

# MODELING STRENGTH LOSS IN WOOD BY CHEMICAL COMPOSITION. PART I. AN INDIVIDUAL COMPONENT MODEL FOR SOUTHERN PINE

*Jerrold E. Winandy*

Research Wood Scientist

and

*Patricia K. Lebow*

Mathematical Statistician  
USDA Forest Service  
Forest Products Laboratory<sup>1</sup>  
One Gifford Pinchot Drive  
Madison, WI 53705-2398

(Received May 2000)

## ABSTRACT

In this study, we develop models for predicting loss in bending strength of clear, straight-grained pine from changes in chemical composition. Although significant work needs to be done before truly universal predictive models are developed, a quantitative fundamental relationship between changes in chemical composition and strength loss for pine was demonstrated. In particular, this study explored a linear independent-component modeling approach. The models were evaluated across a range of environmental exposure conditions known to cause strength loss and with several chemical treatments capable of causing hydrolytic chemical degradation in wood. Simple linear models developed reasonably accurate predictions of strength loss of clear, straight-grained southern pine wood based on changes in its chemical composition. Side-chain sugars of hemicellulose were the most susceptible to acid hydrolysis. The extent of their degradation was a sensitive predictor of early strength loss. Those sugars associated with the hemicellulose backbone were the next most susceptible, but they were strongly correlated between themselves. This is known as collinearity and, as such, data from either mannose or xylose, or from Klason lignin or glucose, often precluded the need for the other in the models. A linear three-parameter model using changes in a side-chain hemicellulose (arabinose), a main-chain hemicellulose (mannose), and glucose as an indicator of the extent of cellulose degradation reasonably predicted bending strength loss. We believe that with further work, residual strength or serviceability models based on a linear accumulation of the changes in chemical composition of wood during microbiological attack, thermochemical treatments, or severe environmental exposures can be developed to provide sensitive predictors of post-treatment or in-service strength loss.

*Keywords:* Strength, wood chemistry, properties.

## INTRODUCTION

For all their differences, most softwoods and hardwoods share many basic similarities in gross structure, general chemical composition, and mechanisms by which they are de-

graded. This study is the first in a series that will explore the development of predictive models that exploit these commonalities in structure, chemistry, and degradation mechanism. Our modeling approach relates proportional changes in wood strength, hereafter termed residual strength (R-ratio), to a linearized cumulative function of the fractional proportions of the individual chemical components. The R-ratio represents the proportion of remaining or residual strength after various

---

<sup>1</sup> The Forest Products Laboratory is maintained in cooperation with the University of Wisconsin. This article was written and prepared by U.S. Government employees on official time, and it is therefore in the public domain and not subject to copyright.

exposures and/or treatments. An R-ratio of 1.0 is equal to 100% of the average strength of 30 untreated, unexposed controls. An R-ratio of 0.6 means 60% of the strength is remaining (i.e., 40% strength loss). This study is part of a larger program to develop an integrated, multidisciplinary approach to understanding and preventing wood degradation.

#### BACKGROUND

The physical and mechanical properties of wood are a complex function of cellular and polymeric structure and chemistry (Winandy and Rowell 1984). The chemical composition of the softwood tracheid and the hardwood fiber cell wall depends on which cellular layer is being considered. Each layer contains varying amounts of cellulose, hemicellulose, and lignin (Fig. 1). At the polymeric level, cellulose microfibrils are thought to be encrusted in a lignin-hemicellulose matrix. It is believed that lignin, a large irregular polymer with a complex three-dimensional structure, is covalently bonded to hemicellulose (Whistler and Chen 1991). They also stated that although the hemicelluloses are not covalently associated with the cellulose microfibrils, they are closely associated by either intermixing (i.e., physical entanglement at molecular level), hydrogen bonding, or both.

Anthis (1956) showed that glucose and mannose were covalently linked as the backbone of galactoglucomannan and glucomannan hemicelluloses. Timell (1964) discovered that the structure of hemicelluloses was mostly linear with short side chains and that galactoglucomannans were the major hemicellulose in softwoods with lesser amounts of arbinoglucouronoxylans. Timell (1965) later showed that minute amounts of arbinogalactans in pines were also associated with galactoglucomannans and that in softwoods, the pentoses, such as xylans and arabinans, were more sensitive to degradation during isolation than were the hexoses. Timell (1965) also found that furan-ringed arabinans were extremely sensitive to acid hydrolysis.

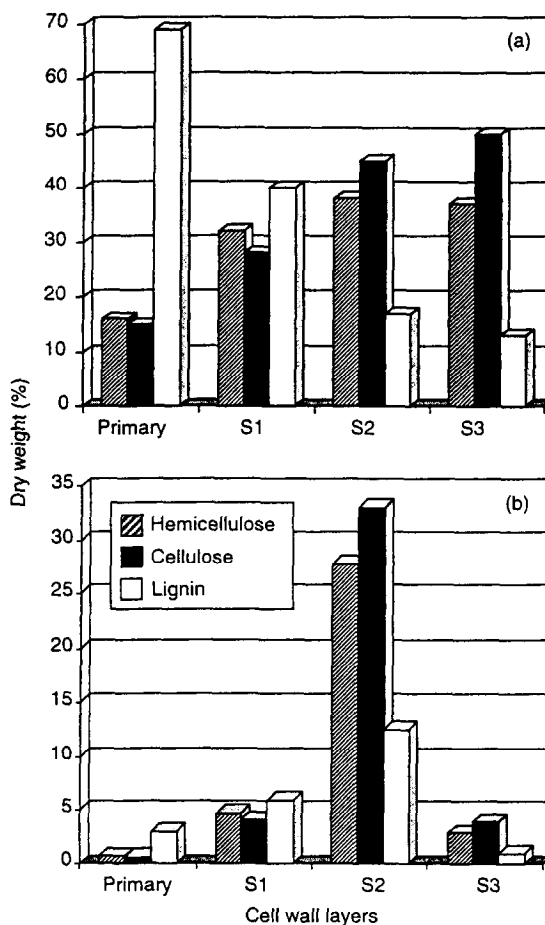


FIG. 1. Distribution of chemical components in conifers (adopted from Winandy and Rowell 1984). (a) Percentage of cell wall by layer, and (b) percentage of cell wall on total weight basis.

Hemicelluloses are generally more readily hydrolyzed by acids than cellulose because of their branched structures and their lower molecular weights (Goldstein 1991). Kolin and Danon (1998) reported that changes in physical properties, such as shrinkage and swelling coefficients, in softwoods and hardwoods were related to losses in acetyl, holocellulose, and lignin contents when these woods were exposed to increasingly more severe temperatures (20°C to 90°C).

In regards to mechanical properties, hemicelluloses are most susceptible to thermochemical degradation (Kollman and Fengel

1965). Loss of toughness in wood has been especially linked to changes in hemicellulose content (Davis and Thompson 1964). Decreases in hemicellulose content and composition of both untreated and fire retardant (FR) treated pine were directly related to early strength loss (LeVan et al. 1990). Similarly for hardwoods, subtle changes in the extractive, hemicellulose, and cellulose contents of maple caused decreased impact bending strength, static bending strength, and stiffness (Reinprecht et al. 1999).

We believe that at its earliest stages, strength loss in wood might be quantitatively modeled as a function of sequential degradation of hemicellulose side-chain sugars (arabinose, galactose) and thereafter, the main-chain hemicellulose sugars (mannose, glucose, xylose). Advanced strength loss is a function of further degradation of residual hemicellulose and initial degradation of cellulose and lignin.

