

Near-IR Identification of Woods for Restoration of Historic Buildings and Furniture

PRAKASH NAIR and ROBERT A. LODDER*

Department of Chemistry and College of Pharmacy, University of Kentucky, Lexington, Kentucky 40506-0055

Frank Lloyd Wright is perhaps the most widely known of American architects. Wright's Prairie period, devoted mostly to residential architecture, is characterized in part by the use of indigenous woods that were neither planed nor painted (though they were sometimes stained). Residences are more likely to be altered over time than are other buildings, often without good documentation of the changes that were made. This research seeks to aid restoration efforts by providing a means of identifying the original woods used in construction.

Index Headings: Near-infrared; Wood; Restoration.

INTRODUCTION

Frank Lloyd Wright is considered America's greatest designer and one of the leading architects of all time. He built over 400 houses and designed at least twice that number in his working career of 72 years.¹ His designs were motivated by his respect for American traditions, landscape, and native materials. He redefined a room from a space enclosed by four walls to a space whose perimeters were set by the changing position of the observer. As a result, the space seems larger, more relaxed, and more varied than its actual size suggests.

Wright was born in Richland Center, Wisconsin, and worked summers on his uncle's ranch. Life on the farm instilled in Wright a respect for wilderness, natural materials, and agricultural life. His career, which was primarily dedicated to residential architecture, is often thought of as divided into three periods of approximately 25 years each. The first period, his Prairie period, was the period in which many of his important residences were constructed. These residences were characterized by extended, low buildings with ample overhanging, low-pitched roofs that integrated into a flat prairie landscape. The materials used in these residences came from the region in which they were constructed, and the woodwork was neither planed nor painted. Often the wood was stained to provide some protection from the weather.

Once a building has been designated a national historical site, any repairs or restorations to the building must restore the building to the original condition. It is occasionally difficult to exactly determine the original condition. In residences, this determination is even more difficult because precise records and photographs are less likely to be retained than is the case for public buildings. In addition, residences are more likely to be modified to suit the needs and wants of their current owners. Residential restoration therefore offers unique opportunities for material identification for chemists involved in restoration work.

Received 4 November 1992.

* Author to whom correspondence should be sent.

THEORY

A population \mathbf{P} in a hyperspace \mathbf{R} represents the universe of possible spectrometric samples (the rows of \mathbf{P} are the individual samples, while the columns are the independent information vectors, such as wavelengths or energies). \mathbf{P}^* is a discrete realization of \mathbf{P} based on a calibration set \mathbf{T} , chosen only once from \mathbf{P} to represent as nearly as possible all the variations present in \mathbf{P} .

\mathbf{P}^* is calculated with the use of a bootstrap process by an operation $\kappa(\mathbf{T})$, and \mathbf{P}^* has parameters \mathbf{B} and \mathbf{C} , where $\mathbf{C} = E(\mathbf{P})$ and \mathbf{B} is the Monte Carlo approximation to the bootstrap distribution.

Each new sample is analyzed by an operation $\psi(\mathbf{T}, \mathbf{B}, \mathbf{X}, \mathbf{C})$, which projects the probability density in hyperspace by many-one mapping onto the vector connecting \mathbf{C} and \mathbf{X} . The directional standard deviation σ is found from the projected probability density:

$$\left\{ \sigma \left| \frac{\int_0^\sigma \left(\int_R \mathbf{P}^* - \overline{\mathbf{C}\mathbf{X}} \right)}{\int_R \mathbf{P}^* - \overline{\mathbf{C}\mathbf{X}}} = 0.68 \right. \right\}. \quad (1)$$

The calculation of a skew-adjusted σ is based on a comparison of the expectation values $\mathbf{C} = E(\mathbf{P})$ and $\mathbf{C}_T = \text{med}(\mathbf{T})$, projected on the hyperline connecting \mathbf{C} and \mathbf{X} .

$$(\mathbf{C} - \mathbf{C}_T) \rightarrow \overline{\mathbf{C}\mathbf{X}}. \quad (2)$$

The result of the corrected projection is an asymmetric σ that provides two measures of the standard deviation along the hyperline connecting \mathbf{C} and \mathbf{X} .

