MULTI-ENVIRONMENT TRIAL ANALYSIS FOR *PINUS RADIATA**

M. DING†, B. TIER,
Animal Genetics and Breeding Unit, University of New England,
Armidale, NSW 2351, Australia

G. W. DUTKOWSKI,
Cooperative Research Centre for Sustainable Forestry, School of Plant Sciences,
University of Tasmania, G.P.O.Box 252-55, Hobart, Tasmania 7001, Australia

H. X. WU,
CSIRO Forest Biosciences**,
P O Box E4008, Kingston ACT 2604, Australia

M. B. POWELL and T. A. McRAE
Southern Tree Breeding Association Inc.
P O Box 1811, Mount Gambier SA 5290, Australia

(Received for publication 20 December 2007; revision 5 February 2008)

ABSTRACT

A stem-diameter data set of five combined trials of *Pinus radiata* D. Don was used to identify and determine the nature of genetics by environment (G×E) interaction. The restricted maximum likelihood approach was applied to handle the main issues of the multi-environment trial analysis:

(1) Testing sources of heterogeneity of variance and lack of between-sites genetic correlation;

(2) Modelling the heterogeneity of error variance among trials and micro-environmental variation within each trial; and

(3) Selecting the best model for prediction of breeding values.

Model comparison was based on the criterion of log-likelihood. The significance of variance components was tested by the likelihood ratio test which showed that all sources of G×E interactions were highly significant, indicating that G×E interactions occurred in these five trials due to both the heterogeneity of variances and the lack of correlation. Estimates of Type B genetic correlations were increased slightly by correcting for the heterogeneity of variances. The full model, which accommodated heterogeneity of error

* Paper originally presented at the inaugural Australasian Forest Genetics Conference "Breeding for Wood Quality", Hobart, 11–14 April 2007.

† Corresponding author: mding@une.edu.au

**Research carried out under the Ensis UJV

variances between trials, spatial variation within trials, and fitting a separate G×E interaction variance for each trial, was superior to other models for this multi-environment trial

Keywords: log-likelihood; G×E interaction; *Pinus radiata*.

INTRODUCTION

Genotype by environment (G×E) interaction refers to differential responses of different genotypes across a range of environments (Kang 2004). These G×E interactions can be distinguished by whether the interactions are treated as either (1) a source of error or bias in assessing a genotype (random, non-repeatable G×E interactions), or (2) a component of variation which is, in part, heritable and exploitable through selection for broad and specific adaptation (repeatable G×E interactions — Delacy, Basford, Cooper, Bull & Mclaren 1996). Only the repeatable G×E interactions are essential and meaningful for breeding strategies (Baker 1988).

Determining the relative proportions of repeatable and non-repeatable G×E interaction effects is an important issue in analysis and interpretation of multi-environment trials. This partitioning was first shown by Robertson (1959). Muir *et al*. (1992) gave methods for partitioning G×E interaction into sources due to heterogeneous variances and lack of correlation. Yang & Baker (1991) used multivariate analysis of variance (MANOVA) and proposed two tests for the significance of the different sources of G×E interaction. These approximate tests are based on unwarranted assumptions about the sampling distributions of estimated variance and covariance components, resulting in a number of undesirable properties such as non-positive definite estimates of genetic variance-covariance matrices. Therefore, Yang (2002) applied a restricted maximum likelihood (REML) approach to estimate genetic parameters and test the significance of different sources of G×E interaction.

The restricted maximum likelihood approach (Patterson & Thompson 1971) has been used for decades to estimate variance parameters based on mixed model theory (Henderson 1984). Mixed model analysis for multi-environment trial data contains frequent approaches in which the variance parameters are estimated using restricted maximum likelihood (Smith *et al*. 2005). Fixed and random effects are estimated using best linear unbiased estimates (BLUEs) and best linear unbiased predictors (BLUPs), respectively. The development of statistical packages such as ASREML (Gilmour *et al*. 1999) allows restricted maximum likelihood estimation of a range of mixed models, and also enables the fitting of more informative and complex models for accommodating different forms of G×E interactions. Cullis *et al*. (1998) allowed for heterogeneity between trials by fitting a separate variety by environment (V×E) interaction variance for each trial. Smith *et al*. (2001)

extended this approach for the analysis of multi-environment trial data which included multiplicative models for the variety effects in each environment. The model provides an approach that accommodates heterogeneity of V×E variance, correlation among V×E interactions, and appropriate error variance structures for individual trials.

In fact, the residual variation can be further partitioned into components due to micro-environment variation and genotype by micro-environment interaction (Nyquist 1991). Variation within trial has been examined by some authors using spatial analysis on single sites (Casanoves *et al*. 2005; Cullis *et al*. 1998; Smith *et al*. 2001). In forestry there is some evidence of gradients and large patch sizes within trials (Costa e Silva *et al*. 2001; Fu *et al*. 1999), and use of a combined spatial model enables an improved analysis of experiment data (Dutkowski *et al*. 2002, 2006; Costa e Silva *et al*. 2001; Hamann *et al*. 2002; Magnussen 1990).

