SRC TR 88-81

Computer-Enhanced Analysis of Multicomponent Fluorescence Spectra

by

N.S. Wang, P.J.B. Rinaudo, and T.J. McAvoy

COMPUTER-ENHANCED ANALYSIS OF MULTICOMPONENT FLUORESCENCE SPECTRA

by

Nam Sun Wang*, Philippe J. B. Rinaudo, and Thomas J. McAvoy

Department of Chemical and Nuclear Engineering

and Systems Research Center

University of Maryland

College Park, MD 20742-2105, U.S.A.

Presented at

Fourth International Congress

on Computer Applications

in Fermentation Technology:

Modelling and Control of

Biotechnical Processes

University of Cambridge, UK

September 25-29, 1988

ABSTRACT

Laser-induced culture fluorescence is one of the few nondestructive measurements that are currently available for the on-line monitoring of the internal state of a microorganism. Many biomolecules, including NADH, exhibit characteristic fluorescence spectra that are dependent on the excitation wavelength. Many fluorophores are simultaneously present in a typical fermentor, and the fluorescence spectra for a mixture of fluorophores must be deconvoluted to yield individual concentrations, after the noisy spectral data are treated in the Fourier domain. Because of the secondary optical effects such as the repeated re-absorption and re-emission of light in a mixture, the principle of superposition does not hold true. Furthermore, the overall emitted signals are shifted toward the higher wavelengths. A rigorous nonlinear model is developed for the most commonly used optical arrangement to relate the detected signal level to the fluorophore concentration. As a model system, the emission spectra of binary and ternary mixtures of simple aromatic amino acids (tyrosine, tryptophan, and phenylalanine) are analyzed with a range of computer algorithms. The predictions resulting from different analytical approaches are compared to the true component concentrations, and the effectiveness of each method is evaluated. In general, it is most difficult to detect a small amount of one component either in the presence of larger concentrations of other fluorophores or in the presence of an unknown fluorophore. The presence of an unknown fluorophore is determined by spectral matching and hypothesis testing. The application of culture fluorescence measurements to fermentation control will be discussed.

BACKGROUND

Armiger and Humphrey wrote in 1979:"One of the major problems in developing mathematical models and control strategies, lies in the inability to measure on-line many of the important process parameters. Significant improvement in existing sensors and the development of new sensors is needed."[2] No solution has been found yet and the challenge of on-line measurements is now a crucial one for industry. To face this challenge, optical techniques seem to be a powerful tool.

The expression "optical techniques" covers a wide variety of techniques that chemists and biologists have been using for concentration or identification measurements for years. Absorption, scattering and fluorescence are some examples of optical techniques that we employ

Optical techniques have particular advantages which make them well-suited for on-line measurements in fermentation and a wide range of complex chemical processes. These advantages include instantaneous response, selectivity of response, sensitivity, and nondestructiveness. The recent progresses in fiber optics, lasers, and photodetectors have further increased their domain of use as well as their convenience. Among all the optical techniques available, fluorescence is the most sensitive and selective for molecular measurements. Because many biological compounds fluoresce, fluorescence is a powerful method for measuring compositions in a biological system. This is especially important in fermentation systems which contain many fluorescent compounds, including amino acids, cofactors, vitamins, and antibiotics. Fluorescence has been widely used in chemistry as well as in biology. Although fluorescence measurement requires a more sensitive and expensive setup than absorption measurement does, it is now a popular technique due to its sensitivity and selectivity. The power of fluorescence also lies in the wide variety of measurements one can make with a fluorescence setup. Changing the excitation

and detection wavelengths and measuring the fluorescence decay in time or phase provides another dimension from which the parameters of interest can be estimated. The use of a laser, as a light source, enhances the properties of fluorescence. The high intensity as well as the narrow bandwidth of the laser light increases the quality of the measurements.

Despite all the advantages and power, fluorescence is not widely used in industry. The main reason for this lack of use is the great complexity of the information given by molecular fluorescence techniques. Interactions between components and sensitivity to environmental conditions have thus far limited its use to systems containing only one pure fluorescent component. For the fluorescence technique to be useful, one must be able to sort out the interactions before the composition of a sample can be estimated.

"SMART SENSING" USING LASER-INDUCED FLUORESCENCE

However, lack of linearity and specificity should not prevent one from obtaining interesting information from fluorescence measurements. Today, with the use of computers in data processing, valid information can be extracted from very nonlinear and modulated measurements. The requirement of linearity and specificity that a sensor traditionally had to meet is no longer needed. Selectivity is not required, but what is important is "selective sensitivity," which means that the parameter to be estimated should influence the set of sensors in some way. Today's sensor must be "smart." A sensor is no longer restricted to a device that gives a signal proportional to a physical/chemical parameter. Rather, it is a system which uses a great deal of data and takes advantage of our knowledge on the sensing process. This approach, called "smart sensing," is the one that will give the capability to fully use fluorescence measurements. The realization of a "smart sensing" system requires interdisciplinary knowledge. Knowledge in spectroscopy, biotechnology as

well as in data processing should be combined in the study of a fluorescence sensor for fermentation monitoring.

The project described here was developed in collaboration with the National Bureau of Standards (NBS) in Gaithersburg, Maryland, USA. This work attempts to develop a general numerical technique for signal deconvolution.