$$\left\{ \begin{matrix} +\sigma \\ -\sigma \end{matrix} \left| \frac{\int_0^{+\sigma} \left(\int_R \mathbf{P}^* - \overline{\mathbf{C}\mathbf{X}} \right)}{\int_R \mathbf{P}^* - \overline{\mathbf{C}\mathbf{X}}} = 0.34 \right. \right\} \quad (3)$$

in the direction of \mathbf{X} in hyperspace, and

$$\left\{ \begin{matrix} -\sigma \\ +\sigma \end{matrix} \left| \frac{\int_0^{-\sigma} \left(\int_R \mathbf{P}^* - \overline{\mathbf{C}\mathbf{X}} \right)}{\int_R \mathbf{P}^* - \overline{\mathbf{C}\mathbf{X}}} = 0.34 \right. \right\} \quad (4)$$

in the opposite direction along the hyperline connecting \mathbf{c} and \mathbf{x} .

Given two calibration sets \mathbf{P}_1^* and \mathbf{P}_2^* with an equal number of elements n , it is possible to determine whether \mathbf{P}_1^* and \mathbf{P}_2^* are drawn from the same population, even if

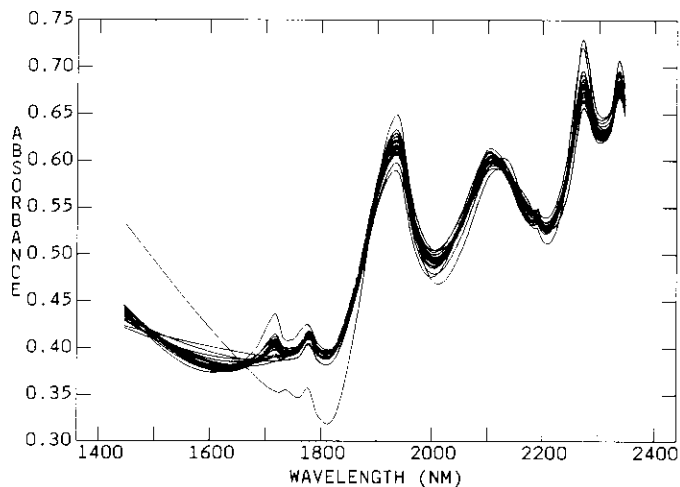


FIG. 1. Average spectrum for each of the 26 wood species examined. The average spectra are calculated from 10 individual spectra of each species. Multiplicative scatter correction was applied to the average spectra. The spectrum that appears markedly different from the rest at shorter wavelengths is the average spectrum of ebony.

σ_1 and σ_2 are less than 3, with the use of quantile-quantile (QQ) plots and a simple correlation test statistic

$$\rho\left(\left\{\int_R \mathbf{P}_1^*\right\}, \left\{\int_R \mathbf{P}_1^*\right\} \cup \left\{\int_R \mathbf{P}_2^*\right\}\right). \quad (5)$$

A bootstrap method is employed to set confidence limits on ρ . The central 68% confidence interval on ρ is also used to calculate σ_ρ , a distance in SDs that is different from σ and is more sensitive to small differences in location and scale between \mathbf{P}_1^* and \mathbf{P}_2^* .

EXPERIMENTAL

Twenty-six different species of wood used in ornamental building and furniture construction were examined spectrophotometrically. The different species were: (1) white ash, (2) beech, (3) birch, (4) cedar, (5) cherry, (6) cypress, (7) ebony, (8) elm, (9) elm (burl), (10) gum, (11) holly, (12) mahogany (Honduras), (13) mahogany (striped), (14) maple, (15) maple (bird's eye), (16) oak, (17) padouk, (18) pearwood, (19) rosewood, (20) satinwood, (21) sycamore, (22) teak, (23) tulip poplar, (24) walnut, (25) walnut (figured), and (26) zebrawood. Each sample was about 1 mm in thickness and had dimensions of about 5 cm × 8 cm. The wood samples were obtained from Schiffer Limited, Exton, PA.

