vantage in such an application.

We have shown that it is possible to obtain two-dimensional separations of protein samples chromatographically. Furthermore, these separations are attainable in an automated system using commercially available equipment, and the analysis time is virtually no longer than that of the first column separation. The method allows for the analysis of the total first column effluent on the second column without significant loss of first column resolution. Frequent sampling of the first column effluent provides peak profiles in both column dimensions. 3-D data representation provides for easy peak identification from run to run and is more reliable than single-column elution times. Most importantly, 2-D separation provides for higher resolution of complex samples and larger peak capacities than single-column chromatography. As pointed out by Giddings (5), the thrust of 1-D chromatographic research is to improve the individual separations, but the thrust of 2-D research will be to improve the combination of separation methods.

ACKNOWLEDGMENT

J.W.J. wishes to thank J. C. Giddings for an insightful discussion on multidimensional separations. We wish to thank Jeff Nystrom for some preliminary work in this area.

LITERATURE CITED

(1) Davis, Joe M.; Giddings, J. C. *Anal. Chem.* **1985**, *57*, 2168-2177.
(2) Davis, Joe M.; Giddings, J. C. *Anal. Chem.* **1985**, *57*, 2178-2182.
(3) O'Farrell, P. H. *J. Biol. Chem.* **1975**, *250*, 4007-4021.
(4) Anderson, N. G.; Anderson, N. C. *Anal. Biochem.* **1978**, *85*, 331-340.
(5) Giddings, J. C. *Anal. Chem.* **1984**, *56*, 1258A-1270A.
(6) Giddings, J. C. *HRC CC, J. High Resolut. Chromatogr. Chromatogr. Commun.* **1987**, *10*, 319-323.
(7) Guiochon, G.; Beaver, L. A.; Gonnord, M. F.; Siouffi, A. M.; Zakaria, M. *J. Chromatogr.* **1988**, *255*, 415-437.
(8) Majors, Ronald E. *J. Chromatogr. Sci.* **1980**, *18*, 571-579.
(9) Johnson, E. L.; Gloor, R.; Majors, R. E. *J. Chromatogr.* **1978**, *149*, 571-585.
(10) Fürst, Peter; Zimmerman, L.; Oulès, Raymond; Yahiel, Veronique; Johnson, Christina; Bergström, Johas *Anal. Chem.* **1982**, *122*, 394-403.
(11) Huber, J. F. K.; Fogy, I.; Fioresi, C. *Chromatographia* **1980**, *13*, 408-412.
(12) Takahashi, Nobuhiro; Takahashi, Yoko; Putnam, Frank W. *J. Chromatogr.* **1983**, *266*, 511-522.
(13) Guiochon, G.; Gonnord, M. F.; Zakaria, M.; Beaver, L. A.; Siouffi, A. M. *Chromatographia* **1983**, *17*, 121-124.
(14) Davis, Gregory C.; Kissinger, Peter T. *Anal. Chem.* **1979**, *51*, 1960-1965.
(15) Lai, S. T.; Nishina, M. M.; Sangermano, L. *HRC CC, J. High Resolut. Chromatogr. Chromatogr. Commun.* **1984**, *7*, 336-337.
(16) Bhikhabhai, R.; Lindblom, H.; Källman, I.; Fägerstam, L. *Am. Lab.* **1979**, May, 76-81.
(17) Dunn, Michael H. *LC-GC* **1989**, *7*, 138-140.
(18) Erni, F.; Frei, R. W. *J. Chromatogr.* **1978**, *149*, 561-569.
(19) White, Jackie G.; St. Claire III, Robert L.; Jorgenson, James W. *Anal. Chem.* **1986**, *58*, 293-298.

RECEIVED for review August 17, 1989. Accepted October 19, 1989. Support for this work was provided by the Hewlett-Packard Corporation and American Cyanamid.

Near-Infrared Diffuse Reflectance Analysis of Athabasca Oil Sand

Robert C. Shaw

Research Department, Syncrude Canada Ltd., Edmonton, Alberta, Canada T6C 4G3

Byron Kratochvil*

Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2G2

The application of near-infrared diffuse reflectance analysis for the determination of bitumen content in samples of Athabasca oil sand is described. Quantitative analysis is performed using calibrations based on linear correlation between spectral features at selected wavelengths and compositional data for a series of known samples. The anhydrous bitumen content of oil sand within a length of core was obtained from spectra recorded at 1-cm sampling intervals. Analyses by a micro Soxhlet extraction procedure provided reference assay data for correlating sample composition and spectral response. The training set included representative samples of oil sand, indurated clay, coal, and interbedded regions. A model based on the first derivative of absorbance spectra provided the best relationship between spectral features and hydrocarbon concentration over a wide range of oil sand grades. This model was used to predict the bitumen content of samples along the entire length of the test core section, thus providing a high-resolution profile of grade variability as a function of depth.

Near-infrared diffuse reflectance (NIR-DR) spectroscopy has become widely accepted for the quantitative measurement of organic species, mineral constituents, and moisture in a variety of solid substrates (*1-4*). The technique is rapid, is nondestructive, and in many instances requires little or no sample preparation. Recent advances in instrumentation and software permit simultaneous multicomponent determinations to be performed on a single sample. Comprehensive reviews are available (*5, 6*).

The wavelength region of interest for NIR-DR work is typically 1000–2500 nm. Bitumen absorption in this region is relatively weak. As a consequence, instrumentation requirements are stringent. High signal to noise ratios require special consideration to factors such as source stability, optical design, sample positioning, and detector sensitivity. Unfortunately, structural information that can be deduced from spectra obtained in the mid-infrared region is often obscured in the near-infrared due to the multiplicity and overlap of higher frequency overtone bands. Further, specular reflection and the effects of particle size, sample packing, and surface irregularities can have a substantial impact on both the overall intensity of reflected radiation and the contrast between base-line and absorbance peaks.

To alleviate these apparent shortcomings, quantitative near-IR analysis is performed using a calibration curve developed by correlating detector response at a number of selected wavelengths to known compositions for a series of standard samples (training set). Multivariate regression techniques and sophisticated algorithms that correct for base-line anomalies, extract and reconstruct component

0003-2700/90/0362-0167$02.50/0

spectra from complex mixtures, and optimize the number of wavelength readings required for a robust and unbiased estimate for the concentration of each constituent in the sample serve to enhance the versatility of the method (*7–13*).

In connection with a sampling study of the heterogeneity of the Athabasca oil sand deposit, a need arose to perform a large number of measurements for bitumen content on relatively small portions of material. Methods currently in use for quantitative analysis require large samples and are slow and expensive (*14*). Accordingly, we investigated the use of near-infrared diffuse reflectance spectroscopy to provide compositional information on small quantities of oil sand.

