Genotype by Environment Interaction of *Pinus radiata* in New Zealand.

T.M. McDonald[1] and L.A. Apiolaza[1]

[1] New Zealand School of Forestry, University of Canterbury,
Private Bag 4800, Christchurch 8140, New Zealand
E-mail: tmm39@student.canterbury.ac.nz

Abstract

In New Zealand, a formal tree improvement and breeding programme for *Pinus radiata* (D.Don) commenced in 1952. A countrywide series of progeny trials was progressively established on over sixty sites, and is managed by the Radiata Pine Breeding Company. Diameter at breast height data from the series was used to investigate genotype by environment interaction with a view to establishing the need for partitioning breeding and deployment efforts for *P. radiata*. More than 280,000 measurements made this study one of the largest for genotype by environment interaction ever done.

Bivariate analyses were conducted between all pairs of sites to determine genetic correlations between sites. This enabled construction of a proximity matrix by subtracting each correlation from unity. The process of constructing the matrix highlighted issues of connectedness between sites.

Multiple regression on resemblance matrices was carried out by regressing a number of environmental correlation matrices on the diameter at breast height correlation matrix. Genotype by environment interactions were found to be driven by total rainfall and extreme maximum and minimum temperatures.

In addition, a method from Graph Theory using proximity thresholds was utilised as a form of clustering. However, in this study, minimal grouping of sites was observed.

Future work, incorporating ASReml to conduct the univariate and bivariate analyses, is expected to produce clearer results than experienced in this pilot study.

Introduction

A formal tree improvement programme for *Pinus radiata* (D.Don) commenced in New Zealand in 1952 (Burdon *et al*., 1997; Shelbourne *et al*., 1986). The programme was originally limited in terms of sites and provenances, but it became progressively more comprehensive with the establishment of a countrywide series of provenance tests during the 1980s on a large number of sites. Since the inception of the programme, over 2650 trees have been selected (based on external traits such as diameter, branch size, and stem straightness) and progeny tested (Jayawickrama *et al*., 1997).

The improvement programme is currently managed by the Radiata Pine Breeding Company (RPBC) Ltd., which involves several major New Zealand and Australian forestry companies. The RPBC aims to provide superior radiata pine to its shareholders and customers in Australasia[1].

---

[1] http://www.rpbc.co.nz/ accessed 25th September 2007 at 10:45am

The forests owned/managed by the New Zealand-based RPBC shareholders encompass a vast number of heterogeneous climatic zones, stretching from subtropical in the far north (34° S) to cool temperate in the south (46° S), and form a combined plantation area of more than 800,000 hectares (NZFOA, 2008). The sites in this study are spread over approximately 1300 km and are situated at elevations between 20 m and 780 m above sea level. Annual rainfall varies with latitude, but is between 600 mm and 1600 mm per annum and snow falls are unusual in the coastal areas of New Zealand.

The intention in this study was to explain the genotype by environment interaction observed from the measurements made (mostly between the ages of 7-10 years old) on over sixty New Zealand sites (see Figure 1) and more than 2,500 distinct genotypes. In New Zealand, partial attempts have been made to explain the genotype by environment interactions of *P. radiata* (Burdon, 1977; Carson, 1991; Johnson & Burdon, 1990). However, these studies have been limited by the scale of data studied.
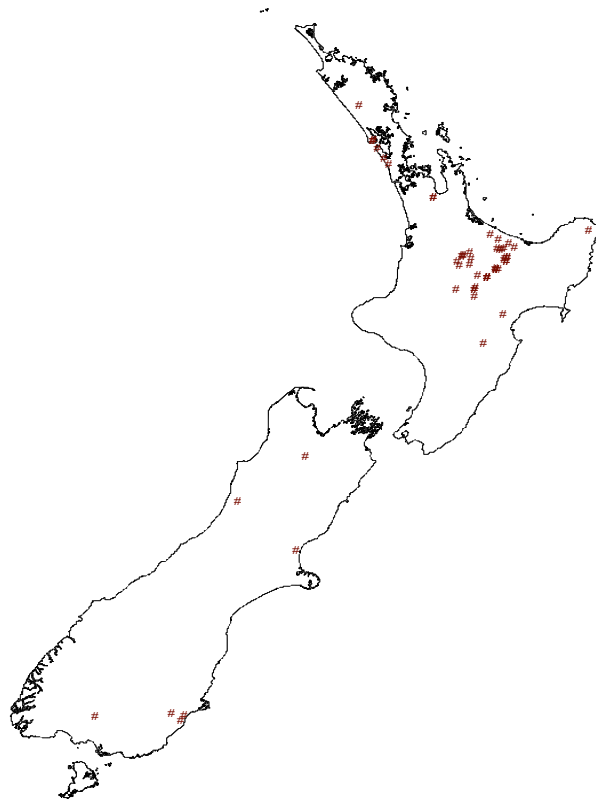
Figure 1 Location of RPBC progeny trials in New Zealand

