

Abstract

Next Generation Breeding Values

**Dutkowski, GW¹, Kerr, RJ¹, Tier, B², Li, L², Costa e Silva, J³, Ivković, M⁴, Potts, BM⁵,
and McRae, TA¹.**

¹ PlantPlan Genetics, c/- School of Plant Science, University of Tasmania, Locked Bag 55, Hobart 7001, Australia
E-mail: Greg.Dutkowski@plantplan.com

² Animal Genetics and Breeding Unit, University of New England, Armidale, Australia

³ Instituto Superior de Agronomia, Lisbon, Portugal

⁴ CSIRO, Canberra, Australia

⁵ School of Plant Science and CRC for Forestry, University of Tasmania, Hobart, Australia.

Short Abstract

The current benchmark for genetic value prediction is program-wide multivariate Best Linear Unbiased Prediction of measured traits, from which harvest age traits are predicted and then combined into economic indices. This is a very flexible framework, but is demanding to implement and not many programs have done so. Models for hybrid populations, calculation of standard errors for harvest traits and indices for genotypes, families and larger deployment units, economic models for risk traits, and integration of synthetic variables derived from markers information into evaluation are all in the process of operational adoption. There are other advances, ranging from accounting for indirect genetic effects to integration of genome-wide molecular information for which there is still much work needed to allow them to be used on a program-wide scale. For operational implementation, both the computational systems and their supporting database systems need to be developed to store and process the increasing amounts of information used.

Current Generation

The benchmark in current generation genetic value prediction is program-wide massively multivariate Best Linear Unbiased Prediction (BLUP) of all measured traits incorporating group effects, individual breeding values by using the additive relationship matrix, family specific combining abilities (SCA), and within family non-additive effects. All within trial design features are used, the data is standardised across trials by the additive genetic variance, and heritability is allowed to vary by trait within trial. The variates used represent different traits, measured at different ages, and on different site types. This allows the most precise estimation possible by using all the correlated data from different traits and ages, while accounting for genotype by environment interaction, as well as accounting for selection that has taken place on these correlated traits. These measured traits are then transformed into harvest age stand economic traits as tree breeding involves indirect selection, and then economic indices are used to properly weight selection towards those traits that have the highest economic impact.

Spatial analysis using autocorrelated residuals is routinely used in this framework to better account for within trial environmental variation. This is done in a two stage process, by initially deriving the best environmental model for each trait on each site, and then adjusting the data for those effects. Testing of direct incorporation of the spatial model into program-wide evaluation indicates that while it is computationally feasible, it is probably not computationally efficient as little extra gain in accuracy is made.

With current climate change scenarios and the long-rotation lengths of tree crops, there is increasing interest in making selections for future abiotic and biotic environments. This can largely be achieved within the current framework by appropriate classification of test sites that relates site characteristics to the likely future environments. Traits that have some better relationships with future environments (such as physiological measurements, or pest and disease susceptibility) can simply be added as extra traits into the multivariate genetic value prediction. Such traits will be most effective where they are better related to future environments than environments of previous trials and/or they can be more cheaply or more rapidly assessed than testing in new environments where this is possible.

Not many tree breeding programs have reached this benchmark, but more are doing so as the tools and skills develop. The approach is not without challenges, notably in estimation of appropriate variances and correlations (upon which BLUP depends), especially involving the harvest age traits which are not routinely measured in progeny trials.

Next Generation

The next generation of genetic values have expanded into new areas, but the tools for them are not fully developed or applied.

In the TREEPLAN software, we have enabled the application of a model for hybrid breeding programs. This initial model allows for inter-species interaction (heterosis - positive or negative) and assumes that thereafter the performance of an advanced generation hybrid synthetic population is the average of the pure species performance (and heterosis) weighted by the proportion of the parental species. The genetic variances are also assumed to be a weighted average of the parental species variances. We will be testing this model in the next year with Queensland southern pine plantation species, although it is unclear if sufficient appropriate test material exists to explicitly test all of the assumptions of the model. In hybrids that have more complex interactions, more complex models will no doubt be necessary.

The end product of breeding is family (and mix) or clonal (mix) deployment material, however it is usually not possible to compare different alternative deployment units and say whether they are likely to be different from one another. The estimates of the genetic value of these deployment units may be the weighted average sum of the different types of genetic values that the type of deployment exploits. Estimates of the prediction error variances and covariances of the group, additive, SCA and within family non-additive effects are used to derive standard errors for each type of deployment unit to help breeders in their decision making. These standard errors have also been extended from the measured traits to the harvest traits and then to the indices, as these are the final basis of decision making. Standard errors of genetic values for individual genotypes are often not presented as they seem very large. However the error associated with composite deployment units (such as orchard mixes) that are used will be smaller. The use and understanding of these error estimates in operational programs is still to be tested.