The spectral data were collected with an InfraAlyzer Flex near-IR spectrometer (Bran+Luebbe, Elmsford, NY), which scans the samples at 19 different wavelengths ranging from 1445 nm to 2348 nm. Each sample was scanned ten times, and every scan covered a different portion of the sample surface to obtain an estimate of variation due to sample presentation. The experiment was carried out in a dark room to reduce noise due to light, and care was taken to maintain other factors such as temperature constant. The data were collected from the spectrometer to a MicroVax II computer (Digital Equipment Corp., Maynard, MA) with the use of a program written in Speakeasy IV Epsilon (Speakeasy Computing Corp., Chicago, IL), and analysis was done on an

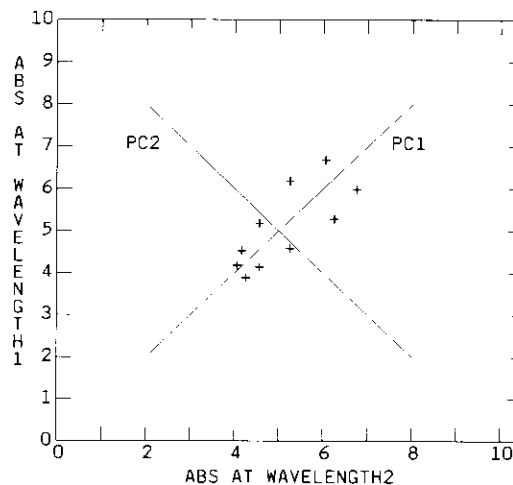


FIG. 2. Formation of principal axes from hypothetical absorbance data (shown as pluses) is accomplished through a translation and rotation of the original wavelength coordinate system.

IBM 3090-600J parallel vector supercomputer (IBM, Armonk, NY), also with programs written in Speakeasy. A multiplicative scatter correction was performed on the spectra, and the average of the 10 spectra for each species was graphed as Fig. 1, which shows that all wood species are at least superficially similar.

Principal Components Analysis. After the distances in multidimensional standard deviations (SDs) between types of wood were determined, the raw data were converted into factor scores via principal-axis transformation (PAT) as an aid to data interpretation. In PAT, a rigid rotation of the original axes is undertaken to maximize the variation along one new axis, which is called the first principal component (PC) (see Fig. 2). The second new axis, which by definition is at right angles to the first, is rotated to maximize description of the variation remaining after formation of the first PC. In raw near-IR data, the first and second PCs usually represent almost 95% of the variations. A plot of the first two axes often gives us enough information about the location and scale of the clusters of spectra in hyperspace to differentiate among the species.

When the raw data were transformed into their principal component values, the first two axes provided almost 99% of the total spectral variation, and the first six axes provided about 99.9% of the total spectral variation. A plot of the first two principal component values for all the spectra suggested that most of the species could be identified correctly with these two PCs alone. Figure 3 shows a PC plot for a sampling of five of the 26 samples of wood—white ash, beech, birch, cedar, and cherry. Two different spectra from the same species of wood occupy a similar position in spectral hyperspace, while two different spectra from two different species of wood occupy locations farther apart.

Multiplicative Scatter Correction (MSC). The spectra of the 26 species of wood showed a wide variation in light scattering both between species and within each species. It was unclear *a priori* what scattering information was necessary to differentiate among the wood samples, so distances between species spectra were calculated with and without MSC on the raw spectra. MSC was also

