

USE OF ANALYTIC FACTOR STRUCTURE TO INCREASE HERITABILITY OF CLONAL PROGENY TESTS OF *Pinus taeda* L.

Jaime Zapata-Valenzuela¹

Advanced variance-covariance structures are commonly used in genetic evaluation of crops to account for micro-site variability and achieve higher accuracy of predictions to increase selection efficiency. Various genetic variance-covariance structures were explored to predict best linear unbiased genetic merits of 453 loblolly pine (*Pinus taeda* L.) cloned progeny tested at 16 different locations in the southern U.S. Statistical models were compared using model fit statistics, variance components and genetic parameters. Among the models explored, spatial autoregressive error correlation with independent residual term for the *R* side with a factor analytic structure for the *G* side of the mixed model was superior. The model produced one of the smallest fit statistics (LogL equal to -2694), a small error variance (12.72), and the highest broad-sense heritability (0.45), compared with the default homogeneous error and genetic variance-covariance structure (statistical significance at $P < 0.05$). We concluded that the combination of specific structure for error and genetic design was effective to remove spatial-related variance, and to increase the accuracy of predictions of clonal genetic values, which could be used as analytical tool for increasing the selection efficiencies in forest genetic trials.

Key words: Linear mixed model, quantitative forest genetics, genetic variance.

The incorporation of clonal forestry as an operational breeding and development option for forestry programs in the southern United States could potentially result in large volume gains for loblolly pine (*Pinus taeda* L.), with estimates of 60-70% gain above open-pollinated families or mass-controlled pollination systems (Whetten and Kellison, 2010). The accurate prediction of the genetic value of clones becomes a critical step in determining the ranking of the phenotypes within a selection program. To maximize the cost-efficiency of forest field testing, the data should be analyzed using the most appropriate statistical methods.

Many model approaches assume simple, often inappropriate, within-test error structures, such as randomized incomplete block design (RIBD), and a common error variance for all tests. Gezan *et al.* (2006) showed that appropriate choice of experimental designs can yield considerable improvements in clonal testing. Using simulated data based on a single site test of 256 clones arranged in single-tree plots with no missing observations, it was found that row, column and RIBD with a block size of eight trees had the highest individual broad-sense heritability estimates (H^2). In another study using linear mixed models,

Isik *et al.* (2005) analyzed clonal field tests that were established at two sites using rooted cuttings from 450 clones of eight full-sib families of loblolly pine. At age four, there were significant differences among full-sib families and among clones within families for all traits analyzed. Clonal selection yielded considerable genetic gains for volume over the means of all the clones at both South Carolina and Florida sites. Also, Isik *et al.* (2008) used 12-yr-old data from 43 clones from nine full-sib families of loblolly pine to measure the variation of microfibril angle using a quadratic mixed model fitted over wood rings as predictor variable. This statistical approach had the advantage of choosing a better residual variance-covariance structure that correctly fit the data.

Mixed model analyses of unbalanced data can include modeling of genetic variance-covariance structures. Examples of the alternatives studied in previous studies include regression analysis, additive main effects and multiplicative interaction (AMMI) model, and factor analytic models (Gauch, 1992; Meyer, 2009; Raman *et al.*, 2011). The AMMI model was originally proposed as a fixed effects model. Assuming genotypes as random, the genetic by environment interaction ($G \times E$) can be analyzed in a mixed-model framework with a factor analytic covariance structure to model the multiplicative terms. The factor analytic form framework has been increasingly used in plant breeding (de los Campos and Gianola, 2007; Meyer, 2009; Raman *et al.*, 2011), because there is interest in structures that utilize the principal components

¹North Carolina State University, Department of Forestry and Environmental Resources, Raleigh, North Carolina 27695-8008, USA. (jzapata@arauco.cl)

Received: 7 November 2011.

Accepted: 9 July 2012.

of a covariance matrix. It has become accepted that such structures can be fit directly within the mixed models approaches commonly applied for both estimation of variance-covariance components and the prediction of genetic merit for selection purposes.

