Multi-environment trial analysis on *P. radiata*: step by step to search "real" GxE interaction

Meimei Ding[1], Bruce Tier[1] and Gregory W. Dutkowski [2]

e-mail: mding@une.edu.au; Tel: +61 2 67733780

[1] Animal Genetics and Breeding Unit[*], University of New England, Armidale, NSW, 2351, Australia

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

[*] AGBU is a joint venture of NSW Department of Primary Industries and the University of New England

Abstract

Five combined trials data set of *P. radiata* were used to identify and determine the "real" GxE interaction. REML approach was applied to handle three main issues of the multi-environment trials (METs) analysis, 1). testing and modeling heterogeneity of genetic variance and error variance between trials; 2). testing the source of lack of correlation; and 3). modeling micro-environmental variation within each trial. Three series of mixed models were used to test the sources of GxE interaction, measure the relative size of GxE interactions and select the best model for prediction of genetic values, respectively. Modeling impacts were also investigated by evaluating the ratio (**K**) of family by trial variance to family variance and type B genetic correlation. Model comparison based on the criteria of log-likelihood. The significance was tested by likelihood ratio test (LRT). The LRTs show all the sources of GxE interactions are highly significant, indicating the GxE interactions occur in these five trials due to two parts, one associates with the heterogeneity of variances, one associates with lack of correlation. The full model which accommodates heterogeneity of error variances between trials with the spatial variation within trial, and allowing correlation forms among the GxE interactions is superior to other models for this METs. The modeling impacts are clearly reflected by the estimates of variance parameters and ratios (**K**), which decrease from 197.7% to 137.7% and 122.6%, by assuming the simplest homogeneous model to accommodating trial heterogeneity and further allowing the spatial variation within trial. The estimates of type B genetic correlations increase slightly after correcting the heterogeneity of variances.

Key word: "real" GxE interaction, REML, *P. radiata*

Introduction

Genotype by environment interaction (GxE interaction) refers to differential responses of different genotypes across a range of environments (Kang, 2004). However, only the repeatable GxE interactions which cause the ranking of genotypes to change across the macro-environments are essential and meaningful for breeding strategy (Baker, 1988).

These GxE interactions can be distinguished on whether the interactions are treated as either: (1) a source of error or bias in assessing a genotype (random, non-repeatable GxE interactions) or (2) as a component of variation which is, in part, heritable and exploitable through selection for broad and specific adaptation (repeatable GxE interactions, Hammer et al. 1996). Determining the relative proportion of repeatable and non-repeatable GxE interaction effects is an important issue in analysis and interpretation of METs. This partitioning was first shown by Robertson (1959). Muir et al. (1992) gave methods for partitioning the GxE interaction into the sources due to heterogeneous variances and lack of correlation. The lack of correlation is the component which directly impacts on the efficiency of selection whereas that due to heterogeneity of variances has little influence (Delacy et al., 1996). Yang and Baker (1991) used multivariate analysis of variance (MANOVA) and proposed two tests for the significance of the different sources of GxE 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 significance of different sources of GxE interaction.

REML (Patterson and Thompson, 1971) has been used for decades to estimate variance parameters based on mixed model theory (Henderson, 1984). Mixed model analysis for METs data contain frequentist approaches in which the variance parameters are estimated using REML and fixed and random effects are estimated using best linear unbiased estimates (BLUEs) and best linear unbiased predictors (BLUPs), respectively (Smith et al., 2005). The development of statistical packages such as ASREML (Gilmour, 1999) allows REML estimation of a range of mixed models and also enables them to fit more informative and complex models for accommodating different forms of GxE interactions.

Cullis et. al. (1998) allowed for heterogeneity between trials by fitting a separate variety by environment interaction (VxE) variance for each trial. Smith et. al. (2001) extended this approach for the analysis of METs data which included multiplicative models for the variety effects in each environment. The model provides an approach that accommodates heterogeneity of VxE variance, correlation among VxE interactions, and appropriate error variance structures for individual trial. In fact, the residual variation can be further partitioned into components due to micro-environment variation and genotype by micro-environment interaction (Nyquist, 1991). The variation within trial has been examined by some authors using spatial analysis in single sites (Casanoves et al., 2005; Cullis et al., 1998; Smith et al., 2001). Some evidence in forestry indicates that gradients and large patch sizes were found within trees (Costa e Silva et al., 2001; Fu, 1999), and also showed that using a combined spatial model enables to improved analysis of experiment data (Dutkowski et al., 2002, 2006; Hamann et al., 2002; Costa e Silva et al., 2001; Magnussen et al., 1990).

