Title
Chemometric analysis of frequency-domain photon migration data: quantitative measurements of optical properties and chromophore concentrations in multicomponent turbid media.

Journal
Applied Optics, 39(10)

ISSN
0003-6935

Authors
Berger, AJ
Venugopalan, V
Durkin, AJ
et al.

Publication Date
2000-04-01

DOI
10.1364/ao.39.001659

Chemometric analysis of frequency-domain photon migration data: quantitative measurements of optical properties and chromophore concentrations in multicomponent turbid media

Andrew J. Berger, Vasan Venugopalan, Anthony J. Durkin, Tuan Pham, and Bruce J. Tromberg

Frequency-domain photon migration (FDPM) is a widely used technique for measuring the optical properties (i.e., absorption, μ_a, and reduced scattering, μ_s', coefficients) of turbid samples. Typically, FDPM data analysis is performed with models based on a photon diffusion equation; however, analytical solutions are difficult to obtain for many realistic geometries. Here, we describe the use of models based instead on representative samples and multivariate calibration (chemometrics).

FDPM data at seven wavelengths (ranging from 674 to 956 nm) and multiple modulation frequencies (ranging from 50 to 600 MHz) were gathered from turbid samples containing mixtures of three absorbing dyes. Values for μ_a and μ_s' were extracted from the FDPM data in different ways, first with the diffusion theory and then with the chemometric technique of partial least squares. Dye concentrations were determined from the FDPM data by three methods, first by least-squares fits to the diffusion results and then by two chemometric approaches. The accuracy of the chemometric predictions was comparable or superior for all three dyes. Our results indicate that chemometrics can recover optical properties and dye concentrations from the frequency-dependent behavior of photon density waves, without the need for diffusion-based models. Future applications to more complicated geometries, lower-scattering samples, and simpler FDPM instrumentation are discussed.

OCIS codes: 170.1580, 170.5280, 170.4090.

1. Introduction

Frequency-domain photon migration (FDPM) is a well-characterized technique for quantitatively measuring optical properties of homogeneous bulk samples. In a FDPM experiment, intensity-modulated light is launched into a sample and detected at one or more locations away from the source.[1] The bulk optical properties of the sample, namely, the absorption coefficient μ_a and the reduced scattering coefficient μ_s', are calculated from the FDPM data. Numerous groups have shown that the FDPM technique can be used to extract estimates of μ_a and μ_s'.[1–4] From these, other properties of interest can be calculated. For example, by estimating μ_a at multiple wavelengths, one obtains an absorption spectrum from which can be calculated the concentrations of different chemical or biological constituents.[1] In the medical optics community, there has been recent interest in using FDPM as a means of characterizing human tissue for various medical purposes, such as oxygenation monitoring and tumor assessment.[5–9]

Light transport in tissue at source–detector separations of more than a few millimeters is well described by a diffusion approximation to the equation of radiative transport. In the frequency domain, this corresponds to the diffusion of photon density waves emanating from the light source, with μ_a and μ_s' determining the propagation properties of the waves.[10] To extract μ_a and μ_s' from FDPM data,

A. J. Berger, V. Venugopalan, T. Pham, and B. J. Tromberg (tromberg@bli.uci.edu) are with the Beckman Laser Institute, University of California, Irvine, Irvine, California 92612. V. Venugopalan is also with the Department of Chemical and Biochemical Engineering. A. J. Durkin is with Candela Corporation, Wayland, Massachusetts 01778.

Received 15 October 1999; revised manuscript received 5 January 2000.

0003-6935/00/101659-09$15.00/0

one must invert this model. The ease and accuracy with which one can perform this inversion and thereby extract fundamental parameters from FDPM data with the diffusion theory depend considerably on the measurement geometry. For infinite and semi-infinite measurement geometries, analytical solutions to the diffusion equation are readily obtained and can be inverted to provide μ_a and μ_s'.[11] However, for more complicated geometries, such as those presented by small-pathlength cuvettes or human body parts, finding analytical solutions to the diffusion equation becomes prohibitively difficult. Numerical inversions can be attempted for particular geometries, but these often are computationally intensive. Beyond the mathematical difficulties, the physical validity of the diffusion approximation suffers in measurements involving highly absorbing media or short source–detector separations.[12] In many geometries, therefore, FDPM data are easy to acquire but impossible or awkward to invert with diffusion models. Alternative data-analysis methods are needed that can model the complicated geometries with greater ease, as has been suggested.[13]

In this paper, we develop a FDPM data-analysis technique based on a multivariate calibration algorithm and a set of training samples rather than on diffusion theory alone. Multivariate techniques and neural networks have previously been used to extract tissue optical properties from spatially resolved steady-state diffuse reflectance profiles[13–15]; to the best of our knowledge, this is the first application of such methods to the frequency domain. Multifrequency (i.e., 50–600 MHz), multiwavelength FDPM curves are obtained for each training sample. Correlations between photon density wave phase and amplitude behavior and various parameters of interest (μ_a, μ_s', and dye concentration) are sought, guided by physical insight from the diffusion model. In such an approach, geometry-dependent effects can be modeled empirically. The correlation trends are then used to predict the same parameter(s) from other FDPM curves. This empirical approach to extracting information is often called chemometrics. Our major goals in this paper are one, to develop a general framework for applying chemometrics to multifrequency, multiwavelength FDPM data, and two, to compare the performance of diffusion- and training-set-based analysis techniques in extracting optical properties and concentrations from a FDPM test set.

2. Materials and Methods

A. Sample Preparation

Forty samples with varying absorption and scattering characteristics were created from mixtures of three stock solutions of absorbing dyes (nigrosin, Aldrich Co.; naphthol green B [naphthol], Sigma Chem.; and copper phthalocyanine tetrasulfonic acid [CPTA], Sigma Chem.), a scattering solution (Intralipid-20% suspension), and distilled water. Figure 1 depicts the sample preparation sequence.

