

B. J.: Disk Electrophoresis II: Method and application to human serum proteins. *Ann. N. Y. Acad.* **121**: 404–427 (1964). — FINS, L. and LIBBY, W. J.: Population Variation in Sequoiadendron: Seed and Seedling Studies, Vegetative Propagation, and Isozyme Variation. *Silvae Genet.* **31**: 102–110 (1982). — FURNIER, G. R., MOHN, M. S. C. A. and CLYDE, M. A.: Geographic Patterns of Variation in Allozymes and Height Growth in White Spruce. *Can. J. For. Res.* **21**: 707–711 (1991). — GIBSON, J. P. and HAMRICK, J. L.: Genetic Diversity and Structure in *Pinus pungens* (table Mountain Pine) Population. *Can. J. For. Res.* **21**: 635–642 (1991). — GULLBERG, U. et al.: Allozymes Variation in Scots Pine (*Pinus sylvestris* L.) in Sweden. *Silvae Genet.* **34**: 193–201 (1985). — GURIES, R. P. and LEDIG, F. T.: Genetic Diversity and Population Structure in Pitch Pine (*Pinus rigida* MILL.). *Evolution* (Lawrence, Kans.) **36**: 378–420 (1982). — HAMRICK, J. L.: Gene Flow and Genetic Variation in Plant Populations in Differentiation Patterns in Higher Plants. Edited by K. URBANSKA. Academic Press, New York. Pp. 53–67 (1987). — HAMRICK, J. L. and GODT, J. W.: Allozyme Diversity in Plant Species. In: Plant Population Genetic, Breeding and Gene Resources. Edited by A. D. H. BROWN, M. T. CLEGG, A. L. KAHLER and B. S. WIER: Sinauer Associates Inc., Sunderland, Massachusetts. Pp. 43–63 (1989). — HUANG, Q. Q. et al.: Genetic Control of Isozyme Variation in Masson Pine, *Pinus massoniana* LAMB.

Silvae Genet. **43**: 285–291 (1994). — KIMURA, M.: Evolutionary Rate at the Molecular Level. *Nature* (London) **217**: 624–626 (1968). — O'MALLEY, D. M.: Inheritance of Isozyme Variation and Heterozygosity in *Pinus ponderosa*. *Biochemical Genetics* **17**: 233–250 (1979). — ORNSTEIN, L.: Disk Electrophoresis-II: Background and Theory. *Ann. Acad. Sci.* **121**: 321–349 (1964). — PLESSAS, M. E. and STRAUSS, S. H.: Allozyme Differentiation among Populations, Stands, and Cohorts in Monterey Pine. *Can. J. For. Res.* **16**: 115–164 (1986). — SCHILLER, G. et al.: Isozyme Variation among Nature Stand and Plantations of Aleppo Pine in Israel. *Isr. J. Bot.* **35**: 161–174 (1986). — S. GE et al.: The Genetic Variation in Natural stands of Masson pine (Chinese). *Scientific Silvae Sinicae* **24**(4): 399–408 (1988). — SHIRAIISHI, S.: Inheritance of Isozyme Variation in Japanese Black Pine, *Pinus thunbergii* parl. *Silvae Genet.* **37**: 93–100 (1988). — SUYAMA: Inheritance of Isozyme Variation in Natural Populations of *Abies mariesii* MAST. Thesis for Doctorate in University of Tsukuba (1992). — WHEELER, N. C. and GURIES, R. P.: Populations Structure, Genetic Diversity and Morphological Variation in *Pinus contorta* DOUGL. *Can. J. For. Res.* **12**: 595–606 (1982). — YAZDANI, R., MUONA, O., RUDIN, D. and SZMIDT, A. E.: Genetic Structure of a *Pinus sylvestris* L., Seed-tree Standard Naturally Regenerated Understorey. *For. Sci.* **31**: 430–436 (1985).

Optimising Early Selection Using Longitudinal Data

By L. A. APIOLAZA¹⁾²⁾³⁾, D. J. GARRICK¹⁾ and R. D. BURDON²⁾

(Received 17th January 2000)

Abstract

This study analysed the use of longitudinal data, i.e. repeated assessment of the same individuals at different ages, in the context of early selection. Autoregressive relationships, banded correlations and unstructured ('unsmoothed') matrices were used to model the additive genetic covariance matrix (G_0) for 10 total height measurements of a *Pinus radiata* open-pollinated progeny test. We examined the effects on response to selection of inferred covariance structure, mass versus combined selection, one or multiple assessments, and two breeding-delay intervals. End results are expressed as predicted average gain per year. The patterns of predicted response to selection vary widely between inferred covariance structures. Considering the autoregressive model (based on logarithm of age ratios between assessments) as an example, the effect of combining information from relatives on response to selection is more important (16% to 41% extra gain) than using extra measurements (2% to 25%), when predicting individual breeding values, although the economics of extra gain vs extra assessment costs must be carefully analysed. It is expected that using multiple assessments could be advisable for datasets with lower genetic auto-correlations. An approximate comparison across covariance models showed the autoregressive model to exhibit the best ability to produce 'correct' selections as well as the highest predicted response to selection.

¹⁾ Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand.

²⁾ New Zealand Forest Research Institute, Private Bag 3020, Rotorua, New Zealand.

³⁾ Current address: CRC for Sustainable Production Forestry, School of Plant Science, University of Tasmania, GPO Box 252-55, Hobart, Tasmania 7001, Australia.

Key words: longitudinal data, early selection, covariance structures, tree model, *Pinus radiata*.

