# CLASSIFICATION OF WOOD SURFACE FEATURES BY SPECTRAL REFLECTANCE[1]

*Patricia K. Lebow*

Mathematical Statistician
USDA Forest Service
Forest Products Laboratory
One Gifford Pinchot Drive
Madison, WI 53705

*Charles C. Brunner*

Associate Professor

*Alberto G. Maristany*

Research Scientist

Department of Forest Products
Oregon State University
Corvallis, OR 97331

and

*David A. Butler*

Professor
Department of Statistics
Oregon State University
Corvallis, OR 97331

(Received March 1995)

ABSTRACT

A database of spectral-reflectance curves of Douglas-fir veneer surface features is presented and analyzed via principal-component analysis. The paper describes how such analysis can be used to model and classify the spectral-reflectance curves by feature type. For modeling (curve-reconstruction) purposes, three principal components were sufficient by most criteria. For classification purposes, seven principal components achieved classification accuracies (with quadratic discriminant analysis) on the order of 99%, comparable to the accuracies achieved with the raw spectral data. The best seven principal components were not those associated with the largest variation in the data. This paper suggests how comparable classification accuracies might be achieved in a system operating at production speeds in a mill.

*Keywords:* Principal-component analysis, discriminant analysis, classification, spectral reflectance, color, Douglas-fir, wood.

## INTRODUCTION

The wood-products industry is increasingly adopting computer-automated manufacturing technologies to increase efficiency and recovery. Color is an important aspect of the appearance of most wood features, and color-based wood-scanning systems are being employed with increasing frequency to identify defects difficult to identify with gray-scale scanning systems.

The primary determinant of the color of an

[1] This is Paper 3067 of the Forest Research Laboratory, Oregon State University, Corvallis, OR 97331.

*Wood and Fiber Science,* 28(1), 1996, pp. 74–90

object is its spectral reflectance, that is, the percentage of incident light of each wavelength reflected from its surface. An object's spectral reflectance, measured at potentially hundreds of wavelengths, contains much more information than its chromaticity, which is usually given by three numbers (regardless of color space). The color signal output from a video camera depends upon the spectral reflectance of the object in view, but is influenced by two other factors as well: the spectral power distribution of the light source and the spectral sensitivity of the sensor. Thus, in transforming from spectral reflectance to chromaticity, much spectral information is inevitably lost, and information from the influences of lighting and sensor spectral sensitivity is added.

A knowledge of the spectral reflectances of various wood-surface features can guide the proper selection of cameras, lights, filters, and color representation to ensure that the essential information for the specific application is retained. In addition to aiding the selection of camera/lighting components, such knowledge can foster a better understanding of wood color images through the use of physical models that describe how surfaces appear.

This paper describes how principal-component analysis (PCA) can be used to model spectral-reflectance curves and to classify wood-surface features on the basis of their spectral reflectances. The literature on wood spectral reflectance and methods for analyzing spectral-reflectance curves, including PCA, is first reviewed. The set of spectral-reflectance curves, each of a common wood feature, that were collected as part of this study are then described and characterized and classified with PCA. Finally, some conclusions, including potential use of this research in a video-camera system for defect identification and classification, are offered.

## LITERATURE REVIEW

### *Wood color and wood spectral reflectance*

Color is an important aspect of the appearance of most wood features. Brunner et al. (1990) described the potential of color in applications of machine vision to wood processing, and its utility for detecting defects in images of wood has been empirically demonstrated by several researchers (Conners et al. 1985; Butler et al. 1989). Research has shown that at least two components of a color space are required to adequately separate normal from defective wood (Conners et al. 1985; Brunner et al. 1992).

The primary determinant of the color of an object is its spectral reflectance. As noted, a knowledge of the spectral reflectances of various wood-surface features can guide the proper selection of cameras, lights, filters, and color representation to ensure that the essential information for the specific application is retained. For example, Maristany et al. (1991) showed that the representation of color data affects the performance of computer vision algorithms. Their results indicated that transforming the camera's primary RGB color representation into either the CIELAB or CIELUV color spaces enhanced the performance of two tested classification routines. Experiments with filter adjustments have been successful in enhancing the detection of certain wood-surface defects (Matthews and Beech 1976; Soest 1984; Conners et al. 1985; Soest and Matthews 1985).

Solid-state video cameras can sense light of wavelengths from approximately 300 to 1,100 nm. This ability to effectively operate in the near-ultraviolet and near-infrared regions of the electromagnetic spectrum outside the visible range (360 to 830 nm) provides additional data that may improve imaging system performance. There is also potential for better understanding wood color images through physical models that describe how surfaces appear on the basis of their reflectance properties (Kanade and Ikeuchi 1991; Maristany et al. 1993, 1994).

### *Linear models*

The spectral-reflectance curves of naturally occurring materials are typically smooth (MacAdam 1981). This smoothness suggests

that they might be characterized by finite-dimensional linear models, that is, as linear combinations of a finite number of fixed curves. Let $\boldsymbol{x} = [x_1, x_2, \ldots, x_p]'$ be the individual measurements that represent a given spectral-reflectance curve. (In this paper $p = 41$, because the spectral reflectance of each sample was measured at 41 equally spaced wavelengths.) Using a finite-dimensional linear model assumes that any such $\boldsymbol{x}$ can be adequately represented as

$$\boldsymbol{x} = G\boldsymbol{\gamma} + \epsilon = \sum_{k=1}^{q} \gamma_k \boldsymbol{g}_k + \epsilon \qquad (1)$$

where $G = [\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_q]$ and $\boldsymbol{g}_k = [g_{k1}, g_{k2}, \ldots, g_{kp}]'$ and $\epsilon$ is a residual error vector. The $\boldsymbol{g}_k$ are referred to as basis functions, and each spectral-reflectance curve is approximated by a linear combination of these basis functions, with weights given by $\boldsymbol{\gamma}$. Among the assumptions of model (1) are that the errors $\epsilon$ are additive and independent. Many researchers have represented spectral reflectances by finite-dimensional linear models, including Maloney (1986), Buchsbaum and Gottschalk (1984), and Stiles et al. (1977). Choices for basis functions have included Legendre polynomials (Healey and Binford 1987), band-limited trigonometric functions (Stiles et al. 1977), unit step functions (Stiles and Wyszecki 1962), and empirical orthogonal functions (Cohen 1964).

### *Principal-component analysis*

In principal-component analysis (PCA) (Mardia et al. 1979), the basis functions $\boldsymbol{g}_k$ are chosen to be mutually orthogonal and are empirically determined from a sample of response curves $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]'$ gathered from the population of interest. (Because $G$ depends on the data, the model is nonlinear.)

PCA has been applied to many kinds of spectral data, including spectral-reflectance curves (Cohen 1964; Maloney 1986; Healey 1989; Parkkinen et al. 1989; Tominaga and Wandell 1989), spectral response curves of neurons in monkeys (Young 1986), absorbance curves (Rao 1964; Cochran and Horne 1977), transmittance curves (Parkkinen and Jaaskelainen 1987), and mass spectra (Rozett and Petersen 1976; Hoogerbrugge et al. 1983). Simonds (1963) appears to have laid much of the foundation for many of these applications when he described the use of PCA for hypothetical photographic and optical response curves.

PCA offers data reduction and, in the single-population case, the ability to transform correlated variables into uncorrelated ones. Following the terminology of Mardia et al. (1979), assume one has a random vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_p]'$ from some population with mean $\boldsymbol{\mu}$ and covariance $\Sigma$. The principal-component transformation is defined as

$$\boldsymbol{y} = \beta'(\boldsymbol{x} - \boldsymbol{\mu}) \qquad (2)$$

where $\beta = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_p]$ is a $p \times p$ orthonormal matrix ($\beta\beta' = \beta'\beta = I$) such that $\beta'\Sigma\beta = \Lambda$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$. The column vectors $\boldsymbol{\beta}_i$ are the eigenvectors of $\Sigma$, and the $\lambda_i$ are the associated eigenvalues. The $i$th principal component of $\boldsymbol{x}$ is $\boldsymbol{\beta}'_i(x - \boldsymbol{\mu})$. The eigenvector $\boldsymbol{\beta}_i$ is called the $i$th vector of principal-component loadings. Because of the way the principal components are constructed, they are uncorrelated, the variance of the ith principal component is $\lambda_i$, and no standardized linear combination of $\boldsymbol{x}$ has a variance larger than $\lambda_1$ (Mardia et al. 1979).

