# Quantifying the Optical Properties and Chromophore Concentrations of Turbid Media by Chemometric Analysis of Hyperspectral Diffuse Reflectance Data Collected Using a Fourier Interferometric Imaging System

**TUAN H. PHAM, CHARLOTTA EKER, ANTHONY DURKIN, BRUCE J. TROMBERG, and STEFAN ANDERSSON-ENGELS***

*Department of Physics, Lund Institute of Technology, P.O. Box 118, Lund, SE-22100, Sweden (T.H.P., C.E., S.A.-E.); Beckman Laser Institute and Medical Clinic, Laser Microbeam and Medical Program (LAMMP), Irvine, California 92612 (T.H.P., B.J.T.); and Candela Corporation, Wayland, Massachusetts 01778 (A.D.)*

**A non-contact Fourier transform interferometric imaging system was used to collect hyperspectral images of the steady-state diffuse reflectance from a point source in turbid media for the spectral range of 550–850 nm. Steady-state diffuse reflectance profiles were generated from the hyperspectral images, and partial least-squares (PLS) regression was performed on the diffuse reflectance profiles to quantify absorption ($\mu_a$) and reduced scattering ($\mu_s'$) properties of turbid media. The feasibility of using PLS regression to predict optical properties was examined for two different sets of spatially-resolved diffuse reflectance data. One set of data was collected from 40 turbid phantoms, while the second set was generated by convolving Monte Carlo simulations with the instrument response of the imaging system. Study results show that PLS prediction of $\mu_a$ and $\mu_s'$ was accurate to within ±8% and ±5%, respectively, when the model was trained on turbid phantom data. Moreover, PLS prediction of optical properties was considerably faster and more efficient than direct least-squares fitting of spatially-resolved profiles. When the PLS model was trained on Monte Carlo simulated data and subsequently used to predict $\mu_a$ and $\mu_s'$ from the diffuse reflectance of turbid phantom, the percent accuracies degraded to ±12% and ±5%, respectively. These accuracy values are applicable to homogenous, semi–infinite turbid phantoms with optical property ranges comparable to tissues.**

Index Headings: **Partial least squares; PLS; Turbid media; Diffuse reflectance; Optical properties; Multispectral imaging.**

## INTRODUCTION

The absorption and reduced scattering parameters ($\mu_a$ and $\mu_s'$, respectively) of turbid media, such as tissue, can be used to characterize its composition and structure. For instance, $\mu_a$ and $\mu_s'$ of tissue can provide information on a variety of physiological processes. Wavelength-dependent absorption is used to quantify the concentration of biologically important chromophores, such as hemoglobin, myoglobin, water, fat, and near-infrared absorbing drugs.[1–6] Wavelength-dependent scattering properties offer insight into the composition, density, and organization of tissue structures, such as cells, subcellular organelles, and connective tissue/extracellular matrix.[7–11] Since changes in these components generally accompany pathologic transformations and physiologic processes, techniques for non-invasively quantifying $\mu_a$ and $\mu_s'$ *in vivo* have generated intense interest.

One approach for characterizing $\mu_a$ and $\mu_s'$ of turbid media is to use spatially resolved, steady-state diffuse reflectance. Two strategies for spatially resolved reflectance measurements of turbid, semi-infinite media have been demonstrated: contact probe detection using fiber-optic arrays[12–19] and image reflectometry.[20–23] Systems based on contact probe detection often have the capacity to measure the reflectance for a continuous range of wavelengths but require that the detection fibers touch the tissue. Imaging reflectometry has an advantage in that measurements can be performed remotely. Noncontact detection can be particularly useful in clinical settings where fiber-optic probes may contribute to a number of problems, including contamination of sterile fields and alteration of regional tissue perfusion (via probe-induced pressure).

Imaging reflectometry systems often have limited spectral range, typically consisting of only a few discrete wavelengths.[20,21,24] For many tissue applications, broad spectral range is desirable. This is due to the fact that detailed absorption and scattering spectra can be used to determine the physiologic and structural properties of tissues and are diagnostically valuable in characterizing pathology.[25–27] Calibration procedures are generally difficult and complex because the instrument response depends not only on the spatial and wavelength performance of the CCD, but also on the point spread function of the imaging system. Consequently, techniques that extend the wavelength range and facilitate rapid, accurate calibration are of considerable interest for spectroscopy and imaging of turbid media.

In a previous study, a Fourier transform interferometric imaging system (FTIIS) was used to collect hyperspectral images of the steady-state diffuse reflectance from tissue-like phantoms containing fat emulsion scatterers and dye absorbers. We have demonstrated that Monte Carlo simulations can be fit to the image-generated diffuse reflectance profile in order to quantify optical properties over a continuous and broad spectral range.[28] However, the iterative fitting procedure is computationally intensive and therefore not practical for real-time property determination. In this study, we investigate the feasibility of using multivariate analysis, specifically, partial least-squares (PLS) regression, to quantify $\mu_a$ and $\mu_s'$ from spatially resolved diffuse reflectance data obtained from hyperspectral images. PLS regression was selected because it is a robust and effective multivariate approach.[29,30]

Received 4 August 2000; accepted 16 April 2001.
* Author to whom correspondence should be sent.