Introduction

Trees included in breeding programs are often evaluated in progeny trials to predict their genetic value. Results from testing determine the participation of the trees in the breeding population, as well as their use as parents of future plantations (ZOBEL and TALBERT, 1984; WHITE, 1987). The breeding objective includes tree characteristics at harvest age (e.g. volume and wood density); however, progeny tests are assessed at one or more early ages, often less than half the rotation age. The problem of early selection arises with the existence of less than perfect genetic and phenotypic correlations between performance at early assessments and performance at harvest age.

Extending the testing period increases accuracy of selection, i.e. the correlation between predicted and real breeding values, but also increases financial costs and time delays to achieve gain. Optimising response to selection for a given objective involves finding the appropriate combination of accuracy and evaluation time. Traditionally, this has been achieved by calculating gain for different selection ages using the formula for correlated response to mass or index selection, which includes the heritability at early and mature ages and genetic correlation between the ages (SEARLE, 1965; FALCONER and MACKAY, 1996). The selection age that maximises either a biological criterion (response per year, e.g. LAMBETH, 1980) or an economic criterion (net present value, e.g. NEWMAN and WILLIAMS, 1991) may be chosen.

Genetic and phenotypic covariance structures used for calculating correlated response are often estimated from studies based on a few measurements per tree. Heritabilities are interpolated and extrapolated, perhaps using regression or splines (GWAZE et al., 1997), while genetic correlations are often modelled adapting LAMBETH's empirical phenotypic relationship (LAMBETH, 1980), based on either phenotypic or genetic correlations (LAMBETH, 1980; BURDON et al., 1992; GWAZE et al., 1997). WEI and BORRALHO (1996) proposed an alternative model for heritability based on the concept of repeatability. Positive definite additive genetic and phenotypic covariance matrices are not automatically assured by using any of these methods. Concurrently, MAGNUSSEN (1988, 1993) put forward different approaches to early selection, based on the size class distribution of the phenotypes at different ages. However, these latter procedures do not take into account genetic information.

Genetic evaluation tools in forest genetics have undergone a progressive refinement, from evaluation based on family-average (e.g. HATCHER et al., 1981) to the use of Best Linear Prediction (BLP, e.g. WHITE et al., 1987; WHITE and HODGE, 1988) and Best Linear Unbiased Prediction (BLUP, e.g. BORRALHO, 1995; JARVIS et al., 1995). In spite of this, determination of optimum selection time is often based only on response to selection using the most recent measurement of individual performance, even when more assessments were available at the time of analysis. BURDON (1989) suggested the use of longitudinal data, i.e. repeated assessment of the same individuals at different ages, to increase accuracy and therefore response to selection. These assessments can be integrated into a selection index.

This research analyses the implications of using longitudinal data when selecting at an early age. In the course of that we consider the effects of assuming different models for additive genetic variance, the use of repeated assessments combined in a selection index, and the use of mass and combined selection on the prediction of genetic gain. End results are expressed in terms of average predicted response per year.

Materials and Methods

Dataset

Genetic and phenotypic covariance matrices for constructing the indexes were estimated from a radiata pine (*Pinus radiata* D. DON) open-pollinated progeny test with 10 assessments at ages 1, 2, 4 to 9, 12, and 15 years from planting. The test included 45 open-pollinated families, planted in 5-tree row-plots within 8 randomised complete blocks. Trees suppressed by early competition were omitted from the analysis, leaving a total of 1526 trees. Observations after mechanical damage to leaders (especially from age 12 onwards) were omitted. Further details are described elsewhere (APIOLAZA et al., 2000).

Statistical model

Considering s assessments on individual i and defining $\mathbf{y}_i = [y_{i1} \ y_{i2} \ \dots \ y_{is}]'$ as the vector of phenotypic observations, the model equation for individual i is:

$$\mathbf{C}_{US} = \begin{bmatrix} 1 & a & b & c \\ a & 1 & d & e \\ b & d & 1 & f \\ c & e & f & 1 \end{bmatrix} \quad \mathbf{C}_{BC} = \begin{bmatrix} 1 & g & h & i \\ g & 1 & g & h \\ h & g & 1 & g \\ i & h & g & 1 \end{bmatrix} \quad \mathbf{C}_{ARlog} = \begin{bmatrix} 1 & r^{|\log(2/1)|} & r^{|\log(3/1)|} & r^{|\log(4/1)|} \\ r^{|\log(2/1)|} & 1 & r^{|\log(3/2)|} & r^{|\log(4/2)|} \\ r^{|\log(3/1)|} & r^{|\log(3/2)|} & 1 & r^{|\log(4/3)|} \\ r^{|\log(4/1)|} & r^{|\log(4/2)|} & r^{|\log(4/3)|} & 1 \end{bmatrix}$$

Figure 1. – Example of the unstructured (US), banded correlations (BC) and autoregressive with time on a natural logarithm scale (ARlog) models. Correlations with the same letter represent the same value. The BC model assumes similar correlations for measurements with equal time between expressions.

