Forest&Wood Products Australia
Knowledge for a sustainable Australia

RESOURCES

PROJECT NUMBER: PNC076-0809 JULY 2011

Industry wide genetic analysis of tree breeding data using TREEPLAN®

This report can also be viewed on the FWPA website
www.fwpa.com.au
FWPA Level 4, 10-16 Queen Street,
Melbourne VIC 3000, Australia
T +61 (0)3 9927 3200 F +61 (0)3 9927 3288
E info@fwpa.com.au W www.fwpa.com.au

Industry wide genetic analysis of tree breeding data using TREEPLAN®

Prepared for

Forest & Wood Products Australia

by

R.J. Kerr, T.A. McRae, L. Li, B. Tier, G.W. Dutkowski
and J. Costa e Silva

Publication:

Industry wide genetic analysis of tree breeding data using TREEPLAN®

Project No: PNC076-0809

This work is supported by funding provided to FWPA by the Australian Government Department of Agriculture, Fisheries and Forestry (DAFF).

ISBN: 978-1-921763-39-7

Researcher/s:

Southern Tree Breeding Association Inc.:	**T.A. McRae, D.J. Pilbeam and P.A. Buxton**
PlantPlan Genetics Pty Ltd:	**R.J. Kerr and G.W. Dutkowski**
Animal Genetics and Breeding Unit (AGBU):	**L. Li and B. Tier**

Final report received by FWPA in July 2011

Forest & Wood Products Australia Limited
Level 4, 10-16 Queen St, Melbourne, Victoria, 3000
T +61 3 9927 3200 F +61 3 9927 3288
E info@fwpa.com.au
W www.fwpa.com.au

Executive Summary

TREEPLAN software is used for genetic evaluation of tree breeding data on an industry and species wide basis. It has the ability to summarise all data and information simultaneously from many genetics trials spread across the plantation estate. This results in a much better understanding of the genetic potential of a given tree and its value for use in further breeding and deployment i.e. it enables us to identify the best trees for a given situation.

Its advanced functionality has been broadened to cover the full range of plantation species, using advanced statistical and genetic models.

The new version TREEPLAN V3 was used by the STBA for national evaluations in Radiata pine and Blue gum. TREEPLAN V3 was tested with data from 159 *P. radiata* trials spread across the pine estate. Genetic and breeding values were calculated in a multi-generation, multi-site, multi-trait and multi-age analysis for 30 measured traits and four harvest age breeding objective traits: growth (MAI m^3/ha/yr), branch size (cm), stem straightness (mm/m) and timber stiffness (GPa). Economic indices were calculated for different production zones using economic weights for an integrated industry, as were regionalised genetic values which account for genotype by environment interaction (GE).

TREEPLAN V3 was used to analyse data from 139 *E. globulus* trials spread across temperate Australia. Genetic and breeding values were calculated for 15 measured traits and three harvest age breeding objective traits: volume growth (m^3/ha), wood basic density (kg/m^3) and kraft pulp yield (percent). Economic indices were calculated for different production zones using economic weights for industries which grow forests for chip, domestic or export kraft pulp markets. Genetic values and economic indices are being generated on a regionalised and water availability basis to account for GE.

TREEPLAN was also used for an industry wide analysis of *E. nitens* (Shining gum) data.

The software now has the utility for analysing pure and hybrid populations and is ready for application to industry data for the sub-tropical softwoods (*P. elliottii*, *P. caribaea* and hybrids).

It is being used in Sweden for national evaluations in *P. sylvestris* (Scots pine), *P. abies* (Norway spruce), *P. contorta* (Lodgepole pine) and *B. pendula* (Silver birch).

A successful pilot scale evaluation used data from the French tree improvement cooperative for *P. pinaster* (Maritime pine).

Competitive interactions among trees in trials for limited resources can limit the value that individual tree measurements have in predicting stand performance in plantations. It is particularly evident in trials located on sites with high productivity and when thinning does not occur until after later age measurement are taken. Our modelling of competition effects has been vastly improved and we have a much better understanding of the inter-relationships between mortality, trial design and the prediction of competitive effects at both the genetic and non-genetic (residual) levels. It is likely that confounding due to competition at pre-thinning ages is having only a minor effect on our ability to predict mature age, whole stand productivity.

Other enhancements include better modelling and removal of systematic bias; more comprehensive modelling of global effects (such as inbreeding depression and propagation effects); pooled measurements; non-additive gene action (dominance and epistasis);

estimation of genetic parameters for reproductive and fitness traits; accuracies and standard errors of genetic values for selection criteria (measured traits), harvest age breeding objective traits and genetic worth indices which are used to rank genetic material for use in different situations.

TREEPLAN was used to integrate molecular data into routine genetic evaluation in trees for the first time. Molecular based selection criteria were derived as a function of SNP allele affects and allele frequencies. DNA based data were combined with performance records of other phenotypic traits for growth, wood density and pulp yield to predict genetic values and the economic worth of trees in *E. nitens*. The results were encouraging but also show the importance of valuing different types of information through its impact on an economic objective, rather than as traits considered in isolation.

STBA and its members use the TREEPLAN results routinely to make selections for inclusion in breeding arboreta and seed orchards to produce seed and plants. It is important the broader industry adopts TREEPLAN for data analysis to improve efficiencies and enhance genetic gain in other species. Non-members of the STBA should source genetically improved material developed by the cooperative or at least ensure TREEPLAN values are available for choosing genetic material for deployment.

Benefits of the project to industry include better utilisation of the national genetic resources for plantation species, resulting in improved forestry resources better matched to a range of production and processing requirements. This project alone is expected to deliver an additional $10 million net present value (NPV) pa to industry through improved genetics as a result of more accurate genetic values. With broader industry adoption of TREEPLAN V3, the marginal gain in economic benefits can be as high as $25 million NPV pa.

Table of Contents

Executive Summary i

Introduction 1

The objectives of the Research Project 1

The extent to which the Research Project objectives were achieved 2

Software Development 3

Results and Discussion 6

Details of all Intellectual Property Rights and Confidential Information created in the course of carrying out the Research Project 14

Commercial implementation of the results 15

The difficulties encountered 17

Conclusions 17

The Project Manager's recommendations for any further research or actions needed to further assist commercial exploitation of the results of the Research Project 19

Any other details the Company may reasonably require to be included in the report 19

Scientific papers or publications resulting from the Research Project 19

Acknowledgements 20

Researcher's Disclaimer 20

Appendices 20

- Appendix A: Competition effects in forestry .... (24 pages)
- Appendix B: Testing of the SAR[2] model .... (4 pages)
- Appendix C: Genetics of Female Reproductive Ability in Radiata Pine .... (19 pages)
- Appendix D: Specification of TREEPLAN models .... (1 page provided of 16)
- Appendix E: Computing accuracies for genetic values .... (12 pages)
- Appendix F: Optimal testing of molecular breeding .... (27 pages)

Introduction

TREEPLAN is an advanced analytical tool providing accurate and precise predictions of genetic values to operational tree breeders. This project aimed to exploit TREEPLAN's analytical power for the benefit of all sections of the forest growing industry; in particular, the deployment sector and programs not traditionally associated with the Southern Tree Breeding Association. The outputs include: the routine inclusion of reproductive and fitness traits into the assessment framework; better prediction of stand performance through competition models; the undertaking of pilot studies using data from large industry programs; and the better modelling of the genetic structures of hybrid populations.