REML approaches based on mixed models allow more flexible variance structures, which are helpful for fitting GxE interactions. However, most applications in forestry focus mainly on quantifying the relative size of GxE interaction (Carson, 1991; Haapanen, 1996; Johnson and Burdon, 1990; Matheson and Raymond, 1984; Pederick, 1990), few formally identify and partition the sources of GxE interaction despite it has been recognized for a long time that they influence on the efficient decision of breeding programs and rapid genetic advance.

The goal of this study is to identify the repeatable GxE interaction, focuses on three aspects: 1). testing the sources of GxE interaction; 2). selecting the best models for METs; 3). investigating the modeling impact on variance parameters, genetic parameters, and the parameter used for measuring relative magnitude of GxE interaction.

Materials and data preparing

The genetic materials originated from an Australia-wide diallel mating experiments. The details were described by Wu et al. (2005). Five typical sites (PT5459, RAD211, VRC060, RS27A and RS27B) were chosen and combined for this study. They are distributed in four regions in Australia, containing 12460 genotypes from 169 to 216 full-sib families. Each

trial was a randomized incomplete-block design (RIB) with 3 replicates and four-tree row plots at a spacing of 3.0mx3.0m, excepting VRC060 (3.6mx2.3m). Trials PT5459, RAD211 and VRC060 have the same block numbers within each replicate. In trials RS27A and RS27B, the block numbers are continuous across trial. The growth trait of diameter at breast height (DBH) was measured at 10.5 years of age. The details of trials are presented in Table 1.

All check lots were eliminated from the data set before analysis. For spatial analysis, each individual tree was marked on a grid of R rows within C columns in each trial. For trials with an irregular shape, the data were expanded to construct a complete rectangular matrix by inserting missing value using BLOCKIT (Dutkowski, 2004). The proportion of missing values in these rectangular ranged from 0 to 61%.

Statistical models

The linear models were fitted at family level and individual tree level, separately.

In matrix notation, the general linear mixed model is

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

where **y** is the vector of observations for DBH, **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** has sub-vectors of block (within replicate for PT5459, RAD211 and VRC060 but not within replicate for RS27A and RS27B), family, family x trial and family x replicate (nested within trial) interaction. For the individual tree models, the random effects include block, plot, tree, family, tree x trial and family x trial interaction. This full model is only suitable for the case when tree and family effects are uncorrelated between sites. Nevertheless, tree and family terms were dropped from the model when the correlations between trials were given. Separate model terms in **u** were assumed to be uncorrelated.

It is assumed that the joint distribution of the random effects is 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 direct sum of the variance-covariance matrices of each of the random effects. For the family model, the **u** and **e** being,

$$\mathbf{Var}\begin{pmatrix}\mathbf{inb}\\ \mathbf{f}\\ \mathbf{t.f}\\ \mathbf{t.r.f}\\ \mathbf{e}\end{pmatrix}=\begin{bmatrix}\mathbf{G}_{\mathbf{inb}} & 0 & 0 & 0 & 0\\ 0 & \mathbf{G}_{\mathbf{f}} & 0 & 0 & 0\\ 0 & 0 & \mathbf{G}_{\mathbf{t.f}} & 0 & 0\\ 0 & 0 & 0 & \mathbf{G}_{\mathbf{t.r.f}} & 0\\ 0 & 0 & 0 & 0 & \mathbf{R}\end{bmatrix}$$

