WOOD RESEARCH
62 (3): 2017
389-404

MODELLING THE EFFECT OF EUCALYPTUS GENOTYPES IN THE PULPING PROCESS WITH GENERALISED ADDITIVE MODELS AND FRACTIONAL POLYNOMIAL APPROACHES

Partson Tinarwo, Temesgen Zewotir, Delia Nort
University of Kwazulu-Natal, School of Mathematics, Statistics and Computer Science
Durban, South Africa

(Received May 2016)

ABSTRACT

The advent of lean (waste reduction), six sigma (process variation minimisation) and proper raw material selection are the essential challenges to achieve the required quality on the overall industrial processes. Accordingly a laboratory experiment for the dissolving wood pulping process was conducted on nine Eucalyptus genotypes to measure the change in lignin, viscosity and α-cellulose at each of the six pulp processing stages. The changes to these properties were modelled using the Generalised Additive Models (GAM) and Fractional Polynomial (FP) models. These models proved to be equally important in their unique ways and produced complementary results. The results revealed that Emearnsii genotype produced the best results for both α-cellulose and viscosity, while Enitens genotype was selected for the optimal lignin reduction. Egrandis genotype is the only genotype that proved to have adverse effects on the viscosity property.

KEYWORDS: Lean, six sigma, α-cellulose, dissolving pulp, Eucalyptus genotype, genotype effect, interactions, lignin, viscosity.

INTRODUCTION

Manufacturing processes aim to produce products that are of the best quality using the most efficient operational procedures with minimal costs to satisfy the customer and simultaneously yielding huge profits coupled with a well structured and best resource utilisation system (Moreno 2008). In a bid to achieve this, some companies are moving towards the implementation of such methods as lean manufacturing which aims at improving the quality by eliminating waste and

hence a reduction in downtime and costs of production (Manzouri et al. 2014). Some companies on the other hand use the combination of the waste reduction methodologies (Lean) and process variation minimisation (Six Sigma) led to the term "Lean-Six Sigma" (Pyzdek 2003). Having these systems in place complimenting each other, does not necessarily mean a manufacturing or production process runs perfectly or guarantees quality and optimal resource utilisation (Fursule et al. 2012).

In the pulping process, the α-cellulose indicates undegraded and a higher molecular weight in pulp (Tappi 1999). The Lignin removal determines the hardness, bleachability and other pulp properties (Tappi 2002). The lignin removal process degrades the cellulose molecular weight (Tappi 1999). Hence, measuring viscosity gives an average degree of cellulose polymerisation. The existence of different tree genotypes and consequently their different genetic makeup prompted the need to realise the actual influence they have on the products manufactured from them.

To understand the genotype effect Melesse and Zewotir (2013, 2015) looked at the different growth rates of the *Eucalyptus* genotypes but did not study the effect of genotypes on to the chemical properties in the pulping process. Bodhlyera (2014, 2015) outlined the effect of each sub process of the chemical pulping on the reactivity. Bodhlyera et al. (2014, 2015) did not take into account the variation within each genotype, the resultant simultaneous effect on lignin, viscosity and the α-cellulose. Kristina (2005) instead, considered a multivariate characterisation and analysis of the reactivity and spectroscopic properties in dissolving pulp revealing the short cellulose chains and low molecular weight in high reactivity pulp. Again the focus was on the viscosity alone yet solely relying on a single chemical property to determine the α-96 cellulose product quality may not suffice. Similarly, Grzeskowiak et al. (2000) predicted the pulp strength properties in eucalyptus plantations using densitometry and image analysis techniques where they found out that pulp density is correlated with bulk, burst and tensile compared with anatomical properties. However, neither the stage nor the genotypes were compared in both cases. On the other hand

Behin et al. (2008) shifted their attention to non-wood raw material (corn) to produce cellulosic dissolving pulp and this exploration could not highlight the contribution that the corn will bring to the cellulosic product quality in the pulping process.

Studies on the effect of genotype, and pulping process on the pulp chemical properties are limited. Hence this study attempted to give insight into the effect of nine genotypes on three chemical properties across the pulp processing stages within the lean six-sigma objective.

MATERIALS AND METHODS

The study design

The study was based on secondary data obtained from Council for Scientific and Industrial Research (CSIR)-Durban where the pulping and bleaching work was conducted. The project for generating the data was financially supported by CSIR and Sappi Saiccor of South Africa. The results for the pulping process were obtained experimentally in a laboratory set up.

The pulping process

The dissolving wood pulp (DWP) process can either be a sulphite or pre-hydrolysis sulphate process. The sulphite process produces pulp with cellulose content of up to 92% whereas the pre-hydrolysis sulphate process produces pulp with cellulose content of up to 96%. The study focused on the α-96 cellulose which is used to make rayon yarn for industrial products such as tire cord,

rayon staple for high-quality fabrics, and various acetate and other specialty products. The DWP undergoes six stages during processing (Tab. 1). The stage numbering is just for statistical analysis purposes only.

Tab. 1: The six stages of pulp processing.