0003-7028/01/5508-1035$2.00/0

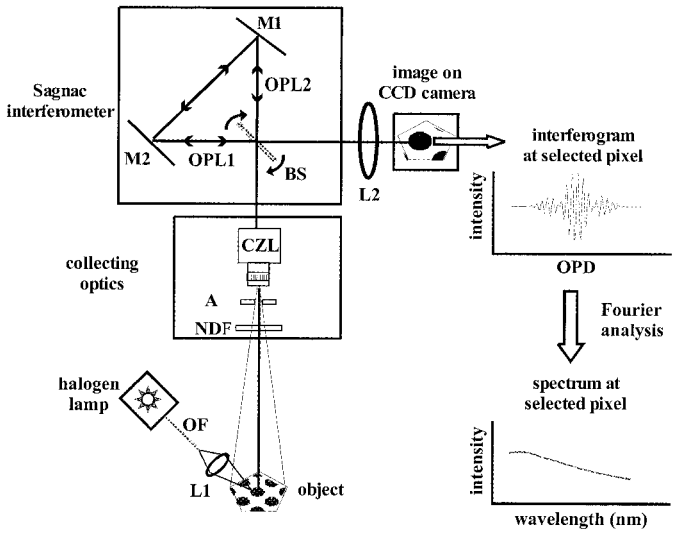



FIG. 1. Experimental setup of the Fourier transform interferometric imaging system used to collect hyperspectral images of the diffuse reflectance from turbid media. The principle parts of the system are the halogen white light source, the collection optics, a Sagnac interferometer, and a cooled 12-bit CCD camera (CCD). Interference patterns of the object are collected for a sequential series of optical path differences (OPD) by incrementally rotating the beamsplitter (BS). An interferogram is generated at each pixel, which corresponds to a particular region of the object. Inverse Fourier transform of the interferograms yields the hyperspectral images. Data acquisition and processing were performed using a PC. Abbreviations: optical fiber (OF), aperture (A), camera zoom lens (CZL), neutral density filter (NDF), lens to focus the white light source (L1), lens to image the recombined beams onto the CCD (L2), mirror (M), and optical pathlength (OPL).

Moreover, this analytical technique was recently used to successfully analyze spectra collected from scattering media.[31,32]

We examined three approaches to PLS regression and verification for two different data sets. The first data set consisted of measurements from 40 turbid phantoms. The second data set was generated by convolving Monte Carlo simulations with the instrument response of FTIIS (MCSI⊗IR). The three approaches examined were: (1) perform PLS regression and model validation on turbid phantom data (labeled as trainTPD–valTPD); (2) perform PLS regression and model validation on Monte Carlo simulated data (trainMCSI–valMCSI); and (3) perform PLS regression on simulated data and then use the model to predict optical properties from the diffuse reflectance of turbid phantom (trainMCSI–predTPD). Predicted $\mu_s'$ and $\mu_a$ spectra, in turn, were used to quantify, respectively, the fat emulsion and dye concentrations.

## MATERIAL AND METHODS

**Fourier Transform Interferometric Imaging System (FTIIS).** Figure 1 schematically illustrates the main components of the hyperspectral imaging device. White light from a broadband halogen source (LS1, Ocean Optics Inc., Dunedin, FL) is coupled into a 0.16 NA, 200 μm diameter optical fiber. Light exiting the fiber (2 mW) is re-focused to form a 250 μm diameter point source that is launched onto tissue-like phantoms at an incidence angle of 25° relative to the surface normal. At this angle, specularly reflected light is not collected by the imaging optics, and the diffuse reflectance distribution is minimally affected.[21] Hyperspectral images are collected using a Fourier transform interferometric imaging system (SpectraCube® imaging system, Applied Spectral Imaging, Migdal HaEmek 10511, Israel).

The FTIIS employs a cyclic (Sagnac) interferometer[33] to construct hyperspectral images. The principle components of the FTIIS are the collection optics, a Sagnac interferometer, and a 12-bit charge-coupled device (CCD) camera operating at −32 °C (Model TE-CCD-512-EFT/UV, Princeton Instruments, Inc., Trenton, NJ). The collection optics consist of a commercial camera zoom lens (a 50 mm focal length, $f/1.2$ to $f/16$) and a 1.5 O.D. neutral density filter (NDF). The camera is oriented perpendicular to the surface of the phantom, as shown in Fig. 1. The achromatic zoom lens allows for adjustable image magnification and acceptance angle. Neutral density filters were placed in front of the camera lens to prevent saturation of the CCD pixels that image the regions near the source where the intensity is maximal.

Light entering the collection optics is split into two beams and directed along different optical paths (OPL1 and OPL2). The two beams are then recombined at the CCD. The optical path difference (OPD) between the beams, which is a function of the beamsplitter's angular position, generates interference patterns of the object on the CCD. To obtain hyperspectral images, interference patterns are collected for sequences of OPDs, generated by step-wise rotation of the beam splitter. In this manner, an interferogram is constructed for each pixel in the image. With the use of the software that accompanies the FTIIS, Fourier analysis of the interferograms produces the wavelength spectra. The selected magnitude, range, and spacing of the OPDs determine spectral range and resolution. Images can occupy up to 512 × 512 pixels (for a given wavelength), and, depending on the selected magnification, objects with sizes on the order of 1–50 mm can be imaged. In this study, we set the FTIIS parameters so that hyperspectral images for wavelengths ranging from 550–850 nm were generated; this wavelength range corresponds to the spectral range of the FTIIS. The images obtained were unevenly spaced within the wavelength range. On average, the spacing was 15 nm. Image size was 170 × 170 pixels, corresponding to a physical (object) dimension of 25 × 25 mm. At these settings, each set of 24 hyperspectral images was acquired in 40 s.