When independence of the sub-vectors **b, f, t.f and t.r.f** are assumed, then $\mathbf{Var(inb)} = \mathbf{I}_{\mathbf{inb}}\boldsymbol{\sigma}^2_{\mathbf{inb}}$, $\mathbf{Var(f)} = \mathbf{I}_{\mathbf{f}}\boldsymbol{\sigma}^2_{\mathbf{f}}$, $\mathbf{Var(t.f)} = \mathbf{I}_{\mathbf{t.f}}\boldsymbol{\sigma}^2_{\mathbf{t.f}}$ and $\mathbf{Var(t.r.f)} = \mathbf{I}_{\mathbf{t.r.f}}\boldsymbol{\sigma}^2_{\mathbf{t.r.f}}$, where $\mathbf{I}_{\mathbf{inb}}$, $\mathbf{I}_{\mathbf{f}}$, $\mathbf{I}_{\mathbf{t.f}}$ and $\mathbf{I}_{\mathbf{t.r.f}}$ are identity matrices of appropriate order, with constant variances $\boldsymbol{\sigma}^2_{\mathbf{inb}}$, $\boldsymbol{\sigma}^2_{\mathbf{f}}$, $\boldsymbol{\sigma}^2_{\mathbf{t.f}}$ and $\boldsymbol{\sigma}^2_{\mathbf{t.r.f}}$ for block (**inb**), family (**f**), family x trial (**t.f**) and family x replicate within trial (**t.r.f**). $\otimes$ is the Kronecker product (Searle et al., 1992).

For the individual tree model, additional random effects of plot (**p**), tree (**g**) and tree x trial (**t.g**) were included, the additional variance structures to be defined, $\mathbf{Var(p)} = \mathbf{I}_{\mathbf{p}}\boldsymbol{\sigma}^2_{\mathbf{p}}$, $\mathbf{Var(g)} = \mathbf{A}\boldsymbol{\sigma}^2_{\mathbf{a}}$, $\mathbf{Var(t.g)} = \mathbf{I}_{\mathbf{t.g}}\boldsymbol{\sigma}^2_{\mathbf{t.g}} = \mathbf{I}_{\mathbf{t}}\mathbf{A}\boldsymbol{\sigma}^2_{\mathbf{a}}$, where $\boldsymbol{\sigma}^2_{\mathbf{a}}$ is the additive genetic variance, **A** is the numerator relationship matrix, $\mathbf{I}_{\mathbf{p}}$ and $\mathbf{I}_{\mathbf{t}}$ 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 x trial (or tree x trial) interaction can be expressed as, $\mathbf{Var(F)} = (\boldsymbol{\sigma}^2_{\mathbf{f}}\mathbf{J}_{\mathbf{t}} + \boldsymbol{\sigma}^2_{\mathbf{t.f}}\mathbf{I}_{\mathbf{t}}) \otimes \mathbf{I}_{\mathbf{f}}$, where $\mathbf{J}_{\mathbf{t}}$ is the **t x t** unit matrix (i.e. with all elements equal to one). This variance structures is known as a compound symmetry structure. It implies that the all 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). Thus, the magnitude of GxE 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 full model can be expressed as,

$$\mathbf{Var}\begin{pmatrix}\mathbf{f}_1\\ \mathbf{f}_2\\ \vdots\\ \mathbf{f}_i\end{pmatrix}=\begin{bmatrix}\sigma^2_{f1} & \sigma_{f12} & \cdots & \sigma_{f1j}\\ \sigma_{f21} & \sigma^2_{f2} & \cdots & \sigma_{f2j}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{fi1} & \sigma_{fi2} & \cdots & \sigma^2_{fi}\end{bmatrix}\otimes\mathbf{I}_f$$

where the diagonal element $\boldsymbol{\sigma}^2_{fi}$ is the family variance in environment **i**, off-diagonal element $\boldsymbol{\sigma}_{fij}$ is the family covariance between environments **i** and **j**, $\mathbf{I}_f$ is **f x f** identity matrix. The covariance can be expressed as, $\boldsymbol{\sigma}_{fij}=\boldsymbol{\rho}_{fij}\boldsymbol{\sigma}_{fi}\boldsymbol{\sigma}_{fj}$, where $\boldsymbol{\rho}_{fij}$ is the correlation between environment **i** and **j**.