Stage	Process	Description
1	Raw Pulp	Wood is chemically converted into pulp
2	O_2	Delignification, targeting mainly lignin removal
3	D_1	Bleaching and γ-cellulose removal together with lignin
4	E_0	Bleaching and γ-cellulose removal together with lignin
5	D_2	Bleaching and γ-cellulose removal together with lignin
6	P/H	Finishing stage where either peroxide (P) or hydroxide (H) is used to chemically peel or cut α-cellulose.

Fig. 1: The flow diagram of the different stages in the pulping process.

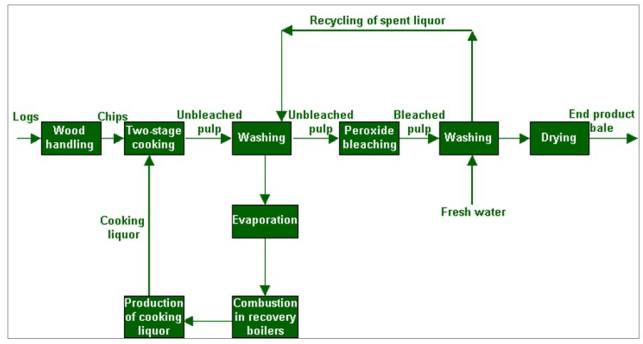

Fig. 1: The pulping process.
Source: http://home4.swipnet.se/~w-49687/broschyr.htm

From Tab. 1 and Fig. 1, the process can be classified as delignification, bleaching and finishing. Delignification is acid bisulphite pulping and takes place in a digester whereby wood chips are circulated in bubbling SO_2MgO slurry to produce cooking liquor. Bleaching and finishing entail the bleaching of the O_2 delignified pulp samples to a target of 96α grade in the following sequence: D_1 stage (ClO_2 treatment), E stage (NaOH treatment), D_2 stage (ClO_2 treatment), and a peroxide stage.

The data

The experimental results recorded viscosity, lignin, γ-cellulose, β-cellulose, α-cellulose, α+β-cellulose, copper number, glucose, xylose, mannose, klason lignin and acid solution lignin from nine genotypes namely: *E.dunnii, E.smithii, E.grandis, Macarthurii, E.mearnsii, E.nitens, GCG438, GUA380 and GUW962*. A randomisation process was employed that used 16 different trees from the nine genotypes and these were sampled from eight different climatic conditions (site qualities) ranging from warm to cold. The different samples (258 observations) were also grouped into three categories which were basically the different bleaching conditions (Fig. 2). Hence, each sample was characterised by its genotype, site quality, tree and the bleaching condition.

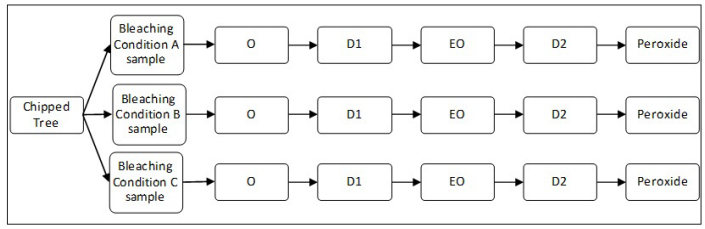

Fig. 2: Bleaching condition and the Processing stages.

Viscosity is an indirect measure of the degree of polymerization of cellulose chains in fibres and indicates the degree of chemical damage to fibres (Tappi 1994). Lignin is a complex organic polymer deposited in the cell walls of many plants, making them rigid and woody. The k-number method was used and the principle is based on the direct oxidation of lignin in pulp by standard potassium permanganate and back titrating the excess permanganate with ferrous ammonium sulphate (Mohr's salt) standard solution (Tappi 2013). The α-cellulose is based on the extraction of carbohydrates with sodium hydroxide followed by oxidation with potassium dichromate (Tappi 2000). Viscosity, lignin and α-cellulose are the core chemical properties and hence this study focussed on these three chemical properties.

RESULTS AND DISCUSSION

Fig. 3 shows the effect of each genotype on the three response variables suggesting that Generalized Additive Model and Fractional Polynomial models are likely to describe the depicted behaviour.

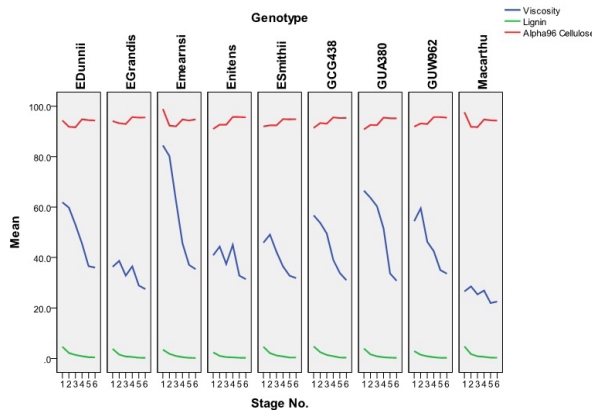

Fig. 3: Multiple line graph of lignin, viscosity and α-96 cellulose.