EXPERIMENTAL SECTION

Samples. Samples were obtained from the Athabasca oil sand deposit in a region operated by Syncrude Canada Ltd. (*15*). The oil sand used in preliminary studies was freshly mined, stored in sealed containers, and ground with a mortar and pestle just prior to analysis. Samples used for subsequent work were obtained from a core supplied by Syncrude Canada Ltd., representing a typical vertical transect of the Athabasca deposit. This core, originally drilled in 1985, was approximately 120 m in length and encased in a plastic tube having an inside diameter of approximately 6.4 cm and wall thickness of 0.1 cm. As the result of earlier processing, the core had been cut in half longitudinally and a small V-notch sample removed from along the center line. The slabbing operation provides a uniformly textured flat surface for diffuse reflectance measurements. Grades within the ore zone of the core ranged from approximately 1 to 15% by weight bitumen on an anhydrous basis.

A small section of this core, just over 4 m in length, was selected for detailed NIR-DR analysis. The section was chosen to represent what appeared to be typical of grade variations seen over the entire length of core. The core sample, stored under ambient conditions since it was first analyzed, was almost completely dehydrated.

Instrumentation. Preliminary studies were conducted with a Nicolet 7199 Fourier transform infrared (FT-IR) spectrophotometer equipped with a diffuse reflectance attachment. Sample spectrta were recorded in absorbance mode ratioed against powdered KBr.

The extensive data base required for detailed inspection of core necessitated use of a spectrophotometer having more rapid data acquisition and processing capabilities. This was provided by a Quantum 1200 near-infrared analyzer (LT Industries, Inc.). Spectra were recorded over the range 1200–2400 nm in reflectance mode. Data were transformed as log $(1/R)$ to provide the corresponding absorbance spectra. The high-speed optics permitted spectra to be obtained at the rate of 5 scans/s with a spectral resolution of 1 nm. A Compaq Deskpro 286 computer ran the software driving the data acquisition, mathematical processing, modeling, and graphing functions for the analyzer.

The instrument was positioned on end such that the incident radiation could be directed downward onto the surface of the sample. A sampling table, consisting of a core holder, registration guide, and bridge assembly for mounting the spectrophotometer, was constructed to permit core to be rastered beneath the viewing port at 1-cm intervals. The viewing port itself was masked to provide a square sample window measuring 1 cm on a side. A schematic of the sampling table is provided in Figure 1.

Detector response was found to be extremely sensitive to sample elevation. Care was taken to ensure consistent positioning of the sample surface normal to the incident light and at a standard distance from the mask. Spectra generated by averaging the signal over 50 scans limited noise to approximately 0.5% relative. This was considered adequate for all subsequent work. Carbon black was used as reference material.

Digitized sample spectra were obtained for oil sand at 1-cm intervals along the length of the test core section and stored on hard disk in the computer. To establish a model relating the spectral features of these samples to oil sand composition, a reference library of spectra for known samples was assembled. With the spectra of samples in the core on file, it became a matter of analyzing a selected number to provide the requisite assay information. This training set was then used to generate correlations between spectra characteristics and sample composition.

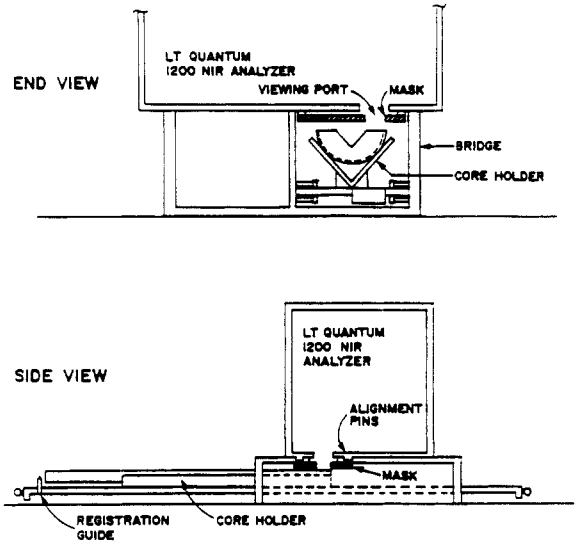

Figure 1. Schematic illustration of the sampling table constructed for core analysis.

Micro Soxhlet Extraction Procedure. A total of 48 samples were selected to represent the broad compositional range of material contained within the test section of core. Samples, each weighing on the order of 1–2 g, were removed from the core by using a scalpel, and their position along the core was labeled so that each could be matched with a corresponding NIR-DR spectrum.

The training set samples were ground in a mortar and pestle. A subsample of the homogenized material was then extracted in a micro Soxhlet assembly by using a 10 × 50 mm cellulose thimble and 10 mL of toluene. Extraction was complete in 10 min. The solvent extract was quantitatively transferred to a tared evaporating dish, allowed to dry overnight at room temperature, and then weighed to obtain the bitumen value. Previous experiments had shown that traces of toluene remaining in the bitumen (as measured by IR upon dissolving the residue in trichloroethylene) can be considered negligible. The tared thimble was dried at 120 °C for 1.5 h and reweighed to obtain the solids content.

Results of replicate determinations on a variety of oil sand samples in which both the thimbles and evaporating dishes were dried in a forced draft oven for 24 h at 120 °C showed consistently low mass closures (the ratio of the sum of the measured bitumen and solids mass to the original sample mass), indicating that measures for the bitumen and/or solids content of the sample are underestimates of the true composition. Sources of error investigated included loss of bitumen light ends upon drying the solvent extract at elevated temperatures, loss of hygroscopic water from thimbles and oil sand samples, and incomplete retention of fine particles in the thimble. The first two sources of bias would lead to low recoveries. The latter problem, if present, would not be detected by examining mass closure data. Solids lost through the thimble would ultimately be measured as bitumen. The toluene used for extraction contained no measurable residue upon evaporation.

A comparison of bitumen weights obtained from sample extracts evaporated under ambient conditions and then dried in an oven for 16 h at 120 °C indicated an average weight loss upon heating on the order of 0.01–0.02 g for bitumen residues weighing approximately 0.2 g. Blank thimble extractions were also done to see whether adsorbed water could affect assay results. Weight losses on the order of 0.01 g were seen upon drying thimbles for 1.5 h at 120 °C. To minimize the effects of these systematic errors, sample thimbles must be dried prior to analysis and extract solutions evaporated to dryness at room temperature.