$$\mathbf{y}_i = \mathbf{X}_i \mathbf{m} + \mathbf{Z}_{Bi} \mathbf{b} + \mathbf{Z}_{Pi} \mathbf{p} + \mathbf{Z}_{Ti} \mathbf{a}_i + \mathbf{e}_i$$

where $\mathbf{m} = [m_1 \ m_2 \ \dots \ m_s]'$ is the vector of fixed effects (overall mean at each age), $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_{bxs}]'$ is the vector of $b \times s$ random block effects, $\mathbf{p} = [p_1 \ p_2 \ \dots \ p_{pxs}]'$ is the vector of $p \times s$ random plot effects, $\mathbf{a}_i = [a_{i1} \ a_{i2} \ \dots \ a_{is}]'$ is the vector of additive genetic values, $\mathbf{e}_i = [e_{i1} \ e_{i2} \ \dots \ e_{is}]'$ is the vector of random residuals, and \mathbf{X}_i , \mathbf{Z}_{Bi} , \mathbf{Z}_{Pi} and \mathbf{Z}_{Ti} are the incidence matrices for fixed, block, plot and additive genetic effects respectively. If we think of a non-inbred individual with all measurements (i.e. no missing observations) the dispersion matrices are:

$$\text{Var}[\mathbf{b}] = \mathbf{B}_0, \text{Var}[\mathbf{p}] = \mathbf{P}_0, \text{Var}[\mathbf{a}_i] = \mathbf{G}_0 \text{ and } \text{Var}[\mathbf{e}_i] = \mathbf{R}_0$$

with typical elements σ_{bjk} , σ_{pjk} , σ_{ajk} , σ_{ejk} respectively. For individuals with missing observations (and the vector \mathbf{y}_i reduced accordingly) the corresponding rows and columns from \mathbf{R}_0 are omitted.

The multivariate model equation for the N individuals in the progeny test is:

$$\mathbf{y} = \mathbf{X} \mathbf{m} + \mathbf{Z}_B \mathbf{b} + \mathbf{Z}_P \mathbf{p} + \mathbf{Z}_T \mathbf{a} + \mathbf{e}$$

where $\mathbf{y} = [y_1' \ y_2' \ \dots \ y_N']'$, $\mathbf{a} = [a_1' \ a_2' \ \dots \ a_N']'$ and $\mathbf{e} = [e_1' \ e_2' \ \dots \ e_N']'$. In addition $\mathbf{X} = [\mathbf{X}_1' \ \mathbf{X}_2' \ \dots \ \mathbf{X}_N']'$,

$\mathbf{Z}_B = [\mathbf{Z}_{B1}' \ \mathbf{Z}_{B2}' \ \dots \ \mathbf{Z}_{BN}]'$, $\mathbf{Z}_P = [\mathbf{Z}_{P1}' \ \mathbf{Z}_{P2}' \ \dots \ \mathbf{Z}_{PN}]'$ and $\mathbf{Z}_T = \Sigma_{\oplus} \mathbf{Z}_{T_i}$, where Σ_{\oplus} denotes direct sum (SEARLE, 1982).

The expected value and dispersion matrices considering a multivariate normal distribution and zero covariance between random factors (blocks, plots and trees) are:

$$E[\mathbf{y}] = \mathbf{X} \mathbf{m}$$

$$\text{Var}[\mathbf{b}] = \mathbf{B} = \Sigma_{\oplus} \mathbf{B}_0, \text{Var}[\mathbf{p}] = \mathbf{P} = \Sigma_{\oplus} \mathbf{P}_0, \text{Var}[\mathbf{a}] = \mathbf{G} = \mathbf{A}_N \otimes \mathbf{G}_0 \text{ and } \text{Var}[\mathbf{e}] = \mathbf{R} = \Sigma_{\oplus} \mathbf{R}_0$$

$$\text{thus } \text{Var}[\mathbf{y}] = \mathbf{Z}_B \mathbf{B} \mathbf{Z}_B' + \mathbf{Z}_P \mathbf{P} \mathbf{Z}_P' + \mathbf{Z}_T \mathbf{G} \mathbf{Z}_T' + \mathbf{R}$$

where \mathbf{A}_N is the numerator relationship matrix (HENDERSON, 1984) and \otimes denotes direct product (SEARLE, 1982).

Best Linear Unbiased Prediction of the breeding values (\mathbf{a}) were calculated using HENDERSON's mixed model equations (HENDERSON, 1984) and assumed values of covariance components. Values for all covariance components were obtained by Restricted Maximum Likelihood (REML) using ASReml (GILMOUR et al., 1998).

APIOLAZA et al. (2000) analysed the dataset comparing five models for describing the additive genetic covariance matrix \mathbf{G}_0 : unstructured (US), autoregressive with time in a natural logarithm time scale (ARlog), banded correlations (BC), random regressions (RR), and uncorrelated (UC). The UC model is equivalent to independent univariate analyses and does not provide direct estimates for genetic correlations between ages (i.e. assuming zero between-trait genetic correlations). The RR model did not converge; thus its estimates of genetic parameters are not reliable. Therefore we calculated response to selection from only three models: US, BC and ARlog.

Expressing $\mathbf{G}_0 = \mathbf{S} \mathbf{C} \mathbf{S}$, where \mathbf{S} is a diagonal matrix with elements equal to the standard deviations for each assessment and \mathbf{C} the correlation matrix between assessments, these covar-

iance models can be seen as using \mathbf{C} with different sets of restrictions (see *Figure 1*). In the US model \mathbf{C}_{US} has no restrictions except for being positive definite. The BC model assumes similar correlations for measurements with equal time between expressions, creating bands of identical correlations in \mathbf{C}_{BC} . The autoregressive model assumes that in $\mathbf{C}_{\text{ARlog}}$ the correlation between two assessments at ages j and k has form $r^{|\log(k)-\log(j)|} = r^{|\log(k/j)|}$. Additive genetic and phenotypic covariance parameters for these three models are presented in *table 1* (adapted from APIOLAZA et al., 2000).