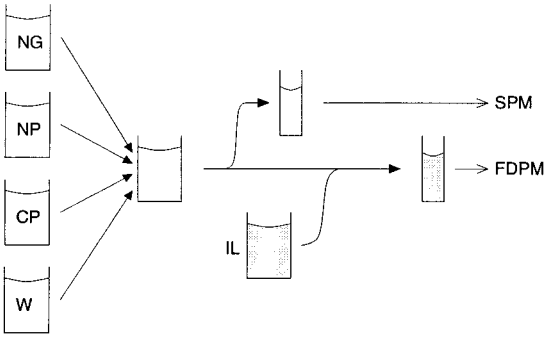

Fig. 1. Sample preparation sequence for creation of 40 mixtures with varying μ_a and μ_s' values. NG, nigrosin; NP, naphthol; CP, CPTA; W, water; IL, intralipid; SPM, spectrophotometer.

Two samples were created for each combination of dyes: a turbid one for FDPM analysis and a nonturbid one (lacking Intralipid) for spectrophotometer analysis.

The mixtures were designed to have uncorrelated values of μ_a and μ_s'; otherwise, information about one parameter could spuriously improve the chemometric prediction of the other. For the same reason, the relative concentrations of the three dyes were chosen to vary randomly.[16] Figure 2 plots μ_a versus μ_s' reference values at 674 nm for the 40 samples and shows that there is no correlation between these parameters. As shown, the μ_a values ranged from 0.009–0.09 mm^{-1} and the μ_s' values from 0.5–2 mm^{-1}. This corresponded to sample albedos $[\mu_s'/(\mu_a + \mu_s')]$ ranging from 0.930 to 0.999.

B. Spectrophotometer Data

Absorption spectra from 600–1000 nm were measured for the three stock dye solutions and also for the 40 nonturbid dye mixtures just described. Spectra were obtained on a Beckman DU-650 single-beam spectrophotometer with 1-cm pathlength cuvettes and distilled water used as a reference. These spectra were used to calculate μ_a values at the seven laser

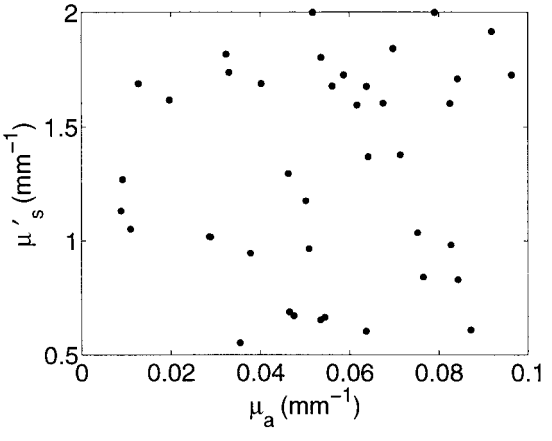

Fig. 2. Plot of μ_a versus μ_s' values for the 40 samples in this study, showing the lack of correlation between the two optical parameters.

wavelengths used by the FDPM instrument (674, 782, 803, 849, 894, 947, and 956 nm), as described below.

First, values of μ_a at 674 nm (selected arbitrarily from among the seven laser wavelengths) were determined for each stock solution from the measured absorbance A. Because the stock solutions were too dense to measure absorbances directly, portions of them were diluted with distilled water by factors of either 32 or 48. Calculation of μ_a from A came from the relation $I/I_o = \exp(-\mu_a p) = 10^{-A}$, where I/I_o is the ratio of output-to-input intensities traversing a cuvette of internal pathlength p. With $p = 10$ mm, the calculation of μ_a from A therefore becomes

$$\mu_a = \frac{A \ln(10)}{10 \text{ mm}}. \tag{1}$$

Values for the diluted solutions were multiplied by 32 or 48 as appropriate to generate $\mu_a^{(674)}$ for the nondiluted stock solutions. Linear combinations of these values were then used to predict μ_a for the 40 mixtures by means of

$$\mu_a^{(674)} = \sum_{d=1}^{3} (\mu_a)_d^{674} \frac{V_d}{V_{\text{tot}}}, \tag{2}$$

where d is an index over the three dyes and (V_d/V_{tot}) is the volume fraction of stock containing dye d used in making the sample. We computed the μ_a values at the other six laser wavelengths λ by scaling $\mu_a^{(674)}$ to the relative absorbances at those wavelengths:

$$\mu_a^{(\lambda)} = \mu_a^{(674)} \frac{A^{(\lambda)}}{A^{(674)}}. \tag{3}$$

These μ_a values agreed to within 1% with values calculated directly from the samples' absorbances with Eq. (1), indicating that the chromophores interacted minimally in scattering-free solutions. Final values for μ_a were obtained by inclusion of the absorption due to water,[17] which was absent in the spectrophotometer data because of the water in the reference cell.

Reference values for $\mu_s'^{(\lambda)}$ were calculated for each turbid sample on the basis of the concentration of Intralipid.[18] The set of reference values associated with each sample therefore contained μ_a and μ_s' at each wavelength of interest, along with the concentrations of nigrosin, naphthol, and CPTA.

C. Frequency-Domain Photon Migration Data

FDPM data from the 40 samples were obtained with the optical-fiber-based infinite geometry shown in Fig. 3. Amplitude-modulated laser light was generated and detected with instrumentation described elsewhere[8]; briefly, the system includes diode lasers, a network analyzer for rf signal generation and detection, and an avalanche photodiode for conversion of the detected light into a rf electrical signal. Measured quantities are the phase shift, ϕ, and the ac amplitude demodulation, A, of the avalanche photodiode current relative to the rf signal that modulates the laser output. The two optical fibers (source and detection) were placed several centimeters deep within the phantoms to create approximately infinite boundary conditions. The source–detector separation distance r was varied by means of micrometer control. Measurements of ϕ and A were acquired at 20 modulation frequencies f between 50 and 572.5 MHz, in increments of 27.5 MHz, at $r = 6$ and $r = 10$ mm. The accuracies of the phase and the amplitude measurements were determined to be $\pm 0.30°$ and $\pm 3.5\%$, respectively.[19]

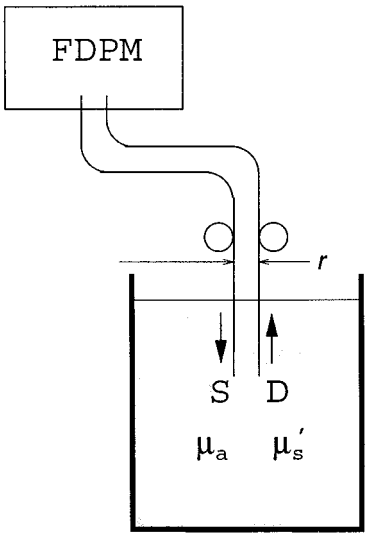

Fig. 3. Infinite geometry for FDPM data acquisition in phantoms.

