J. B. F. Lloyd,[1] *Ph.D.; I. W. Evett,*[2] *B.Sc., Dip.Stat.; and J. M. Dubery,*[1] *M.A., M.S.*

Examination of Petroleum Products of High Relative Molecular Mass for Forensic Science Purposes by Synchronous Fluorescence Spectroscopy. II: Discrimination Within an Arbitrary Set of Representative Samples

REFERENCE: Lloyd, J. B. F., Evett, I. W., and Dubery, J. M., **"Examination of Petroleum Products of High Relative Molecular Mass for Forensic Science Purposes by Synchronous Fluorescence Spectroscopy. II: Discrimination Within an Arbitrary Set of Representative Samples,"** *Journal of Forensic Sciences,* JFSCA, Vol. 25, No. 3, July 1980, pp. 589-602.

ABSTRACT: Synchronous fluorescence and ultraviolet absorbance spectroscopy are used in the examination of a collection of high relative molecular mass petroleum products that contains 118 samples drawn from forensic science work. Data from 22 defined spectral features of each sample are processed to show that within this sample set, and with a considerable reduction in the data finally used, the spectral patterns can be efficiently retrieved with a high level of sample discrimination and with an accompanying low level of spurious choices. The system gives results in good agreement with earlier conclusions drawn from case work on the basis of visual comparisons of spectral patterns of control samples and contact traces and will enable future evidential data to be more efficiently collected and its significance to be more precisely assessed.

KEY WORDS: criminalistics, petroleum products, spectroscopic analysis

The application of luminescence spectroscopy in the examination of contact traces containing high molecular mass petroleum products has been discussed in general terms in Part I of this series of papers [*1*]. Since the introduction of these techniques to forensic science, particularly in the form of synchronous fluorescence spectroscopy [*2*], they have been used for essentially qualitative comparisons based on uncorrected spectra, which are highly dependent in form on the instrumentation that produces them. Such spectra tend to vary with time and consequently are not suited to the production of data usable in the long term. Evidential significance, therefore, has been assessed on a subjective basis.

Commercial instrumentation that generates relatively undistorted and stable spectra is now widely available. It is thus feasible to define spectral features upon which a data collection appropriate to forensic science use can be based, provided that other variables

Received for publication 17 Nov. 1979; revised manuscript received 26 Jan. 1980; accepted for publication 28 Jan. 1980.

[1]Senior principal scientific officer and senior scientific officer, respectively, Home Office Forensic Science Laboratory, Gooch St. North, Birmingham, United Kingdom.

[2]Principal scientific officer, Home Office Central Research Establishment, Aldermaston, Reading, Berkshire, United Kingdom.

are adequately controlled [*1*]. (The defined features, given in Table 1, are derived from synchronous fluorescence spectra, in deoxygenated solvent at ambient temperatures, corrected for instrumental and inner filter effects, and also from ultraviolet absorbance spectra). The present paper is concerned with the effectiveness with which such features may be used in the differentiation and retrieval of spectrally derived data in a file representing a range of products typical of and largely drawn from forensic science work. We emphasize that the results refer to commonly encountered materials in a general sense and not to a specific type of case. Before data relevant to a particular set of circumstances can be assembled, and subsequently efficiently retrieved, it is in our view essential that the relationships between the different features be understood in order that redundant or ineffective features may be eliminated, that key features may be identified, and, hence, that the most efficient use may be made of the necessary computer-based retrieval techniques.

A considerable number of other techniques are available for the examination of petroleum products of the sort considered here [*3*]. Many of these techniques are ap-

TABLE 1—*Defined spectral features in the synchronous fluorescence and UV absorbance spectra of petroleums.*

Feature	Spectrum[a]	Definition[b]
1	UV	(sum of absorbances at the excitation and emission wavelengths of Feature *2*) ÷ (sample concentration in units of g/mL)
2	D34	(maximum intensity before 282 nm) − (intensity of blank at corresponding wavelength)
3	D34	(maximum intensity in range 292 to 302 nm) − (minimum intensity in range 282 to 292 nm)
4	D34	(intensity at 313 nm) − (minimum intensity in range 302 to 308 nm)
5	D34	(intensity at 324 nm) − (blank intensity)
6	D34	(maximum intensity in range 343 to 356 nm) − (minimum intensity in range 336 to 343 nm)
7	D34	wavelength of Feature *2*
8	D34	wavelength of Feature *6*
9	D20	(maximum intensity in range 266 to 277 nm) − (blank intensity)
10	D20	(intensity at 278 nm) − (minimum intensity in range 278 to 283 nm)
11	D20	(maximum intensity in range 283 to 298 nm) − (minimum intensity in range 278 to 283 nm)
12	D20	(maximum intensity in range 303 to 312 nm) − (minimum intensity in range 298 to 303 nm)
13	D20	(maximum intensity in range 323 to 331 nm) − (minimum intensity in range 316 to 323 nm)
14	D20	(maximum intensity in range 379 to 385 nm) − (minimum intensity in range 370 to 379 nm)
15	D20	wavelength of Feature *11*
16	D7	(maximum intensity in range 270 to 284 nm) − (blank intensity)
17	D7	(maximum intensity in range 298 to 312 nm) − (minimum intensity in range 284 to 298 nm)
18	D7	(maximum intensity in range 324 to 334 nm) − (minimum intensity in range 312 to 324 nm)
19	D7	(maximum intensity in range 351 to 359 nm) − (blank intensity)
20	D7	(maximum intensity in range 388 to 398 nm) − (blank intensity)
21	D34 UV	(Feature *2*) ÷ [(absorbance sum from Feature *2*) × (D34 solvent Raman intensity at 324 nm)]
22	D34 UV	(Feature *1*) × (Feature *21*) × 10^{-3}