Generalised additive models (GAM)

A generalised additive model is a generalised linear model but the slight difference is that the linear predictor exists as a sum of smooth functions of covariates which are then considered to be related to the response variable (Wood 2006). Additive models have the best transformations that are determined simultaneously without parametric assumptions associated with their form (Faraway 2006). To establish a relationship between the mean of the response variable and the smooth function (f) of the explanatory variables, the GAM uses a link function

(Guisan et al. 2002, Zuur et al. 2007). The GAM derives its strengths from the ability to deal with highly non-linear and monotonic relationships between the response and the explanatory variables (Yee and Mitchell 1991). The technique uses an iterative procedure called a local scoring algorithm and is applicable to any likelihood-based regression model (Hastie and Tibshirani 1986). If we suppose that a response variable y is explained by predictors $x_1,..., x_p$ then the generalised additive model is given by

$$y = \beta_0 + \sum_{j=1}^{p} f_j(x_j) + \varepsilon \quad (1)$$

The **mgcv** package automatically chooses the smoothing amount (f) and has the advantage of a wider functionality (Wood 2006). The smoothing functions can have both parametric and nonparametric components that may result in semi-parametric models. Hence the semi-parametric models compromise the restrictive nature of parametric models and too much flexibility that comes with nonparametric models (Fan and Li 2004). An additive model is estimated by a penalised least squares approach which is also applied in the **mgcv** package (Wood 2006). The linear model "wiggliness" is controlled by a penalty $\lambda \int \left[f_j''(x)\right]^2 dx$ to the least squares to minimise

$$(\boldsymbol{Y} - \boldsymbol{X\beta})^T (\boldsymbol{Y} - \boldsymbol{X\beta}) + \lambda \int \left[f_j''(x)\right]^2 dx \quad (2)$$

For a mixture of categorical variables and continuous variables (1) can be modified to $y = \beta_0 + \sum_{j=1}^{p} f_j(x_j) + \varepsilon$, where design matrix **Z** denotes the variables for the non-additive part of the model. These variables can either be quantitative or qualitative and the regression parameters are then represented by **γ**.

Generalised additive model (GAM) results

For the GAM application, the processing stages were coded from 1 up to 6 for the first stage and the final stage respectively. For the parametric terms, the smoothing functions used three degrees of freedom (Tab. 2) of which several (df = 1, 2, 4 and 5) were tried but all yielded higher Akaike Information Criterion (AIC) and Generalised Cross-Validation (GCV) values.

Tab. 2: Summary of fitted GAM models.

CELLULOSE	df	F	p-value
Genotype	8	12.137	2.33e-14
ns(Stage, df = 3)	3	20.172	1.34e-11
Genotype: ns(Stage, df = 3)	24	4.422	1.04e-09
R-sq.(adj) = 0.485 Deviance explained = 55.5%			
AIC: 963.1154 GCV = 2.4815 Scale est. = 2.1353 n = 258			
VISCOSITY	df	F	p-value
Genotype	8	11.428	1.51e-13
ns(Stage, df = 3)	3	28.904	8.14e-16
Genotype: ns(Stage, df = 3)	24	2.262	0.00109
R-sq.(adj) = 0.593 Deviance explained = 64.9%			
AIC: 1959.134 GCV = 117.85 Scale est. 101.41 n = 258			
LIGNIN	df	F	p-value
Genotype	8	24.673	< 2e-16
ns(Stage, df = 3)	3	501.222	< 2e-16
Genotype: ns(Stage, df = 3)	24	5.239	5.44e-12
R-sq.(adj) = 0.942 Deviance explained = 95%			

The predictors (stage and genotype) explained 55.5% of the variation in α-96 cellulose; viscosity (64.9%) and 95% in lignin. Fig. 4 shows the predicted models together with the associated confidence intervals for the processing stage effect. Fig. 5 displays the natural spline ns( ) smoother on the vertical axis instead of the smoothing function f values. The parametric coefficients of the models provided the intercepts for α-96 cellulose, viscosity and lignin to be 94.2894; 61.8681 and 4.518202 respectively. Both lignin and viscosity levels decreased from the onset to the final stage hence all the smoothers are negative. The α-96 cellulose smoother signs alternated from negative to positive suggesting that the α-96 cellulose increases and decreases during the processing stages. Irrespective of genotype, the overall effect of genotypes per stage indicated α-96 cellulose drop in the final peroxide.

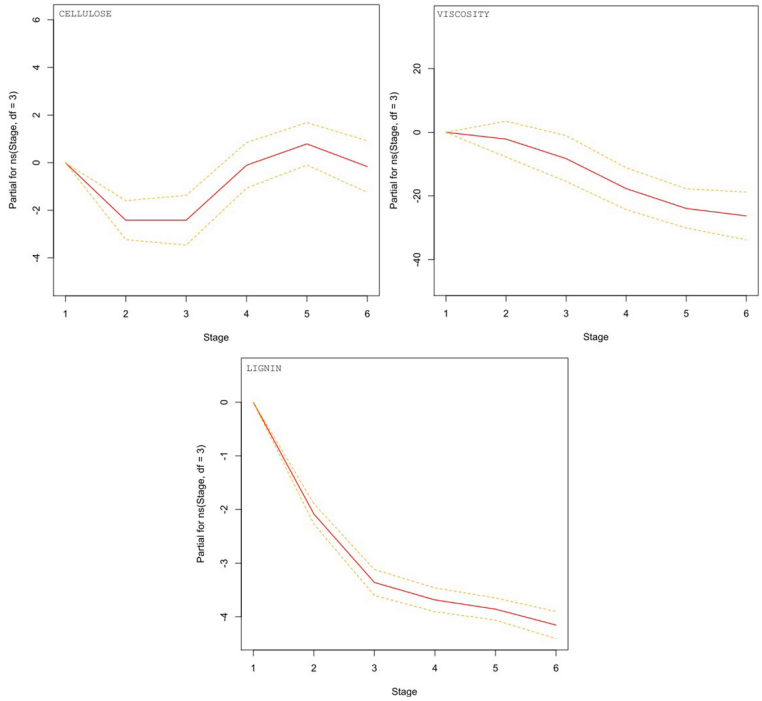

Fig. 4: Effect of stage on the responses.