D. Data-Analysis Techniques

1. *Removal of Instrumental Effects*

According to diffusion theory, the measured phase shift, ϕ, and ac amplitude, A, at source–detector separation r in an infinite geometry such as Fig. 3 are related to μ_a and μ_s' by the following expressions[2]:

$$\phi = k_i r + \phi_I, \tag{4}$$

$$A = C_I \frac{\exp(-k_r r)}{r}, \tag{5}$$

where ϕ_I is the instrumental phase offset, C_I is an instrument-dependent multiplicative constant affecting the amplitude, the subscript I indicates terms that do not depend on r, and the wave numbers k_r and k_i are the following functions of ω, μ_a, and μ_s'

$$k_r = \left[\frac{3}{2}\mu_a(\mu_a + \mu_s')\right]^{1/2}\left\{\left[1 + \left(\frac{\omega}{\mu_a c}\right)^2\right]^{1/2} + 1\right\}^{1/2}, \tag{6}$$

$$k_i = \left[\frac{3}{2}\mu_a(\mu_a + \mu_s')\right]^{1/2}\left\{\left[1 + \left(\frac{\omega}{\mu_a c}\right)^2\right]^{1/2} - 1\right\}^{1/2}, \tag{7}$$

where ω is the modulation frequency in radians ($\omega = 2\pi f$) and c is the speed of light in the medium.

One method of removing the instrumental effects is

to perform measurements at two different values of r while holding all other parameters fixed, thus obtaining the measured quantities $\phi^{(1)}$, $\phi^{(2)}$, $A^{(1)}$, and $A^{(2)}$, where the superscript here indicates that $r_1 = 6$ mm or $r_2 = 10$ mm. By transforming variables to

$$\Phi \equiv \phi^{(1)} - \phi^{(2)} = k_i(r_1 - r_2), \tag{8}$$

$$\mathcal{A} \equiv A^{(1)}/A^{(2)} = \frac{\exp[-k_r(r_1 - r_2)]}{r_1/r_2} \tag{9}$$

the new variables Φ and $\mathcal{A}$ depend solely, to within the validity of Eqs. (4) and (5), on the optical properties at the measurement wavelength (through k_r and k_i) and the measurement parameters ω, r_1, and r_2, with instrumental artifacts suppressed.

We note for later use that Eqs. (8) and (9) can be rewritten to express k_r and k_i in terms of the measured quantities r, Φ, and $\mathcal{A}$,

$$k_r = \frac{-\ln\left(\frac{r_1}{r_2}\mathcal{A}\right)}{r_1 - r_2}, \tag{10}$$

$$k_i = \frac{\Phi}{r_1 - r_2}, \tag{11}$$

and that Eqs. (6) and (7) can be inverted to express μ_a and μ_s' in terms of ω, k_r, and k_i,

$$\mu_a = \left(\frac{\omega}{2c}\right)\frac{k_r^2 - k_i^2}{k_r k_i}, \tag{12}$$

$$\mu_s' = \frac{k_r^2 - k_i^2}{3\mu_a} - \mu_a. \tag{13}$$

2. Prediction of Optical Coefficients

Diffusion approach. One way to estimate μ_a and μ_s' from FDPM data is to compare the experimental curves of Φ and $\mathcal{A}$ versus ω to the diffusion theory predictions of Eqs. (8) and (9) for different values of μ_a and μ_s'. Selecting the optimal fit is an iterative, nonlinear procedure. In the study described here, a Levenberg–Marquardt fitting criterion was used to achieve a simultaneous fit to both Φ and $\mathcal{A}$.[8] Data from each sample were fit twice, with different random starting guesses for the optical properties, and in all cases both fits converged to the same final values for μ_a and μ_s'. This process was repeated for data at each laser wavelength, so the products of the analysis were seven-wavelength μ_a and μ_s' spectra.

Chemometric approach. The central tool for the previous approach is a diffusion model that provides an explicit link between the FDPM measurements and the parameters μ_a and μ_s'. By contrast, in a chemometric approach, the emphasis is placed on empirical correlations, with no adherence to any *a priori* model. The correlations are found by analysis of FDPM data from a training set of representative samples whose optical properties are already known. After this calibration step, the optical properties of other samples can be predicted from FDPM data alone, just as can be done with a theoretical model.

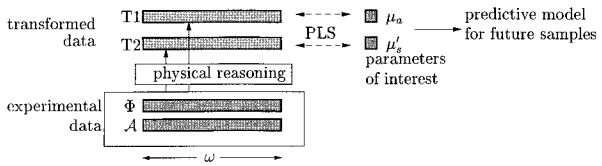

Fig. 4. Logic of transforming original FDPM data so that it can be analyzed by PLS.

For this study, the chemometric method of partial least-squares (PLS) method was chosen. PLS has found widespread use in the field of analytical chemistry in the past 15 years and has been applied to problems of biomedical interest; a recent, comprehensive review of its uses is provided as a reference.[20] In essence, PLS derives a basis set of near-orthogonal functions that, when linearly superposed, accurately fit the measured responses of the training samples. Regression of the fitting coefficients against a target parameter (*e.g.*, optical constant or concentration) yields a linear formula that can be used to predict the value of that same parameter in other samples. The rank (number of functions) of the model is usually chosen to be as small as possible while still providing a low root-mean-squared error of prediction (RMSEP). The rank should approximately equal the number of independent parameters affecting the measured data (in this case, the number of dyes present). Several excellent tutorials and introductions to PLS can be found in the literature.[21–23] To the extent that it tries to fit the measured FDPM data closely, PLS is similar to the *a priori* diffusion approach mentioned above. However, in the case of PLS the *a priori* information comes instead from the training set, which gives PLS an inherent robustness against background and instrumental fluctuations that might not be described in a theoretical model. PLS also uses an explicit formula rather than an iterative process to predict the target parameter.

