	39.60	29.85
14	5	8	44.29	47.70	46.19
15	5	9	54.44	59.40	56.85
16	5	8	42.19	47.10	41.89
17	5	9	55.72	59.70	56.51

*) DIALL is the correct coefficient obtained from the DIALL program. Means is the estimate using arithmetic means in the equation [rs(n–2)/(n–1)] and harmonic means is the estimate using harmonic means in the equation [rs(n–2)/(n–1)]. Where r, s, and n represent the number of replications, crosses per parent, and parents.

More Complex Models

In many cases one deals with disconnected half-diallels and may wish to pool the analysis over half-diallels. In such cases the model would include parents and crosses nested within diallels. If one was to analyze disconnected diallels as a single diallel the GCA component may become inflated. This inflation would occur if there was a significant difference among the genetic values of the disconnected diallels due to sampling. The sampling effect would be confounded with the GCA component.

The SAS program can be further expanded to multiple sites by expanding the model to include sites, site by cross, and the site by dummy variable interactions (*Table 3*). As in the simple example, two VARCOMP procedures are run; the first to obtain the cross and site · cross variance components; the second to obtain the SCA and SCA interaction components. The equalities for the cross, GCA, and SCA estimates are true for the interaction components. For our 5-parent example:

$\sigma^2_{\text{Site x Cross}} = {}^1/_3\sigma^2_{\text{site x additive}} + {}^1/_4\sigma^2_{\text{site x dominance}}$
.... $\sigma^2_{\text{Cross x Site}}$ from the first VARCOMP
$\sigma^2_{\text{Site x sca}} = {}^1/_4\sigma^2_{\text{site x dominance}}$
.... $\sigma^2_{\text{Cross x Site}}$ from the second VARCOMP

Example Runs

In order to examine the impact of unbalanced data on equation [3], age-7 height and diameter were used to compare the unbiased results from the DIALL program with the GCA variance calculated using equation [3]. In order to increase the number of comparisons each diallel was analyzed with reps 1 to 23 in one analysis and reps 24 to 46 in another to give a total of 56 comparisons. Differences were recorded as percentage of the DIALL estimates and are shown in *table 4*.

Both DIALL and SAS gave the same estimates for σ^2_d and σ^2_e. Estimates of σ^2_a were very similar and never appreciably differed even though the diallels averaged 8.1 completed crosses out of the possible 10 (ranged from 6 to 10 (*Table 2*)). Standard deviation estimates of σ^2_a variance components generated by DIALL were always at least one-half the estimate of σ^2_a (sometimes greater than the σ^2_a estimate); and the largest discrepancy between SAS and DIALL estimates of σ^2_a was 5.25%. The average discrepancy was only -0.27% with a standard deviation of 1.87% (*Table 4*). When the analysis was pooled over all diallels and run over all replications the discrepancy between methods was only 0.13%. Thus the imbalance in the data did not appreciably affect the SAS variance component estimates.

Discussion

SAS was able do all the necessary data manipulations and produce the required ANOVA mean squares. The drawback to SAS was that it could not produce an accurate estimate of the GCA expected mean square (EMS) since our model did not have a single entry for GCA, but used dummy variables to pull out the appropriate sum-of-squares. The standard formulae for estimating the coefficients of the variance components in the EMS are not accurate when the data are unbalanced. However, a close approximation of σ^2_a could be obtained without knowing the coefficient since the two runs of PROC VARCOMP produced two estimates from which σ^2_a could be derived.

The imbalance in our example data set was a result of mortality (85% survival) and in not completing diallels (6 to 10 crosses per diallel). The estimates produced using the method described here were not appreciably different from those produced by the DIALL program which adjusts for imbalance when computing ANOVA based variance components.

In severely unbalanced situations, like those regularly found in forest tree breeding, maximum likelihood estimates provide superior estimates (SWALLOW and MONAHAN, 1984; SEARLE *et al.*, 1992). The ANOVA based estimators remain unbiased but are no longer of minimum variance (READ, 1961). An advantage to using SAS procedures is that maximum likelihood estimation procedures are available. SAS also offers the flexibility of altering the model to reflect fixed and random effects.

Imbalance can either be in the number of crosses per parent or the number of individuals per cross. Missing crosses can be

Table 3. – Expected mean squares of a disconnected half diallel planted on multiple sites