Restricted maximum likelihood approaches based on mixed models allow more flexible variance structures, which are helpful for fitting G×E interactions. However, most applications in forestry have focused on quantifying the relative size of G×E interaction (Carson 1991; Haapanen 1996; Johnson & Burdon 1990; Matheson & Raymond 1984; Pederick 1990). There are few studies on identifying and partitioning the sources of G×E interaction, despite the fact that it has been recognised for a long time that they influence efficient decision-making in breeding programmes, and rapid genetic advance.

The goal of this study was to identify repeatable G×E interaction and it focused on three aspects:

(1) testing the sources of G×E interaction;

(2) selecting the best models for multi-environment trials;

(3) investigate the impact of fitting alternative models on estimates of variances, genetic parameters, and the parameter used for measuring relative magnitude of G×E interaction.

MATERIALS AND METHODS

The genetic materials originated from an Australia-wide diallel mating experiment. The details have been given by Wu & Matheson (2005). Five typical sites (PT5459, RAD211, VRC060, RS27A, and RS27B) were chosen and combined for this study. They were distributed in four regions in Australia and contained a total of 12460 genotypes, with from 165 to 216 full-sib families represented at each site. Each trial was a randomised incomplete-block design with three replicates and four-tree row plots at a spacing of 3.0 × 3.0 m, excepting VRC060 at 3.6 × 2.3 m). Trials PT5459, RAD211, and VRC060 had the same block numbers within each replicate. In trials RS27A and RS27B, the block numbers were continuous across

the trial. Each individual tree was marked on a grid of R rows within C columns in each trial. Diameter at breast height (dbh) was measured at 10.5 years of age (for details see Table 1).

All check lots were eliminated from the data set before analysis. For trials with an irregular shape, the data were expanded to construct a complete rectangular matrix by inserting missing values using BLOCKIT (Dutkowski 2004) for spatial analysis. The proportion of missing values in these rectangles ranged from 0 to 61%.

TABLE 1–Design information for five trials, and mean diameter at breast height with standard deviation

Site	PT5459	RAD211	VRC060	RS27A	RS27B
Observations	2536	2624	2592	3260	1652
Missing values	236	554	565	1643	541
Replicates	3	3	3	3	3
Blocks	18	18	18	108	39
Plots	648	656	648	648	546
Plot size (m^2)	4	4	4	4	4
Rows	96	97	36	83	49
Columns	60	68	72	55	54
Spatial rate (%)	56	61	0	29	38
Number of families	216	216	216	216	169
Spacing (m)	3 × 3	3 × 3	3.6 × 2.3	3 × 3	3 × 3
Measured age (years)	10.5	10.5	10.5	10.5	10.5
DBH (mm)	174 ± 22	158 ± 34	201 ± 44	233 ± 41	242 ±3 5

Models

Family and individual tree models were fitted. The general linear mixed model in matrix notation is:

$$\mathbf{y = Xb + Zu + e} \quad [1]$$

where **y** is a vector of observations for diameter at breast height,

b and **u** are vectors of fixed (trial, replicate within trial) and random effects, respectively,

X and **Z** are design matrices relating the observations to the fixed and random effects, respectively,

e is a vector of random residuals.

For the family models, the random effect vector **u** had sub-vectors of block (within replicate for PT5459, RAD211, and VRC060 but not within replicate for RS27A and RS27B), family, family × trial, and family × replicate (nested within trial) interaction. For the individual tree models, the random effects include block, plot, tree, family, tree × trial, and family × trial interaction. Separate model terms in **u** were assumed to be uncorrelated. It was assumed that the joint distribution of the random effects was Gaussian with zero mean and variance matrix,

$$\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} \sim \mathbf{N}\left[\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{G}\ \mathbf{0}\\ \mathbf{0}\ \mathbf{R}\end{bmatrix}\right]$$

where **R** is the variance-covariance matrix of the residuals and **G** is the variance-covariance matrices of the random effects. For the family model, the variances for effects in **u** and **e** are:

$$\mathbf{Var}\begin{bmatrix}\mathbf{inb}\\ \mathbf{f}\\ \mathbf{t.f}\\ \mathbf{t.r.f}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G_{inb}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G_f} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{G_{t.f}} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{G_{t.r.f}} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}\end{bmatrix} \quad [2]$$

where inb, f, t.f, and t.r.f are the effects for block, family, family by trial, and family by replicate within trial, respectively. When independence of the sub-vectors is assumed, then $\mathrm{Var}(\mathbf{inb}) = \mathbf{I_{inb}}\sigma^2_{inb}$, $\mathrm{Var}(\mathbf{f}) = \mathbf{I_f}\sigma^2_f$, $\mathrm{Var}(\mathbf{t.f}) = \mathbf{I_{t.f}}\sigma^2_{t.f}$, and $\mathrm{Var}(\mathbf{t.r.f}) = \mathbf{I_{t.r.f}}\sigma^2_{t.r.f}$, where $\mathbf{I_{inb}}$, $\mathbf{I_f}$, $\mathbf{I_{t.f}}$, and $\mathbf{I_{t.r.f}}$ are identity matrices of appropriate order, with constant variances σ^2_{inb}, σ^2_f, $\sigma^2_{t.f}$, and σ^2_{trf}.