Because $\Phi(\omega)$ and $\mathcal{A}(\omega)$ vary nonlinearly with changes in μ_a and μ_s' [cf. Eqs. (8) and (9)], PLS would be a poorly suited tool for modeling these relations directly. Although PLS can be used as a black box to analyze nonlinear systems with some success,[22] its robustness is unreliable. If the dependence on μ_a and μ_s' is at least partially understood, one should use this physical information to motivate a mathematical transformation of the experimental data, specifically a transformation that creates linear relations to the parameter(s) of interest. This approach is depicted conceptually in Fig. 4. We stress that this is a general approach; it does not require a complete mathematical description of the diffusion process, as is available in the test case presented here. The great merit of PLS is its ability to compensate for incomplete knowledge of a system by modeling the effects of interferents. As long as the data can be transformed such that their dependence on the target parameter becomes (approximately) lin-

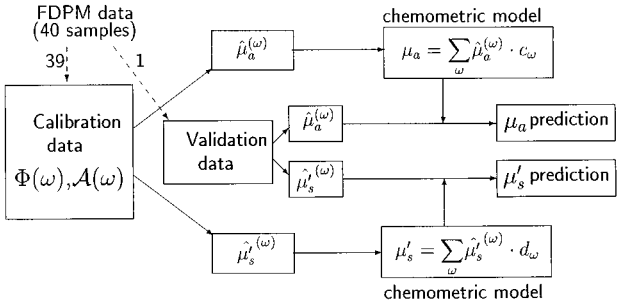

Fig. 5. Data-processing sequence for chemometric prediction of μ_a and μ_s' from FDPM data.

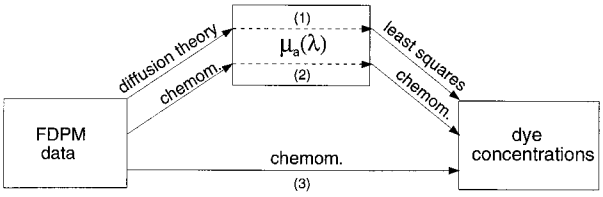

Fig. 6. Schematic of different ways to compute dye concentrations from FDPM data.

ear, PLS can model the effects of complicated boundaries, instrumental artifacts, and other interferents.

For the measurements presented here, a very accurate linearization is available, because of the simple geometry and the samples' homogeneity. One can use Eqs. (12) and (13) to compute estimates, at each modulation frequency, of μ_a and μ_s'. The FDPM measurements can thus be transformed into two data vectors containing estimates of μ_a and μ_s'. Naturally, these vectors have a very linear dependence on the actual values of μ_a and μ_s', degraded only by measurement errors. We shall call these vectors $\hat{\mu}_a^{(\omega)}$ and $\hat{\mu}_s'^{(\omega)}$ to highlight this relation. We note, however, that this transformation (and the steps to follow) can also be performed on data from more complicated measurements. For instance, if the boundary conditions were not infinite, then Eqs. (12) and (13) would generate estimates of μ_a and μ_s' convolved with boundary effects. PLS analysis would still be feasible in such a case, for the reasons suggested above.

After the FDPM data have been linearized as best as possible, PLS analysis is performed to predict μ_a and μ_s'. For this study's limited data set to be used most efficiently, the results reported below were generated by leave-one-out cross validation; *i.e.*, each sample's properties were predicted with the other 39 as a training set.[16] A full iteration of the cross-validation procedure, using the manipulations just described, is summarized in the flow chart of Fig. 5. Unlike the diffusion approach, which produces simultaneous estimates of μ_a and μ_s', the chemometric approach generates separate predictions of the two parameters. In addition, we note that reference μ_s' values are not needed to generate μ_a predictions and vice versa.

3. Prediction of Dye Concentrations

Three methods of predicting dye concentrations from the FDPM data were explored; their schematic differences are highlighted in Fig. 6 and discussed below. As indicated, the first two techniques make use of the $\mu_a(\lambda)$ spectra that were derived in the Subsection 2.D.2, whereas the third approach bypasses μ_a calculations.

Diffusion or least-squares approach. The conventional method for calculating the concentration of one absorber from a sample's absorption spectrum has been least-squares fitting.[8] In this approach, the sample's spectrum is modeled with a linear combination of the absorption spectra of the individual chromophores with weightings that minimize the difference between the measured and the predicted spectra. By combining this approach with diffusion theory, concentration predictions are extracted from the FDPM data of the sample completely *a priori*, without recourse to a training set. This method was used to extract concentration predictions of nigrosin, naphthol, and CPTA from the seven-wavelength μ_a spectra provided by diffusion theory. Unit absorption spectra of the three dyes were obtained with the spectrophotometer, and absorption values for water were taken from the literature.[17]

Chemometric approach with μ_a. Dye concentrations can also be linearly extracted from the μ_a spectrum of a sample by use of chemometrics. PLS was again selected as the method. Combined with the chemometric prediction of μ_a described above, this provides a two-step chemometric method of converting FDPM data to dye concentration predictions, as shown in the middle path of Fig. 6. Predictions are no longer *a priori*, as they depend on the existence of a training set. However, the concentration of each dye is predicted independently; *e.g.*, no reference data on CPTA or naphthol is needed to predict nigrosin. This is a potentially important advantage over least-squares fitting, which requires reference data on all absorbers even if a prediction is desired for only one of them.

Chemometric approach with FDPM data. Chemometrics offers a second, substantially different method of processing the original FDPM data to give estimates of dye concentration, one in which the middle step of calculating μ_a is bypassed (see path 3 in Fig. 6). Instead, a PLS regression is performed directly between the dye concentrations and the FDPM data transformed into $\hat{\mu}_a^{(\omega)}$ values (computed at all wavelengths as well as at all modulation frequencies).