The precision of the analyses (means based on eight measurements, assay results expressed in weight percent) for low-grade oil sand was as follows: bitumen, 6.05 ± 0.22; solids, 93.56 ± 0.25; mass closure, 0.9962 ± 0.0006. Corresponding results for medium-grade oil sand were as follows: bitumen, 8.54 ± 0.26; solids, 90.69 ± 0.29; mass closure, 0.9923 ± 0.0009. Results for high-grade oil sand were as follows: 16.74 ± 0.19; solids, 82.78 ± 0.18; mass

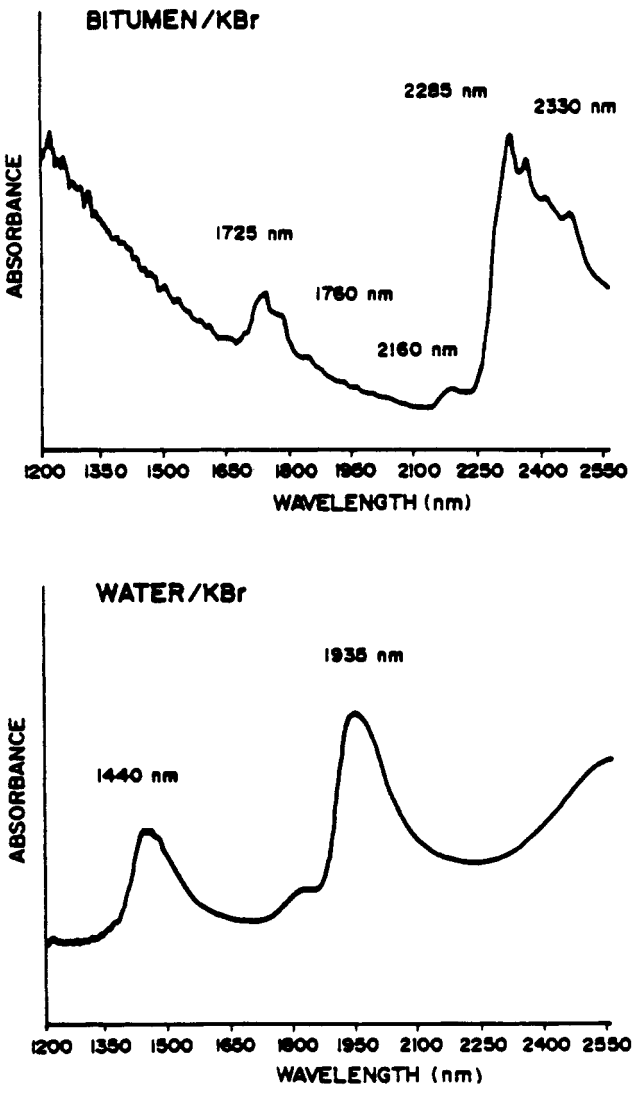

Figure 2. Near-IR diffuse reflectance spectra for bitumen and water dispersed in KBr.

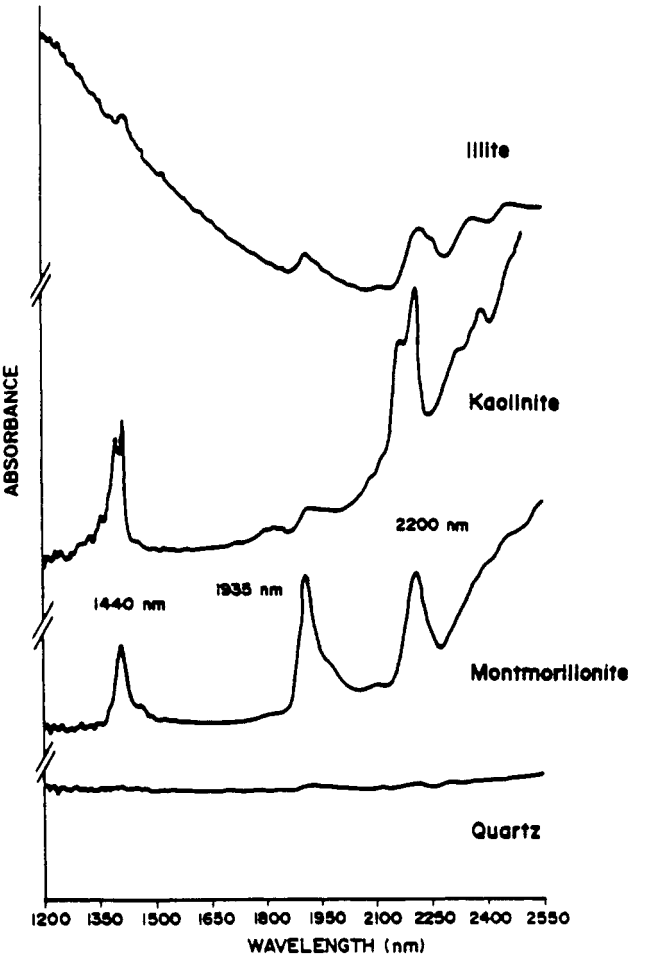

Figure 3. Near-IR diffuse reflectance spectra for various clay minerals and quartz.

closure, 0.9952 ± 0.0005. Uncertainties are 1 standard deviation. It should be noted that differences among these replicate sample analyses probably reflect sampling variability more than actual analytical variability.

RESULTS AND DISCUSSION

Preliminary Studies with the Nicolet 7199 Spectrophotometer. Spectra obtained for bitumen, water, quartz, and a variety of clays normally present in oil sand are shown in Figures 2 and 3. The bitumen spectrum shows two distinctive sets of doublets, one with major peaks at 2285 and 2330 nm and the other with peaks at 1725 and 1760 nm. These absorbances, along with the much weaker bands at 2400 and 2460 nm, can be ascribed to overtones and combinations of the fundamental vibration frequencies within consitituent methyl and methylene groups. Although the absorbance peak at 2160 nm may signify the presence of aromatic hydrocarbon species, it could also be attributed to contamination by clay, most probably kaolinite.

The diffuse reflectance spectrum for water indicates two strong absorbance peaks at 1440 and 1935 nm. These bands, representing combinations of fundamental vibrations associated with free water molecules, are also seen in the spectra for illite, kaolinite, and montmorillonite. These three spectra also show a major band in the region around 2200 nm, which is diagnostic of aluminum-bearing clays.

Differences in the spectral patterns for the various clays are significant. In the spectrum of kaolinite, for example, peaks centered at 1440 and 2200 nm are resolved as doublets. Only single peaks at these wavelengths are observed for illite and montmorillonite. The splitting of the 1440-nm absorbance band may indicate differences in coupling between the vibrational frequencies of free hygroscopic water and those associated with water of hydration. In contrast with the spectrum for montmorillonite, the free water absorbance at 1935 nm for kaolinite and illite is relatively weak. Further, illite exhibits a progressive increase in absorbance (diminishing reflectance) toward shorter wavelengths in a manner similar to that of bitumen, whereas the opposite trend is evident in the spectra for other types of clay.

The near-infrared diffuse reflectance spectra for fresh and dehydrated high-grade oil sand (approximately 13% bitumen content) in Figure 4 show many of the significant features of the individual component spectra. The free water absorbance band at 1935 nm is absent in the spectrum for the dehydrated sample. Overall reflectivity is affected by the water content of the material. Moisture in fresh oil sand contributes to an increase in absorbance readings at all wavelengths, presumably due to internal reflections within thin films of water. To eliminate uncertainties in absorbance measurements, subsequent spectra were obtained by using aged oil sand samples dried by exposure to the atmosphere. Samples were ground to a fine powder with a mortar and pestle and placed in the sample cup without compaction.