The outputs are designed to benefit orchard managers who require accurate predictions of individual tree genetic values for phenological and reproductive type traits. These values are vital for the selection of productive orchard clones and for the optimum design of orchards. Inclusion of reproductive and survival traits into TREEPLAN can also aid in the monitoring of the genetic fitness of populations. The prediction of epigenetic effects, which are effects on the phenotype due to genetic mechanisms such as gametic imprinting and DNA methylation are considered. Routine estimations of such effects will benefit deployment. Gametic imprinting has a direct bearing on the decision to use a genotype primarily as a seed or pollen parent in orchards. Epigenetic effects (via DNA methylation) usually find their expression in heritable traits such as cold hardiness and bud set. Production of seed in multiple environments can be enhanced by selection of genotypes that display adaptations to local environmental conditions.

The outputs will benefit plantation managers who require greater knowledge on the relationships between individual tree and stand performance. Individual tree genetic values will no longer be confounded with competition effects, which can lead to over- or under-estimation of the value of improved genetics to plantations.

Use of TREEPLAN across more programs will advance a single genetic evaluation "currency" being used across the entire industry. This will improve efficiencies in tree improvement across Australasia. A deliverable has been especially designed for tree improvement programs that deploy hybrid or synthetic genotypes. The current TREEPLAN analytical model, though capable of analysing across species data, will be enhanced to model more correctly the complex genetic structures of hybrid populations.

This initiative builds on national capability servicing the forestry industry through improved genetics.

The objectives of the Research Project

- Enhance TREEPLAN and DATAPLAN software to account for competition effects and indirect genetic effects.
- Estimate genetic values for reproduction and fitness traits for use in breeding and deployment systems for plantation species.
- Develop models and algorithms for handling pure species and hybrid data in genetic evaluation.

- Benchmark on a national basis genetic material tested in different programs and locations.
- Undertake scoping studies in other tree species using TREEPLAN.
- Consolidate genetic resources on a national and species wide basis, while reducing costs and improving efficiencies in tree improvement.

The extent to which the Research Project objectives were achieved

This project aimed to exploit TREEPLAN's analytical power for the benefit of all sections of the forest growing industry; in particular, the deployment sector and programs not traditionally associated with the Southern Tree Breeding Association.

The project has been largely successful in achieving this objective with new features made operational after rigorous testing.

TREEPLAN software used for analysing field data from trees grown in genetic trials spread across the plantation estate has been upgraded with new innovations. Its advanced functionality has been broadened to cover the full range of plantation species.

We have been able to apply the results and findings of the project in a number of species: *Pinus radiata*, *Eucalyptus globulus*, *Eucalyptus nitens*, *Pinus pinaster*, *Pinus sylvestris*, *Picea abies*, *Pinus contorta* and *Betula pendula*. The TREEPLAN software now has the utility to analyse data for all the main plantation species in Australia, including the sub tropical hybrid pines (*P. elliottii*, *P. caribaea* and hybrids), using the most advanced methods.

TREEPLAN is used routinely for national evaluations in *P. radiata* (Radiata pine), *E. globulus* (Blue gum) and E. nitens (Shining gum). TREEPLAN was used (PlantPlan Genetics) for a combined analysis of *E. nitens* data for Gunns Limited and Forestry Tasmania. Plans are underway to run a species wide TREEPLAN analysis for *E. nitens* (incorporating new data) in the coming year. These species account for more than 80 percent of the plantation estate in Australia.

An objective of the current project was to broaden its application to other species, including inter-specific hybrids and synthetic populations typical of the sub-tropical softwoods (*P. elliottii*, *P. caribaea* and their hybrids) grown in southern Queensland by Forestry Plantations Queensland (FPQ). The new V3 version of TREEPLAN was successfully trialled using simulated data for an industry scale multi-species hybrid tree improvement program. A pilot was completed for INRA using data for *P. pinaster* (Maritime pine). Data for trees in ten field trials was used to predict breeding values for 84,964 genotypes spread across generations. This species is important for drier areas in WA.

Negotiations will continue with other parties (RPBC, SFNSW and DAFWA) for piloting and adoption of TREEPLAN for genetic evaluation in other programs and species.

Modelling competition effects was a major sub-project. A choice of alternative models was identified, and the identification of a set of suitable research trials for testing was completed. A paper detailing the findings will be submitted for publication in due course.

Other enhancements include modelling and removal of systematic bias associated with different operators assessing different trees in trials. Global effects (other than additive and family genetic) common to all trials in an analysis, such as inbreeding depression and propagation effects, can be modelled. Data for a pooled measurement where the observation is scored on a sample of material sourced from more than one genotype can also be included. For example, where cores or disks from more than one tree are ground and combined into a single sample for wood quality testing. Effects due to non-additive gene action (dominance and epistasis) are also better modelled.

Genetic parameters for reproductive and fitness traits in *P. radiata* and *E. globulus* were studied. These traits are important for monitoring population fitness and predicting reproductive capacity of parent trees in seed orchards. TREEPLAN provided accuracies of genetic values for selection criteria (measured traits) to reflect their reliability. Standard errors are now available, as well as accuracies for breeding objective traits (derived from multiple selection criteria) and genetic worth indices which are used to rank genetic material for use in different situations. This can be important in breeding and deployment as values of similar magnitude may have different reliabilities.

TREEPLAN was used to integrate molecular data into routine genetic evaluation in trees for the first time. The genotypic data for *E. nitens* used in the study belonged to the Hottest 100 FWPA project. Molecular based selection criteria were derived as a function of SNP allele affects and allele frequencies. These DNA based data made available by CSIRO were combined with performance records of other phenotypic traits for growth, wood density and pulp yield to predict genetic values and the economic worth of all trees in the genetic resources of Gunns Limited and Forestry Tasmania. The results were encouraging and show the importance of valuing different types of information through its impact on an economic objective, rather than as traits considered in isolation.

The source code of TREEPLAN software was upgraded to incorporate the research findings and is released as the operational version for routine genetic evaluation.

Software Development

The TREEPLAN® Analytical Software ("TREEPLAN®") is a FORTRAN program designed for the efficient and accurate prediction of genetic values in trees for breeding and deployment purposes. It is designed to apply best practice analytical technologies to data associated with large scale tree improvement programs like those of the Southern Tree Breeding Association for *Pinus radiata* and *Eucalyptus globulus*.

TREEPLAN® allows data with genetic linkage across generations, sites, years and age classes to be combined into a single multi-trait analysis to produce, for all genotypes, a complete list of genetic (breeding) values for each trait and environment combination. TREEPLAN® fits a reduced individual tree model for purposes of efficiency. TREEPLAN® can model multiple genetic groups, handle individual and clonal data, fit multi-trait models with many traits, accommodate heterogeneous variances (allows for differing heritability), fit site specific statistical and genetic mixed models, and also weights information across environments to account for genotype by environment interaction (GxE), and across time to allow for age:age correlations.

The TREEPLAN® FORTRAN code interacts with a modern web based data management system DATAPLAN® that efficiently handles large databases with full pedigree information for trees and plants. DATAPLAN® acts as a dynamic repository for data and pedigree information, genetic and environmental parameters, and has the capability to handle multiple species. The system not only facilitates efficient storage and retrieval of data for genetic evaluation, but also the delivery of genetic values and other information to nominated parties through the Internet. As new trials and traits are assessed by breeders, the data are validated and entered into a database, and preliminary analyses are done on a single site basis using other software to estimate variance and covariance components which are used to allow for differences in productivity and heritability between sites.