Fig. 5 shows the partial effect of each genotype on all the three response variables (α-96 cellulose, viscosity and lignin). *E.mearnsii* and *Macarthurii* gave the best results for α-96 cellulose. E.mearnsii again provided better molecular weight (viscosity). Both *E.mearnsii* and *Macarthurii* have low lignin content requirements and *E.nitens* proved to be the best candidate for lignin performance followed by GUA962. Tab. 3 provides a summary of statistical tests to prove the significance of the partial effects and the interaction parametric coefficients. The t-tests show negative parametric coefficients indicating a significant decrease and similarly a positive indicating a significant increase due to the corresponding genotype either partially or in an interaction.

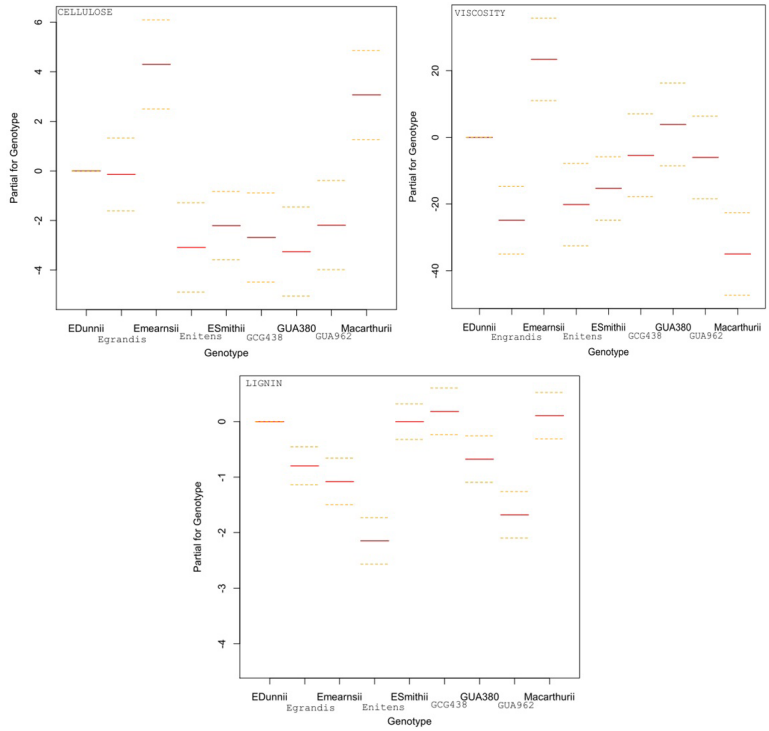

Fig. 5: Partial effect of genotype on GAM.

Tab. 3: Summary of GAM significant parametric coefficients.

Significant parametric coefficients				
Genotype	Effect	α-96 cellulose	Viscosity	Lignin
	Intercept	+94.28946***	+61.8681***	+4.518202***
E.grandis	Partial	-00.14418	-24.8681***	-0.798835***
	Interaction	+00.08067	+15.0980*	+0.522957*
E.mearnsii	Partial	+04.29638***	+23.3700***	-1.079449***
	Interaction	-00.93581	-18.8849*	+0.285214
E.nitens	Partial	-03.08706***	-20.1767**	-2.149111***
	Interaction	+01.18804	+13.4285.	+1.174368***
E.smithii	Partial	-02.20506**	-15.3525**	-0.003231
	Interaction	+00.37767	+09.0001	+0.011616
GCG438	Partial	-02.68527**	-05.3926	+0.182779
	Interaction	+00.28877	+00.7950	-0.351072
GUA380	Partial	-03.25620***	+03.8550	-0.676967**
	Interaction	+00.83352	-12.2309	+0.445468.
GUW962	Partial	-02.18720*	-06.0492	-1.679949***
	Interaction	+00.54344	+02.3362	+0.689617**
Macarthurii	Partial	-03.06249***	-34.9901***	+0.103992
	Interaction	-00.68496	+19.2814*	+0.263641
Sig. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Ideally, Tab. 3 columns for α-96 cellulose and viscosity should all be containing significant coefficients with positive signs (+) and all negatives (-) for those of lignin. Only *E.mearnsii* satisfied all the three objectives making it the best candidate for the genotype that consistently have a good effect on all the three chemical properties.