Source	df	Expected Mean Square
Sites	s-1	
Replication(Site)	s(r-1)	
Diallel	d-1	
Diallel x Rep(Site)	s(r-1)(d-1)	
Cross (Diallel)	d{[n(n-1)/2]-1}	$\sigma^2_e + r\,\sigma^2_{Cross\ x\ Site} + rs\,\sigma^2_{Cross}$
GCA (Diallel)*	d(n-1)	$\sigma^2_e + r\,\sigma^2_{Cross\ x\ Site} + rs\,\sigma^2_{sca} + rs(n-2)\,\sigma^2_{gca}$
SCA (Diallel)*	d[n(n-3)/2]	$\sigma^2_e + r\,\sigma^2_{Cross\ x\ Site} + rs\,\sigma^2_{sca}$
Cross (Diallel) x Site	(s-1)d{[n(n-1)/2]-1}	$\sigma^2_e + r\,\sigma^2_{Cross\ x\ site}$
GCA(Diallel) x Site	(s-1)d(n-1)	$\sigma^2_e + r\,\sigma^2_{sca\ x\ site} + r(n-2)\,\sigma^2_{gca\ x\ site}$
SCA(Diallel) x Site	(s-1)d[n(n-3)/2]	$\sigma^2_e + r\,\sigma^2_{sca\ x\ site}$
Error	(s-1)d(r-1){[n(n-1)/2]-1}	σ^2_e
Total	sdrn-1	

*) The GCA and SCA lines are not obtained in the first run of VARCOMP. The SCA line is the CROSS line in the second VARCOMP run. The GCA sums of squares is obtained by summing the dummy variable sum-of-squares in the second VARCOMP run. where: s, d, r, and n refer to the number of sites, diallels, replications, and parents per disconnected diallel respectively.

Table 4. – Percent deviation of SAS variance component estimates from DIALL estimates examined over 56 analyses.

Variance Component	Mean Percentage Difference	Percentage Difference Range	Standard Deviation of Differences
σ^2_a	-0.27%	-5.25 to 3.37	1.87
σ^2_d	0%	0	0
σ^2_e	0%	0	0

a major problem if it severely reduces the degrees of freedom in estimating the SCA component. Reducing the degrees of freedom will reduce the accuracy of the SCA estimate (or the GCA estimate) since the variance of the estimate is inversely related to the degrees of freedom.

The described analysis can be used with other software packages which have the same capabilities as SAS. Other programs designed to analyze diallels have usually been limited to a single statistical package (DEAN and CORRELL, 1988; LINDA, 1993).

By using standard packages, one has greater ability to manage the data and can also overcome model restrictions that are common in special software programs for diallel analysis. For example, using SAS, pooled estimates are readably attainable, effects can be either fixed or random (the MIXED procedure), and different estimation methods are available. These options may not be an available with some specialty programs.

Acknowledgement

We would like to acknowledge Dr. M. D. WILCOX who arranged for the original "875" selections and provided the mating design for this important series. We wish to thank all the people of the FRI and Tasman Forestry Ltd who have helped with this series over the years. We also wish to thank Drs. P. Baradat, COLIN MATHESON, RICHARD SNIEZKO, and BARBARA MCCUTCHAN for reviewing the manuscript. We would like to acknowledge the input and contribution of Dr. A. D. YANCHUK and NANCY MANDEL.

References

BARADAT, P. and DESPREZ-LOUSTAU, M. L.: Analyze diallèle et intégration de la sensibilité à la rouille courbeuse dans le programme d'amélioration du in maritme. Ann. Sci. For. **54**, 83–106 (1997). — DEAN, C. A. and CORRELL, R. L.: Analysis of diallel matings with missing values. Silvae Genet. **37**, 187–192 (1988). — GARRETSEN, F. and KEULS, M.: A general method for the analysis of genetic variation in complete and incomplete diallels and North Carolina II (NC II) designs. Part II. procedures and general formulas for the fixed model. Euphytica **27**, 49–68 (1978). — GRIFFING, G.: Concept of general and specific combining ability in relation to diallel crossing systems. Aust. J. Biol. Sci. **9**, 463–493 (1956). — HUBER, D. A., WHITE, T. L., LITTELL, R. C. and HODGE, G. R.: Oridinary least squares estimation of general and specific combining abilities from half-diallel mating designs. Silvae Genet. **41**, 263–273 (1992). — KEMTHORNE, O. and CURNOW, R. N.: The partial diallel cross. Biometrics **17**, 229–250 (1961). — KEULS, M and GARRETSEN, F.: A general method for the analysis of genetic variation in complete and incomplete diallels and North Carolina II (NC II) designs. Part I. procedures and general formulas for the random model. Euphytica **26**, 537–551 (1977). — KING, J. N., CARSON, M. J. and