For the individual tree model, additional random effects of plot (p), tree (g), and tree × trial (t.g) were included, the additional variance structures to be defined, $\mathrm{Var}(\mathbf{p}) = \mathbf{I_p}\sigma^2_p$, $\mathrm{Var}(\mathbf{g}) = \mathbf{A}\sigma^2_a$, $\mathrm{Var}(t.g) = I_{t.g}\sigma^2_{t.g}$, where σ^2_a is the additive genetic variance, **A** is the numerator relationship matrix, $\mathbf{I_p}$ and $\mathbf{I_t}$ are identity matrices.

There are different possible forms for the genetic variance matrix. The above independent assumption for genetic variance matrix of family (or tree) and family × trial (or tree × trial) interaction can be expressed as $\mathrm{Var}(\mathbf{F}) = (\sigma^2_f\mathbf{J}_t + \sigma^2_{t.f}\mathbf{I}_t)\ \mathrm{O}\mathbf{I}_f$, where $\mathbf{J}_t$ is a t × t unit matrix (i.e., all elements equal to one and Θ is the Kronecker product (Searle *et al.* 1992)). This variance structure is known as a compound symmetry structure (Smith *et al.* 2001). It implies that all the interaction effects have the same variance and for different families (or trees) are uncorrelated, and interaction effects for different pairs of environments all have the same covariance (Smith *et al.* 2005). The magnitude of G×E interactions can be estimated through the size of the estimates of variance components. The most general form for genetic (co)variance matrices allows correlations between trials, contains p(p+1)/2 parameters, e.g., the family (co)variance structure for i trials under the family model can be expressed as:

$$\mathrm{Var}\begin{bmatrix}\mathbf{f_1}\\ \mathbf{f_2}\\ \vdots\\ \mathbf{f_i}\end{bmatrix} = \begin{bmatrix}\sigma^2_{f1} & \sigma_{f12} & \cdots & \sigma_{f1j}\\ \sigma_{f21} & \sigma^2_{f2} & \cdots & \sigma_{f2j}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{fi1} & \sigma_{fi2} & \cdots & \sigma^2_{fi}\end{bmatrix} \mathbf{\Theta}\ \mathbf{I_f} \quad [3]$$

where the diagonal element σ^2_{fi} is the family variance in environment i, off-diagonal element σ_{fij} is the family covariance between environments i and j, $\mathbf{I}_f$ is an f × f identity matrix.

When accommodating heterogeneity of residual variances between trials, **R** can be expressed as:

$$\text{Var}\begin{bmatrix}\mathbf{e_1}\\ \mathbf{e_2}\\ \\ \mathbf{e_i}\end{bmatrix} = \begin{bmatrix}\mathbf{R_1} & & & \\ & \mathbf{R_2} & & \\ & & \ddots & \\ & & & \mathbf{R_i}\end{bmatrix} \quad [4]$$

where $\mathbf{R_1}$, $\mathbf{R_2}$, ..., $\mathbf{R_i}$ represent the residual variance matrices for site 1, 2, …, i, respectively. Allowing within-trial error variance, $\mathbf{R_i}$ have a different structure based on a decomposition of **e** into spatially dependent (**ç**) and independent (**î**) residuals (Costa e Silva *et al*. 2001; Dutkowski *et al*. 2006). The $\mathbf{R_i}$ matrices are:

$$\mathbf{R_i} = \sigma^2_{çi}[\textstyle\sum_c(\tilde{n}_{col}) \Theta \sum_r (\tilde{n}_{row}))] + \sigma^2_{îi}\ \mathbf{I} \quad [5]$$

where $\sigma^2_{çi}$ is the spatial residual variance in site i, $\sigma^2_{îi}$ is the independent residual variance of the "white noise" process in site i, **I** is an identity matrix, and $\sum(\rho)$ is a first-order autoregressive correlation matrix with autocorrelation ρ, for a random factor spatial ordered in one dimension with n levels, the form is:

$$\text{AR}(\rho) = \begin{bmatrix} 1 & \tilde{n} & \tilde{n}^2 & \tilde{n}^3 & \dots & \tilde{n}^{n-1} \\ \tilde{n} & 1 & \tilde{n} & \tilde{n}^2 & \dots & \vdots \\ \tilde{n}^2 & \tilde{n} & 1 & \tilde{n}^3 & \dots & \vdots \\ \tilde{n}3 & \tilde{n}2 & \tilde{n}3 & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{n}^{n-1} & \dots & \dots & \dots & \dots & 1 \end{bmatrix} \quad [6]$$

The independent errors in trials PT5459, RAD211, and VRC060 were significant in the previous single site spatial analysis, which was fitted in multi-site analysis.