The principal components of a sample are defined in a manner analogous to the principal components of a population (Flury 1988). Define the sample data matrix from the population as $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]'$, where each $p$-vector $\boldsymbol{x}_i$ is an observation and $n$ is the number of observations. Let

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})' \qquad (3)$$

be the sample covariance matrix, where

$$\bar{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \qquad (4)$$

Let $VDV'$ be the spectral decomposition of $S$, i.e.,

$$S = VDV' = \sum_{k=1}^{p} \boldsymbol{d}_k \boldsymbol{v}_k \boldsymbol{v}_k' \tag{5}$$

the $i$th sample principal component (without adjustment for the mean) can be written as $\boldsymbol{u}_i = X\boldsymbol{v}_i$.

For several populations, PCA is not as well defined as when there is a single population. Assuming a common covariance structure for the populations simplifies matters, but there remains the decision of what estimator of $\Sigma$ to use (Krzanowski 1984; Manly and Rayner 1987; Flury 1988).

The application of PCA when other relationships besides equality are assumed between population covariance matrices has been investigated by several researchers. Krzanowski (1979) mathematically compared principal-component subspaces between groups to obtain a measure of similarity. Flury (1988), who has summarized much of this work, constructed a hierarchy of covariance matrix relationships that includes proportionality, common eigenvector structures, and partially common eigenvector structures.

Graphical procedures for comparison as well as formal tests for the equality of covariance matrices exist (Seber 1984). See Lebow (1992) for details.

### *Data reduction with PCA*

One of the benefits of PCA is data reduction. Such reduction is achieved by writing each observation as a linear combination of the eigenvectors $\boldsymbol{v}_k$ of $S$:

$$\boldsymbol{x}_i = \sum_{k=1}^{q} \boldsymbol{v}_k u_{ki} + \boldsymbol{\epsilon}_i \tag{6}$$

where $\boldsymbol{\epsilon}_i$ is a residual error vector and $q < p$. Replacing the $p$-dimensional vector $\boldsymbol{x}_i$ by its $q$-dimensional vector of principal-components $\boldsymbol{u}_{\cdot i}$ provides the data reduction. One attempts to choose $q$ so that the remainder, $\boldsymbol{\epsilon}_i$, can be considered noise. This modeling approach is based on the fact that the smaller components contribute less toward the total variance than do the larger ones, a relationship implying that the smaller components are of less importance and essentially noise.

If one assumes that the sample eigenvalues are in decreasing order, $d_1 \geq d_2 \geq \ldots \geq d_p \geq 0$, many procedures for the selection of $q$ exist, including:

(i) choosing $q$ by Cattell's scree test, which plots the eigenvalues (or their contribution to total variation) in decreasing order. An "elbow" in the plot, i.e., where the change in the ordered eigenvalues takes a noticeable jump, indicates the number of components, $q$, to keep (Mardia et al. 1979).

(ii) choosing the minimum $q$ such that

$$\left(\sum_{i=1}^{q} d_i\right) \Big/ \left(\sum_{i=1}^{p} d_i\right) > a \tag{7}$$

where $a$ defines some percentage (Mardia et al. 1979).

(iii) Kaiser's criterion: choosing the maximum $q$ such that

$$d_q \geq \bar{d} \tag{8}$$

where $\bar{d}$ is the average of all the eigenvalues (Mardia et al. 1979).

(iv) choosing the maximum $q$ such that the test of the hypothesis $H_0$: $\lambda_q = \lambda_{q+1} = \ldots = \lambda_p$ fails. [This test is known as the isotropy test, Barlett's sphericity test, or Barlett's test of homogeneity (Mardia et al. 1979; Flury 1988).]

(v) choosing $q$ via PRESS, [predicted residual sum of squares (Wold 1978; Krzanowski 1987)], which provides a measure of prediction error associated with a particular principal-components model. The PRESS statistics for different models are compared by an F-test or other ratio criteria to determine a suitable value for $q$.

With the exception of Barlett's test, these tests are subjective. Using these and other criteria, researchers working with relatively

smooth spectral-reflectance curves (such as those of wood) have suggested values for $q$ ranging from 3 to 9 (Cohen 1964; Maloney 1986; Parkkinen et al. 1989).

### *Discriminant analysis*

Suppose there are $g$ distinct groups of $p$-dimensional populations, each with an associated probability density $f_j(\cdot)$, and one wants to classify an unknown observation $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$ on the basis of its measurements. Discriminant analysis is the development of rules separating $\boldsymbol{R}^p$ into disjoint regions in such a way that any point $\boldsymbol{x}$ is assigned to the group $k$ having the highest likelihood in the region containing x, i.e.,

$$f_k(\boldsymbol{x}) = \max_j\{f_j(\boldsymbol{x})\} \quad (9)$$

If the populations are known to be normally distributed with means $\boldsymbol{\mu}_j$ and common covariance matrix $\Sigma$, the maximum-likelihood discriminant rule allocates $\boldsymbol{x}$ to the group that minimizes the square of the Mahalanobis distance, which is defined as follows:

$$D^2 = \min_j(\boldsymbol{x} - \boldsymbol{\mu}_j)'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_j). \quad (10)$$

This rule establishes linear boundaries separating the groups. Linear discriminant analysis represents the data in such a way that the ratio of the between-group variation to the within-group variation is nearly maximized for each pair of groups (Mardia et al. 1979).

Suppose the population parameters $\boldsymbol{\mu}_j$, $\boldsymbol{j} = \boldsymbol{1}, \boldsymbol{2}, \ldots, \boldsymbol{g}$, and $\Sigma$ are unknown, but that $n_j$ observations are made from each population $j$, $j = 1, 2, \ldots, g$. The sample maximum-likelihood discriminant rule (Mardia et al. 1979) is

$$D^2 = \min_j(\boldsymbol{x} - \bar{\boldsymbol{x}}_j)'S^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_j) \quad (11)$$

where S is the sample covariance matrix. This rule also establishes linear between-group boundaries, and so classifying observations by it is known as linear discriminant analysis.

If the populations have different covariance structures, then the estimated Mahalanobis distance is

$$D^2 = \min_j(\boldsymbol{x} - \bar{\boldsymbol{x}}_j)'S_j^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}_j) \quad (12)$$

This rule establishes quadratic boundaries separating the groups, because the intersecting contours of the covariance structures are nonlinear. Thus, classifying observations by this rule is called quadratic discriminant analysis.

### *Discriminant analysis error rates*

The maximum-likelihood discriminant rule is a special case of the Bayes discriminant rule when prior probabilities of classification are equal. In this context, the posterior probability of classification associated with allocating a random observation to a particular population can be discussed (Mardia et al. 1979). Posterior probabilities of misclassification, referred to as error rates, help assess the effectiveness of classification or discriminant rules. They are estimated by various techniques including resubstitution and cross-validation, among others.

Resubstitution procedures estimate misclassification probabilities by the proportion of observations from each population that, when classified by the sample discriminant rule, is allocated to another population. The resubstitution estimate is called an apparent error rate. Resubstitution methods tend to be overly optimistic, i.e., biased toward zero.