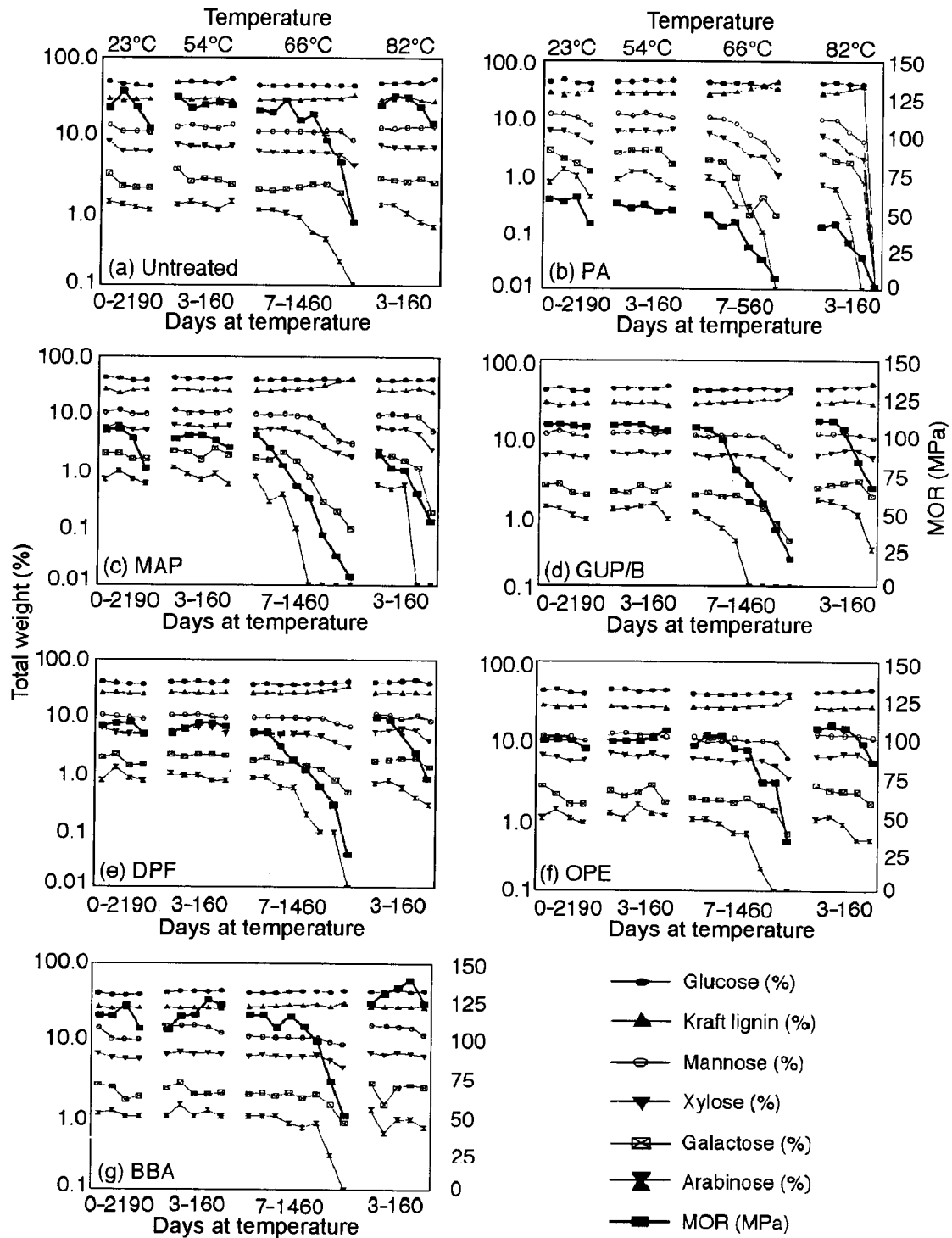
This qualitative relationship between hemicelluloses and treatment-induced or thermal-induced degradation of wood strength was recently defined in a comprehensive series of studies. The effects of six different FR treatments on the bending properties of more than 3,500 density-matched southern pine (*Pinus* spp.) specimens exposed to ambient and to elevated steady-state temperatures for up to 4 years were examined by LeVan et al. (1990), Winandy (1995), and Lebow and Winandy (1999). Increasing exposure to higher temperatures with time produced a progressive reduction in hemicellulosic sugar content, and these reductions appeared to be directly related to a corresponding loss in strength (Fig. 2). Arabinose showed the earliest direct relationship to strength loss, followed by galactose and then by mannose and xylose (LeVan et al. 1990; Winandy 1995). Generally, cellulose and lignin were not measurably affected until strength losses exceeded 30% to 40%.

This was unexpected because a commonly accepted theory suggests that the acids in wood hydrolyzed the cellulose chains, especially when accelerated by acidic chemical treatments and/or exposures to high tempera-

tures. Since cellulose is often thought to be primarily responsible for the wood fiber's strength, reducing the length of the cellulose molecules, known as degree of polymerization, should cause a reduction in macro-strength properties. This theory of hydrolytic cellulose depolymerization has been advanced by Ifju (1964) and has widely been accepted (Mark 1967; Kass et al. 1970).

Leopold and McIntosh (1961) measured the tensile strength of individual fibers, which may or may not extrapolate, to indicate the strength behavior of solid wood, and they found no relationship between degree of polymerization and strength. Ifju (1964), on the other hand, examined larger specimens (thin microtomed sections) cut into 2.5- by 100-mm rectangles. Random cellulose depolymerization was induced by gamma irradiation, followed by strength tests. Ifju postulated that the lignin should have been unaffected by the irradiation due to its aromatic structure. However, hemicellulose, which should have been at least as susceptible to radiation-induced depolymerization as cellulose, was not measured. After irradiation, the cellulose was nitrated, and isolated from the lignin and hemicellulose. While it is certainly true that a reduction in the degree of polymerization of cellulose was observed along with a reduction in tensile strength, it was unclear whether the reduction in degree of cellulose polymerization was causative or merely incidental to the strength loss. Another explanation, although unmeasured, could be that the observed strength loss was primarily caused by reductions in the degree of hemicellulose polymerization. Such an interpretation agrees with the findings of Davis and Thompson (1964) who showed that heat treatments primarily affected hemicelluloses as toughness decreased.

The next question from our previous analysis of the data (LeVan et al. 1990; Winandy 1995; and Lebow and Winandy 1999) was whether or not degree of cellulose polymerization could be reduced without measurable compositional loss in glucose. Sweet and Winandy (1999) showed that reductions in degree



of cellulose polymerization (i.e., cellulose chain length) were not occurring and thus were not related to this initial 30% to 40% strength loss, while degradation of hemicellulose was highly correlated with it. The overriding conclusion of these studies was that hemicellulose degradation alone, independent of any measurable degrade in cellulose or lignin contents or loss in degree of cellulose polymerization, accounted for initial strength losses of up to 40% for thermally degraded wood, regardless of whether it was untreated or FR-treated wood (LeVan et al. 1990; Winandy 1995; Lebow and Winandy 1999; Sweet and Winandy 1999).

