A Proposed Method for Estimation of Genetic Parameters on Forest Trees Without Raising Progeny: Critical Evaluation and Refinement

By D. KUSNANDAR[1)2)] and N. GALWEY[1)]

(Received 12th August 1999)

Abstract

A method for estimation of genetic parameters of forest trees without raising progeny, the SAKAI and HATAKEYAMA method proposed in 1963, is evaluated. This method assumes that all variation not genetic in origin follows a spatial pattern, and that genetic variation is randomly distributed in space. It is applied here to simulated data and to a progeny trial of maritime pine (*Pinus pinaster* AIT.). Even though the method has been widely used, it has an inadequate theoretical foundation and seems rarely to give good estimates. With simulated data, the method failed to provide a reasonable estimate of genetic variance when the spatial trends in the data were weak. With data from the progeny trial, the method overestimated the genetic variance. A modification to the SAKAI and HATAKEYAMA method is proposed, using a modern approach to spatial analysis. The modified method has a better theoretical foundation, and gives slightly better estimates of genetic variance when applied to the simulated data. Nevertheless, the estimates only approached the true values when the number of genotypes was large and either the spatial autocorrelation in the error term was strong or the ratio of genetic to error variance was large. The modified method gave a worse estimate of genetic variance than the unmodified SAKAI and HATAKEYAMA method when applied to the progeny trial. It is concluded that the use of the SAKAI and HATAKEYAMA method, either its original form or the modified form, can rarely if ever be recommended. However, if limited information on the genetic relationships between trees is available, either from pedigrees or from molecular-genetic markers, more modern methods of spatial analysis should be of great value in the estimation of genetic parameters.

Key words: Autocorrelation, heritability, progeny trial, SAKAI and HATAKEYAMA method, simulation, spatial analysis, variance components.

Introduction

Knowledge of genetic parameters, such as genetic variance, narrow- and broad-sense heritability and genetic correlation, is essential to the efficient operation of forest-tree breeding programs. These parameters are important in predicting the magnitude and direction of the response of a given trait to selection on the basis of the same trait or another trait (ROFF, 1997; FALCONER, 1989). Many methods have been proposed for the estimation of genetic parameters, and these can be applied in forestry as in other biological systems. However, they nearly all require pedigree information on each tree in the forest. Therefore such methods are only applicable in a specially designed forest. In some forestry trials no such pedigree information is available. The use of these methods is also constrained by the long rotations and long breeding cycles of forest tree species, which limit the opportunities for comparison of parents and progeny.

In an attempt to overcome the limitations of the standard methods, SAKAI and HATAKEYAMA (1963) proposed a method for estimation of genetic parameters in forest trees without raising progeny. They claimed that, provided certain conditions were fulfilled, the method could be applied in any forest plantation to estimate the broad-sense heritability of a variable and the genetic correlation between variables. Their method was based on a study by SMITH (1938) on agricultural crops to examine the relationship between between-plot variance and the size of the plot. The method assumes that neighbouring individuals will share a more similar environment than individuals further apart. This phenomenon has often been observed during the analysis of uniformity trials (e.g. SMITH, 1938; MODJESKA and RAWLINGS, 1983), and is referred to as positive spatial correlation.

The assumption that neighbouring individuals will share a similar environment is also the basis of a set of methods known as spatial analysis (WILKINSON *et al.,* 1983; GLEESON and CULLIS, 1987; CULLIS and GLEESON, 1991). Spatial analysis is now widely used in the breeding of annual crops to identify variation that is common to plots in a particular part of a field trial and to eliminate this variation from the error associated with comparison between genotypes. The rapid development of spatial analysis techniques during the last two decades has made them applicable in a variety of research disciplines. In particular, SAKAI and HATAKEYAMA's method can be modified to take advantage of modern spatial-analysis principles and techniques. This gives the method a better theoretical foundation. Moreover, whereas the original method only permits the estimation of variance components, such a modification permits estimation of the genetic effects for individual trees. Both in SAKAI and HATAKEYAMA's method and in other methods of spatial analysis, the assumption is made that genotypes are distributed randomly in space, and hence that spatial patterns are not due to genetic effects. This assumption is almost certain to be invalid in a natural forest, but is reasonable in a forest plantation, as it is in crop field trials.

SAKAI and HATAKEYAMA's method has been widely used (SUHAENDI, SOERIANEGARA and NASOETION, 1976; SANTOSO, 1981; KUSNANDAR, 1982 and more recent unpublished studies in Indonesia), but it appears that its validity has not previously been the subject of a thorough investigation. In this paper, SAKAI and HATAKEYAMA's method is evaluated using simulated data and data from a progeny trial of maritime pine (*Pinus pinaster* AIT.). An alternative method of estimating genetic parameters is also proposed and evaluated, namely a modification of Sakai and Hatakeyama's method that utilises a modern approach to spatial analysis.

Theoretical Background

The SAKAI and HATAKEYAMA method

SAKAI and HATAKEYAMA's method assumes that the genotypes are randomly distributed over the research field and that the field has a certain spatial pattern of environmental variation.