When TREEPLAN® is run on a full data set, breeding values for all trees (genotypes) in a specified population are updated using all information available on the individual genotype itself, any correlated traits, and information from all relatives in the pedigree.

TREEPLAN® versions V1, V2 and V3

TREEPLAN® V1, V2 and V3 are FORTRAN software programs. The FORTRAN source code and binaries is one component of the overall genetic evaluation system. Significant resources and intellectual property has been invested in other components of the TREEPLAN® model. These other components include: (i) the data management system (DATAPLAN®) and its integration with the FORTRAN program(s); (ii) the estimation of genetic parameters; (iii) the system for running the FORTRAN binaries via the web; (iv) validating the genetic values; and (v) data output management and transfer of results to industry.

TREEPLAN® V1 features included: (a) genetic groups to take account of founder effects, (b) mapping of measured traits to a smaller number of selection criteria to handle genotype by environment interaction, and provide flexibility to the system for adding new traits, (c) up to 20 different traits in the model, (d) handling of heterogeneous environment and genetic variances, (e) flexible mixed models so that experimental design effects, such as REP, BLOCK and PLOT can be treated as fixed or random effects in each trial by breeders, and (f) a reduced individual tree model that is efficient because of the high proportion of progeny (non-parents) in tree breeding data.

The system has also been integrated with the STBA data management system (DATAPLAN®) that is operated via a web based interface. This facilitates the regular update of breeding and genetic values as more information and data are collected.

TREEPLAN® V1 is based on related animal breeding technologies in so far as algorithmic ideas and procedures. The source code for TREEPLAN® V1 began from a clean slate with “TREEDRIVER” preparing data that is unique to forest tree genetics trials. The algorithms in the programs “TREEINDEX” and “TREESOLVE” rely heavily on technologies developed over the last 25 years. However, they differ from analogous animal programs due to the innovative reduced, implicit individual tree (animal) model that was used, and because of the unique features of forest tree data.

TREEPLAN® V2 features included: (i) All of TREEPLAN® V1's essential features; (ii) a much faster and more efficient iterative method --- preconditioned conjugate gradient (PCG) instead of Gauss-Seidel (GS) to solve the mixed model equations (MMEs); (iii) accommodates more than 20 different traits in the multivariate model; (iv) accommodated replicates of cloned genotypes; (v) accommodated more complex pedigree structures where parentage is uncertain - such as genetic aggregates of male and/or female parents, including partial selfing, pollen mixes, seed orchard pollen and bulks; (vi) non-additive gene effects within family are calculated for the genotypes with observations; (vii) analyses of categorical and binary data; (viii) incorporates DNA marker information; and (ix) included spatial modelling of environmental effects.

Inclusion of these features has required significant changes to both DATAPLAN® and its interface with the TREEPLAN® software. Further enhancements to DATAPLAN® are required in order to fully capitalize on these new features.

TREEPLAN® V3 includes:

- All of the features of TREEPLAN® V1 and TREEPLAN® V2
- the ability to analyse multi-species data and data collected on their cross derivatives
- the ability to correct for competition at the residual level
- the ability to combine molecular based data with phenotypic based data
- ability to specify global effects, including
    - inbreeding depression
    - propagation effects
    - major gene effects
    - epi-genetic effects
- the ability to specify a variance-covariance matrix for genetic group effects (as opposed to assuming the same correlations between traits as for the additive effects and assuming the genetic group variance is 0.2 times the additive variance)
- the ability to specify a variance-covariance matrix for individual dominance effects, which are often referred to as “delta” effects (as opposed to assuming the same correlations between traits as for the family effects and assuming the delta variance is 3 times the family variance)
- the ability to include data measured on non-clonal plot measurements
- standard errors as well as accuracies for individual breeding and total genetic values at the
    - selection criteria level
    - breeding objective trait level
    - profit index level
- standard errors as well as accuracies for aggregated breeding and total genetic values (e.g. a family, family mix, mating aggregate) at the
    - selection criteria level
    - breeding objective trait level
    - profit index level

In addition, our knowledge base for determining the correct statistical model on a trial by trial basis using ASReml and other software has increased dramatically over the course of the project. Determination of the most accurate statistical model and the extraction of within trial variances is a very important precursor to the global type analysis performed by TREEPLAN®. The research team is at the forefront of genetic trial analysis and is continually building on its repertoire of tools to aid the preparation of data and interpretation of results. Projects of this type are important for increasing our knowledge base and adding to the suite of software tools.

Results and Discussion

The project had the following key objectives:

1. Competition effects accounted for in genetic analysis. TREEPLAN breeding and genetic values for survival and production traits accurately predicting harvest performance, without bias due to competition in small plot trials
2. Incorporation of reproductive and fitness traits in TREEPLAN
3. Methods for modelling epigenetic and indirect genetic effects
4. TREEPLAN enhanced to facilitate combined pure- and cross-species breeding and selection
5. Complex non-additive models in TREEPLAN and other miscellaneous functionality
6. Implementation and scoping studies in non-STBA programs

During the course of the project the provision of accuracies for breeding objective traits and economic worth indices ($NPV), at both the genotype and family levels, became an important priority. This was added as objective 7.

It should be emphasised that objectives 1 and 2 are objectives of an exploratory type and did not necessarily require substantial changes to TREEPLAN source code.

In Objective 1 we undertook a comprehensive investigation into the appropriate models for use in competition analysis in forestry. Simulated data was used to check the practicality and accuracy of various models. From our simulation studies the main findings are:

- Autoregressive structures are equal, if not superior to fitting neighbourhood structures, in terms of modelling competition at the residual level.
- Use of intensity of competition factors when fitting neighbourhoods are a convenient way to integrate mortality, unequal intra-row and inter-row tree spacing, and competition into the one model.
- Competitive additive variance, if not accounted for in the model, will under most conditions be absorbed into the error variance. Thus we may be at risk of underestimating heritability for growth traits.
- If there is a strong negative correlation between genetic direct and competitive effects, we may be compromising our selection response by not exploiting the "hidden" heritable variance due to genetic competition.

Our competition models were then applied to observed field data. In the three species examined: *Pinus radiata*; *Picea abies*; and *Eucalyptus globulus*; similar patterns were evident. The main findings are:

- Competition at the genetic level is rarely evident at pre-thinning ages.
- At post-thinning ages genetic competition effects are really only evident in site types with high productivity.
- At pre-thinning ages competition at the residual level is evident only in site types with high productivity.

- A simple fitting of the presence or absence of a neighbour, together with an autoregressive structure to the residuals, is also a good method to account for both mortality and competition at the residual level.
- A third order equal roots autoregressive (EAR3) model is an appropriate model for fitting both local trend and competition at the residual level.
- A second order equal roots autoregressive model (EAR2) did not fit the model any better than a standard first order autoregressive model (AR1).

Given these findings we decided to implement EAR3 in TREEPLAN V3, so that both local trend and competition at the residual level can be modelled. Competition at the genetic level requires more study, using both simulated and field data. A draft manuscript, describing in more detail the methods, aims and conclusions reached is presented as APPENDIX A. The results of the testing of the implementation of EAR3 into TREEPLAN V3, also referred to as SAR2, is presented as APPENDIX B.