By gaining greater insight into genotype by environment interactions of *P. radiata*, it is hoped that breeders will be able to more accurately predict the performance of genotypes across the range of New Zealand environments. Consequently, the probability of selecting the most suitable parent, or group of parents, for any site will be improved. Furthermore, an understanding of the likely performance of genotypes on a range of sites will enable a decision to be made regarding the breeding strategy for New Zealand. Is it to the forestry industry's advantage to create regional breeds or is it more efficient to produce a national breed from a central location and deploy it across the country?

Materials and Methods

In order to determine the environmental variables that are most influential on the genotype by environment interaction of *P. radiata* in New Zealand and to identify interacting sites and genotypes, the performance of genotypes at each site in terms of diameter at breast height (over bark at 1.4m above ground, dbh) was calculated.

Construction of the proximity matrix

First, each of the sixty trials was analysed in a univariate fashion to obtain family means. This was followed by bivariate analysis of all pairs of trials, taking into account differences for experimental and mating designs. All linear mixed model analyses were conducted using SAS.

Once the genetic correlations had been calculated for all pairs of sites, they were used to populate a proximity matrix showing the "distance" (the correlation subtracted from unity) between any two sites in terms of dbh.

Having quantified the genotype by environment interaction using these measures of proximity, a multi-dimensional scaling plot was used to gain a visual representation of the similarity of sites. In addition, a level plot was used to view genotype performance across sites.

Isolating environmental variables that are driving GxE

Environmental data were extracted from the National Climate Database (CliDB) operated by New Zealand's National Institute of Water and Atmospheric Research (NIWA) using the CliFlo web service[2]. NIWA store data from numerous weather stations around New Zealand and the Pacific. Hawth's geospatial analysis tools in ArcMAP® 9.2 (ESRI, 2006) were used to select weather stations situated closest to each trial's GPS location. In some cases, weather stations that were further from the trial coordinates than the closest station were selected, in order to ensure that sufficient data for climatic variables were available. The average straight line distance a weather station was from a trial was 25km, with a maximum distance of 103km.

As many variables as were available were extracted from CliDB for the period between and including the year of establishment of the trial and the time of measurement of the trial. It was decided to restrict variables used in the analysis to those that were available for most sites. Environmental variables included: total rainfall (mm), mean air temperature (°C), mean daily maximum air temperature (°C), mean daily minimum air temperature (°C), extreme maximum air temperature (°C), extreme minimum air temperature (°C), mean vapour pressure (hPa), and maximum 24-hour rainfall (mm). For many trials, a combination of 2 or 3 weather stations were used to ensure climatic data were available for the desired time period. In addition, altitude was calculated by intersecting the trial GPS coordinates with the underlying raster from LandCare Research's Digital Elevation Model.

Dissimilarity matrices were constructed for each environmental variable by calculating the distance between two sites as the value at site two subtracted from the value at site one. These (independent or explanatory) environmental matrices were then compared with the

---

(dependent or response) proximity matrix using a procedure known as multiple regression on distance matrices or MRM (Legendre *et al.*, 1994; Lichstein, 2007; Smouse *et al.*, 1986).

MRM has evolved from the work of Mantel (1967), who was investigating time-space clustering of leukemia. Compared to traditional Mantel analysis, MRM offers the opportunity for separating environmental variables into individual distance matrices to allow inferences to be made without fear of dilution by unimportant variables (Lichstein, 2007). Here, it was used to assess the impact of various environmental variables on the performance of genotypes as measured by a phenotypic growth response. In addition, calculations for fitting a MRM are the same as those for a multiple regression with standard datasets. However, due to dependence issues in a distance matrix, significance of results for MRM is usually tested through permutation rather than using Fisher's Z-transformation (Dow & Cheverud, 1985).