[1)] Faculty of Agriculture, The University of Western Australia, Nedlands, Western Australia 6907, Australia

[2)] Permanent address: Faculty of Agriculture, Tanjungpura University, Jl. A. Yani, Pontianak 78124, Indonesia

This pattern causes neighbouring observations to have similar values, as noted by SMITH (1938), but does not follow any more specific model (PEARCE, 1976). The trees in the field are repeatedly grouped into clusters, the size of which is increased at each iteration. At each iteration, the between-cluster mean square is calculated. The values obtained are fitted to the following model:

$$MS_x = G + x^b E \quad \ldots\ldots\ldots\ldots \quad (1)$$

where MS_x is the between-cluster mean square for cluster size x; G and E are interpreted as the genetic and environmental variances, respectively; x is the cluster size; and b is a constant with a value between zero and one (SAKAI and HATAKEYAMA, 1963).

In order to estimate the parameters of Model (1), SAKAI and HATAKEYAMA set several values of b between zero and one by trial and error. The model could then be fitted using the simple linear regression technique. Estimates of G and E were obtained from the iteration that gave the best fit. Possibly the model was fitted in this way because of the limitations of the computing devices available at the time. With the more recent development of computers and statistical packages, estimate of b, G and E can be obtained simultaneously using non-linear regression techniques.

The spatial model

Spatial analysis is usually applied to n data obtained from a rectangular field that is indexed by the rows and columns of an r x c array. The model used in spatial analysis is similar to the ordinary general linear model, but with the addition of spatial terms. Let **y** be the vector of observations in field order (e.g. in column-wise order). An ordinary general linear model for **y** is

$$\mathbf{y} = \mathbf{X}\tau + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad \ldots\ldots\ldots \quad (2)$$

The spatial model for **y** is

$$\mathbf{y} = \mathbf{X}\tau + \mathbf{Z}\mathbf{u} + \xi + \eta \quad \ldots\ldots \quad (3)$$

where τ is a vector of length t of fixed effects associated with an n x t design matrix **X**; **u** is a vector of length b of random effects associated with an n x b design matrix **Z**; **e** = ξ + η is the vector, of length n, of individual error values; ξ is a vector of spatially dependent random error values; η, sometimes called a nugget effect, is a vector of random error values with mean zero whose elements are pair-wise independent. The vector **u** is composed of a series of q sub-vectors, corresponding to the q random factors in the model. Thus $\mathbf{u}' = (\mathbf{u}_1', \mathbf{u}_2' \ldots \mathbf{u}_q')$ where $\mathbf{u}_i$ is a vector comprising the effects of the levels of the ith random factor. In practice, in addition to dummy variables specifying the experimental design, matrices **X** and **Z** in the spatial model may include smoothing splines and/or polynomial functions of the spatial coordinates (GILMOUR, CULLIS and VERBYLA, 1997).

The assumption used in the implementation of Model (3) is as follows. We define

$$\gamma' = \left[\frac{\sigma_1^2}{\sigma^2}, \frac{\sigma_2^2}{\sigma^2} \cdots \frac{\sigma_q^2}{\sigma^2}\right] \quad \ldots\ldots\ldots \quad (4)$$

where σ_i^2 = variance component due to the ith random term; and σ^2 = error variance. The joint distribution of (**u,** ξ, η) is assumed to follow a multivariate-Normal distribution with mean zero and variance-covariance matrix **V**, where

$$\mathbf{V} = \sigma^2 \begin{bmatrix} \mathbf{G}(\gamma) & 0 & 0 \\ 0 & \Sigma(\alpha) & 0 \\ 0 & 0 & \psi\mathbf{I} \end{bmatrix} \quad \ldots\ldots\ldots \quad (5)$$

Matrix **G** may be completely general, although in most cases it is a block diagonal matrix, i.e., $\mathbf{G} = \oplus_{i=1}^{q} \mathbf{G}_i$. For example, in an ordinary anova model $\mathbf{G}_i$ is taken to be $\gamma_i\mathbf{I}$, where γ_i is the common variance component ratio of the ith random factor. α is a vector of spatial covariance parameters (see below). Σ is a covariance matrix corresponding to the chosen model of ξ, and hence specified by α. GILMOUR *et al.* (1997) recommended that the model of ξ be chosen from the class of separable processes (MARTIN, 1990; CULLIS and GLEESON, 1991). Finally, $\psi = \frac{\sigma_\eta^2}{\sigma^2}$.

Many statistical packages are now available to analyse spatial data. For this paper spatial analysis was carried out using a beta version of ASREML for Windows 95 released in April 1999 (GILMOUR *et al.,* 1999).

Modification of SAKAI and HATAKEYAMA's method

In modifying the SAKAI and HATAKEYAMA method we adopt their assumptions, stated above. The spatial pattern in the data is modeled using Model (3) with an appropriate variance-covariance matrix Σ. Non-spatial sources of variation, including the genetic variation, will then be attributed to the independent error term η (the nugget effect). When the environmental variation can mostly be explained by spatial variation, η will provide an estimate of genetic effects. However, if the spatial variation only explains a part of the environmental variation, the resulting estimate of genetic variance will be confounded with the unexplained environmental variation, and hence will be biased upward.

