

# Modeling Wood Quality Using Random Regression Splines

Luis Alejandro Apiolaza

School of Forestry, University of Canterbury, Private Bag 4800, New Zealand.  
Email: Luis.Apiolaza@canterbury.ac.nz

**Abstract.** During the last five years there has been a surge of wood quality data describing variation from pith to bark. This data has been generated by low-cost repeated assessment of individuals or by SilviScan-type analyses of increment cores. Traditionally, breeders have used multiple univariate analyses to study the evolution of genetic control and estimate age-age correlations.

In this paper I random regression splines to model longitudinal data. After introducing theoretical considerations, I apply the method to two SilviScan data sets from open-pollinated progeny trials, using microfibril angle ( $^{\circ}$ ) for 188 trees from an 11 year old *Eucalyptus globulus* trial in Australia and basic density for 425 trees from an 28 year old *Pinus radiata* trial in New Zealand.

The random regressions approach permits modeling the mean wood quality curve and individual curves at the family and tree level. These curves show significant variation from the mean, which can be exploited for selection purposes, particularly when quality thresholds are important. In addition, their use produces a smoothed additive genetic covariance matrix.

## Introduction

Tree breeding aims to improve the profit of plantations by deploying superior genotypes that, in combination with appropriate site selection and silviculture, will turn into superior phenotypes. For a long time the definition of superiority (the breeding objective) was dominated by growth and form traits. However, since the 1990s wood properties have played an increasingly important role, in particular wood basic density (DEN,  $\text{kg m}^{-3}$ ), which was considered the canonical quality trait.

Plantation growers in New Zealand target mostly the structural and appearance wood radiata pine markets. Besides growth and form traits, with a big influence on wood quantity, there are other traits that affect quality. For example, in the case of structural wood there are wood stiffness requirements, a characteristic related mostly to microfibril angle (MFA,  $^{\circ}$ ) in corewood, and to MFA and DEN in outerwood. Dimensional stability is also important, where low MFA—and a small MFA gradient—together with percentage of galactan (Floyd and Stanish, 2004) are important predictors.

An interesting development has been the mounting evidence that the relationship between DEN and wood quality is weak in juvenile trees (for example, Chauhan and Walker 2006; Apiolaza 2009), exactly the time when breeders perform early selection. Therefore, response to selection for actual wood quality when selecting for DEN has been small. The situation for *Eucalyptus* solid wood production is not that dissimilar, having been dominated by DEN and with the realization that other traits (for example, MFA gradient) may be more important.

Most breeding programs have tackled wood attributes using a 'weighted average' approach. For example, the calculation of DEN by water displacement is in fact producing an average of the densities for rings 1, 2, etc. weighted by their respective volumes. Apiolaza (2009) put

forward the importance of dealing with some wood quality traits as ‘time to reach a threshold’, making possible much earlier characterization and selection of superior trees. There is, however, another family of traits for which we are interested not in their average, but in their gradient or difference between two or more assessments of the trait (an example is dimensional stability). To study these traits breeders use either repeated assessments on time or, most likely, a single sample that contains information on multiple times, like increment cores. A trajectory of wood quality traits—or successive assessment of the same individual—can be seen as a case of longitudinal data where the observations have an intrinsic order that has to be included in the model (see, for example, Apolaza and Garrick 2001).

This paper presents the analysis of wood quality trajectories using random regressions with cubic splines. The variation of trajectory is explored and families with ‘flatter’ profiles identified. Finally, the implications of the results for breeding programs are discussed.

## Materials and methods

This paper deals with two data sets previously published in partial form: Apolaza et al. (2001) presented an analysis of X-ray densitometry (obtained using SilviScan) for 425 trees from a 28 year old radiata pine progeny trial in New Zealand. However, in that analysis the authors calculated a weighted average for groups of five rings. In this paper I use individual ring DEN values. Apolaza et al. (2005) analyzed the variation of physical and chemical wood properties for 188 trees from *Eucalyptus globulus* in a field trial in North-Western Tasmania. All analyses were done using the average values for increment cores. At the same time, there was SilviScan data (that is, every 2 mm from pith to bark) for MFA, which was not included in the paper. The latter is analyzed in this article.