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

$$\mathbf{R}=\mathbf{Var}\begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\\ \vdots\\ \mathbf{e}_i\end{bmatrix}=\begin{bmatrix}\mathbf{R}_1 & & & \\ & \mathbf{R}_2 & & \\ & & \ddots & \\ & & & \mathbf{R}_i\end{bmatrix}$$

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, $\mathbf{R}_i$ (**i=1, 2,…**) have a different structure based on a decomposition of **e** into spatially dependent ($\boldsymbol{\eta}$) and independent (ξ) residuals (Costa e Silva et al., 2001; Dutkowski et al., 2002). The $\mathbf{R}_i$ matrices are

$$\mathbf{R}_i=\boldsymbol{\sigma}^2_{\xi i}[\mathbf{AR1}(\boldsymbol{\rho}_{col})\otimes\mathbf{AR1}(\boldsymbol{\rho}_{row})]+\boldsymbol{\sigma}^2_{\eta i}\mathbf{I}$$

where $\boldsymbol{\sigma}^2_{\xi i}$ is the spatial residual variance in trial i, $\boldsymbol{\sigma}^2_{\eta i}$ is the independent residual variance of the "white noise" process in trial **i**, **I** is an identity matrix, and **AR1(** $\boldsymbol{\rho}$ **)** represents a first-order autoregressive correlation matrix which, for ordered coordinates of size n, has the form:

$$\mathbf{AR1}(\boldsymbol{\rho})=\begin{bmatrix}1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{n-1}\\ \rho & 1 & \rho & \rho^2 & \cdots & \vdots\\ \rho^2 & \rho & 1 & \rho^3 & \cdots & \vdots\\ \rho^3 & \rho^2 & \rho^3 & 1 & \cdots & \vdots\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ \rho^{n-1} & \cdots & \cdots & \cdots & \cdots & 1\end{bmatrix}$$

where $\boldsymbol{\rho}$ is the autocorrelation parameter. The independent errors in trials PT5459, RAD211 and VRC060 were significant in the previous single site spatial analysis, which was fitted in multi-site spatial model.

Estimates of the fixed and random effects in [1], the solutions to the mixed model equations are given by Henderson (1984), which is

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \qquad [2]$$

This lead to BLUEs of the fixed effects and BLUPs of random effects. We compared the modeling impacts on BLUPs of the genetic effects. The parameters in **G** and **R** are replaced by estimates from data, using ASREML (Gilmour et al., 1999).

Statistical analyses

Three series of models were used. Series one included five alternative family models for testing the sources of GxE interaction. Series two included three family models with both family and family x trial effects to measure the relative size of GxE interaction. Series three fitted five individual tree models to select the best model for obtaining BLUPs of genetic values (parent and offspring tree). Table 2 lists the series one models with specific constraints by giving different forms of **R** and **G** structures. The different forms of correlation structures are listed in appendix A.

For series one, the statistical procedures started from full model to 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 GxE 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, allowed the residual and family variances to be heterogeneous, the family covariances were the same and correlation was constrained to be 1. Reduced model 3 (S1M3) with the hypothesis of the family variances were homogeneous among trials, allowed the heterogeneity of error variance between trials and family covariances were different, but family variances were the same among five trials. Reduced model 4 (S1M4)

with the hypothesis of error variances were homogeneous, allowed the family variances and covariances to be different among trials.

For series two, the procedures started from the simple model, assuming the residual variance are homogeneous (S2M1). Model 2 (S2M2) accommodated heterogeneity of error variance between trials. Model 3 (S2M3) included the spatial variation within each trial nested in model 2. The random effects included family and family x trial interactions. The variance structure is known as a uniform (or compound symmetric) structure. The relative size of GxE interactions can be estimated by ratio (**K**) which is the interaction variance by family variance.

For series three, the models were at the individual tree level. The first three procedures (S3M1, S3M2, S3M3) have 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 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,

LRT=-2*(logL. of full model - logL. of reduced model) [3]

For series two and three,

LRT=-2*(log L. of model (i) - logL. of model (i+1)) [4]

where model **i** and model **i+1** are with and without the tested component or structure, respectively. Under the null hypothesis, LRT 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 and Stuart, 1979). When **LRT** $> \chi^2_q$, it indicates the estimated variance component is significant.

Results

The results of series one with log-likelihood, LRTs and estimates of variance parameters are given in Table 3. The LRTs show all the tested sources are highly significant. The first test indicates the GxE interactions are present across five sites. Moreover, the significance of heterogeneity of genetic and error variances, and lack of correlation indicate the GxE

interactions occur in this METs attribute to the heterogeneity of genetic and error variances, and lack of correlation. In most sites, the family variances in full model (S1M0) are higher than the estimates in other reduced models, whereas the estimates of error variances in full model are lower than in the reduced models.