COMMERCIAL PROBE CHARACTERIZATION

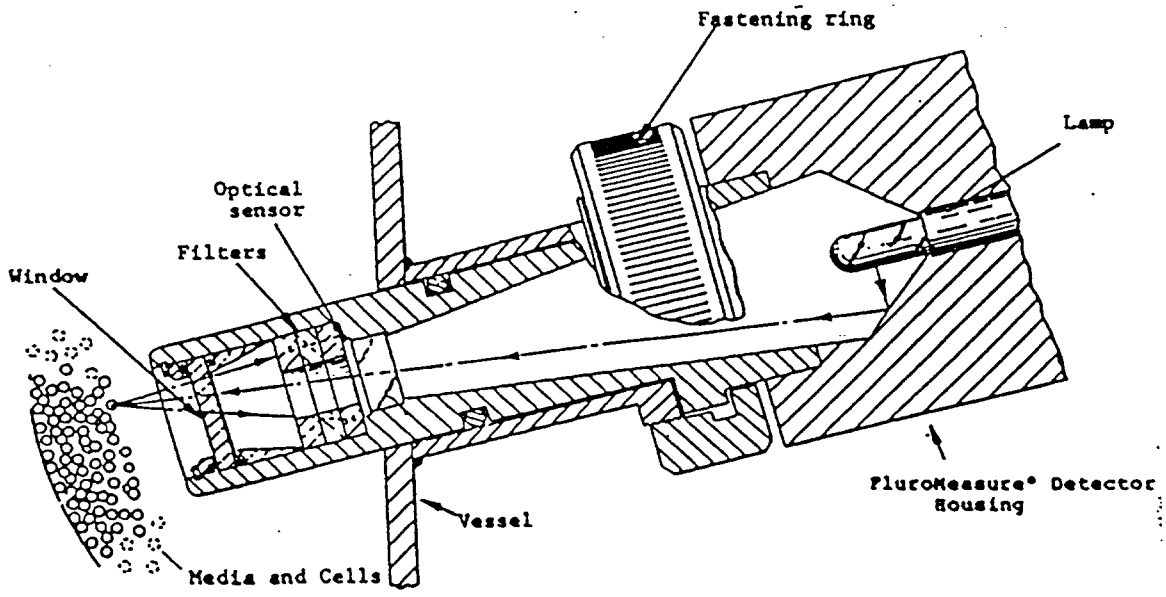


Figure 1. Schematic of the NADH probe. Reproduced from [MacBride et al. 1986].[10]

A sterilizable probe which can monitor the fluorescence of a microbial fermentation has recently become commercially available from two manufacturers: Ingold, Inc.[7] and BioChem Technology, Inc.[10] A schematic of the probe produced by BioChem Technology, Inc. is shown in Figure 1. The probe has already been used in many applications, which have been reviewed elsewhere.[3, 4, 7] The operation of the probe is based on fluorescence of microbial cells within the fermentor. The

probe was designed so that it could detect the fluorescence of the reduced cofactor, NADH, which is an electron carrier central to all cells' energy metabolism. The fluorescence signal, however, is influenced by a number of factors. The manner in which these factors affect the fluorescence signal has not been well characterized. In this study, the response of this probe was modeled in order to account for some of these factors. The model developed by one of the authors earlier[11] can be used as a basis for analyzing the fluorescence signal measured by the probe. Similar analysis can be extended to general fluorescence measurements in other configurations.

THEORY

FLUORESCENCE

Fluorescence results when a molecule, excited by absorption of light, emits light at a lower frequency when returning to its ground state. There are usually many closely spaced energy levels in both the excited and ground electronic states due to vibrational energy levels and interactions with surrounding molecules. The molecule is usually excited from the lowest energy level of the ground state. Radiationless decay also occurs along with fluorescence as the excited molecules climb down closely spaced energy levels. The time constant for this radiationless decay is on the order of 10^{-15} s, whereas the time constant for fluorescence is on the order of 10^{-4}–10^{-10} s. Thus, the excited molecules lose some energy via radiationless decay before and after emitting radiation. In addition, although the majority of the molecules that absorb the excitation light are in the lowest energy level, the emission of fluorescence returns the excited molecule to *any* of the available energy levels in the ground state. For these two reasons the emission wavelength is always statistically longer than the excitation wavelength. In general, not *all* of the excited molecules emit light because they may also relax via radiationless decay or undergo a photochemical reaction to return to the ground state. Often there is a metastable triplet state

that causes the emitted light to be long-lived, on the order of seconds or minutes. This type of radiation is called phosphorescence.

FLUORESCENCE MEASUREMENT MODEL

We will start with a model that is independent of the geometric configuration of the measurement and absorbance of the sample. First, we will assume that the intensity of excitation light at a point within the sample is proportional to the intensity of excitation light at the source. The proportionality constant is assumed to be a function of position and absorbance at excitation wavelength:

$$I = I_0 g(x, y, z, A^{\lambda_{ex}}). \tag{1}$$

The amount of light at wavelength, λ_{em}, emitted by a fluorophore within a differential volume is proportional to the amount of light absorbed by the fluorophore, which is, in turn, proportional to the fluorophore concentration:

$$dF_{i,\mathrm{TOTAL}}^{\lambda_{em}}(x, y, z) = \phi_i^{\lambda_{em}} a_i^{\lambda_{ex}} c_i I_0 g(x, y, z, A^{\lambda_{ex}})\, dv. \tag{2}$$

The fraction of this emitted light that is actually measured by the detector is also assumed to be a function of position within the sample and absorbance of the sample at the emission wavelength, λ_{em}:

$$dF_i^{\lambda_{em}}(x, y, z) = h(x, y, z, A^{\lambda_{em}}) \cdot dF_{i,\mathrm{TOTAL}}^{\lambda_{em}}(x, y, z). \tag{3}$$

Integrating over the volume of the sample and summing over all fluorescent species, gives the total amount of emitted light measured by the detector:

$$\boldsymbol{F}^{\lambda_{em}} = \sum_{i=1}^{n} \boldsymbol{F}_i^{\lambda_{em}} = \sum_{i=1}^{n} \phi_i^{\lambda_{em}} a_i^{\lambda_{ex}} c_i I_0 \int_V g(x, y, z, A^{\lambda_{ex}}) h(x, y, z, A^{\lambda_{em}})\, dv. \tag{4}$$

A new function, Γ, can be defined as:

$$\Gamma(A^{\lambda_{ex}}, A^{\lambda_{em}}) = \int_V g(x, y, z, A^{\lambda_{ex}}) h(x, y, z, A^{\lambda_{em}})\, dv. \tag{5}$$

This function depends only on the geometry of the system. One can now express the fluorescence divided by Γ as a linear function of the fluorophore concentrations:

$$\frac{F^{\lambda_{em}}}{\Gamma(A^{\lambda_{ex}}, A^{\lambda_{em}})} = I_0 \sum_{i=1}^{n} \left(\phi_i^{\lambda_{em}} a_i^{\lambda_{ex}}\right) c_i, \tag{6}$$

where, the factors, $\left(\phi_i^{\lambda_{em}} a_i^{\lambda_{ex}}\right)$, can be found from the fluorescence spectra of the pure components.