[a]UV is the ultraviolet absorbance spectrum; D34, D20, and D7 are the synchronous fluorescence spectra, where the number indicates the scanning interval.

[b]The given wavelengths are excitation wavelengths; the corresponding emission wavelengths are determined by the scanning interval. These and further details are given in Ref *1*.

plicable to forensic science. For example, a variety of chromatographic techniques may be used in the separation of lower molecular mass components, particularly additives, from small samples, and some components may be determined by spectroscopic techniques applied to the samples directly. But most of the fluorescence intrinsic to these products is due to a highly complex mixture of polyalkylated aromatic compounds that, to date, has yielded few discrete chromatographic peaks, although some fractionation is possible [*4–6*].

The fluorescence technique, applied directly, has given results rapidly from samples that in some circumstances could not be analyzed in other ways because, for instance, the sample was very small, or associated with materials interfering in other techniques, or was not extractable from a matrix without loss of components upon which alternative techniques depended. On other occasions, the fluorescence technique may be subject to similar restrictions. Again, it may sometimes seem reasonable to assume that the fluorescence results are uncorrelated to those from other techniques. But the validity of the assumption, as well as the value of the fluorescence technique relative to others under various circumstances, and the evidential significance of its results cannot be objectively assessed without an appropriate numerical representation of the spectra.

Data Sources and Processing

Spectra

Full details of the experimental technique have been given previously [*1*]; they are summarized here.

The fluorescence spectra were recorded with a Perkin-Elmer MPF-4 spectrometer fitted with a corrected spectra accessory, and the ultraviolet (UV) spectra with a Pye-Unicam SP-8000 spectrometer. For the synchronous fluorescence spectra (in the corrected excitation mode), at scanning intervals of 34, 20, and 7 nm, the samples were dissolved at a concentration usually of 20 μg/mL in purified cyclohexane, the solution was deoxygenated with a stream of nitrogen in a 5-mm cuvette at ambient temperature, and the spectra were recorded at a nominal optical depth of 2.5 mm with right-angle illumination.

The various features (Table 1) were measured on the spectra and corrections made as necessary for any reagent blank and for instrumental and inner filter effects [*1*]. The fluorescence intensity values (that is, Features *2–6, 9–14, 16–20;* Table 1) were finally normalized to give a value for their sum of 10^4.

The spectra, mostly recorded from independent duplicate dilutions of each sample, were collected over 22 months. Alongside these spectra, during each operational day, were run duplicate dilutions of a standard oil chosen to show all of the common spectral features. The variation with time in instrumental performance detected in the latter spectra resulted in a trend of not greater than 3% for any of the features. A correction to the sample data was made for this, but the effect is negligible relative to the between-sample variation.

Data

The file of data, referred to as the test file or test data, is based on the sets of 22 features (Table 1) from 118 samples. These are 24 motor vehicle oils of known manufacture and 44 of unknown manufacture; 17 samples of various types of grease; 15 typical base oils; 2 paraffin waxes; 8 bitumens; 3 cutting oils; and 5 miscellaneous oils. Excepting the base oils, which were included as samples characteristic of oil stocks in general use, most of the samples were submitted to the laboratory as control samples in connection with

case work. Overall, the collection is taken to represent the distribution of sample type to which the technique under consideration may be applied.

Computation

A Trivector Systems Ltd. Vector-80, which is a microcomputer based on an Intel 8080 microprocessor, was used. The system has 32 K byte of memory and dual floppy disk storage for programs and data. Input and output is through an ASR teletype.

The data files were stored on disk, with one file used for each category of material. A BASIC interpreter with software floating-point arithmetic is supplied with the system and used for the data input and the main statistical calculation programs. To minimize computation times the inner loops of the programs were written in 8080 assembler language using the floating-point arithmetic package directly. Run times of about 90 min were obtained for the more complex calculations.

Results and Discussion

Spectral Characteristics

Some examples of different types of materials and circumstances taken from case work are shown in Figs. 1 to 3. These spectra are excitation-corrected. The emission-correction and inner-filter effects are without significant influence on the spectral profiles [*1*] although they are corrected for prior to data processing.