It appears that interpolymeric load sharing is reduced between hemicellulose chains as side-chain constituents (arabinose and galactose) of the hemicelluloses and thereafter as main-chain hemicellulose constituents (xylose and mannose) are degraded. On further degradation, load sharing between hemicellulose and ligneous and/or cellulosic polymers is systematically reduced and overall product strength diminishes. The orderly and progressive nature of these results clearly showed that the degradation of individual hemicelluloses was qualitative. More importantly, this relationship was correlated to the degree that it might quantitatively predict incipient strength loss, especially the increased brittleness of chemically treated or thermally degraded wood.

However, it is not just thermal or hygrothermal processes that result in strength loss via hemicellulose degradation in untreated and FR-treated wood. Winandy (1994) found that thermochemical degradation of hemicelluloses was highly correlated to initial strength loss for preservative-treated wood. Winandy and

Morrell (1993) showed that arabinose followed by galactose both preferentially removed prior to measurable degradation of mannose, glucose, or xylose during the first stages of incipient brown-rot decay of Douglas-fir heartwood. This later work suggested that microbiological processes caused hydrolytic degradation patterns that were similar to thermal and/or hygrothermal hydrolytic processes with respect to the hemicellulose-strength relationship.

These studies support the qualitative concept of a universal strength degradation mechanism and hence a universal strength degrade model that might apply to both hardwoods and softwoods. At the same time, it shows a mechanistic relationship exists that can be used to model strength loss, especially during its early stages, as a function of changes in chemical composition of wood.

#### TWO PROPOSED CHEMICAL COMPOSITION-STRENGTH MODELS

From a modeling standpoint, several competing methodologies exist to predict the relationship between changes in chemical composition and wood strength. Two of these competing modeling approaches are the independent-component method and the grouped-component method. Both methods represent theoretical models based on the varying hydrolytic sensitivity of each component and the known compositional arrangements of those components within the wood cell wall.

##### *Independent-component method*

This modeling method relates the residual wood strength (R-ratio), defined as a fraction ranging from 0 to 1, to a linearized function

←

FIG. 2. Changes in chemical composition of matched southern pine specimens, untreated or treated with one of six fire-retardant chemicals, then exposed for various durations to one of four temperatures scenarios. Exposure durations were 3, 7, 21, 60, and 160 days at 23°C, 54°C, or 82°C or 7, 21, 60, 160, 290, 560, 1095, and 1460 days at 66°C. (a) untreated, (b) phosphoric acid (PA), (c) monoammonium phosphate (MAP), (d) guanylurea phosphate-boric acid (GUP/B), (e) dicyandiamide-phosphoric acid-formaldehyde (DPF), (f) organophosphonate ester (OPE), (g) borax-boric acid (BBA).

of each chemical component's fractional contribution to total weight. Each chemical component is estimated as a fractional factor of its contribution to total weight (ranging from 0 to 1). Models were analyzed both with and without selected first-order interactions. The four selected first-order interactions were selected based on known hemicellulose associations such as arabinan-xylose, glucon (in hemicellulose)-mannan, galactan-mannan, and galactan-glucon (in hemicellulose). This relationship can now broadly be defined as either

$$\begin{aligned} \text{R-ratio} = & b_0 + b_1\text{Arb} + b_2\text{Gal} + b_3\text{Xyl} \\ & + b_4\text{Man} + b_5\text{Glu} + b_6\text{Klig} \end{aligned} \quad (1)$$

or

$$\begin{aligned} \text{R-ratio} = & b_0 + b_1\text{Arb} + b_2\text{Gal} + b_3\text{Xyl} \\ & + b_4\text{Man} + b_5\text{Glu} + b_6\text{Klig} \\ & + b_{13}\text{Arb} \times \text{Xyl} + b_{24}\text{Gal} \times \text{Man} \\ & + b_{25}\text{Gal} \times \text{Glu} + b_{45}\text{Man} \times \text{Glu} \end{aligned} \quad (2)$$

where R-ratio = residual strength;  $b_i$  = least-squares fitted parameter(s); Arb = arabinose; Gal = galactose; Xyl = xylose; Man = manose; Glu = glucose; Klig = Klason lignin.

#### *Grouped-component method*

The grouped-component method relates residual strength (R-ratio) to known groupings of the individual chemical components, especially as this pertains to various hemicellulose agglomerations or types. The construction and composition of each individual carbohydrate macromolecule are assumed as described by Sjoström (1981) and Pettersen (1984). Each estimated that softwood hemicellulose consists of arabinoglucouronoxylans (AGUX at 7% to 10% total wood weight (tww)), galactose-rich galactoglucomannan (GGM at 5% to 8% tww), and galactose-poor galactoglucomannan (gGM at 10% to 15% tww).

If we limited our consideration to the relationship of strength to recognized polymeric

groupings of carbohydrates and lignin, then this relationship can broadly be defined as

$$\text{R-ratio} = f(\delta \text{AGUX}, \delta \text{GGM}, \delta \text{gGM}, \delta \text{Lig}, \delta \text{Cel}) \quad (3)$$

where R-ratio = residual strength;  $\delta$  = change in; AGUX = arabinoglucouronoxylan (1:3:13 ratio); GGM = galactose-rich galactoglucomannan (1:1:3 ratio); gGM = galactose-poor galactoglucomannan (0.1:1:4 ratio); Lig = lignin; Cel = cellulose.

Each of these subcomponents is assumed to contribute some quantity towards residual strength in direct relationship to its molar mass, spatial function, and relative accessibility. Because lignin and crystalline cellulose are less affected in the early stages of hydrolytic chemical degradation, their importance to incipient changes in strength appears minimal. Thus, because we are concerned with modeling incipient strength loss, then Eq. (3) can be rewritten as

$$\begin{aligned} @\text{R-ratio} = & \delta f(\delta \text{AGUX}, \delta \text{GGM}, \delta \text{gGM}) \\ & \text{for R-ratio} \geq 0.6 \end{aligned} \quad (4)$$

In this form, low molecular weight carbohydrates are critical predictors for early strength loss.

#### OBJECTIVE

In this report, we explored the independent-component modeling approach. The objective was to develop independent-component models and to predict strength loss from chemical compositional data. These models were evaluated across a range of environmental exposure conditions known to cause strength loss and hydrolytic chemical degradation in wood. Subsequent reports will address our long-term objectives by exploring the grouped-component method and offer comparisons to the independent-component modeling approach presented in this study.

#### METHODS

In this study, highly matched mechanical properties and chemical data from the three

prior studies were merged and quantitatively analyzed (LeVan et al. 1990; Winandy 1995; Lebow and Winandy 1999) (Fig. 2). In these three studies on defect-free, straight-grained, FR-treated southern pine (*Pinus* spp.) wood, about 4,600 specimens (9.5 by 25 by 255 mm) were sorted into 154 density-matched groups of 30 specimens. Seven FR treatments and four long-term exposure temperatures (27°C 54°C 66°C and 82°C) with durations from 3 to 160 days at 54°C and 82°C up to 4 years at 66°C and up to 6 years at 27°C were studied. The seven treatments were phosphoric-acid (PA), monoammonium-phosphate (MAP), guanylurea-phosphate/boric-acid (GUP/B), di-cyandiamide/phosphoric acid/formaldehyde (DPF), organo-phosphate-ester (OPE), borax/boric acid (BBA), and untreated (UNT). After its allotted treatment and thermal exposure, each specimen was equilibrated to constant weight at 23°C and 65% relative humidity and then destructively tested in flat-wise bending across a span of 22.9 mm using center-point loading applied at a loading rate of 4.8 mm/min. Load-deflection data were continuously recorded and used to calculate static modulus of elasticity (MOE), bending strength, which is commonly called modulus of rupture (MOR), and work to maximum load (WML). After mechanical testing, specimen moisture content and density were measured for each piece.