Fractional polynomial (FP) model results

Dupont (2010) affirmed that whenever the fractional polynomial models fit the data well, they will take preference than the GAM models. The FP models provide a wider range of mean functions and more so using only a few terms with similar results from the other power transformations family (Weisberg 2005). Preserving the continuous nature of the covariates in regression analysis together with the suspicion of non-linear may also call for fractional polynomials (Amber and Benner 2015). GAM always requires graphical representations to understand them but FPs are linear models of which their components can readily be interpreted. A fractional polynomial regression model takes the form

$$y = \beta_0 + \beta_1 x^p \tag{3}$$

Where y is the response variable; the covariate $x > 0$; β_j the regression coefficients and power $p \in S$ where $S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$. The order of the fractional polynomial is determined by the number of powers (p) in the model. For a single power p_1 the order is 1 and denoted by FP1 in (3) with order 2 (FP2) given by $y = \beta_0 + \beta_1 x^{p_1} + \beta_2 x^{p_2}$ or in the case of repeated powers, $y = \beta_0 + \beta_1 x^p + \beta_2 x^p \log x$ which can be generalised to the m^{th} order model (FPm) that is expressed as

$$y = \beta_0 + \sum_{j=1}^{m} \beta_j h_j(x) \tag{4}$$

Where: $h_j(x) = \begin{cases} x^{p_j} & , \quad p_j \neq p_{j-1} \\ h_{j-1}(x)\log x & , \quad p_j = p_{j-1} \end{cases}$ and $p_j \in S$

Royston and Sauerbrei (2008) indicated that and all these powers are from the set S. power forms include Box-Tidwell and exponential functions (Royston et al. (2008), this study focussed on the powers based on the set S as it contains the most commonly used powers. Fractional polynomial model requires that the covariates be positive (Royston et al. 2008). Covariate transformations are applied in some cases before model fitting (Royston and Altman 1994). The x values are scaled and centred to the form

$$x^* = \left(\frac{x \pm a}{b}\right)^p$$

to reduce numerical underflow or overflow in extreme cases. Constants a (centring) and b (scaling) are automatically determined by the software with the power p estimation not affected by the scaling. Centering should be avoided at an early stage because it produces different results (Royston et al. 2008). The suggestion was to first scale; estimate the powers and then centre. Test algorithm procedures are used for selecting the model (Meier-Hirmer et al. (2003), Sauerbrei et al. (2006), Ambler and Benner (2015)).

For the fractional polynomial models, the processing *stage* was scaled from 0 to 5 instead of 1 to 6. To solve the problem of division by zero, the centring was applied (automatically by

software) resulting in simply adding +1 to the covariate *Stage* in order to avoid the term to be undefined in the first stage when the covariate is zero. The transformed covariate (x*) becomes:

$$x^* = \left(\frac{x \pm a}{Scale}\right)^p = \left(\frac{Stage + 1}{10^1}\right)^p$$

Tab. 4 shows that using a single covariate (*Stage*) and without any interaction with the factor variable (*Genotype*), the best fitted fractional polynomial model for *α-96 cellulose* was an FP2 with $p_1 = -2$ and $p_2 = -1$. Both lignin and viscosity were modelled by FP1s with $p_1 = -0.5$ for *Lignin* and $p_1 = +1$ for *Viscosity*.

Tab. 4: Power transformations of the fitted fractional polynomials without interactions.

Coefficients:				
α-96 cellulose	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	98.93651	0.92288	107.204	< 2e-16 ***
I(((Stage + 1)/10)^-2)	0.17839	0.03582	4.981	7.65e-06 ***
I(((Stage + 1)/10)^-1)	-2.14392	0.43084	-4.976	7.77e-06 ***

AIC: 208.87	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
Viscosity	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	58.634	3.836	15.286	< 2e-16 ***
I(((Stage + 1)/10)^1)	-45.983	9.850	-4.669	2.17e-05 ***

AIC: 428.78	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
Lignin	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.8959	0.2380	-7.967	1.44e-10 ***
I(((Stage + 1)/10)^-0.5)	1.5874	0.1178	13.479	< 2e-16 ***

AIC: 92.53	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

All the model terms were significant and the corresponding diagnostic plots confirmed no violation of the assumptions underlying the model. The fitted models (without genotype interaction) were

$$\alpha - 96Cellulose = 98.93651 + 0.17839\left(\frac{Stage + 1}{10}\right)^{-2} - 2.14392\left(\frac{Stage + 1}{10}\right)^{-1}$$
$$\alpha - 96Cellulose = 98.93651 + 0.17839\,c1 - 2.14392\,c1 \quad (5)$$

where: $c1 = \left(\frac{Stage + 1}{10}\right)^{-2}$ $c2 = \left(\frac{Stage + 1}{10}\right)^{-1}$

To understand the actual behaviour of the α-96 cellulose the application of differential equations was used to determine the rate of change

$$f'_{\alpha-Cellulose} = -0.035678\left(\frac{Stage + 1}{10}\right)^{-3} + 0.214392\left(\frac{Stage + 1}{10}\right)^{-2} \quad (6)$$

Fig. 5, a plot of (6) revealed that the α-96 cellulose drops drastically with a steep gradient (-14.24 %) from 95.34 % (raw pulp) down to 92.68 % (first delignification). This also confirms

the α-96 cellulose drop at the first delignification stage noted in the GAM approach (Fig. 6), although this method could not clearly quantify the change. Thereafter, the α-96 cellulose kept on increasing but at slower rates towards the final stage with a significant increase at the first bleaching stage. The increase in the α-96 cellulose indicates that the proportion of low molecular weight carbohydrates and degraded cellulose becomes less at each subsequent stage.