Figure 1 shows the spectra of some oily debris (A) present on a stolen electric battery, (B) found on the clothing of the suspected thief, and (C) from a drip tray from which the suspect said the material on his clothing had come. The spectra clearly discount the latter explanation but not the proposition that A and B are of the same origin.

Figure 2 is from a case of buggery. The spectra (B) demonstrate the presence of a mineral oil lubricant in an anal swab from the child involved and agree (after allowance for the blank, C, resulting mainly from the solvent Raman band) with those from a sample (A) of Vaseline® found in the possession of the defendant. Although these spectra permit little discrimination between different samples of this product, a decisive opinion can be given that a lubricant was present. (In view of the small amount of the lubricant on the swab—about 1 μg—relative to other materials, it is not apparent that the result could have been obtained in any other way with this degree of certainty, in the time available.)

The spectra in Fig. 3 are from a case of murder. Spectrum A is of a control sample from a vehicle allegedly driven at the deceased. The other spectrum (B), of oil on the deceased, is in close agreement with the control sample.

The evidential significance of these results obviously depends on their probability given the circumstances to which they refer and has hitherto rested on visual comparisons between relevant spectra. As these figures show, the same fluorescence features are present in most samples. Sample differentiation depends on relative fluorescence intensities. In Table 1 are listed and defined the fluorescence intensity features used in the following discussions as well as some UV absorbance features. Some fluorescence wavelengths are also used as features. Many of the features are identified by the underlined numbers in Fig. 3.

Within- and Between-Sample Variation

The means and the between- and within-day standard deviations from an analysis of variance of 32 sets of duplicated results from the standard oil used are given in Table 2. The between-day expectedly exceed the within-day standard deviations at levels of signi-

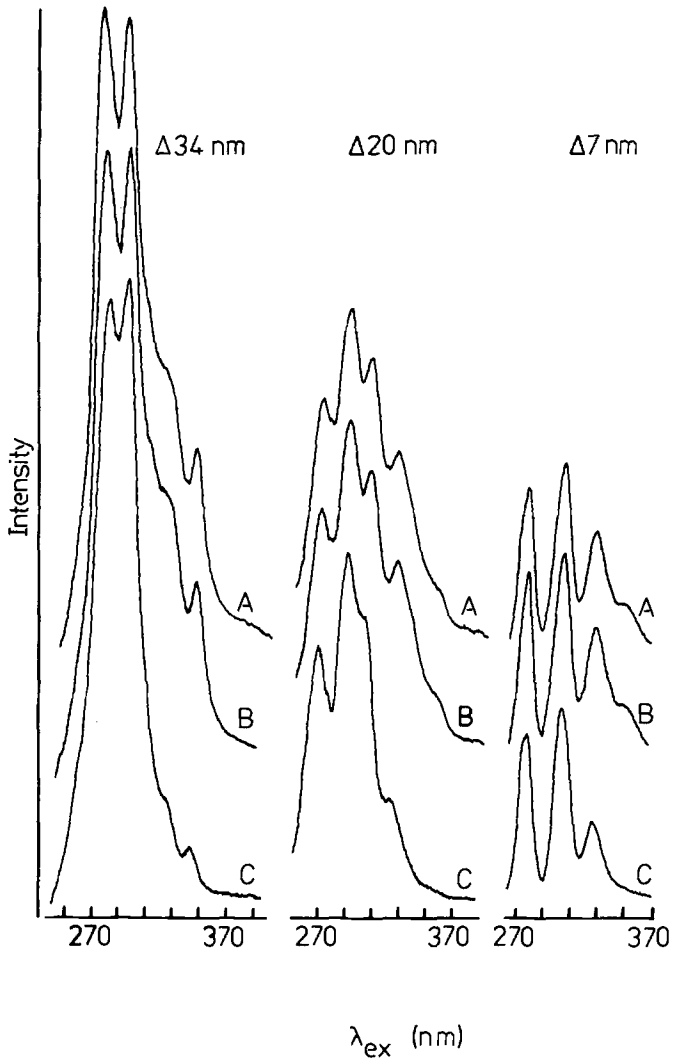

FIG. 1—*Synchronous fluorescence spectra at scanning intervals of 34, 20, and 7 nm of oily debris from a stolen electric battery* (A), *the clothing of a suspect* (B), *and a drip tray* (C). *All the spectra were run at the same sensitivity.*

ficance varying from feature to feature. Expressed as coefficients of variation, the respective averages are 5.3 and 3.0%.

In Table 3 are given the means and the between- and within-sample standard deviations of the data in the test file. The average coefficient of within-sample variation is 6.4%. Because the duplicate analyses from which these within-sample deviations have been calculated were made over the relatively short period of time for which most of the samples were available, these figures relate most closely to the within-day variation of Table 2. Their increase is attributable to variation in the state of the samples examined and in the accuracy with which the features could be measured in different types of spectral pattern as well as to large contributions from some samples in which differences between duplicates were small relative to the high values of the features concerned but large relative to other samples.