Similarly, typical field designs used for genetic testing of forest trees have been randomized complete/incomplete block design, or occasionally split-plot designs, but spatial analysis has not been widely utilized in forest genetic tests (Costa e Silva *et al.*, 2001). The main goal is to model the variation related to the physical location of the experimental units, given by the arrangements of columns and rows within the field test. The spatial component is fit as the form of separable two-dimensional autoregressive residuals. It is expected that spatial models would give significant improvement over standard design models allowing further genetic gain. Dutkowski *et al.* (2006) analyzed height and diameter from 55 tests to gain a better understanding of the utility of spatial analysis in a more wide ranging scenario. The spatial model yielded a more realistic and satisfying description of the site variability and gave a better understanding of the nature of that variation.

The objective of this study was to compare different matrix structures for the genetic random effects to fit the best statistical model in estimating the genetic merit for 453 clones. We analyzed 5-yr-old data of volume in clonal field tests of loblolly pine growing in multiple sites across southern United States.

MATERIALS AND METHODS

Genetic material and pedigree information

For the study, 28 parents were used to produce 23 full-sib families and two open-pollinated families. Though the majority of crosses were single-pair mating, several families were related by a common male or female parent. From each cross of female and male parents, clones were produced via somatic embryogenesis (Bettinger *et al.*, 2009). A deep pedigree file was designed, containing a section of the identity of each parent used to produce the full-sib families or open-pollinated families, and a second section containing the identity of each clone generated from their corresponding parent's combinations.

Study sites, experimental design, and data collection

Three series of tests were planted between 2000 and 2002, nine sites were planted in Georgia (GA), five sites were planted in Mississippi (MS), and two sites were planted in South Carolina (SC). A total of 16 sites were planted with 526 clones. The experimental design used was alpha-lattice incomplete block design with single tree plots. The average of clones/site was 163. Out of the 526 clones, we removed 73 clones that had fewer than 10 living ramets/clone, and/or clones were planted on fewer than three

sites. The final data set included 453 clones, and a total of 14 704 loblolly pine trees were measured. Calipers were used to obtain stem diameter at 1.4 m above ground line. Total height was measured with a precision of 0.03 m using an hypsometer (Vertex IV, Haglof Inc., Langele, Sweden). Volume was calculated using the metric form of the formula given in Sherrill *et al.* (2008): volume = $[0.95456089 + (0.28241587 \times \text{diameter}^2 \times \text{height}/10)]$. Using a SAS script, we standardized the data using a mean of 100 for height and volume variances calculated from the data (SAS Institute, 2010). As required for spatial analysis, the data were organized into a grid to be read correctly in the model. Table 1 contains the main descriptive information for the 16 sites assigned to this study. The different sites were presented with their information of year of planting, location, and number of clones.

Statistical analysis

Variance-covariance structures explored for this study were summarized in Table 2. The linear mixed model used for all the different structures comparison was described under the matrix form:

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

where y is the vector of observations representing the trait of interest (volume); the vector of parameters b and u which are the fixed and random effects to be estimated, respectively. Within the random effects, the genetic variance is due to genetic differences among clones ($\sigma_c^2 = \frac{1}{2} \sigma_a^2 + \frac{3}{4} \sigma_d^2 + \sigma_i^2$), and genetic difference among families ($\sigma_f^2 = \frac{1}{2} \sigma_a^2 + \frac{1}{4} \sigma_d^2$) where σ_a^2 , σ_d^2 , and σ_i^2 are additive, dominance, and epistatic genetic variances, respectively (Isik *et al.*, 2008). We included the clone effect with expectations $\sim N(0, A\sigma_c^2)$, where A is the numerator relationship matrix and $G = A\sigma_c^2$ is a nonsingular matrix; the variance due to the family had expectations $\sim N(0, I\sigma_f^2)$, where I is an identity matrix. Also, we included

Table 1. Descriptive statistics for the 16 sites analyzed. The mean and coefficient of variation (CV) of volume were presented for the raw data, to have a dimension of the variable dispersion among sites. The CV for volume ranged from 36.1% to 65.3%.

Site number	Year of planting	State ¹	Number of clones	Mean (m ³) ²	CV ²
1	2000	GA	135	0.034	36.1
2	2000	GA	135	0.024	47.3
3	2000	MS	133	0.023	39.8
4	2000	MS	128	0.025	39.5
5	2000	GA	86	0.023	42.4
6	2001	GA	182	0.012	65.3
7	2001	GA	192	0.017	48.7
8	2001	GA	187	0.026	40.2
9	2001	GA	182	0.025	45.5
10	2001	SC	193	0.023	49.8
11	2001	MS	191	0.026	47.3
12	2002	GA	174	0.014	56.3
13	2002	GA	175	0.006	58.2
14	2002	SC	176	0.018	43.1
15	2002	MS	176	0.020	43.2
16	2002	MS	175	0.016	47.2

¹GA: Georgia, MS: Mississippi, SC: South Carolina.