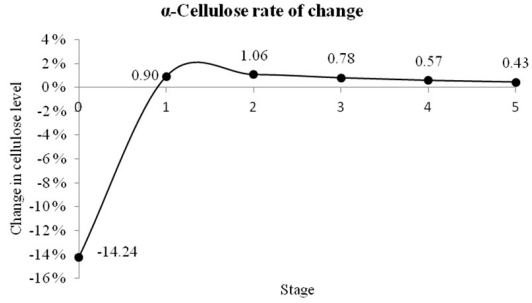

Fig. 6: Fractional polynomial model derivative plot for α-96 cellulose rate of change.

The viscosity FP1 result shows that

$$Viscosity = 58.634 - 45.983\left(\frac{Stage+1}{10}\right)^{+1}$$
$$Viscosity = 58.634 - 45.983v1 \qquad where \quad v1 = \left(\frac{Stage+1}{10}\right)^{+1}, \tag{7}$$

The behaviour of viscosity changes was described by a simple linear regression equation. On average (of all the genotypes), the initial viscosity level of 54.0357 (stage 0) drops by 4.5983 (first derivative of 7) at each stage and this translate to an average of 35.6425 (Stage 5) molecular weight that can be achieved at the final stage irrespective of the genotype effect.

$$Lignin = -1.8959 + 1.5874\left(\frac{Stage+1}{10}\right)^{-0.5}$$
$$Lignin = -1.8959 + 1.5874l1 \qquad where \qquad l1 = \left(\frac{Stage+1}{10}\right)^{-0.5} \tag{8}$$

It can be noted (8) that the content level is constantly dropping at a *k-number* of 1.8959 irrespective of the stage. There is a sharp lignin decrease in the first delignification process and the subsequent stages do not seem to have a huge impact. By also considering the differential equation of (8), gives the model for the rate of lignin change (9).

$$f'_{Lignin} = -0.07937\left(\frac{Stage+1}{10}\right)^{-\frac{3}{2}} \tag{9}$$

The greater amount of the complex organic polymer is removed in the first delignification stage than in the subsequent stages (Fig. 6).

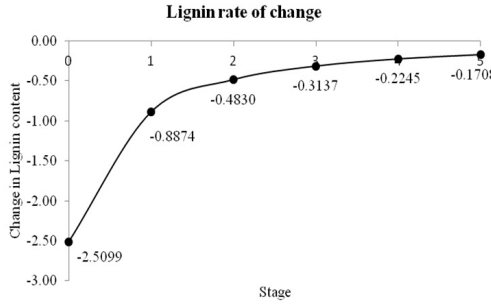

Fig. 7: Fractional polynomial model derivative plot for lignin rate of change.

The genotype interactions produced another set of parametric coefficients summarised in Tab. 5. The α-96 cellulose model and as the main process output shows that its variation is not significantly affected by the individual (partial term) genotypes. This means that the effect of each genotype on α-96 cellulose is not significantly different but some notable effects have been

Tab. 5: Parametric coefficients of the new variables in the interaction terms.

Genotype	Coefficients	α-96 cellulose	Viscosity	Lignin
	Intercept	100.19495***	+59.1000***	-2.45279***
$c1=\left(\frac{Stage+1}{10}\right)^{-2}$		+000.27185***		
$c2=\left(\frac{Stage+1}{10}\right)^{-1}$		-002.89155***		
$v1=\left(\frac{Stage+1}{10}\right)^{+1}$			-14.5714	
$l1=\left(\frac{Stage+1}{10}\right)^{-0.5}$				+2.09738***
Genotype	Coefficients	α-96 cellulose	Viscosity	Lignin
		c1 | c2	v1	l1
E.grandis	Partial	-0.64971	-014.9333.	-0.48677
	Interaction	-0.05774 +0.61179	-009.7143	+0.11821
E.mearnsii	Partial	-2.56046	+042.1667***	+1.26896*
	Interaction	-0.16746* +1.25580	-110.7500***	-0.99720***
E.nitens	Partial	-0.97223	-011.4317	+1.97506**
	Interaction	-0.14296. +0.88863	-008.5464	-1.51083***
E.smithii	Partial	-0.82818	-013.1333	-0.90997
	Interaction	-0.04559 +0.50360	-012.7143	+0.42221
GCG438	Partial	-1.12121	+007.1917	+1.09868.
	Interaction	-0.05498 +0.66274	-053.3750*	-0.62875*
GUA380	Partial	-2.13872	+011.8333	-00.3955
	Interaction	-0.18281* +1.31948	-045.7143*	-0.27542
GUW962	Partial	-0.47979	-000.66833	+0.67497
	Interaction	-0.02973 +0.34625	-030.6071	-0.65689*
Macarthurii	Partial	-2.57567	-025.2017**	+1.43025*
	Interaction	-0.15987. +1.14041	-11.2821	-1.06115***
Sig. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

realised in the interactions as explained by Grace-Martin (2000). *E.mearnsii*, *E.nitens*, *GUA380* and *Macarthurii* are significantly decreasing with the inverse of the square of Stage within each genotype.