ARlog was selected as the best model from a penalised-likelihood standpoint (i.e. the log-likelihood value LogL penalised, in this case, according to the number p of independent parameters of the model), using Akaike's Information Criterion (AIC, WADA and KASHIWAGI, 1990):

$$\text{AIC} = -2 \text{LogL} + 2 p$$

The best model using this criterion has the lowest value of AIC. A detailed description of the models and model selection criteria can be found in APIOLAZA and GARRICK (2001).

Predicted response to selection

Consider the breeding objective (H) or aggregate genotype as a linear combination of n additive genetic values $\mathbf{a} = [a_1 a_2 \dots a_n]'$ weighted by their relative economic importance $\mathbf{v}' = [v_1 v_2 \dots v_n]$ where all elements of \mathbf{v} are different from zero:

$$H = \mathbf{v}'\mathbf{a}$$

The selection index used to predict the aggregate genotype is:

$$I = \mathbf{c}'\mathbf{y}^*$$

where $\mathbf{c}' = [c_1 c_2 \dots c_m]$ is the vector of index weights, calculated using economic and genetic information to maximise the correlation between H and I , and $\mathbf{y}^* = [y_1^* y_2^* \dots y_m^*]'$ is the vector of phenotypic assessments on the trees adjusted for fixed effects (i.e. $\mathbf{y} - \mathbf{X}\mathbf{m}$ for BLP or, in our case, $\mathbf{y} - \mathbf{X}\hat{\mathbf{m}}$ for BLUP). Index weights are calculated using (HAZEL, 1943):

$$\mathbf{c} = \mathbf{P}^{-1}\mathbf{G}\mathbf{v}$$

where \mathbf{P} and \mathbf{G} represent the phenotypic and additive genetic covariance matrices for the traits. These weights maximise the accuracy of selection, i.e. the correlation between I and H (r_{IH}) when the fixed effects, \mathbf{P} and \mathbf{G} are known.

Predicted response to selection (ΔG), considering one generation, is a function of accuracy of selection (r_{IH}), the variance of the selection target (σ_H^2) and the intensity of selection (i , related to the proportion of trees selected):

$$\Delta G = i r_{IH} \sigma_H$$

H can be partitioned in such a way that $H = a_1 I_1 + a_2 I_2 + \dots + a_n I_n$, where each index estimates the breeding value for a different trait. VILLANUEVA et al. (1993) extended this concept partitioning matrices \mathbf{P} and \mathbf{G} to facilitate generalisation of the estimation process, where each submatrix corresponds to a trait. The more general case, for m selection criteria (characters and/or measurements with typical elements j and k) and n traits in the aggregate genotype (with typical element q) is:

Table 1. – Genetic parameters for unstructured (US), banded correlations (BC) and autoregressive with time on a natural logarithm scale (ARlog) models. Additive genetic (σ_a^2) and phenotypic variances (σ_y^2), additive genetic covariances (σ_{ajk} , above diagonal) and phenotypic covariances (σ_{yjk} , below diagonal).

Age	σ_a^2	σ_p^2	Age (years)										
			1	2	4	5	6	7	8	9	12	15	
US													
1	0,001	0,016		0,002	0,007	0,011	0,016	0,019	0,019	0,018	0,018	0,024	
2	0,004	0,051	0,022		0,016	0,023	0,029	0,032	0,033	0,026	0,020	0,027	
4	0,070	0,379	0,054	0,113		0,093	0,120	0,132	0,136	0,118	0,105	0,139	
5	0,132	0,752	0,071	0,147	0,473		0,169	0,194	0,193	0,178	0,166	0,213	
6	0,227	1,105	0,080	0,167	0,548	0,837		0,264	0,273	0,266	0,253	0,339	
7	0,315	1,306	0,083	0,170	0,563	0,863	1,127		0,330	0,333	0,358	0,463	
8	0,349	1,570	0,083	0,173	0,579	0,887	1,170	1,331		0,349	0,394	0,511	
9	0,367	2,014	0,089	0,183	0,616	0,946	1,270	1,464	1,633		0,444	0,581	
12	0,580	3,402	0,102	0,199	0,678	1,065	1,439	1,678	1,924	2,241		0,764	
15	0,936	6,968	0,130	0,256	0,878	1,399	1,893	2,224	2,597	3,014	4,354		
BC													
1	0,001	0,016		0,002	0,007	0,008	0,010	0,010	0,010	0,009	0,008	0,005	
2	0,004	0,052	0,022		0,018	0,023	0,028	0,029	0,028	0,025	0,020	0,015	
4	0,080	0,382	0,054	0,114		0,105	0,130	0,138	0,138	0,127	0,115	0,103	
5	0,142	0,759	0,071	0,148	0,478		0,175	0,190	0,193	0,178	0,160	0,135	
6	0,223	1,112	0,080	0,168	0,553	0,844		0,241	0,251	0,234	0,214	0,207	
7	0,268	1,311	0,082	0,171	0,569	0,871	1,134		0,279	0,266	0,256	0,249	
8	0,298	1,572	0,083	0,174	0,585	0,894	1,176	1,335		0,284	0,284	0,273	
9	0,278	2,008	0,089	0,185	0,623	0,954	1,274	1,464	1,632		0,288	0,281	
12	0,337	3,368	0,101	0,201	0,685	1,073	1,443	1,673	1,916	2,223		0,373	
15	0,468	6,900	0,129	0,256	0,877	1,393	1,880	2,197	2,564	2,970	4,296		
Arlog													
1	0,002	0,016		0,002	0,008	0,012	0,015	0,018	0,019	0,019	0,023	0,028	
2	0,004	0,051	0,022		0,015	0,021	0,028	0,032	0,034	0,035	0,041	0,050	
4	0,064	0,380	0,055	0,115		0,088	0,117	0,133	0,141	0,148	0,174	0,211	
5	0,126	0,754	0,072	0,149	0,474		0,168	0,192	0,203	0,212	0,250	0,304	
6	0,231	1,108	0,082	0,170	0,549	0,839		0,264	0,280	0,293	0,344	0,419	
7	0,312	1,309	0,084	0,173	0,565	0,865	1,129		0,331	0,345	0,406	0,494	
8	0,360	1,573	0,084	0,176	0,583	0,891	1,173	1,333		0,376	0,443	0,538	
9	0,402	2,016	0,090	0,187	0,622	0,953	1,274	1,465	1,636		0,473	0,575	
12	0,590	3,399	0,105	0,208	0,702	1,097	1,476	1,709	1,954	2,269		0,718	
15	0,913	6,959	0,134	0,266	0,906	1,435	1,933	2,258	2,630	3,046	4,366		