²Values were reported for non-standardized data. Number of blocks per site was 8 except for sites 6 and 9 where there were four blocks at each site.

Table 2. Variance-covariance structures used for clonal data analyses and estimated model fit statistics (LogL) for each structure. The structures with spatial adjustment had the smallest LogL values. Also, the structure AR1+ σ^2_{η} +FA1 was found statistically significant ($P < 0.05$).

Structure name	R error structure	G variance structure	G correlation structure	LogL
IID+FA1	Homogeneous	Factor analytic, heterogeneous	Heterogeneous	-3010
AR1+ σ^2_{η} +CS	AR1 + independent error, heterogeneous	Compound symmetry	Homogeneous	-2751
AR1+ σ^2_{η} +CH	AR1 + independent error, heterogeneous	Heterogeneous	Homogeneous	-2711
AR1+ σ^2_{η} +FA1	AR1 + independent error, heterogeneous	Factor analytic, heterogeneous	Heterogeneous	-2694

the block term as random with expectations $\sim N(0, I\sigma^2_b)$. The denoted X and Z are the design or incidence matrices. The term e represented the error component or vector of residuals with expectations $N(0, I\sigma^2_e)$, where I is the identity matrix, σ^2_e is the error variance, and $R = I\sigma^2_e$ is a nonsingular matrix (Lynch and Walsh, 1998; Balding *et al.*, 2007). The variance-covariance matrix of observations is given by the expression $\text{Var}(y) = V = ZGZ' + R$. Fixed and random effect solutions were obtained by solving the mixed-model equations (Henderson, 1984). The best linear unbiased predictions (BLUP) of each clone were obtained by solving mixed model equations as $\hat{u} = GZ'V^{-1}(y - Xb)$, where $b = (X'V^{-1}X)^{-1}X'V^{-1}y$ is the generalized least-square solution for b (Lynch and Walsh, 1998).

Definition of matrices structure and model development

In general, R and G structures are typically formed as a direct product (\otimes) of particular variance models. When the model is specified, the order of terms in a direct product must agree with the order of effects in the corresponding model term. Also, variance models may be correlation matrices or variance matrices with equal or unequal variances on the diagonal (Gilmour *et al.*, 2009). In our study, the default and simplest homogeneous R structure was that residual effects were independently and identically distributed (IID). The IID form was designed as matrix form where residuals were homogeneous across 16 sites: $R = I_1\sigma^2_{e1} \otimes I_2\sigma^2_{e2} \otimes \dots \otimes I_{16}\sigma^2_{e16}$, where I is an identity matrix and $\sigma^2_{e1} = \sigma^2_{e2} = \sigma^2_{e3} = \dots = \sigma^2_{e16}$.

Different forms of R can be used in the model definition and we used the identity matrix for the default IID form, and compared it with an autoregressive (AR1) form correlation structure. The AR1 variance-covariance matrix of the spatial design was defined as: $\text{Var}(e) = R = \sigma^2 \Sigma(\alpha) = \sigma^2 \Sigma_r(\alpha_r) \otimes \Sigma_c(\alpha_c) + I\sigma^2_{\eta}$, where $\text{Var}(e)$ is the error variance-covariance matrix, with σ^2 being a spatially dependent variance and σ^2_{η} is the independent error variance. The Σ denotes the spatial correlation matrix as a function of the number of α parameters and has associated variance σ^2 . The autoregressive form is feasible when the data from observations that are close together are more similar than those that are further apart (Gilmour *et al.*, 2009). As the sites were arranged in rectangular arrays, we used a two-dimensional coordinate system to define the location of each observation in the row (r) and column

(c), as usually given in other studies (Dutkowski *et al.*, 2002; Cumbie, 2010).

