Genetic Correlations between Environments with Genetic Groups Missing in Some Environments

By R. D. Burdon

Ministry of Forestry, Private Bag 3020, Rotorua,
New Zealand

Summary

Genetic correlations between environments are a valuable way of studying genotype-environment interactions, but missing or unbalanced data can cause problems in estimating the correlations. It is suggested that the input statistics from each environment that will usually offer the most convenient and reliable correlation estimates are (i) genetic group means and (ii) calculated error variance of estimating group means based on all groups represented. Situations where this proposal offers advantages are discussed.

Key words: genotype-environment interaction, genetic correlation, missing data.

Introduction

Genotype-environment interaction can be studied by estimating genetic correlations between environments (BURDON, 1977) for the trait in question (e.g. height growth). The interaction is manifested in a combination of (a) departures of between-environment genetic correlations from unity and (b) differences between environments in the extent to which genetic differences are expressed, the former usually being of prime interest.

As BURDON (1977) noted, a genetic correlation (r_{gij}) between a pair of environments (e.g. sites) i and j, can be estimated as:

either $\quad r_{gij} = r_{pij}/(h_i^- h_j^-)$ $\qquad$ (1)

or $\qquad r_{gij} = \hat{c}\hat{o}v_{gij}/(\hat{\sigma}_{gi}\,\hat{\sigma}_{gj})$ $\qquad$ (2)

where: $\quad r_{pij} =$ phenotypic correlation between means of genetic groups in i and j respectively

$h_i^2{}^-$ and $h_j^2{}^- =$ heritabilities of group means in the respective environments.

$cov_{gij} =$ covariance for group performance between two environments which, assuming no common maternal or environmental effects, is estimated directly by mean cross-products of group means

$\sigma^2{}_{gi}$ and $\sigma^2{}_{gj} =$ between-group variances in the respective environments

and where groups may represent clones, full-sib families, half-sib families, etc., provided the groups all belong to one such category.

Such genetic correlation estimates (denoted Type B by BURDON (1977) are necessarily based on information from different individuals (ramets or seedlings) within groups for the respective 'traits'. In this context the 'traits' represent the performances for the conventional trait in the respective environments.

In practice, the analysis will often be complicated by unbalanced classifications in the data. WELLER and RON (1987), addressing Type A correlation estimates (in which both traits of a pair are normally measured on each

individual), demonstrated the sensitivity of correlation estimates to missing data. They concluded that the most reliable estimates come from data subsets representing only those individuals without missing data, namely, the individuals with information on both traits of a pair.

With Type B correlation estimates, missing or unbalanced data poses its own problems. One important situation is where gaps in the data specifically entail some missing genetic groups in certain environments, such that different subsets of groups will be common to different pairs of environments. The only way of basing all the input statistics specifically on groups common to any particular pair of environments would be to carry out separate ANOVAS on every different subset of groups that is common to a pair of environments. This is cumbersome and, as discussed later, may still sacrifice good information.

The Proposal

To avoid the above problems it is proposed that the following basic statistics may be used from each environment:

(i) means of individual groups (as already required for Equations 1 and 2)

(ii) error variance of estimating group means.

These would then be used as follows:

$r_{gij} = \hat{C}\hat{o}v_{gij}\,([SSi/(k-1) - \hat{\sigma}^2\varepsilon_i]\,[SSj/(k-1) - \hat{\sigma}^2\varepsilon_j])^{-1/2}$ $\quad$ (3)

where: $\quad$ k is the number of groups common to the two environments

SSi and SSj are the corrected sums of squares for the inputted means of the k groups in the respective environments

$\sigma^2\varepsilon_i$ and $\sigma^2\varepsilon_j =$ error variances of estimating group means in the respective environments.

An estimate of $\sigma^2\varepsilon$ within a particular environment can be obtained as

$\hat{\sigma}^2\varepsilon =$ (groups mean square)/n $- \hat{\sigma}^2{}_g$ $\qquad$ (4)

where: $\quad$ n = coefficient for between-groups variance in the expected mean square for groups in the analysis of variance within the environment.

This formulation has the advantage of not depending on the existence of a straightforward F-test for group differences.