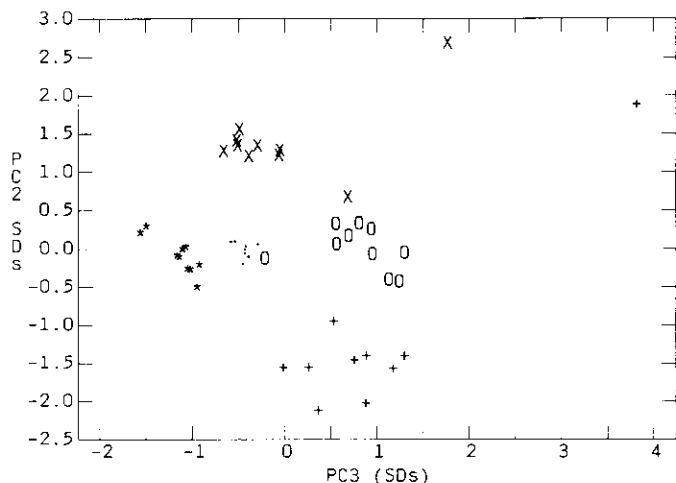


FIG. 3. A principal component plot of 5 wood species' spectra showing the individual spectral scans as single points. The average spectrum of each species is simply the center of each cluster of 10 points (period = white ash, plus = beech, circle = birch, cross = cedar, asterisk = cherry). The axes have units of SDs.

calculated on the raw spectra in two ways: on the entire array of 260 spectra simultaneously and for 26 separate MSCs (a different MSC for each wood species). Thus three sets of spectra were analyzed by the supercomputer algorithms: the 260 raw spectra as well as 260 scatter-corrected spectra and 26 groups of ten scatter-corrected spectra.

Contour Plotting. A bootstrap algorithm was used to calculate confidence limits on ρ_c , the correlation value between two sets of spectra drawn from the same underlying population distribution. A ρ_c value was calculated for each group of ten spectra from each species. The calibration set and validation set for this algorithm are formed from the same group of ten spectra. The bootstrap algorithm creates one set of replicates from the input spectra ($B = 100$ replicates for all species) that becomes the calibration set. The algorithm then creates many validation sets, each containing $B = 100$ replicates

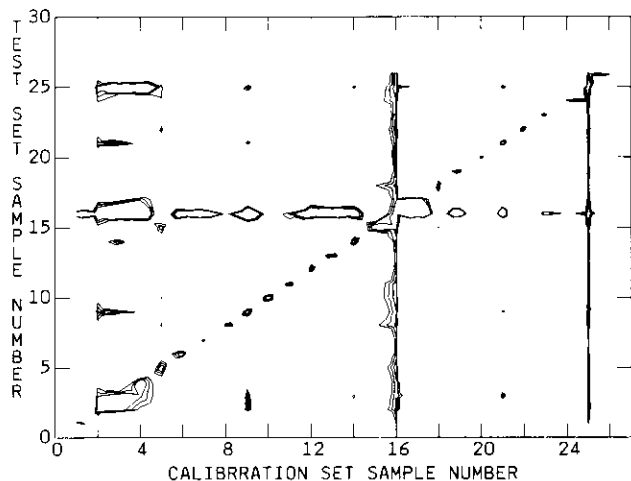


FIG. 4. A contour plot of the distance matrix (distances given in SDs) with contours drawn at 1, 2, and 3 SDs. The 26 wood species appear on the y-axis as the calibration set in subcluster detection. The 26 wood species on the y-axis are used as the test set in subcluster detection.

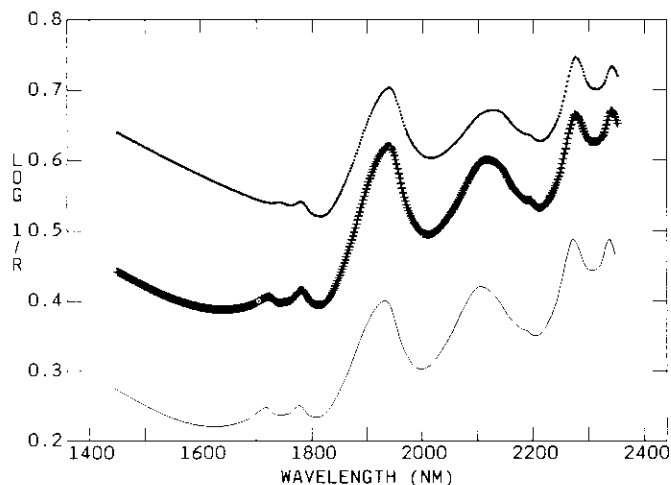


FIG. 5. Averages of 10 raw spectra of white ash (solid line), ebony (dotted line), and walnut (dashed line) showing differences in baselines between species.

as well, and calculates the correlation between the calibration set and each validation set. The mean and standard deviation of the correlation values are then calculated, and these parameters are used to produce the confidence level for each correlation ρ_c . The values for the mean correlation are ordinarily very close to one, because the calibration set and validation sets come from replicates of the same set of spectral data points. (Occasionally the mean is substantially lower than one, indicating that the calibration set is drawn from more than one underlying distribution.)