Cross-validation is a sample re-use method. It estimates error rates by using portions of the original sample. Cross-validation repeatedly divides the original sample into two exclusive groups, a training set and an evaluation set. The discrimination rule is calculated on the basis of the training set, and then the observations in the evaluation set are classified by this discrimination rule. The observations in the evaluation set are then added to the training set, and another subset of the training set that has yet to be evaluated becomes the new evaluation set. This process is repeated until all observations in the sample have been classified. The estimates of the posterior probabilities of misclassification are calculated by the percentages of observations that have been misallocated. When the evaluation set is a sin-

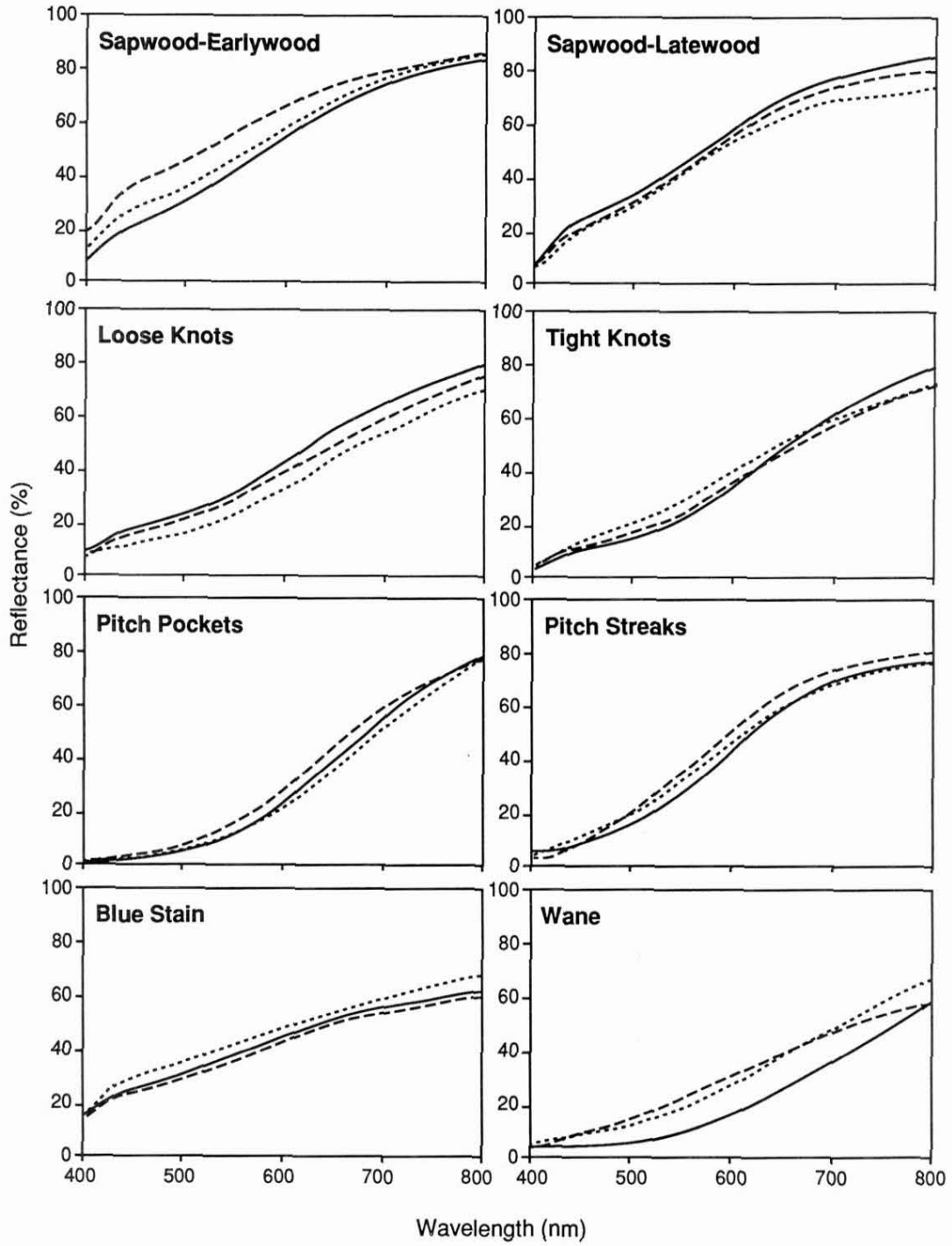


FIG. 1. Typical reflectance curves, by wood group. Each graph shows three randomly selected curves from the group.

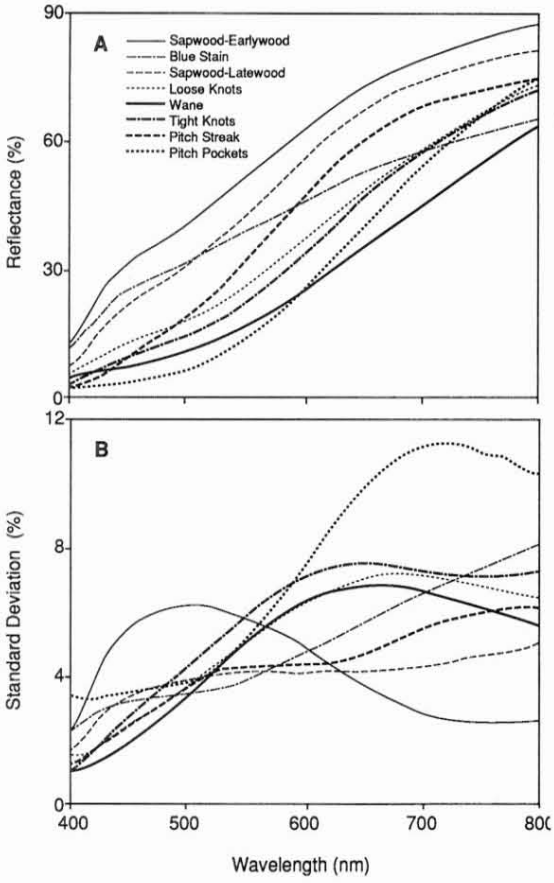


FIG. 2. (A) Mean spectral reflectance curves by wood group. (B) Standard deviations of spectral reflectance curves, by wood group.

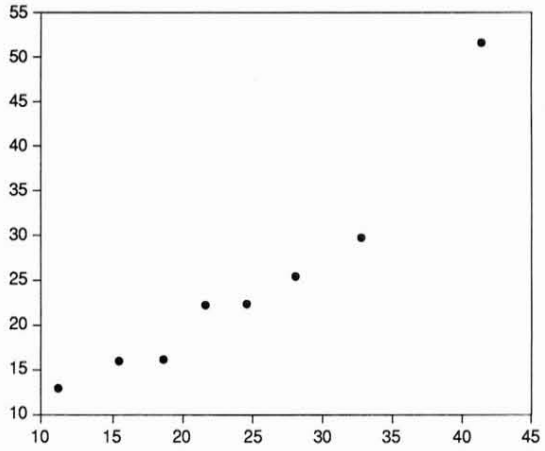


FIG. 3. Probability plot of the traces of the mean-corrected sum-of-product matrices for each wood group. Graph points are dimensionless quantities. When points fall on a straight line, it suggests that the group covariance matrices are similar.

gle observation, this method is referred to as the leave-one-out method.

Error rates (misclassification probabilities) are computed for classification rules to determine the effectiveness of a particular rule for separating groups in a data set and to compare the different rules. Sample re-use methods are generally acknowledged to be better estimators of error rates than resubstitution. However, these procedures are computationally intensive, especially with a large number of variables or observations. Snapinn and Knoke (1988) provide a good review, as well as comparison, of many estimators of misclassification error rates.

Response-curve data tend to be highly correlated, and there may be more variables than observations (Hoogerbrugge et al. 1983). Applying linear or quadratic discriminant analysis to this type of data will result in unstable discriminant-function estimates, since the discriminant functions depend on the inverse of $\Sigma$. Authors have proposed a variety of data-reduction methods prior to classification, including principal-component analysis, in order to deal with this problem. See Lebow (1992) for a discussion.