In Objective 2 our CSIRO collaborator Dr Harry Wu was supplied with fecundity data (counts of female flowers) scored in the Australia wide diallel series of trials. Project resources were used to enter this data into DATAPLAN for further analysis. A moderate negative genetic correlation (-0.3 to -0.4) between fecundity and growth was observed in this large population of Radiata pine. This indicates that improvement of wood volume of trees could reduce fecundity of Radiata pine and vice versa (reduction of the fecundity could divert energy to wood production), which is consistent with theories of reproductive cost. There were also reciprocal and maternal effects evident. The implications of these findings on selection and breeding strategy need to be further studied. A draft manuscript on this topic is presented as APPENDIX C.

Dr Richard Kerr examined survival analysis methods using computer simulations. Restricted maximum likelihood (REML) and Bayesian Gibbs Sampling (GS) were tested for their accuracy in estimating the genetic based variance components for survival. REML has been found to be an inconsistent and biased estimator for this type of trait. The GS approach has been identified as a superior method and has been defined as the standard for the present time. Genetic variances for survival have now been estimated for many trials within the national databases for *E. globulus* and *P. radiata*. REML has been tested as a platform for providing genetic correlations between survival and growth/form/wood quality traits. Simulation studies indicated unbiased estimates can be obtained by fixing the variance for survival using the estimate obtained from the GS. However, few statistically significant estimates have been obtained from the analyses to date. A conclusion is that REML is unsuitable for this task. AGBU aims to expand the GS capability to analyse multiple, binary traits. A manuscript on the findings of this study is currently being prepared.

Objectives 3 to 5 are objectives which required major adjustments to the source code. A thorough description of current TREEPLAN models is provided as APPENDIX D. This confidential document explains the new feature of the source code that we call “global effect fitting”. It is through “global” effects that we are able to model hybrid and synthetic data, inbreeding depression, propagation effects, epigenetic effects, major gene effects and other miscellaneous functionality.

Given that much of the effort of the project was in designing and implementing the new features in the source code, APPENDIX D reflects much of the activity associated with these three objectives. Testing the new source code has been an important activity within the scope of the overall project. An exact replica of TREEPLAN is maintained in Hobart, TAS, but written in OCTAVE. OCTAVE is a free computer program for performing numerical computations. It derives solutions using direct manipulation of the various matrices defined in the mixed model equations. It can only compute solutions for small test data sets. TREEPLAN is written in FORTRAN and uses indirect solving techniques, which are needed for the industrial size datasets of the STBA and other clients. As new functionality is introduced into TREEPLAN quality control is ensured by synchronising the two software approaches.

The following represents the relevant findings for objectives 3 to 5.

Objective 3 does not require much explanation. The project team was not able to demonstrate an analysis, which included epigenetic effects, by means of an actual case study. Resources were diverted to other areas such as accuracies, which were considered a higher priority for tree improvement programs. However, the technology is available in TREEPLAN V3 and we are hoping to pursue leads with colleagues in Norway for provision of suitable test data. There was no testing phase for this objective.

Objective 4. Here we present the results of our testing phase.

Modelling hybrid data in TREEPLAN V3 software makes the genetic evaluation system much more flexible to handle the pure species and cross-species data in one analysis. The corresponding modification of source code has been done for this function. The new functionality has been tested extensively step by step from a simple scenario to more complicated scenarios with simulated data sets. The TREEPLAN OCTAVE program was used to compare the integrity of the solutions for the same data set for all scenarios. The following scenarios (SC) have been tested:

Table 1. The scenarios tested for hybrid data in TREEPLAN V3

Scenario	Number of species	Fixed heterosis	Random heterosis	Cross type	Aggregates in pedigree	Mixed families present	Clonal data present	SCA effects fitted
SC 1	2	Yes	No	F1[a]	No	No	No	No
SC 2	2	Yes	Yes	F1	No	No	No	No
SC 3	2	Yes	Yes	F1,F2[b],BC[c]	No	No	No	No
SC 4	3	Yes	Yes	F1,F2,BC	Yes	No	No	No
SC 5	3	Yes	Yes	F1,F2,BC	Yes	Yes	No	No
SC 6	3	Yes	Yes	F1,F2,BC	Yes	Yes	Yes	No
SC 7	3	Yes	Yes	F1,F2,BC	Yes	Yes	Yes	Yes

[a] The F1 is a first generation cross between two species
[b] The F2 is a first generation cross between F1 individuals
[c] The BC is a back-cross between the F1 and either one or both of the parental species

From the simplest situation (SC 1) with two species and only fixed effects for heterosis in the model to the most complicated one (SC 7), exactly the same solutions were obtained from TREEPLAN V3 and the independent octave program. Though, the testing in this part was focused on the heterosis effect in the cross species data, the new function to handle mixed family data in TREEPLAN V3 was also successfully tested.

The implications for updating DATAPLAN software to handle hybrid data and associated pedigree, and for specifying a BLUP model for the analysis of such data and pedigree were documented.

Objective 5

More complex models of non-additive gene action

In TREEPLAN V2 the proportion of the phenotypic variance due to non-additive gene effects is modelled solely due to dominance. One quarter of the dominance variance is accounted for by fitting a family effect. The remaining three quarters of the dominance variance is accounted for by fitting a dominance Mendelian sampling effect, which we call a "delta" effect.

In TREEPLAN V3 the following changes have been implemented:

- The variance due to a delta effect can now be specified rather than defaulting to three times the family variance. Thus extra non-additive variance over and above just dominance variance can be modelled (ie. epistatic variance).
- Each selection criteria trait can have a different variance due to delta effects.
- A correlation matrix among selection criteria due to delta effects can be specified.
- SCA (specific combining ability) effects are predicted for the component families of a family mix or aggregate. (In TREEPLAN V2, the SCA effect was predicted for an umbrella "pseudo" type family which represented the collection of families in the mix or aggregate.)

Inbreeding depression

Fitting inbreeding depression is an example of fitting a "global effect", i.e. an effect common across all trials. Previously in TREEPLAN V2, an effect other than additive and family genetic effect, could only be fitted within a trial.

A detailed specification for Hutchinson Software has been completed. This specification describes what changes are needed within DATAPLAN software for users to specify the different types of global effects. This specification will be implemented as Hutchinson Software directs resources into this activity.

A series of published papers concerning inbreeding depression in *E. globulus* (co-authored by Dr João Costa e Silva and Prof. Brad Potts) and *P. radiata* (authored by Dr Harry Wu), has been used as the basis for refining how we model inbreeding depression in TREEPLAN. Firstly, the evidence suggests the relationship between depression in performance and inbreeding coefficients is linear. Secondly, the evidence also suggested there is an interaction between sub- or land-race and the parameters of this relationship (intercept and slope). On a practical level, this suggests current databases may be insufficient for estimating inbreeding depression parameters for every selection criteria trait which we suspect is susceptible to this phenomenon.

What this has meant for our implementation in TREEPLAN V3 is that we have allowed multiple regressions within selection criteria to be fitted, which account for the differences among sub- or land-races in the expression of inbreeding depression. In advanced generations where an individual has genes deriving from multiple sub-races, the coefficients used to penalize performance due to inbreeding will be a weighted average of the individual sub-race coefficients.

For those selection criteria with inadequate data to predict a slope separate from other selection criteria, the coefficients can be constrained to equal the coefficients which are estimated for selection criteria with adequate data.

