Laser Remote Sensing of Pollution of the Sea: 
a Quantitative Approach

T. Hengstermann and R. Reuter
University of Oldenburg - Physics Department
2900 Oldenburg, Germany

ABSTRACT

Identification of oil types from laser fluorosensor measurements necessitates a systematic investigation of the fluorescence behaviour of a representative number of different oils. Based on laboratory measurements on 52 oil samples, a catalogue of optical signatures was assembled. From these data the fluorescence signatures of characteristic oil types are evaluated.

The analysis of these spectra has shown that approximately 99% of the spectral information can be represented by only four abstract eigenfunctions. Oil classification based on these four eigenfunctions leads to a considerable reduction of spectral data involved in the identification process. Moreover, they allow an optimization of the design of laser fluorosensors with respect to a minimum number of spectral detection channels, necessary for gathering significant information.

From computer simulations a method has been derived which allows for the identification of the fluorescence signature of the pure oil also in cases of optically thin films, where the measured spectra are obscured by contributions from the water column.

As a result, the airborne laser fluorosensing technique has been found to be capable of filling a gap in the range of airborne instrumentation available for marine oil spill monitoring. Particularly its potential of a substance specific classification, and its sensitivity in a thickness range of few micron, makes it a very compromising technique within surveillance programmes according to international regulations as, for example, the Marpol Agreement (Hengstermann and Reuter, 1990).

To allow full usage of these capabilities, laboratory analyses and theoretical investigations have been carried out. The result of these investigations will be implemented into the laser fluorosensor which is currently under development (Hengstermann et al., 1991). The presentation of these results is the aim of this paper.

1. FLUORESCENCE PROPERTIES OF MINERAL OIL

The spectral variability of fluorescence signatures of oils is substantially based on the statistical distribution of a complex mixture of numerous organic compounds. The distribution of these compounds is widely determined by the genesis, depth and geological position of the deposit. However the common nomenclature of oils, which is mainly based on empirical, macroscopic properties as, for example, the gravity, the viscosity or the spreading behaviour, does not consider any fluorometric characteristics.

Hence, it is not possible to predict a certain spectral signature or feature from the nomenclature. It is therefore necessary to investigate the optical properties of a representative number of different oil types in the laboratory, and to get an overview of the spectral variability of the fluorescence signatures.

In practice, it is impossible to investigate the optical properties of all different oils transported at sea. This has
to be done with a number of representative oils, whereby the term 'representative' does not necessarily has a global meaning. In most cases local or operational aspects should also be taken into consideration.

Due to this fact, it is not the primary goal of the classification to identify the name of the spilled oil. Instead, an accurate estimation of the spilled volume and an assessment of its ecological risk shall be obtained from fluorescence measurements.

Based on these investigations a division of the oils into different classes with a common spectral signature has to be performed, whereby other physical and chemical properties are of secondary relevance.

The results reported in the following are based on fluorescence spectra of 52 different crude oils and refined products. The spectra were taken with a commercially available Perkin Elmer 650-40 fluorometer equipped with an front-surface-assembly. They are corrected for the spectral characteristic of the excitation source and the spectral sensitivity of the detection system. A fluorescence standard of quinine sulfate has been used to eliminate long term instabilities of the detector system.

Fig. 1 shows as an example the fluorescence spectra of a crude oil, a diesel, and a vegetable oil. To allow for an easy comparison of the shape of these spectra, they have been normalized to unit fluorescence efficiency.

Together with the absorption coefficient, which is an essential parameter in the estimation of the film thickness and which must therefore be derived from the identification, a catalogue of optical properties has been assembled. This catalogue forms the basis for the definition of a classification scheme and the development or improvement of classification algorithms. It also provides information on the optimization of the technical layout of the detection system.

2. EXTRACTION OF DISCRIMINATING FEATURES

The first and most important step in the definition of a classification scheme is the extraction of features which efficiently allow to distinguish between different objects or classes of objects. The selection of these features strongly affects the design of the classifier, and in most cases has direct consequences on the design of the detector unit.

In a preprocessing step all spectra have been normalized to unit total fluorescence intensity observed in the entire spectral range from 311 to 700nm. This has the advantage that the detector system must not be calibrated in absolute units. But the classification of course does only consider the shape of the spectrum and makes no use of the absolute fluorescence intensity, which, in some critical cases can be an additional discriminating feature.