When discriminant analysis is performed on the principal components of a data set, those components associated with the most variation are not necessarily the best discriminators. Common forward variable-selection procedures in discriminant analysis were outlined in Habbema and Hermans (1977). These included Wilk's Lambda ($U$ statistic), the $F$ statistic, and maximal estimated correct classification rate.

## SPECTRAL-REFLECTANCE DATABASE

The wood spectral-reflectance curves analyzed in this paper came from eight types of surface features in Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco] veneer. The eight

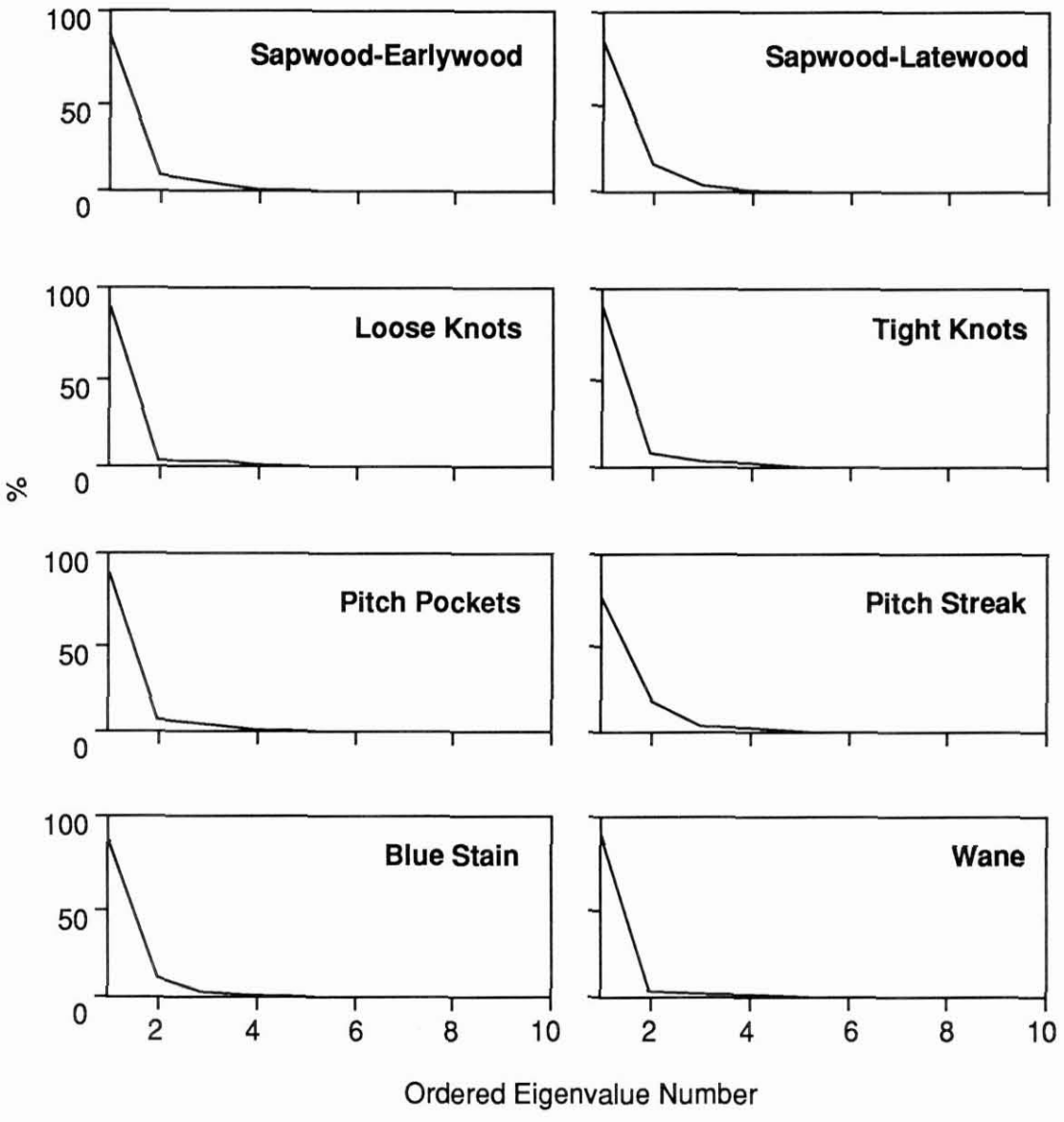


FIG. 4. Scree plots of the largest 10 (ordered) eigenvalues as percentages of the total sample variances, by wood group.

feature types were sapwood-earlywood, sapwood-latewood, loose knots, tight knots, pitch pockets, pitch streaks, blue stain, and wane (cambial interface). Five specimens were obtained for each feature type. Measurements were made at 40 random locations on each specimen, yielding a total of 1,600 spectral-reflectance curves.

Measurements were taken with a LI-COR 1800 spectroradiometer connected to a UV quartz receptor by a quartz fiber-optic light-guide. The spectroradiometer collected radiant energy from a circular area 0.9 mm in diameter. This measurement area, although smaller than those of some previous studies (Klapstein et al. 1989; Phelps et al. 1983), was chosen so