After static bending tests, a small section from each specimen was cut near the failure point. These sections were ground to 30 mesh (0.595-mm openings), and chemical analyses for sugars, acid soluble lignin, and Klason lignin were done generally following the procedures of Pettersen and Schwandt (1991), TAPPI Method 250 (TAPPI 1982), and Effland (1977), respectively. Individual chemical components were determined as a percentage of total wood weight.

We then used the independent-component modeling method to predict residual strength (i.e., strength loss) in southern pine wood as a function of fractional changes in chemical composition based on the individual contri-

butions of each lignocellulosic component to total weight.

#### MODELING

We evaluated the independent-component method using ordinary least-squares (OLS) regression techniques, both with and without previously obtained kinetic rate constant ( $E_a$ ) information for untreated wood and for wood treated with each of the six FR chemicals (Lebow and Winandy 1999). We also evaluated the benefits of transforming the data to facilitate model fitting using basic transformations, additions of quadratic terms, and segmentations.

Determining common relationships among groups with simple regression models involves fitting a sequence of models and then examining reductions in error-terms associated with the various models. Modeling error is often evaluated by comparing terms such as residual sum of squares, root mean square error (RMSE), and standard error of prediction ( $SE_{pred}$ ) (Draper and Smith 1998). However, as the number of predictors and the number of groups increase, the evaluation of "error" in the model-building process can become less clear. We chose to use a hierarchical approach as suggested by Draper and Smith (1998).

Initially, individual regressions were fit within each FR treatment group and examined for commonalities. Next, in our model building and analytical exercise, we used two fit-test techniques. The first technique identified which treatments could be assumed to be similar and thus analyzed together. The second technique then merged similar groups and concentrated on defining optimum model form and the importance of individual model factors.

In the first technique, we separated the observations by treatment groups. We systematically selected a training set with five or six of the seven treatment groups. Then, in a repeated series of fit-test analyses, a model, as given in either Eq. (1) or (2) was "fit" to the observations in a training set. This fitted mod-

el was then tested against a test set of the one or two groups of observations that were not included in the training set. This procedure was repeated until all combinations of fit and test groups were evaluated.

This procedure helped determine commonality between groups as well as simple model selection. To do this, subjective visual comparisons of residual plots were used to check for randomness of residual values and to check for systematic sources of bias between and within groups. Subjective comparisons of RMSE and  $SE_{\text{pred}}$  were also made. Prediction error for these models was based on an iterative "leave one observation out at a time" exclusion technique using only observations from the five or six groups included in the training set.

The second technique was intended to finalize the best form of the model, identify the appropriateness of specific factors, and parameterize those factors. We used a more formal "sample-reuse" method known as cross-validation (Davison and Hinckley 1997). Cross-validation ignored the original groupings and analyzed the data as a single universal set, enabling us to develop a more robust model allowing more universal applicability.

In our cross-validation procedure, we randomly removed about 15% to 20% of the data and evaluated fit by comparing the predicted strength values to the actual strength results of unused data to validate the models. We would then add that data back into the model, randomly remove another 15% to 20% of the data, and repeat the evaluation until each data value had been removed once and only once. This process has been used primarily for model selection with some group determination based on comparisons of prediction error (Davison and Hinckley 1997).

#### RESULTS AND DISCUSSION

The primary model (Eq. (1)) evaluated was an independent-component method using only data on Klason lignin (Klig), glucose (Glu), mannose (Man), xylose (Xyl), arabinose

(Arb), and galactose (Gal) without first-order interactions between the main factors.

Pair-wise correlations (i.e., scatterplots) provide an effective way to show the basic relationships between each of the variables. Pair-wise correlation plots, also known as scatterplots, have long been used in statistical disciplines to quickly compare the bivariate relationships between each of a large number of variables (Cleveland 1994). In Fig. 3, pair-wise correlation plots are given for all potential first-order combinations in the individual-component model. The value in the upper right and lower left corner of each column or row label represents the upper limit and lower limit, respectively, of the range for that weight-percent variable when used as either the  $x$  or the  $y$  axis depending on the individual correlation being examined. For example, the data range for Arb goes from 0 to 0.02 when used as the  $x$  axis of the Man-Arb correlation. Likewise, the range of Arb goes from 0 to 0.02 when used as the  $y$  axis of the Xyl-Arb correlation. A thorough study of all 21 correlations shown in Fig. 3 allows the reader to comprehend the interdependence between many of the variables, and these correlations will be the basis for many of the assessments made in the following discussions.

The pair-wise comparisons in Fig. 3 clearly show that Arb, Gal, Man, and Xyl are positively correlated with residual strength. As any one of these hemicellulose building blocks is degraded, that wood material's bending strength is reduced. With the exception of the PA (the most acidic group) and BBA (the only basic group) treatment groups, visual inspection of the pair-wise correlation graphs indicated similar rate relationships between residual strength and the individual chemical components for untreated and the remaining four FR treatment groups (Fig. 3).

Transformations of the predictors and residual strength to improve the models were suggested by visual analysis of the pair-wise plots. However, these transformations did not produce substantial gains in prediction or understanding, and the primary compositional