Whole plot measurements

A whole plot measurement is when a single observation is scored on multiple stems within the trial. The stems belong to many different genotypes, either of the same or different families. TREEPLAN V3 handles this type of measurement by giving each stem in whole plot measurement the same value for the observation. The alternative was to apportion the single observation across the multiple stems, using fractions in the incidence matrices that assign observations to effects in the model. Given that the implications of this alternative was a major rewrite of the source code, and given that the differences between the two methods is only minor in terms of accuracy, the alternative was not warranted.

Non-spatial autocorrelation

Analysis of data associated with the INRA pilot with *Pinus pinaster* and the gene conservation project with *Pinus radiata* has uncovered some examples of non-spatial autocorrelation. In most of these cases the effect is due to the assessor or assessors. In most of these cases the effect can be removed pre-TREEPLAN evaluation using the appropriate within-site ASReml model. The scientific team is now much more experienced in noticing and removing assessor effects in data. This knowledge has been documented and will be useful in all future analyses. There has been no need to add additional functionality to TREEPLAN to account for assessor effects.

Objective 6

TREEPLAN V3 was used by STBA for a national genetic evaluation in *P. radiata* and *E. globulus.*

TREEPLAN was also used for an industry wide analysis of *E. nitens* (Shining gum) data for Gunns and Forestry Tasmania. It is anticipated further analyses will be done in the coming year to incorporate recently collected data and new information.

In *E. nitens* TREEPLAN was successfully used to integrate information at the DNA level. A subset of the larger industry-wide genetic evaluation was used for this task. This subset contained 27 trials including: the two genetics trials, in which trees were assayed for a suite of six single nuclear polymorphism (SNP) genotypes; and other progeny trials, in which trees were not assayed for the SNP, but contained the ortets of the 92 seed orchard clones that had been assayed. The results of the pilot study indicated unequivocally that when screening genotypes for deployment, use of molecular information will change what is selected. More importantly, the marginal gain (in terms of economic value) above baseline productivity due to improved genetics will increase by as much as 11% by including this additional information. This additional gain is part of the marginal gain above baseline productivity, not an 11% increase in total productivity or economic value. Note: 89% of the total genetic gain is due to the phenotypic data alone. This pilot study is a remarkable achievement and is probably the world's first practical demonstration of the incorporation of molecular data into routine genetic evaluation in forestry. A confidential report on this project has been provided as APPENDIX F. This work will be further developed and reported on as part of FWPA project PNC220-1011.

TREEPLAN is being used by Skogforsk in Sweden for its national evaluations in *P. sylvestris* (Scots pine), *P. abies* (Norway spruce), *P. contorta* (Lodgepole pine) and *B. pendula* (Silver birch).

The software has the utility for pure and hybrid populations and is ready for application to industry data belonging to FPQ for the sub-tropical softwoods (*P. elliottii*, *P. caribaea* and hybrids). Data is being compiled for this purpose.

A pilot study using data from the French tree improvement cooperative for *P. pinaster* (Maritime pine) was successfully completed in 2011. Data for trees in ten field trials was used to predict breeding values for 84,964 genotypes spread across generations. It is anticipated TREEPLAN will be adopted for national evaluations in France for this and other species. Maritime pine is also an important species for drier areas in Western Australia, and data could be similarly analysed.

TREEPLAN was also used to analyse data on 28,837 unimproved genotypes for a series of provenance trials as part of the FWPA gene conservation project (PNC135-0809). We anticipate this and other data will be incorporated into the national evaluations in due course.

Objective 7

Accuracies and standard errors of breeding objective traits and economic worth indices ($NPV).

A thorough description of the methods to compute accuracies and standard errors for breeding objective traits, and economic worth indices ($NPV), at the level of both genotype and family is presented as APPENDIX E.

The code has been written and Hutchinson Software is now implementing the necessary changes in DATAPLAN for clients and breeders to use the functionality.

In TREEPLAN accuracies and standard errors are only approximated. Correct accuracies and standard errors cannot be made available because they require inverting the coefficient matrix of the mixed model equations. This is not possible, due to its size. Elements of the inverse are approximated using techniques described in APPENDIX D. In order to test the approximation method a real inverse coefficient matrix was obtained using a subset of a TREEPLAN *P. abies* system involving three trials and three selection criteria. Accuracies were then derived using: the approximation method to obtain inverse elements (the prediction error variance and covariances); and then using the actual elements from the derived inverse. Accuracies derived using the approximation method and using the true method was then compared.

Tables 1 and 2 show the results of the comparison.

Table2. Summary of the comparison of accuracies and standard errors of genotype values (breeding values and total genetic values) at selection criteria, breeding objective trait (BOT) and $NPV levels, derived using the approximation method used in TREEPLAN V3 (Approx) and using a true inverse coefficient matrix (True).

Trait type	Value type	Accuracy or Standard error	Method	**Abbreviation**[c]	**N**	**Mean**	**SD**	**Min**	**Max**	**Correlation coefficient**
Selection criteria	Breeding value	acc	Approx	**geno_sc_ebv_acc**	16002	0.736	0.066	0.561	0.879	0.987
		acc	True	**geno_sc_ebv_acc1**	16002	0.695	0.077	0.495	0.876	
		s.e.	Approx	**geno_sc_ebv_se**	16002	0.734	0.079	0.523	0.907	0.985
		s.e.	True	**geno_sc_ebv_se1**	16002	0.779	0.083	0.528	0.952	
	Total genetic value[b]	acc	Approx	**geno_sc_egv_acc**	16002	0.713	0.071	0.572	0.930	0.993
		acc	True	**geno_sc_egv_acc1**	16002	0.698	0.080	0.541	0.949	
		s.e.	Approx	**geno_sc_egv_se**	16002	0.876	0.109	0.465	1.038	0.994
		s.e.	True	**geno_sc_egv_se1**	16002	0.892	0.122	0.399	1.064	
BOT[a]	Breeding value	Acc	Approx	**geno_bot_ebv_acc**	16002	0.921	0.059	0.768	0.981	0.997
		Acc	True	**geno_bot_ebv_acc1**	16002	0.909	0.066	0.740	0.980	
		S.e.	Approx	**geno_bot_ebv_se**	16002	0.396	0.142	0.212	0.702	0.996
		S.e.	True	**geno_bot_ebv_se1**	16002	0.426	0.148	0.219	0.737	
	Total genetic value	Acc	Approx	**geno_bot_egv_acc**	16002	0.916	0.063	0.776	0.985	0.998
		Acc	True	**geno_bot_egv_acc1**	15999	0.910	0.067	0.760	0.985	
		S.e.	Approx	**geno_bot_egv_se**	16002	0.472	0.170	0.220	0.798	0.996
		S.e.	True	**geno_bot_egv_se1**	15999	0.487	0.176	0.218	0.822	
$NPV	Breeding value	Acc	Approx	**geno_idx_ebv_acc**	5334	0.916	0.018	0.876	0.960	0.988
		Acc	True	**geno_idx_ebv_acc1**	5334	0.911	0.019	0.871	0.962	
		S.e.	Approx	**geno_idx_ebv_se**	5334	124.121	13.376	86.930	150.258	0.988
		S.e.	True	**geno_idx_ebv_se1**	5334	131.914	14.160	87.873	157.582	
	Genetic value	Acc	Approx	**geno_idx_egv_acc**	5334	0.911	0.020	0.881	0.977	0.994
		Acc	True	**geno_idx_egv_acc1**	5333	0.910	0.023	0.879	0.984	
		S.e.	Approx	**geno_idx_egv_se**	5334	147.114	19.763	77.131	170.561	0.996
		S.e.	True	**geno_idx_egv_se1**	5333	150.4389	23.14847	65.917	175.521	

[a] Breeding objective trait

[b] Total genetic value is the sum of breeding value, family effect and individual dominance effect

[c] Abbreviations are constructed by concatenating: the entity for which the genetic values is predicted, e.g. "geno" for genotype; the type of trait, e.g. "sc" for breeding objective trait, or "bot" for breeding objective trait; the type of genetic entity, e.g. "egv" for estimated total genetic value, or "ebv" for estimated breeding value; and the statistic used, e.g. "se" for standard error, or "acc" for accuracy. Thus "geno_sc_ebv_acc" represents the accuracy for estimated breeding value computed for a genotype corresponding to a selection criteria trait.