Figure 5 presents spectra obtained for rich-, medium-, and low-grade oil sands. The reduction in the peak heights of the major bitumen absorbance bands (relative to an imposed base line) with decreasing oil saturation is evident.

Differences in the spectra patterns between ores of marine and estuarine origin can be seen in the near-IR diffuse re-

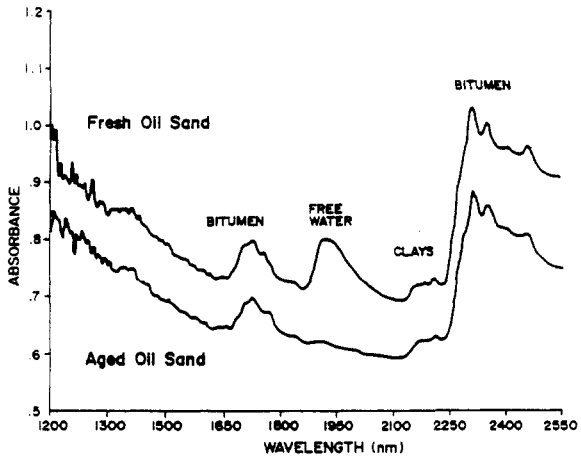

Figure 4. Comparison of near-IR diffuse reflectance spectra for fresh and aged samples of oil sand.

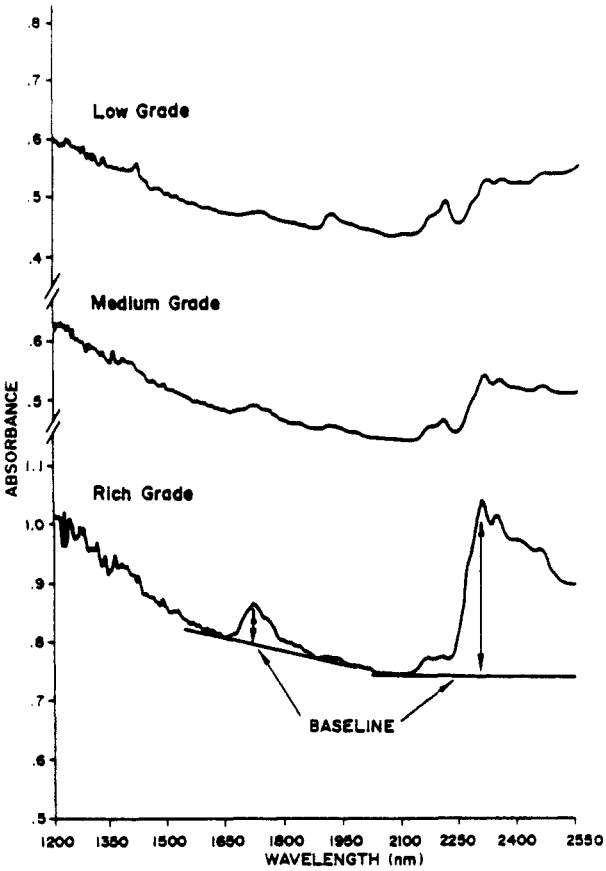

Figure 5. Comparison of near-IR diffuse reflectance spectra for oil sands of varying bitumen content.

flectance spectra presented in Figure 6. Although the two test portions studied here are of similar composition (approximately 13% bitumen content), the total amount of light reflected from the marine sample is markedly less than that from the estuarine material. Further, the spectrum for the marine ore tends to have a shallower slope at short wavelengths, leading to less pronounced curvature in the spectrum toward the higher wavelength region. The ratio of the absolute absorbance at 1200 nm to that of the bitumen peak at 2285 nm is near unity for the estuarine ore and somewhat greater for the marine ore. These traits may be useful to distinguish between oil sands from differing depositional environments.

The reproducibility of diffuse reflectance measurements is affected by the nature of the sample surface presented for analysis. Replicate spectra recorded for a sample of oil sand under various conditions of surface preparation are shown in

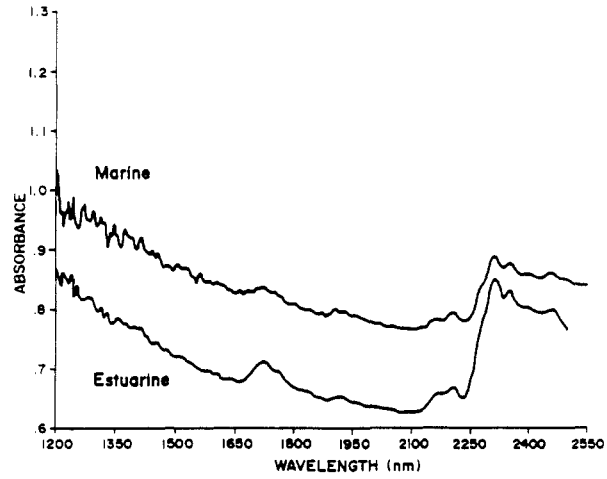

Figure 6. Comparison of near-IR diffuse reflectance spectra for oil sands of marine and estuarine origin. Both contain approximately 13% bitumen.

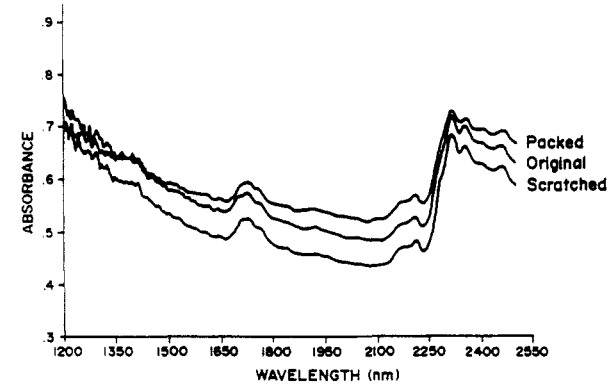

Figure 7. Effect of surface preparation procedure on the diffuse reflectance spectra of an oil sand sample.

Figure 7. The original surface is that obtained by placing the powdered material into the sample cup without compaction. The surface was leveled by screeding the sample with a straight-edge blade. The spectrum for the sample generated after tamping the material into the sample cup with a spatula shows a general decrease in absolute reflectance over much of the wavelength range. At first glance, this would not constitute a serious problem since quantitative analysis is concerned with the relative absorbance from a base line that is drawn with respect to the shape and positioning of the curve. Closer inspection of the spectrum, however, reveals that the peak height of the absorbance band at 2285 nm is diminished in relation to that of the original, uncompacted sample surface. This would lead to an underestimation of bitumen content if a calibration curve had been developed by using the initial sample handling procedure. Roughening of the packed sample by scratching the surface with a pin appears to restore the integrity of the spectrum, albeit with a higher overall reflectivity. Results indicate that methods used to prepare samples for diffuse reflectance analysis must be consistent from sample to sample.