JOHNSON, G. R.: Analysis of disconnected diallel mating designs: II – Results from a third generation progeny test of the New Zealand radiata pine improvement programme. Silvae Genetica **47**, 80–87 (1998). — LINDA, S. B.: GRIFFING – a SAS macro implementing GRIFFING's analysis of diallel crossing systems. HortScience **28**, 61 (1993). — READ, R. R.: On quadratic estimates of the interclass variance for unbalanced designs. J. Royal Statist. Soc. B **23**, 493–497 (1961). — SANDERS, W. L.: Analysis of diallel cross designs with empty diagonals via SAS PROC GLM and PROC REG. In: Statistical Considerations in Genetic Testing of Forest Trees. Southern Cooperative Series Bull. No. 324. Dept. of For., School of Forest Resources and Conservation, Univ. of Fla., Gainesville (1987). SAS Institute Inc.: SAS/STAT User's Guide, Version 6, 4th Edition. Cary, NC: SAS Institute Inc. 1686 p. (1990). — SCHAFFER, H. G. and USANIS, R. A.: General least squares analysis of diallel experiments. A computer program – DIALL. N.C. State Univ., Genet. Dep., Res., Rep., 1 (1969). — SEARLE, S. R.: Topics in variance component estimation. Biometrics **27**, 1–76 (1971). — SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E.: Variance Components. John Wiley & Sons Inc., New York. 501p. (1992). — SHELBOURNE, C. J. A., BURDON, R. D., CARSON, S. D., FIRTH, A. and VINCENT, T. G.: Development plan for radiata pine breeding. Ministry of Forestry, Forest Research Institute, Rotorua, NZ. 142 pp. (1986). — SNYDER, E. B. and NAMKOONG, G.: Inheritance in a diallel crossing experiment with longleaf pine. USDA For. Serv. Res. Pap. SO–140 (1978). — SWALLOW, W. H. and MONAHAN, J. F.: Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components. Technometrics **26**(1): 47–57 (1984). — TALBERT, J. T.: An advanced-generation breeding plan for the North Carolina State University – Industry Pine Tree Improvement Cooperative. Silv. Genet. **28**: 72–75 (1979). — YEH, F. C. and HEAMAN, J. C.: Estimating genetic parameters of height growth in seven-year old coastal Douglas-fir from disconnected diallels. Forest Science **33**: 946–957 (1987).

Appendix 1

Routine to Produce Dummy Variables for Partitioning the Sum-of-Squares and Produce the Variance Component Estimates

```
**** OBTAIN A DATA SET WHICH HAS ALL THE PARENTS
IN EACH DIALLEL ****;
DATA ONE;
  SET ORIGINAL;
        KEEP FEMALE PARENT DIALLEL;
        PARENT = FEMALE;
        DROP FEMALE;
RUN;
DATA TWO;
  SET ORIGINAL;
        KEEP MALE PARENT DIALLEL;
        PARENT = MALE;
        DROP MALE;
RUN;
DATA ALL;
 SET ONE TWO;
RUN;
PROC SORT DATA=ALL;
        BY DIALLEL PARENT;
RUN;
DATA ALL;
  SET ALL;
  BY DIALLEL PARENT;
        IF FIRST.PARENT THEN DO;
   RANK+1;
   OUTPUT;
   END;
        ELSE DELETE;
RUN;
****  RECODE PARENT NAMES TO STANDARD NAMES
FOR EACH DIALLEL ****;
PROC RANK DATA=ALL OUT=NEWALL;
        BY DIALLEL;
        VAR PARENT
        RANKS NEW_PAR;
RUN;
**** MERGE STANDARD PARENT NAMES ONTO ORIGI-
NAL DATA ****;
PROC SORT DATA=NEWALL;
        BY DIALLEL PARENT;
RUN;
DATA NEWALL;
  SET NEWALL;
        FEMALE = NEWPAR;
RUN;
PROC SORT DATA=ORIGINAL;
        BY DIALLEL FEMALE;
RUN;
DATA ORIGINAL;
  MERGE ORIGINAL NEWALL;
        BY DIALLEL FEMALE;
        MA = NEWPAR;
        DROP PARENT;
RUN;
DATA NEWALL;
  SET NEWALL;
        MALE = NEWPAR;
        DROP FEMALE;
RUN;
PROC SORT DATA=ORIGINAL;
        BY DIALLEL MALE;
RUN;
DATA ORIGINAL;
  MERGE ORIGINAL NEWALL;
        BY DIALLEL MALE;
        PA = NEWPAR;
        DROP PARENT;
RUN;
****  CONSTRUCT THE DUMMY VARIABLES ****;
DATA ORIGINAL;
  SET ORIGINAL;
        IF PA=1 OR MA=1 THEN P1=1; ELSE P1=0;
        IF PA=2 OR MA=2 THEN P2=1; ELSE P2=0;
        IF PA=3 OR MA=3 THEN P3=1; ELSE P3=0;
        IF PA=4 OR MA=4 THEN P4=1; ELSE P4=0;
        IF PA=5 OR MA=5 THEN GP=1; ELSE P5=0;
RUN;
*** RUN THE FIRST VARCOMP ***;
PROC VARCOMP METHOD=TYPE1;
 CLASS REP  CROSS DIALLEL;
 MODEL HEIGHT = REP DIALLEL CROSS(DIALLEL);
 RUN;
*** RUN THE SECOND VARCOMP ***;
PROC VARCOMP METHOD=TYPE1;
 CLASS REP  CROSS DIALLEL P1 P2 P3 P4;
  MODEL HEIGHT = REP DIALLEL P1(DIALLEL) P2(DIAL-
  LEL)  P3(DIALLEL)
P4(DIALLEL) CROSS(DIALLEL);
 RUN;
```