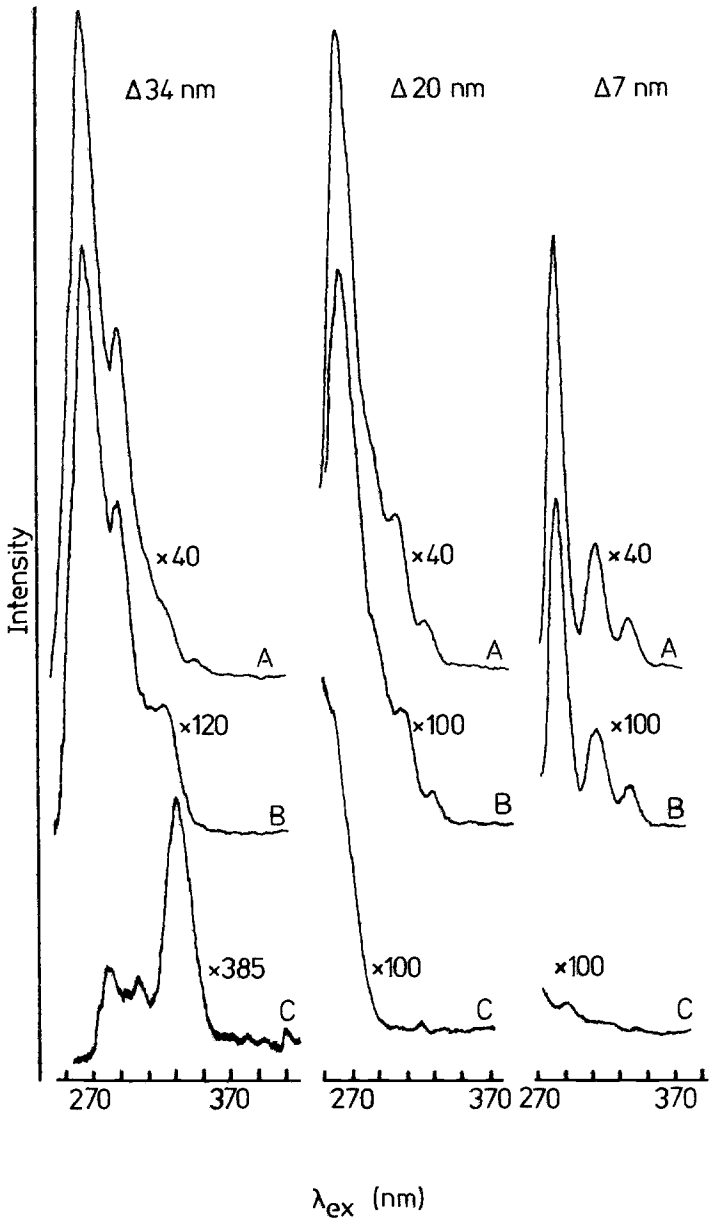

FIG. 2—*Synchronous fluorescence spectra at scanning intervals of 34, 20, and 7 nm of a control sample of Vaseline* (A) *and extracts from an anal swab* (B) *and from an unused swab* (C). *Relative instrumental sensitivities are marked on the spectra.*

From the data given in Tables 2 and 3 it is apparent that the features vary widely in their ability to differentiate samples. In the case of the wavelength measurements (Features *7, 8, 15*) the between-sample variations are negligibly greater than the within-sample variations. But in other cases, mostly involving fluorescence intensities, highly significant between-sample variations are apparent despite the increased within-sample variation, relative to the standard data.