APPLICATION OF MEASUREMENT MODEL

By considering the geometric configuration of a given system, we approximate the functional form of $\Gamma(A^{\lambda_{ex}}, A^{\lambda_{em}})$.

1. *NADH Probe*

First, we will consider the NADH probe, shown in Figure 1. We will use the Beer-Lambert law and an idealized model of the light-source/detector configuration. We will consider a one-dimensional model, in which we can assume that the excitation light is a collimated beam perpendicular to the probe surface, that the intensity of the excitation light depends only on the distance from the probe tip, and that the emission light must travel this same distance to reach the detector. According to the Beer-Lambert law, this one-dimensional model gives:

$$I = I_0 \exp\left(-x \sum_{i=1}^{n} a_i^{\lambda_{ex}} c_i\right). \tag{7}$$

Thus, noting that $A^{\lambda_{ex}} = \sum_{i=1}^{n} a_i^{\lambda_{ex}} c_i$ and comparing the above equation to equation (1), we can express the function g as:

$$g(x, y, z, A^{\lambda_{ex}}) = \exp\left(-x A^{\lambda_{ex}}\right). \tag{8}$$

The intensity of the emitted fluorescent light is proportional to amount of light emitted, according to the Beer-Lambert law. However, not all of the light is aimed back at the probe, where the detector is. As shown in Figure 2, the cone angle, φ_2',

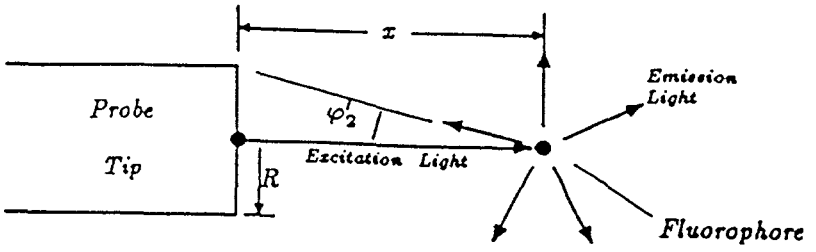


Figure 2. Amount of total light emitted that is aimed back at the detector surface as defined by the cone angle, φ_2' or the distance from the probe, x, and the radius, R, of the effective detector surface.

---

defines the "monitoring efficiency" at a given point within the fermentor. This cone angle is a function of the distance from the probe tip, x. Assuming the direction of the emitted light is completely random, the "monitoring efficiency" is equal to the fractional area of a sphere of arbitrary radius, r, that is bounded by the cone angle, φ_2', where φ_2' ranges from $\varphi_2' = 0$ as $x \to \infty$ to $\varphi_2' = \pi/2$ when $x = 0$. This area is equal to $2\pi r^2(1 - \cos\varphi_2')$. Since the area of the entire sphere is equal to $4\pi r^2$, the fractional area is equal to: $\frac{1}{2}(1 - \cos\varphi_2')$. Noting that $\cos\varphi_2' = \left(\frac{x}{\sqrt{x^2 + R^2}}\right)$, where R is the radius of the probe tip, the "monitoring efficiency" is given by:

$$f(x, R) = \frac{1}{2}\left[1 - \frac{x}{\sqrt{x^2 + R^2}}\right]. \tag{9}$$

Thus, the amount of light resulting from the fluorescence of species i in the differential slice that reaches the detector is given by:

$$dF_i^{\lambda_{em}} = \frac{1}{2}\left[1 - \frac{x}{\sqrt{x^2 + R^2}}\right] \exp\left(-x\sum_{k=1}^{N}(a_k^{\lambda_{em}} c_k)\right) dF_{i,\mathrm{TOTAL}}^{\lambda_{em}}. \tag{10}$$

Comparing this equation to equation (3) and noting that $A^{\lambda_{em}} = \sum_{k=1}^{N}(a_k^{\lambda_{em}} c_k)$, we can express the function h as:

$$h(x, y, z, A^{\lambda_{em}}) = \frac{1}{2}\left[1 - \frac{x}{\sqrt{x^2 + R^2}}\right] \exp\left(-xA^{\lambda_{em}}\right). \tag{11}$$

Now, we can express the function Γ, as defined in equation (5), as:

$$\Gamma(A^{\lambda_{ex}}, A^{\lambda_{em}}) = \int_0^L \frac{1}{2}\left[1 - \frac{x}{\sqrt{x^2+R^2}}\right] \exp\left(-x\left[A^{\lambda_{ex}} + A^{\lambda_{em}}\right]\right)\, dx, \qquad (12)$$

where L is the path length of the sample. In order to evaluate this integral analytically, we will approximate the factor, $\left[1 - \frac{x}{\sqrt{x^2+R^2}}\right]$, with an exponential factor, $\exp(-xS)$, in which S is an adjustable parameter. After substituting this approximation, the integral is evaluated as:

$$\Gamma(A^{\lambda_{ex}}, A^{\lambda_{em}}) = \left(\frac{1}{2}\right) \frac{1 - \exp\left(-L\left[A^{\lambda_{ex}} + A^{\lambda_{em}} + S\right]\right)}{A^{\lambda_{ex}} + A^{\lambda_{em}} + S}. \qquad (13)$$

For a mixture of fluorescent compounds, the function $\Gamma(A^{\lambda_{ex}}, A^{\lambda_{em}})$ can be expressed as a function of concentrations:

$$\Gamma(c_i) = \frac{1}{2} \frac{\left[1 - e^{-L\left[\sum_{j=1}^{N}\left(a_i^{\lambda_{ex}} + a_i^{\lambda_{em}}\right)c_i + S\right]}\right]}{\sum_{i=1}^{N}\left(a_i^{\lambda_{ex}} + a_i^{\lambda_{em}}\right)c_i + S}. \qquad (14)$$