Simulation Procedures

Data were simulated based on a rectangular field with 10 x 15 observations. The model used in simulation was as follows:

$$y_{ij} = \mu + g_k + e_{ij} \quad \ldots\ldots\ldots\ldots \quad (6)$$

where y_{ij} is the simulated observation at the location (i, j); μ is the general mean which is set to be zero for each simulation; g_k is the genetic effect associated with the kth genotype, allocated by a random permutation to location (i, j); e_{ij} is the error term which is simulated so as to be spatially correlated in both directions following the first-order separable autoregressive process, which is designated by AR1 × AR1 (MARTIN, 1990). The variance of e is set to be one for each simulation. The correlation between values of e in adjacent rows is determined by the autoregressive parameter ρ_r, and that between values in adjacent columns by ρ_c. In the present simulations these parameters are taken to be the same, i.e. $\rho_r = \rho_c = \rho$.

Ten simulations were generated for each of the 112 possible combinations of the following parameter values: number of genotypes (n_g = 2, 3, 5, 6, 10, 15, and 30); ratio between genotype and error variances (γ = 0.25, 0.5, 1, and 4); and autoregressive parameter of the error terms (ρ = 0.1, 0.3, 0.5, and 0.7). The genetic effects were generated based on a Gaussian distribution with mean zero and the variance required in order to give the chosen variance ratio. The spatially correlated error terms were generated by firstly generating a 10 x 15 matrix of independent values $\{e_{ij}\}$ from a Gaussian distribution with mean zero and variance one. The elements of the first column of this matrix were then modified, so as to make them spatially correlated, by the following formula

$$e_{i,1} \leftarrow \rho_r \times e_{i-1,1} + \sqrt{(1-\rho_r^2)} \times e_{i,1}$$

$$\text{for } i = 2 \ldots 10 \quad \ldots\ldots\ldots\ldots \quad (7)$$

where ρ_r is the autoregressive parameter for row. The elements of the first row of the matrix were similarly modified by the following formula:

$$e_{1,j} \leftarrow \rho_c \times e_{1,j-1} + \sqrt{(1-\rho_c^2)} \times e_{1,j}$$

$$\text{for } j = 2 \ldots 15 \quad \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

where ρ_c is the autoregressive parameter for column. The elements of the rest of the matrix were then modified using the formula:

$$e_{i,j} \leftarrow \rho_r \times e_{i-1,j} + \rho_c \times e_{1,j-1} - \rho_c \times \rho_r \times e_{i-1,j-1} +$$

$$\sqrt{(1-\rho_c^2)}\sqrt{(1-\rho_r^2)} \times e_{i,j}$$

$$\text{for } i = 2 \ldots 10,\ j = 2 \ldots 15 \quad \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

After these transformations, the new matrix will conform to the AR1 × AR1 process with autoregressive parameters ρ_r and ρ_c.

Estimation Procedures

Simulated data were analysed using the SAKAI and HATAKEYAMA method. The location of each individual was mapped on graph paper, and this display was used to group the individuals into clusters. At each cluster size, each individual was allocated to exactly one cluster. This process was performed for cluster sizes x = 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, and 20. For each cluster size, except 1, the process was repeated two or three times. The parameters of Model (1) were estimated on the resulting data set using the FITNONLINEAR directive of Genstat 5 (Genstat 5 Committee, 1993).

The simulated data were also analysed using the modified SAKAI and HATAKEYAMA method. To achieve this, the vector of genetic effects ($\mathbf{g}' = [g_1, g_2 \ldots g_{n_g}]'$) was excluded from the model and a spatially correlated error term and a nugget effect (η) were fitted: that is, the fitted model was

$$\mathbf{y} = \mu + \xi + \eta \quad \ldots\ldots\ldots\ldots\ldots\ldots (10)$$

Since the genetic effects are not explicitly included, they are expected to be estimated by the nugget effect. The reliability of the estimation is evaluated by examining the correlation between the estimates of nugget-effect values and the true values of gi, and by comparing the corresponding variance components.

Simulation Results

For many simulated data sets, the model-fitting process in the SAKAI and HATAKEYAMA method failed to converge. The average frequency of convergence was 63.5%, and the frequency varied between 20% and 100% over the 112 combinations of parameter values. The frequency of convergence was high in cases where the autoregressive parameter was large (sometimes 100%, average 76%), and low when the autoregressive parameter was small ($\rho \leq 0.3$; frequency sometimes as low as 20%, average 60%.)

In addition to the cases where model-fitting failed to converge, there were others in which the process converged to a boundary value of a parameter – either a variance component value of zero or a b value of zero or one. Only those cases in which model-fitting converged to non-boundary values were considered further and are presented here. These comprise 25.8% of all simulations.