In breeding programs wood quality traits are sampled from genetic trials, which superimpose an experimental design and a family structure. Linear mixed models allow for a generic formulation that includes:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{u} + \mathbf{e} \quad [1]$$

where  $\mathbf{y}$  is the vector of phenotypic observations,  $\mathbf{b}$  is the vector of fixed effects,  $\mathbf{a}$  the vector of random additive genetic effects,  $\mathbf{u}$  the vector of additional random effects (e.g. blocks) and  $\mathbf{e}$  the vector of residuals.  $\mathbf{X}$ ,  $\mathbf{Z}_1$ , and  $\mathbf{Z}_2$  are incidence or linkage matrices joining the vectors to the response variable. In this context, all restrictions to randomization (e.g. blocks, plots) can be considered as random effects that do not require significance testing. In addition, genetic effects are also considered random, leaving only the intercept and locality (the latter only for the *Eucalyptus* data set) as fixed effects.

Model 1 can be extended from a single trait to a multiple trait (or multiple assessments) situation, where all the measurements of an individual are treated as a vector of observations (see Apolaza and Garrick 2001 for a detailed explanation). The additive genetic effects ( $\mathbf{a}$ ) can also be modeled with a function  $f(\mathbf{a})$  that can be expressed in a linear form. Popular choices for functional forms are orthogonal polynomials and cubic splines (for example Meyer 1998; White et al. 1999).

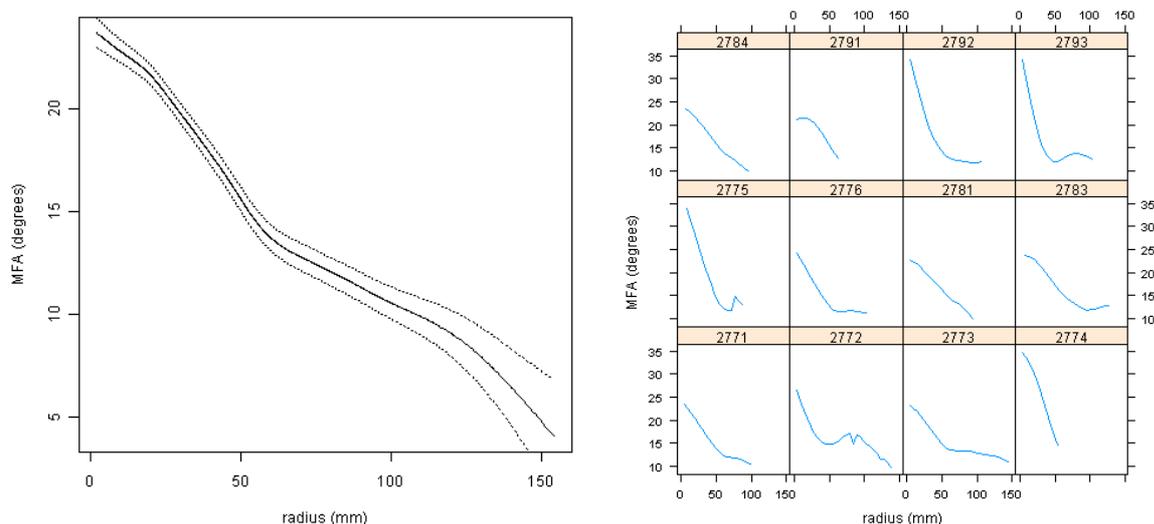
Both the trajectory and the covariance structures can be tackled in different parts of the model (either at the random effect level or at the residual level). In this paper, I will model the overall mean with a spline, tree level deviations with splines and additional effects with an autoregressive process.

All analyses were performed with a beta version of asreml-R 3 (Butler et al. 2007), which is a package for the R statistical system, based on ASReml (Gilmour et al. 2002). Model testing was performed using Akaike Information Criterion, accounting for residual maximum likelihood and number of parameters, aiming to obtain parsimonious models.