Similarly, the simplest structure for G matrix in the linear model defined in Equation [1], was represented as a general matrix form as:

$$G = \begin{bmatrix} \sigma^2_{c1} & \sigma_{1,2} & \dots & \sigma_{1,16} \\ \sigma_{2,1} & \sigma^2_{c2} & \dots & \sigma_{2,16} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{16,1} & \sigma_{16,2} & \dots & \sigma^2_{c16} \end{bmatrix} \otimes I_{nc} = I_{16} \otimes A\sigma^2_c \quad [2]$$

The diagonal of this matrix was represented by the number of sites with 16 levels. Each site represented a sub-matrix. In this structure the genetic variance (σ^2_c) associated with the clone effect was assumed homogeneous across 16 sites ($\sigma^2_{c1} = \sigma^2_{c2} = \sigma^2_{c3} = \dots = \sigma^2_{c16}$), but the genetic variance can be heterogeneous, different ($\sigma^2_{c1} \neq \sigma^2_{c2} \neq \sigma^2_{c3} \neq \dots = \sigma^2_{c16}$). The off-diagonal elements (e.g., $\sigma_{1,2}$) are the covariances of individuals between pair of sites. The covariances can be homogeneous ($\sigma_{1,2} = \sigma_{1,3} = \dots = \sigma_{1,16}$) or heterogeneous ($\sigma_{1,2} \neq \sigma_{1,3} \neq \dots \neq \sigma_{1,16}$) to model GxE interactions. We tested a diagonal matrix containing homogeneous variance components associated with the genetic effect and homogeneous correlations of the individuals across the sites. This structure is commonly named compound symmetry or CS (Gilmour *et al.*, 2009). Also, we tested heterogeneous genetic variances and homogeneous covariances (CH structure). The full G matrix is the product of two matrices with the dimension of 16×16 (number of sites) and the identity matrix I with dimensions of number of clones (I_{nc}) at each site. Since individuals are not independent but genetically related, I matrix at each site can be substituted by the numerator relationship matrix A .

Finally, we specified both heterogeneous genetic variances and genetic covariances for G structure. This was achieved using a factor analytic form or FA1 structure (de los Campos and Gianola, 2007). According to Meyer (2009) and Raman *et al.* (2011), the objective of the factor analytic approach is oriented to GxE interaction experiments, where there is accounting for the genetic covariances among sites in terms of smaller number of k unknown factors. In FAK model, the covariance matrix was modeled as $\Sigma FA = \Gamma\Gamma' + \Psi$ where Γ is the matrix of loadings on the covariance scale, and $\Psi = \text{diag}\{\psi_i\}$ is a diagonal vector of specific variances. For $k = 1$ environmental covariate or factor, the matrix representation was:

$$\Gamma = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_{16} \end{bmatrix} \quad \Psi = \begin{bmatrix} \sigma_{c1}^2 & 0 & \cdot & 0 \\ 0 & \sigma_{c2}^2 & \cdot & 0 \\ 0 & 0 & \cdot & \cdot \\ 0 & 0 & \cdot & \sigma_{c16}^2 \end{bmatrix}$$

$$\Sigma FA = \begin{bmatrix} \sigma_{c1}^2 + \sigma_1 \sigma_1 & \sigma_1 \sigma_2 & \cdot & \sigma_1 \sigma_{16} \\ \sigma_1 \sigma_2 & \sigma_{c2}^2 + \sigma_2 \sigma_2 & \cdot & \sigma_2 \sigma_{16} \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_1 \sigma_{16} & \sigma_2 \sigma_{16} & \cdot & \sigma_{c16}^2 + \sigma_{16} \sigma_{16} \end{bmatrix} \quad [3]$$

where σ_1 is the loading on a covariance scale for the first site. Also, $\sigma_{c1}^2 \neq \sigma_{c2}^2 \neq \sigma_{c3}^2 \neq \dots, \sigma_{c16}^2$. Similarly, In FAK model, the variance-covariance matrix $\Sigma^{(w \times w)}$ was modeled on the correlation scale as $\Sigma = DCD$, where $D^{(w \times w)}$ is diagonal such that $DD = \text{diag}(\Sigma)$. The component $C^{(w \times w)}$ is a correlation matrix of the form $LL' + E$, where $L^{(w \times k)}$ is a matrix of loadings on the correlation scale and E is diagonal defined such that $\text{diag}(LL' + E)$ is identity. For $k > 1$, higher-order models are required but they were not tested in our study. For all the matrix structures compared, the covariance was obtained from the definition of the genetic correlation (r_g), which was defined by Falconer and Mackay (1996). All the mixed model equations were solved by ASReml version 3 (Gilmour *et al.*, 2009), which was used in this work due to its flexibility to analyze unbalanced designed experiments, multi-environment trials, and irregular spatial data.