A very common way to extract such features is the definition of some simple logical or arithmetical relations between two or more detection wavelengths as, for example, differences or/and ratios.

It was shown that already by defining a contrast function between only three fluorescence wavelengths enables to distinguish three different kinds of oil classes, heavy oils, crude oils and light products. (Anderson et al., 1987).

But with those methods most parts of the information stored in the spectra is lost. Changes in the spectral layout of the detector system necessitate a revision of the whole catalogue of spectra. On the other side, supplementing new oils to the catalogue may result in a new layout of the detection system. Finally, the efficiency of these features depends on the intuitive skillness of the operator.

A more systematic approach in defining those features without loosing important information is to perform a principal component analysis. This method was developed in the field of social sciences. A very thorough discussion
of this method can be found by Hotelling (1933). It has also widely been extended to biological, chemical and physical sciences and has become a standard tool in pattern recognition techniques. Especially in cases in which large data sets have to be handled this method has found to be extremely useful in data reduction (Anderson, 1984; Fukunaga, 1972; Hangac, 1982; Watanabe, 1980).

Given n objects, on which measurements of m variables have been performed, each object being represented by a pattern vector in a m-dimensional feature space,

\[ \mathbf{v} = (v_1, v_2, v_3, \ldots, v_m) \]

and where the \( v_i \) represent the value measured on each of the m variables.

Since these variables generally show more or less correlation they do not form an orthogonal basis of the feature space. This means that there is a certain degree of redundancy in the information of each variable which leads to an unintentional increase of the weight of those variables and can result in a coverage of important information.

The approach of principal component analysis now is to find a new set of variables, which are linear combinations of the original ones, and which will form an orthogonal lower dimensional basis of the feature space. In a geometrical interpretation this approach leads to a linear projection of a high-dimensional manifold onto a lower dimensional hyper plane under the constrain that the distance between the feature vector and its picture in the hyper plane is minimal, which also means that the loss of information is minimized (Bock, 1974).

It can be shown that the eigenvectors of the empirical covariance matrix calculated among all feature vectors optimally fulfil these requirements. The set of eigenvectors are equivalent to the principal axes of the variance hyper ellipsoid. Hence distinguishing objects of a certain data set also means looking at the data in terms of variances.

Since these new variables are orthogonal, the covariance between any two of them for each sample will be zero. The total variance of the data set can therefore be expressed as the sum of the variance of these new variables. The intrinsic dimensionality of the data set can be obtained by discarding those variables (eigenvectors) which show the lowest variances. These variables can therefore be considered to contain less information e.g. mostly noise. In most cases this leads to a considerable reduction of data.

Following the idea of the principal component analysis, each spectrum of the catalogue can be expressed by a certain number of statistically independent variables,

\[ \mathbf{I}_i = \sum_{j=1}^{n} c_{ij} \mathbf{E}_{ij} \]

where \( \mathbf{E}_{ij} \) denotes the i-th eigenvector of the covariance matrix in the i-th detection channel and \( c_{ij} \) its weight in approximating the entire spectrum. Due to its spectral character, \( \mathbf{E}_j \) will be denoted as an eigenspectrum in the following.

In case of linear mixing of different fluorescent substances these eigenspectra can be related to the fluorescence spectra of these compounds; the weighting factors \( c_{ij} \) represent their concentration. But due to the very high absorption coefficient of oils being in the order of 1/\( \mu \text{m} \) this assumption does not hold. Therefore the calculated eigenspectra do not necessarily have any physical or chemical meaning. Hence in this state the method can only be regarded as a mathematical tool in handling the data.

The eigenanalysis of the covariance matrix of 51 spectra of the catalogue has shown that only 4 eigenspectra contain already 99% of the whole information of the catalogue. The total variance was distributed with 74% on the first 17% on the second, 4% on the third and 4% on the fourth eigenspectrum, respectively. This means that from a fluorometric point of view the investigated oils can be represented by a 4-tupel of loadings on the eigen- spectra which is individual of each oil.

As an example of the quality of this method, Fig. 2a shows the original spectrum of an Arabian crude oil (full line) and the spectrum calculated on the basis of this four eigenspectra (dashed line) which are also shown (dotted lines). Obviously, almost the whole information of the entire spectral range can be regained. The same holds for a diesel oil which is shown in Fig. 2b. Only in those regions where high frequency spectral features can be found there are higher discrepancies between the curves. In case of the fuel Jet-A1 shown in Fig.2c the information can completely be regained.