It is important to compare the experimental constraints imposed by the three data-analysis methods described above. For the diffusion or least-squares approach, the measurement geometry needs to be fairly simple, and the spectrum of each significant absorber in the system must be known. If the chemical content of the system is known incompletely, and if there is an additional unknown absorber not described by the spectral model, subsequent calcula-

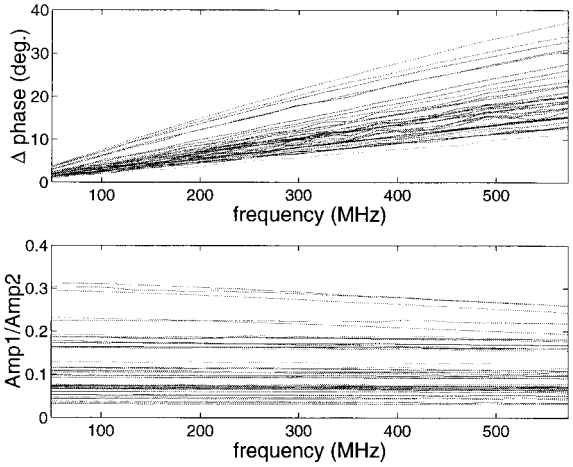

Fig. 7. Typical FDPM data for the samples used in this study. All data were taken with the 674-nm diode laser at modulation frequencies between 50 and 600 MHz. Top, phase difference between $r = 10$ and $r = 6$ mm [see Eq. (8)]. Bottom, amplitude ratio between $r = 10$ and $r = 6$ mm [see Eq. (9)].

tions may be grossly affected. For the two-step chemometric analysis, the constraints on measurement geometry and absorption line shapes are lifted; instead, reference values for μ_a and for one dye's concentration must be known for each sample in a training set. For the one-step chemometric analysis, the constraints are reduced still further, as only the dye concentration must be known. Both chemometric models are inherently robust against incomplete knowledge of the full list of absorbers, since only one chemical is considered at a time.

3. Results and Discussion

A. Data

Plots of the instrument-corrected variables, Φ and $\mathcal{A}$, for the first ten samples scanned, are shown in Fig. 7. Although an uncorrected phase measurement include shifts of as much as 10^4 degrees owing to optical and electronic delays, the two-distance calculation of Φ suppresses this instrumental factor and emphasizes sample-dependent effects. The high sensitivity of FDPM measurements to intersample variations in optical properties can be observed directly in the instrumental data.

B. Predictions of μ_a and μ_s'

The results of predicting μ_a at 674 nm with diffusion fitting and chemometric modeling are shown in Fig. 8 in which the predictions are plotted versus the reference values. As the plots show, the linearity of the predictions is high in both cases (r^2 is 0.98 for diffusion and 0.99 for chemometrics). The diffusion model predictions are systematically low, with a least-squares estimated slope of 0.79 relative to the reference values, whereas the chemometrically predicted μ_a values inherently fall along the line of unit slope.

Very few loading vectors were needed for the PLS

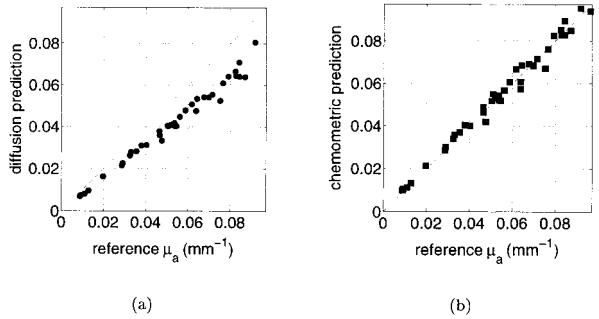

Fig. 8. Predictions of μ_a values at 674 nm from FDPM data of 40 turbid mixtures. (a) Plot obtained with the best fit to the phase and amplitude data for each sample independently. (b) Plot derived with chemometric cross validation as described in the text. The r^2 values for the plots are 0.98 for the top and 0.99 for the bottom, respectively.

calibrations because, as expected, the $\hat{\mu}_a^{(\omega)}$ estimates strongly resembled the actual μ_a values against which they were regressed, as shown in Fig. 9. The five shorter wavelengths between 674 and 849 nm required only a single loading vector to achieve optimal prediction of μ_a and thus indicated no competing effects. At the higher wavelengths (947 and 956 nm), a few additional loading vectors (1 extra at 947, 3 extra at 956) were needed because the signal level dropped lower, and the instrumental background effects became nonnegligible.

The associated predictions of μ_s' at 674 nm are

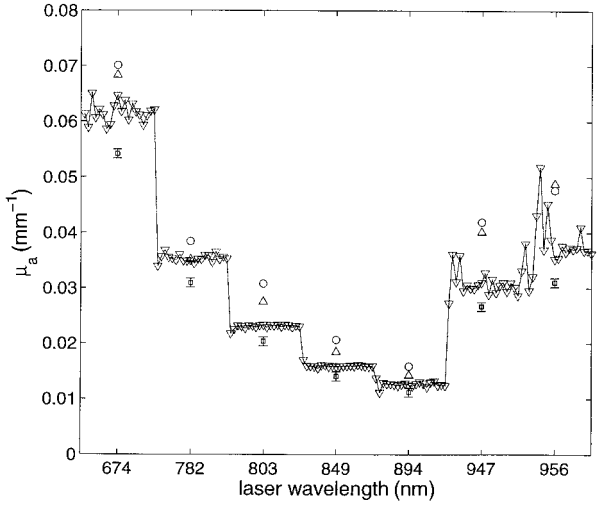

Fig. 9. Comparison of data vectors analyzed for extraction of dye concentrations from a typical sample. Note that the x axis is not linear with respect to wavelength. Circles represent reference μ_a values as determined by spectrophotometer. Squares with error bars represent μ_a predictions from diffusion theory analysis of FDPM data. Upwards triangles represent μ_a predictions from chemometric analysis of FDPM data. Connected downwards triangles represent $\hat{\mu}_a^{(\omega)}$ vector created from FDPM data. Note that the $\hat{\mu}_a^{(\omega)}$ vector contains 20 data points plotted per laser wavelength, corresponding to frequencies from 50 to 600 MHz. Error bars on the diffusion predictions are $\pm 8 \times 10^{-4}$ mm^{-1}, as measured in an earlier study.[19]