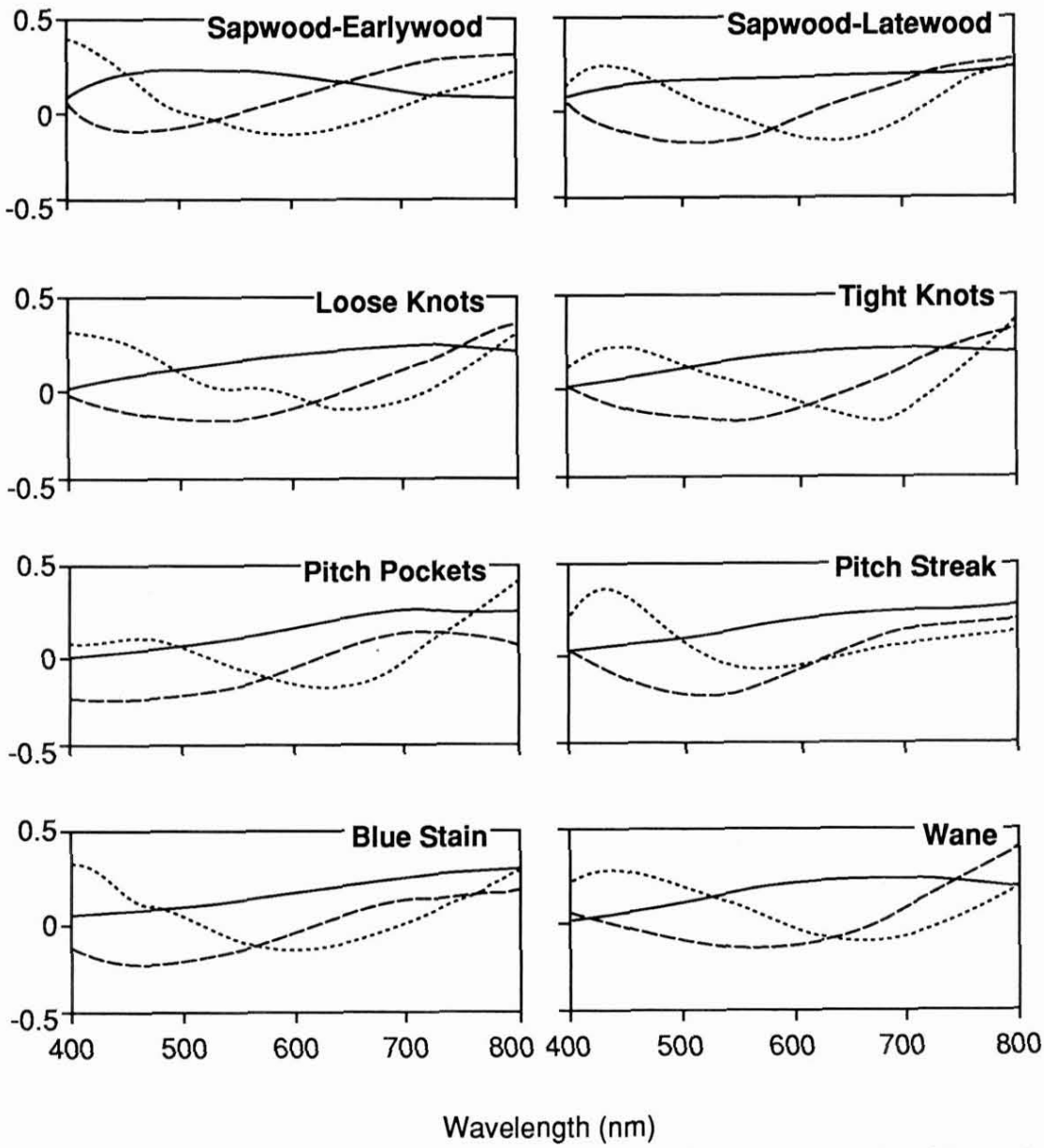


FIG. 5A. Plots of the first (solid line), second (dashed line), and third (dotted line) eigenvectors of the sample covariance matrices, by wood group.

that the spectroradiometer could focus on single earlywood/latewood bands. Reflectance curves from the veneer specimens were obtained with a standard 45/0 geometry, in which each sample was illuminated with an incandescent light source at a 45° angle to the surface normal (Hunter and Harold 1987). Measurements were made at the normal, with illumination parallel to the wood-fiber direction. All wood measurements were preceded by a reference measurement from a Labsphere Spectralon℠ diffuse reflectance standard. The spectral-reflectance factor for each wavelength was obtained by dividing wood energy measure-

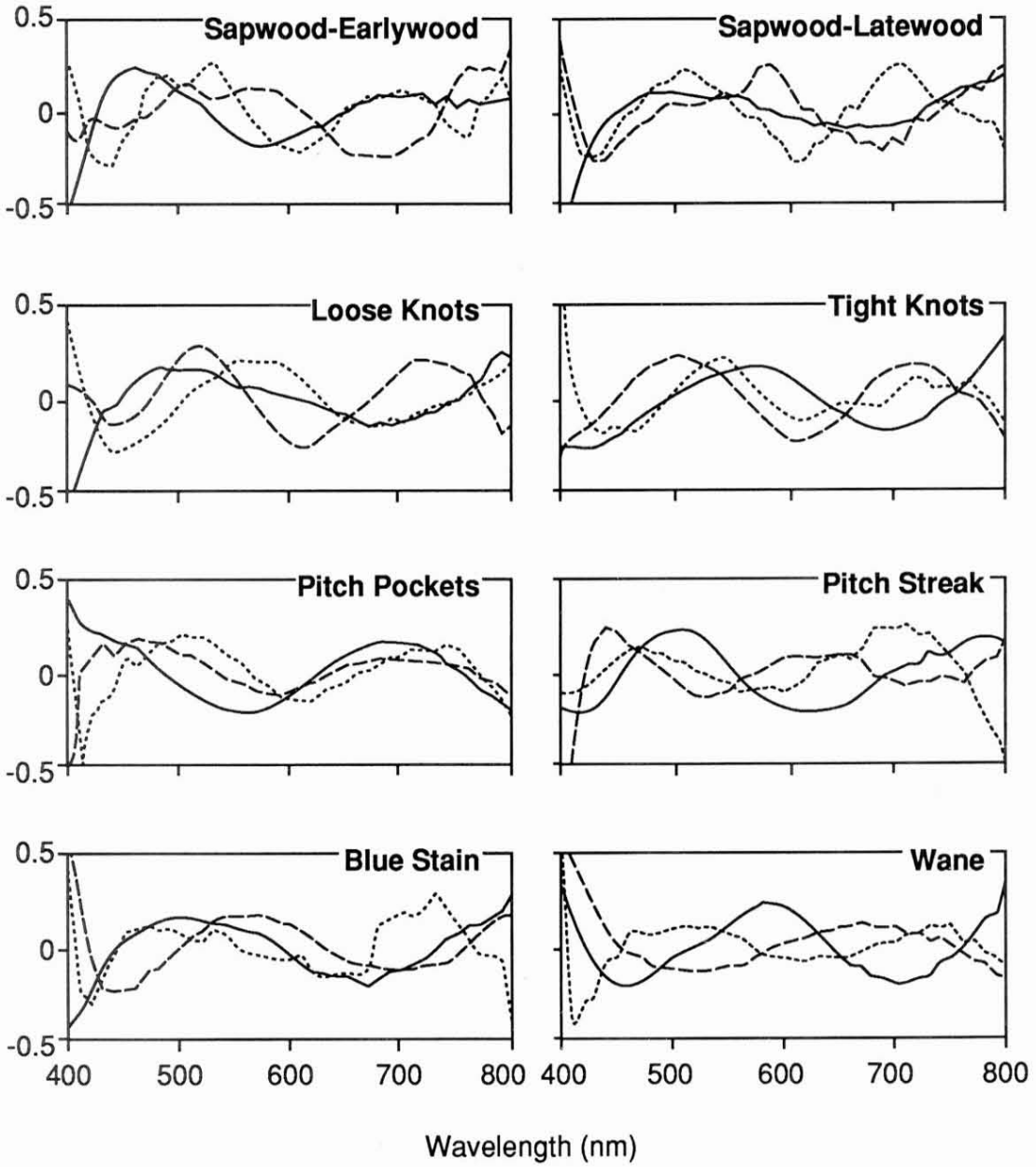


FIG. 5B. Plots of the fourth (solid line), fifth (dashed line), and sixth (dotted line) eigenvectors of the sample covariance matrices, by wood group. Y-ordinates are dimensionless quantities. Eigenvectors shown have unit Euclidean norm.

ments by reference energy measurements at each wavelength from 400 to 800 nm.

Although the spectroradiometer used to obtain the curves was capable of measuring at 1-nm increments, a preliminary experiment showed that the wood-feature curves were smooth enough that measuring at 10-nm increments was sufficient. Linearly interpolating between the data points of the 10-nm curves produced an RMS error of 1.2%, relative to

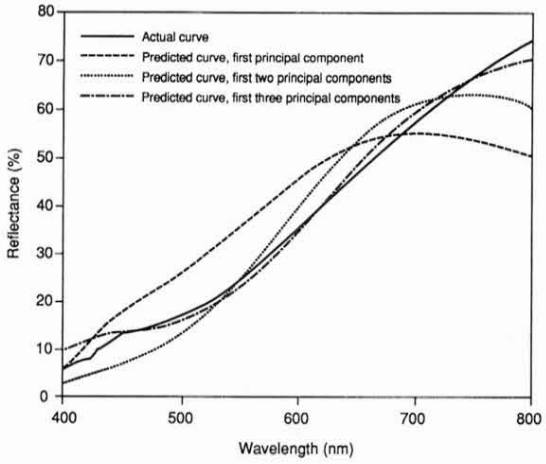


FIG. 6. Actual reflectance curve and predicted curves from principal-component models.