Model fit statistics

According to Gilmour *et al.* (2009), a general method for comparing the fit of nested models fitted by restricted maximum likelihood (REML) is the REML likelihood ratio test (LRT). The fixed effects for models must be the same for LRT to be valid but also the same parameterization is required. If L_{R2} is the REML log-likelihood of the more general model and L_{R1} is the REML log-likelihood of the restricted model (that is, the REML log-likelihood or simply LogL under the null hypothesis), then the LRT can be defined as the following statistic: $D = 2 \log(L_{R2}/L_{R1}) = 2 [\log(L_{R2}) - \log(L_{R1})]$. If R_i is the number of parameters estimated in model i , then the asymptotic distribution of the LRT under the restricted model is χ^2 with degrees of freedom = $R2 - R1$. The LRT is implicitly two-sided. The approximate P value for the LRT statistic D , is $0.5\{1 - \Pr(\chi^2 \leq d)\}$ where d is the observed value of D . Different models were compared using the LRT calculated for each model and tested against the χ^2 distribution and $P < 0.05$. Results of the model fit statistics are listed in Table 2.

Heritability estimates and G×E interaction

We accounted for relatedness among clones using pedigree relationships in estimations of variance components and standard errors. The variance components estimates were used to estimate single site broad-sense heritability (H^2) for each fitted variance-covariance structure definition, using the following general ratio of the genetic variance

and phenotypic variance: $H^2 = [\sigma_c^2 + \sigma_f^2 / (\sigma_c^2 + \sigma_f^2 + \sigma_e^2)]$, where the different variances were defined in Equation [1]. For spatial AR1+ σ_η^2 with a disentangled residual variance into a spatial residual and independent residual variance, we used the independent residual variance (σ_η^2) for the heritability estimation. Heritability estimates and their standard errors were estimated using the Delta method (Lynch and Walsh, 1998), and implemented using ASReml (Gilmour *et al.*, 2009). The results of the main variance components and heritability estimates were listed in Table 3. As a graphical measurement of G×E interaction, a linear plot between site loadings and BLUP for a random subset of 20 clones was produced from the solution file of AR1 + σ_η^2 + FA1 structure, to examine the G×E interaction for each site loading used in the modeling (Meyer, 2009), as shown in Figure 1.

RESULTS AND DISCUSSION

We reported a series of statistical approaches to analyze clonal replicated progeny for volume in loblolly pine. The best model fit was achieved by controlling the error variances using a spatial design plus its combination with a factor analytic structure for the clonal and family co(variances).

Variance-covariance structures comparison

In our study, any model that made R and G structures different from default IID assumption for standard

Table 3. Clonal variance (σ_c^2), error variance (σ_e^2), block variance (σ_b^2), and broad-sense heritability (H^2) according to different variance-covariance structures. The structure AR1+ σ_η^2 +FA1 increased the heritability and decreased both the error and block variances, compared with the default IID+FA1 model.

Structure	σ_c^2	σ_e^2	σ_b^2	H^2
IID+FA1	8.95 ¹	14.79 ²	1.25	0.38
AR1+ σ_η^2 +CS	9.07	12.67 ³	0.78	0.42
AR1+ σ_η^2 +CH	9.06 ¹	12.71 ³	0.78	0.43
AR1+ σ_η^2 +FA1	9.05 ¹	12.72 ³	0.79	0.45

¹Value was the mean of 16 clonal variances across sites.

²Value was the mean of 16 residual variances across sites.

³Value reported was the independent error variance.

The standard error of the heritability estimates ranged from 0.02 to 0.06.

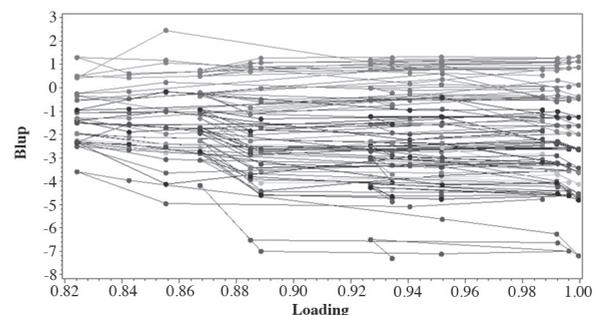


Figure 1. Estimated breeding values using best linear unbiased predictions (BLUP) vs. 16 site loadings (two loading values were equal) of a random subset of 20 clones after fitting of AR1+ σ_η^2 +FA1 structure. The general flat trend of lines for most of the predictions across different site loadings values was a clear sign of small G×E interaction.