Since the spectra have been measured in the range from 310 to 700 nm with a resolution of 1 nm, the application of the principal component analysis superficially leads to a considerable reduction of the data set by a factor of 100.

Limitation on the first two eigenspectra, which means a retention of 91% of the total variance, allows a graphical representation of the feature space in a plane. In Fig. 3 the loadings on the eigenspectra E1 versus E2 are plotted for all 52 samples of oil. Groups of oils can be identified in this plot by regions with higher density of points.
Two general tendencies are found in Fig. 3:
- All crude oils including the heavy residuals are located in a homogeneous area along line G1. Light crude oils like the Nigerian Light or the Brent crude are located in the lower part of this line while the heavy crude and residual oils reside in its upper part. The course of line G1 can therefore be regarded as a measure of the density of the oil.
- The refined products form a considerable heterogeneous cluster along the line G2. Within this cluster additional subclusters can be found which are the class of the lubricating and diesel oils in the lower part of G2; the petrels are in the middle, and the very light products in the upper part of that line. This tendency of the line can be taken as an information on the degree of fractionation of these oils.

In addition the plot clearly indicates that the discrimination of the vegetable oils from the refined products on the basis of only two eigenspectra might be difficult to achieve. A proper discrimination of these oils can only be
done by using all four eigenspectra or by taking into account the total fluorescence intensity as a fifth feature as well.

In contrast to the refined products displayed in Fig. 3 it is not possible to define some subclusters within the cluster of the crude oils. But a detailed k-means cluster analysis of all 52 oil samples has shown that on the basis of all four eigenspectra it is possible to define eight different oil classes, which can be distinguished from each other. The typical spectra of these oil classes are shown in Fig. 4a and b.

In addition to these findings analysis of the catalogue of fluorescence signatures allows to find an optimal choice of the number and spectral position of detection wavelength. Since the characteristic features $c_i$ according to equation (2) are independent of the detection wavelength, four detection channels should be in principle sufficient for the evaluation of these four characteristic features. However, due to signal fluctuations or measurement errors sufficient results could only be obtained by using 12 detection channels.

An information on the optimal distribution of these detection channels within the entire spectrum can be obtained from the variance spectrum calculated among all fluorescence signatures as shown in Fig. 5. This spectrum shows high values in the UV region near the excitation wavelength, followed by a plateau in the range from 370 to 430 nm approximately. The variance shows a minimum at 500 nm and slightly higher values in the red. This curve can be taken as a measure of the density of detection wavelength selected within a certain spectral range. An optimized detection system will therefore require a high number of detection channels in the UV region, but only a few channels in the green and a little more in the red.

As an example, Fig. 6 shows the original spectrum of a Diesel oil and an Arabian Heavy crude oil (full lines). The dots indicate the spectral position of the detection channels used with the newly developed laserfluorosensor system (Hengstermann et al., 1991). Also the spectra

Fig. 4a,b - Fluorescence spectra of different oil classes obtained from a k-means clustering algorithm. Classification of oil pollutions should be done according to these classes. Similar to the spectra shown in Fig. 1 these spectra are also normalized to unit total fluorescence intensity.

Fig. 5 - Variance spectrum calculated from the catalogue spectra. The variance observed in a certain wavelength intervall can be taken as a measure of the density of detection channels, which have optimally to be positioned in this spectral region.
calculated on the basis of the first four dominant eigenspectra are shown as dashed lines for comparison. For the calculation only the values of the original spectrum at the indicated wavelength have been considered.

![Graph showing fluorescence spectra](image)

**Fig. 6 - Original spectra of the an Arabian Heavy crude oil and an Diesel oil (full lines). The dotted lines show the spectra recalculated on the basis of the four dominant eigenspectra. For the calculation only the intensities at the wavelengths indicated by the dots have been considered. These wavelengths correspond to the spectral position of detection channels of the new laser-fluorosensor (Hengstermann 1991).**

Comparison of Fig. 6 with the data from Fig. 2a, b and c, where the whole information of the entire original spectrum has been considered in the calculation of the individual weight factors shows that nearly the whole spectral information could be regained. Hence, it becomes evident that an optimized sensor system may consist of only twelve detection channels, which have to be positioned in the entire spectrum considering the variance spectrum.