E.grandis has got the highest impact on decreasing the overall molecular weight. *E.mearnsii* produced very high values for viscosity followed by GCG438 and GUA380. Although *E.mearnsii* viscosity values are very high above the other genotypes, they are slowly decreasing within this genotype across the stages.

Lignin fractional polynomial model shows that inclusion of the genotypes *E.mearnsii, E.nitens,* GCG438 and *Macarthurii* significantly affected the lignin content and they also have a significant variation within themselves. A correlation analysis of the regression coefficients of the significant parametric terms for *E.mearnsii, E.nitens,* GCG438 and *Macarthurii* shows that there is a strong negative correlation (-0.9693) between the partial and interaction regression coefficients. Similarily, *E.nitens,* GCG438 and *Macarthurii* are almost equally as competitive as *E.mearnsii* in the lignin reduction of dissolving pulp process.

CONCLUSIONS

The GAM approach is a convenient way of modelling that splits the data into splines such that localised models are fitted and later added together in order to give the full model. The accuracy of the model depends on the type and amount of smoothness (Wood 2006). The fitted model also comes with a confidence level which makes it very informative. The GAM smoothing fine tunes the response profile and traces the changes at each stage precisely. The application of the GAM model to the pulp dissolving process initially produced a clear profile of each individual response variable irrespective of the genotype effect.

The results show that the genotype *E.mearnsii* proved to be the most ideal candidate for producing both high α-96 cellulose proportion and viscosity property. *E.nitens* gives the product with the lowest amount of lignin content. These genotypes met the chemical property requirements as described by the Tappi methods. The actual behaviour of α-96 cellulose has been realised to drop in the final stage. The GAM partial plot for the genotype effect indicated the position of each species relative to each other and this allowed the identification of the best genotype. However, this lacked the ability to show the variation within each genotype. More so, the smoothing function in GAM involves some complex mathematical concepts (Wood 2006). The FP shifted from such complicated method and complimented the GAM efforts in the flexibility of interpreting the rate of change as well as the interaction terms which are the core area in this study. These interactions expand the understanding of the relationships among variables in a model and enable more hypotheses to be tested (Grace-Martin 2000).

The results of the FP model show that the viscosity model was linear and this confirms the random coefficient model findings by Bodhyera et al. (2014). The best genotype to produce pulp with the most desirable molecular weight was *E.mearnsii* and contrary to this, *E.grandis* is the only genotype that proved to have adverse effects on viscosity. The FP model also revealed that there is no genotype that is producing significantly better α-96 cellulose results than others. However, indicated a huge drop in α-96 cellulose in the first delignification stage and a gradual increase in the subsequent stage due to less amount of lignin being removed. Similarly, Yang and Wayman (2004) also discovered cellulose digestibility for flow through to be related to lignin removal in corn stover cellulose. The α-96 cellulose decreases significantly within each individual genotype for *E. mearnsii*, *E. nitens*, GUA380 and *Macarthurii*. The genotypes *E.mearnsii*,

E.nitens, GCG438 and *Macarthurii*. responded well to lignin reduction by their inclusion into the model and also vary significantly within themselves. The significant changes within *GUW962* had nothing to do with the overall change in the lignin content. Again, if the plantations were all favourable to *E.mearnsii* growth, this genotype would be the most commonly grown species in the world and in plantations in the vicinity of the pulp processing plants according to the fractional polynomial approach results. Such knowledge helps in growing more of what is required and the encouragement of local community engagement in forestry by providing them with free suitable tree seedlings and technical advice (Meadows 1999).

With this study having looked at different techniques to understand the genotype effect, future researchers may consider mixed models or joint models to unravel more detail on the role played by the genetic makeup of the Eucalyptus species on the pulping process.

REFERENCES

1. Ambler, G., Benner, A., 2015: Multivariate fractional polynomials. Repository CRAN. Available at https://cran.rproject.org/web/packages/mfp/vignettes/mfp_vignette.pdf
2. Behin, J., Mikaniki, F., Fadaet, Z., 2008: Dissolving pulp (alpha-cellulose) from corn stalk by kraft process, Iranian Journal of Chemical Engineering 5(3):14-28.
3. Bodhlyera, O., Zewotir T., Ramroop, S., 2014: Random coefficient model for changes in viscosity in dissolving pulp, Wood Research 59(4): 571-582.
4. Bodhlyera, O., Zewotir, T., Ramroop, S., Chunillal, V., 2015: Analysis of the changes in chemical properties of dissolving pulp during the bleaching process using piecewise linear regression models, Cellulose Chemistry and Technology 49(3): 317-332. Available at http://www.cellulosechemtechnol.ro/pdf/CCT3-4(2015)/p.317-332.pdf (Accessed 10 April 2015).
5. Dupont, W.D., 2010: Review of multivariable model-building: A Pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables, by Royston and Sauerbrei, The Stata Journal 10: 297-302.
6. Fan, J., Li, R., 2004: New estimation and model selection procedures for semi-parametric modelling in longitudinal data analysis, Journal of the American Statistical Association 99: 710-723.
7. Faraway, J.J., 2006: Extending the linear model with R. Generalised Linear, mixed effects and nonparametric regression models, Chapman & Hall/CRC, 331 pp.
8. Fursule, N.V., Bansod, S.V., Fursule, S.N., 2012: Understanding the benefits and limitations of six sigma methodology, International Journal of Scientific and Research Publications 2(1): 2250-3153.
9. Grace-Martin, K., 2000: StatNews #40: Interpreting interactions in regression. Available at https://www.cscu.cornell.edu/news/archive.php (Accessed 28 October 2015)
10. Grzeskowiak, V., Turner, P., Megown, R.A., 2000: The use of densitometry and image analysis techniques to predict pulp strength properties in Eucalyptus plantations, TAPPSA Conference African Paper week 2000 and beyond, Durban, South Africa, 17 – 20 October.
11. Guisan, A., Edwards, T., Hastie, T., 2002: Generalized linear and generalized additive models in studies of species distributions: Setting the scene, Ecological Modelling 157: 89-100.
12. Hastie, T., Tibshirani, R., 1986: Generalized additive models, Statistical Science 1(3): 297-318.