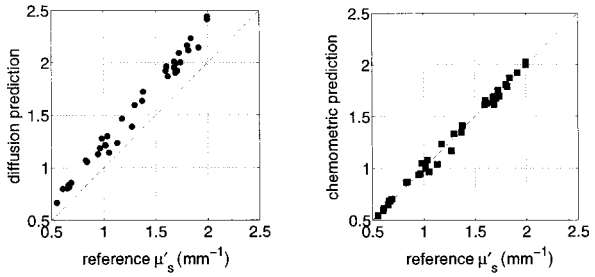

Fig. 10. Predictions of μ_s' values at 674 nm from FDPM data of 40 turbid mixtures. (a) Plot obtained with the best fit to the phase and amplitude data for each sample independently. (b) Plot derived from the chemometric cross validation as described in the text. The r^2 value for both plots is 0.99.

shown in Fig. 10. Once again, both plots are highly linear (r^2 of 0.99 for both). The PLS models required, in general, one more loading vector than for absorption, with total numbers ranging from two to four for optimal predictions. With scattering, the diffusion model overpredicted the μ_s' values by a factor of 1.14, whereas it underpredicted μ_a by nearly the same amount; this is a consistent pattern at all wavelengths, as evidenced in Table 1. This suggests that the diffusion model interpreted some absorption losses as scattering losses. However, the discrepancy could also be because of a systematic error in the preparation of the samples, and in any case the essential feature to be compared between these plots is the linearity, as described by the r^2 value. As these plots show, by use of data gathered in an infinite geometry, the PLS technique can extract optical parameters from FDPM data with an accuracy comparable with that of the diffusion theory.

C. Prediction of Dye Concentrations

Figure 9 displays examples of input data for the three different methods of concentration prediction. The squares (with error bars) are a μ_a spectrum estimated from FDPM data with the diffusion theory; the upward triangles, the μ_a spectrum estimated by use of chemometric analysis; and the solid line, the same FDPM data converted into $\hat{\mu}_a^{(\omega)}$ values computed as

Table 1. Slopes of Diffusion-Predicted μ_a and μ_s' Values at Each Wavelength in the Study[a]

λ (nm)	μ_a Slope	μ_s' Slope	Mean Slope
674	0.79	1.14	0.97
782	0.88	1.14	1.01
803	0.77	1.17	0.97
849	0.81	1.17	0.99
894	0.82	1.17	1.00
947	0.74	1.24	0.99
956	0.81	1.22	1.02

[a]The consistent overprediction and underprediction in nearly equal amounts at each wavelength suggests that the diffusion model ascribed some of the absorption losses to scattering.

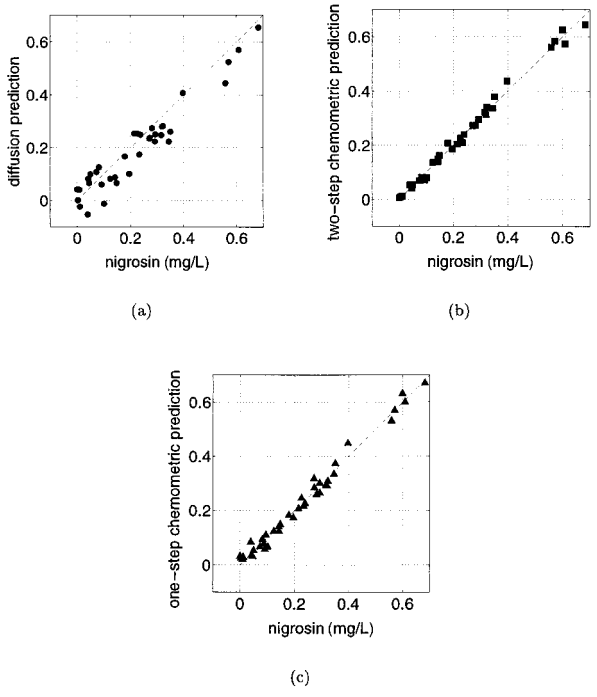

Fig. 11. Predictions of nigrosin concentrations from the FDPM data.

described in Section 2.D.2. For reference, μ_a values calculated from spectrophotometer data as described in Section 2.A are also shown (circles). The error bars on the diffusion theory calculations indicate reproducibility and are due to system noise and the stability of the fitting procedure[19]; the other values plotted have errors of similar magnitudes.

The results of the three different methods of predicting the nigrosin dye concentrations are shown in Fig. 11, with the corresponding results for naphthol and CPTA appearing in Figs. 12 and 13. Each dye or method combination has an associated RMSEP value that represents the typical error in the predictions; these nine values are presented graphically in Fig. 14. As Fig. 14 shows, the chemometric methods (center and right groups in the plot) produce RMSEP values that are consistently as low or lower than the diffusion–least-squares method (left group). This result shows that in situations in which the FDPM data are gathered in an infinite geometry and the diffusion theory is therefore readily applied chemometrics can predict chromophore concentrations as accurately as conventional diffusion analysis and least-squares curve fitting.

In addition, note that the two different chemometric methods of predicting dye concentration generate very similar RMSEP values. This indicates that it is unnecessary to use the two-step method in which μ_a is first predicted; one can proceed directly from FDPM data to dye concentrations without substantial loss of prediction accuracy.