The relationship between the estimate of genetic variance and the number of genotypes, obtained with different values of variance ratio and of the autoregressive parameter of the error term, is presented in the upper part of *figure 1.* The corresponding relationship for the error variance is presented in the lower part of the figure. Estimates of both genetic and error variances usually (but not always) approached the true values when the number of genotypes was large. SAKAI and HATAKEYAMA's method generally estimated genetic and error vari-

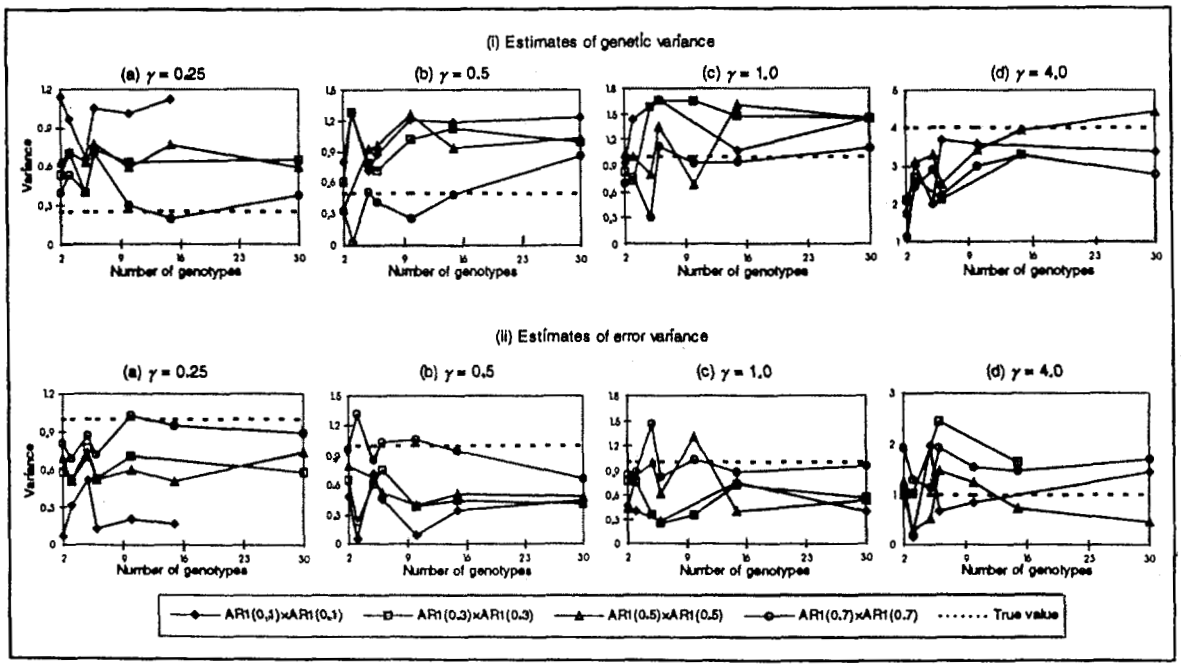

Figure 1. – Estimates of genetic and error variances for different values of γ and ρ obtained by the SAKAI and HATAKEYAMA method.

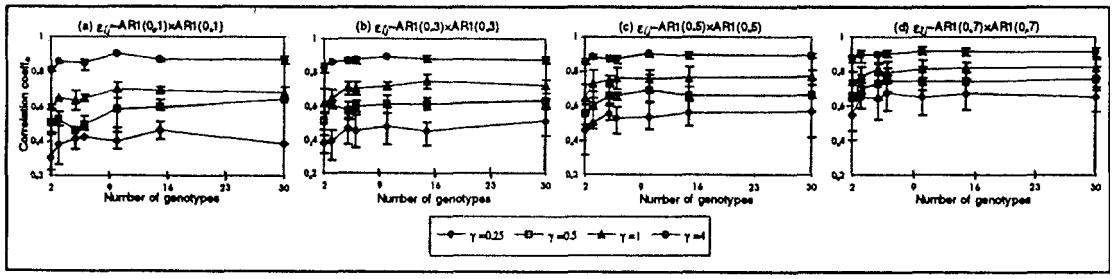

Figure 2. – Correlation between nugget effects and the true values of genetic effects for different values of γ and ρ.

ance fairly well when the variance ratio was small ($\gamma \leq 1$) and the autoregressive parameter in the error term was high (ρ = 0.7). However, when autocorrelation was weak, the method failed to estimate the variance components correctly. It generally over-estimated genetic variance and under-estimated error variance.

When the modified SAKAI and HATAKEYAMA method was fitted to simulated data, the nugget variance component was constrained to be positive: that is, it was given a boundary value of zero. The average frequency of convergence was 87% of cases. A higher frequency was achieved when the data were simulated with a high autoregressive parameter (ρ = 0.5 and ρ = 0.7) and a small variance ratio ($\gamma \leq 1$). In these cases, the frequency of convergence was sometimes 100%. However, problems occurred when the data were simulated with a low autoregressive parameter (ρ = 0.1). In these cases, even though the frequency of convergence was about 80%, the fitting process usually converged to the boundary value. As in the case of the SAKAI and HATAKEYAMA method, further interpretation was based only on the analyses of simulated data that converged to non-boundary values, which comprised 56% of the total.