Table 4 shows the results estimated from series two. The variances of family ($\boldsymbol{\sigma}_{\mathbf{f}}^{\mathbf{2}}$) are stable across models, and the trend of standard error decreases by fitting the homogeneity (S2M1), heterogeneity (S2M2) and spatial model (S2M3). The family variances are 27.2, 27.1 and 27.2 with the standard error 6.9, 6.1 and 6.0, respectively. This indicates that relaxing the heterogeneity of error variance does not influence the estimates of family variances. However, the variances of family x trial ($\boldsymbol{\sigma}_{\mathbf{fe}}^{\mathbf{2}}$) decrease, from 53.7 to 37.4, 33.4, with the decreasing standard errors, resulting in the ratio decreasing 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 accommodates heterogeneity of residual variance much better than S3M1, with a large increase in log-likelihood (LRT=977.2 on 4 df, $p<0.001$). Maintaining heterogeneity, from S3M2 to S3M5 any further fitting get significant model improvement. Nested the heterogeneity model, accommodating spatial variation within each trial, the model improves significantly (LRT=21.4, on 3 df, $p<0.01$). Containing S3M2, giving the correlation structure in the model, there are also significant model improvement (LRT=29.1, on 15 df, $p<0.01$). The model S3M5 which accommodates heterogeneity of error variance between trials, spatial variation within trial and correlation structure for individual tree and family, is superior to all the above models.

The estimates of genetic correlations between 10 pairs of sites are presented in Table 6. The averages of genetic correlations change slightly by different estimates, which the (co)variances are constrained in different ways. It is lowest (0.34) when assuming all the variances are homogeneous (all the heterogeneities confounded). Correcting the heterogeneity of error variance between trials, the average type B genetic correlation increases to 0.38. They are 0.35 and 0.36 by correction for both the heterogeneity of error and genetic variances, and spatial variation within trials, respectively. The highest genetic correlation occurs between site RAD211 and VRC060 across all approaches. The genetic

correlation between RS27A and RS27B increases after correcting for heterogeneity of error variance, from 0.51 to 0.75. For other pairs, the trend is similar by different estimates. The pairs of VRC060 and RAD211, VRC060 and PT5459 are a little variable. The analysis fails to correct the heterogeneity of family variances because of the singularity problem after the first iteration.

Discussion

It is of great value for breeding program to partition the GxE interaction component, particular in the situation which the GxE interactions attribute 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 component of selection for specific selection (Delacy, 1996). Our study detected that the GxE interactions associated with the sources, heterogeneity of genotype variance and lack of correlation. Furthermore, the heterogeneity of error variance between trials and spatial variation within each trial were identified and investigated, they might be the sources which are confounded with the GxE interactions as well.

The evidence was reflected by the ratios (**K**) of family x trial interaction variance to family variance. The estimates of ratio decreases approximate 30% by correcting for the heterogeneity of residual variance between sites. Further partitioning and removing the spatial variation within each site, the ratio goes down approximate 11%.

Eisen et al. (1983) showed that heterogeneous variances can cause the estimates of GxE interaction components to be biased upward, and genetic correlation between trials to be biased downward. In our study, the genetic correlations between trials are also detected to be biased by the heterogeneity of error variance but change slightly. It is more likely since that the variation between and within sites was excluded from the total variance so as to increase the similarity of environments, leading to the increased genetic correlations. The correction for the heterogeneity of family variance occurred singularity problem after the first iteration, failing to investigate the influence on the estimates. However, if removing both of the heterogeneity of family and error variances, type B genetic correlation decrease a bit rather than increase, not matching the expectation.

The power of REML approach allows different forms of **R**, **G** as well as correlation structure. As a result, a range of more complex and informative models can be used to better explore the GxE interactions. In our study, the flexible correlation structures are used with constrained and/or unconstrained (co)variances so that the sources of heterogeneity and correlation can be partitioned and tested. Moreover, the genetic correlations between sites can also be estimated by correcting for heterogeneity of error and genetic variances. These analyses should be more realistic since the possibility of variance heterogeneity among GxE interaction effects have been recognized by many authors (Patterson et al., 1992; Frensham et al., 1997; Cullis et al., 1998). However, the risks of fitting complex models cause some computational problems. The individual tree model did not converge initially when the complex variance structures were used. The convergence troubles could be more frequent particularly for the spatial model with independent error variance. Fortunately, Dutkowski et al. (2006) suggested some effective ways to achieve convergence. Singularity after the first iteration happened when poor models were fitted with unusually constrained error variances, leading to poor estimates of variance parameters. Therefore, care must be taken of some constraints, they may only be used for some special tests not advocated for estimation of parameters.