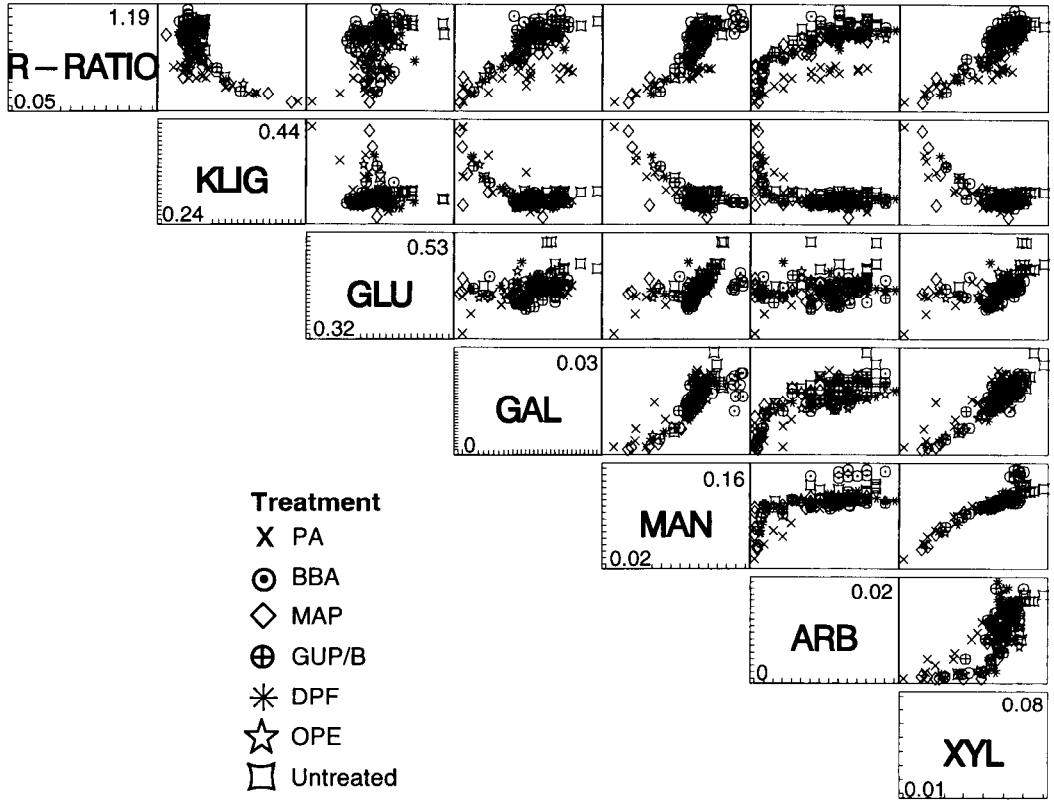


FIG. 3. Pair-wise correlation plots, also known as scatterplots, for all potential first-order combinations in the individual-component model. The value in the upper right and lower left corner of each column or row label represents the upper and lower limit, respectively, of the range for that variable for either the  $x$  or  $y$  axis depending on the individual correlation being examined. For example, the data range for arabinose (Arb) goes from 0 to 0.02 when used as the  $x$  axis of the mannose (Man)-Arb correlation or when used as the  $y$  axis of the xylose (Xyl)-Arb correlation.

data were not transformed. In fact, Arb acted as a naturally segmented predictor. A naturally segmented predictor is highly correlated to only a limited segment of the entire range of another variable. For example, in reviewing the R-ratio vs. Arb relationship (Fig. 3), it was evident that Arb was only related across a limited range of R-ratios in that Arb went from 100% to 0% composition when residual strength had not yet approached 50% (Fig. 3). Our data indicated that all the Arb had been degraded before R-ratios decreased below 0.5.

No simple correlation exists between Glu content and early strength loss (Fig. 3). Recall that most of the Glu in pine is associated primarily with cellulose. Because of cellulose's

unbranched, linear structure and its crystalline nature, its critical  $\beta$ -1-4 glucosidic linkages are inherently less accessible and accordingly more resilient than are those of the hemicelluloses (Sjostrom 1981). Thus, Glu is not easily or initially degraded.

Klig content is negatively correlated with residual strength. One reason for this is that Klig is not as susceptible to acid hydrolysis as are the carbohydrates. As the carbohydrates are selectively removed, the percentage of Klig in the residual material appears to increase, resulting in a negative correlation with residual strength (Fig. 3).

Hemicellulose carbohydrates were highly correlated with residual strength (Fig. 3). In

general terms, the R-ratio can be thought of as the ratio between a specimen's actual strength and the average strength of the untreated, unexposed controls. In Fig. 3, the R-ratio-Gal and R-ratio-Arb relationships appear linear right from the start, but the R-ratio-Xyl, R-ratio-Man, and R-ratio-Klig relationships seem unaffected during the first 10% to 20% strength loss. Arb was found to be strongly related to, and a good predictor of, early strength loss in the R-ratio range of 1.0  $\rightarrow$  0.5. Gal content was strongly related to strength across the whole specimen (R-ratio range 1.0  $\rightarrow$  0.0), but it did not predict initial strength loss nearly as well as Arb. Also, after an initial 40% strength loss, the R-ratio-Xyl and R-ratio-Man relationships each appeared to become strongly correlated to strength loss as residual strength dropped from 0.6 to 0. This was another example of naturally segmented predictors. This implies that strength loss in pine appears to have a sequential relationship with hemicellulose degradation. First Arb is affected, then Man and Xyl, which provide inference on the condition of the hemicellulose main chain. The R-ratio-Glu relationship does not appear to be as strongly related as other components. Cellulose, which is composed entirely of glucose, 'does not appear to be involved in the earliest stages of strength loss. This is further confirmed by recent work of Sweet and Winandy (1999), which indicated that degree of cellulose polymerization was not strongly related to the first 30% to 40% loss in strength. Lignin also appears to be less involved in early strength loss. These results led us to speculate that at its earliest stages, wood strength is reduced by decomposition of the branched side chains and later on by decomposition of the backbone components of hemicelluloses.

Some segregation between the treatment groups (six FR treatments and the untreated pine group) is apparent in several plots (Fig. 3). Individual regression analyses by treatment group did not select a consistent set of regression variables. However, it was felt that these differences were not significant enough to sup-

port independent models for each group when they appear visually similar. Although each individual regression had high  $R^2$  values (0.63 to 0.96), several of the independent variables (Klig, Glu, Man, and Xyl) were found to be moderately collinear. Collinearity occurs when one variable is a linear function of one or more of the others (Draper and Smith 1998). In general terms, this means that if variables A and B are collinear, they are linear functions of one another. By entering variable A into a model, this effectively also adds information from variable B into that model. In severe cases, this may completely preclude the need to add B into the model. The variables associated with the higher condition indices, a measure of collinearity, included the intercept, Klig, Glu, Xyl, and Man. These later factors are each associated with lignin, cellulose, or the backbone of the hemicelluloses and not the branched side chains of hemicellulose.

Several predictors repeatedly had high variance inflation factors. Variance inflation factors are statistical comparisons that provide a measure of the linear relationship between any one predictor and other predictor variables, excluding the intercept (Draper and Smith 1998). High variance inflation factors were noted for Klig, Gal, Xyl, and Man. The high variance inflation factor also suggests collinearity and results from the high pair-wise correlation between these four factors (Fig. 3). This may explain why Gal was not found to be a critical model parameter because using either Man or Xyl partially accounted for Gal-related modeling information.

#### *Grouped analysis*

To further our understanding, we grouped all the treatments to evaluate an aggregated relationship between the chemical components and strength. From our first analysis, we recognized that while the degradation of individual carbohydrate components of hemicellulose was related to strength loss, the specific relationships were complex. The inclusion of a covariate to standardize the influence of the

TABLE 1. Ordinary least-squares regression fit statistics for adjusted coefficient of determination ( $R^2$ ), root mean square error (RMSE), and predictive error ( $SE_{pred}$ ) based on only the six "main factor" chemical components of pine, both with and without  $E_a$ .

Excluded groups	n	Residual MOR (R-ratio)	Adjusted $R^2$		Relative RMSE (%)		Relative $SE_{pred}$ (%) <sup>a</sup>	
			Without $E_a$	With $E_a$	Without $E_a$	With $E_a$	Without $E_a$	With $E_a$
None	151	0.765	0.75	0.81	16.8	14.7	17.2	15.2
BBA	129	0.731	0.74	0.85	17.7	13.8	18.6	14.3
PA	132	0.823	0.85	0.86	10.1	9.7	10.3	10.0
BBA, PA	110	0.795	0.88	0.88	9.5	9.2	9.9	9.6

<sup>a</sup> Relative  $SE_{pred}$  is determined using an iterative "leave one observation out" withdrawal technique performed n times.

various FR treatments was considered to enhance the universality of the model. Previous work had shown that the various FR treatment chemicals affected strength loss differently. Figures 2a through 2g suggest that individual FR treatments, or more precisely the hydrolytic potential of each FR treatment to reduce strength, might act as covariates. The kinetic activation energy ( $E_a$ ) values for each FR treatment as derived by Lebow and Winandy (1999) were used to normalize the rate of strength loss data. Accordingly,  $E_a$  values for each FR treatment or for untreated wood were then added to the models (Eq. (1) and (2)). However, the  $E_a$  values were not independent of the current data set and could, as a result, yield positive results.