models, was superior in terms of LogL parameter. The lowest LogL value (more positive) was obtained for the combination of spatial design and factor analytic form ($AR1 + \sigma^2_{\eta} + FA1$). After calculating the LRT between this model and each of the others, it was found a highly statistical significant difference ($P < 0.05$). The standard model has been used in previous studies (Dutkowski *et al.*, 2002; Smith *et al.*, 2002; Raman *et al.*, 2011) as a benchmark model for simplicity reasons or maybe due to the more difficult implementation of advanced models. We found that the only constraint to use advanced models was a longer computational time required for convergence (mainly for structures CH and FA1). The CS structure was the simplest structure for genetic co(variances). It only required the estimation of a correlation parameter and one genetic pooled variance, but it also implied that all sites had the same genetic variance and all pairs of sites had the same covariance (hence the correlation was the same, as also assumed for CH structure). This may not be a realistic scenario for progeny tests of forest trees tests, especially because they were planted in different edaphic, climatic conditions. The CS structure was also found to have a high Akaike Information Criterion value when it was compared with heterogeneous variance-covariance structures for treatment, site factors in crops (Raman *et al.*, 2011). They pointed out that treatments evaluated in different environments (considered as random factor), had different levels of error variation as the tests were conducted with different levels of precision, which was not captured by a simple structure of co(variances) as CS.

We preferred the FA1 structure for genetic co(variances) because it allowed to model the full G matrix, with heterogeneous co(variances). Moreover, FA1 structure combined with spatial design increased the model performance (best and significant LogL value), by increasing the clonal variance mean for 16 sites and decreasing the residual variances. The better estimates for this structures resulted in increased clonal mean repeatabilities for almost all sites (not shown), which represented an important gain by allowing variances to be different for each site. A fair comparison with FA1 structure should consider to fit a similar advanced structure. One example is the unstructured form of G matrix. It can be used to calculate the genetic correlations among pairs of sites. However, Smith *et al.* (2002) and Isik *et al.* (2008) suggested that unstructured G may be inefficient or unstable even for large number of sites or environments. We tested an unstructured G form that lead to less precise parameter estimation due to the large number of parameters (183 in our experiment), in addition to the presence of missing values. Therefore, the factor analytic form represented a more parsimonious representation and hence was reported in this study.

This experiment demonstrated the feasibility of using factor analytic form for the G matrix. The advantage is that fitting this structure to the covariance matrix

was flexible enough to accommodate heterogeneity of variances and differences in genetic correlations between environments, which tended to be small. Also, the structure was parsimonious enough to allow estimation of the parameters involved with high accuracy. Additional improvement could be achieved using FAK models of second, or third order, but when $k > 1$, it is necessary to impose constraints to ensure uniqueness of the structure and avoid interpretation problems of the factor analytic model. Smith *et al.* (2002) demonstrated that a FA2 model accounted for 77% of the genetic variation against 50% accounted by a FA1 model for a multi-environmental test study.

In addition, the use of a spatial design adding an independent residual variance component allowed the removal of spatially dependent residual noise from the model, which was expected according to the theory of models that include both sources of non-genetic variation (Stefanova *et al.*, 2009). This finding was important evidence of the need to separate spatial from the independent residual variances, in order to capture the random tree-to-tree variation. Dutkowski *et al.* (2002) pointed out that the independent residual term is almost always present in forestry trials and accounting for it is necessary. This is also made to avoid inflating the additive genetic variance and avoid misleading conclusions from fitting an inappropriate model.

Variance components and genetic parameters

Table 3 summarized the main variance components for the four scenarios under analysis. To facilitate the comparison, an average of variance was used for the models with heterogeneous genetic variances (CH and FA1). It was noticeable that the standard error mean was stable for different model scenarios. Also, the residual variance and block variance for each scenario was reported, where both non-genetic variances were reduced for the $AR1 + \sigma^2_{\eta}$ form with any combination with a G structure.