Each matrix was symmetric, so that the upper right triangle and the lower left triangle of the matrix were reflections of each other. As the entries on the main diagonal represented the distance between a site and itself, they were all zero. Therefore, one of the triangular portions of each matrix was regarded as redundant and the main diagonal (containing self-distances) was discarded, leaving $n(n-1)/2$ distances. The remaining distances in all matrices were then unfolded in the same sequence to form vectors of distances. The vector of distances for dbh was then regressed against the explanatory distance vectors using the GLM procedure in SAS, resulting in a series of t-statistics.

To avoid the use of traditional parametric tests of significance, the rows and columns of the dbh proximity matrix were randomly permuted 2000 times. At the end of each iteration, t-statistics were calculated and used to develop null distributions for the test statistic of each explanatory variable.

Identifying interacting sites

Due to the sparseness of the proximity matrix for dbh, many of the more common clustering techniques were unable to be applied to the RPBC data. However, use of some basic definitions from Graph Theory allowed the formation of an algorithm for grouping sites.

An adjacency matrix is defined as the n x n matrix in which the entry in row *i* and column *j* is the number of edges joining the vertices *i* and *j* (Aldous & Wilson, 2000). For the purposes of this study, the existence of a connection between two sites was thought more important than the exact "length" of that connection *per se*. Therefore, an initial threshold of 0.02 was selected and the proximity matrix was converted to a form of adjacency matrix by replacing threshold-bound entries with unity, indicating an incidence or strong connection between the two sites. Missing entries and threshold-exceeding entries were replaced with zero to represent no or low incidence (or connection) between those two sites.

It is also true that the number of walks of length *k* from vertex *i* to vertex *j* is equal to the entry in row *i* and column *j* of kth power of the adjacency matrix (Aldous & Wilson, 2000). For this reason, the adjacency matrix was then multiplied by the nth (i.e. 64th) power, replacing all positive entries with unity at each matrix-mulitplication, to give blocks of connectedness. From the resulting matrix, clusters were identified as those row (or column, due to symmetry) numbers in each column (or row) where the matrix entry was unity.

The threshold was then slackened by 0.01 and the process repeated. Each change in threshold presented an expanded set of clusters. That is, more sites were included in existing clusters, new sites formed a cluster of their own, and/or two or more clusters merged into one cluster.

Results and discussion

Construction of the proximity matrix

The proximity matrix highlighted an issue of connectedness between sites, with only 665 out of the 2016 cells in the (triangular portion of the 64 x 64) matrix populated. The lack of connectedness exists due to sites having insufficient families in common. Given that the trial sites were established over such a long time period, it is likely that trials were planted with the families that were considered the potential "winners" at the time. As subsequent trials were established, opinions and objectives changed and the group of "winners" appears to have been adjusted to fit with the thinking of the day.