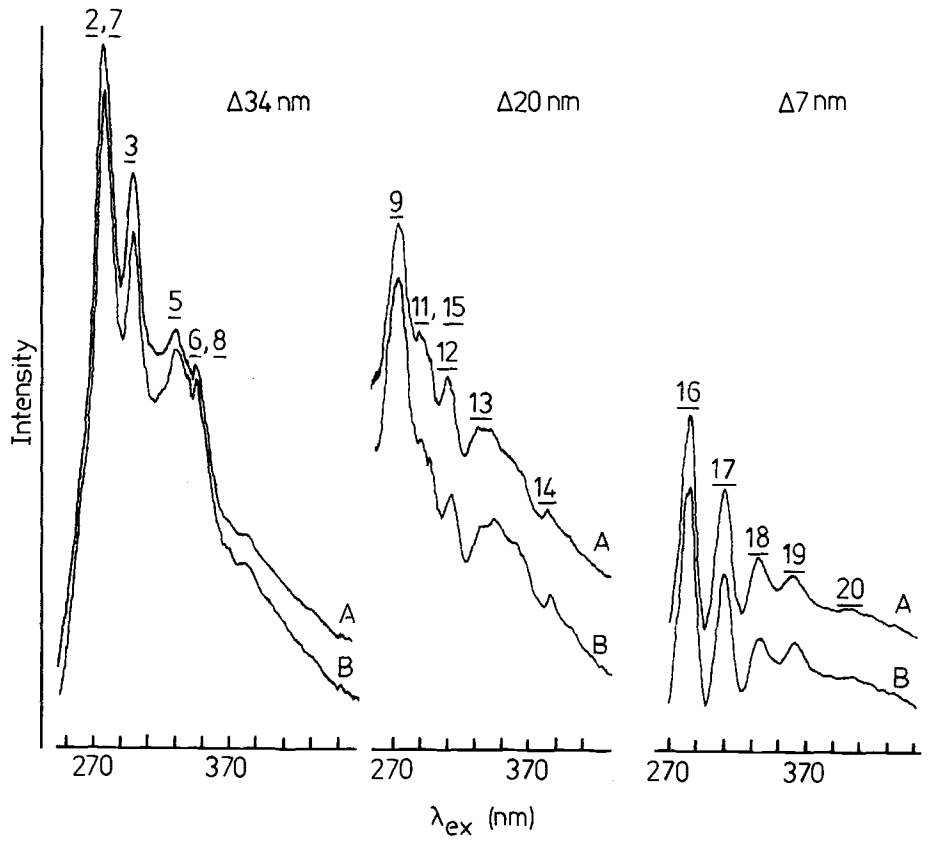

FIG. 3—*Synchronous fluorescence spectra at scanning intervals of 34, 20, and 7 nm of a control sample of oil* (A) *and material found on a victim of murder* (B). *All the spectra are run at the same sensitivity. The underlined numbers indicate some of the features identified in Table 1.*

Correlation

A between-feature correlation matrix from the test file is shown in Table 4. Only those features generating one or more coefficients with absolute values equal to or greater than 0.7 are included. Some features are predictably correlated. Thus, *19* and *20* are spectrally adjacent and undoubtedly contribute intensity to one another. Similarly, *9* and *16* represent the same spectral region of the spectra run at different scanning intervals. Such features may also be chemically correlated, because the range over which a compound emits corresponds approximately to the scanning interval [*7*]. The relatively low correlation between Features *1*, *21*, and *22* (the coefficients, not included in Table 4, are for *1* and *21*, −0.40; for *1* and *22*, 0.48; and for *21* and *22*, 0.11), which represent, respectively, the absorbance of the sample in the 280- to 310-nm region, the fluorescence intensity emitted in the 310-nm region relative to the absorbance, and the latter fluorescence relative to the sample mass, supports the previous conclusion [*1*] that only a small part of the UV-absorbing components of petroleums are fluorescent.

The D *Statistic*

The efficiency of the various features and their combinations in differentiating samples is considered in terms of the extent to which they differentiate each sample in the test

TABLE 2—*Variation of fluorescence features of the standard oil.*

Feature[a]	Mean[b]	Within-Day Standard Deviation	Between-Day Standard Deviation	*F* Ratio[c]
1	969.1	29.5	119.3	16.4
2	3449.8	40.7	71.3	3.1
3	803.3	36.5	49.2	1.8
4	−285.7	34.7	34.9	1.0
5	1127.5	20.0	45.9	5.3
6	106.2	7.6	15.2	4.0
7	276.6	0.6	0.6	1.0
8	348.3	0.5	0.6	1.4
9	1640.8	30.7	69.8	5.2
10	133.7	14.4	32.7	5.2
11	249.4	29.5	85.0	8.3
12	216.2	25.3	86.6	11.7
13	125.7	9.1	11.4	1.6
14	7.9	4.5	8.3	3.4
15	290.4	0.6	1.0	2.8
16	1138.0	20.1	46.8	5.4
17	769.2	13.3	25.0	3.5
18	257.4	7.2	13.3	3.4
19	190.4	13.8	20.1	2.1
20	70.3	10.4	70.3	45.7
21	2090.6	72.4	259.7	12.9
22	2026.0	81.8	189.1	5.3

[a]As defined in Table 1 and in Ref *1*.
[b]From 32 pairs of duplicate runs.
[c]$P = 0.05$, $F = 1.82$; $P = 0.01$, $F = 2.34$, where F is the variance ratio and P the corresponding probability level.

file from all of the others. Given two n-dimensional data points, x and y, then a simple measure of the distance between them is the n-dimensional Euclidean distance:

$$\left[\sum_{i=1}^{n} (x_i - y_i)^2\right]^{1/2}$$

This is an unsatisfactory measure, however, when the magnitudes of the components widely differ: the contributions from components of large magnitude completely mask those of smaller components. The effect may be countered by "autoscaling" [*8*], in which the overall mean and standard deviation are calculated for each of the n components and used to weight each of the components of the individual data points. Although there is some justification for the procedure when the underlying population distribution can be assumed to be multivariate Gaussian, if, as here, that assumption cannot be made, there is no reason to suppose that the procedure is in any sense optimal. Indeed, the effect on the configuration of the clusters of the data points in n space may be deleterious [*9*].