Substituting this expression into equation (6), we can predict the following dependence of fluorescence on fluorophore concentrations:

$$F^{\lambda_{em}} = \frac{1}{2} I_0 \sum_{i=1}^{n} \phi_i^{\lambda_{em}} a_i^{\lambda_{ex}} c_i \frac{\left[1 - e^{-L\left[\sum_{j=1}^{N}\left(a_i^{\lambda_{ex}} + a_i^{\lambda_{em}}\right)c_i + S\right]}\right]}{\sum_{k=1}^{N}\left(a_k^{\lambda_{ex}} + a_k^{\lambda_{em}}\right)c_k + S}. \qquad (15)$$

2. *Laser-Induced Fluorescence – Absorbance Correction*

Now, we will consider the laser induced fluorescence setup, shown in Figure 3. In this case, since our samples were not highly absorbent at the emission wavelengths, we can neglect the absorption of this light. In addition, since the path length of the sample was only 1 cm, we can assume that the monitoring efficiency is a constant value, k. In this case, the function, $\Gamma(A^{\lambda_{ex}}, A^{\lambda_{em}})$, can be expressed as:

$$\Gamma(A^{\lambda_{ex}}, A^{\lambda_{em}}) = \int_0^1 k \cdot \exp\left(-xA^{\lambda_{ex}}\right)\, dx = \frac{k\left[1 - \exp\left(-A^{\lambda_{ex}}\right)\right]}{A^{\lambda_{ex}}}. \qquad (16)$$

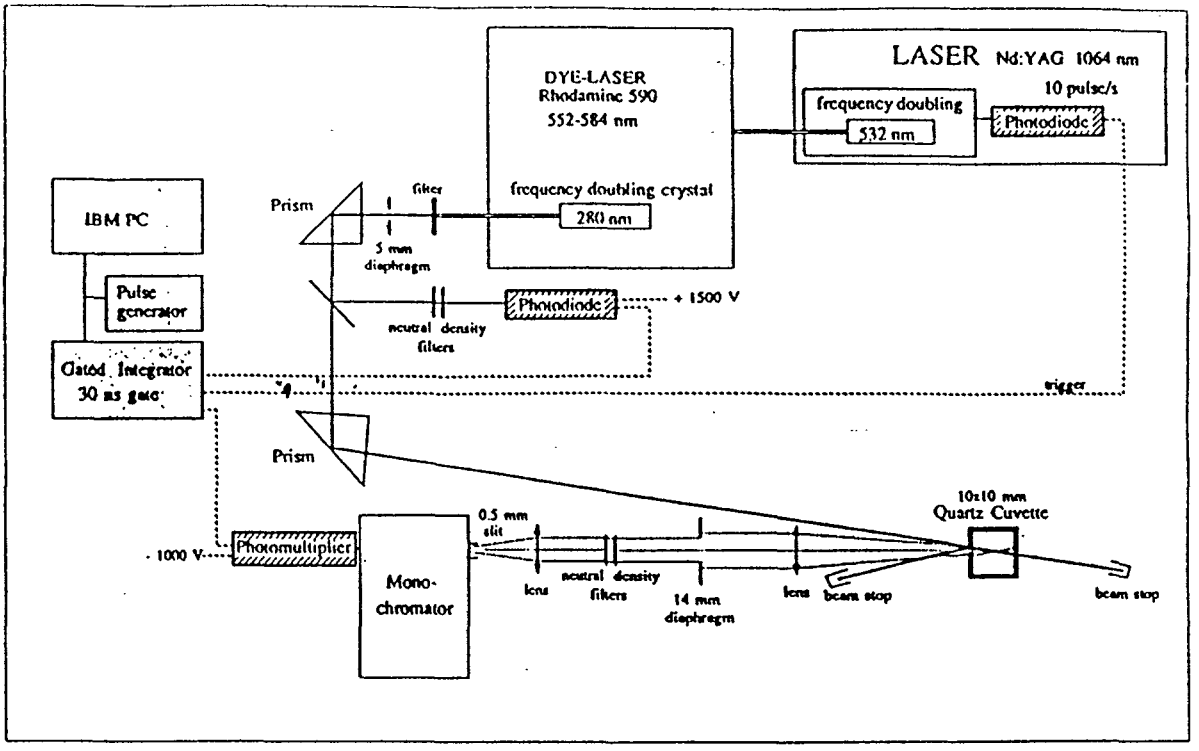


Figure 3. Laser induced fluorescence experimental setup.

MULTIVARIATE ANALYSIS TECHNIQUES

In order to resolve a mixture, many multivariate analysis techniques have been considered. Two of them are presented below. The problem addressed by these methods can be stated as follows:

⋆ Estimate the concentration of the main fluorescent compounds in a mixture.

We can assume the following:

⋆ All the fluorescent compounds are known.

⋆ The spectra of the pure components are known as well as the spectra of some reference mixtures if needed.

These assumptions are reasonable for the analysis of fluorescence in fermentation processes. The important fluorophores are known and often well studied. It is not a

problem to get the spectra of the pure components as well as some mixtures spectra.

MULTIPLE LINEAR REGRESSION

Multiple linear regression is the most classical way to analyze experiments when one variable y is related to a number of $\mathbf{x}$ variables. The model is written as follows :

$$y = \mathbf{X}\mathbf{b} + e. \tag{17}$$

In the present study, $\mathbf{y}$ is a column vector of length n representing the spectrum of the mixture, $\mathbf{X}$ is a $n \times m$ matrix which columns represent the pure components spectra. $\mathbf{b}$ is a column vector of m coefficients. These coefficients are the ones of the linear combination of pure components spectra:

$$y_i = b_1 x_{i1} + b_2 x_{i2} + \cdots + b_m x_{im} + e_i. \tag{18}$$

$\mathbf{e}$ is a residual vector. Using the hypothesis of linearity for the mixture spectrum, the b_j should be equal to the concentration of the component j in the mixture divided by its concentration in the pure sample x_{ij}. Since the spectra are defined with up to 3000 points and only 2 or 3 components are considered, $m \ll n$ and there is no exact solution. But one can get a solution by minimizing the norm of the residual vector e. The problem is then an optimization problem:

$$\min_{\mathbf{b}} \sum_{i=1}^{n} e_i^2. \tag{19}$$

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}. \tag{20}$$

The solution of this least squares problem is well known:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{21}$$

The limitations of this method are the assumption that the matrix $\mathbf{X}$ is exact and the inversion of $\mathbf{X}'\mathbf{X}$. Indeed as soon as m gets large some collinearity problems appear and the matrix inversion is impossible. To overcome the collinearity

problem, some rank reducing methods such as principal component regression or partial least squares regression are needed. The two methods presented below are slightly different from multiple linear regression in the way they use the data. They use a set of reference measurements rather than only the pure components spectra. These reference measurements are done with mixtures of known concentration of each component. The pure components spectra were the reference measurements of multiple linear regression but no mixture spectra could be used in this method due to collinearity problems which would appear. Since the other methods do not have this problem, the use of mixtures data to build the model is possible. Some nonlinear effects can be taken into account by these two methods due to the use of mixture data.