## Results and discussion

Phenotypic means for MFA ranged from 23.71 (0.73) at 2 mm from the pith to 4.05 (2.75) at 154 mm from the pith. Average MFA trends are clearly non-linear, with a sharp drop during the first 50 mm of radius, followed by a slower change from that point (see Figure 1a). Nevertheless, there is still a large amount of variation around the overall trend. Figure 1b displays splines fitted at the individual tree level for a sample of twelve individuals.

Some trees display a desirable profile for dimensional stability purposes, with a low initial MFA followed by a gradual reduction. Tree 2773, for example, starts with an MFA of 23 degrees, which gradually diminishes to 10 degrees. In contrast, other trees start with a very high MFA, followed by a very steep decline. For example, tree 2792 starts with 35 degrees and drops 20 degrees in less than 50 mm.

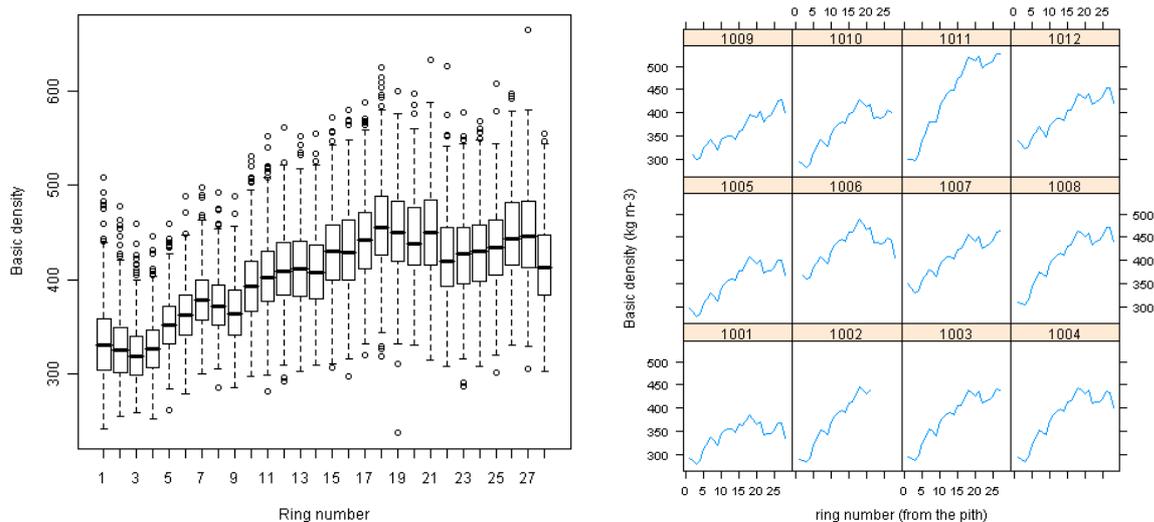


**Figure 1:** (a) population average MFA, and its confidence interval, for a *Eucalyptus globulus* progeny trial; (b) tree-level splines for a sample of 12 individuals. There is substantial variation on MFA profiles that can be used for selection purposes.

Additive genetic effects for the starting MFA value were statistically significant, as well as the linear trend of MFA reduction. In addition, tree-level random splines—expressing the deviation from the linear trend—were also significant.

Concerning the radiata pine trial, phenotypic means for DEN ranged from 321.36 (32.27) at ring 3 to 459.34 (50.43)  $\text{kg m}^{-3}$  at ring 18. There is strong non-linearity for the average trajectory for this trait, with DEN starting at 334.98 (44.37) in ring 1, then dropping to the minimum, followed by an increase at age 7 and then another drop (Figure 2).

The number of observations is the lowest at the extremes, first because of increment cores missing the pith ( $n = 369$ ) and last due to mortality ( $n = 317$ ). Most years cover 425 observations.



**Figure 2:** (a) Variability of wood basic density ( $\text{kg m}^{-3}$ ) with ring number in a *Pinus radiata* progeny trial; notice the non-linear trend and the large degree of within-year dispersion; (b) tree-level splines for a sample of 12 individuals, showing large variation for DEN profiles.