As mentioned in the Subsection 2.D.2, the key advantages of the one-step chemometric method are that it requires neither a simple geometry nor a ref-

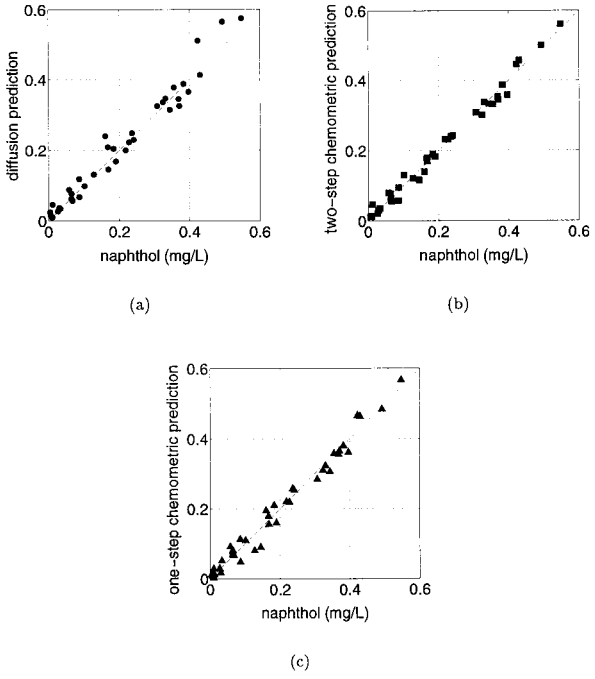

Fig. 12. Predictions of naphthol concentrations from the FDPM data.

erence measurement of μ_a. This second condition provides experimental convenience, as μ_a is usually difficult to establish for turbid samples. These advantages do not, of course, ensure that the method predicts concentrations from FDPM data successfully. As noted above, when the measurement is not

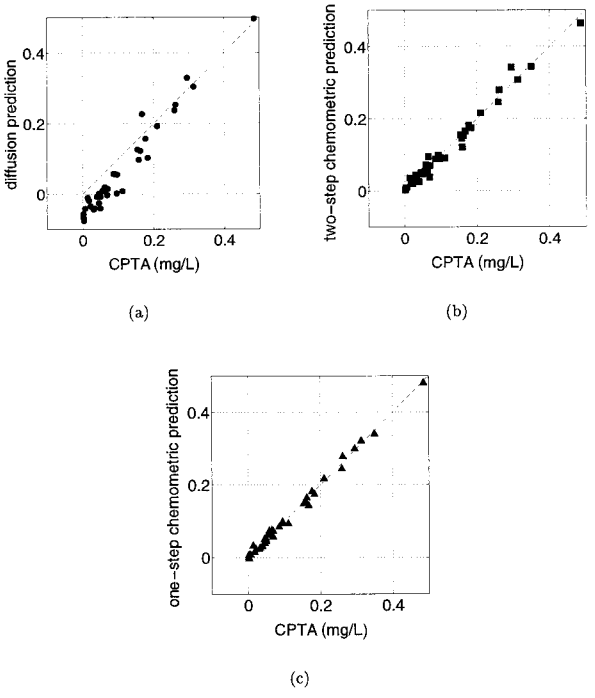

Fig. 13. Predictions of CPTA concentrations from the FDPM data.

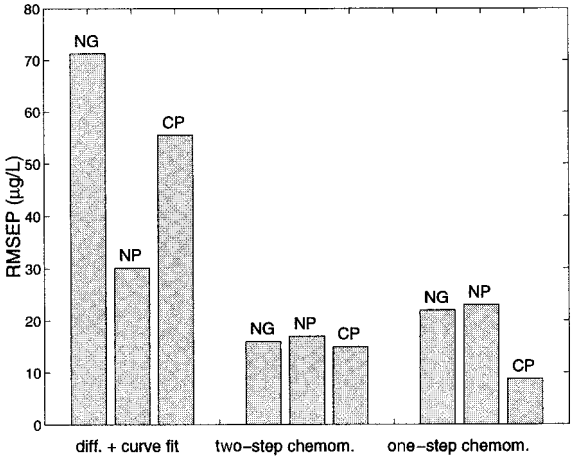

Fig. 14. Summary of dye concentration predictions with the three strategies depicted in Fig. 6.

conducted in an infinite geometry, $\hat{\mu}_a^{(\omega)}$ as given by Eq. (12) is no longer a good estimate of μ_a. However, if the perturbation introduced by more complicated measurement geometries is not too great, some linearity between μ_a and the perturbed $\hat{\mu}_a^{(\omega)}$ should be preserved. In such cases, PLS can exploit this relation and extract concentration predictions. Further studies are needed to identify which measurement geometries can be handled robustly by the chemometric methods suggested here. In particular, geometries that are experimentally practical but mathematically ill defined should be investigated.

It will be important to investigate measurement schemes in which the amount of input FDPM data is greatly reduced. In the calculations presented here, 20 modulation frequencies and seven laser wavelengths were used at each of two source–detector separations. Additional analysis is needed to establish the accuracy of predicting μ_a, μ_s', and chromophore concentration with fewer modulation frequencies, source–detector positions, and optical wavelengths. It may even be possible to deduce chromophore concentrations chemometrically even when the amount of input data is fundamentally insufficient to calculate μ_a. Reducing the amount of necessary input data could lead to significantly simpler instrumentation. These investigations will form the basis of further studies.

4. Conclusions

A general framework for processing FDPM data with chemometrics has been developed. The approach is to gather both phase and amplitude data, construct new variables that vary linearly with a property of interest, and perform regressions such as PLS to obtain correlations and to correct for interfering signals, including instrumental artifacts. Physical insight (such as that provided by diffusion theory for simplified cases) should be used to motivate the change of variables. This study has shown that chemometric analysis can predict optical and chemical properties

with an accuracy comparable with or superior to that of diffusion-based analysis in an infinite geometry.

The results suggest that chemometric, training-set-based methods may broaden the range of techniques employed to analyze multifrequency FDPM data in two senses. First, by providing an alternative means of converting FDPM data to μ_a and μ_s' values, chemometrics can analyze data without requiring fits to a diffusion model. This approach should better tolerate complicated boundary conditions, short source–detector separations, and other instances for which the diffusion approximation becomes invalid. In addition, chemometrics frees the user from viewing FDPM data solely in terms of the embedded μ_a and μ_s' values. Instead, the FDPM data vector itself can be treated as a source of information from which predictions are directly extracted. This capitalizes on the exquisite sensitivity of overdamped photon density waves to optical loss mechanisms without requiring an exact physical model of the measurement geometry. Analysis can be performed even when the full list of significant absorbers is unknown. These insights may prove useful in various future FDPM applications, ranging from *in vitro* studies to noninvasive *in vivo* measurements. Finally, chemometric techniques may be helpful in reducing the needed number of FDPM measurements per sample, thus leading to simpler instrumentation.