PARTIAL LEAST SQUARES REGRESSION

New notations are used for the partial least squares method. Now $\mathbf{Y}$ is a matrix $n \times p$, each line of which is a vector of fluorophore concentrations in a reference mixture. $\mathbf{X}$ is a $n \times m$ matrix which lines are the fluorescence spectra of reference mixtures. For this method, the variables are usually mean-centered and scaled to unit variance before processing.

Mean-centered variables are obtained by subtracting to every data point x_{ij} or y_{ij} the mean of its column. The unit variance scaling is obtained by dividing every data point by the standard deviation of its column. These two processes improve the conditioning of the data. If some variables are less important or significant than the others, one can reduce the variance of the corresponding column by multiplying it by a weighting factor.

Partial least squares modeling, which has been developed by H. Wold et al.,[12] is a rank reducing technique like principal component analysis. The difference between partial least squares regression and principal component regression is that

the $\mathbf{Y}$ data are used to determine a decomposition of $\mathbf{X}$ in lower rank matrices which is optimal for prediction. The complete model uses three equations, two for the decompositions of $\mathbf{X}$ and $\mathbf{Y}$ called outer relations, and a relation between the two preceding decompositions called inner relation. The three equations are the following :

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E}. \tag{22}$$

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{F}. \tag{23}$$

$$\mathbf{U} = \mathbf{TB}. \tag{24}$$

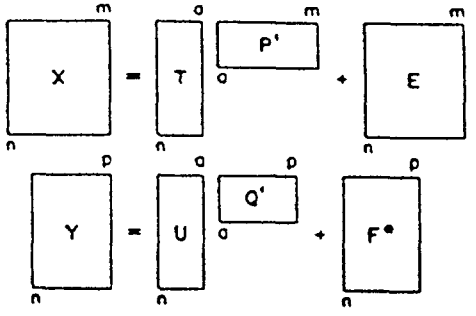


Figure 4. The two outer relations of partial least squares model.

Where $\mathbf{B}$ is a diagonal $a \times a$ matrix and the others are represented in Figure 4. In this model, the matrix $\mathbf{T}$ is a projection of $\mathbf{X}$ as in principal component regression but is calculated both to approximate $\mathbf{X}$ and to predict $\mathbf{Y}$. The algorithm is given in Figure 5. The properties of the partial least squares factors are the following:

- $\star$ $\mathbf{p}_h'$ and $\mathbf{q}_h'$ have unit length
- $\star$ $\mathbf{t}_h$ and $\mathbf{u}_h$ are centered around zero for each h
- $\star$ $\mathbf{w}_h$ and $\mathbf{t}_h$ are orthogonal

The PLS algorithm

It is assumed that X and Y are mean-centered and scaled:

For each component: (1) take u_{start} = some y_j.

In the X block: (2) $w' = u'X/u'u$

(3) $w'_{new} = w'_{old}/\|w'_{old}\|$ (normalization)

(4) $t = Xw/w'w$

In the Y block: (5) $q' = t'Y/t't$

(6) $q'_{new} = q'_{old}/\|q'_{old}\|$ (normalization)

(7) $u = Yq/q'q$

Check convergence: (8) compare the t in step 4 with the one from the preceding iteration. If they are equal (within a certain rounding error) go to step 9, else go to step 2. (If the Y block has only one variable, steps 5—8 can be omitted by putting $q = 1$, and no more iteration is necessary.)

Calculate the X loadings and rescale the scores and weights accordingly:

(9) $p' = t'X/t't$

(10) $p'_{new} = p'_{old}/\|p'_{old}\|$ (normalization)

(11) $t_{new} = t_{old}\|p'_{old}\|$

(12) $w'_{new} = w'_{old}\|p'_{old}\|$

(p', q' and w' should be saved for prediction; t and u can be saved for diagnostic and/or classification purposes).

Find the regression coefficient b for the inner relation:

(13) $b = u't/t't$

Calculation of the residuals. The general outer relation for the X block (for component h) is

$E_h = E_{h-1} - t_h p'_h$; $X = E_0$

The mixed relation for the Y block (for component h) is

$F_h = F_{h-1} - b_h t_h q'_h$; $Y = F_0$

From here, one goes to Step 1 to implement the procedure for the next component. (Note: After the first component, X in steps 2, 4 and 9 and Y in steps 5 and 7 are replaced by their corresponding residual matrices E_h and F_h.)

Figure 5. The partial least squares algorithm. Reproduced from [Geladi and Kowalski, 1986].[6]

The algorithm was developed upon more or less intuitive arguments and its theoretical interpretation is not completely clear yet. Many papers try to clarify the theory of this algorithm.[9] The main part of the algorithm (steps 2 to 7) consists of a powerful method to find the eigenvectors of $\mathbf{X'YY'X}$ and therefore it proves a relation with the singular value decomposition of $\mathbf{X'Y}$. The method used to compute the eigenvectors is not completely safe. Indeed the algorithm can diverge in

case of very close eigenvalues. However it converges fast for almost all the matrices. The use of **Y** in the computation of the eigenvectors gives to the method a higher predictive power than principal component regression.[5, 8] The number of components to take into account in the model is computed by using a cross validation technique. The prediction step is very easy. **p**, **q**, **w** and **b** from the calibration step have been saved for this purpose. The algorithm for prediction is:

1. $\mathbf{E}_0 = \mathbf{X}\ \ \mathbf{Y} = 0$

2. for $h = 1$ to a:

$$\mathbf{t}_h = \mathbf{E}_{h-1}\mathbf{w}_h.$$

$$\mathbf{E}_h = \mathbf{E}_{h-1} - \mathbf{t}_h\mathbf{p}'_h.$$

$$\mathbf{Y} = \mathbf{Y} + \mathbf{b}_h\mathbf{t}_h\mathbf{q}'_h$$

where a is the number of components to be included in the model. Applications have shown the power of partial least squares and its wide range of application. Lindberg used it to analyze spectrofluorometric data from mixtures of humic acid and ligninsulfonate,[8] and Frank and Kowalski applied partial least squares modeling to the prediction of wine quality and geographic origin from chemical measurements.[5]

EXPERIMENTAL METHODS

All the experiments presented in this study were made at the National Bureau of Standards facilities in Gaithersburg, Maryland, under the direction of Hratch G. Semerjian and John J. Horvath.