The relationship between the true genetic effects and the estimates obtained from the modified SAKAI and HATAKEYAMA method was also investigated. The correlation coefficient between the two was higher when the number of genotypes was large, and reached a plateau when the number of genotypes was about 10. This relationship, for different values of the variance component ratio and of the autoregressive parameter of the error term, is illustrated in *figure 2.* Vertical lines at each data point indicate the range of the correlation coefficient. These results indicate that the nugget effect can be used as an alternative estimate of genetic effects if the variance component ratio is high. When the variance component ratio is low, the nugget effect may still be used as an appropriate estimate of the genetic effect provided that the spatial correlation in the error term is strong.

The estimates of nugget variance (variance due to $\eta = \sigma^2_\eta$) and spatial variance (variance due to ξ) approached the true values of the genetic variance and error variance respectively when the number of genotypes was large (*Figure 3*). The difference between the estimates and the true values was smaller when there was a medium to strong spatial correlation in the error structure ($\rho \geq 0.3$). The results indicate that in cases where the variance ratio is low (up to γ = 1), the nugget variance can be used as an estimate of the genetic variance only when both these conditions are met: large number of genotypes and strong spatial correlation in the error structure. For a high value of the variance ratio (γ = 4) a strong spatial correlation is not necessary when the number of genotypes is large.

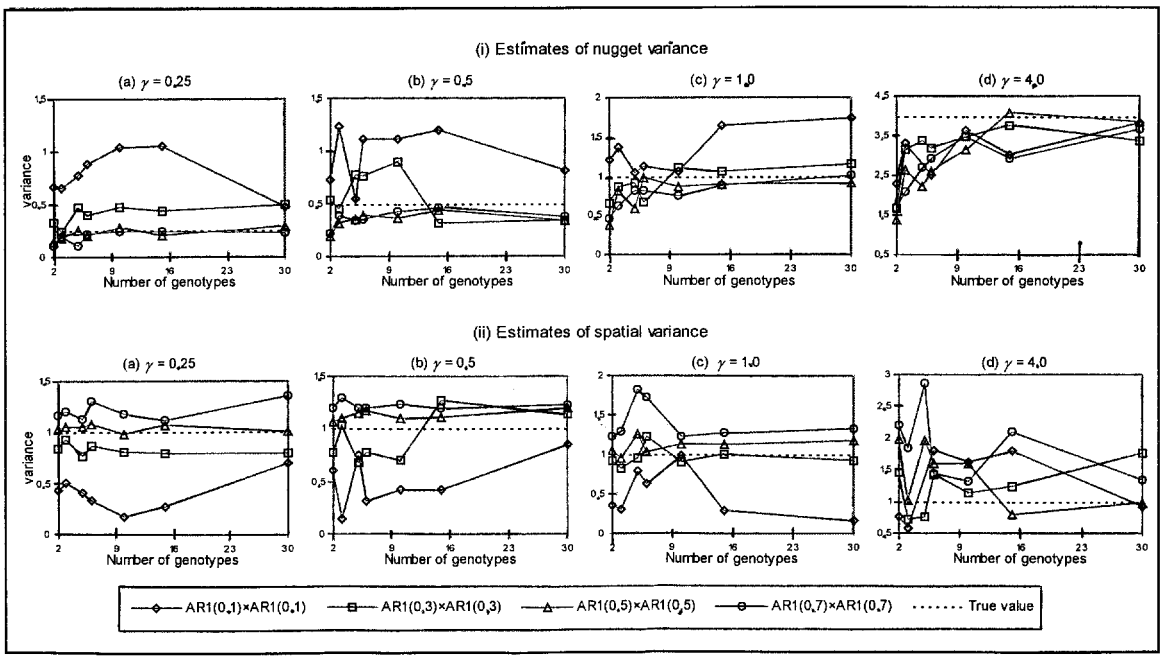

Figure 3. – Estimates of nugget and spatial variances for different values of γ and ρ obtained by the modified SAKAI and HATAKEYAMA method.

Table 1. – Overview of the methods applied and models fitted to maritime pine data.

Estimation method	Fixed effects	Variance Model		Variance component used as *var*(error)
		Global Variation	Natural variation	
1. Anova	fertiliser	family + rep.block + rep.block.plot	Identity	residual variance from anova
2. Spatial analysis (1)	fertiliser	family + rep.block + rep.block.plot	AR1×AR1	var(ξ)
3. Spatial analysis (2)	fertiliser	family + rep.block	AR1×AR1 + η	var(ξ) + var(η)
4. Modified Sakai & Hatakeyama	fertiliser	rep.block	AR1×AR1 + η	var(ξ)
5. Sakai & Hatakeyama	—	—	—	E from Equation 1

Application to a Progeny Trial of Maritime Pine

The SAKAI and HATAKEYAMA method and the modified method were used to estimate the genetic variance of height in a progeny trial of maritime pine (*Pinus pinaster* AIT.) in Gnangara Plantation, Wanneroo, Western Australia. The results were compare to those obtained by a conventional method of estimation, using information about the pedigree of each tree. The data were kindly provided by Dr. TREVOR BUTCHER of the Department of Conservation and Land Management, Western Australia. The trial consist of 15 full-sib families, derived from five male parents crossed with three female parents in all combinations, arranged in four replications, each consisting of six blocks of 90 trees. The blocks were separated by buffers consisting of two rows/columns of trees. The trees in the buffers were a mixture of several full-sib families. Six fertiliser treatments were applied, one to each block in each replication. Each block was divided into six plots, each containing one tree of each of the 15 families, but no treatment factor was assigned to the plots. Analyses of height at age 6 years are presented here. Other variables were also measured: further details of the experiment are given by HOPKINS and BUTCHER (1994).