Equation 5 was coded into an algorithm² that was used to calculate the distance between two wood species' spectra in hyperspace. This algorithm uses one group of spectra as its calibration set and another group of spectra as its test set. Replication ($B = 100$) is done on both the calibration set and the test set spectra, and the distance between the centers of the two sets is calculated and returned as a test ρ .

Subcluster-detection analysis was performed on all 26 groups of 10 spectra. Each group of 10 spectra from one species was considered as the calibration set once, and the test sets comprised the 26 groups of spectra taken one species at a time. Thus, a correlation was determined between samples from the population distribution of each wood species with itself and with the other wood species. The ρ values formed a 26×26 array. When the two sets of spectra were similar in location and scale, their correlation was very close to one. When the two groups of spectra were different in location and/or scale, the correlation value from the subcluster detection algorithm was less than one. The significance of the correlation is decided with the use of the ρ value from the subcluster-detection algorithm and the mean ρ_c (μ_{ρ_c}) and the standard deviation of ρ_c (σ_{ρ_c}) from the bootstrap algorithm that calculates confidence limits on ρ_c .

A distance in standard deviations is used to tell the difference between the various species. This distance is

$$SDs = (\rho - \mu_{\rho_c}) / \sigma_{\rho_c} \quad (6)$$

For two groups of spectra from the same wood species, μ_{ρ_c} and ρ will be very close to one, so the distance will be

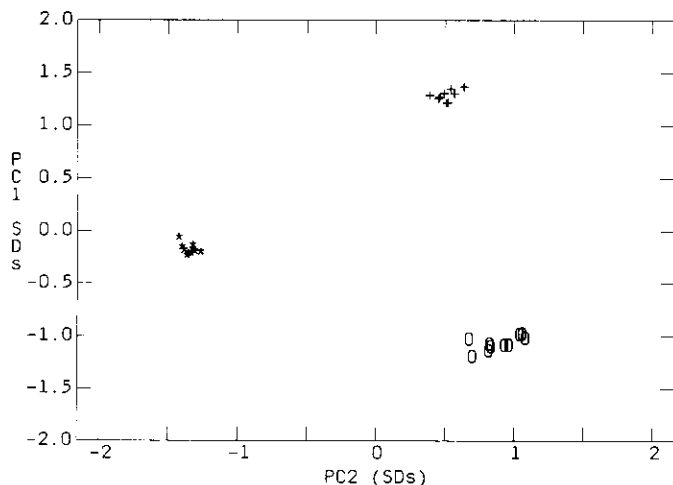


FIG. 6. PC plot of 10 spectra of white ash (pluses), 10 spectra of ebony (circles), and 10 spectra of walnut (asterisks). The fact that the clusters do not fall on the same line indicates that baseline differences are not the only differences between the samples.

very small (usually less than 3 SDs). When the spectral groups are drawn from different wood species, the distance between groups is higher than 3 SDs and we can say that the groups are different. The distance between groups in SDs is depicted in the contour plot in Fig. 4, in which contours are drawn at 1, 2, and 3 SDs, outlining the wood species that appear similar to the near-IR method. The spectral comparisons across the diagonal are comparisons between the same species, and all produced less than 3 SDs. Oak (#16) could not be differentiated from any of the other wood species one way (vertically), while oak produced several distances greater than 3 SDs when the test was performed the other way (horizontally).

Because the subcluster-detection test is not symmetric, testing the spectra of a sample **A** and sample **B** to see whether they are different requires that **A** be used as the calibration set and **B** as the test set to get one ρ ; then **B** must be used as the calibration set and **A** as the test set to get the other ρ . Both the upper and lower triangles of the distance array must be calculated to be certain that two groups of spectra are really the same.