**Turbid Phantom Data.** The turbid phantom data set consisted of the hyperspectral images of the diffuse reflectance and the expected $\mu_a$ and $\mu_s'$ values of 40 phantoms. The phantoms were prepared from a unique and randomly selected combination of three absorbing dyes (Nigrosin, Janus green B, and Prussian blue, Sigma, St. Louis, MO) and fat emulsion scatterers (Intralipid, 200 mg/mL (20%), Pharmacia & Upjohn, Sverige AB, Stockholm, Sweden). The dye and Intralipid concentrations were selected to yield optical property ranges comparable to those of tissues.[34] This produced phantoms with unique and randomly distributed $\mu_a$ and $\mu_s'$ values within the optical property range of $1.0 \times 10^{-3}$–0.20 $mm^{-1}$ for $\mu_a$ and 0.20–2.50 $mm^{-1}$ for $\mu_s'$. For each 500 mL phantom, known concentrations of aqueous dye solutions were prepared from distilled water and dyes, followed by the addition of Intralipid scatterers. Conventional absorption

spectrophotometry was performed on a small sample (2 mL) collected from the aqueous dye mixture in order to independently measure $\mu_a$. These $\mu_a$ values deviated less than ±3% from the expected phantom absorption values that were calculated from the dye concentrations and extinction coefficients.

The corresponding expected values for reduced scattering ($\mu_s'$) were calculated according to van Staveren et al.,[35] who used Mie theory to relate the scattering coefficient and anisotropy factor of Intralipid to the optical wavelength. According to that work, the scattering coefficient ($\mu_s$) and the anisotropy factor ($g$) of 10% Intralipid can be expressed as a function of the optical wavelength, $\lambda$:

$$\mu_s(\lambda) = 16\lambda^{-2.4} \quad (1)$$

$$g(\lambda) = 1.1 - 0.58\lambda \quad (2)$$

where $\lambda$ is given in micrometers and $\mu_s$ in $mm^{-1}$. Values obtained from Eqs. 1 and 2 were shown to have an accuracy of ±6% as compared to Mie theory prediction. The reduced scattering coefficient for 10% Intralipid was calculated from $\mu_s$ and $g$ using the similarity relation $\mu_s' = \mu_s(1 - g)$. For Intralipid, this can be expressed as $\mu_s' = \sigma C_{\%}$, where $\sigma = \mu_s(1 - g)/10$ and $C_{\%}$ is the Intralipid concentration in percent volume. This relationship has been shown valid for Intralipid concentrations of less than 10% by volume.[35] These expected $\mu_a$ values determined from dilute solution absorption spectrophotometry and $\mu_s'$ values deduced using the method above were used as the ''gold standard'' for calibration and validation of the PLS model.

Hyperspectral images of the diffuse reflectance were collected immediately following the preparation of the phantoms to minimize the possibility of any changes in optical properties over time. Data from the hyperspectral images were preprocessed before performing PLS modeling. The image of the diffuse reflectance at each wavelength was binned circumferentially. That is, the image was divided into concentric rings, and pixel values in each ring were summed and normalized to the pixel count of the ring. Circumferential binning of data conferred two advantages: (1) pixel values from different areas but with the same radial distance from the source were averaged, and thus noise was reduced; and (2) the computational demands of PLS regression were reduced because the diffuse reflectance data were of one spatial dimension instead of two. We performed a natural logarithmic transform of the diffuse reflectance profile in order to obtain an approximately linear relationship between the optical properties and the logarithm of the diffuse reflectance intensity. PLS regression optimally models relationships that are linear.[29,31] The natural logarithmic transform was selected because the diffuse reflectance can be approximated by the function

$$R \propto \frac{C_1 \exp(-C_2\rho)}{\rho^m} \quad (3)$$

where $C_1$, $C_2$, and $m$ are parameters dependent on the optical properties of the medium as well as the source-detector ($\rho$) separations.[18,36] The specific functional relationship between the diffuse reflectance and optical properties of the medium is well described in the literature, for instance, see Refs. 18, 21, 37, and 38.

**Monte Carlo Simulated Data.** In addition to the turbid phantom training set, we constructed another training set based on Monte Carlo simulated data. Monte Carlo simulations of light propagation in turbid media were used to generate diffuse reflectance data for homogenous, semi-infinite media. We chose to use Monte Carlo simulations,[38] instead of the diffusion approximation, because the numerical approach readily accounts for the effects of boundary, source distribution, and media geometry on the diffuse reflectance. In addition, reflectance data from Monte Carlo simulations are more accurate for highly absorbing media, as well as for data close to the source. The Henyey–Greenstein phase function,[39] which describes the probability that a photon is scattered in a particular direction, was used in the simulations. Various investigators have used this function to successfully model the scattering phase function of fat emulsions. In order to obtain two data sets of equal size, 40 unique and randomly generated pairs of $\mu_a$ and $\mu_s'$ were generated for each of the 24 wavelengths. The $\mu_a$ and $\mu_s'$ values used in the simulations encompassed optical property ranges comparable to those of the phantoms. Specifically, $\mu_a$ and $\mu_s'$ ranges were $1 \times 10^{-3}$–0.30 $mm^{-1}$ and 0.20–2.50 $mm^{-1}$, respectively. The refractive index was set to the value corresponding to water ($n = 1.33$), and the anisotropy factor was set to the value corresponding to fat emulsion ($g = 0.7$).[35] Because of the natural cylindrical symmetry, Monte Carlo data for diffuse reflectance were organized using a cylindrical coordinate system. For example, diffuse reflectance data were arranged in concentric rings surrounding the source. The radial bin thickness of the concentric rings was 50 μm. The number of photons was empirically set to $N = (\mu_a/\mu_s')^{1/4} \times 10^7$ so that the noise characteristics were similar across the data set.[21] The diffuse reflectance was recorded as a function of the radial distance from the source and stored as photon probability per unit area.