We obtained higher broad-sense heritabilities or H^2 estimates for sites when more complex genetic and environmental variances were used. For example, the mean broad-sense heritability across sites was 0.38 for IID scenario, and increased from 0.42 to 0.45 for AR1 plus a complex (co)variance structure. In general, the clone broad-sense heritability was higher than other studies for growth traits in loblolly pine (Gezan *et al.*, 2006). This was due to the large number of trees used for each clone in this study, resulting in more reliable estimates of the variances. The statistical analysis was able to reduce error and minimize bias in the estimates. Thus, the use of clonally replicated progeny tests reduced the environmental noise and enabled more accurate prediction of genetic merit for different clones. The use of repeated measures of the clonal data is an experimental design suitable for being applied in other experiments such as the study of the change in a trait or variable measured at

different times, or individuals exposed to different level of the same treatment (White *et al.*, 2007).

Since we had data from 16 different sites, we used FA1 to include an advanced structure for modeling G×E interaction. Genetic correlations were high, with a mean of 0.86 and a range of 0.65 to 0.99 (the standard error ranged from 0.01 to 0.04). They were obtained for the best fitted $AR1 + \sigma^2_{\eta} + FA1$ structure. The high values of genetic correlations obtained for FA1 structure were evidence that there were no strong G×E interaction in our clonal data. The small G×E interaction effect was reflected in the flat trend of estimated breeding values of a subset of the clones for each of the 16 site loadings values obtained from FA1 structure plotted in Figure 1. The same trend was observed for all 453 clones, but for visual simplicity Figure 1 only plotted 20 clones. If interaction was present, the cross of lines for different loadings would have been evident. The analysis of G×E is relevant in forestry breeding programs due to the diverse range of environments where tree can be cultivated. The main advantage of FA1 model was to account for the (co) variances in terms of smaller hypothetical factors (de los Campos and Gianola, 2007). The FA1 approach could be used to obtain a class of structures for the genetic variance matrix *G*. The model is also postulated in terms of the unobserved clonal effects in different sites, which had associated loading factors (Meyer, 2009).

CONCLUSIONS

The use of spatial structure combined with a heterogeneous factor analytic structure for genetic variances and covariances was more efficient than simple IID assumption for variances commonly used in forestry tests, because of better controlling of site variability and removing of spatial-related variance. The factor analytic structure was found to increase the estimates genetic variances and broad-sense heritability due to effective accounting for heterogeneity of effects among pair of sites. The use of advanced linear mixed models for clonal data in a commercial pine species was proved to be more efficient than simple assumptions because of increased broad-sense heritability, reliable breeding values were obtained, and issues of model convergence, computational time were avoided for larger data sets.

ACKNOWLEDGEMENTS

The author thanks Plum Creek Timber Company, Inc., CellFor, Inc. for field measurement data, Dr. Fikret Isik, Dr. Ross Whetten, and Dr. Gary Hodge for their useful input in the data analysis. This work was supported by the North Carolina State University (NCSU) Cooperative Tree Improvement Program, and the Department of Forestry and Environmental Resources at NCSU.

Uso de la estructura de factor analítico para aumentar la heredabilidad de ensayos clonales de *Pinus taeda* L. Diversas estructuras avanzadas de varianzas-covarianzas se han utilizado comúnmente en análisis genético de cultivos agrícolas para detectar la variabilidad del micro-sitio y lograr una alta precisión en las predicciones del valor genético, mejorando la eficiencia de la selección. Diferentes estructuras de varianza-covarianza fueron exploradas para predecir el mejor predictor lineal insesgado del valor genético en 453 clones de *Pinus taeda* L., evaluados a través de 16 sitios en el sureste de Estados Unidos. Los modelos estadísticos fueron comparados usando parámetros de diagnóstico, componentes de varianza y parámetros genéticos. De los modelos comparados, el mejor modelo fue el escenario con ajuste de la varianza residual independiente de la forma autoregresiva para la matriz *R* más una estructura de factor analítico para la matriz *G*. El modelo generó el menor valor del parámetro de diagnóstico (LogL igual a -2694), una baja varianza residual (12,72), y la más alta heredabilidad en sentido amplio (0,45), comparada con las estructuras básicas de varianzas-covarianzas homogéneas, a un nivel de significancia de $P < 0,05$. Se concluye que la combinación de una estructura específica para el efecto genético y residual resultó efectiva para remover la variabilidad relacionada con el espacio, e incrementar la exactitud de la predicción de los valores genéticos, lo cual podría usarse como herramienta analítica para aumentar la eficiencia de la selección en ensayos genéticos forestales.