However, sufficient information was available to be able to produce a multi-dimensional scaling plot to represent the similarity of sites to each other (Figure 2). Sites which are located closely to each other encourage similar performance rankings of the families that are planted on them. Conversely, sites that are further apart contain family rankings that differ from each other. No obvious pattern was noted in the MDS plot.
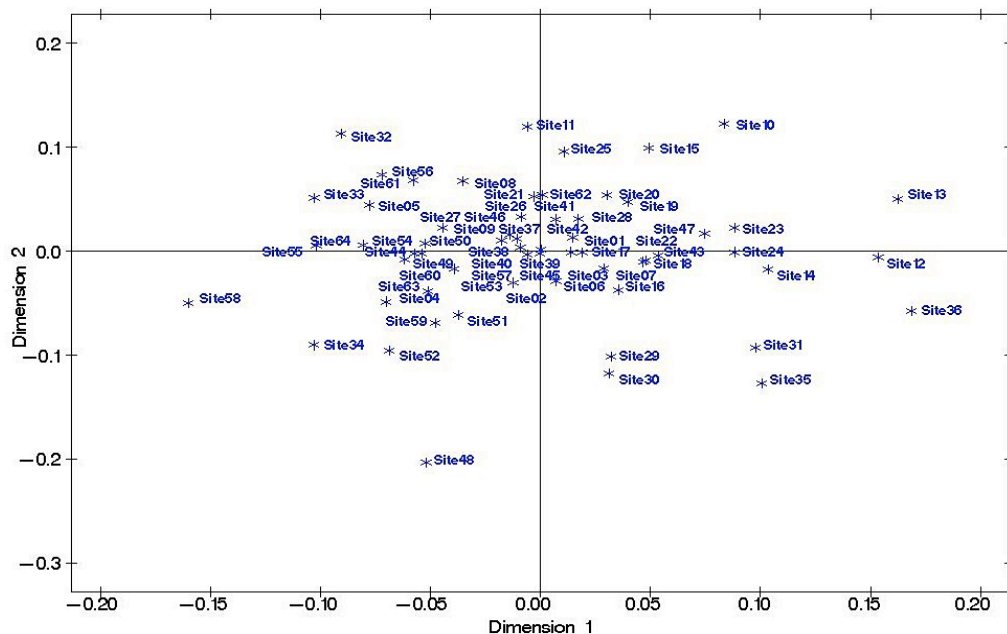

Figure 2 Multi-dimensional scaling plot of dbh

In addition, genotype performance across sites was compared using a level plot (Figure 3). Those genotypes whose rows are predominantly darker shades tended to be ranked in the bottom of the diameter distribution at each site. In contrast, those genotypes with lighter shades were consistently ranked in the top of the diameter distribution across sites. Where a genotype displays a mix of darker and lighter shades, the rank of that genotype is fluctuating significantly between sites, and hence, it is subject to strong genotype by environment interactions.

More specifically, it can be seen that family F5 is an under-performer on most sites on which it is present. Family F17, on the other hand, is consistently ranked in the top twenty percent of families on the sites at which it is present (Note that a family such as this does not necessarily represent a family that is most desired in a breeding programme. It could be argued that such a family will be unable to respond to improved site conditions). Families

F13 and F46 display drastic changes in ranking depending on at which site they are established. The latter two families are clearly expressing strong genotype by environment interactions.
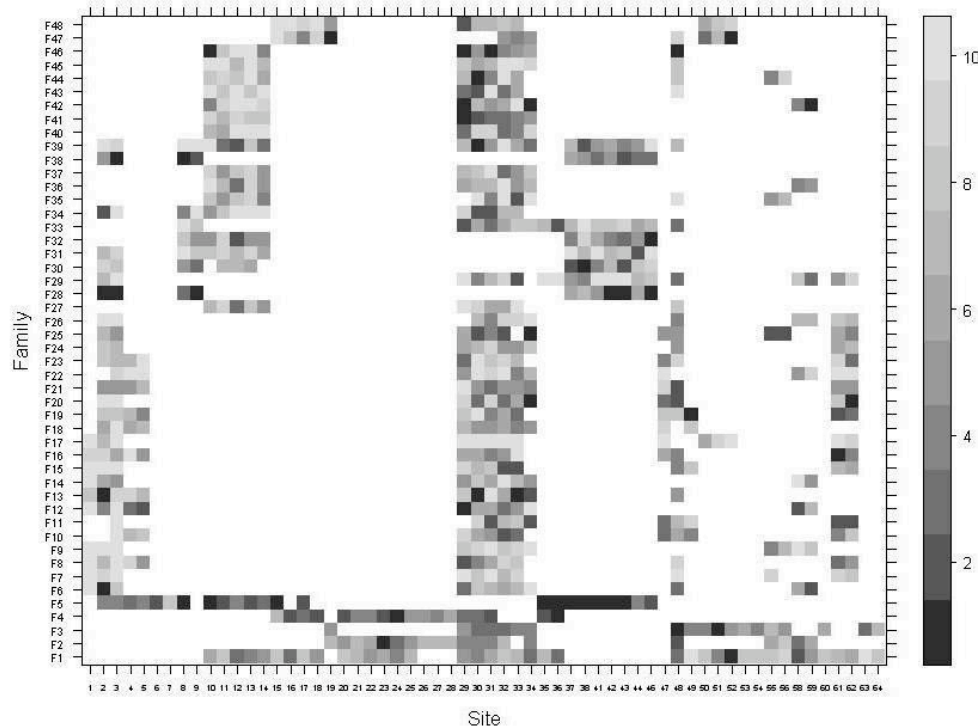