Table 3. Summary of the comparison of accuracies (acc) and standard errors (se) of family values (breeding value and genetic value) at selection criteria, breeding objective trait (BOT) and index levels, derived using the approximation (Approx.) method used in TREEPLAN V3 and using a true inverse coefficient matrix (True)

Trait type	Value type	Accuracy	Source	**Abbreviation**	**N**	**Mean**	**SD**	**Min**	**Max**	**Correlation coefficient**
Selection criteria	Breeding value[a]	Acc	Approx	**fam_sc_ebv_acc**	5853	0.865	0.147	0.592	0.996	0.990
		Acc	True	**fam_sc_ebv_acc1**	5853	0.768	0.144	0.470	0.959	
		S.e.	Approx	**fam_sc_ebv_se**	5853	0.360	0.177	0.079	0.674	0.978
		S.e.	True	**fam_sc_ebv_se1**	5853	0.509	0.123	0.236	0.738	
	Genetic value[b]	Acc	Approx	**fam_sc_egv_acc**	5853	0.813	0.138	0.556	0.989	0.986
		Acc	True	**fam_sc_egv_acc1**	5853	0.726	0.134	0.442	0.947	
		S.e.	Approx	**fam_sc_egv_se**	5853	0.485	0.145	0.133	0.743	0.975
		S.e.	True	**fam_sc_egv_se1**	5853	0.594	0.111	0.288	0.802	
BOT	Breeding value	Acc	Approx	**fam_bot_ebv_acc**	5853	0.958	0.054	0.781	0.999	0.880
		Acc	True	**fam_bot_ebv_acc1**	5853	0.899	0.131	0.322	0.992	
		S.e.	Approx	**fam_bot_ebv_se**	5853	0.206	0.116	0.030	0.523	0.914
		S.e.	True	**fam_bot_ebv_se1**	5853	0.319	0.145	0.107	0.792	
	Genetic value	Acc	Approx	**fam_bot_egv_acc**	5853	0.944	0.057	0.765	0.997	0.919
		Acc	True	**fam_bot_egv_acc1**	5853	0.927	0.075	0.622	0.994	
		S.e.	Approx	**fam_bot_egv_se**	5853	0.266	0.118	0.069	0.576	0.949
		S.e.	True	**fam_bot_egv_se1**	5853	0.355	0.149	0.119	0.822	
$NPV	Breeding value	Acc	Approx	**fam_idx_ebv_acc**	5334	0.9164	0.0179	0.876	0.964	0.987
		Acc	True	**fam_idx_ebv_acc1**	5334	0.9105	0.0188	0.871	0.961	
		S.e.	Approx	**fam_idx_ebv_se**	5334	124.120	13.375	86.930	150.258	0.987
		S.e.	True	**fam_idx_ebv_se1**	5334	131.913	14.160	87.873	157.581	
	Genetic value	Acc	Approx	**fam_idx_egv_acc**	5334	0.911	0.020	0.880	0.976	0.994
		Acc	True	**fam_idx_egv_acc1**	5333	0.9103	0.022	0.879	0.983	
		S.e.	Approx	**fam_idx_egv_se**	5334	147.114	19.763	77.131	170.561	0.995
		S.e.	True	**fam_idx_egv_se1**	5333	150.438	23.148	65.917	175.521	

[a] Breeding value for a family is the average of the parental breeding values

[b] Genetic value for a family is the family breeding value plus the specific combining ability (SCA)

There is generally good concordance between true accuracies and approximate accuracies, as shown in Tables 2 and 3. The correlation between true accuracies and approximate accuracies was generally greater than 0.98. The correlation was only approximately 0.9 for accuracies of breeding objective traits for family breeding and genetic values.

Testing of TREEPLAN V3 using an industrial scale dataset

Prior to releasing TREEPLAN V3, it is important to ensure integrity of all new functionality and consistency with previous versions of TREEPLAN. This is best achieved by analysing an industrial scale data set with TREEPLAN and comparing the analysis with a previous analysis.

An industrial sized test data set was obtained by extracting from DATAPLAN the most recent system developed for *Pinus radiata* (PRAD201105). A total of 497,897 genotypes, including first-, second- and third-generation trees from 159 trials are included. Data were observed on a total of 311,533 stems, for a total of 143 measured traits, which are mapped to 27 selection criteria. The data set was analysed using both TREEPLAN V2 and TREEPLAN V3. The genotype EBV and family SCA values were summarized and the correlation coefficients (R) between two versions were calculated. The results are shown in Table 4.

Table 4. Summary of genotype EBV and family SCA value from TREEPLAN V2 and V3 with *Pinus radiata* data (PRAD201105)

Source	N	Mean	SD	Min	Max	R
EBV_V2	9048435	0.3093	0.6260	-5.9693	5.2284	0.94
EBV_V3	9048435	0.2968	0.6036	-7.8880	6.3339	
SCA_V2	63655	0.0002	0.1431	-2.2421	1.2988	0.93
SCA_V3	63655	0.0001	0.1344	-2.2323	1.3067	

The result shows that the difference in the mean and standard deviation of EBV and SCA between versions 2 and 3 is small. However the correlations are 0.93 and 0.94 which indicate there may be some changes in rankings. The reason for this is most likely caused by fitting probabilistic SCA effects rather than fitting SCA effects for pseudo families that represent a pairing between an individual parent and an aggregate parent, or a pairing between two aggregate parents. Also family mixes existed in the pedigree prepared for TREEPLAN V3, which do not exist in the pedigree prepared for TREEPLAN V2.

Details of all Intellectual Property Rights and Confidential Information created in the course of carrying out the Research Project

The Research Project enhanced the existing TREEPLAN® software Licensed to the Researcher. The enhanced TREEPLAN® software ("TREEPLAN® V3") will remain the property of the Owners. In terms of the existing Licence for TREEPLAN®, the Researcher will retain its exclusive worldwide licence to use, commercialise and exploit TREEPLAN®

V3 in the trees and forestry industry. The Company (Forest and Wood Products Australia Limited) provides the Owners, including the Researcher and PlantPlan Genetics Pty Ltd with exclusive worldwide commercialisation rights to TREEPLAN® V3 without further written agreement.