Conclusion

REML approach is powerful and flexible for partitioning and testing the sources of GxE interaction. The sources of heterogeneity of genetic variance and error variance, and lack of correlation are all significant, indicating the GxE interactions occur in the five trials associate with two parts, repeatable (lack of correlation) and non-repeatable (heterogeneity) GxE interaction.

The model accommodating the trial heterogeneity with the form of correlation structure is good model for METs data analysis. Nested this model, allowing the spatial variation within each site is superior to the other models.

The heterogeneity of variances is biased upward or downward of estimating GxE interaction. The ratios (**K**) decrease by correcting for heterogeneity of variances. The estimates of genetic correlation between sites change slightly by correcting the heterogeneity sources.

Acknowledgements: M. Ding gratefully acknowledges STBA and FWPRDC for the funding support for her study. The authors thank Mike Powell for providing the data kindly.

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 x environment interaction and optimal number of progeny test sites for improving Pinus radiata in New Zealand. N. Z. J. For. Sci. 21: 32-49.

Casanoves, F., R. Macchiavelli, and M. Balzarini. 2005. Error variation in multienvironment Peanut trials: within-trial spatial correlation and between–trial heterogeneity. Crop Science. 45: 1927-1933.

Costa. e Silva, G. W. Dutkowski, and A. R., Glimour. 2001. Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. Can. J. For. Res 31:1887-1893.

Cullis, B. R., B. J. Gogel, A. P. Verbyla, and R. Thompson. 1998. Spatial analysis of multi-environment early generation trials. Biometrics 54:1-18.

Delacy, I. H., K. E., Basford, M. Cooper, J. K. Bull, and C. G. Mclaren. 1996a. Analysis of multi-environment trials- an historical perspective, *In* M. Cooper and G. L. Hammer, ed. Plant Adaptation and Crop Improvement. CAB INTERNATIONAL, Cambridge.

Delacy, I. H., M. Cooper, and P. K. Basford. 1996b. Relationships among analytic methods used to study genotype-by-environment interactions and evaluation of their impact on response to selection. In: Kang, M. S. and H. G. Gauch (eds) Genotype-by-Environment Interaction: New Perspectives. CRC Press, Boca Raton, Florida, pp. 51-84

Dutkowski, G. W., J. Costa e Silva, A. R, Gilmour, and G. A. Lopez. 2002. Spatial analysis methods for forest genetic trials. Can. J. For. Res 32:2201-2214.

Dutkowski, G. W., J. Costa e Silva, A. R. Gilmour, H. Wellendorf, and A. Aguiar, 2006. Spatial analysis enhances modeling of a wide variety of traits in forest genetic trials. Can. J. For. Res. 36: 1851-1870.

Eisen, E. J., and A. M. Saxton. 1983. Genotype by environment interactions and genetic correlation involving two environmental factors. Theor. Appl. Genet. 67: 75-86.

Frensham, A.B., B. R. Cullis, and A. P. Verbyla, 1997b. Genotype by environment variance heterogeneity in a two- stage analysis

Frensham, A. B., B. R. Cullis, and S. D. Pelham, 1997a. The analysis of 10 years of South Australian oat evaluation data. Euphytica 99:43-56.

Fu Y. B., A. D.Yanchuk, and G. Namkoong. 1999. Spatial patterns of tree height variations in a series of Douglas-fir progeny trials: implications for genetic testing. Can. J. For. Res. 29:714-723.

Gilmour, A. R., B. R. Cullis, S. J. Wielham, and R.Thomson, 1999. ASREML reference manual NSW Agric. Biom, Bull. 3., New South Wales, Austrialia.

Johnson, G. R., and R. D. Burdon. 1990. Family-site interactions in Pinus radiata: Implications for progeny testing strategy and regionalized breeding in New Zealand. Silvae Genet. 39: 55-62.

Magnussen, S. 1990. Application and comparison of spatial models in analyzing tree-genetics ield trials. Can. J. For. Res 20.