For these reasons we define a statistic D such that:

$$D = \left[\sum_{i=1}^{n} (x_i - y_i)^2/(x_i + y_i)\right]^{1/2}$$

Statistically, for a population of known distribution the probability distribution of D is inevitably complex. In the present application, where the population distribution is

TABLE 3—*Variation of fluorescence features of samples in the test file.*

Feature	Mean[a]	Within-Sample Standard Deviation	Between-Sample Standard Deviation	*F* Ratio
1	2083.2	256.3[b]	5127.2[c]	400.2
2	3669.7	81.9	1009.0	151.8
3	590.7	163.5	516.5	10.0
4	−267.9	43.2	307.3	50.6
5	938.6	49.0	841.5	294.9
6	78.1	11.3	146.4	167.9
7	274.7	0.7	2.3	10.8
8	345.0	0.8	3.2	16.0
9	1838.4	73.6	885.6	144.8
10	224.2	43.5	551.5	160.5
11	197.9	41.9	484.0	133.4
12	110.7	24.7	144.9	34.4
13	99.6	13.2	164.7	155.7
14	20.9	9.9	89.3	81.4
15	287.3	0.6	3.4	32.1
16	1251.9	53.1	625.8	138.9
17	755.3	23.9	308.8	166.9
18	239.5	14.1	129.0	83.7
19	165.6	13.0	262.2	406.8
20	86.7	12.8	240.7	353.6
21	2499.1	152.0	2095.7	190.1
22	3047.4	322.4[b]	3947.8[c]	149.9

[a] 118 samples.
[b] 69 degrees of freedom; 81 for other entries in this column.
[c] 90 degrees of feedom; 117 for others in this column.

unknown, the justification of the use of the statistic is empirical: it performs effectively in practice. The statistic is computationally simple to use, which is an important advantage when the available computer hardware is of limited capacity.

Feature Selection

The difficulties involved in the reduction of the dimensionality and in the selection of the most efficient set of features for the discrimination of samples from one another are discussed, for instance, by Pichler and Perone [*10*]. Ideally, to reduce the dimensionality of the problem from, for example, 22 to 10, the performance of every combination of 10 features selected from the 22 should be determined; this was not computationally feasible within the limits of this study. Hence, an empirical approach was adopted based on the previously considered within- and between-sample variation, correlation of the features, and on individual *D* values.

In Table 5 are shown the threshold values of *D*, for the features considered individually, below which fall 99% (the value is arbitrarily selected) of all possible "pairwise" comparisons in the standard file. Initially the threshold values were obtained by estimation from the numbers of pairs of samples found whose *D* values were in excess of conjectural values. In every case the result was very close to the mean plus three standard deviations of all of the values from the pairwise comparisons. Accordingly, the latter thresholds, which were the more readily obtained, are shown in Table 5 and used in following discussions.

An appreciable variation occurs between the thresholds of the individual features (Table 5), which is evidently related to the type of measurement concerned. Wavelength

TABLE 4—*Correlation matrix of features with coefficients greater than or equal to ±0.70 from data in the test file.*

Feature	*1*	*4*	*5*	*7*	*9*	*10*	*11*	*13*	*15*	*16*	*18*	*19*	*20*
20	0.72	0.60	0.65	0.35	−0.56	−0.29	0.19	0.62	0.32	−0.57	0.59	0.92	1
19	0.72	0.77	0.85	0.51	−0.75	−0.44	0.33	0.83	0.46	−0.76	0.79	1	...
18	0.52	0.71	0.90	0.49	−0.85	−0.56	0.47	0.85	0.44	−0.85	1	...	...
16	−0.52	−0.63	−0.86	−0.60	0.99	0.82	−0.76	−0.82	−0.64	1	...	...	...
15	0.31	0.27	0.47	0.71	−0.64	−0.58	0.69	0.40	1	...	...	...	...
13	0.56	0.91	0.88	0.46	−0.81	−0.55	0.41	1	...	...	...	...	...
11	0.28	0.19	0.44	0.51	−0.77	0.70	1	...	...	...	...	...	...
10	−0.31	−0.33	−0.56	−0.48	0.84	1	...	...	...	...	...	...	...
9	−0.51	−0.61	−0.86	−0.59	1	...	...	...	...	...	...	...	...
7	0.30	0.28	0.54	1	...	...	...	...	...	...	...	...	...
5	0.51	0.74	1	...	...	...	...	...	...	...	...	...	...
4	0.51	1	...	...	...	...	...	...	...	...	...	...	...
1	1	...	...	...	...	...	...	...	...	...	...	...	...

TABLE 5—*Individual threshold* D *values of features from data in the standard file.*

Feature	*D*, 99%[a]	Feature	*D*, 99%[a]
1	9.42	*12*	3.14
2	2.83	*13*	3.39
3	4.59	*14*	5.49
4	6.25	*15*	0.11
5	2.52	*16*	2.68
6	4.87	*17*	2.20
7	0.08	*18*	2.37
8	0.06	*19*	2.45
9	3.06	*20*	3.32
10	5.63	*21*	13.06
11	8.97	*22*	7.65

[a]The values quoted are the mean plus three standard deviations of each feature's *D* values from 1275 comparisons within 51 samples.

features (*7, 8,* and *15*) give particularly low values, whereas UV absorbance features (*1* and *21*) tend to give high values. As mentioned earlier, small components in the sort of summation involved here, when features are combined into sets, are susceptible to masking by larger ones. Hence, the wavelength features were excluded from further consideration. Although some form of scaling might permit their inclusion, it is apparent from Table 3 that these features are unlikely to contribute much discrimination.