Table 1 presents an overview of the methods used and the models fitted to the data. Model 1 is the anova model corresponding to the design of the experiment. The residual variance in this model will include a component due to genetic variation between trees within each family. This, like any other source of within-family variation, will interfere with the estimation of family effects in all models. Model 2 is a spatial analysis in which the effect of families is included explicitly (unlike the modified SAKAI and HATAKEYAMA model), but in which it is assumed that all error variation is spatially distributed. Model 3 allows for the possibility that this assumption may be incorrect by introducing a nugget effect. Model 4, the modified SAKAI and HATAKEYAMA model, does not include the effect of families explicitly and attempts to account for it by means of the nugget effect. Model 5 is the unmodified SAKAI and HATAKEYAMA model. In each model, each random term was judged to be significant if it had a large value of

$$t = \frac{\text{estimated variance component}}{\text{SE}_{\text{estimate}}} \qquad (11)$$

which is approximately a t statistic, and/or a large value of the change in residual deviance due to inclusion of the term, which is approximately a χ^2 statistic with 1 DF. *Table 2* presents the REML log-likelihood obtained with each model.

For the purpose of Estimation Methods 2 to 5 the data were treated as if they were obtained from an experiment with 36 rows × 60 columns of trees, ignoring the fact that buffer trees were present. This procedure may have affected the estimates of spatial parameters. However, some such adjustment was necessary as no measurements were taken on the buffer trees, and the solution adopted had the merit of making the analyses comparable to Estimation Method 1.

For each estimation method, an estimate of heritability (broad sense) was derived from the variance components obtained. In the case of Estimation Methods 1 to 3, the variance among families was partitioned into components due to the male parent, the female parent and the male × female interaction (σ^2_m, σ^2_f and σ^2_{mf}). This permitted estimation of the additive genetic variance (V_A) and the variance due to dominance effects (V_D) (see KUSNANDAR *et al.*, 1998 for details of these calculations). The heritability was then calculated as

$$h^2 = \frac{V_A + V_D}{V_P} \qquad (12)$$

where

$$V_P = \sigma^2_m + \sigma^2_f + \sigma^2_{mf} + var(\text{error}), \qquad (13)$$

the phenotypic variance. In the case of Estimation Method 4 and 5, the variance among families was not specifically estimated, and the non-spatial variation was taken to be an estimate of the total genetic variance. The heritability was then calculated as

$$h^2 = \frac{var(\text{genetic})}{var(\text{genetic}) + var(\text{error})} \qquad (14)$$

The estimates of variance components obtained by each method, and the heritabilities, are presented in *table 3*. Because only a small number of parents are involved, all these methods may give imprecise estimates of the true genetic vari-

Table 2. – REML log-likelihoods of models fitted to maritime pine data.

Estimation method	Number of variance parameter	REML log-likelihood
1. Anova	4	-194.7
2. Spatial analysis (1)	6	-178.4
3. Spatial analysis (2)	6	-157.8
4. Modified Sakai & Hatakeyama	5	-197.3
5. Sakai & Hatakeyama	3	—

Table 3. – Estimates of variance components of maritime pine data.

Estimation method	Estimate of variance components						
	Family or genetic	Error	ρ_{col}	ρ_{row}	nugget	b	h^2 [a]
1. ANOVA	0.0198	0.4010	—	—	—	—	15.0
2. Spatial analysis (1)	0.0212	0.4092	0.110	0.068	—	—	15.9
3. Spatial analysis (2)	0.0218	0.0989	0.660	0.741	0.3433	—	14.7
4. Modified Sakai & Hatakeyama	0.3692 [b]	0.0917	0.670	0.750	— [b]	—	80.1
5. Sakai & Hatakeyama	0.2850 [c]	0.2411 [d]	—	—	—	0.777	54.2

[a]) Estimate of broad-sense heritability (%) [b]) Nugget variance was used as the estimate of family variance
[c]) Estimate of G in equation (1) [d]) Estimate of E in equation (1)

ance. However, the available data can legitimately be used to *compare* estimates of genetic variance and genetic effects. The values of the autoregressive parameters in Model 2, and the presence of a substantial nugget effect in Model 3, show that there was only a weak spatial pattern in the data. However, inclusion of the AR1 × AR1 error model significantly improved the fit. It also slightly increased the estimate of family variance relative to that obtained by the ANOVA estimation method.