This research was conducted at the Beckman Laser Institute at the University of California, Irvine, with support from the National Institutes of Health (NIH) Laser Microbeam and Medical Program (grant RR-01192), NIH grant R29-GM50958, Department of Energy grant DE-FG03-91ER61227), and Office of Naval Research grant N00014-91-C-0134. A. J. Berger gratefully acknowledges postdoctoral support from the George E. Hewitt Foundation for Medical Research.

References

1. S. Fantini, M. A. Franceschini, J. B. Fishkin, B. Barbieri, and E. Gratton, "Quantitative determination of the absorption spectra of chromophores in strongly scattering media: a light-emitting-diode based technique," Appl. Opt. **33,** 5204–5213 (1994).
2. B. J. Tromberg, L. O. Svaasand, T.-T. Tsay, and R. C. Haskell, "Properties of photon density waves in multiple-scattering media," Appl. Opt. **32,** 607–616 (1993).
3. D. A. Boas, M. A. O'Leary, B. Chance, and A. G. Yodh, "Scattering and wavelength transduction of diffuse photon density waves," Phys. Rev. E **47,** R2999–R3002 (1993).
4. B. W. Pogue and M. S. Patterson, "Frequency-domain optical absorption spectroscopy of finite tissue volumes using diffusion theory," Phys. Med. Biol. **39,** 1157–1180 (1994).
5. E. M. Sevick, B. Chance, J. Leigh, S. Nioka, and M. Maris, "Quantitation of time- and frequency-resolved optical spectra for the determination of tissue oxygenation," Anal. Biochem. **195,** 330–351 (1991).
6. B. W. Pogue, M. Testorf, T. McBride, U. Österberg, and K. Paulsen, "Instrumentation and design of a frequency-domain diffuse optical tomography imager for breast cancer detection," Opt. Express **1,** 391–403 (1997).
7. S. Fantini, S. A. Walker, M. A. Franceschini, M. Kaschle, P. M. Schlag, and K. T. Moesta, "Assessment of the size, position, and optical properties of breast tumors *in vivo* by noninvasive optical methods," Appl. Opt. **37,** 1982–1988 (1998).
8. B. J. Tromberg, O. Coquoz, J. B. Fishkin, T. Pham, E. R. Anderson, J. Butler, M. Cahn, J. D. Gross, V. Venugopalan, and D. Pham, "Non-invasive measurements of breast tissue optical properties using frequency-domain photon migration," Philos. Trans. R. Soc. Lond. **352,** 661–668 (1997).
9. L. O. Svaasand, T. Spott, J. B. Fishkin, T. Pham, B. J. Tromberg, and M. W. Berns, "Reflectance measurements of layered media with diffuse photon-density waves: a potential tool for evaluating deep burns and subcutaneous lesions," Phys. Med. Biol. **44,** 801–813 (1999).
10. J. B. Fishkin, E. Gratton, M. J. Vandeven, and W. W. Mantulin, "Diffusion of intensity modulated near-IR light in turbid media," in *Time-Resolved Spectroscopy and Imaging of Tissues*, B. Chance and A. Katzir, eds., Proc. SPIE **1431,** 122–135 (1991).
11. R. C. Haskell, L. O. Svaasand, T.-T. Tsay, T.-C. Feng, M. S. McAdams, and B. J. Tromberg, "Boundary conditions for the diffusion equation in radiative transfer," J. Opt. Soc. Am. **11,** 2727–2741 (1994).
12. V. Venugopalan, J. S. You, and B. J. Tromberg, "Radiative transport in the diffusion approximation: an extension for highly absorbing media and small source-detector separations," Phys. Rev. E **58,** 2395–2407 (1998).
13. T. J. Farrell, B. C. Wilson, and M. S. Patterson, "The use of a neural network to determine tissue optical properties from spatially resolved diffuse reflectance measurements," Phys. Med. Biol. **37,** 2281–2286 (1992).
14. J. S. Dam, P. E. Andersen, T. Dalgaard, and P. E. Fabricius, "Determination of tissue optical properties from diffuse reflectance profiles by multivariate calibration," Appl. Opt. **37,** 772–778 (1998).
15. A. Kienle, L. Lilge, M. S. Patterson, R. Hibst, R. Steiner, and B. C. Wilson, "Spatially resolved absolute diffuse reflectance measurements for noninvasive determination of the optical scattering and absorption coefficients of biological tissue," Appl. Opt. **35,** 2304–2314 (1996).
16. E. V. Thomas, "A primer on multivariate calibration," Anal. Chem. **66,** 795A–804A (1994).
17. G. M. Hale and M. R. Querry, "Optical constants of water in the 200-nm to 200-μm wavelength region," Appl. Opt. **12,** 555–63 (1973).
18. H. J. van Staveren, C. J. M. Moes, J. van Marle, S. A. Prahl, and M. J. C. van Gemert, "Light scattering in Intralipid-10% in the wavelength range of 400–1100 nm," Appl. Opt. **30,** 4507–4514 (1991).
19. T. H. Pham, O. Coquoz, J. B. Fishkin, E. Anderson, and B. J. Tromberg are preparing a manuscript to be called, "A broad bandwidth frequency domain instrument for quantitative tissue optical spectroscopy."
20. S. D. Brown, S. T. Sum, F. Despagne, and B. K. Lavine, "Chemometrics," Anal. Chem. **68,** R21–R61 (1996).
21. P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," Anal. Chim. Acta **185,** 1–17 (1986).
22. D. M. Haaland and E. V. Thomas, "Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information," Anal. Chem. **60,** 1193–1202 (1988).
23. H. Martens and T. Næs, *Multivariate Calibration* (Wiley, New York, 1989).