Estimates of the fixed and random effects in [1], the solution to the mixed model equations, have been given by Henderson (1984):

$$\begin{bmatrix}\mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z + G^{-1}}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{b}} \\ \hat{\mathbf{u}}\end{bmatrix} = \begin{bmatrix}\mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y}\end{bmatrix} \quad [7]$$

This leads to best linear unbiased estimates of the fixed effects and best linear unbiased predictors of random effects. We compared the impact of fitting alternative models on best linear unbiased predictors of the genetic effects. The parameters in **G** and **R** were replaced by estimates from data, using ASREML (Gilmour et al. 1999).

Statistical Analyses

Three series of models were fitted. The series one models with specific constraints are listed in Table 2, giving different forms of **R** and **G** structures. Series one included five alternative family models for testing sources of G×E interactions. Series two included three family models with both family and family × trial effects to measure

TABLE 2–Full model and four reduced models for testing significance of G×E interaction, and examining causes of G×E interaction due to heterogeneity of variances and lack of genetic correlations between sites

Model	Null hypothesis	Constraint of R* and C† structures	Estimated parameters
S1M0	All the sources were allowed to be heterogeneous	R structure; CORGH† correlation structure	σ^2_{f1} σ^2_{f2}…σ^2_{f5}; ρ_{ij}; σ^2_{e1}…σ^2_{e5}
S1M1	No G×E interaction	R structure; $\rho = 1$; CORUV† structure; σ^2_f is the same across sites	σ^2_f; σ^2_{e1}…σ^2_{e5}
S1M2	Perfect genetic correlation between sites	R structure; $\rho = 1$; CORUH† structure;	σ^2_{f1} σ^2_{f2}…σ^2_{f5}; σ^2_{e1}…σ^2_{e5}
S1M3	Homogeneity of family variances	R structure; CORGV† structure; σ^2_f is the same across sites	σ^2_f; ρ_{ij}; σ^2_{e1}…σ^2_{e5}
S1M4	Homogeneity of error variances across sites	σ^2_e is the same across sites; CORGH† structure	σ^2_{f1} σ^2_{f2}…σ^2_{f5}; ρ_{ij}; σ^2_e

* Residual variance structure; R structure, heterogeneity of residual variances across sites.
† Correlation structure; details are given in Appendix A.

the relative size of G×E interactions. Series three fitted five individual tree models to select the best model for obtaining best linear unbiased predictors of genetic values. The different forms of correlation structures are listed in Appendix A.

For series one, the statistical procedures started from the full model to a series of reduced models. The full model (S1M0) allowed the residual variances to be heterogeneous in each trial; the family variances and covariances were different and correlated. Reduced model 1 (S1M1), with the hypothesis of no G×E interaction, allowed heterogeneity of error variances between trials but constrained the family variances and covariances to be the same across trials. Reduced model 2 (S1M2) had the hypothesis of perfect family correlation among trials, the family covariances were the same, and correlation was constrained to 1 but allowed the residual and family variances to be heterogeneous. Reduced model 3 (S1M3) had of the family variances being homogeneous among trials; family variances were the same among five trials but allowed the heterogeneity of error variance between trials and family covariances to be different. Reduced model 4 (S1M4) had the hypothesis that residual variances were homogeneous; residual variances among trials were constrained to be the same but allowed the family variances and covariances to be different among trials.

For series two, the procedures started from the simple model, assuming residual variances were homogeneous (S2M1). Model 2 (S2M2) accommodated heterogeneity of error variance between trials. Model 3 (S2M3) included the spatial variation within trial nested in model 2. The random effects included family and family × trial interactions. The relative size of G×E interactions can be estimated by ratio (K) which is the interaction variance by family variance.

For series three, the first three procedures (S3M1, S3M2, S3M3) had the same **R** structures as in series two. Model 4 (S3M4) was nested S3M2, giving correlation structure, allowing the genetic variances in each trial and covariances between the pair of trials to be different. Model 5 (S3M5) was nested model 3 (S3M3) with the same correlation structure as model 4.

Models were compared using the criterion of log-likelihood. Significance was tested by likelihood ratio test (LRT). For series one, the likelihood test ratio is:

$$\text{LRT} = -2*(\text{logL. of full model} - \text{logL. of reduced model}) \quad [8]$$

For series two and three, the likelihood test ratio is:

$$\text{LRT} = -2*(\text{log L. of model (i)} - \text{logL. of model (i+1)}) \quad [9]$$

where model i and model i+1 are with and without the tested component or structure, respectively. Under the null hypothesis, the likelihood test ratio is expected to be distributed as χ^2_q with degrees of freedom (q) given by the difference between the numbers of variance and covariance parameters (Kendall & Stuart 1979). When the likelihood test ratio $>\chi^2_q$, the additional estimated variance component is significant.