13. Kristina, E.C., 2005: Dissolving pulp-multivariate characterisation and analysis of reactivity and spectroscopic properties, Umeå University, Sweden. Available at http://www.diva-portal.org/smash/get/diva2:143367/FULLTEXT01.pdf (Accessed 18 June 2015)
14. Manzouri, M., Ab-Rahman, M.N., Zain, C.R.C.M., Jamsari, E.A., 2014: Increasing production and eliminating waste through lean tools and techniques for Halal Food Companies, Sustainability 6: 9179-9204; www.mdpi.com/journal/sustainability
15. Meadows, D.G., 1999: Growing the forest in South Africa, TAPPI Journal 82(7): 60-67
16. Meier-Hirmer, C., Ortseifen, C., Sauerbrei, W., 2003: Multivariate fractional polynomials in SAS, An algorithm for determining the transformation of continuous covariates and selection of covariates, Inst. of medical Biometry, University of Freiburg Germany, 18pp.
17. Melesse, S.M., Zewotir, T., 2013a: The effect of correlated climatic factors on the radial growth of eucalyptic trees grown in coastal Zululand of South Africa, African Journal of Agricultural Research 8(14): 1233-1244.
18. Melesse, S.M., Zewotir, T., 2013b: Path models-approach to the study of the effect of climatic factors and tree age on radial growth of juvenile Eucalyptus hybrid clones, African Journal of Agricultural Research 8(22): 2685-2695.
19. Melesse, S.M., Zewotir, T., 2015: Fitting three parameter growth curves using a nonlinear mixed effects modelling approach, South African Statistical Journal 49(2): 223-240.
20. Moreno, C.W., 2008: Optimization in production operations: Optimal "Lean Operations" in manufacturing, Ultramax Corporation, Cincinnati, Ohio, USA
21. Pyzdek, T., 2003: The six sigma handbook, A complete guide for green belts, black belts, and managers at all levels. Mc Graw Hill, 848pp.
22. Royston, P., Altman, D.G., 1994: Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion), Applied Statistics 43(3): 429-467.
23. Royston, P., Sauerbrei, W., 2008: Multivariable model-building. A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables, John Wiley & Sons, Ltd, 267pp.
24. Sauerbrei, W., Meier-Hirmer, C., Benner, A., Royston P., 2006: Multivariable regression model building by fractional polynomials: Description of SAS, STATA and R programs, Computational Statistics & Data Analysis 50: 3464-3485.
25. Tappi T222 OM-02, 2002: Acid-insoluble lignin in wood and pulp.
26. Tappi T235 OM- 60, 2000: Alkali solubility of pulp.
27. Tappi T203 CM-99, 1999: Alpha, beta and gamma cellulose in pulp.
28. Tappi UM251, 2013: Permanganate number of pulp, Useful method 251.
29. Tappi T230 OM-99, 1999: Viscosity of pulp (Capillary method).
30. Tappi T230 OM-94, 1994: Viscosity of pulp.
31. Weisberg, S., 2005: Applied linear regression, 3rd Ed. John Wiley & Sons
32. Wood, S.N., 2006: Generalised additive models: An introduction with R. Boca Raton, FL: Chapman & Hall/CRC, 410 pp.
33. Yang, B., C., 2004: Effect of xylan and lignin removal by batch and flow through pretreatment on the enzymatic digestibility of corn stover cellulose, Biotechnology and Bioengineering 86(1): 88-98
34. Yee, T.W., Mitchell, N.D., 1991: Generalized additive models in plant ecology, Journal of Vegetation Science 2: 587- 602.
35. Zuur, A.F., Ieno, E.N., Smith G.M., 2007: Analysing ecological data. Springer Verlag, 672 pp.

*Partson Tinarwo, Temesgen Zewotir, Delia Nort
University of Kwazulu-Natal
School of Mathematics, Statistics and Computer Science
Private Bag X 54001
Durban, 4000
South Africa
Phone: 031 260 301
Corresponding author: partsont@gmail.com