An analysis using the six main factors (Eq. (1)) found that  $E_a$  was sometimes a significant parameter that improved model fit as shown by adjusted  $R^2$  and standard error of prediction (Table 1). However, the inclusion of  $E_a$  primarily improved the fit of the tested models that included the PA data, evaluated with a

graphical analysis of predicted vs. residual plots (not shown).

The combined data set provided us with more observations to better explore a main-effects model that included selected first-order interactions (Eq. (2)). We evaluated this type of model with and without  $E_a$ . The addition of first-order interactions into the model did little to improve fit and appreciably increased condition indices. As with the previous main-effects model, including  $E_a$  provided substantial improvements in fit for the PA group (Table 2). However, including  $E_a$  into an analysis of main factors and first-order interactions did not substantially improve model fit for the other FR treatments or untreated wood especially when evaluated by residual plots.

Concerns about the effects of collinearity and inherent differences between treatment groups led us to evaluate the inclusion of  $E_a$  and specified interactions using the systematic fit-test scheme by group(s). This allowed us to examine the influence of the different treatment groups by pair-wise exclusion. Differ-

TABLE 2. Ordinary least-squares regression fit statistics for adjusted coefficient of determination ( $R^2$ ), root mean square error (RMSE), and predictive error ( $SE_{pred}$ ) based on only the six "main factor" chemical components of pine, selected first-order interactions, both with and without  $E_a$ .

Excluded groups	n	Residual MOR (R-ratio)	Adjusted $R^2$		Relative RMSE (%)		Relative $SE_{pred}$ (%) <sup>a</sup>	
			Without $E_a$	With $E_a$	Without $E_a$	With $E_a$	Without $E_a$	With $E_a$
None	151	0.765	0.75	0.81	16.7	14.5	18.0	15.7
BBA	129	0.731	0.74	0.84	18.0	13.9	19.9	14.9
PA	132	0.822	0.87	0.88	9.5	9.0	10.7	10.3
BBA, PA	110	0.795	0.89	0.89	9.0	8.9	10.5	10.2

<sup>a</sup> Relative  $SE_{pred}$  is determined using an iterative "leave one observation out" withdrawal technique performed n times.

ences were noted between the strongly acidic PA group and the weakly basic BBA group with the untreated and other four slightly acidic remaining groups (Fig. 3). The PA group did not exhibit a similar relationship to the other treatments in the R-ratio-Arb and R-ratio-Gal relationships. More strikingly, BBA differed from all others in all Man relationships (Fig. 3). BBA was included in the original FR-degrade experiments because the chemical behavior associated with weakly basic BBA differs from acid-based systems that use phosphates for fire retardancy. Thus, while BBA differences are not striking like the PA differences in terms of prediction, they were different enough to cause concern.

The grouped analysis primarily helped us to identify which treatment groups could be modeled together (Tables 1 and 2). The results of the grouped analysis can be summarized as follows:

- 1) BBA should not be included in cross-validation analyses in order to remove the uncertain influence it might have in the model building process because its basic chemistry is dissimilar to the acidic chemistry of the phosphate groups.
- 2)  $E_a$  should not be included in subsequent cross-validation analyses because, except for PA treatments, it did not seem to substantially improve the fit of the model of the other weakly acid FRs or untreated wood when judged by residual plots.
- 3) First-order interactions should not be included in our subsequent cross-validation analyses because they did little to improve the fit of our models and appreciably increased our condition indices.

The grouped analysis provided a data set for the cross-validation analysis, which was used to identify the preferred model form and the modeling factors and to parameterize those factors.

#### *Cross-validation analysis*

The six-fold cross-validation scheme, which randomly removed 1/6 of the data at a time,

TABLE 3. Model selection and fit statistics for six-fold cross-validation procedure based on six components without BBA observations using a "remove 1/6 of the data at a time" iterative technique.

Number of predictors	Predictors	Adjusted $R^2$	Relative $SE_{pred}$ (%)
1	Xyl	0.69	19.4
1	Man	0.65	20.8
1	Arb	0.53	24.2
1	Gal	0.49	25.1
1	Klig	0.39	27.3
1	Glu	0.12	32.8
2	Arb, Xyl	0.72	18.5
2	Glu, Xyl	0.70	19.3
2	Arb, Man	0.70	19.3
3	Arb, Xyl, Gal	0.74	18.3
3	Arb, Xyl, Glu	0.73	18.2
4	Arb, Xyl, Glu, Gal	0.74	18.2
4	Arb, Xyl, Glu, Klig	0.72	18.4
5	Arb, Xyl, Glu, Gal, Klig	0.74	18.3
5	Arb, Xyl, Man, Gal, Glu	0.74	18.4
6	All	0.74	19.0

showed minor variability on component selection but generally produced fairly consistent coefficients of determination and standard error of predictions (Eq. (1) without  $E_a$ , without interactions, without BBA, but with PA). Table 3 shows the results of such a model, the significance of various factors, and the error associated with variously factored models. The overall six-fold cross-validation scheme would choose the four-component model based on Arb-Xyl-Glu-Gal based on its having the lowest  $SE_{pred}$ , (Table 3). The top two-, three-, and five-component models would be Arb-Xyl, Arb-Xyl-Gal, and Arb-Xyl-Glu-Gal-Klig, respectively.

Recall that the PA treatment had severe effects on strength. Those strength losses approached 50% initially after treatment (Fig. 2b). The other treatments also had substantial strength loss (Fig. 2a, 2c-g), but those losses were closer to 20% to 25%, which was only half as much as the strength loss from the PA. These differences imposed much higher error for any model that included PA (compare  $SE_{pred}$  of about 18% in Table 3 (with PA) to the  $\approx 10\%$  in Table 4 (without PA)).

Excluding PA from further cross-validation

TABLE 4. Model selection and fit statistics for five-fold cross-validation procedure based on six components without BBA and PA observations using a "remove 1/5 of the data at a time" iterative withdrawal technique.

Number of predictors	Predictors	Adjusted $R^2$	Relative $SE_{pred}$ (%)
1	Xyl	0.75	13.5
1	Man	0.75	13.5
1	Gal	0.72	14.8
1	Arb	0.68	15.6
1	Klig	0.46	20.0
1	Glu	0.05	26.5
2	Arb, Man	0.86	10.4
2	Arb, Xyl	0.84	11.2
3	Arb, Man, Glu	0.87	9.8
3	Arb, Man, Klig	0.87	10.0
4	Arb, Man, Glu, Klig	0.87	9.7
4	Arb, Man, Glu, Xyl	0.88	9.8
4	Arb, Man, Glu, Gal	0.87	9.8
5	Arb, Man, Glu, Xyl, Klig	0.88	9.7
5	Arb, Man, Glu, Gal, Klig	0.88	9.7
6	All	0.88	9.7

modeling provided fewer specimens and necessitated using a five-fold cross-validation scheme (removed 1/5 of the data at a time) instead of the six-fold scheme (Eq. (1) without  $E_a$ , without interactions, without BBA, and without PA). As would be expected after removing the outlying group, the five-fold cross-validation results had lower variability and exhibited less variation in component selection. Table 4 shows the results of several multiple-factor models. The significance of various factors and the error associated with variously factored models, including the fully parame-