A calibration procedure was performed in order to account for the instrument response (IR) of the imaging system, including such instrumental factors as the point spread function, numerical aperture, spectral sensitivity of the CCD, and uniformity of the CCD response. The calibration procedure consisted of three main steps. First, we collected an image of a marked ruler and counted the number of pixels corresponding to a unit length on the ruler. The ratio of length-to-pixel count is the factor that relates the pixel count to the physical size of the object. Second, hyperspectral images of the diffuse reflectance were acquired from a reference phantom, which was constructed from Intralipid and dyes, in order to determine the IR at each measured wavelength. Calibration images were obtained by using the same settings as those used to perform data collection so that, for example, numerical aperture and magnification factor were identical to those used to construct hyperspectral images of the test samples. Finally, the IR of the system was calculated using the calibration images and the known optical properties of the reference phantom. For this purpose, the calibration image at each wavelength was Fourier transformed into the spatial-frequency domain (SFD). Likewise, the Monte Carlo simulated image (MCSI) corresponding to

the reference phantom optical properties was transformed into the SFD. The overall instrument response of the FTIIS at each wavelength was calculated from the spatial transforms of the calibration and MCSI via the relation

$$\mathcal{F}(IR) = \mathcal{F}(I)/\mathcal{F}(MCSI) \quad (4)$$

where $\mathcal{F}(I)$, $\mathcal{F}(MCSI)$, and $\mathcal{F}(IR)$ are the spatial Fourier transforms of the calibration image, the Monte Carlo simulated image, and the instrument response, respectively.

To construct a training set based on Monte Carlo simulations, the MCSI corresponding to a particular set of optical properties was effectively convolved with the instrument response at the evaluated wavelength, yielding MCSI⊗IR where ⊗ denotes a convolution. However, the convolution was performed indirectly in the spatial-frequency domain. MCSI was first transformed into the SFD, $\mathcal{F}(MCSI)$. The spatial-frequency transform of MCSI was multiplied by $\mathcal{F}(IR)$ to account for the instrument response. The product, $\mathcal{F}(MCSI)\mathcal{F}(IR)$, was inversely Fourier transformed into the spatial domain to yield MCSI⊗IR. The Monte Carlo training set consisting of MCSI⊗IR data was subjected to the same preprocessing scheme, (i.e., circumferential binning and logarithmic transformation) as that performed on the turbid phantom data.

**Chemometric Analysis.** Diffuse reflectance data and expected optical property spectra were organized into matrices, denoted respectively as matrices **X** and **Y** in Fig. 2a. The rows of **X** and **Y** correspond to the samples (turbid phantoms or MCSI⊗IR). The columns of **X** correspond to the logarithmic intensity profiles as a function of the radial distance at the measured wavelengths, while the optical property ($\mu_a$ or $\mu_s'$) spectra make up the columns of **Y**. Partial least-squares regression was implemented in order to develop models that predict the optical properties (**Y**) based on the diffuse reflectance measurements (**X**).

Partial least-squares regression is a widely accepted multivariate method. In this method, a linear model is used to correlate the measured variable (**X**) to the parameters of interest (**Y**).[29] A PLS model is developed empirically from the calibration data. The PLS model, in turn, is used to predict parameter **Y** from future measurements of **X**. An abundant amount of literature has been published on PLS regression.[40,41] Briefly, the basic idea of PLS regression is to decompose the matrix containing the measured variable **X** into a linear combination of near-orthogonal vectors denoted as the scores ($t_i$) and loadings ($p_i$), as illustrated in Fig. 2b. Similarly, PLS regression decomposes the **Y** matrix into two sets of basis vectors, which are denoted as the score ($u_i$) and loading ($q_i$) vectors. Basis vectors, when linearly combined, comprise the information content of matrix **X** or **Y**, while the residual matrices **E** and **F** contain the ''noise'' components of **X** and **Y**, respectively. A key feature of PLS regression is that the **X** and **Y** matrices are decomposed in a manner that optimally correlates the basis vectors of **X** and **Y**.[29] The weight matrix **W** relates the regression of **X** to **Y**, and is used with the loading matrices **P** and **Q** to predict the value of **Y** in future samples. Specifically, matrix **Y** is predicted from measured variable **X** as[29]