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} & \cdots & \mathbf{P}_{1m} \\ \mathbf{P}_{21} & \mathbf{P}_{22} & \cdots & \mathbf{P}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{m1} & \mathbf{P}_{m2} & \cdots & \mathbf{P}_{mm} \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_{11} & \mathbf{g}_{12} & \cdots & \mathbf{g}_{1n} \\ \mathbf{g}_{21} & \mathbf{g}_{22} & \cdots & \mathbf{g}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{g}_{m1} & \mathbf{g}_{m2} & \cdots & \mathbf{g}_{mn} \end{bmatrix}$$

\mathbf{P}_{jk} refers to the phenotypic covariance matrix between selection criteria j and k , and \mathbf{g}_{jq} represents the vector of covariances between the selection criteria j and breeding value for trait q . Breeding values on H can be estimated with different selection indices (selection criteria) representing different selection schemes:

- Mass selection, considering only the own record of individual i for trait j (y_{ij}^*), representing the simplest index:

$$I_{\text{mass}_j} = c_{1j} y_{ij}^*$$

- Combined selection includes individual and family-average information for trait j , whereby:

$$I_{\text{comb}_j} = c_{1j} y_{ij}^* + c_{2j} \bar{y}_{\cdot,j}^*$$

where: $\bar{y}_{\cdot,j}^*$ is the phenotypic average (adjusted for fixed effects) for the half-sib family, including the assessment on y_{ij}^* , on trait/measurement j , and c_{1j} and c_{2j} are index coefficients for individual and family information respectively.

This notation is fairly general, allowing for any number of selection criteria and traits in the breeding objective, and can be readily extended to other types of relatives (see KERR, 1998, for examples based on full-sib mating over a number of generations). Consider now an objective including a single breeding value of the individual (e.g. height at 15 years) and m phenotypic assessments at earlier ages (e.g. height at age ≤ 15 years). \mathbf{P} and \mathbf{G} for a given selection method contain $m \times m$ submatrices and m subvectors respectively. The submatrices and subvectors are:

- Mass selection (in this case \mathbf{P}_{jk} and \mathbf{g}_{jq} have dimension 1×1 , i.e. they are scalars):

$$\mathbf{P}_{jk} = \text{Cov}(y_{ij}^*, y_{ik}^*) = \sigma_{y_{jk}}$$

$$\mathbf{g}_{jq} = \text{Cov}(y_{ij}^*, a_{iq}) = \text{Cov}(a_{ij}, a_{iq}) = \sigma_{a_{jq}}$$

- Combined selection (individual - y_{ij}^* - and average of half-sib family - $\bar{y}_{\cdot,j}^*$ - with t individuals):

$$\mathbf{P}_{jk} = \begin{bmatrix} \text{Cov}(y_{ij}^*, y_{ik}^*) & \text{Cov}(y_{ij}^*, \bar{y}_{\cdot,k}^*) \\ \text{Cov}(\bar{y}_{\cdot,j}^*, y_{ik}^*) & \text{Cov}(\bar{y}_{\cdot,j}^*, \bar{y}_{\cdot,k}^*) \end{bmatrix}$$

$$\mathbf{g}_{jq} = \begin{bmatrix} \text{Cov}(y_{ij}^*, a_{iq}) \\ \text{Cov}(\bar{y}_{\cdot,j}^*, a_{iq}) \end{bmatrix}$$

where:

$$\text{Cov}(\bar{y}_{\cdot,j}^*, y_{ik}^*) = \text{Cov}(y_{ij}^*, \bar{y}_{\cdot,k}^*) = \text{Cov}(\bar{y}_{\cdot,j}^*, \bar{y}_{\cdot,k}^*) = (\sigma_{p_{jk}} + 0.25(t-1)\sigma_{a_{jk}})/t$$

$$\text{Cov}(y_{ij}^*, a_{iq}) = \sigma_{a_{jq}}$$

$$\text{Cov}(\bar{y}_{\cdot,j}^*, a_{iq}) = \frac{\sigma_{a_{jq}}}{2}$$

and the phenotypic covariance is:

$$\sigma_{y_{jk}} = \sigma_{p_{jk}} + \sigma_{p_{jk}} + \sigma_{a_{jk}} + \sigma_{e_{jk}}$$

Further generalisation to an objective with n traits would involve $m \times n$ subvectors to define \mathbf{G} .