The experimental setup, used in this research, is designed to measure fluorescence spectra in the ultraviolet and visible region. A front surface detection geometry was used. In this configuration, the fluorescence is measured on the illuminated surface, in the opposite direction of the excitation beam. This configuration allows measurement in very opaque or turbid solutions. Furthermore, the design of a probe

with fiber optics is very simple, so this geometry is likely to be the one selected in any industrial device.

The ultraviolet light beam, generated by a laser, hits the surface of the cuvette with less than 10° of incidence. The fluorescent light emitted by the sample is then collected perpendicularly to the illuminated surface by a set of lenses, and directed into the slit of a monochromator. The monochromatic light is measured by a photomultiplier tube. The excitation light intensity is also measured by a photodiode.

Light Source

The light source is composed of two lasers. The first one, a Nd:YAG laser (Quantel YG581C), generates pulsed light at 1064 nm. After frequency doubling, the light at 532 nm excites a dye laser (Quantel TDL) with a Rhodamine dye which produces light at a tunable wavelength between 552 and 584 nm. Another frequency doubling gives the ultraviolet light usually tuned at 280 nm.

Sample Illumination

The beam which leaves the laser, crosses a quartz plate, which reflects 4% of the light to a photodiode used to record the laser power. The light is then directed to a $1\,cm \times 1\,cm$ quartz cuvette which contains the sample. A high quality quartz cuvette is indispensable because ordinary glass absorbs the light at 280 nm and even quartz may absorb a little and fluoresce at the wavelengths used in the experiments if it contains any impurity. The fluorescence is collected by a set of lenses. The first lens collimates the light coming from the cuvette in a parallel beam. The focal point of the first lens is the middle of the cuvette. A 14 mm diaphragm is placed in the beam path before a set of neutral density filters used to reduce the intensity of the light to a value acceptable to the electronics. Another lens focuses the beam on the slit of the monochromator. The monitoring efficiency is about 0.03%.

Measurement Devices

The measurement of the fluorescence is made by a monochromator and a photomultiplier, and the laser intensity is recorded through a photodiode. The monochromator (GCA/McPherson EU-700) is composed of two parabolic mirrors, one for collimating and the other for focusing. The wavelength dispersing element is a plane diffraction grating. The photomultiplier tube (Hamamatsu R955) has multialkali (Na-K-Sb-Cs) photocathode which, combined with a fused-silica window, gives a very wide spectral response from the ultraviolet to the near infrared region with a quantum efficiency varying between 15-25% and a sensitivity of 60-70 mA/W in our range of measurement. The electron multiplier is a 9 stage circular-cage, giving an amplification of 10^7 for a rise time of 2.2 ns.

Electronic Processing

The photomultiplier tube and the photodiode described earlier, are connected to a boxcar integrator. This electronic system performs an integration of the two signals over the time period of a pulse, approximately 30 ns. A photodiode, installed in the Nd:YAG laser, triggers this integration. The boxcar also allows to make an averaging of 3, 5, 10 or 30 measurements. The averaging reduces the noise on the measurements but has some drawbacks. Indeed, since the monochromator keeps scanning, the averaging is made with measurements taken at different wavelengths all shorter than the one the average value is affected at. This method of averaging results in a shift of the measurements towards longer wavelengths. If v is the scanning speed of the monochromator in nm/s, the step between two flashes of the laser is $(v/10)\,nm$. Therefore if the boxcar averages N measurements, the computer can take a sample every $(Nv/10)\,nm$. This rate of measurement allows that every pulse of the laser is used for one and only one measurement, preserving the independence of the measurements and taking advantage of all the information

available. When this averaging procedure of the boxcar is used, one should take care to shift all the measurements by $-(Nv/20)\,nm$ in order to get an unbiased spectrum. The amount of noise is then reduced by $\sqrt{N}$. This reduction of noise is only obtained by a reduction of information (the number of measurements is divided by N) and is with this respect similar to smoothing techniques described below, even if it is directly accomplished by the electronics.

An IBM AT controls all the measurement process. The user can choose the scanning speed of the monochromator and the rate of data acquisition through a program written by researchers at the National Bureau of Standards. The data acquisition rate is limited to 10 readings per second since the measurements are integrated over one pulse and there are 10 pulses per second.

The fluorescence was recorded between 260 and 460 nm. This range contains all the interesting peaks (tryptophan at 340 nm, tyrosine at 305nm, etc.). The scanning speed was usually 10 $\AA/s$. Sometimes the internal averaging of the boxcar was used. In this configuration 30 measurements were averaged and 2 readings per second were taken. More often the maximum rate of data acquisition of 10 readings per second was selected. Naturally no averaging was made at this rate.

Data Smoothing

In order to get a signal that is easier to analyze, the signal was smoothed using a low pass filter. The filtering is done by using a Fourier transform. Fast Fourier transform is performed on the spectrum, and the high frequencies are set to zero. An inverse transform gives the smoothed spectrum. Many experiments were recorded with 2000 points. The scattering peak was removed from the spectrum, and then a 2048-point fast Fourier transform was used.