The Researcher will use TREEPLAN® V3 to provide genetic evaluation services to the forestry industry. The Researcher has Members who are Levy Payers involved in the forestry industry. In consideration of the Company's contribution to the Research Project, a royalty payment for use of the TREEPLAN® V3 software in the forestry industry will be due to the Company. Use of the TREEPLAN® V3 software for the provision of genetic evaluation services to Members of the Researcher will be royalty free. The royalty payment for use of TREEPLAN® V3 software for genetic evaluation services provided to the trees and forestry industry (excluding the Researcher's Members) will be one and a quarter percent (1.25%) of the Net Exploitation Revenue received for use of TREEPLAN® V3 software by the Researcher. This condition replaces the Royalty percentage defined in Clause 1.1 (33).

The Researcher may enhance TREEPLAN® V3 software after the Project Term without prior permission of the Company. Following any additional enhancements to TREEPLAN® V3, future royalty payments due to the Company will be determined on the basis of proportional contributions. The royalty (one and a quarter percent) will be revised, based on the ratio of additional cash and in-kind contributions for subsequent enhancements as a proportion of the overall development costs. For the purpose of calculating future royalties due to the Company if additional enhancements are made to the TREEPLAN® V3 software, the contributions of the parties providing Background Technology at the completion of this Research Project are ninety eight and three quarters of one percent and the Company one and a quarter percent.

Any genetic values produced in the Research Project and associated with the Researchers genetic material will remain the confidential property of the Researcher, PlantPlan Genetics Pty Ltd and their clients and Members. However, genetic information including population parameters (such as variances, co-variances, heritabilities and genetic correlations) will be available for publication in reports and scientific journals subject to the prior written consent of the Researcher. Consent for publication will not be unreasonably withheld.

The national genetic evaluation of the Researchers tree improvement programs in *Pinus radiata* and *Eucalyptus globulus* were completed using the TREEPLAN® V3 software in conjunction with the Research Project. The results of these evaluations are confidential to the Researcher and its Members. The results of genetic analyses in *E. nitens*, *P. pinaster* and other species as part of the research project are confidential to the parties contributing the data.

Commercial implementation of the results

The TREEPLAN software has broad market penetration with genetic evaluation being completed for the main plantation species in Australia. The TREEPLAN platform also provides an integrated framework to deliver benefit from a number of breeding programs and associated research projects in genetics.

The source code of TREEPLAN software was upgraded to incorporate the research findings and is released as the operational version (TREEPLAN V3) for routine genetic evaluation.

TREEPLAN V3 was used by STBA for a national genetic evaluation in *P. radiata*. Results show an improvement in the accuracy of the breeding values due to the fitting of superior genetic and statistical models. TREEPLAN was run with data for 312,015 trees from 159 trials spread across the pine estate. Data was modelled for some 4,373 families (open- and control-pollinated) and 27 genetic groups. This multi-generation, multi-site, multi-trait and multi-age analysis generated genetic values for 30 selection criteria traits covering growth, form, wood properties, and resistances to pests and disease. Genetic and breeding values were calculated for the four harvest age breeding objective traits: growth (MAI m^3/ha/yr), branch size (cm), timber stiffness (GPa) and stem straightness (mm/m). Economic indices were calculated for different production zones using economic weights for an integrated industry, as were regionalised genetic values which account for genotype by environment interaction (GE). GE is where the best genotypes for one region are not necessarily the best for another. TREEPLAN was also used to analyse data on 28,837 unimproved genotypes for a series of provenance trials as part of the FWPA gene conservation project (PNC135-0809). We anticipate this and other data will be incorporated into the national evaluations in due course. The results are stored and accessed via the web based DATAPLAN system, subject to access restrictions.

TREEPLAN V3 was used by STBA for a national genetic evaluation in *E. globulus*. TREEPLAN was run with data for 245,468 trees from 139 trials spread across temperate Australia. Data was modelled for some 3,078 families and 32 genetic groups (localities and sub-races). This multi-generation, multi-site, multi-trait and multi-age analysis generated genetic values for 15 traits covering growth, form, wood and pulp properties, and disease resistance. Genetic and breeding values were calculated for the three harvest age breeding objective traits of volume growth (m^3/ha), wood basic density (kg/m^3) and kraft pulp yield (percent). Economic indices were calculated for different production zones using economic weights for an industry which grows forests for chip, domestic or export kraft pulp markets. Genetic values and economic indices are being generated on a regionalised and water availability basis to account for GE interactions. The results are stored and accessed via the web based DATAPLAN system, subject to access restrictions.

TREEPLAN was also used for an industry wide analysis of *E. nitens* (Shining gum) data for Gunns and Forestry Tasmania. It is anticipated further analyses will be done in the coming year to incorporate recently collected data and new information from industry partners.

TREEPLAN is being used by Skogforsk in Sweden for its national evaluations in *P. sylvestris* (Scots pine), *P. abies* (Norway spruce), *P. contorta* (Lodgepole pine) and *B. pendula* (Silver birch).

The software has the utility for pure and hybrid populations and is ready for application to industry data belonging to FPQ for the sub-tropical softwoods (*P. elliottii*, *P. caribaea* and hybrids). Data is being compiled for this purpose.

A pilot study using data from the French tree improvement cooperative for *P. pinaster* (Maritime pine) was successfully completed in 2011. Data for trees in ten field trials was used to predict breeding values for 84,964 genotypes spread across generations. It is anticipated TREEPLAN will be adopted for national evaluations in France for this and other species. Maritime pine is also an important species for drier areas in Western Australia, and data could be similarly analysed.

STBA used the TREEPLAN results to make selections for grafting into breeding arboreta and seed orchards of industry partners (STBA members and seedEnergy). Industry has used the results to guide harvesting of orchard seed, and choice of seed and plants for deployment through nurseries. Updated results, incorporating historical information and new trial data will be used for deployment and breeding activities on an ongoing basis. Industry accesses the outputs through the web based DATAPLAN system. The systems framework of TREEPLAN is enabling tree breeders and research scientists to efficiently work toward common objectives. The results of TREEPLAN V3 analyses will be used for matching genetics to site types as part of the new SEEDPLAN II project (PNC211-1011).

TREEPLAN was used to integrate molecular data into routine genetic evaluation in trees for the first time. The genotypic data for *E. nitens* used in the study belonged to the Hottest 100 FWPA project. Molecular based selection criteria were derived as a function of SNP allele affects and allele frequencies. These DNA based data made available by CSIRO were combined with performance records of other phenotypic traits for growth, wood density and pulp yield to predict genetic values and the economic worth of all trees in the genetic resources of Gunns Limited and Forestry Tasmania. The results were encouraging and show the importance of valuing different types of information through its impact on an economic objective, rather than as traits considered in isolation. TREEPLAN V3 software will also be used to evaluate further options for testing the utility of molecular breeding as part of the new project PNC220-1011.

Benefits to industry include better utilisation of the national genetic resources for plantation species, resulting in improved forestry resources better matched to a range of production and processing requirements.

The difficulties encountered

There were no major issues impacting on the project.

It was necessary to reorganise some activities to ensure the research objectives were aligned with the needs of the various tree improvement programs. Dr Richard Kerr (PPG) also researched some aspects of the project which were previously allocated to Dr Li Li (AGBU).

David Pilbeam (Eucalypt breeder) assisted Peter Buxton (Pine breeder) with data collation for the national TREEPLAN run for Pine (completed 27 June 2011). The system provides transparency and ongoing security of data and associated information.

Collating data for the subtropical pines has been problematic due to the floods, and structural and personnel changes within FPQ. These issues are resolved, and TREEPLAN can now be used for genetic evaluation in this hybrid tree breeding program.