RESULTS

The results of series one with log-likelihoods, likelihood ratio tests, and estimates of variance parameters are given in Table 3. The likelihood ratio tests showed all the tested sources were highly significant. The first test indicated that G×E interactions were present across five sites. Moreover, the significance of heterogeneity of genetic and error variances, and lack of correlation, indicated the G×E interactions occurring in this multi-environment trial — the heterogeneity of genetic and error variances, and lack of correlation. In most sites, the family variances in the full model (S1M0) were higher than the estimates in other reduced models, whereas the estimates of error variances in the full model were lower than in the reduced models.

The results estimated from the second series of models are shown in Table 4. The family variances ($\hat{\sigma}^2_f$ = 27.2, 27.1, and 27.2) were stable across the three models, with a slight decrease in the standard error from fitting the homogeneity (S2M1), heterogeneity (S2M2), and spatial model (S2M3). This indicated that relaxing the heterogeneity of error variance did not influence the estimates of family variances. However, the interaction variance ($\hat{\sigma}^2_{fe}$) decreased from 53.7 to 37.4 and 33.4, resulting in a decreased ratio of interaction to family variance from 197.7% to 137.7% and 122.6% in models S2M1, S2M2, and S2M3 respectively.

The results of series three are given in Table 5. The model S3M2 with heterogeneous residual variances provided a much better fit than S3M1, with a large increase in log-likelihood (LRT = 977.2 on 4 df, $p < 0.001$). Accommodating spatial variation within each trial, the model S3M3 gave a further significant improvement (LRT = 30.4 on 3 df, $p < 0.01$). The nested model S3M4 with a correlation structure provided a better fit than model S3M2 (LRT = 37.9 on 15 df, $p < 0.01$). Model S3M5, which accommodated heterogeneity of error variance between trials, spatial variation within trials, and correlation structure for individual trees and families was the best model for estimating breeding values.

The estimates of genetic correlations between pairs of sites are presented in Table 6. The average of genetic correlations changed slightly when fitting different models, while the (co)variances were constrained in different ways. It was lowest (0.34) when assuming all the variances were homogeneous (all the heterogeneities confounded). After correcting for heterogeneity of error variance between trials, the average genetic correlation increased to 0.38. It was 0.35 and 0.36 after correction for the heterogeneity of error and genetic variances, and spatial variation within trials, respectively. For all models, the highest genetic correlation occurred between sites RAD211 and VRC060. The genetic correlation between sites RS27A and RS27B increased after correcting for heterogeneity of error variance, from 0.51 to 0.75. For other pairs of sites, the trend of change resulting from fitting different models was similar.

TABLE 3–Likelihood ratio tests for heterogeneity and lack of correlation and the estimates of variance parameters in model series one

Model	Sources of G×E interaction	logL.	LRT†	$\hat{\sigma}^2_{f1}$	$\hat{\sigma}^2_{f2}$	$\hat{\sigma}^2_{f3}$	$\hat{\sigma}^2_{f4}$	$\hat{\sigma}^2_{f5}$	Mean ñ‡	$\hat{\sigma}^2_{e1}$	$\hat{\sigma}^2_{e2}$	$\hat{\sigma}^2_{e3}$	$\hat{\sigma}^2_{e4}$	$\hat{\sigma}^2_{e5}$
S1M0	Total G×E interaction	–6066.63		43.2	102.1	45.0	126.6	171.1	0.35	419.9	1013.0	1152.9	1735.1	1441.4
S1M1	Presence of G×E interaction	–6100.05	66.84 ***	$\hat{\sigma}^2_f$ = 33.7					0.999	414.8	1032.5	1138.3	1777.4	1554.7
S1M2	Lack of perfect correlation	–6095.88	58.50 **	25.5	66.5	36.5	39.1	1.35	0.999	416.6	1013.0	1134.9	1780.2	1569.1
S1M3	Heterogeneity of family variance	–6077.25	21.24 **	$\hat{\sigma}^2_f$ = 66.9					0.38	416.6	1026.8	1143.6	1770.3	1502.7
S1M4	Heterogeneity of error variance	–6519.60 **	905.94	3.3	93.6	54.0	221.0	228.4	§	$\hat{\sigma}^2_e$ = 1047.2				

* P<0.05; ** P<0.01; *** P<0.001.
† LRT (likelihood ratio tests) are comparisons between the full model and reduced models
‡ Average of genetic correlation between sites.
§ Not sufficient data to support the analysis

TABLE 4–Estimates of the variance components in model series 2 with standard error and variance ratio ($\sigma^2_{fe} / \sigma^2_f$)

Model	LRT†	$\hat{\sigma}^2_{fe}$	$\hat{\sigma}^2_f$	K (%)
S2M1	0	53.7±11.0	27.2±6.9	197.7
S2M2	976.9 ***	37.4±8.1	27.1±6.1	137.7
S2M3	33.3 ***	33.4±7.9	27.2±6.0	122.6

* P<0.05; ** P<0.01; *** P<0.001

† Likelihood ratio test compared with previous model, e.g., S2M2 compared with S2M1; S2M3 compared with S2M2.