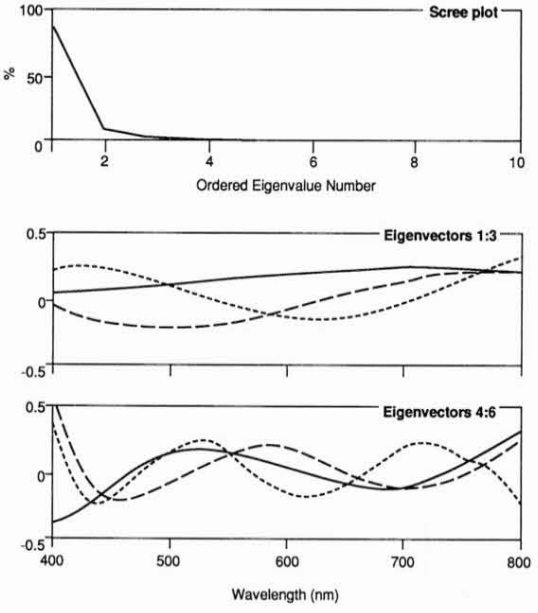


FIG. 7. Eigenstructure of the pooled wood-group sample covariance matrix. Top graph is a scree plot of the largest 10 (ordered) eigenvalues as percentages of total sample variance. Middle graph shows the first (solid line), second (dashed line), and third (dotted line) eigenvectors. Bottom graph shows the fourth (solid line), fifth (dashed line), and sixth (dotted line) eigenvectors. Y-ordinates of middle and bottom graphs are dimensionless quantities. Eigenvectors shown have unit Euclidean norm.

TABLE 1. *Ordered eigenvalues and percentage of variation explained by them.*

| Index | Eigenvalue | Cumulative % of total variation |
|---|---|---|
| 1 | 1,059.36 | 86.27 |
| 2 | 133.77 | 97.17 |
| 3 | 25.75 | 99.26 |
| 4 | 5.57 | 99.72 |
| 5 | 1.79 | 99.86 |
| 6 | 0.74 | 99.92 |
| 7 | 0.19 | 99.94 |
| 8 | 0.14 | 99.95 |

the 1-nm curves. Cubic spline interpolation reduced this relative RMS error to 0.4%. Figure 1 shows typical reflectance curves for each wood-feature category. The three curves shown in each graph were randomly selected from the collection of measured spectral-reflectance curves for the category. Figure 2A shows the mean response vectors for each feature group. The mean curves are fairly distinct in shape, although the mean sapwood-earlywood curve is similar to the mean sapwood-latewood curve and the mean loose-knot curve is very similar to the mean tight-knot curve. For both pairs, the shapes of the curves are similar but one is shifted upward, indicating a brightness difference rather than a hue difference—that is, sapwood-earlywood has, on the average, a somewhat brighter appearance than sapwood-latewood. Likewise, loose knots generally are somewhat brighter appearing than tight knots, although the difference is minimal. The standard deviations of spectral reflectance at each wavelength are presented in Fig. 2B. Although this figure does not illustrate the relationships between the variability of spectral reflectances at different wavelengths, the curves do show that the variance structures of the feature groups seem to differ substantially.

Although Fig. 2B appeared to show differences in the variance structures of the wood groups, as a further check Gnanadesikan and Lee's graphical test for equality of covariance matrices [as described by Seber (1984)] was performed. This graphical procedure is based on a probability plot of the traces of the mean-

TABLE 2. *PRESS statistics for progressively larger principal-component models.*

| $q$ | PRESS ($q$) |
|---|---|
| 1 | 90.6563 |
| 2 | 20.1134 |
| 3 | 2.9862 |
| 4 | 0.2312 |
| 5 | 0.0995 |
| 6 | 0.0619 |
| 7 | 0.0542 |
| 8 | 0.0471 |
| 9 | 0.0357 |
| 10 | 0.0279 |

TABLE 3. *Number of principal components retained under various selection criteria.*

| Criterion | No. components retained |
|---|---|
| Cattell's scree | 2 |
| 95% of total variation | 2 |
| 99% of total variation | 3 |
| 99.9% of total variation | 6 |
| Kaiser's criterion | 2 |
| PRESS | 32 |

corrected sum-of-products matrices. Figure 3 shows the resulting plot. The points fall roughly on a straight line, indicating that the group covariance matrices are similar.

## CURVE CHARACTERIZATION BY PCA

As a first step in analyzing the wood spectral-reflectance curves by PCA, eigenstructure plots (i.e., scree plots and eigenvector plots) were made for each wood group separately and for all groups pooled together. The eigenstructure plots for the individual wood groups are given in Figs. 4 and 5. The eigenvalues and eigenvectors, although not identical for each group, are similar. The eigenvectors were compared geometrically by using Krzanowski's (1979) between-groups comparison of principal-component subspaces. When six-dimensional principal-component subspaces of the eight wood groups were compared, the subspaces appeared to capture similar major sources of variation among the groups. With six-dimensional subspaces, the maximum angle between the groups' first eigenvectors and the vector defined as the closest to them was 0.8°. The maximum angles between the groups' second through sixth eigenvectors and the vectors defined as the closest to them were 1.7°, 3.7°, 8.7°, 17.1°, and 39.9°, respectively.

Figure 6 shows how well a typical spectral-reflectance curve is modeled on the basis of a principal-component model with common population covariances. From the figure, it appears that more than three components are necessary to adequately model this curve (which is from a pitch pocket). With six principal components, the modeled curve is virtually indistinguishable from the actual curve except for being slightly flattened at 450 nm.

The eigenstructure for all of the group covariance matrices pooled together is illustrated in Fig. 7. The pooled eigenstructure shown in this figure appears similar to that of the individual eigenstructure plots of Figs. 4 and 5. The elbow in the top graph suggests that, according to Cattrell's scree test, one or two principal components are sufficient. Table 1 gives the cumulative percentage of total variation explained by the ordered eigenvalues. Because the first two components contain over 95% of the total variation while the remainder of the components' contributions decrease geometrically, two components seem sufficient in explaining the variation.

Under Kaiser's criterion, the number of eigenvalues that exceed the average eigenvalue is two. (The average eigenvalue in the decomposition of the pooled covariance is 29.95.)

Let $\boldsymbol{x}_{ij}$ be the $i$th response curve observed from the $j$th group. If $\hat{\boldsymbol{x}}_{ij}$ is the corresponding predicted response curve from a $q$-dimensional principal-component model, then the PRESS statistic can be given by

$$\text{PRESS}(q) = \frac{1}{np} \sum_{j=1}^{g} \sum_{i=1}^{n_j} ||\hat{\boldsymbol{x}}_{ij} - \boldsymbol{x}_{ij}||^2 \quad (13)$$

where $g$ is the number of groups and $n = \Sigma\, n_j$ is the total number of observed response curves. Notice that PRESS($q$) is the mean-squared er-

TABLE 4. *Estimated misclassification rates (%) based on raw spectral-curve data.*

| Type of discriminant analysis | Feature type | | | | | | |
|---|---|---|---|---|---|---|---|
| | Sapwood | Knot | Pitch pocket | Pitch streak | Blue stain | Wane | Overall |
| Linear (LDA) | 1.0 | 2.8 | 23.0 | 0.0 | 8.0 | 1.5 | 5.0 |
| Quadratic (QDA) | 1.0 | 0.0 | 1.0 | 0.5 | 1.0 | 4.5 | 1.1 |

ror. Table 2 summarizes these average squared deviations based on progressively larger principal-component models. Application of the PRESS statistic by following the leave-one-out scheme as outlined previously results in 32 components being retained.

The number of components retained for each of the preceding methods is summarized in Table 3. The results of the methods for choosing the number of meaningful components vary considerably. As is common with many analyses, the researcher must decide which number of components adequately represents the data. Between three and six principal components probably are sufficient; with six components, predictability is substantially improved.

## CURVE CLASSIFICATION BY PCA

The preceding section showed how the wood curves in this data set can be characterized by a small number of principal components. Principal components can also be used to classify the curves by feature type. The usefulness of PCA for classification was investigated by comparing a number of methods for classifying the spectral curves by feature type. First, both linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) were performed on the raw spectral curves. Table 4 shows the estimated rates of misclassification for each type of analysis. The rates were estimated by the cross-validation, leave-one-out method discussed earlier. In this table the two sapwood groups (early and late) have been merged for reporting purposes, as have the two knot groups (loose and tight), because for many practical purposes, it is not necessary to distinguish between these groups, even though they may differ in their spectral-reflectance characteristics. For example, if grain pattern is not important, there is no advantage in knowing if a normal-wood pixel is earlywood or latewood. The same can be said for loose and tight knot when neither is allowed in a cutting or when both are treated equally for grading purposes. Without merging, the overall misclassification rate for LDA increased from 5.0% to 7.1%; for QDA the rate increased from 1.1% to 1.6%.