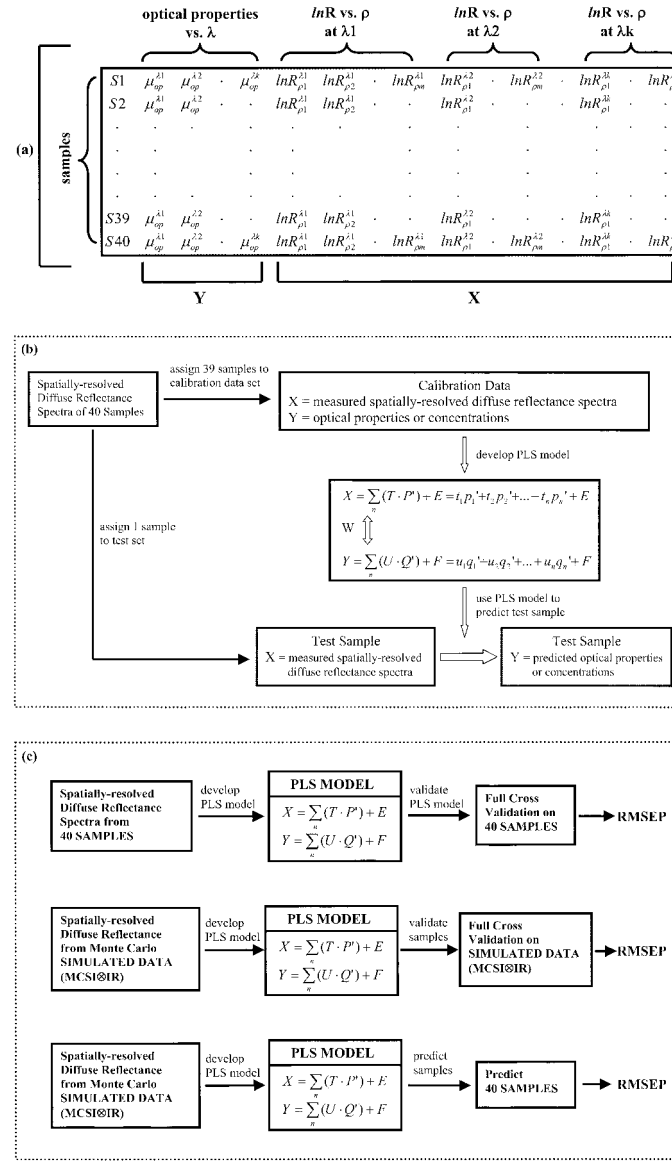



FIG. 2. (**a**) A schematic showing the organization of matrices **X** and **Y**. The columns of **X** consist of the diffuse reflectance profiles at measured wavelengths. The columns of **Y** consist of expected optical properties ($\mu_a$ or $\mu_s'$) vs. wavelength. Rows of **X** and **Y** correspond to the number of samples. (**b**) Explanation of the basic concepts of PLS regression and the full-cross validation (leave-one-out) procedure. PLS regression decomposes matrices **X** and **Y** into score and loading vectors (scores $t_i$ and loadings $p_i$ for **X**; scores $u_i$ and loadings $q_i$ for **Y**) in a manner that optimally correlates **X** to **Y**. The resulting PLS model is used to predict **Y** values from future **X** measurements. In full-cross validation, one sample is left out of the training set while PLS regression is done on the remaining samples. Prediction is then made on the omitted sample. The procedure is iterated for the next sample and so on until all samples have served as test specimens. (**c**) An outline of the three approaches to PLS regression and verification of the two different training sets.

$$\mathbf{Y} = \mathbf{W}(\mathbf{P}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{Q}^{\mathrm{T}}\mathbf{X}. \quad (5)$$

In this study, the PLS2 algorithm[29] was implemented using the software package The Unscrambler (Camo ASA, Oslo, Norway). The PLS2 algorithm was selected because the **Y** matrix was multivariate (i.e., consisting of optical property spectra).The performance of a PLS model is judged by how well the model predicts optical properties from measurements of the diffuse reflectance. Ide-

ally, measurements of the diffuse reflectance from a new set of samples, (i.e., the test set) should be used to validate the model. An acceptable alternative for model verification is to use full-cross validation (leave-one-out),[42] which efficiently utilizes all data from the training set for PLS regression. Figure 2b schematically shows the full-cross validation procedure performed on a data set consisting of forty samples. Thirty-nine samples are used for PLS regression, while one is omitted to serve as a test sample. The PLS model based on the 39 samples is used to predict the optical property of the omitted sample. Iteration is performed on subsequent samples until all 40 samples have served as test specimen. For each test sample, the errors between PLS predicted and expected optical properties are calculated. The root mean square error of prediction (RMSEP) is calculated from these errors and represents the accuracy of PLS prediction. The number of loading vectors, or rank, of the PLS model is selected so that the RMSEP is minimal.[29,42] In addition to RMSEP, other merit indicators of the PLS model are the slope of the plot between the predicted versus expected values and its corresponding correlation coefficient.

To evaluate the feasibility of using PLS regression to predict optical properties from the diffuse reflectance data, three strategies for PLS regression were used with two different training sets: turbid phantom data and Monte Carlo simulated data. The three strategies are outlined in Fig. 2c. Slope, correlation coefficient, and RMSEP were calculated for each approach.

**Concentrations.** Partial least-squares regression and least-squares solution were both used to determine the dye and Intralipid concentrations from PLS predicted $\mu_a$ and $\mu_s'$ spectra, respectively. Specifically, a second stage PLS regression predicted concentration (dye or Intralipid) from the optical property ($\mu_a$ and $\mu_s'$) spectra. Least-squares solution was an alternative approach to extract concentration from optical property spectra. For instance, $\mu_a$ spectra are related to the dye concentrations as expressed by the matrix

$$\begin{bmatrix} \mu_a^{\lambda 1} \\ \mu_a^{\lambda 2} \\ \cdot \\ \cdot \\ \mu_a^{\lambda n} \end{bmatrix} = \begin{bmatrix} \varepsilon_{d1}^{\lambda 1} & \varepsilon_{d2}^{\lambda 1} & \cdot & \cdot & \varepsilon_{dm}^{\lambda 1} \\ \varepsilon_{d1}^{\lambda 2} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \varepsilon_{d1}^{\lambda n} & \varepsilon_{d2}^{\lambda n} & \cdot & \cdot & \varepsilon_{dm}^{\lambda n} \end{bmatrix} \begin{bmatrix} C_{d1} \\ C_{d2} \\ \cdot \\ \cdot \\ C_{dm} \end{bmatrix} \quad (6)$$

where $\epsilon_{di}^{\lambda i}$ is the extinction coefficient (ml/mg mm$^{-1}$) of the dye $di$ at wavelength $\lambda i$, and $C_{di}$ is the concentration (mg/ml) of the dye $di$. In this study, the concentration vector consisted of three elements because three dyes were used to make the phantoms. Extinction coefficients of Nigrosin, Janus green, and Prussian blue over the wavelength range were determined from absorption spectrophotometer measurements. A least-squares solution for Eq. 6 was determined for the concentration vector with the constraint that the concentration values be positive.[43] Within a certain range, the macroscopic $\mu_s'$ is proportional to the scatterer concentration and can be expressed in matrix form similar to Eq. 6. PLS predicted $\mu_s'$ spectra were then used to determine Intralipid percent concentration in a manner analogous to the dye concentration calculation.