Relationships between the bitumen content and absorbance for peaks centered at 1725 and 2285 nm for a variety of oil sand samples are shown in Figure 8. Differences in the slope of the calibration curves for oil sands of marine and estuarine origin are clearly evident. A significant inverse relationship was also found to exist between absorbance measurements for clays in the region of 2200 nm and bitumen concentration for samples of estuarine ore. The correlation between measured bitumen assay and that predicted from a linear combination of spectral readings from all three wavelengths for this type of oil sand is illustrated in Figure 9.

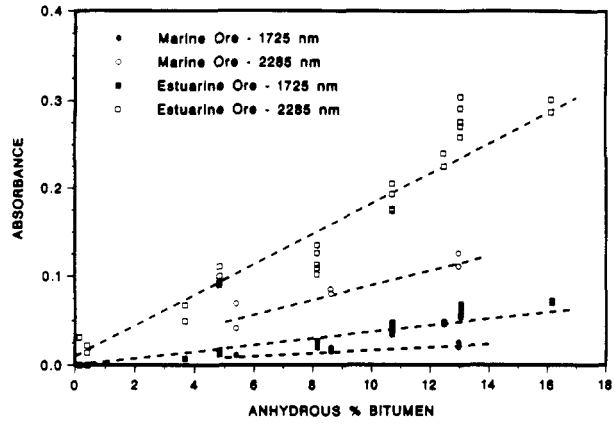

Figure 8. Correlation between anhydrous bitumen content and relative absorbance at 1725 and 2285 nm for oil sand samples of marine and estuarine origin. Replicate absorbance readings obtained for each sample provide an indication of the uncertainty of the measurement.

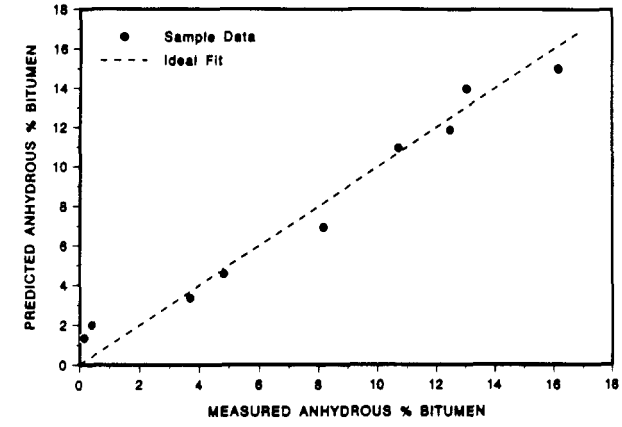

Figure 9. Comparison of predicted versus measured bitumen content for estuarine ore samples based on linear regression of absorbance measurements at 1725, 2200, and 2285 nm.

Results generated by using the Nicolet FT-IR indicated that the bitumen content of oil sand samples can be determined from near-infrared absorbance measurements. Combining the basic technique with more sophisticated data handling and multivariable statistical analysis procedures for correlating spectral response and bitumen assay can greatly enhance the utility of the method. The remainder of this report deals with the analysis of oil sand in a core using a microprocessor-based instrument specifically designed for NIR-DR analysis.

Reproducibility of Sample Spectra Obtained with the Quantum 1200 Near-IR Analyzer. A total of 10 replicate analyses were performed on samples of core (oil sand, clay, and interbedded material) and carbon black to evaluate the precision of absorbance measurements. The core and background samples were removed from the sample table and repositioned between scans. The mean, minimum, and maximum absorbance spectra for each of the core samples are presented in Figure 10. Features characteristic of the hydrocarbon and mineral components of the samples are clearly evident in the spectra. It is also apparent that differences in the overall reflectivity of a sample can be used to distinguish oil sand and clay. The spectra of the interbedded sample (a stratified mixture of oil sand and clay) reveal the traits of both parent materials.

Curvature in plots of standard deviation versus wavelength (not shown) indicates that the precision of absorbance values for a given sample is not constant. The magnitude of the uncertainty varies with wavelength and tends to be inversely related to absorbance.

To provide a common basis for comparing the variability among spectra for each type of core sample, spectra were

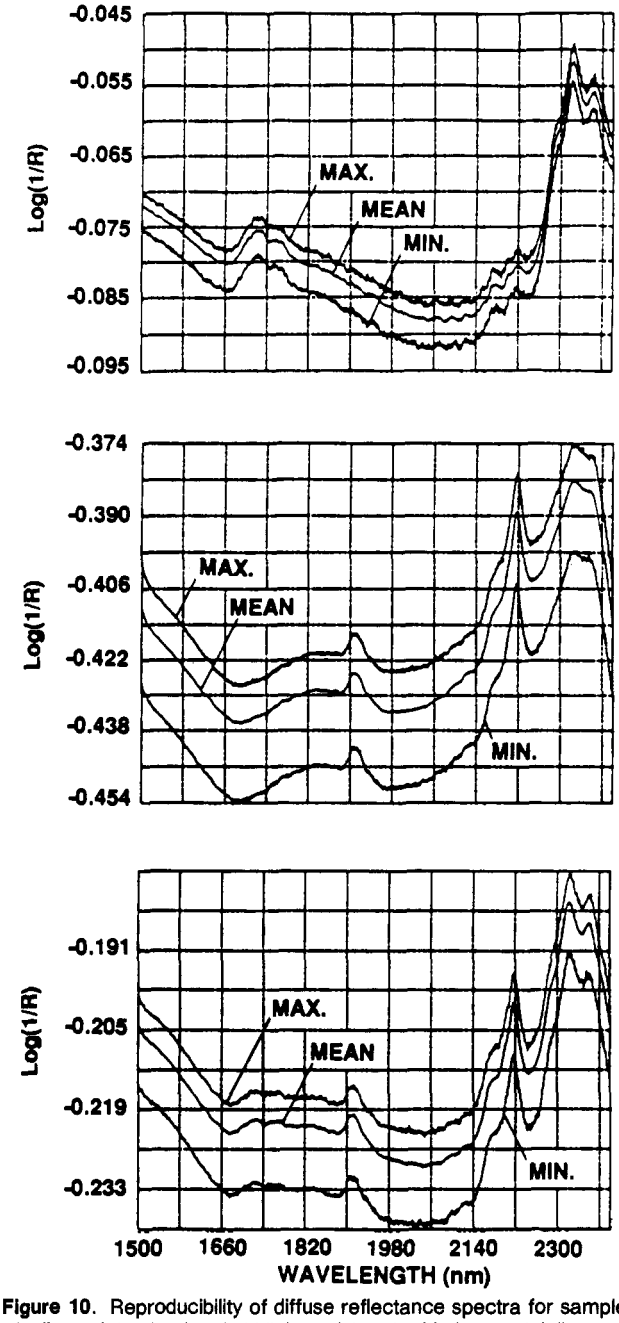

Figure 10. Reproducibility of diffuse reflectance spectra for samples of oil sand (top), clay (middle), and interbedded material (bottom). Results are based on 10 replicate analyses (background is carbon black).

characterized in terms of an average absorbance and standard deviation calculated over the wavelength range 1500–2400 nm based on data points taken at 50-nm intervals. Results were −0.0780 ± 0.0019, −0.2161 ± 0.0047, and −0.4203 ± 0.0077 for oil sand, interbedded material, and clay, respectively. The uncertainty among replicate sample spectra is on the order of 2% relative.