The data points of the raw reflectance curves are highly correlated, and calculation of the discriminant functions on such data can result in ill-behaved discriminators, because their calculation depends on the inverse of the pooled sample covariance matrix. For these data, the

TABLE 5. *Estimated misclassification rates (%) of QDA as based on principal components.*

| PCs used in QDA | Feature type | | | | | | |
|---|---|---|---|---|---|---|---|
| | Sapwood | Knot | Pitch pocket | Pitch streak | Blue stain | Wane | Overall |
| 2 | 26.5 | 36.0 | 20.5 | 96.5 | 9.5 | 54.0 | 38.2 |
| 2,3 | 9.5 | 20.5 | 16.0 | 1.5 | 29.0 | 46.0 | 19.1 |
| 2,3,1 | 1.3 | 13.5 | 12.5 | 3.0 | 4.6 | 12.5 | 7.8 |
| 2,3,1,4 | 1.3 | 6.5 | 7.5 | 1.5 | 4.5 | 10.5 | 5.0 |
| 2,3,1,4,5 | 1.8 | 3.0 | 6.5 | 0.5 | 5.0 | 5.5 | 3.4 |
| 2,3,1,4,5,6 | 1.0 | 1.3 | 2.0 | 1.0 | 3.0 | 3.5 | 1.8 |
| 2,3,1,4,5,6,9 | 0.0 | 1.0 | 1.0 | 0.5 | 2.5 | 3.0 | 1.1 |

TABLE 6. *Confusion (predicted vs. actual) matrix for QDA based on the best seven PCs.*

| Actual feature type | Predicted feature type | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Sapwood-earlywood | Sapwood-latewood | Loose knot | Tight knot | Pitch pocket | Pitch streak | Blue stain | Wane |
| Sapwood-earlywood | 199 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sapwood-latewood | 2 | 198 | 0 | 0 | 0 | 0 | 0 | 0 |
| Loose knot | 0 | 0 | 180 | 19 | 0 | 0 | 0 | 1 |
| Tight knot | 0 | 0 | 29 | 169 | 0 | 0 | 0 | 2 |
| Pitch pocket | 0 | 0 | 0 | 0 | 200 | 0 | 0 | 0 |
| Pitch streak | 0 | 1 | 0 | 0 | 0 | 199 | 0 | 0 |
| Blue stain | 5 | 1 | 1 | 0 | 0 | 0 | 193 | 0 |
| Wane | 0 | 0 | 1 | 2 | 2 | 0 | 0 | 195 |

condition number (the ratio of the largest eigenvalue to the smallest) of the pooled sample covariance matrix was $5.4 \cdot 10^5$, indicating that the matrix was near-singular. This number can be interpreted as indicating that changes in the fifth significant digit of the pooled sample covariance would be sufficient to make it singular (Stewart 1987). The effects of measurement density on the stability of the sample covariance matrix and on the ability to classify were investigated by repeating the discriminant analyses and using only every fifth wavelength reading for each spectral curve. Doing so affected the results only modestly. The condition number of the pooled sample covariance matrix decreased to $1.1 \cdot 10^4$, the overall LDA misclassification rate increased from 5.0% to 6.6%, and the overall QDA misclassification rate actually decreased from 1.1% to 0.8%.

Discriminant analysis was also performed on various combinations of the curves' principal components. Table 5 shows the misclassification rates for quadratic discriminant analysis, as estimated with the cross-validation, leave-one-out method. The principal components for each row of the table were chosen via a forward stepwise variable-selection criterion that considered the first 10 principal components, PC1 through PC10 (rank-ordered by eigenvalue). Note that the PCs associated with the larger eigenvalues were often, but not always, the best discriminators. For example, the second principal component (PC2) proved to be the best single component for discrimination, in that it produced the lowest overall misclassification rate for the merged set of feature classes. However, it alone was able to correctly classify only about 60% of the reflectance curves. The second best principal component (when used in conjunction with PC2) was PC3. The addition of PC3 in the analysis dramatically reduced the pitch-streak error rate (although at the expense of blue stain) and also substantially reduced the sapwood error rate. Adding PC1 to PC2 and PC3 brought the blue-stain error rate back to a reasonable level and further reduced the sapwood error rate. Adding PC4 to PC1, PC2, and PC3 improved the knot and pitch-pocket error rates.

Linear discriminant analysis was also performed on the principal components, but as with the raw data, it proved significantly inferior. Merging the two sapwood groups and knot groups prior to QDA also resulted in somewhat higher misclassification rates.

For those reflectance curves that were misclassified, it is instructive to consider into what classes they were assigned. The QDA confusion matrix, i.e., the matrix of actual versus predicted classifications by feature type for the unmerged classes (based on the best eight principal components), is given in Table 6. As the table shows, the two types of knots are often confused, but the two types of sapwood are not. Furthermore, blue stain has a tendency to be misclassified as sapwood-earlywood. Wane also has a tendency to be misclassified, but not to any single other group.

## DISCUSSION AND CONCLUSION

A database of spectral-reflectance curves of Douglas-fir veneer surface features was presented and analyzed via principal-components analysis. For characterization (modeling) purposes, the choice of which principal components to consider becomes the choice of how many principal components to retain, because those associated with the largest eigenvalues are the best choices. For this data set, the various methods of determining the number of components to keep for characterizing the curves yielded widely different results, ranging from 2 to 32.

For classification purposes, the choice of which principal components to consider is more complicated than merely a choice of how many principal components to retain. For this data set, the principal components associated with the largest eigenvalues were not always the best for classification purposes. When quadratic discriminant analysis was used to classify the curves, raw spectral reflectance measurements spaced 50 nm apart (9 readings per curve) achieved approximately the same misclassification rates as when QDA was applied to a comparable (7) number of principal components. Quadratic discriminant analysis proved substantially superior to linear discriminant analysis, whether operating on the raw data or on a subset of principal components.

Even though the classification accuracies of QDA on the raw data were comparable to those for QDA on the best seven principal components, practical considerations favor the PCA approach. Direct spectral-reflectance measurements are difficult or impossible to make at production speeds in a mill. However, if one is given a set of $p$ principal-component loading curves (eigenvectors), it is theoretically possible to design a $p$-dimensional image-acquisition system whose $p$ color values at a point are precisely the desired principal components for the spectral reflectance of the object at that point (Vora and Trussell 1993; Vrhel and Trussell 1993). This design is possible because the major principal-component loading curves are fairly smooth functions of wavelength. A simple system would consist of a gray-scale camera and a filter wheel accommodating $p$ custom-made filters. A real-time system would use several spatially registered sensors, each with its own filter, or a single sensor fitted with a liquid crystal tunable filter (Hoyt 1995). Such systems are called multispectral or extended-color systems because they usually employ more than the usual three sensors of a conventional color camera, i.e., $p > 3$. At least one practical extended-color system is already being proposed (Trussell 1994). Although this study is based on limited data and its conclusions will need further verification, its results suggest that extended-color systems may be able to achieve, in a mill environment, classification accuracies comparable to those found here.

## REFERENCES


BRUNNER, C. C., G. B. SHAW, D. A. BUTLER, AND J. W. FUNCK. 1990. Using color in machine vision systems for wood processing. Wood Fiber Sci. 22(4):413–428.

———, A. G. MARISTANY, D. A. BUTLER, D. VANLEEUWEN, AND J. W. FUNCK. 1992. An evaluation of color spaces for detecting defects in Douglas-fir veneer. Ind. Metrol. 2:169–184.