The 15 features selected for initial examination are shown in the first row of Table 6. From the standard file, the 99% threshold for this combination is 11.68. In the test file the number of pairs of samples for which the value of *D* is less than this, that is, which are not differentiated under this criterion, was found to be 35 out of a possible 6903 pairs. (Because the test data are subject to a greater within-sample variation than the standard data, an actual 99% threshold for the test data would be greater than 11.68. The standard threshold is simply an arbitrary point of reference and comparison that can be readily computed for the various combinations of features used.)

In the subsequent rows of Table 6 is shown the effect of removing features from the initial set. For each number of features is shown only the set that gave the minimum number of undifferentiated pairs among all of the sets examined. The sets shown were finally arrived at by trial and error, conditioned by the results already given (Tables 3 to 5). The pertinent threshold *D* values, arbitrarily taken as 99%, were calculated from the standard file. From Table 6, if the efficiency of each set is taken to be reciprocally proportional both to the number of features present and to the number of undifferentiated pairs found, and if computing time may be discounted, then there is little to choose between the sets containing nine or more features. If computing time or memory is limited, clearly the smaller sets are the more useful.

Retrieval of Rerun Samples

To test the performance of the system and the stability of the spectra, the spectra of 16 samples were rerun 12 to 18 months later than the original spectra. The samples were seven common unused motor oils, and the others a selection of base oils. These samples were used purely because of their availability in a well-kept state (the smaller samples had been refrigerated, the larger ones stored in filled containers at room temperature in darkness). The test file, searched for patterns corresponding to the new spectra, gave the results shown in Table 7.

From Table 7, the number of correctly chosen samples, with the 99% threshold,

TABLE 6—*Combinations of features giving the maximum discrimination between sample pairs in the test data.*

Number of Features	Identity of Features															Comparisons < Threshold[a]
15	*2*	*3*	*5*	*6*	*9*	*10*	*11*	*12*	*13*	*14*	*16*	*17*	*18*	*19*	*20*	35
14	*2*	*3*	*5*	*6*	*9*	*10*	*11*	*12*	*13*	...	*16*	*17*	*18*	*19*	*20*	36
13	*2*	*3*	*5*	*6*	*9*	*10*	...	*12*	*13*	...	*16*	*17*	*18*	*19*	*20*	41
12	*2*	*3*	*5*	*6*	*9*	*10*	...	*12*	*13*	...	*16*	*17*	*18*	*19*	...	59
11	*2*	*3*	*5*	...	*9*	*10*	...	*12*	*13*	...	*16*	*17*	*18*	*19*	...	59
10	*2*	...	*5*	...	*9*	*10*	...	*12*	*13*	...	*16*	*17*	*18*	*19*	...	58
9	*2*	...	*5*	...	*9*	...	...	*12*	*13*	...	*16*	*17*	*18*	*19*	...	59
8	*2*	*3*	*5*	...	...	...	...	*12*	...	...	*16*	*17*	*18*	*19*	...	79
7	*2*	*3*	*5*	...	...	...	...	...	...	...	*16*	*17*	*18*	*19*	...	101
6	*2*	...	*5*	...	*9*	...	...	...	...	...	*16*	*17*	...	*19*	...	115
5	*2*	...	*5*	...	*9*	...	...	...	...	...	*16*	*17*	...	...	...	

[a]From 6903 comparisons within 118 samples; a 99% threshold was used (mean plus three standard deviations) from the standard file.

TABLE 7—*Retrieval of rerun samples from the test file.*

Number of Features[a]	Threshold[b] %	Number of Correct Choices	Total Number of Choices	Incorrect ÷ Correct Choices
15	99	11	30	1.7
14	99	12	39	2.3
13	99	8	23	1.9
12	99	11	28	1.5
11	99	11	29	1.6
10	99	10	28	1.8
9	99	13	26	1.0
9	99.3	16	33	1.1
9	99.3	16	40	1.5

[a]The combination of features used are given in Table 6.
[b]The thresholds are the means plus 3 (99), 3.33 (99.3), and 4 (99.9) standard deviations of the *D* values within the standard file.

tends to vary only slightly as the number of features is reduced from 15 to 9. Within this range the largest number of correct choices is obtained with the smallest set. As the last column of Table 7 shows, each one of these particular choices is accompanied by one incorrect choice, on average, out of all of the 6903 possible choices within the file. When the thresholds are increased to the 99.3 and 99.9% values (as defined, for the standard file, in the footnote to the table), with the nine-feature set, all of the required samples are found, with only slight increases in the average number of spurious choices. Although it is not intended to assess here evidential significance pertinent to any particular event, we note that a valuable level of specificity is implicit in these and the foregoing results. Thus, the test file included 68 motor vehicle oils. Each of the seven rerun samples of these enabled the original data to be identified with each of the thresholds used, with an average of 1.4 spurious choices (99.3% threshold) within this group of samples and 1.6 within the whole file.