Next generation genetic values will increasingly rely on additional information that is coming from molecular genetic studies, which must be effectively integrated with information with phenotypic information. The molecular information can come in a variety of forms. Information from a relatively limited number of neutral markers can be used to identify with either certainty or some probability the true parents of open pollinated or bulk families. This information can then be used to modify the additive relationship matrix that is used in breeding value prediction, and are easily incorporated into existing models.

Breeding values incorporating non-neutral marker information have been developed using a synthetic pseudo-continuous trait. Multi-locus marker-trait responses are combined into a prediction of genetic merit for this trait based on genotypes which have both phenotypic and marker information. In the estimation process, the reliability, variance and correlation of the synthetic trait with other phenotypically based traits is estimated. This prediction can then be made for genotypes with only marker information and is relatively easily integrated into normal genetic value prediction. Application of this method to wood quality traits in a *Eucalyptus nitens* seed orchard has demonstrated the potential to increase gain by culling of the orchard where limited phenotypic information on the genotypes exists.

Factor analytic models are becoming more widespread in dealing with across trial genetic value prediction and are available through a variety of software packages. They offer the ability to understand patterns of genotype by environment interaction, as well as to estimate genetic values with a reduced parameterisation that is computationally more efficient. Genetic values can thus be predicted along one or more environmental gradients, as well as for each test site. However the models can break down when there is a very high degree of disconnection between the trials.

Where traits are only sporadically expressed, then it is necessary to work out how much selection pressure needs to be applied to those traits. Calculating the economic cost of traits such as foliar infection and drought mortality under a variety of scenarios in *Pinus radiata* has enabled them to be incorporated into the economic models. This will enable selections that appropriately balance these traits against other production traits and increase the population means for both traits at a rate appropriate for the risk involved.

Future Generation

The large amounts of molecular information which are increasingly becoming available can be incorporated into prediction in a variety of ways. Rapid advances in genotyping technology will soon make it feasible to test individuals for hundreds of thousands of single nucleotide polymorphisms simultaneously at a reasonable price. A genomic relationship matrix based on the markers can be merged with the average coancestry matrix currently used. This will increase the accuracy of breeding value prediction. Low marker density scans of an individual can probably be used when close relatives have high marker densities.

Where these whole genome scans reveal alleles that have large effect, these can be directly incorporated into the model. Such detection is computationally much more demanding, with Bayesian methods often being used. Once detected, their incorporation into the models to use all phenotypic data is relatively straightforward. This promises to be very useful for traits, such as important harvest age traits, that are only expressed late in the tree's life, or are only expressed rarely, such as pest and disease traits, or are very expensive to measure. This has the potential to markedly increase gain. Continued collection of phenotypic data will be required to ensure precision of the estimates.

Current models do not produce genetic values for every trial, but produce values for as many correlated site types as the breeder (and forest owner) can effectively identify and the inter-site type correlations are low enough (or areas large enough) to justify separate genetic value prediction. This model does not allow for low repeatability of performance within some site types (identified through low genetic correlations between sites in a site type) as fitting an individual genotype by trial interaction term would be computationally infeasible. We are however looking at fitting a parent (and cloned genotype) by trial interaction term as an approximation. This should be computationally feasible, and can be used to shrink genetic values on site types with low repeatability. This repeatability will also tell the breeder that they

need more trials on such site types. Insufficient numbers of trials on different site types would make it difficult to identify appropriate site types and estimate the necessary between and within site type correlations.

Indirect genetic effects have been shown to have potential importance in tree breeding. In essence the genes in an individual can affect both the individual's performance, and those of its neighbours. The net stand heritability for a trait such as productivity can be calculated from the size of the indirect genetic effects and the correlation between direct and indirect genetic effects, and the spatial arrangement of the genotypes. In some trials it has been found that the indirect genetic effect for tree size can be large and negatively correlated to direct genetic effects to the extent that the stand heritability for the trait is close to zero. Fortunately this indirect genetic effect has not been found to be statistically significant in many trials, but this may be a power issue and such effects are of concern. Realised gain trials have shown stand level responses to individual size selection, which suggests that such a low stand heritability is not always the case, but could explain why in some circumstances realised gains are lower than predicted. For some disease traits the correlation between direct and indirect genetic effects is expected to be positive so that the net heritability is increased. Selection for low disease damage may reduce not only your own disease damage, but also those of your neighbours, resulting in an increased response to selection.

Incorporation of these indirect genetic effects gives the potential for better estimation of harvest age stand traits. Stands are however dynamic: competition can increase with age, but then reduce again if the stands are thinned or small trees die. The rate at which competition develops will be determined by both productivity and stand management practices. It is unclear how to incorporate this dynamism into the translation from the measured traits to harvest age traits. Indeed, more work needs to be done to understand the prevalence and dynamics of these indirect genetic effects. Direct and indirect genetic effects are also likely to be present for other key traits such as survival and these are likely to be correlated to the genetic effects for size, increasing the complexity of incorporation into stand productivity estimates.