BUCHSBAUM, G., AND A. GOTTSCHALK. 1984. Chromaticity coordinates of frequency-limited functions. J. Opt. Soc. Am. A 8:885–887.

BUTLER, D. A., C. C. BRUNNER, AND J. W. FUNCK. 1989. A dual-threshold image sweep-and-mark algorithm for defect detection in veneer. Forest Prod. J. 39(5):25–28.

COCHRAN, R. N., AND F. H. HORNE. 1977. Statistically weighted principal component analysis of rapid scanning wavelength kinetics experiments. Anal. Chem. 49: 846–853.

COHEN, J. 1964. Dependency of the spectral reflectance curves of the Munsell color chips. Psychon. Sci. 1:369–370.

CONNERS, R. W., C. W. MCMILLIN, AND C. N. NG. 1985. The utility of color information in the location and identification of defects in surfaced hardwood lumber. Pages XVIII-1 to XVIII-33 *in* Proc., First International Conference on Scanning Technology in Sawmilling. Miller-Freeman Publications, San Francisco, CA.

FLURY, B. 1988. Common principal components and related multivariate models. John Wiley and Sons, New York, NY.

HABBEMA, J. D. F., AND J. HERMANS. 1977. Selection of variables in discriminant analysis by F-statistic and error rate. Technometrics 19:487–493.

HEALEY, G. 1989. Using color for geometry-insensitive segmentation. J. Opt. Soc. Am. A 6:920–937.

———, AND T. O. BINFORD. 1987. The role and use of color in a general vision system. Pages 599–613 *in* Proc., ARPA Image Understanding Workshop, DARPA. University of Southern California, Los Angeles, CA.

HOOGERBRUGGE, R., S. J. WILLIG, AND P. G. KISTEMAKER. 1983. Discriminant analysis by double stage principal-component analysis. Anal. Chem. 55:1710–1712.

HOYT, C. C. 1995. Toward higher res, lower cost quality color, and multispectral imaging. Adv. Imaging 10(4): 53–55.

HUNTER, R. S., AND R. W. HAROLD. 1987. The measurement of appearance. 2nd ed. John Wiley and Sons, New York, NY.

KANADE, T., AND K. IKEUCHI. 1991. Introduction to the special issue on physical modeling in computer vision. IEEE Trans. Pattern Anal. Mach. Intell. PAMI. 13:609–610.

KLAPSTEIN, K. D., F. L. WEICHMAN, W. N. BAUER, AND D. J. KENWAY. 1989. Optical characteristics of wood stains and rot. Appl. Opt. 28(20):4450–4452.

KRZANOWSKI, W. J. 1979. Between-groups comparison of principal components. J. Am. Stat. Assoc. 74:703–707.

———. 1984. Principal-component analysis in the presence of group structure. Appl. Stat. 33:164–168.

———. 1987. Cross-validation in principal-component analysis. Biometrics 43:575–584.

LEBOW, P. K. 1992. Estimation of discriminant analysis error rate for high dimensional data. Doctoral thesis, Department of Statistics, Oregon State University, Corvallis, OR.

MACADAM, D. L. 1981. Color measurement: Theme and variations. Springer-Verlag, Berlin, Germany.

MALONEY, L. T. 1986. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. J. Opt. Soc. Am. A 3:1673–1683.

MANLY, B. F. J., AND J. C. W. RAYNER. 1987. The comparison of sample covariance matrices using likelihood ratio tests. Biometrika 74:841–847.

MARDIA, K. V., J. T. KENT, AND J. M. BIBBY. 1979. Multivariate analysis. Academic Press, Orlando, FL.

MARISTANY, A. G., P. K. LEBOW, C. C. BRUNNER. 1994. Application of the dichromatic reflection model to wood. Wood Fiber Sci. 26(2):249–258.

———, D. A. BUTLER, C. C. BRUNNER, AND J. W. FUNCK. 1991. Exploiting local color information for defect detection on Douglas-fir veneer. Pages 1–7 *in* Proc., Fourth International Conference on Scanning Technology in Sawmilling. Miller-Freeman Publications, San Francisco, CA.

———, P. K. LEBOW, C. C. BRUNNER, D. A. BUTLER, AND J. W. FUNCK. 1993. Classifying wood-surface features using dichromatic reflection. Pages 56–64 *in* J. A. DeShazer and G. E. Meyers, eds. Optics in agriculture and forestry. Proc., SPIE 1836. SPIE—The International Society for Optical Engineering, Bellingham, WA.

MATTHEWS, P. C., AND B. H. BEECH. 1976. Method and apparatus for detecting timber defects. U.S. Patent No. 3,976,384.

PARKKINEN, J. P. S., AND T. JAASKELAINEN. 1987. Color representation using statistical pattern recognition. Appl. Opt. 26:4240–4245.

———, J. HALLIKAINEN, AND T. JAASKELAINEN. 1989. Characteristic spectra of Munsell colors. J. Opt. Soc. Am. A 6:318–322.

PHELPS, J. E., E. A. MCGINNES, JR., H. E. GARETT, AND G. S. COX. 1983. Growth-quality evaluation of black walnut wood. Part II—Color analyses of veneer produced on different sites. Wood Fiber Sci. 15(2):177–185.

RAO, C. R. 1964. The use and interpretation of principal-component analysis in applied research. Sankhya A. 26: 329–358.

ROZETT, R. W., AND E. M. PETERSEN. 1976. Classification of compounds by the factor analysis of their mass spectra. Anal. Chem. 48:817–825.

SEBER, G. A. F. 1984. Multivariate observations. John Wiley and Sons, New York, NY.

SIMONDS, J. L. 1963. Application of characteristic vector analysis to photographic and optical response data. J. Opt. Soc. Am. 53:968–974.

SNAPINN, S. M., AND J. D. KNOKE. 1988. Bootstrapped and smoothed classification error rate estimators. Commun. Stat.—Simul. Comput. 17:1135–1153.

SOEST, J. 1984. Optical scanning technique for defect detection. Scanning technology for the eighties. Spec. Publ. No. SP-21:120–123. Forintek Canada Corp., Vancouver, BC, Canada.

———, AND P. C. MATTHEWS. 1985. Laser Scanning technique for defect detection. Pages XVII-1 to XVII-4 *in* Proc. First International Conference on Scanning Technology in Sawmilling. Miller Freeman Publications, San Francisco, CA.

STEWART, G. W. 1987. Collinearity and least squares regression. Stat. Sci. 2(1):68–100.

STILES, W. S., AND G. WYSZECKI. 1962. Counting metameric object colors. J. Opt. Soc. Am. 52:58–75.

———, ———, AND N. OHTA. 1977. Counting metameric object-color stimuli using frequency-limited spectral reflectance functions. J. Opt. Soc. Am. 67:779–784.

TOMINAGA, S., AND B. A. WANDELL. 1989. Standard surface-reflectance model and illuminant estimation. J. Opt. Soc. Am. A 6:576–584.

TRUSSELL, H. J. 1994. Personal communication. Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC.

VORA, P. L., AND H. J. TRUSSELL. 1993. A mathematical method for designing a set of colour scanning filters. Pages 322–329 *in* J. Bares, ed. Color hard copy and graphics arts II. Proc., SPIE 1912. SPIE—The Inter-

national Society for Optical Engineering, Bellingham, WA.

VRHEL, M. J., AND H. J. TRUSSELL. 1993. Optimal scanning filters using spectral reflectance information. Pages 404–411 *in* B. E. Rogowitz and J. P. Allebach, eds. Human vision, visual processing, and digital display IV. Proc., SPIE 1913. SPIE—The International Society for Optical Engineering, Bellingham, WA.

WOLD, S. 1978. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics 20:397–405.

YOUNG, R. A. 1986. Principal-component analysis of macaque lateral geniculate nucleus chromatic data. J. Opt. Soc. Am. A 3:1735–1742.