### 3. RECONSTRUCTION OF THE FLUORESCENCE SPECTRUM

The identification of only those oil types requires the knowledge of the fluorescence signature of the pure oil. But in airborne applications only spectra, which are taken over optically thick parts of an oil spill, represent the fluorescence spectrum of the pure oil.

In case of discharges of small volumes of oil this condition is only rarely fulfilled. Very thin oil films are partially transparent to the laser light and the fluorescence signal from the oil will be more or less superimposed by fluorescence contributions from naturally occurring organic matter (gelbstoff) in the water column. To overcome this limitation, an algorithm has to be provided, which allows to isolate the fluorescence from this fluorescence background in case of optically thin films.

The mathematical model of the measuring process is based on the integration of the lidar equation of Browell (1977). Application of this equation on a two layer model of the oil/water column allows to express the fluorescence signal \( P \) obtained from an optically thin oil layer with film thickness \( d \) at a certain point within the oil spill in terms of the fluorescence signal expected from the pure oil \( \sigma_O/a_O \) and the water column \( \sigma_W/k_W \) (Kung and Itzkan, 1976):

\[
P_{O,W} = \frac{A/(mH)^2}{\sigma_O/a_O + (\sigma_W/k_W - \sigma_O/a_O) \exp(-a_O * d)}
\]

with \( A \) being a constant including instrumental parameters, \( H \) the aircraft flight altitude and \( m \) the refractive index of the air/water or air/oil interface, respectively. \( \sigma_O \) and \( \sigma_W \) denote the fluorescence quantum efficiency of the oil and the water, \( a_O \) the sum of the absorption coefficients of the oil at the excitation and detection wavelength, and \( k_W \) the attenuation coefficient of the water.

Assuming that the oil spill can be found in an area where there are only small variations in the concentration of the organic substances in the water column and that the composition of the oil will not change within the spill, equation (3) can be rewritten in terms of the fluorescence properties of the water column and the oil, respectively:

\[
P_{O,W} = P_O + (P_W - P_O) \exp(-a_O \cdot d)
\]

where \( P_W \) is the fluorescence signal of the oil free water column and \( P_O \) the fluorescence signal expected from an optically thick part of oil slick.

An estimation of the oil specific fluorescence \( \sigma_O/a_O \) from the signals is difficult, if no optically thick part within the slick can be found. Then an estimate of the absorption coefficient of the oil cannot be derived from the fluorescence signatures. A possible way to overcome this problem makes use of the redundancy in the different detection channels if film thickness data have to be derived.

Information about the film thickness distribution within the oil slick can preferably be obtained from the water Raman scattering, which in the case of an oil-free water surface provides an intense, well-defined peak at 344 nm when excited at 308 nm. Due to its very narrow bandwidth when compared to the other structures in the spectrum, this signal can easily be isolated from the underlying fluorescence background arising from gelbstoff and mineral oil (Hoge and Swift, 1980).
Normalization of this signal $R_{0,W}$ to the water Raman intensity $R_w$ obtained outside the slick over the oil-free water column yields:

$$\exp \left( - a_0^R d \right) = \frac{R_{0,W}}{R_w}$$  \hspace{1cm} (5)

Hence a solution for the film thickness $d$ requires an estimate of the oil specific absorption coefficient $a_0^R$, which also must be derived from the catalogue of optical signatures.

Inserting the information on the oil film thickness from eq. (4) into eq. (5) yields:

$$P_{0,W} = P_0 + \left( P_W - P_0 \right) \left( \frac{R_{0,W}}{R_w} \right)^{a_0^R/a_0^R}$$  \hspace{1cm} (6)

Successive measurements at different locations of the oil slick result in a system of nonlinear equations with 3 unknown parameters $P_0$, $a_0$, and $a_0^R$. Assuming a sufficient number of measurement points these parameters can be extracted by a nonlinear least-square method.

It is known from experimental measurements, that the absorption coefficient of oils decreases monotonously with increasing wavelength. Since $a_0$ and $a_0^R$ are the sum of the absorption coefficients at the excitation and the detection wavelength respectively calculation of the exponent $(a_0 / a_0^R)$ can be done under the constraint

$$0.5 < (a_0 / a_0^R) < 2.0$$  \hspace{1cm} (7)

Taking the fluorescence background at the Raman scattering wavelength, equation (6) would be reduced to a simple linear expression.