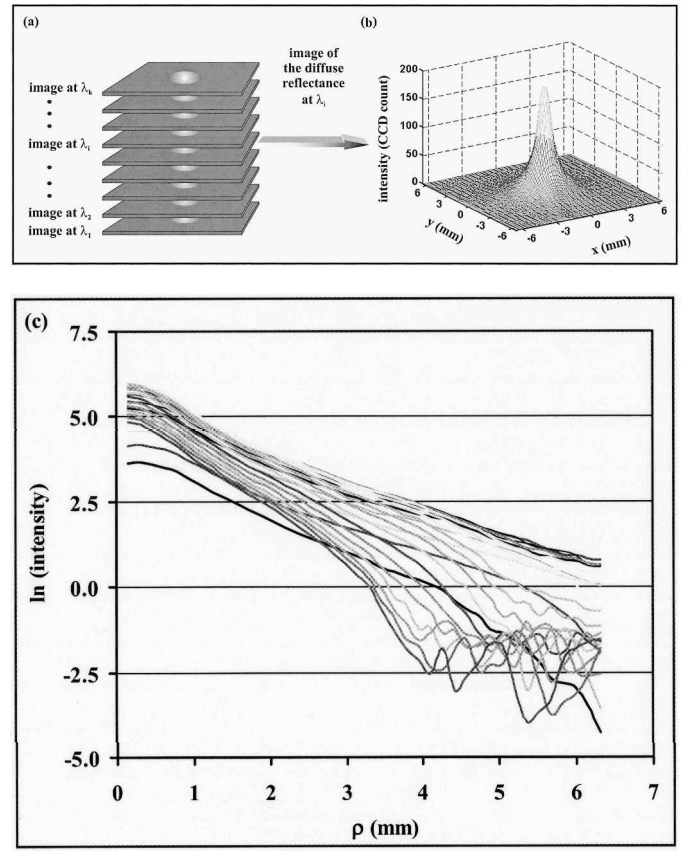
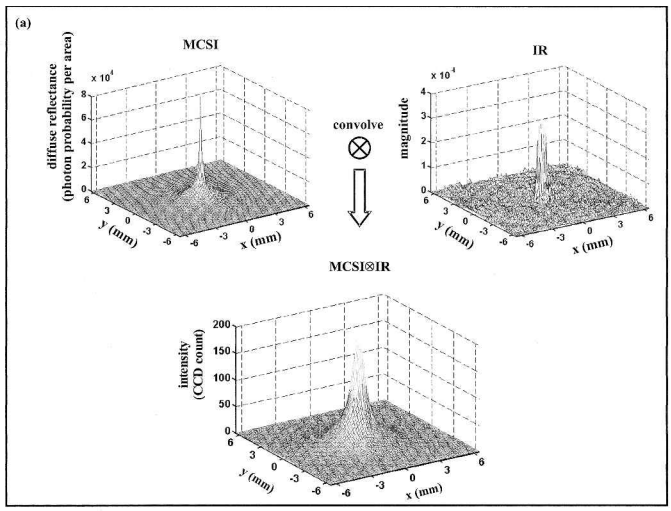



FIG. 3. (**a**) A schematic showing the data organization of hyperspectral images obtained from the FTIIS measurements. Image at each wavelength may contain up to 512 × 512 pixels at 12–bit dynamic range. (**b**) A typical image of the diffuse reflectance collected on a phantom with $\mu_a$ = 0.0281 mm$^{-1}$ and $\mu_s'$ = 1.163 mm$^{-1}$ at 630 nm. Image intensities were circumferentially binned and log transformed. (**c**) Plots of representative log-transformed diffuse reflectance data at 24 wavelengths from one of the turbid phantoms.

## RESULTS AND DISCUSSION

The hyperspectral images of 40 tissue-like phantoms were examined to select an outer distance to be used in PLS regression. An outer distance of 6.5 mm was selected to guarantee that the image intensities used in data fitting were higher than the CCD dark noise and the signal-to-noise ratio was on average greater than 2 at the outer distances (~6.5 mm). Figure 3a schematically presents the organization of the hyperspectral image data. In Fig. 3b, image intensity (in units of CCD counts) is plotted as a function of spatial coordinates (x, y) for a representative diffuse reflectance image at 630 nm. The images were circumferentially binned, and the intensity profiles were natural log transformed. Figure 3c shows representative log transformed diffuse reflectance profiles at 24 wavelengths for one of the samples. These profiles combined with the expected optical properties at 24 wavelengths of the forty samples constituted, respectively, the **X** and **Y** matrices of the turbid phantom training set.

Monte Carlo simulated data were used as the training set for PLS regression. To account for the instrument factors, Monte Carlo simulated image was convolved with the instrument response at the appropriate wavelength to form MCSI⊗IR. Figure 4a shows a typical example of



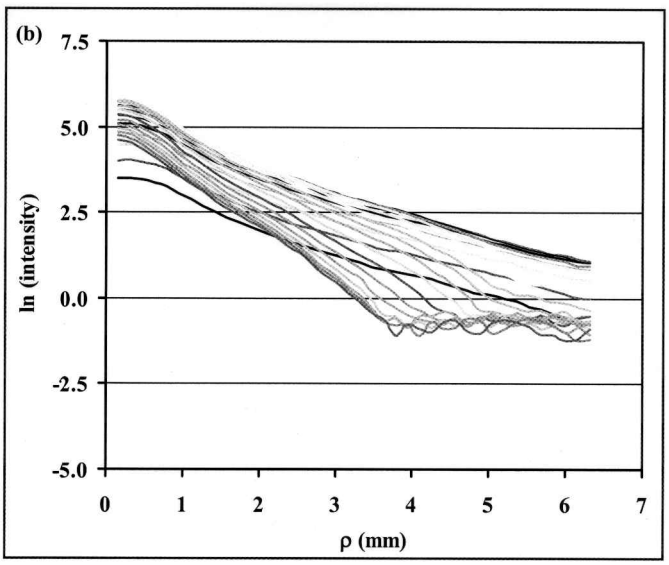




FIG. 4. (**a**) Examples of the Monte Carlo simulated image for $\mu_a$ and $\mu_s'$ values of 0.028 $mm^{-1}$ and 1.16 $mm^{-1}$, respectively, and the FTIIS instrument responses (IR) at 630 nm. IRs were calculated from the hyperspectral images collected on a reference phantom and the corresponding MCSI generated from its optical properties. MCSI was convolved with IR at the appropriate wavelength to generate MCSI⊗IR. A Monte Carlo training set was constructed from MCSI⊗IR data. (**b**) Representative examples of the circumferential binned and log-transformed MCSI⊗IR data.