Palabras clave: modelo lineal mixto, genética forestal cuantitativa, varianza genética.

LITERATURE CITED

- Balding, D.J., M. Bishop, and C. Cannings. 2007. Handbook of statistical genetics. 3rd ed. 1445 p. Wiley-InterScience, Wiltshire, England.
- Bettinger, P., M. Clutter, J. Siry, M. Kane, and J. Pait. 2009. Broad implications of southern United States pine clonal forestry on planning and management of forests. *International Forestry Review* 11:331-345.
- Costa e Silva, J., G.W. Dutkowski, and A.R. Gilmour. 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.
- Cumbie, W.P. 2010. Association genetics for growth, carbon isotope discrimination, and stem quality in loblolly pine. 153 p. PhD. thesis. North Carolina State University, Raleigh, North Carolina, USA.
- De los Campos, G., and D. Gianola. 2007. Factor analysis models for structuring covariance matrices of additive genetic effects: a Bayesian implementation. *Genetics Selection Evolution* 39:481-494.
- Dutkowski, G.W., J. Costa e Silva, A.R. Gilmour, and G.A. Lopez. 2002. Spatial analysis methods for forest genetic trials. *Canadian Journal of Forest Research* 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. *Canadian Journal of Forest Research* 36:1851-1870.

- Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 464 p. Longman Group, Essex, England.
- Gauch, H.G. 1992. Statistical analysis of regional yield trials: AMMI analysis of factorial designs. 278 p. Elsevier Science Publishers B.V., Amsterdam, The Netherlands.
- Gezan, S.A., T.L. White, and D.H. Huber. 2006. Achieving higher heritabilities through improved design and analysis of clonal trials. *Canadian Journal of Forest Research* 36:2148-2156.
- Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2009. ASReml user. Guide release 3.0. 372 p. VSN International Ltd., Hemel Hempstead, UK.
- Henderson, C.R. 1984. Applications of linear models in animal breeding. 462 p. University of Guelph, Guelph, Canada.
- Isik, F., B. Goldfarb, A. LeBude, B. Li, and S. McKeand. 2005. Predicted genetic gains and testing efficiency from two loblolly pine clonal trials. *Canadian Journal of Forest Research* 35:1754-1766.
- Isik, F., M. Gumpertz, B. Li, B. Goldfarb, and X. Sun. 2008. Analysis of cellulose microfibril angle using a linear mixed model in *Pinus taeda* clones. *Canadian Journal of Forest Research* 38:1676-1689.
- Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. 980 p. Sinauer Associates Sunderland, Massachusetts, USA.
- Meyer, K. 2009. Factor-analytic models for genotype \times environment type problems and structured covariance matrices. *Genetics Selection Evolution* 41:21.
- Raman, A., J.K. Ladha, V. Kumar, S. Sharma, and H.P. Piepho. 2011. Stability analysis of farmer participatory trials for conservation agriculture using mixed models. *Field Crops Research* 121:450-459.
- SAS Institute. 2010. SAS User's guide. Version 9.1. 7886 p. SAS Institute, Cary, North Carolina, USA.
- Sherrill, J.R., T.J. Mullin, B.P. Bullock, S.E. McKeand, R.C. Purnell, M.L. Gumpertz, and F. Isik. 2008. An evaluation of selection for volume growth in loblolly pine. *Silvae Genetica* 57:22-28.
- Smith, A., B. Cullis, and R. Thompson. 2002. Exploring variety-environment data using random effects AMMI models with adjustments for spatial field trend: Part 1: Theory. *In* Kang, M.S. (ed.) Quantitative genetics, genomics and plant breeding. Oxford University Press, Cary, North Carolina, USA.
- Stefanova, K.T., A.B. Smith, and B.R. Cullis. 2009. Enhanced diagnostics for the spatial analysis of field trials. *Journal of Agricultural, Biological, and Environmental Statistics* 14:392-410.
- Whetten, R.W., and R. Kellison. 2010. Research gap analysis for application of biotechnology to sustaining US forests. *Journal of Forestry* 108(4):193-201.
- White, T.L., W.T. Adams, and D.B. Neale. 2007. Forest genetics. 702 p. CABI Publishing, Cambridge, Massachusetts, USA.