The variability among spectra for a particular sample may arise from uncertainties associated with the background correction or inconsistencies in positioning the sample surface (spatial displacement, either lateral or vertical). Given an independent estimate for the standard deviation among replicate absorbance spectra of the reference material, it is possible to determine the relative contribution of each source of error to the total based on the additivity of variances. Replicate absorbance spectra obtained for carbon black (background = 1.0) indicated that the standard deviation is reasonably constant over the wavelength range 1500–2400 nm,

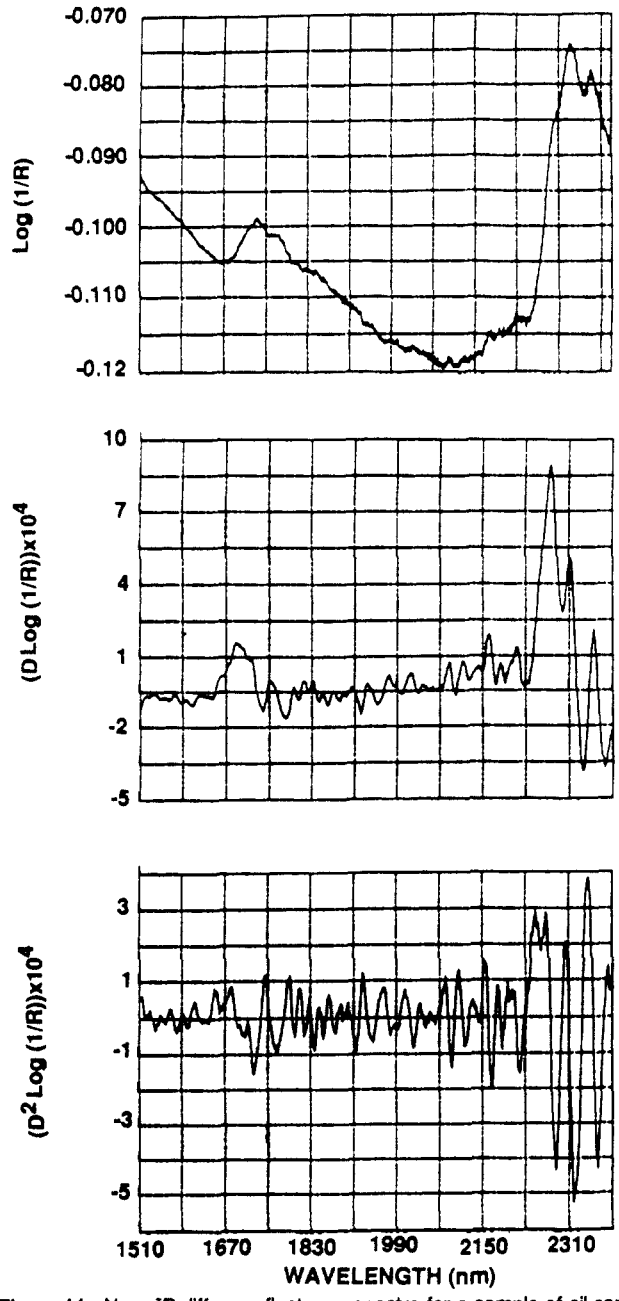

Figure 11. Near-IR diffuse reflectance spectra for a sample of oil sand in the training set illustrating log (1/R) (top), first derivative (middle), and second derivative (bottom) transformations.

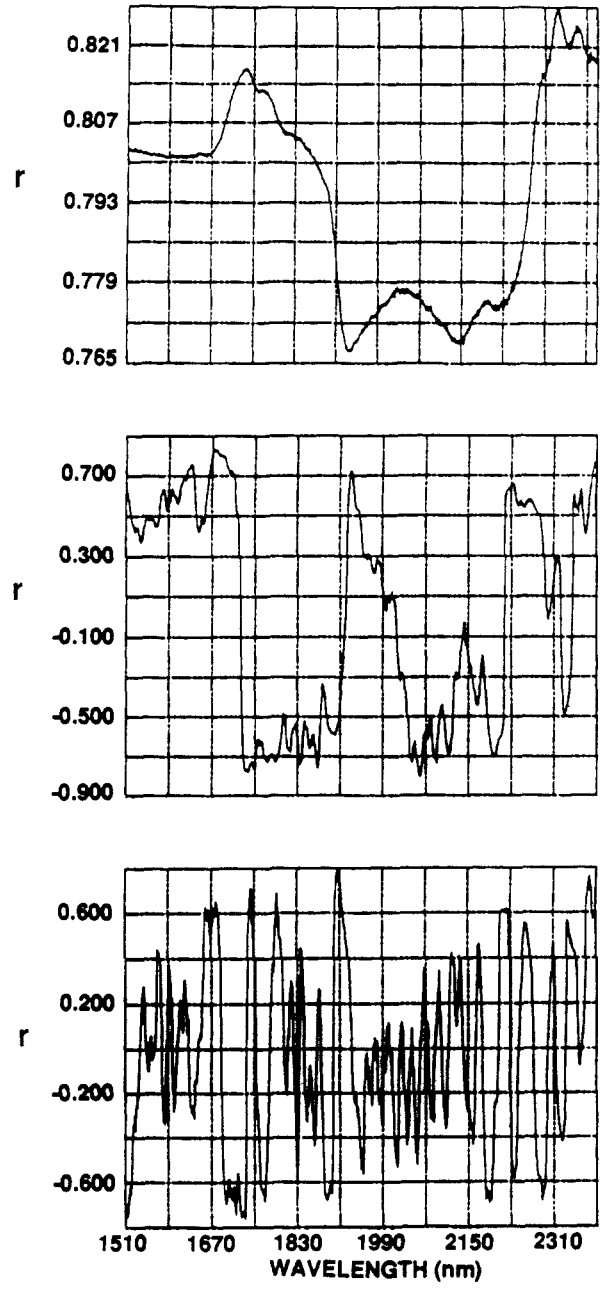

Figure 12. Correlation coefficients for the relationships between compositional data and spectral features as a function of wavelength for samples in the training set based on log (1/R) (top), first derivative (middle), and second derivative (bottom) data.

averaging 0.00045 absorbance unit (obtained in the same manner as previously discussed for the core samples). Since this value is small in comparison to the standard deviations calculated for the background-corrected spectra of core samples, it appears that uncertainties due to sample positioning are the dominant source of error.

Training Set and Data Modeling. Typical spectra obtained for an oil sand sample included in the training set are illustrated in Figure 11. The first- and second-derivative spectra were generated from log (1/R) transformed data.