RESULTS AND DISCUSSION

MULTIPLE LINEAR REGRESSION

The first attempt at resolving a multiple components mixture has been made using a simple multiple linear regression. Table 0 shows the results. The results are almost all reproducible to within 10% for duplicate samples. However, the estimates are all wrong, the errors varying between 10% and 300%. The direction of the errors is clearly the same for all the measurements. Therefore the errors are systematic, and it should be possible to figure out the cause of this very poor analysis. The tyrosine peak is higher than the tryptophan one in pure solution but in a mixture the tyrosine peak is much smaller than the tryptophan peak. What can produce this effect

	Real		Estimated	
Sample	Tryptophan	Tyrosine	Tryptophan	Tyrosine
Mix1a	0.50	0.50	0.778	0.239
Mix2a	0.50	0.50	0.747	0.217
Mix3a	0.75	0.25	0.798	0.071
Mix4a	0.75	0.25	0.851	0.097
Mix5a	0.25	0.75	0.517	0.459
Mix6a	0.25	0.75	0.502	0.461
Mix7a	0.10	0.90	0.297	0.698
Mix8a	0.10	0.90	0.273	0.836

Table 1. Multilinear estimation of the composition of some tryptophan-tyrosine mixtures. All the concentrations are given in mM. The pure components reference spectra used in the regression are 10^{-3} *M.*

The reabsorption of tyrosine fluorescence by tryptophan is a possibility. But, in fact, the absorption by tryptophan in the 300 *nm* range is too low for this, only

a tenth of its absorbance at 280 nm. Also, if tryptophan would absorb the tyrosine fluorescence, then it should reemit approximately 20% of the absorbed light; this would give a tryptophan peak only slightly higher than the one predicted by the linear combination. Therefore a reason other than reabsorption should be found. For this purpose, let us consider the model derived earlier in this paper.

CORRECTION FOR ABSORPTION

The mixture spectrum can be written as a function of the pure component spectra as follows:

$$\frac{F}{\Gamma(A)} = \frac{c_1}{c_{10}}\frac{F_{10}}{\Gamma(A_{10})} + \frac{c_2}{c_{20}}\frac{F_{20}}{\Gamma(A_{20})}. \tag{25}$$

The concentrations c_1 and c_2 can then be deduced by a simple linear regression. As a matter of fact, this is the same regression that was done in the multiple linear regression. The multiple linear regression gives the two coefficients α and β such that:

$$F = \alpha F_{10} + \beta F_{20}. \tag{26}$$

Therefore :

$$c_1 = \alpha c_{10}\frac{\Gamma(A_{10})}{\Gamma(A)}. \tag{27}$$

$$c_2 = \beta c_{20}\frac{\Gamma(A_{20})}{\Gamma(A)}. \tag{28}$$

In order to estimate the performance of this model on the mixture measurements presented before, an estimation of Γ is needed.

The estimation of Γ given earlier was:

$$\Gamma(A) = \frac{k}{A}(1 - \exp(-A)). \tag{29}$$

For the pure components and mixtures measurements presented before, an estimation of the absorbance is possible from the known concentrations and therefore an

estimation of Γ is possible. The absorptivities of the pure components were obtained from the literature. The absorptivities of the mixtures were calculated by

Tryptophan (10^{-3} mol/l)	Tyrosine (10^{-3} mol/l)	Absorbance	$\Gamma(A)$
1.00	0.00	11.50	0.087
0.00	1.00	3.45	0.280
0.75	0.25	9.49	0.105
0.50	0.50	7.47	0.134
0.25	0.75	5.46	0.182
0.10	0.90	4.26	0.231

Table 2. Estimation of the absorbance of some mixtures of tryptophan and tyrosine.

	Real		Estimated	
Sample	Tryptophan	Tyrosine	Tryptophan	Tyrosine
Mix1a	0.50	0.50	0.51	0.50
Mix2a	0.50	0.50	0.49	0.45
Mix3a	0.75	0.25	0.66	0.19
Mix4a	0.75	0.25	0.71	0.26
Mix5a	0.25	0.75	0.25	0.71
Mix6a	0.25	0.75	0.24	0.71
Mix7a	0.10	0.90	0.11	0.85
Mix8a	0.10	0.90	0.10	1.01

Table 3. Multilinear estimation of the composition of some tryptophan-tyrosine mixtures corrected for absorption. The concentrations are given in mM. The pure components reference spectra used in the regression are 10^{-3} M.

linear combination.

Table 2 presents these estimations. Using these values of Γ, it is possible to compute the concentrations of the mixtures from the results of the multiple linear regression given in Table 1. Table 3 presents the new results. The quality of the results is surprising when one considers the numerous approximations made in the computation of Γ and in the estimation of the absorbances. The estimates of the concentration are as good as they can be considering the precision of the measurements. The errors are almost all under 10% and seem random. It proves that, in these amino acids mixtures, the nonlinearity is mainly introduced by the variation of absorbance at the excitation wavelength. The simple model developed can eliminate this nonlinearity if a measurement of the absorption at the excitation wavelength is provided.

PARTIAL LEAST SQUARES REGRESSION

A partial least squares regression program has been written using the algorithm presented by Geladi and Kowalski.[6] It was used to estimate the concentrations of the mixtures already studied in the preceding section. From each smoothed spectrum thirty values are taken, one every $5\,nm$ from 285 up to $430\,nm$. The reference spectra set needed for partial least squares is composed of the two pure component spectra, a mixture 0.5 mM tryptophan and 0.5 mM tyrosine, a mixture 0.75 mM tryptophan and 0.25 mM tyrosine and a mixture 0.25 mM tryptophan and 0.75 mM tyrosine. The five spectra are stored in a 5×30 matrix and the corresponding concentrations in a 5×2 matrix. Each variable is mean-centered and scaled to unit variance. The partial least squares algorithm is then used to determine a model at the maximum order 5. The **p**, **q** and **w** vectors are saved for the prediction.