Number and timing of measurements

We consider from one to three assessments only for progeny tests, owing to economic and practical considerations, to predict height performance at 15 years (then $\mathbf{v} = 1$, a scalar). Thus \mathbf{P} contains between 1×1 and 3×3 submatrices, \mathbf{G} contains between 1 and 3 subvectors, and the breeding objective considers only 1 trait. We generate all combinations for 1, 2 and 3 measurements out of the 10 assessments used by APIOLAZA et al. (2000) and calculate predicted response to selection for each combination. We do not interpolate intermediate ages of assessments, but use only years actually assessed. For all calculations we assume 200 families and 40 trees per family using forwards selection, i.e. selection of the progeny rather than of the parents. The selection is 200 out of 8000, for an intensity i of 2.338 (FALCONER and MACKAY, 1996).

The best option is chosen based on response per unit of time, i.e. response divided by generation interval, ignoring measurement costs. Generation interval is calculated as selection delay (age of the latest measurement) plus breeding delay (time between selection and propagation of sufficient offspring for planting). Two levels for breeding delay are considered: 5 and 8 years.

Results and Discussion

APIOLAZA et al. (2000) determined that, based upon AIC value, the ARlog model was the best for the data set analysed. Nevertheless, they also pointed out that small differences in statistical model selection criterion could conceal large differences in genetic parameters. Given that the dataset available comprises only one generation, it is not possible to compare the covariance structures in terms of empirical gain, but only in terms of predicted response to selection based upon estimates of genetic parameters. Consequently, at each time one of the models was assumed as the 'true' one and response to selection calculated accordingly. Because of this, results are not directly comparable across covariance models.

The pattern of predicted response to selection, considering any number of measurements, vary widely among covariance structures. Both US and ARlog models achieve similar maximum gain per year, but with three years of difference in timing (4 vs 7 years, Figure 2). The ARlog model consistently achieves higher gains when selecting under age 9 years. Predicted response for the ARlog model tends to be dominated by the high level of autocorrelation, while in the US model seems to follow the trend for heritability of height. The predicted response to selection of the US model based on early measurements seems to fluctuate more erratically. APIOLAZA et al. (2000), suggest the use of a much larger experiment to obtain more reliable estimates of the genetic correlations. Predicted response from the BC model is far lower than with the other two models but follows a trend similar to the ARlog model. In the former model heritability and genetic correlation estimates are most of the time smaller than in the ARlog and US models.

The trends of the effects of additional measurements and extra information from relatives were similar for all covariance models. We will use the ARlog model for illustration purposes. The effect of integrating information from relatives in the genetic evaluation is far greater than that of including extra measurements when predicting the breeding values (see Figure 3),

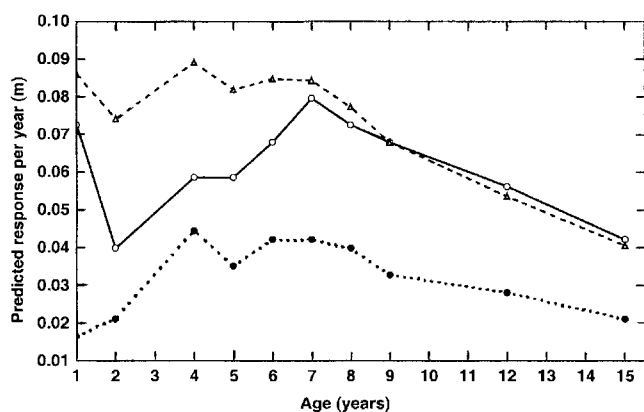


Figure 2. – Predicted response per year to forward mass selection for different covariance structures: unstructured (US, —○—), banded correlations (BC, ····) and autoregressive with time on a natural logarithm scale (ARlog, ---▲---) considering 5 years breeding delay.

with little or no extra cost (except those related to the use of more sophisticated analyses). Results for multiple assessments presented in figure 3 correspond to the best combination of two ages lower or equal to the age reported (Table 2). From age 1 to 8 years using an extra measurement increases predicted response between 2% and 5%, and only from age 9 onwards

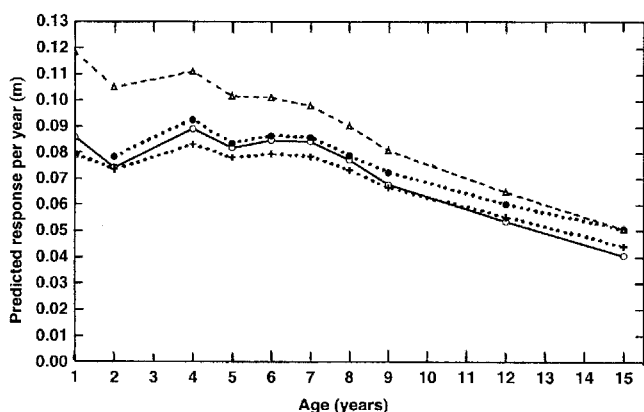


Figure 3. – Predicted response per year to forward selection for the autoregressive model with time on a natural logarithm scale (ARlog) considering: single assessment-mass selection (—○—), two assessments-mass selection (····) and single assessment-combined selection (---▲---) for 5 years of breeding delay. Response considering single assessment-combined selection for 8 years of breeding delay (···+···).

Table 2. – Predicted response to selection and best combinations of assessments for the autoregressive model with time on a natural logarithm scale (ARlog), for mass selection with 1 and 2 measurements.