Oak was differentiated from cherry (#5), elm (#8), gum (#10), birds-eye maple (#15), pearwood (#18), satinwood (#20), teak (#22), walnut (#24), and zebrawood (#26). However, oak cannot be differentiated from the other species. When a calibration set has only a few members ($n = 10$ here), the presence of one or two outlier spectra in the calibration set can make the set cluster large enough to overlap many other species. Figured walnut (#25) appears similar to many other species when tested vertically in Fig. 4, but actually is different from most of them when tested horizontally. The patterns in

TABLE I. The ρ values for white ash, ebony, and walnut.

	Calibration set		
	White ash	Ebony	Walnut
White ash	0.99071	0.62389	0.62645
Ebony	0.65485	0.98504	0.66503
Walnut	0.76587	0.77421	0.98548

TABLE II. Mean and standard deviation values on ρ for white ash, ebony, and walnut.

	Calibration set		
	White ash	Ebony	Walnut
Mean	0.98329	0.98565	0.99194
Standard deviation	0.0052157	0.0067985	0.0034402

figured walnut make it easy to generate apparent outliers in small samples of spectra.

Testing for subcluster correlations in both ways is important because the subcluster-detection algorithm is dependent upon the location and scale of the test set with respect to the location and scale of the calibration set. The detection algorithm has maximum sensitivity when the calibration set population distribution is larger in scale than the test set calibration distribution.² The algorithm is less sensitive to the reverse case, in which the test set population distribution is larger in scale than the calibration set population distribution. The results of the subcluster detection algorithm should only be considered valid after both ρ values have been calculated. Both ρ values must be less than 3 before the two groups of spectra can be said to be the same, but only one ρ must be greater than 3 for the two groups to be called different.

The same set of subcluster-detection calculations was performed on the two sets of MSC corrected spectra. We chose to define the best scatter correction as the one that gave the largest number of accurate species identifications in the 26×26 distance matrix. Of the three different distance matrices generated, the one that had the smallest number of values in the "less than 3 SDs" range was the matrix calculated from the raw spectra. We conclude that the scattering information plays an important role in differentiating among the wood samples, and that correcting scattering to an average level with an algorithm like the MSC is inadvisable for strict identification of planed wood surfaces in the restoration context.

White ash, ebony, and walnut were chosen as specific examples from the 26×26 distance matrix to illustrate the process of forming the distance matrix. Figure 5 shows the average spectra of these three species without scatter correction, and Fig. 6 shows their PC plots with all 10 spectra of each species. From the figures it is evident that the spectra of three species differ from one another. The difference is not simply a difference in baseline, because the three groups in the PC plot do not fall on a line. Table I shows ρ values for the three wood species calculated by the subcluster-detection algorithm. Table II shows the mean and standard deviation of ρ_c for the three wood species, and Table III shows the corresponding distances in SDs obtained for the three species.

The subcluster-detection algorithm also calculates the

TABLE III. Distances in standard deviations for white ash, ebony, and walnut.

	Calibration set		
	White ash	Ebony	Walnut
White ash	-1.4225	52.8650	103.7300
Ebony	63.4240	0.0895	93.1980
Walnut	43.3440	32.0260	1.8789

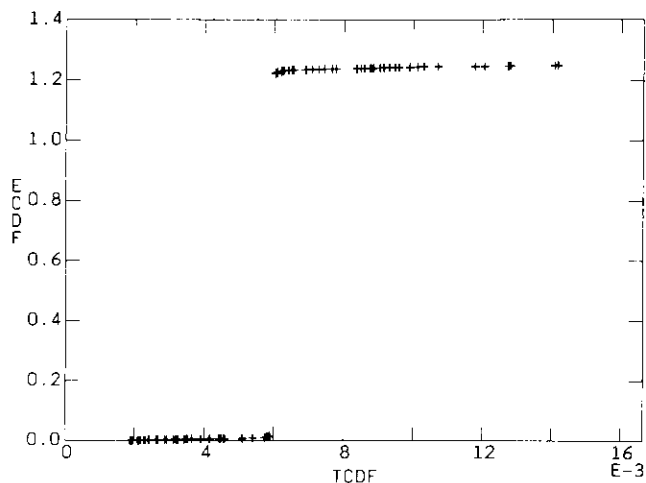


FIG. 7. QQ plot of the TCDF (from the calibration set) and ECDF (from the union of calibration and test sets) used to calculate ρ in Eq. 5. White ash and ebony are the calibration set and test set.