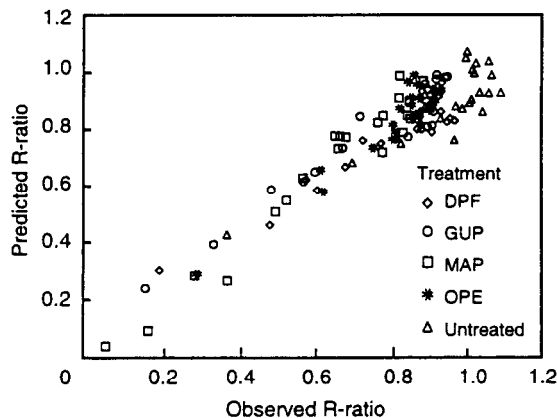


FIG. 4. Accuracy of six-parameter individual-component model as judged by predicted vs. residual plot of R-ratio for model including Klason lignin (Klig), glucose (Glu), xylose (Xyl), galactose (Gal), arabinose (Arb), and mannose (Man).

terized six-factor model, are shown. The overall five-fold cross-validation scheme, when used for model selection, would select Arb-Man, Arb-Man-Glu, Arb-Man-Glu-Klig, and Arb-Man-Glu-Xyl-Klig as the best two- to five-parameter models, respectively.

Excluding both the highly buffered BBA observations and the highly acidic PA observations simplifies the model building process and interpretation of the results. The resulting six-factor model (Klig, Glu, Man, Xyl, Gal, Arb) appeared to have very good predictive abilities with truly random-looking residuals whether judged collectively or by treatment group (Fig. 4; Table 5). However, the collin-

TABLE 5. Parameter estimates and standard errors (in parentheses) for the full (six-factor) model and the two selected three-factor models without BBA and PA observations using a "remove 1/5 of the data at a time" iterative withdrawal technique.

	All components (six factors)	Three components	Three components
Klig	-0.74 (0.45)		-1.18 (0.38)
Glu	-0.85 (0.35)	-1.11 (0.30)	
Xyl	3.09 (2.52)		
Man	4.93 (1.49)	8.19 (0.65)	5.89 (0.74)
Gal	2.35 (3.34)		
Arb	20.51 (2.79)	22.58 (2.51)	23.67 (2.54)
Intercept	0.48 (0.17)	0.24 (0.11)	0.34 (0.16)
Adjusted $R^2$	0.88	0.87	0.87
$SE_{pred}$ (%)	9.7	9.8	10.0

earity diagnostics for this six-factor model still indicated moderate collinearity between average strength, Klig, and Man. There also appears to be weaker dependency between Xyl, Man, and Gal. The individual pair-wise correlations among Xyl, Man, and Gal were high (Fig. 3). The Pearson's correlation coefficients were 0.94 for Xyl-Man, 0.87 for Man-Gal, and 0.92 for Xyl-Gal. Most importantly, Arb was strongly associated with the low order condition indices, suggesting that its importance in explaining early strength loss was not dependent upon the other factors.

In summary, several two-factor models were strong contenders compared with the larger five- and six-factor models (Table 4). However, real improvements in fit and reduced error were evident with several three-factor models. In particular, two of the three-factor models had virtually equivalent performance to the larger six-factor model. Average strength (i.e., the intercept term), Arb, either Man or Xyl (not both), and either Glu or Klig (not both) resulted in a fairly stable model with similar accuracy to the six-factor models discussed earlier (Table 4). Two of the three-parameter models were indicated by adjusted  $R^2$  and  $SE_{\text{pred}}$  as the best choices when all aspects of maximized fit and minimized prediction error were considered. Parameter estimates for the Arb-Man-Glu and Arb-Man-Klig models are given in Table 5. The residuals of the Arb-Man-Klig model were less random than those of the Arb-Man-Glu model and thus the Arb-Man-Glu model was selected as our best three-parameter model and our preferred overall model.

#### *Theoretical mechanism*

Our initial hypothetical model proposed that microbial or thermochemical degradation of solid wood occurred initially at side-chain structures, then at hemicellulose main-chain locations, and finally with cellulose and lignin. Limiting our empirical analysis of the data to the first 50% loss in original strength (i.e., from 100% to 50% residual strength) showed

that a direct relationship clearly existed between strength and mannose, galactose, xylose, and arabinose contents. From a qualitative viewpoint, Arb content, or change thereof, was the single most important predictive parameter at the earliest stages of strength loss in pine. This clearly supports our claim that degradation of hemicellulose side-chains is a primary event in early strength loss. Entering Man as the second model parameter seemed to add more than entering either Xyl or Gal separately. Because Man was highly correlated with both Xyl and Gal (Fig. 3), adding Man into the model appeared to implicitly add information from all three components (Man, Xyl, Gal). Mechanistically, it implies that the extent of degradation of the hemicellulose main chains is the next determinative event. Finally, Tables 3, 4, and 5 show that the final sequential piece of information involves the integrity of the cellulose or lignin matrix. Entering either Glu or Klig provided inference on the chemical matrix of wood. Our analysis supports using Glu as the third parameter, although Klig can also function in this role. This qualitative approach of choosing Arb-Man-Glu as the three most appropriate parameters to use in a three-parameter model also agreed with the quantitative analyses using stepwise (i.e., forward) ( $\alpha \leq 0.05$ ) and backward regressions ( $\alpha \leq 0.05$ ). An evaluation of sequential residual plots, which are qualitatively similar to a forward selection procedure, also supported this three-parameter Arb-Man-Glu model. When Glu and Klig were compared as the third parameter added to a two-parameter model, two outlying observations appeared to heavily influence the Klig relationship, whereas Glu appeared more consistent across the range of observations (Fig. 5).

In summary, a fundamental relationship exists between changes in chemical composition and strength loss for pine. Significant work still needs to be done to verify the exact relationships for other wood species, material quality levels, and degradation pathways (i.e., mechanisms) before truly robust predictive models can be developed. Robust models are

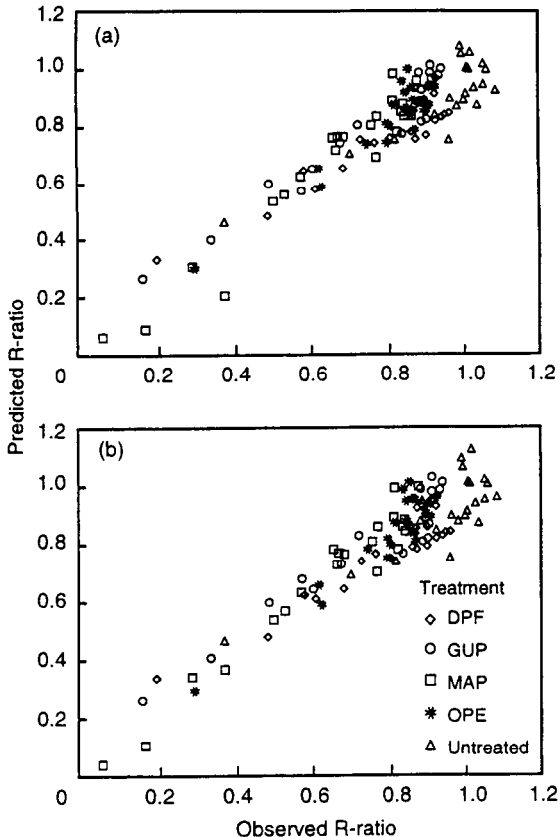


FIG. 5. Accuracy of the two best three-parameter individual-component models as judged by a plot of predicted vs. residual strengths for (a) a model including glucose (Glu), arabinose (Arb), and mannose (Man) and (b) a model including Klason lignin (Klig), arabinose (Arb), and mannose (Man).