Age of selection (years)	Mass selection, 1 measurement		Mass selection, 2 measurements	
	Response per year (m)	Response per year (m)	Response per year (m)	Combination of measurements (ages)
1	0.086			
2	0.074	0.078		[1 2]
4	0.089	0.092		[2 4]
5	0.082	0.083		[4 5]
6	0.085	0.086		[2 6]
7	0.084	0.086		[2 7]
8	0.077	0.079		[7 8]
9	0.068	0.072		[7 9]
12	0.054	0.060		[7 12]
15	0.040	0.051		[7 15]

predicted response increases from 7% at 9 years to 25% at 15 years (Figure 3). A similar trend is observed for combined selection with 2 measurements (results not presented). In spite of this, the optimum selection time does not change from age 4 years, when the additional predicted response is 4%. The results from including a third measurement are not presented, because the gain is marginal.

The use of family information increased predicted response to forwards selection by 16% to 41%, especially at early ages. Additionally, the optimum selection times reduces from 4 years to 1 year. The reduction of selection age also applies to the US model, where optimum selection is at one year. Selection time is not affected in the BC model.

The only case when combined selection is inferior to mass selection using two assessments (at ages 7 and 15) is for predicted response selecting at age 15. This is caused by the low accuracy at age 15 (low heritability) compared to an index that integrates information from that year with information from age 7 (the age of the highest heritability).

When considering selection at very early ages (e.g. 1 year) a breeding delay of 5 years could be far too optimistic, given current biological constraints. Breeding delay includes the time needed for flowering, the delay between flowering and seed production, and time for multiplication. While the last two are independent from selection age, the first one is probably interdependent with age during the first 5 years. Therefore, the use of a uniform breeding delay for all selection ages should be considered as a simplifying assumption. If the first component of breeding delay is addressed with more detail selection ages would tend to be pushed forward. As expected, the effect of increasing breeding delay (from 5 to 8) is larger for early selection ages than for late selections (Figure 3). The reduction of predicted per-year response resulting from the increased breeding delay, using single-assessment combined selection, ranges from 33% at age 1 to 13% at age 15 (Figure 3). This trend of reductions is close to linear ($R^2 = 0.95$) and very similar for all covariance structures.

Final Considerations

Although including more assessments increases predicted response to selection, especially after age 8, the extra response does not match the gain attained using a single assessment with combined selection. Nevertheless, it is expected that using multiple assessments could be advisable for datasets with lower genetic autocorrelations or strong age-age environmental correlations. A decision on single versus multiple measurements for selection should take into account the gain in response (weighted by the number of hectares deployed with material from the breeding program) versus the costs of extra measurements. Three further considerations are: that several measurements might reduce optimal generation interval (increasing accuracy at a given age and ascertaining the stability of rankings, making earlier selection more appealing), that selection for seed orchards can be continually updated and could make use of additional measurements (e.g. last three combinations of Table 2), and that measuring costs often increase with age (e.g. NEWMAN and WILLIAMS, 1991).

We anticipate that reducing breeding delay (through overcoming biological constraints upon age of flowering) would drive optimum selection to earlier ages, because the denominator of response per unit of time would be dominated by selection age. Reducing the generation interval from 9 years (i.e. selection at 4 years) to 6 years (i.e. selecting at 1 year) most probably will affect profit when considering net discounted value.

It seems that it is still possible in tree breeding to obtain additional gains through the use of more sophisticated methods of genetic evaluation, without resorting to extra assessments. Dealing with more than one generation and/or more complex crossing designs (e.g. controlled pollination) will imply using selection criteria including the estimated breeding values of the parents. However if selection operates through several generations other factors must be considered: gametic linkage phase disequilibrium (BULMER effect), genetic drift, mutational variance and effective population size (WEI et al., 1996). An extra advantage of considering data over several generations is the opportunity to determine the best covariance model on terms of realised response to selection, rather than on predicted values.

An important aspect, but beyond the scope of this paper, is the consideration of the risk involved in early selection. Deviations from predicted gain (either overestimation or underestimation) can potentially alter both the selection age(s) and the economic results of a breeding program; and the variance of the response should be taken into account. Risk may arise, among other reasons, because of low accuracy of prediction, traits not being expressed at selection age (e.g. effect of *Cyclaneusma* needle cast, cf. BURDON, 1989), differences between performance at final assessment (15 years) and at rotation age (20 years), and the effect of faster reduction of effective population size due to more frequent generation turnover. Several approaches to deal with risk have been proposed in tree and animal breeding literature including simulation of predicted gain using stochastic sampling of genetic correlations (e.g. NEWMAN and WILLIAMS, 1991; MAGNUSSEN and YANCHUK, 1993), quadratic programming (SCHNEEBERGER et al., 1982) and Bayesian decision theory (e.g. WOOLLIAMS and MEUWISSEN, 1993). A comprehensive risk analysis will probably need to consider the effects of early selection on variance of predicted response across several generations.

It is appropriate to emphasise that the results presented in this research relate to a small number of families growing in one site. Therefore they should not be considered as the 'standard' results for radiata pine. It will be necessary to extend the analyses of longitudinal data to datasets including multiple sites and more families, to be sure of the reliability of estimates of genetic parameters. Accordingly, this study provides an illustration of methodology rather than definitive guidelines for early selection.

Acknowledgements

LUIS APIOLAZA was funded with New Zealand Overseas Development Assistance and New Zealand Forest Research Institute scholarships during the development of this study. Thanks are due to CLAUDIO BALOCCHI and the breeding team of Bosques Arauco S.A., who kindly provided the dataset used in this study.