empirical and theoretical cumulative distribution functions (ECDF and TCDF) for quantile-quantile (QQ) plotting. A QQ plot of the ECDF vs. TCDF gives information regarding the scale relationship between the test set and the calibration set as well as the distance between them. If the two sets are the same, the QQ plot is a straight line with a slope of 1. Larger slopes in QQ plot line segments correspond to larger cluster scales. The TCDF is the calibration set, while the ECDF is the union of the calibration set and test set. Larger distances between two line segments in the QQ plot indicate better separation in SDs between spectral clusters in hyperspace. Figures 7 and 8 show the QQ plots of white ash/ebony and white ash/walnut, respectively.

CONCLUSIONS

The identification of woods during restoration of buildings is particularly important in national historical sites, where it is a legal requirement that any restoration work must restore the structure as nearly as possible to its original form. The materials used in restoration, including woods, paints, bricks, etc., should be the same as those used when the site was originally constructed. The importance of identifying woods is great in restoring the works of Frank Lloyd Wright because Wright relied heavily on native woods in his residential projects.

Near-IR spectrometry is a useful analytical technique for these identification purposes. No sample preparation is required, and any sample in solid form can be scanned

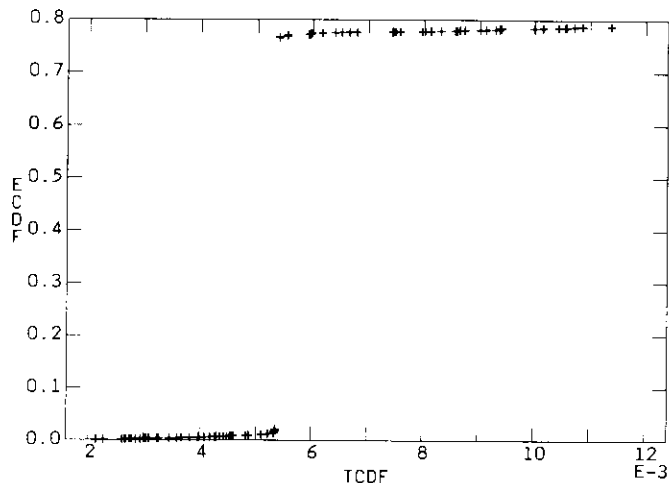


FIG. 8. QQ plot of white ash and walnut.

directly. Most near-IR instruments are sufficiently rugged and portable to be used at a restoration site. With the use of fiber-optic probes, near-IR instruments can go to restoration sites and scan woods and paints from floors to high ceilings quickly, obviating the need for removal or destruction of any material from the site.

The algorithms used for data analysis are very reliable, quick, and effective. The total amount of time required for calibration and analysis of the 260 spectra was less than an hour on the IBM 3090-600J. From the three different sets of calculations (two with MSC and one without), it appears that scattering information is important in the identification of planed woods. Apart from oak, which had apparent calibration set outliers and was difficult to differentiate from the rest of the wood species, most of the wood samples were correctly identified by the near-IR method. Ebony gave by far the most distinctive spectra due to intense electronic absorptions contributing to the near-IR baseline. Identification of woods with the use of the same algorithms with acoustic-resonance spectrometry appears to hold great promise, and will be investigated in future work.

ACKNOWLEDGMENTS

This work was supported through NSF NYI No. 9257998 and Grant No. STI-9108764.

1. H. A. Brooks, "Frank Lloyd Wright," in *Master Builders: A Guide to Famous American Architects*, D. Maddex, Ed. (Preservation Press, Washington, D.C., 1985), pp. 119-123.
2. R. A. Lodder and G. M. Hieftje, *Appl. Spectrosc.* **42**, 1500 (1988).