the instrument response (630 nm) and the convolution procedure used to generate a simulation-based training set. Representative log transformed diffuse reflectance profiles from MCSI⊗IR data are shown in Fig. 4b for 24 wavelengths.

The regression of $\mu_a$ spectra vs. diffuse reflectance profiles was carried out separately from that of $\mu_s'$ spectra vs. diffuse reflectance. PLS regression with full-cross validation of the turbid phantom training set resulted in PLS models (trainTPD–valTPD) with rank (i.e., the number of components where RMSEP was minimal) 10 for $\mu_a$ and rank 5 for $\mu_s'$. Similarly, PLS regression with full-cross validation was carried out on the MCSI⊗IR training set, resulting in PLS models (trainMCSI–valMCSI) with rank 8 for $\mu_a$ and rank 5 for $\mu_s'$. In the third case (i.e., trainMCSI–predTPD), PLS models were developed from an MCSI⊗IR training set and were subsequently used to predict the optical properties of turbid phantoms. Optimal predictions were achieved using PLS models with a rank 5 for both $\mu_a$ and $\mu_s'$.

Figure 5 plots the results of predicted (from absorption spectrophotometry) vs. expected $\mu_a$ for the three PLS models. Data at 730 nm are shown for 40 samples and represent typical prediction results for $\mu_a$ at other wavelengths. The merit indicators of $\mu_a$ prediction (i.e., the slope, correlation coefficient, and RMSEP) for the data shown are specified in the figure caption. Likewise, Fig. 6 plots predicted vs. expected $\mu_s'$ for the three PLS models. Data at 640 nm are shown for 40 samples and represent typical prediction results for $\mu_s'$ at other wavelengths. The slope, correlation coefficient, and RMSEP of $\mu_s'$ prediction for the data shown are specified in the figure caption.

Values for the slope, correlation coefficient, and RMSEP at all measured wavelengths were pooled and the mean and standard deviations computed from the pooled values. Figure 7a plots the mean slope and correlation coefficient of $\mu_a$ and $\mu_s'$ predictions for the three PLS models. When the prediction is perfect, the slope and correlation coefficient equal one while RMSEP equals zero. For $\mu_a$ prediction, mean slopes of 1.014, 1.040, and 0.972 were obtained, respectively, for trainTPD–valTPD, trainMCSI–valMCSI, and trainMCSI–predTPD models. The corresponding mean correlation coefficients for the models were 0.958, 0.976, and 0.931. For $\mu_s'$ prediction, the mean slopes for the respective models were 0.974, 0.979, and 0.956. The corresponding mean correlation coefficients were 0.978, 0.995, and 0.980, respectively. Error bars for $\mu_a$ and $\mu_s'$ plotted in Fig. 7a correspond to standard deviations. Note that the standard deviations for $\mu_s'$ prediction are quite small, and consequently, the error bars do not display well. Error bars for $\mu_a$ prediction are large compared to error bars for $\mu_s'$ prediction because the merits of $\mu_a$ prediction are dependent on the wavelength. Compared to the results shown in Fig. 5, $\mu_a$ prediction was more accurate for shorter wavelengths but less accurate for longer wavelengths. In contrast, the merits of $\mu_s'$ prediction were less dependent on the wavelength. Figures 7b and 7c plot the mean RMSEP of $\mu_a$ and $\mu_s'$ prediction, respectively, for the three models. The root mean square errors of prediction for $\mu_a$ prediction were, respectively, 0.0094, 0.0057, and 0.0130 $mm^{-1}$ (or equivalently, ±8%, ±5%, and ±12% in terms of percent accuracy) for trainTPD–valTPD, trainMCSI–valMCSI, and trainMCSI–predTPD. The root mean square errors of prediction for $\mu_s'$ prediction were 0.081, 0.038, and 0.082 $mm^{-1}$ (±5%, ±3%, and ±5%) for the respective models. All merit indicators were optimal when PLS regression and validation were performed on the MCSI⊗IR training set.

The $\mu_a$ and $\mu_s'$ spectra derived from PLS predictions were used to determine the dye and Intralipid concentrations. Specifically, a second stage PLS regression and a least-squares algorithm were used to extract chromophore concentrations from $\mu_a$ spectra. Similarly, Intralipid percent volume of each phantom was calculated from $\mu_s'$ spectra. Concentration calculations obtained by using either PLS regression or least squares solution have comparable accuracy. Accordingly, Fig. 8 shows only repre-