Comparison with Earlier Results

Finally, it has been examined whether the system gives, with case work samples, results that are comparable to the results of the visual comparisons that the system is intended to replace. The test file was searched, with the nine-feature set (99.9% threshold), for patterns corresponding to the spectra of 15 samples of contact traces collected at scenes of crime, from stolen property, and so forth. Data from the control samples in these cases constituted part of the file. The results are as follows:

1. In 11 cases the control and the contact trace were considered, from the original visual comparisons of their spectra, to be matched. The computed result is the same.
2. In each of two of the last-mentioned cases an additional control sample was visually specifically excluded. The computed result is the same.
3. In three cases no definite conclusion was drawn from the visual comparisons. In the computed result no match is found.
4. In one case the control and suspect samples were considered to be comparable visually, although of low evidential significance, but are not matched in the computed result. The material concerned was of an exceptionally variable nature. A computed result in this circumstance would be obtained only with an increased threshold *D* value, which would result in lower differentiation among other samples and in low evidential significance.

Concluding Remarks

It is emphasized again that our object is to develop and assess a standardized technique in order that in future fluorescence spectroscopy may be efficiently exploited in the collection of evidential data relevant to specific circumstances. The foregoing shows that very satisfactory levels of differentiation and retrieval can indeed be obtained within a representative sample set, which includes a large group of common related materials, by the use of selected fluorescence features in spectra collected over an appreciable period of time. Given the development of direct data acquisition techniques [*11*] it will now be possible to assemble relevant data rapidly and efficiently and, given the development of suitably scaled pattern recognition statistics, to incorporate the results of other techniques as mentioned before [*1*].

Summary

The previously discussed [*1*] fluorescence characteristics of high molecular mass petroleum products have been used in a study of an arbitrary set of samples representative of those encountered in case work. The results show that within this sample set, with the techniques applied, common spectral patterns can be efficiently recognized and retrieved with an accompanying low level of spurious data.

Acknowledgment

We are indebted to Dr. P. W. Hall for the spectra in Fig. 2.

References

[*1*] Lloyd, J. B. F., "Examination of Petroleum Products of High Relative Molecular Mass for Forensic Purposes by Synchronous Fluorescence Spectroscopy. I: Appraisal of Experimental Factors," *Analyst,* Vol. 105, No. 2, Feb. 1980, pp. 97-109.

[*2*] Lloyd, J. B. F., "Synchronized Excitation of Fluorescence Emission Spectra," *Nature Physical Science,* Vol. 231, No. 20, May 1971, pp. 64-65.

[*3*] Fick, N. H., "Lubricants, Oils and Greases," *Analytical Chemistry,* Vol. 49, No. 5, April 1977, pp. 238R-240R.

[*4*] Lloyd, J. B. F., "The Nature and Evidential Value of the Luminescence of Automobile Engine Oils and Related Materials. III: Separated Luminescence," *Journal of the Forensic Science Society,* Vol. 11, No. 4, Oct. 1971, pp. 235-253.

[*5*] Wakeham, S. G., "Synchronous Fluorescence Spectroscopy and Its Application to Indigenous and Petroleum-Derived Hydrocarbons in Lacustrine Sediments," *Environmental Science and Technology,* Vol. 11, No. 3, March 1977, pp. 272-276.

[*6*] Wheals, B. B., "High Performance Size Exclusion Chromatography on Microparticulate Silicas and Its Application to Forensic Analysis," *Journal of Liquid Chromatography,* Vol. 2, No. 1, Jan. 1979, pp. 91-110.

[*7*] Lloyd, J. B. F. and Evett, I. W., "Prediction of Peak Wavelengths and Intensities in Synchronously Excited Fluorescence Emission Spectra," *Analytical Chemistry,* Vol. 49, No. 12, Oct. 1977, pp. 1710-1715.

[*8*] Kowalski, B. R. and Bender, C. F., "Pattern Recognition. A Powerful Approach to Interpreting Chemical Data," *Journal of the American Chemical Society,* Vol. 94, 1972, pp. 5632-5639.

[*9*] Duda, R. O. and Hart, P. E., *Pattern Classification and Scene Analysis,* Wiley-Interscience, Inc., New York, 1973, p. 215.

[*10*] Pichler, M. A. and Perone, S. P., "Computerized Pattern Recognition Applications to Chemical Analysis," *Analytical Chemistry,* Vol. 46, No. 12, Oct. 1974, pp. 1790-1798.

[*11*] Dubery, J. M., "A Design for a Data System for Spectrofluorimetry," M.Sc. thesis, University of Aston in Birmingham, United Kingdom, 1979.

Address requests for reprints or additional information to
J. B. F. Lloyd, Ph.D.
Home Office Forensic Science Laboratory
Priory House, Gooch St. North
Birmingham, B5 6QQ, United Kingdom