References

APIOLAZA, L. A. and GARRICK, D. J.: Analysis of longitudinal data from progeny tests: some multivariate approaches. *For. Sci.*, in press (2001). — APIOLAZA, L. A., GILMOUR, A. R. and GARRICK, D. J.: Variance modell-

ing of longitudinal height data from a *Pinus radiata* progeny test. *Can. J. For. Res.* **30**: 645–654 (2000). — BORRALHO, N. M. G.: The impact of individual tree mixed models (BLUP) in tree breeding strategies. P. 141–145. *In*: POTTS, B. M., BORRALHO, N. M. G., REID, J. B., CROMER, R. N., TIBBITS, W. N. and RAYMOND, C. A.: "Eucalypts plantations: improving fibre yield and quality". Proceedings of CRC-IUFRO Conference, 19–24 February, Hobart, Tasmania, Australia (1995). — BURDON, R. D.: Early selection in tree breeding: principles for applying index selection and inferring input parameters. *Can. J. For. Res.* **19**: 499–504 (1989). — BURDON, R. D., BANNISTER, M. H. and LOW, C. B.: Genetic survey of *Pinus radiata*. 5: Between-trait and age-age correlations for growth rate, morphology, and disease resistance. *New Zealand J. For. Sci.* **22**: 211–227 (1992). — FALCONER, D. S. and MACKAY, T. F. C.: Introduction to quantitative genetics. Longman Group Ltd., UK (1996). — GILMOUR, A. R., THOMPSON, R. and CULLIS, B. R.: ASReml user's manual. New South Wales Agriculture, Orange, Australia (1998). — GWAZE, D. P., WOOLLIAMS, J. A. and KANOWSKI, P. J.: Optimum selection age for height in *Pinus taeda* L. in Zimbabwe. *Silvae Genet.* **46**: 358–365 (1997). — HATCHER, A. V., BRIDGWATER, F. E. and WEIR, R. J.: Performance level – standardized score for progeny test performance. *Silvae Genet.* **30**: 184–187 (1981). — HAZEL, L. N.: The genetic basis for constructing selection indexes. *Genetics* **28**: 476–490 (1943). — HENDERSON, C. R.: Applications of linear models in animal breeding. University of Guelph Press, Guelph (1984). — JARVIS, S. F., BORRALHO, N. M. G. and POTTS, B. M.: Implementation of a multivariate BLUP model for genetic evaluation. P. 212–216. *In*: POTTS, B. M., BORRALHO, N. M. G., REID, J. B., CROMER, R. N., TIBBITS, W. N. and RAYMOND, C. A.: "Eucalypts plantations: improving fibre yield and quality". Proceedings of CRC-IUFRO Conference, 19 to 24 February, Hobart, Tasmania, Australia (1995). — KERR, R. J.: Asymptotic rates of response from forest tree breeding strategies using best linear unbiased prediction. *Theor. Appl. Genet.* **96**: 484–493 (1998). — LAMBETH, C. C.: Juvenile-mature correlations in Pinaceae and implications for early selection. *For. Sci.* **26**: 571–580 (1980). — MAGNUSSEN, S.: Minimum age-to-age correlations in early selections. *For. Sci.* **34**: 928–938 (1988). — MAGNUSSEN, S.: A continuous-time Markov chain for early selection. *For. Sci.* **39**: 845–850 (1993). — MAGNUSSEN, S. and YANCHUK, A. D.: Selection age and risk: finding the compromise. *Silvae Genet.* **42**: 25–40 (1993). — NEWMAN, D. H. and WILLIAMS, C. G.: The incorporation of risk in optimal selection age determination. *For. Sci.* **37**: 1350–1364 (1991). — SCHNEEBERGER, M., FREEMAN, A. E. and BOEHLJE, M. D.: Application of portfolio theory to dairy sire selection. *J. Dairy Sci.* **65**: 404–409 (1982). — SEARLE, S. R.: The value of indirect selection: I. Mass selection. *Biometrics* **21**: 682–707 (1965). — SEARLE, S. R.: Matrix algebra useful for statistics. John Wiley & Sons, New York (1982). — VILLANUEVA, B., WRAY, N. R. and THOMPSON, R.: Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. *Anim. Prod.* **57**: 1–13 (1993). — WADA, Y. and KASHIWAGI, N.: Selecting statistical models with information statistics. *J. Dairy Sci.* **73**: 3575–3582 (1990). — WEI, M., CABALLERO, A. and HILL, W. G.: Selection response in finite populations. *Genetics* **144**: 1961–1974 (1996). — WEI, X. and BORRALHO, N. M. G.: A simple model to describe age trends in heritability in short rotation tree species. P. 178–181. *In*: DIETERS, M. J., MATHESON, A. C., NIKLES, D. G., HARWOOD, C. E. and WALKER, S. M. (Ed.): Proceedings of Tree Improvement for Sustainable Tropical Forestry, 27 October to 1 November, Caloundra, Queensland, Australia (1996). — WHITE, T. L.: A conceptual framework for tree improvement programs. *New For.* **1**: 325–342 (1987). — WHITE, T. L., HODGE, G. R. and DELORENZO, M. A.: Best linear prediction of breeding values in forest tree improvement. P. 99–122. *In*: Proceedings of the 1986 Workshop on Statistical Considerations in Genetic Testing. University of Florida, Gainesville, Florida (1987). — WHITE, T. L. and HODGE, G. R.: Best linear prediction of breeding values in a forest tree improvement-program. *Theor. Appl. Genet.* **76**: 719–727 (1988). — WOOLLIAMS, J. A. and MEUWISSEN, T. H. E.: Decision rules and variance of response in breeding schemes. *Anim. Prod.* **56**: 179–186 (1993). — ZOBEL, B. J. and TALBERT, J. T.: Applied forest tree improvement. John Wiley & Sons, New York (1984).