DISCUSSION

This study shows that the G×E interactions are associated with both heterogeneity of variances and lack of correlation. The evidence is reflected by the ratios (K) of estimated family × trial interaction variance to family variance. The estimates of the ratios decrease by approximately 30% after correcting for the heterogeneity of residual variance between sites. By further partitioning and removing the spatial variation within each site, the ratio decreases by approximately 11%. Eisen & Saxton (1983) showed that heterogeneous variances can cause the estimates of G×E interaction components to be biased upward, and genetic correlation between trials to be biased downward. In this study, the genetic correlations between sites were biased by the heterogeneity of error variance but changed slightly. The correction for the heterogeneity of family variances had a singularity problem after the first iteration; the influence was not detected by this procedure. However, if both the heterogeneity of family and error variances are removed, the Type B genetic correlations decrease slightly rather than increase, contrary to expectation.

The power of the restricted maximum likelihood approach allowed different forms of residual covariance matrix (**R**) and genetic covariance matrix (**G**). As a result, a range of more complex and informative models could be used to better explore the G×E interactions. With the flexible correlation structures such as using the ASREML package with constrained and/or unconstrained (co)variances, sources of heterogeneity and lack of correlation can be partitioned and tested. In practice, it is of great value to partition the G×E interaction component, particularly where the G×E interactions are attributable to lack of correlation, because it measures the degree to which performance in one environment fails to predict performance in the other. If this is found to be the case then the strategy of selection for broad adaptation need to be re-evaluated to determine whether it is necessary to add a specific component of selection (Delacy, Cooper & Basford 1996). Meanwhile, the genetic correlations between sites can also be estimated by correcting for heterogeneity of variances. These analyses should be more realistic; the presence of

TABLE 5–Comparison of individual tree models with the variance parameters in model series 3

Model	LogL.	LRT†	$\hat{\sigma}^2_{g1}$ $\hat{\sigma}^2_{g2}$ $\hat{\sigma}^2_{g3}$ $\hat{\sigma}^2_{g4}$ $\hat{\sigma}^2_{g5}$	$\hat{\sigma}^2_{f1}$ $\hat{\sigma}^2_{f2}$ $\hat{\sigma}^2_{f3}$ $\hat{\sigma}^2_{f4}$ $\hat{\sigma}^2_{f5}$	$\hat{\sigma}^2_{e1}$ $\hat{\sigma}^2_{e2}$ $\hat{\sigma}^2_{e3}$ $\hat{\sigma}^2_{e4}$ $\hat{\sigma}^2_{e5}$
S3M1	–6562.90	0	$\hat{\sigma}^2_g$ =23.9	$\hat{\sigma}^2_f$=22.5	$\hat{\sigma}^2_e$ =1027.2
S3M2	–6074.30	977.2***	$\hat{\sigma}^2_g$ =18.6	$\hat{\sigma}^2_f$ =22.2	390.9, 1007.4, 1122.3, 1736.3, 1478.5
S3M3	–6059.12	30.4 **	$\hat{\sigma}^2_g$ =20.5	$\hat{\sigma}^2_f$ =22.4	384.7, 965.5∞, 1120.5, 1736.9, 1478.3
S3M4	–6055.34	37.9‡**	33.2, 43.4, §, 112.5, 197.6	26.7, 80.4, 46.1, 72.0, 75.1	405.5, 996.4, 1156.3, 1680.2, 1341.1
S3M5	–6043.00	32.2‡ **	28.3, 68.9, 1.7, 132.6, 258.6	$\hat{\sigma}^2_f$ =9.6	399.5, 949.9∞, 1154.1, 1681.6, 1339.2

* P<0.05; ** P<0.01; *** P<0.001

† Normally, likelihood ratio tests compared with previous model

‡ S3M4 compared with S3M2, S3M5 compared with S3M3

§ The genetic variance in site 3 was excluded.