Figure 3 Level plot of family performance (dbh) across sites

Isolating environmental variables that are driving GxE

The main drivers of the genotype by environment interactions in this study were found to be total rainfall and extreme maximum and minimum temperature. Altitude has been found to be important for genotype by environment interactions in Australia (Carolyn Raymond, *pers. comm*. 2009). It may be possible that these studies are displaying a similar phenomenon, as rainfall and temperature extremes are closely linked to altitude in New Zealand. These results warrant further investigation into this dataset with a more thorough methodology.

Identifying interacting sites

The threshold technique did not result in a useful separation of the sites. However, this technique has showed a lot of promise during informal testing and will likely be persisted with for the remainder of the study. It is possible that the technique requires the stringency of including the experimental design features.

Future work

It is intended to conduct all linear mixed model analyses using ASReml. In this way, the univariate analyses of each site will be used to obtain starting values for (co)variance components. This will then be followed by bivariate analysis of all pairs of trials, taking into account differences for experimental and mating designs. It is hoped that by incorporating all

trial information in the analyses, more definitive results might be produced by the techniques described above.

Acknowledgments

References

Aldous, J.M., and Wilson, R.J. (2000) *Graphs and Applications: An Introductory Approach*. Springer, London.

Burdon, R.D. (1977) Genetic Correlation as a Concept for Studying Genotype-Environment Interaction in Forest Tree Breeding. *Silvae Genetica*, **26**, 168-175.

Burdon, R.D., Firth, A., Low, C.B., and Miller, M.A. (1997) Native Provenances of *Pinus radiata* in New Zealand: performance and potential. *New Zealand Journal of Forestry*, **41**, 32-36.

Carson, S.D. (1991) Genotype x Environment Interaction and Optimal Number of Progeny Test Sites for Improving *Pinus radiata* in New Zealand. *New Zealand Journal of Forestry Science*, **21**, 32-49.

Dow, M.M. and Cheverud, J.M. (1985) Comparison of Distance Matrices in Studies of Population Structure and Genetic Microdifferentiation: Quadratic Assignment. *American Journal of Physical Anthropology*, **68**, 367-373.

ESRI (2006) ArcMAP 9.2, Environmental Systems Research Institute, Redlands, California.

Johnson, G.R. and Burdon, R.D. (1990) Family-Site Interaction in *Pinus radiata*: Implications for Progeny Testing Strategy and Regionalised Breeding in New Zealand. *Silvae Genetica*, **39**, 55-62.

Legendre, P., Lapointe, F.J., and Casgrain, P. (1994) Modeling Brain Evolution from Behavior – a Permutational Regression Approach. *Evolution*, **48**, 1487-1499.

Lichstein, J.W. (2007) Multiple Regression on Distance Matrices: A Multivariate Spatial Analysis Tool. *Plant Ecology*, **188**, 117-131.

Mantel, N. (1967) The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, **27**, 209-220.

Jayawickrama, K.J.S., Carson, M.J., Jefferson, P.A., and Firth, A. (1997) *Development of the New Zealand Radiata Pine Breeding Population*. In Proceedings of IUFRO '97 Genetics of Radiata Pine, pp 217-225, Rotorua, New Zealand, 1-4 December 1997.

New Zealand Forest Owners Association (2008) New Zealand Forestry Facts & Figures 2007/08. New Zealand Forest Owners Association, 85 The Terrace, GPO Box 1208, Wellington.

Shelbourne, C.J.A., Burdon, R.D., Carson, S.D., Firth, A., and Vincent, T.G. (1986) *Development Plan for Radiata Pine Breeding*. Forest Research Institute, Rotorua, New Zealand.

Smouse, P.E., Long, J.C., and Sokal, R.R. (1986) Multiple-Regression and Correlation Extensions of the Mantel Test of Matrix Correspondance. *Systematic Zoology*, **35**, 627-632.