defined as models that are reasonably reliable and accurate when applied in cases that may violate some of the original modeling assumptions. The preliminary models discussed in this report were reasonably predictive for clear, straight-grained pine with adjusted  $R^2$  values exceeding 0.75. For comparative purposes, the strength-stiffness relationship for clear, straight-grained loblolly pine from the USDA Forest Service, Forest Products Laboratory clear wood database (Forest Products Laboratory 1999) is 0.65. This strength-stiffness relationship, when applied to dimension lumber, has only an  $R^2$  value of about 0.60, but it forms the basis for a majority of the

machine-stress grading of dimension lumber in North America. Thus, the reported relationship between strength and chemical composition may have potential as a nondestructive measure of the residual strength of clear wood and possibly lumber upon further study and modeling. We are working to understand these mechanisms and develop a robust predictive model for predicting strength loss from changes in the chemical composition of wood and lumber.

#### CONCLUSIONS

The empirical analysis clearly supported the purposed theoretical model, which assumed that strength loss starts with chemical degrade at side-chain hemicellulose structures, then in the main-chain structure of hemicellulose, and finally in cellulose and lignin. A linear three-parameter model, using changes in arabinose as an indicator of degradation in side-chain hemicellulose, changes in mannose to indicate main-chain hemicellulose degradation, and changes in glucose content to indicate cellulose degradation, was found to reasonably predict bending strength loss with an  $R^2 \geq 0.75$ .

#### REFERENCES

- ANTHIS, A. 1956. Some carbohydrate linkages in slash pine alpha-cellulose. *Tappi J.* 39:401–405.
- CLEVELAND, W. S. 1994. The elements of graphing data. Hobart Press, Summit, NJ. 297 pp.
- DAVIS, W. H., AND W. S. THOMPSON. 1964. Influence of thermal treatments of short duration on the toughness and chemical composition of wood. *Forest Prod. J.* 14(8):350–356.
- DAVISON, A. C., AND D. V. HINCKLEY. 1997. Bootstrap methods and their application. Cambridge University Press, Cambridge, UK. 582 pp.
- DRAPER, N. R., AND H. SMITH. 1998. Applied regression analysis, 3rd ed. John Wiley & Sons, NY, NY. 706 pp.
- EFFLAND, M. J. 1977. Modified procedure to determine acid-insoluble lignin in wood and pulp. *Tappi J.* 60(10): 143–144.
- FOREST PRODUCTS LABORATORY. 1999. Wood handbook—Wood as an engineering material. Gen. Tech. Rep. FPL-GTR-113. USDA Forest Serv., Forest Prod. Lab. Madison, WI. 463 pp.
- GOLDSTEIN, I. S. 1991. Overview of the chemical composition of wood. Pages 1–5 in M. Lewin, and I. S. Goldstein, eds. Wood structure and composition. Inter-

- national Fiber Science and Technology Series, Vol. 11. Marcel Decker, Inc., New York, NY.
- IFJU, G. 1964. Tensile strength behavior as a function of cellulose in wood. *Forest Prod. J.* 14(8):366-372.
- KASS, A., F. F. WANGAARD, AND H. A. SCHROEDER. 1970. Chemical degradation of wood: The relationship between strength retention and pentosan content. *Wood Fiber* 2(1):31-39.
- KOLIN, B., AND G. DANON. 1998. Influence of temperature upon some physical and chemical properties of wood. *Drevarsky Vyskum* 43(3-4):21-27.
- KOLLMAN, F., AND D. FENGEL. 1965. Changes in the chemical composition of wood by thermal treatment. *Holz Roh-Werkst.* 23:461-473.
- LEBOW, P. K., AND J. E. WINANDY. 1999. Verification of kinetic models for thermal degradation of strength of fire-retardant-treated wood. *Wood Fiber Sci.* 31(1):49-61.
- LEOPOLD, B., AND D. C. MCINTOSH. 1961. Chemical composition and physical properties of wood fibers III. Tensile strength of individual fibers from alkali extracted loblolly pine holocellulose. *Tappi J.* 44(3):235-240.
- LEVAN, S. L., R. J. ROSS, AND J. E. WINANDY. 1990. Effects of fire retardant chemicals on bending properties of wood at elevated temperatures. Res. Pap. FPL-RP-498. USDA, Forest Serv. Forest Prod. Lab., Madison, WI. 24 pp.
- MARK, R. E. 1967. Cell wall mechanics of tracheids. Yale University Press, New Haven, CT. 310 pp.
- PETTERSON, R. C. 1984. The chemical composition of wood. Pages 57-126 in R. M. Rowell, ed. *The chemistry of solid wood*. Advances in chemistry series 207. American Chemical Society, Washington, DC.
- , AND V. H. SCHWANDT. 1991. Wood sugar analysis by anion chromatography. *J. Wood Chem. Technol.* 11(4):490-501.
- REINPRECHT, L., F. KACIK, AND R. SOLAR. 1999. Relationship between the molecular structure and bending properties of chemically and thermally degraded maplewood: I. Decreases in bending properties compared with changes in the basic chemical composition of wood. *Cellulose Chemistry and Technology Bucharest: Editura Academiei Romane* 33(1-2):67-79.
- SJOSTROM, E. 1981. *Wood chemistry: Fundamentals and applications*. Academic Press, New York, NY. 223 pp.
- SWEET, M. S., AND J. E. WINANDY. 1999. The influence of degree of polymerization (DP) of cellulose and hemicellulose on the strength loss of fire-retardant-treated wood. *Holzforschung* 53(3):311-317.
- TAPPI. 1982. TAPPI standards useful method 250, acid-soluble lignin in wood and pulp. Technical Association of the Pulp and Paper Industry, Atlanta, GA.
- TIMELL, T. E. 1964. Wood hemicelluloses. Part I. *Adv. Carbohydr. Chem. Biochem.* 19:247-302.
- . 1965. Wood hemicelluloses. Part II. *Adv. Carbohydr. Chem. Biochem.* 20:409-483.
- WHISTLER, R. L., AND C-C. CHEN. 1991. Hemicelluloses. Pages 287-320 in M. Lewin, and I. S. Goldstein, eds. *Wood structure and composition*. International Fiber Science and Technology Series. Vol. 11. Marcel Decker, Inc., New York, NY.
- WINANDY, J. E. 1994. Effects of long-term elevated temperature on CCA-treated southern pine lumber. *Forest Prod. J.* 44(6):51-55.
- . 1995. Effects of fire retardant treatments after 18 months of exposure at 150°F (66°C). Res. Note FPL-RN-0264. USDA Forest Serv., Forest Prod. Lab. Madison, WI. 13 pp.
- , AND J. J. MORRELL. 1993. Relationship between incipient decay, strength, and chemical composition of Douglas-fir heartwood. *Wood Fiber Sci.* 25(3):278-288.
- , AND R. M. ROWELL. 1984. The chemistry of wood strength. Pages 211-255 in R. M. Rowell, ed. *The chemistry of solid wood*. Advances in Chemistry Series 207. American Chemical Society, Washington, DC.