Gathering of information from two different spectral channels simultaneously necessitates an accurate spectral calibration of the detector system at least in relative units.

To prove its usefulness, the method has been tested under several conditions with computer simulated laser fluorsensor signals. Simulation is done under the assumption, that the fluorescence and Raman signals are normally distributed around a mean value with a given deviation, as it can be observed in flight experiments.

The simulations are done for a system with two detection channels in which one channel corresponds to the water Raman scattering signal and the other can be positioned elsewhere in the spectrum. Normally distributed signals in each detection channel have been generated from a random number generator. Since there is no information available on the preferred statistical distribution of the oil film thickness it is assumed, that the film thickness within the oil spill is equally distributed in a given interval.

A typical result of the simulation is shown in Fig. 7a. Both, the Raman and the fluorescence signal vary around a mean value of 20 scale units. The fluctuation is set to 10% of the mean value which agrees with the typical noise amplitude observed in airborne experiments. The thickness of the oil spill is equally distributed in an interval of 0 to 2 $\mu$m. The optical properties of the oil correspond to the light crude oil class.

![Fig. 7a - Computer simulations of laserfluorsensor signals expected from optically thin oil films. The fluorescence of the water column and the Water Raman scattering have been set to values of 20 scale units with a variance of 10%. The thickness of the oil spill is assumed to be equally distributed in the interval from 0 to 2 $\mu$m. Optical properties were taken from the light crude oil class. The fluorescence of the oil has been set to 5 scale units and the absorption coefficient to 1.0 $\mu$m$^{-1}$ at the excitation wavelength and 0.6 $\mu$m$^{-1}$ and 0.4 $\mu$m$^{-1}$ at the Raman and the fluorescence wavelength, respectively. The optical properties of the oil correspond to the light crude oil class.](image)

![Fig. 7b - A scatter plot of the fluorescence signal as a function of the water Raman intensity (dots) and the result of the non-linear least-square analysis according to eq. (6). The analysis is performed with an implementation of the Levenberg-Marquardt method, which has be-](image)
come a standard in nonlinear data modelling. The implementation of this method works straight forward and no special measures have been provided to prevent the algorithm from taking local minima as a best fit of the desired parameters. But as it can be seen from the legend of Fig. 7b already with this simple approach, the initial values of the water and oil fluorescence can be reproduced by the nonlinear least-square analysis with a deviation better than 5%. A closer analysis under different conditions has shown that, up to a variance of 15% in the Raman and in the fluorescence signal, respectively, sufficient results can be obtained.

CONCLUSION

In laboratory investigations a catalogue of optical properties of different mineral oils has been assembled. This catalogue includes the fluorescence signatures of oils as well as their spectral absorption coefficient, which is an essential parameter in the estimation of the film thickness of slicks on the water surface.

An analysis of the fluorescence signatures has shown, that the information contained in the catalogue can be represented without loss of important information by 4 spectral features. A linear combination of these features with four characteristic weight factors allow a quantitative description of the individual fluorescence spectra, and therefore each oil can be represented by a 4-tupel of these weight factors.

As a consequence of this reduction of the data which are representative for each oil, the experimental efforts for gathering the information necessary to identify a certain oil type can be drastically reduced.

The investigations have shown that twelve detection channels are sufficient to calculate the four significant features from a fluorescence spectrum. This reduction of the data
also will enable data interpretation in real time. By interpreting the variance spectrum of the entire catalogue it is possible to define optimal positions for the detection channels in the wavelength range from 320 to 700 nm.

From computer simulations of laser fluorosensor signal returns expected from optically thin oil films, a method has been derived, which allows for the isolation of the fluorescence signature of the pure oil from the fluorescence background of naturally occurring organic matter.

Although the results of those investigations have not yet been verified in field experiments, there is high confidence that they will represent a basic tool for the interpretation of further data which will come up from the new developed laser fluorosensor data designed for maritime surveillance. A first implementation of routines based on these results will be tested at the end of this year.

Especially it has to be emphasized, that no considerations of weathering effects of the oil on the water surface, which will drastically change the optical properties of the of the oil, have been taken so far. The influence of these effects on the fate of the oil on the water surface will be investigated in future to enable a more quantitative and reliable data analysis.

REFERENCES


Watanabe, S., Pattern Recognition, Human and Mechanical (New York: John Wiley & Sons).