Matheson, A. C., and C. A. Raymond. 1984. The impact of genotype x environment interactions on Australia Pinus radiata breeding programs. Aust. For. Res. 14: 11-25.

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., M. P. Koshy, and G. Namkoong. 2002. Improving precision of breeding values by removing spatially auto-correlated variation in forestry field experiments. Silvae Genet.

Hammer G. L., M. Cooper. 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

Kang, M. S. 2004. Breeding: genotype-by-environment interaction. Encyclopedia of Plant and Crop Science. (edt.) Goodman, R. M. New York Basel.218-221.

Kanzler, A. 2002. Genotype x environment interaction in Pinus patula and its implications in South Africa, North Carolina State University, USA.

Kendall, M. G., and A. Stuart. 1979. The advantage theory of statistics. Vol. 2. Inference and relationship. Griffin and Co., London.

Muir, W., W. E. Nyquist, and S. Xu. 1992 Alternative partitioning of the genotype-by-environment interaction. Theor. Appl. Genet. 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. D. and R. Thompson. 1971. Recovery of interblock information when block sizes are unequal. Biometrika. 58: 545-554.

Patterson, H., and F. Nabugoomu. 1992. REML and the analysis of series of crop variety trials. In Proceedings from the 16th Interactional Biometric Conference, 77-93. Hamilton, New Zealand.

Pederick, L. A. 1990. Family x site interactions Pinus radiata in Victoria, Australia, and implications for breeding strategy. Silvae Genetic. 39: 134-140.

Robertson, A. 1959. The sampling variance of the genetic correlation coefficient. Biometrics. 15: 469-485.

Searle, S. R., G. Gasella, and C. E. McCulloch. 1992. Variance components. John Wiley & Sons, New York.

Smith, A., B. Cullis., and R. Thompson. 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57:1138-1147.

Smith, A. B., B. R. Cullis, and R. Thompson. 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., and A. C. Matheson. 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., and R. J. Baker. 1991. Genotype-environment interactions in two wheat crosses. Crop Science. 31: 83-87

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

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, which is used to test significance of GxE interaction. CORUH structure with different variances among sites but the same covariance between pair of sites, which is used to test lack of correlation. CORGV structure with the same variance in each site, but different covariances between sites, which is used to test heterogeneity of family variance. CORGH structure is with different variances among sites and different covariances between pair of sites.

Table 1. Design information of five trials and mean DBH with standard deviation

Site	PT5459	RAD211	VRC060	RS27A	RS27B
Observation	2536	2624	2592	3260	1652
Missing value	236	554	565	1643	541
Replicate	3	3	3	3	3
Block	18	18	18	108	39
Plot	648	656	648	648	546
Plot size	4	4	4	4	4
Row	96	97	36	83	49
Column	60	68	72	55	54
Spatial rate (%)	56	61	0	29	38
Family number	216	216	216	216	169
Spacing (m)	3 × 3	3 × 3	3.6 × 2.3	3 × 3	3 × 3
Measured Age	10.5	10.5	10.5	10.5	10.5
Dbh (mm)	174 ± 22	158 ± 34	201 ± 44	233 ± 41	242 ± 35

Table 2. Full model and four reduced model for testing the sources of GxE interaction by ASREML

Model	Null hypothesis	Constraint of **R**[a] and **C**[b] structure	Estimated parameters
S1M0	Full model, allowed all the sources	**R** structure; CORGH correlation structure	$\boldsymbol{\sigma}^2_{f1} \ldots \boldsymbol{\sigma}^2_{f5}$; $\boldsymbol{\rho}_{ij}$; $\boldsymbol{\sigma}^2_{e1} \ldots \boldsymbol{\sigma}^2_{e5}$
S1M1	No GxE interaction	**R** structure; $\boldsymbol{\rho} = \mathbf{1}$; CORUV structure; $\boldsymbol{\sigma}^2_f$ is constrained the same across the site	$\boldsymbol{\sigma}^2_f$; $\boldsymbol{\sigma}^2_{e1} \ldots \boldsymbol{\sigma}^2_{e5}$
S1M2	Perfect correlation	**R** structure; $\boldsymbol{\rho} = \mathbf{1}$; CORUH structure;	$\boldsymbol{\sigma}^2_{f1} \ldots \boldsymbol{\sigma}^2_{f5}$; $\boldsymbol{\sigma}^2_{e1} \ldots \boldsymbol{\sigma}^2_{e5}$
S1M3	Homogeneity of family variances across sites	**R** structure; CORGV structure; $\boldsymbol{\sigma}^2_f$ is constrained the same across the site	$\boldsymbol{\sigma}^2_f$; $\boldsymbol{\rho}_{ij}$; $\boldsymbol{\sigma}^2_{e1} \ldots \boldsymbol{\sigma}^2_{e5}$
S1M4	Homogeneity of error variances across sites	$\boldsymbol{\sigma}^2_e$ is constrained the same across the sites; CORGH structure	$\boldsymbol{\sigma}^2_{f1} \ldots \boldsymbol{\sigma}^2_{f5}$; $\boldsymbol{\rho}_{ij}$; $\boldsymbol{\sigma}^2_e$