Discussion

Equation 3 is formally similar to Equation 2, with $SSi/(k-1) - \hat{\sigma}\varepsilon_i$ and $SSj/(k-1) - \hat{\sigma}^2\varepsilon_j$ being alternative estimates of $\sigma^2{}_{gi}$ and $\sigma^2{}_{gj}$ respectively. The equations become fully equivalent only when the classification is completely balanced between environments. The essential difference is that $\hat{\sigma}^2\varepsilon_j$ and $\hat{\sigma}^2\varepsilon_i$ in Equation 3 are based on all groups represented in the respective environments, unlike the corresponding values implied in Equation 2 which are

based on just the k groups in common. One can expect $\hat{\sigma}^2\varepsilon_i$ and $\hat{\sigma}^2\varepsilon_j$, being based on many more degrees of freedom, to be subject to much less bias or sampling errors resulting from differences between environments in samples of groups than would $\hat{h}_i^2$ and $\hat{h}_j^2$ or conventional estimates of $\sigma^2{}_{gi}$ and $\sigma^2{}_{gj}$. Indeed, estimates based on more than k groups are likely to be superior in many situations. Equation 3 therefore offers a method that can give convenient and robust correlation estimates while making full use of available information.

Even without missing or unbalanced data Equation 3 can be convenient to use. This is particularly where one wants to explore between-environment correlations for alternative subsets of genetic groups, since genetic parameters are meaningful only in relation to specified populations. This would, however, depend strongly on $\sigma^2\varepsilon_i$ and $\sigma^2\varepsilon_j$ being constant among subsets.

Equation 3 also has the advantage of requiring exchange of the minimum necessary information between organisations involved in collaborative studies of the genetic correlations.

There will be some situations where using Equation 3 may not be ideal, not that alternative approaches would necessarily be better. Obvious difficulties could arise where the presence or absence of groups in different environments reflects differential truncation selection between the environments. Any truncation selection tends to lead to poorer estimates of genetic correlations. A simulation study by van Vleck (1986) indicated that the problem is one of reduced precision rather than inherent bias; truncation selection was associated with an increased incidence of r_g estimates outside the bounds $+1$ and -1, or even imaginary estimates (resulting from $\hat{\sigma}^2{}_{gi}$ and/or $\hat{\sigma}^2{}_{gj} < 0$). While van Vleck considered only Type A correlation estimates there is every reason to believe that his findings would apply equally to Type B estimates.

In such a situation, though, the use of Equations 1 and 2 would be equally affected, and even the study of interaction by variance component estimates would be subject to bias.

If, within individual environments, different groups are represented by markedly unbalanced numbers of individuals there will also be problems. For Equation 3 single overall values would then not apply to $\sigma^2\varepsilon_i$ and/or $\sigma^2\varepsilon_j$. The problems, however, would not be specific to the use of Equation 3.

If the representation of different groups, however unequal, is in consistent ratios between environments it should then be appropriate to use Equation 2, provided group means are weighted, in calculating the denominators of the equations, according to the numbers of individuals in which the respective groups are represented. With even more troublesome imbalance it would seem appropriate to use analysis of variance with environments as an main effect, and study interaction by the traditional approach of estimating variance components. That, however, might entail accepting that no satisfactory estimates can be obtained of the genetic correlations that reflect the more meaningful aspect of interaction.

Literature Cited

Burdon, R. D.: Genetic correlations as a concept for studying genotype-environment interaction in forest tree breeding. Silvae Genet. 26: 168—175 (1977). —— van Vleck, L. D.: Selection bias in estimation of the genetic correlation. Biometrics 24: 951—962 (1968). —— Weller, J. I. and Ron, M.: A note on the effect of observations with missing data on genetic correlation estimates. Theor. Appl. Genet. 74: 549—553 (1987).

The Estimation of Genetic Parameters for Growth and Stem-Form over 15 Years in a Diallel Cross of Sitka Spruce

By C. J. A. Samuel

Forestry Commission, Northern Research Station,
Roslin, Midlothian, EH25 9SY, Great Britain

Abstract

Regular measurements of total height from the first to the 15th year after planting were made on families comprising a full diallel cross among 7 parent trees of Sitka spruce (Picea sitchensis Bong., Carr.) planted on 2 test sites. Diameter at breast height was measured at intervals after the 8th year and a subjective straightness score was made after 14 years. Although analysis revealed significant proportions of additive variation for height at most assessments, non-additive effects were more highly significant and usually accounted for a greater proportion of variation. In contrast, diameter was under predominantly additive control and straightness appeared to be equally subject to both additive and non-additive effects. The results are discussed with reference to the

current breeding strategy for Sitka spruce in Britain which includes techniques for exploiting the type of genetic control of growth traits suggested by their results.

Key words: Picea sitchensis, diallel cross, variance components.

Introduction

Among the forest tree species of commercial importance in Britain, Sitka spruce (Picea sitchensis Bong. Carr.) is used in by far the greatest proportion; the species accounts for around 70% of current planting. During the last 20 years, the main breeding effort in this species has concentrated on phenotypic selection and half-sib progeny-testing; over this period more than 3500 selec-