Correlation coefficients between compositional data and spectral features as a function of wavelength for samples in the training set are shown in Figure 12. A direct relation between absolute absorbance and bitumen content is observed over virtually all wavelengths in the range 1500–2400 nm. From these results, it appears that a model relating sample bitumen content and absorbance can be based directly on the intensity of light reflected from the sample surface at any given wavelength. Unlike the preliminary correlations established during studies with the Nicolet FT-IR, in this work we are not looking at the peak height of absorbance bands from some imposed base line.

Correlations using the first or second derivative of the absorbance spectrum are considerably more discriminating than those based on absolute reflectivity. By virtue of the mathematical processing, differences in the base-line intensities of the reflected light are normalized. Derivative spectra are thus more descriptive of the actual absorbance peaks. In these instances there is a wide range of variation in the correlation coefficient as a function of wavelength. Models based on derivative spectra are therefore likely to provide a more robust relationship between spectral features and sample composition.

Stepwise linear regression was used to model the data from the training set in the form

$$Y = C_0 + C_1(W_1) + C_2(W_2) + \cdots + C_n(W_n)$$

where Y is the percent bitumen content and W_1, W_2 through

Table I. Regression Statistics for Training Set Models Based on First Derivative of log (1/R) Data

wavelength	model coeff	std error	t value
1984.0	0.7494E+05	0.4727E+04	0.1585E+02
2234.0	0.1403E+05	0.1503E+04	0.9337E+01
1799.0	−0.5040E+05	0.5618E+04	0.8972E+01
2020.0	−0.3181E+05	0.6433E+04	0.4945E+01
intercept	0.1454E+02		
multiple corr coeff (r)			0.9728
F-test			136.6
SEC, % bitumen			1.28
SEP, % bitumen			1.62

wavelength	model coeff	std error	t value
1690.0	0.6326E+05	0.7213E+04	0.8770E+01
2210.0	−0.1082E+05	0.1962E+04	0.5516E+01
2280.0	−0.1177E+05	0.2346E+04	0.5016E+01
intercept	0.1458E+02		
multiple corr coeff (r)			0.9220
F-test			60.5
SEC, % bitumen			2.10
SEP, % bitumen			2.46

W_n represent the magnitude of a spectral feature (absolute absorbance or derivative value) at wavelengths for which there is correspondence with sample composition. The coefficients C_0, C_1, C_2 through C_n are the associated weighting factors.

To reduce the problem to a manageable level, it was assumed that the magnitude of a spectral feature at a wavelength for which there exists a high correlation between response and sample assay serves to explain most of the variability in the data. This simple model is given by the expression

$$Y = C_0 + C_1(W_1)$$

Any lack of fit is caused by second-order effects, which can be modeled by using additional wavelengths. The selection and addition of subsequent terms are based on their ability to reduce residual error (determined by means of an F-test) and are directly related to an improvement in the multiple correlation coefficient.

A preliminary evaluation of composite models derived by using the spectra (obtained in absorbance, first-derivative, and second-derivative modes) and assay information for all samples in the training set indicated that a model based on first-derivative data would provide the best calibration for sample bitumen content. In all cases, the primary relationship between sample composition and spectral feature was determined for each of 10 starting wavelengths. These wavelengths corresponded to the 10 positions of highest correlation between sample assay and spectral response. This initial set having been established, up to four additional terms were added to reduce the sum of residual errors. The best wavelengths selected for modeling the first-derivative spectra were 1984, 2234, 1799, and 2020 nm.

Armed with the knowledge of the preferred wavelengths for calibration, we developed a regression model using the first-derivative spectra and assay results for 36 samples selected at random from the training set. The remaining sample data in the training set would be used to validate the model in terms of its predictive capability.

The standard error of calibration (SEC) and standard error of prediction (SEP) are statistics commonly used to evaluate the performance of regression models (*16–18*). These are defined as

$$\mathrm{SEC} = \left[\frac{1}{n-p-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right]^{1/2}$$

$$\mathrm{SEP} = \left[\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2\right]^{1/2}$$

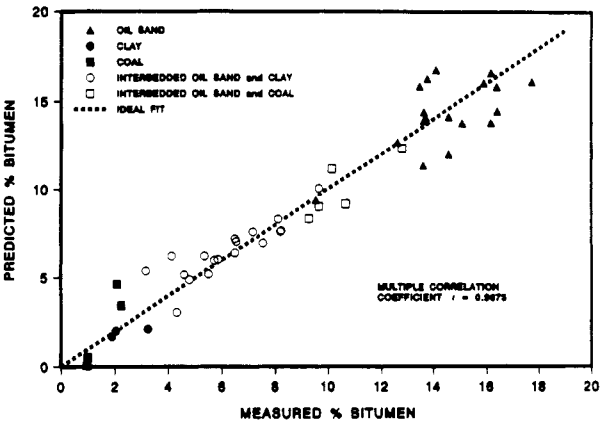

Figure 13. Comparison of predicted versus measured bitumen content for core samples in the training set using the best fit model based on the first derivative of log (1/R) spectral data.

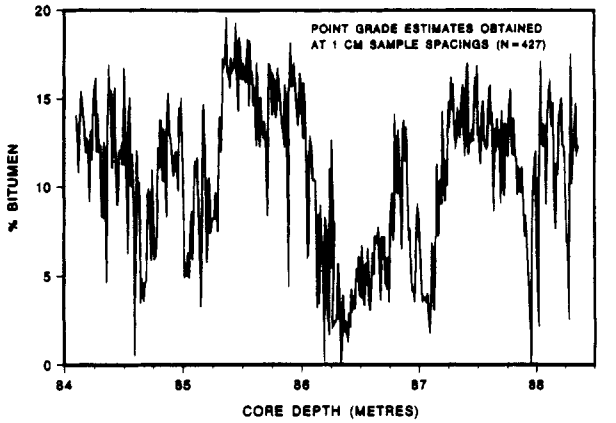

Figure 14. Variation in bitumen content as a function of depth for a 4-m test section of core. Point grade estimates were derived from NIR-DR spectral measurements at 1-cm sample spacings.

where n is the number of calibration samples, p is the number of parameters in the regression equation, m is the number of samples in the validation set, y_i is the measured assay of the ith sample, and $\hat{y}_i$ is the corresponding predicted value from the algorithm.

Regression statistics for the model are presented in Table I. The SEP is comparable to the SEC, providing evidence that the model is reasonable and does not overfit the data. Results are compared to those from a second model developed by using the same grouping of samples from the training set but based on first-derivative values at 1690, 2210, and 2280 nm. These wavelengths represent the points of steepest slope for bitumen and clay absorbances at 1725, 2200, and 2285 nm. Although these three wavelengths have chemical significance, the performance of this model is clearly inferior to that of the former.