∞The spatial errors are 84.7 and 77.4 in S3M3 and S3M5, respectively

TABLE 6–Estimates of genetic correlation of family mean between 10 pairs of sites, based on correcting for heterogeneity of variances

Paired sites	PT5459 RAD211	PT5459 VRC060	T5459 RS27A	T5459 RS27B	RAD211 VRC060	RAD211 S27A	RAD211 RS27B	VRC060 RS27A	VRC060 RS27B	RS27A RS27B	***average***
Model 1	0.35	0.72	0.36	0.01	0.90	0.23	0	0.38	0	0.51	0.34
Model 2	0.41	0.66	0.60	0.08	0.75	0.33	-0.04	0.35	-0.11	0.75	0.38
Model 3	0.39	0.72	0.52	0.09	0.78	0.24	-0.02	0.30	-0.10	0.55	0.35
Model 4	0.45	0.71	0.49	0.08	0.86	0.25	-0.04	0.31	-0.08	0.55	0.36

Model 1 – Assuming homogeneity of family and error variances
Model 2 – Correcting heterogeneity of error variances among trials
Model 3 – Correcting heterogeneity of family and error variances among trials
Model 4 – Correcting heterogeneity of family and error variances among trials, and spatial variation within trial

variance heterogeneity among G×E interactions has been recognised by many authors (Cullis *et al*. 1998; Frensham *et al*. 1997; Patterson & Nabugoomu 1992).

However, fitting complex models can cause some computational problems. For example, the individual tree model did not converge initially when using the complex variance structures. These problems are more frequent, particularly for the spatial model, with independent error variance. Therefore, care must be taken with some constraints; they may be used only for some significance tests and are not universally advocated for estimation of parameters.

CONCLUSION

The restricted maximum likelihood approach is powerful and flexible for partitioning and testing the sources of G×E interaction. The sources of heterogeneity of genetic and error variances, and lack of correlation are all significant, indicating the G×E interactions occurring in the five trials are associated with repeatable (lack of correlation) and non-repeatable (heterogeneity) G×E interaction.

In conclusion, the model that accommodated heterogeneity of error and family (co)variances among trials with a correlation structure was the best model for estimating genetic variances for the five trials analysed here. Incorporation of spatial variation into the model could further increase the precision in estimating breeding values.

REFERENCES

BAKER, R.J. 1988: Analysis of genotype-environmental interactions in crops. *ISI Atlas of Science: Animal and Plant Sciences 1:* 1–4.

CARSON, S. D. 1991: Genotype × environment interaction and optimal number of progeny test sites for improving *Pinus radiata* in New Zealand. *New Zealand Journal of Forestry Science 21:* 32–49.

CASANOVES, F.; MACCHIAVELLI, R.; BALZARINI, M. 2005: Error variation in multienvironment peanut trials: within-trial spatial correlation and between-trial heterogeneity. *Crop Science 45:* 1927–1933.

COSTA E SILVA, J.; DUTKOWSKI, G.W.; GILMOUR, A.R. 2001: Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. *Canadian Journal of Forest Research 31:* 1887–1893.

CULLIS, B.R.; GOGEL, B.J.; VERBYLA, A.P.; THOMPSON, R. 1998: Spatial analysis of multi-environment early generation trials. *Biometrics 54:* 1–18.

DELACY, I.H.; COOPER, M.; BASFORD, P.K. 1996: Relationships among analytic methods used to study genotype-by-environment interactions and evaluation of their impact on response to selection. Pp. 51–84 *in* Kang, M.S.; Gauch, H.G. (Ed.) "Genotype-by-Environment Interaction: New Perspectives". CRC Press, Boca Raton, Florida.

DELACY, I.H.; BASFORD, K.E.; COOPER, M.; BULL, J.K.; MCLAREN, C.G. 1996: Analysis of multi-environment trials — an historical perspective. *In* Cooper, M.; Hammer, G.L. (Ed.) "Plant Adaptation and Crop Improvement". CAB INTERNATIONAL, Cambridge.

DUTKOWSKI, G.W.; COSTA E SILVA, J.; GILMOUR, A.R.; LOPEZ, G.A. 2002: Spatial analysis methods for forest genetic trials. *Canadian Journal of Forest Research 32*: 2201–2214.

DUTKOWSKI, G.W.; COSTA E SILVA, J.; GILMOUR, A.R.; WELLENDORF, H.; AGUIAR, A. 2006: Spatial analysis enhances modeling of a wide variety of traits in forest genetic trials. *Canadian Journal of Forest Research 36:* 1851–1870.

EISEN, E.J.; SAXTON, A.M. 1983: Genotype by environment interactions and genetic correlation involving two environmental factors. *Theoretical & Applied Genetics 67:* 75–86.

FRENSHAM, A.B.; CULLIS, B.R.; VERBYLA, A.P. 1997: Genotype by environment variance heterogeneity in a two-stage analysis. *Biometrics 53:* 1373–1383.

FU Y.B.; YANCHUK, A.D.; NAMKOONG, G. 1999: Spatial patterns of tree height variations in a series of Douglas-fir progeny trials: implications for genetic testing. *Canadian Journal of Forest Research 29:* 714–723.