The estimates of family variance obtained from both the SAKAI and HATAKEYAMA method and the modified SAKAI and HATAKEYAMA method are substantially biased upward. Indeed, in this case the modified method performs worse than the original. This can be explained by the weakness of the spatial pattern in the data. The genetic variance is confounded in the nugget variance with the non-spatial environmental variation, which is a dominant source of variation in this data set, and this combined variance is used as an estimate of genetic variance.

Discussion

Despite the poor rate of convergence, the SAKAI and HATAKEYAMA method performed well in estimating the genetic and error variances when the ratio between these variance components was small and the spatial autocorrelation in the error was strong. However, the method failed to provide reasonable estimates when the autocorrelation was weak. The modification to the SAKAI and HATAKEYAMA method partially overcame these limitations. The frequency of convergence was substantially improved, and the method was also applicable when the spatial autocorrelation was only moderate.

Both the SAKAI and HATAKEYAMA method and our modification of it are based on the assumption that the total variation in the data can be expressed as the sum of genetic variation and environmental variation. Both models further assume that the environmental variation can be accounted for the spatial pattern present in the data. Hence, when the spatial variation has been modeled, the remaining variation is considered to provide an estimate of the genetic variance. Therefore, when aspects of the environment that do not follow a spatial pattern are a major source of variation, they will then be confounded with the genetic variance.

There is an additional, more technical problem that arises in the application of the modified SAKAI and HATAKEYAMA method to some data sets. When the spatial autocorrelation in the error structure is weak, the nugget effect can sometimes not be estimated. This is because in the absence of an autocorrelation term, the distinction between spatially-distributed and nugget error disappears.

In practice, when information about pedigrees is not available, it will generally not be possible to assume that all non-genetic variation follows a spatial pattern. Consequently the SAKAI and HATAKEYAMA model can rarely if ever be recommended. Our modification of the SAKAI and HATAKEYAMA model has a more secure theoretical foundation, and gave reasonable estimates of genetic variance from a wider range of simulated data sets. Nevertheless, the same basic objection applies to it. Our approach to the estimation of genetic effects will, however, have great utility in cases where a small proportion of trees known to have closely related genotypes (preferably an identical genotype, i.e. individuals of the same clone) are distributed through the trial. Such trees will provide an estimate of the magnitude of non-genetic variation, and an indication of the spatial pattern of variation. Spatial analysis is routinely applied in this way to unreplicated field trials of breeding lines of annual crops. Replicate plots of one or a few control varieties are distributed through such trials, often in a regular grid (see for example BESAG and KEMPTON, 1986). Plant breeders have found that when spatial analysis is applied to such a trial design, selection can be practiced at a substantially earlier stage of the breeding program than was formerly possible.

In trials in which no pedigree information is available, spatial analysis methods may still be of value if another source of

Table 4. – Frequency of convergence of methods for estimation of genetic parameters.

Estimation method	Number of simulations	failed to converge (%)	converge to boundary value (%)	converge to non-boundary value (%)
Sakai & Hatakeyama	1120	36.5	37.7	25.8
Modified Sakai & Hatakeyama	1120	12.9	31.2	55.9

information can be found on the degree of genetic relationship between the trees. Such information may be provided by molecular-genetic markers. Information on such markers may be collected in a forestry trial, not only to determine which trees are close relatives but also to determine whether the genetic variation is randomly distributed in space, as has so far been assumed (see Introduction). If this proves to be the case, spatial analysis could be used not only to distinguish genetic, spatial and non-spatial environmental components of variance, much as SAKAI and HATAKEYAMA envisaged, but also to estimate the breeding value of individual trees.

Acknowledgements

We are grateful to AusAID for financial support. We thank Dr. ARTHUR GILMOUR for his suggestions on the simulation processes of correlated error terms. We also thank Dr. TREVOR BUTCHER of the Department of Conservation and Land Management, Western Australia, for making the maritime pine data available. We thank the referee for helpful comments on the draft manuscript.