As with the previous results for MFA, there is considerable between-tree variation for both the initial value of DEN and its change from pith to bark. Once a spline was used to remove the age trend, most of the residual trends tend to be of linear or quadratic form. It would be possible to screen out trees that presented unacceptable levels of either MFA or DEN very early on, saving between 3 and 5 rings with poor quality wood. This difference would be enough to change the profitability of a given rotation age, to reduce rotation age or both.

This paper does not deal with the estimation of genetic parameters, which becomes more flexible (and more complex) than when using univariate models. Random regression models smooth covariance matrices, allowing the estimation of continuous heritabilities and correlations between any two times (see Apiolaza and Garrick 2001 for an example). However, building appropriate standard errors is difficult and results in crude approximations, which makes approaches like bootstrapping appealing, although time consuming.

The application of the methods presented in this paper relies on the availability of data-rich assessments, which is not the norm in most breeding programs. There is a trade-off between the level of detail provided by tools like SilviScan and the number of samples that a breeding program can afford to characterize. Considering that response to selection is directly proportional to selection intensity—and therefore to the number of trees that are screened for a trait—breeders should consider the contrast between increased resolution (and accuracy) for a few individuals and a larger number of individuals with lower quality information. This trade-offs prompt the question ‘Do breeders have alternatives for early screening?’

If the purpose of the assessments were to screen out trees rather than to obtain an accurate estimation of MFA, it would be possible to use simple and cheap techniques. For example, Nakada (2007) pointed out that inducing reaction wood (leaning stems) makes very easy to obtain a ‘quick and dirty’ assessment of MFA using microscopy. This approach would be very useful for clonal testing.

## Conclusions

1. Random regressions were a useful approach for dealing with pith to bark changes for wood quality traits, taking into account the longitudinal nature of the data.
2. Random splines provided a rich description of the genetic control of the trajectory of wood quality traits, including starting point, linear change and deviations from linearity.
3. There was substantial between-tree variation for MFA and DEN that can be exploited with breeding purposes. This variation is amenable to very early screening.
4. Breeders should consider the trade-offs between the level of detail and number of trees that can be screened for wood quality. Alternative, low-accuracy screening methods could be more valuable for breeding than high data density ones like SilviScan.

## Acknowledgements

David Butler (DPI) provided much support with asreml-R, both explaining technicalities and improving the software when I was hitting some limits.

## References

- Apiolaza, L.A. 2009. Very early selection for solid wood quality: screening for early winners. *Annals of Forest Science*. *In press*.
- Apiolaza, L.A. and Garrick, D.J. 2001. Analysis of longitudinal data from progeny tests: some multivariate approaches. *Forest Science* **47**: 129-140.
- Apiolaza, L.A., Raymond, C.A. and Yeo, B.J. 2005. Genetic variation of physical and chemical wood properties of *Eucalyptus globulus*. *Silvae Genetica* **54**: 160-166.
- Butler, D., Cullis, B.R., Gilmour, A.R. and Gogel, B.J. 2007. Analysis of mixed models for S language environments. ASReml-R reference manual. Training Series QE02001.
- Chauhan, S.S. and Walker J.C.F. 2006. Variations in acoustic velocity and density with age, and their interrelationships in Radiata pine. *Forest Ecology and Management* **229**: 388-394.
- Floyd, S.L. and Stanish, M.A. 2004. Methods for quantitatively determining lengthwise shrinkage in wood products. US Patent Application 10814767.457
- Gilmour, A.R., Cullis, B.R., Welham, S.J. and Thompson, R. 2002. ASReml reference manual. New South Wales Agriculture.
- Meyer, K. 1998. Estimating covariance functions for longitudinal data using a random regression model. *Genetics Selection Evolution* **30**: 221-240.
- Nakada, R. 2007. Within-tree variation of wood characteristics in conifers and the anatomical characteristics specific to very young trees. *In The Compromised Wood Workshop, Wood Technology Research Centre, Christchurch, New Zealand, 29-31 January*. 51-68.
- White, I.M.S., Thompson, R. and Brotherstone, S. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science* **88**: 632-638.