When all 48 samples in the training set were merged to enhance the predictive power of the model, the resulting SEC was 1.32% bitumen (multiple correlation coefficient r = 0.9673). Inclusion of the validation samples results in only minor perturbations to the model coefficients and the standard error of the estimate. A plot of the predicted grades versus measured bitumen concentration for the training set samples is shown in Figure 13.

The model was used to establish the bitumen content of all samples analyzed by NIR-DR in the test section of core. The resulting grade profile is shown in Figure 14. A high degree of variability is observed not only on a scale of meters, but also at the centimeter level. This variability is significant and has implication with respect to the design of sampling plans for assessing bitumen content in sections of core.

In summary, NIR-DR analysis appears to be a useful tool for estimating the amount of bitumen in oil sand. Through the use of training sets, multiple wavelength calibration, and mathematical processing of spectra, good correlations can be obtained between NIR-DR measurements and micro Soxhlet extraction results.

ACKNOWLEDGMENT

Special thanks are extended to the staff of the Department of Chemistry Machine Shop, University of Alberta, for the design and construction of the core sampling table.

LITERATURE CITED

(1) Wetzel, D. L. *Anal. Chem.* **1983**, *55*, 1165A-1175A.
(2) Fysh, S. A.; Swinkels, D. A. J.; Fredericks, P. M. *Appl. Spectrosc.* **1985**, *39*, 354-357.
(3) Fredericks, P. M.; Tattersall, A.; Donaldson, R. *Appl. Spectrosc.* **1987**, *41*, 1039-1042.
(4) Crowley, J. K.; Vergo, N. *Clays Clay Miner.* **1988**, *36*, 310-316.
(5) McDonald, R. S. *Anal. Chem.* **1986**, *58*, 1906-1925.
(6) Stark, E.; Luchter, K.; Margoshes, M. *Appl. Spectrosc. Rev.* **1986**, *22*, 335-399.
(7) Honigs, D. E.; Hieftje, G. M.; Hirschfeld, T. *Appl. Spectrosc.* **1984**, *38*, 317-322.
(8) McClure, W. F.; Hamid, A.; Giesbrecht, F. G.; Weeks, W. W. *Appl. Spectrosc.* **1984**, *38*, 322-329.
(9) Hruschka, W. R.; Norris, K. H. *Appl. Spectrosc.* **1982**, *36*, 261-265.
(10) Beebe, K. R.; Kowalski, B. R. *Anal. Chem.* **1987**, *59*, 1007A-1017A.
(11) Montalvo, J. G.; Faught, S. E.; Buco, S. M. *Am. Lab.* **1986**, *18*, 37-55.
(12) Mark, H. *Anal. Chem.* **1986**, *58*, 2814-2819.
(13) Honigs, D. E.; Hieftje, G. M.; Hirschfeld, T. *Appl. Spectrosc.* **1984**, *38*, 844-847.
(14) Bulmer, J. T.; Starr, J. *Syncrude Analytical Methods for Oil Sand and Bitumen Processing*; Alberta Oil Sands Technology and Research Authority: Edmonton, AB, Canada, 1979.
(15) O'Donnell, N. D.; Ostrowski, T. B. *Applied Mining Geology: Ore Reserve Estimation*; Society of Mining Engineers, Inc.: Littleton, CO, 1986; pp 123-135.
(16) Puchwein, G. *Anal. Chem.* **1988**, *60*, 569-573.
(17) Cowe, I. A.; Koester, S.; Paul, C.; McNichol, J. W.; Cuthbertson, D. C. *Chemom. Intell. Lab. Syst.* **1988**, *3*, 233-242.
(18) Wentzell, P. D.; Wade, A. P.; Crouch, S. R. *Anal. Chem.* **1988**, *60*, 905-911.

RECEIVED for review October 26, 1989. Accepted November 6, 1989. The authors are grateful to Syncrude Canada Ltd. for providing facilities and financial support of this project.

Normalized Measure of Overlap between Non-Gaussian Chromatographic Peaks

Eric V. Dose and Georges Guiochon*

Department of Chemistry, University of Tennessee, Knoxville, Tennessee 37996, and Analytical Chemistry Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

A measure of overlap Ω between two chromatographic peak shapes is defined for the general case. Properties of invariance to chromatographic time and concentration scales, commutativity with respect to peak identity, and scaling of Ω to the range [0,1] are demonstrated. Analytical expressions for Ω between Gaussian, Lorentzian, rectangular, right-triangular, and isosceles-triangular peak shapes are derived and discussed. The measure Ω appears to be a natural measure of the power of a preparative-chromatographic method to enrich two components of a solution much as the resolution R_s measures the power of an analytical method to resolve two peak shapes for analyte quantitation.

The theory of chromatography has been applied for over 40 years to improving separations of solution components (*1*, *2*) and to the dependence of degree of separation on column efficiency and selectivity (*2-4*). Most measures of the degree of separation have been limited to peaks of Gaussian (*1*, *2*, *5*) or similar shapes (*6-9*). Non-Gaussian peak shapes obtained in chromatographic systems with nonlinear isotherms or slow retention kinetics sometimes result from unavoidable system conditions such as slow protein diffusion, low stationary-phase capacity, or stationary-phase inhomogeneity. In preparative chromatography, non-Gaussian peak shapes often result from deliberate attempts to overload the column, which may compress the component bands into small eluted volumes, displace other components in order to maximize local competition for sites and thus minimize band overlap, and separate more component material in a single pass. This last practice is designed to increase refined-component production rate with less than a proportional loss of purity.

Several measures of separation have been proposed in the literature. Two survey articles (*9*, *10*) compare several quality criteria, although one (*10*) applies them only to peaks of Gaussian shape. A very early measure, one incorporated in some more complex expressions of later workers, is the fractional overlap at equal impurity (*1*). When fractions are divided at the time where the impurity percentages are equal, fractional overlap is defined as the amount of component A eluting after the cut between early-eluting component A and later-eluting component B, or of B in the fraction containing A. Since the time dividing the fractions is unambiguous, the fractional overlap is uniquely defined. However, it does not account in any way for the distribution over time of the impurity component in a fraction, and so it gives no indication of how much improvement in fraction yield or purity would be possible if fractions were taken differently. The fraction-dividing time which gives equal impurity for the two fractions is also difficult to determine for many non-Gaussian peaks. In fact, where the sample has very different amounts of two closely eluted components, as is common in preparative chromatography, all fractions of the minor component may be less pure than the least pure fraction of the major component. In this case, the fraction-dividing time defined in the above manner does not exist.

Later measures (*1*, *11*) built on the fractional overlap suffer the same limitations. Massart (*11*) applies the concept of the informing power of a signal (*12*) to the amount of resolution between two peaks in a chromatogram. The method requires that the time axis be divided into best fractions and that the level of impurity in each fraction be taken as the fractional overlap between the main component and the other components. The informing power is inversely related to the product of all the fractional overlaps in the chromatogram. The

0003-2700/90/0362-0174$02.50/0