Literature

BESAG, J. and KEMPTON, R.: Statistical analysis of field experiments using neighbouring plots. Biometrics **42**: 231–251 (1986). — CULLIS, B. R. and GLEESON, A. C.: Spatial analysis of field experiments – An extension to two dimensions. Biometrics **47**: 1449–1460 (1991). — FALCONER, D. S.: Introduction to Quantitative Genetics. 3rd ed. John Wiley & Sons, New York (1989). — Genstat 5 Committee: Genstat 5 Release 3 Reference Manual. Oxford University Press Inc., Oxford (1993). — GILMOUR, A. R., CULLIS, B. R. and VERBYLA, A. P.: Accounting for natural and extraneous variation in the analysis of field experiments. J.A.B.E.S. **2**: 269–293 (1997). — GILMOUR, A. R., CULLIS, B. R., WELHAM, S. J. and THOMPSON, R.: ASREML Reference Manual. NSW Agriculture Biometric Bulletin No. 3, N.S.W. Agriculture, Orange, Australia (1999). — GLEESON, A. C. and CULLIS, B. R.: Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. Biometrics **43**: 277–288 (1987). — HOPKINS, E. R. and BUTCHER, T. B.: Improvement of *Pinus pinaster* AIT. in Western Australia. CALMSci. **1**: 159–242 (1994). — KUSNANDAR, D.: Pendugaan Nilai Heritabilitas beberapa Sifat Morfologis *Pinus merkusii* JUNGH et DE VRIES. di Kesatuan Pemangkuan Hutan Pekalongan Barat, Perum Perhutani Unit I, Jawa Tengah (Estimation of heritability on some morphological characters of *Pinus merkusii* JUNGH et DE VRIES.). Thesis Sarjana Pertanian. Departemen Statistika dan Komputasi, Institut Pertanian Bogor, Bogor, Indonesia (1982). — KUSNANDAR, D., GALWEY, N. W., HERTZLER, G. L. and BUTCHER, T. B.: Age trends in variances and heritabilities for diameter and height in maritime pine (*Pinus pinaster* AIT.) in Western Australia. Silvae Genet. **47**: 136–141 (1998). — MARTIN, R. J.: The use of time-series models and methods in the analysis of agricultural field trials. Commun. Statist. A **19**: 55–81 (1990). — MODJESKA, J. S. and RAWLINGS, J. O.: Spatial correlation analysis of uniformity data. Biometrics **39**: 373–384 (1983). — ROFF, D. A.: Evolutionary Quantitative Genetics. Chapman & Hall, New York (1997). — PEARCE, S. C.: An examination of Fairfield Smith's law of environmental variation. J. Agric. Sci. Camb. **87**: 21–24 (1976). — SAKAI, K. I. and HATAKEYAMA, S.: Estimation of genetic parameters in forest trees without raising progeny. Silvae Genet. **12**: 152–157 (1963). — SANTOSO, H.: Heritability of some characters of teak (*Tectona grandis*). Technical Report, Seameo – Biotrop, Bogor, Indonesia (1981). — SMITH, H. F.: An empirical law describing heterogeneity in the yields of agricultural crops. J. Agric. Sci. **28**: 1–23 (1938). — SUHAENDI, H., SOERIANEGARA, I. and NASOETION, A. H.: Penetapan nilai heritabilitas untuk seleksi massa *Pinus merkusii* JUNGH et DE VRIES (Heritability for mass selection of *Pinus merkusii* JUNGH et DE VRIES). Report No. 228, Lembaga Penelitian Hutan, Bogor, Indonesia (1976). — WILKINSON, G. N., ECKERT, S. R., HANCOCK, T. W. and MAYO, O.: Nearest Neighbour (NN) analysis of field experiments. J. R. Statist. Soc. B **45**: 151–211 (1983).

Microsatellite DNA Profiling of Phenotypically Selected Clones of Irish Oak *(Quercus* spp.*)* and Ash (*Fraxinus excelsior* L.)

By F. LEFORT[1) 3)], G. C. DOUGLAS[1)] and D. THOMPSON[2)]

(Received 6th September 1999)

Abstract

Oak and ash are two of the main forest species in Europe. Because of their commercial importance, genetic improvement of such species is considered important. The recent availability of microsatellite sequences for both oak (*Quercus robur, Q. petraea*) and ash (*Fraxinus excelsior*) allowed the characterization of phenotypically selected clones of oaks and ash trees of Irish origin by microsatellite DNA profiling. Oak clones were characterised at nine microsatellite loci and ash clones at 10 microsatellite loci. Allele ranges in selected clones were found to be similar to those observed in natural stands of oaks in Austria and ash in France, but the number of alleles at each locus was higher. Heterozygosity differed between Irish and Austrian oaks at several loci. Analysis of microsatellite profiling provided individual profiles for each clone. Microsatellite data analysis was performed with the software NJBAFD and the calculations of stepwise weighted genetic distance showed the genetic distances between clones. Five clones from a managed oak stand, which were probably from the same source were found to be genetically related to several other Irish sources. Two Irish origins of ash were found to be related to a French source. Microsatellite profiling also showed three bands patterns at several loci in 5 oaks and at one locus in 7 ash trees, suggesting the occurrence of triploidy or aneuploidy. In the latter case, the hypothesis of locus duplication should be checked by crossing studies.

Key words: Fraxinus excelsior, microsatellite, *Quercus petraea, Quercus robur,* single sequence repeats (SSRs).

[1)] Teagasc, Agriculture and Food Development Authority, Kinsealy Research Centre, Biotechnology, Malahide Road, Dublin 17, Ireland.

[2)] Coillte Teoranta, The Irish Forestry Board, Research Laboratory, Newtownmountkennedy, Co. Wicklow, Ireland.

[3)] Laboratory of Plant Molecular Physiology and Biotechnology, Department of Biology, University of Crete, P.O. Box 2208, 714 09 Iraklion, Crete, Greece.

Corresponding author: F. LEFORT, Laboratory of Plant Physiology and Biotechnology, Department of Biology, University of Crete, P.O. Box 2208, 714 09 Heraklion, Crete, Greece.

E-mail : flefort@biology.uch.gr
Ph.: 00 30 81 394 072.
Fax : 00 30 81 394 408