[a]Residual variance structure; **R** structure, heterogeneity of residual variances between sites.

[b]Correlation structure, the details are given in appendix A.

Table 3. LRTs for the heterogeneity and lack of correlation and the estimates of variance parameters (model series one)

Model	Sources of GxE interaction	**logL.**	**LRT**[a]	σ^2_{f1}	σ^2_{f2}	σ^2_{f3}	σ^2_{f4}	σ^2_{f5}	mean ρ [b]	σ^2_{e1}	σ^2_{e2}	σ^2_{e3}	σ^2_{e4}	σ^2_{e5}
S1M0	Total GxE 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 GxE interaction	-6100.05	133.7 ***	σ^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	Family variance heterogeneity	-6077.25	21.24 **	σ^2_f=66.9					0.38	416.6	1026.8	1143.6	1770.3	1502.7
S1M4	Error variance heterogeneity	-6519.60	905.94 ***	3.3	93.6	54.0	221.0	228.4	[c]/	σ^2_e=1047.2				

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

[a] LRTs are the comparisons of full model and reduced models

[b]Average of genetic correlation between sites.

[c]/, Singularity problems after the first iteration of analysis.

Table 4. The estimates of the variance component with standard error and ratio

Model	**LRT**[a]	$\boldsymbol{\sigma^2_{fe}}$	$\boldsymbol{\sigma^2_{f}}$	**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 ***.

[a] LRT compared with previous model, e.g., S2M2 compared with S2M1, S2M3 compared with S2M2.

Table 5. Comparison of individual tree models with the variance parameters

Model	**LogL.**	**LRT**[a]	σ^2_{g1} σ^2_{g2} σ^2_{g3} σ^2_{g4} σ^2_{g5}	σ^2_{f1} σ^2_{f2} σ^2_{f3} σ^2_{f4} σ^2_{f5}	σ^2_{e1} σ^2_{e2} σ^2_{e3} σ^2_{e4} σ^2_{e5}
S3M1	-6562.90	0	σ^2_g =23.9	σ^2_f =22.5	σ^2_e =1027.2
S3M2	-6074.30	977.2***	σ^2_g =18.6	σ^2_f =22.2	390.9, 1007.4, 1122.3, 1736.3, 1478.5
S3M3	-6059.12	30.4 **	σ^2_g =20.5	σ^2_f =22.4	384.7, 965.5[d], 1120.5, 1736.9, 1478.3
S3M4	-6055.34	37.9[b]**	33.2, 43.4, /[c], 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	41.2[b] **	28.3, 68.9, 1.7, 132.6, 258.6	σ^2_f =9.6	399.5, 949.9[d], 1154.1, 1681.6, 1339.2

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

[a]Nomally, LRTs compared with previous model

[b]S3M4 compared with S3M2, S3M5 compare with S3M3.

[c]The genetic variance in site 3 was excluded.

[d]Independent error of RAD211, spatial error is 84.7 and 77.4 in S3M3 and S3M5, respectively.

Table 6. The estimates of genetic correlation of family means between 10 pairwise of sites based on the correcting for the sources of heterogeneity

Pairwise	PT5459 RAD211	PT5459 VRC060	T5459 RS27A	PT5459 RS27B	RAD211 VRC060	RAD211 RS27A	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 variance

Model 2 Correcting heterogeneity of error variance between trials

Model 3 Removing heterogeneity of family and error variances between trials

Model 4 Removing heterogeneity of family and error variances between trials, and spatial variation within each trial