Forests NSW and Forest Products Commission WA initially agreed to participate in pilot projects, but due to restructuring and resource limitations were unable to proceed. This did not adversely impact on the project, but the opportunity was lost for these organisations to become familiar with TREEPLAN and other innovations. The Researcher will continue to foster collaboration with these organisations to improve the efficiency of tree improvement nationally.

Conclusions

The exploratory dimension of the project dealt with the examination of competitive effects in forest genetics trials, in relation to their impact on the prediction of stand performance in plantations. The main conclusion reached was that there was little evidence that competitive effects are having a large impact, at least for the two species studied. Many trials were analysed using competition models and the only trials with clear, unequivocal competitive effects evident, were those trials in sites with the highest productivity, and when thinning had been delayed. Indeed, late thinning (post 18 years) and sites with very high productivity are not the norm in Australia. Therefore, as long as plantation conditions match those in genetics trials in terms of thinning regimes and productivity, there should be little concern that competitive effects are damaging our ability to predict traits such as volume on a per hectare basis at harvest age.

Nevertheless if the competitive breeding value of each individual's genotype can be predicted routinely, this would certainly benefit decision making in regards to deployment, especially of plants to sites with the highest productivity. The project team will continue to improve its understanding of competitive effects at the genetic level, by studying other species, and investigating a greater range of conditions and traits. Our expectation is that future TREEPLAN models will be able to partition the additive genetic effect into its direct and indirect (competitive) components.

The other dimension to the project has been the general improvement of functionality, to enable TREEPLAN's broader application to industry, so that it is not limited just to the analysis of STBA's core species. This dimension has been very successful and has involved rewriting large sections of existing source code and writing new source code. Testing is a very important aspect to the continual improvement of TREEPLAN. The testing phase has been streamlined and efficiencies have been generated that did not exist prior to the project starting.

Genetic evaluation on a large scale is now within reach for all major and minor plantation forest tree species, both nationally and internationally. Each species has its unique features; and we have endeavoured to put into place functionality that is broad enough to cover all bases. "Global effect" fitting is a feature of TREEPLAN V3. Essentially it means the TREEPLAN model can include an effect, which is not a peripheral feature of the trial that creates noise when attempting to predict genetic values, but is inherently important for the prediction of genetic values. Examples include: propagation effects; heterosis effects; epi-genetic effects; and inbreeding depression. Such affects get "added" on to the genetic values, for a better prediction of the deployment value of a genetic entity, such as a clone or family. If epi-genetic effects are a feature of a future TREEPLAN client species; or if exploitation of heterosis is a feature of the breeding program; TREEPLAN analysis can proceed immediately without incurring a research and development phase.

Finally the project team has put in place the provision of reliabilities for all the various types of genetic values produced by TREEPLAN, which are applied to the various types of genetic entities: genotypes; families; parental aggregates; and family mixes (seedlots). Now, for example, seed orchard managers can compare reliabilities for an economic worth index computed for seedlots. Thus if seedlot A has a slightly higher genetic value, but equivalent reliability, compared to seedlot B, and seedlot B is more easily procured, then the orchard manager has a rational basis to choose seedlot B.

The Project Manager's recommendations for any further research or actions needed to further assist commercial exploitation of the results of the Research Project

Support by FWPA through co-investment with STBA in the national tree improvement programs for plantation species can help facilitate broader access and adoption of the results and findings of the research.

Consolidation of data across programs on a species wide basis should improve efficiencies and increase the rate of genetic gain for commercially important traits. This requires cooperation among government and private entities.

The results will be used in further development of breeding (MATEPLAN) and deployment (SEEDPLAN) tools as part of the new STBA initiative with FWPA (PNC211-1011): Optimal use of genetics in deployment and tree breeding.

Using data in DATAPLAN and other trials, with geo-climatic modelling, is needed to explore and quantify the importance of genotype x environment interaction (GxE) in Australia. This will assist the STBA in defining site types and deployment zones for TREEPLAN and SEEDPLAN, ultimately to allow better matching of genetic material to environments and production systems. The Researcher has worked with CSIRO to develop a proposal for consideration by FWPA to undertake this important work.

The project has successfully demonstrated the importance of integrating molecular (SNP) data into routine evaluations with phenotypic data and other economic information. The interface between TREEPLAN genetic evaluation for commercial adoption and molecular data generated from research projects needs further development.

Data for existing and emerging species needs to be consolidated in species databases in DATAPLAN for security and future use on a national basis. More importantly, the data should be comprehensively analysed using TREEPLAN V3 to identify superior genotypes for plantation establishment.

Any other details the Company may reasonably require to be included in the report

None

Scientific papers or publications resulting from the Research Project

Dutkowski, G.W., Almquist, C., Andersson, B., Berlin, M., Ericsson, T., Högberg, K.A., Jansson, G., Karlsson, B., Kroon, J., Persson, T., Rosvall, O., Stener, L.G., Westin, J., Kerr, R.J., McRae, T.A. and Tier, B. (2009). Exploiting all your knowledge and information for breeding value prediction – experience in Norway spruce and Scots pine. Genecar meeting – the next generation in tree breeding Orenäs Slott, Sweden, 15 September 2009.

Dutkowski, G.W., Högberg, K.A., Stener, L.G. and Karlsson, B. (2009). Breeding value prediction when trails are diverse in age and genetic structure. Genecar meeting – the next generation in tree breeding Orenäs Slott, Sweden, 15 September 2009.

Li, L., Kerr, R.J., Dutkowski, G.W., Tier, B., McRae, T.A., Pilbeam, D.J. and Powell, M.B. (2009). Modelling aggregates of different genotypes in genetic evaluation of forest tree species.

McRae, T.A. (2009). TREEPLAN – breeding better genetics for today and tomorrow. News article Australian Forests and Timber News.

McRae, T.A. (2008). TREEPLAN finds top performers faster. Leading Edge article – Forest and Wood Products Australia Limited.

McRae, T.A. (2010). Using economics to match genetics to site types and end products in pine and eucalypts. Forest TECH 2010. Tools and technologies to improve forest planning and operations. Rotorua NZ and Albury Australia.

Pilbeam, D.J., McRae, T.A., Tier, B., Li, L., Dutkowski, G.W. and Kerr, R.J. (2010). STBA TREEPLAN genetic values for *Eucalyptus globulus*. Confidential technical report TR10-01. Southern Tree Breeding Association Inc.

Pilbeam, D.J., Buxton, P.A., McRae, T.A., Dutkowski, G.W., Kerr, R.J., Tier, B., and Li, L. (2011 in press). STBA TREEPLAN® genetic values for *Pinus radiata*. Confidential technical report of Southern Tree Breeding Association Inc.

Acknowledgements

The authors acknowledge the contributions of tree breeders and scientists who were responsible for establishing genetics trials and collection of performance data for various species across decades of tree improvement.

Researcher's Disclaimer

None.

Appendices

The following appendices are provided but are not for distribution at this time.

The results and findings will be reported in the published literature in due course.

Appendix A: Competition effects in forestry (24 pages)

Appendix B: Testing of the SAR[2] model (4 pages)

Appendix C: Genetics of Female Reproductive Ability in Radiata Pine (19 pages)

Appendix D: Specification of TREEPLAN models (1 page provided of 16)

Appendix E: Computing accuracies for genetic values (12 pages)

Appendix F: Optimal testing of molecular breeding (27 pages)