GILMOUR, A.R.; CULLIS, B.R.; WIELHAM, S.J.; THOMSON, R. 1999: ASREML reference manual. *NSW Agriculture Biometric Bulletin 3*.

HAAPANEN, M. 1996: Impact of family-by-trial interaction on the utility of progeny testing methods for Scots Pine. *Silvae Genetica 45:* 130–135.

HAMANN, A.; KOSHY, M.P.; NAMKOONG, G. 2002: Improving precision of breeding values by removing spatially auto-correlated variation in forestry field experiments. *Silvae Genetica 51:* 210–215.

HAMMER G.L.; COOPER, M. 1996: "Plant Adaptation and Crop Improvement". CAB INTERNATIONAL, Wallingford.

HENDERSON, C.R. 1984: Application of linear models in animal breeding values. *Biometrics 32:* 69–83

JOHNSON, G.R.; BURDON, R.D. 1990: Family-site interactions in *Pinus radiata*: Implications for progeny testing strategy and regionalized breeding in New Zealand. *Silvae Genetica 39:* 55–62.

KANG, M.S. 2004: Breeding: genotype-by-environment interaction. Pp. 218–221 *in* Goodman, R.M. (Ed.) "Encyclopedia of Plant and Crop Science". New York, Basel.

KANZLER, A. 2002: "Genotype × Environment Interaction in *Pinus patula* and its Implications in South Africa". North Carolina State University, USA.

KENDALL, M.G.; STUART, A. 1979: "The Advantage Theory of Statistics. Vol. 2. Inference and Relationship". Griffin and Co., London.

MAGNUSSEN, S. 1990: Application and comparison of spatial models in analyzing tree-genetics field trials. *Canadian Journal of Forest Research 20*.

MATHESON, A.C.; RAYMOND, C.A. 1984: The impact of genotype × environment interactions on Australian *Pinus radiata* breeding programs. *Australian Forest Research 14:* 11–25.

MUIR, W.; NYQUIST, W.E.; XU, S. 1992: Alternative partitioning of the genotype-by-environment interaction. *Theoretical & Applied Genetics 84:* 193–200.

NYQUIST, W.E. 1991: Estimation of heritability and prediction of selection response in plant populations. *Critical Reviews in Plant Sciences 10:* 235–322.

PATTERSON, H.; NABUGOOMU, F. 1992: REML and the analysis of series of crop variety trials. Pp. 77–93 *in* "Proceedings from the 16th International Biometric Conference", Hamilton, New Zealand.

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

PEDERICK, L.A. 1990: Family × site interactions *Pinus radiata* in Victoria, Australia, and implications for breeding strategy. *Silvae Genetica 39*: 134–140.

ROBERTSON, A. 1959: The sampling variance of the genetic correlation coefficient. *Biometrics 15:* 469–485.

SEARLE, S.R.; GASELLA, G.; McCULLOCH, C.E. 1992: "Variance Components". John Wiley & Sons, New York.

SMITH, A.; CULLIS, B.; THOMPSON, R. 2001: Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics 57:* 1138–1147.

——2005: The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *Journal of Agricultural Science 143:* 1–14.

WU, H.X.; MATHESON, A.C. 2005: Genotype by environment interaction in an Australia-wide radiata pine diallel mating experiment: implication for regionalized breeding. *Forest Science 51:* 29–40.

YANG, R.C. 2002: Likelihood-based analysis of genotype-environment interaction. *Crop Science 42:* 1434–1440.

YANG, R.C.; BAKER, R.J. 1991: Genotype-environment interactions in two wheat crosses. *Crop Science 31:* 83–87

Appendix A

CORRELATION STRUCTURES IN ASREML

$$\begin{bmatrix} 2 & & & \\ 1 & 2 & & \\ 1 & 1 & 2 & \\ 1 & 1 & 1 & 2 \end{bmatrix} \quad \begin{bmatrix} 2 & & & \\ 1 & 3 & & \\ 1 & 1 & 4 & \\ 1 & 1 & 1 & 5 \end{bmatrix} \quad \begin{bmatrix} 7 & & & \\ 1 & 7 & & \\ 2 & 3 & 7 & \\ 4 & 5 & 6 & 7 \end{bmatrix} \quad \begin{bmatrix} 7 & & & \\ 1 & 8 & & \\ 2 & 3 & 9 & \\ 4 & 5 & 6 & 10 \end{bmatrix}$$

CORUV CORUH CORGV CORGH

Individual numbers in these correlation structure matrices indicate individual variances or covariances. Correlation structure CORUV, with the same variance in each trial and the same covariance between pair of sites, is used to test for the presence of G×E interaction. CORUH structure, with different variances among sites but the same covariance between pairs of sites, is used to test lack of correlation. CORGV structure, with the same variance in each site but different covariances between sites, is used to test heterogeneity of family variance. CORGH structure has different variances among sites and different covariances between pairs of sites.