The tables 4shows the estimates of concentration of the five spectra included in the model as the order of the model increases. The last line of Table 4 shows the sum of the squares of the errors of prediction as the number of components included

	Real	Order 1	Order 2	Order 3	Order 4	Order 5
Tryptophan	1.00	0.868	0.957	0.993	1.0	1.0
Tyrosine	0.00	0.132	0.043	0.007	10^{-16}	10^{-16}
Tryptophan	0.00	-0.090	-0.006	0.004	0.0	0.00
Tyrosine	1.00	1.090	1.006	0.996	1.0	1.00
Tryptophan	0.50	0.630	0.501	0.515	0.5	0.50
Tyrosine	0.50	0.370	0.499	0.485	0.5	0.50
Tryptophan	0.75	0.723	0.802	0.754	0.75	0.75
Tyrosine	0.25	0.277	0.198	0.246	0.25	0.25
Tryptophan	0.25	0.368	0.246	0.234	0.25	0.25
Tyrosine	0.75	0.631	0.754	0.766	0.75	0.75
$\sum$Error		0.114	0.0092	0.0011	0.000	0.000

Table 4. Partial Least Squares estimation of the composition of a tryptophan and tyrosine solution. The concentrations are given in mM. The last row lists the sum of the squares of the estimation errors at different orders.

in the model increases. As expected, the accuracy of the estimates increases with the order of the model and the exact values are obtained when the order is equal to the size of the reference set. The model can be used to estimate the concentration of mixtures not included in the reference set.

Table 6shows the results of this estimation.

The prediction with 3 components is the best one. The errors in prediction are in the order of 0.03 mM and are within the measurement errors. Therefore the result can be considered as very satisfactory.

The algorithm performed very well on this example. The computation of the model was very fast. The convergence in the loop (steps 2 to 8 of the algorithm) was

	Real	Order 1	Order 2	Order 3	Order 4	Order 5
Tryptophan	0.50	0.621	0.536	0.509	0.496	0.502
Tyrosine	0.50	0.379	0.464	0.491	0.504	0.498
Tryptophan	0.75	0.720	0.727	0.732	0.743	0.742
Tyrosine	0.25	0.280	0.273	0.268	0.257	0.258
Tryptophan	0.25	0.351	0.292	0.294	0.297	0.313
Tyrosine	0.75	0.649	0.708	0.706	0.703	0.687
Tryptophan	0.10	0.152	0.128	0.149	0.151	0.148
Tyrosine	0.90	0.848	0.872	0.851	0.849	0.852
Tryptophan	0.10	0.116	0.039	0.094	0.089	0.076
Tyrosine	0.90	0.884	0.961	0.906	0.911	0.923
$\sum$Error		0.057	0.0162	0.0096	0.010	0.0137

Table 5. Partial Least Squares estimation of the composition of a binary mixture of tryptophan and tyrosine. The last row lists the sum of the squares of the estimation errors at different orders.

generally obtained in 2 iterations. The prediction requires only vector and matrix multiplications which are almost instantaneous on a personal computer.

CONCLUSIONS

The purpose of this study is an estimation of the capabilities of fluorescence in fermentation monitoring. Through rigorous sensor modeling and the use of numerical methods, digital data analysis techniques are developed to deconvolute multicomponent laser fluorescence data to estimate composition. The mathematical technique of simple multiple linear regression, model-based multiple linear regression, and partial least squares are employed, and their performance is compared. The use of neural net techniques are currently under investigation.

We have shown that sensor models and proper numerical methods can both be used to vastly improve the estimate of composition. However, there remain many problems to be solved. The choice of a calibration set, the number of calibration points, the number of wavelengths, the optimum wavelengths, and the estimation of nonfluorescent species are some of them. We are currently planning to apply the technique to real-time fermentation systems. economically. The goal in the fermentation study is to estimate the state of the culture so that intelligent control can be applied.

Lack of sensor specificity is a generic problem that is common to many of the new composition sensing techniques. The data processing technique in our smart sensor development is quite general in the sense that it can be easily adopted to other types of spectral analysis or array sensors. These include FTIR and NMR spectra in addition to fluorescence spectra. There are many chemical, clinical, and medical monitoring processes where fast, reliable techniques are indispensable, especially in taxing on-line applications.

ACKNOWLEDGEMENT

The use of experimental facilities at the National Bureau of Standards' Chemical Process Metrology Division under Dr. John Horvath and Dr. Hratch Semerjian is gratefully acknowledged.

REFERENCES

1. **Appellof, C.J. and E.R. Davidson.** 1981. Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents. *Anal. Chem.* **53**: 2053-2056.

2. **Armiger, W. B. and A. E. Humphrey.** 1979. Computer applications in fermentation technology. in Peppler, H. J. and D. Perlman, eds. Microbial Technology, volume 2. Academic Press, New York, NY.

3. **BioChem Technology.** FluroFacts. Malvern,PA.

4. **BioChem Technology.** 1987. The FluroMeasure® System User's Manual. Malvern, PA.

5. **Frank, I. E. and B. R. Kowalski.** 1984. Prediction of wine quality and geographic origin from chemical measurements by partial least squares modeling. *Anal. Chim Acta* **162**: 241-251.

6. **Geladi, P. and B. R. Kowalski.** 1986. Partial least-squares regression : a tutorial. *Anal. Chim Acta* **185**: 1-17.

7. **Ingold Electrodes, Inc.** Ingold Fluorosensor™: Theory and Applications. Wilmington, MA.

8. **Lindberg, W.** 1983. Partial least squares method for spectrofluorimetric analysis of mixtures of humic acid and ligninsulfonate. *Anal. Chem.* **55**: 643.

9. **Lorber, A., L. E. Wangen, and B. R. Kowalski.** 1987. A theoretical foundation for the partial least squares algorithm. *Journal of Chemometrics.* **1**: 19-31.

10. **MacBride, W.R., J.A. Magee, W.B. Armiger, and D.W. Zabriskie.** 1986. Optical Apparatus and Method for Measuring the Characteristics of Materials by Their Fluorescence. *U.S. Patent 4,577,110.*

11. **Simmons, M. B. and N. S. Wang.** 1987. Modeling of a commercial fluorescence probe. AIChE Fall Annual Meeting, New York, NY.

12. **Wold H. A.** 1982. Soft modeling. The basic design and some extensions. in K. G. Jöreskog and H. A. Wold eds. Systems Under Indirect Observation: Causality, Structure, Prediction. Elsevier Science Publishers, New York, NY.

SRC TR 89-18

A Note on the NP-hardness of the Topological Via Minimization Problem

by

C. S. Rim, T. Kashiwabara

and

K. Nakajima