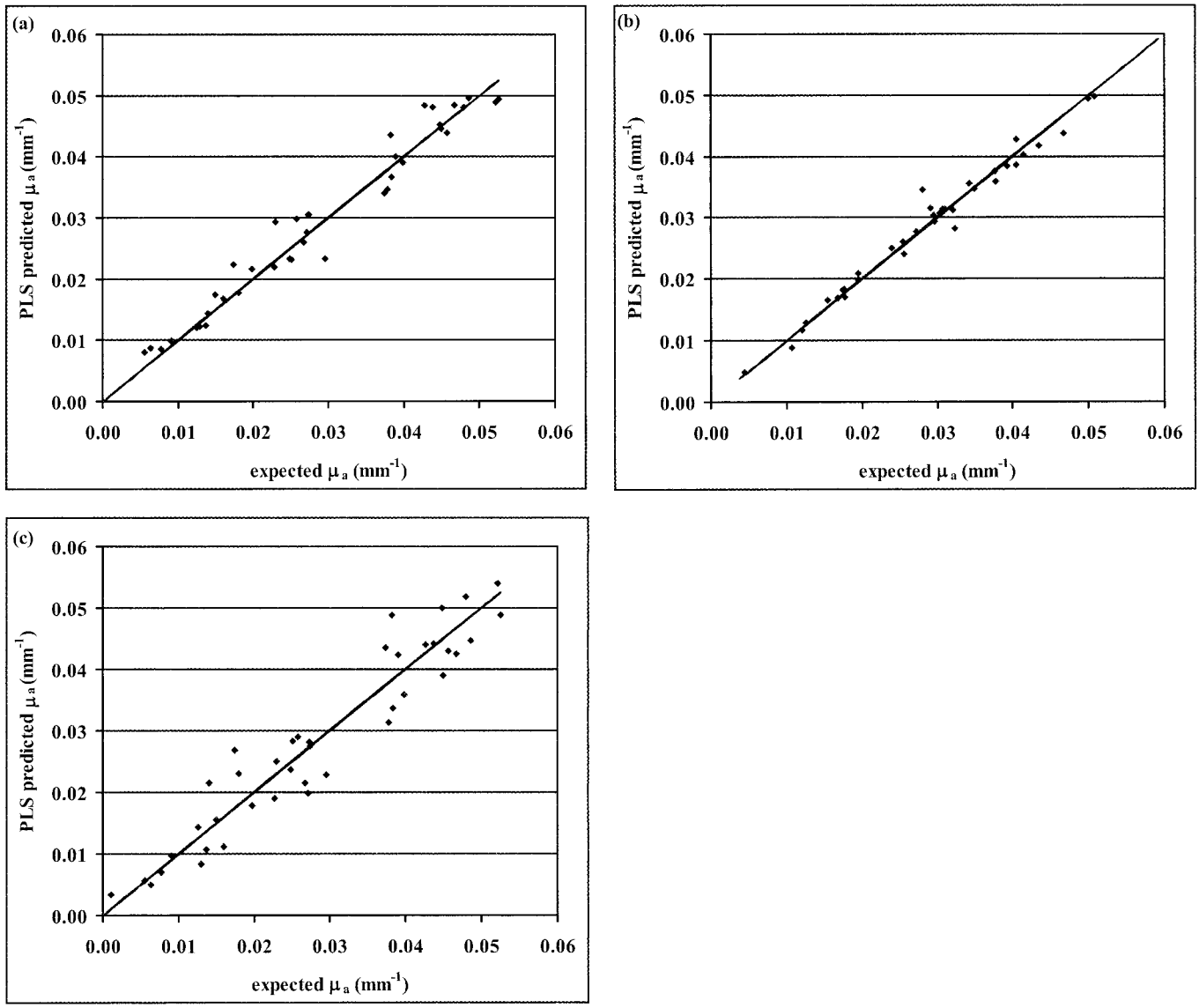



FIG. 5. PLS prediction of $\mu_a$ on all 40 phantoms is shown for the three PLS models at 730 nm. (**a**) PLS regression and full-cross validation were performed on the turbid phantom data set (trainTPD–valTPD). The slope, correlation coefficient, and RMSEP of $\mu_a$ prediction were 1.03, 0.964, and 0.0041 $mm^{-1}$, respectively, for a 10-component model. (**b**) PLS regression and full-cross validation (trainMCSI–valMCSI) were performed on the MCSI⊗IR data set. The slope, correlation coefficient, and RMSEP of $\mu_a$ prediction were 1.05, 0.979, and 0.0029 $mm^{-1}$, respectively, for a 10 component model. (**c**) PLS regression was first performed on the MCSI⊗IR data set; the resulting PLS model was then used to predict $\mu_a$ from the turbid phantom data (trainMCSI–predTPD). The slope, correlation coefficient, and RMSEP of $\mu_a$ prediction were 0.969, 0.954, and 0.0054 $mm^{-1}$, respectively, for a 5-component model.

sentative dye and Intralipid concentrations for the 40 phantoms that were obtained by the least-squares solution: (**a**) predicted vs. expected Prussian blue concentrations, and (**b**) predicted vs. expected Intralipid concentrations. A comparison of the predicted vs. expected concentrations shown in Fig. 8 yielded a prediction accuracy of ±7% for Prussian blue and ±0.5% for Intralipid.

## CONCLUSION

Quantitative hyperspectral imaging of diffuse reflectance from turbid media offers several key advantages over contact fiber-probe strategies. First, the image contains a complete two-dimensional profile of diffuse reflectance over a continuous and adjustable range of distances from the source. This feature may be important in the spectroscopy of biological tissues because the full image offers the possibility of identifying inhomogeneous regions. Second, image detection can be performed in non-contact or remote mode, an important feature well suited for certain clinical measurements. Third, conventional bright field images of the medium can be collected in conjunction with diffuse reflectance, making it possible to simultaneously visualize superficial structure and underlying composition. Finally, the FTIIS has an intrinsically broad spectral dynamic range (visible to near-infrared), thereby overcoming range limitations of single or multi-source imaging devices.

In a previous study, we noted that direct fitting of Monte Carlo simulations to diffuse reflectance profiles is computationally intensive and time consuming.[28] Consequently, the direct fitting approach is not suitable for applications that require rapid processing of hyperspectral data.

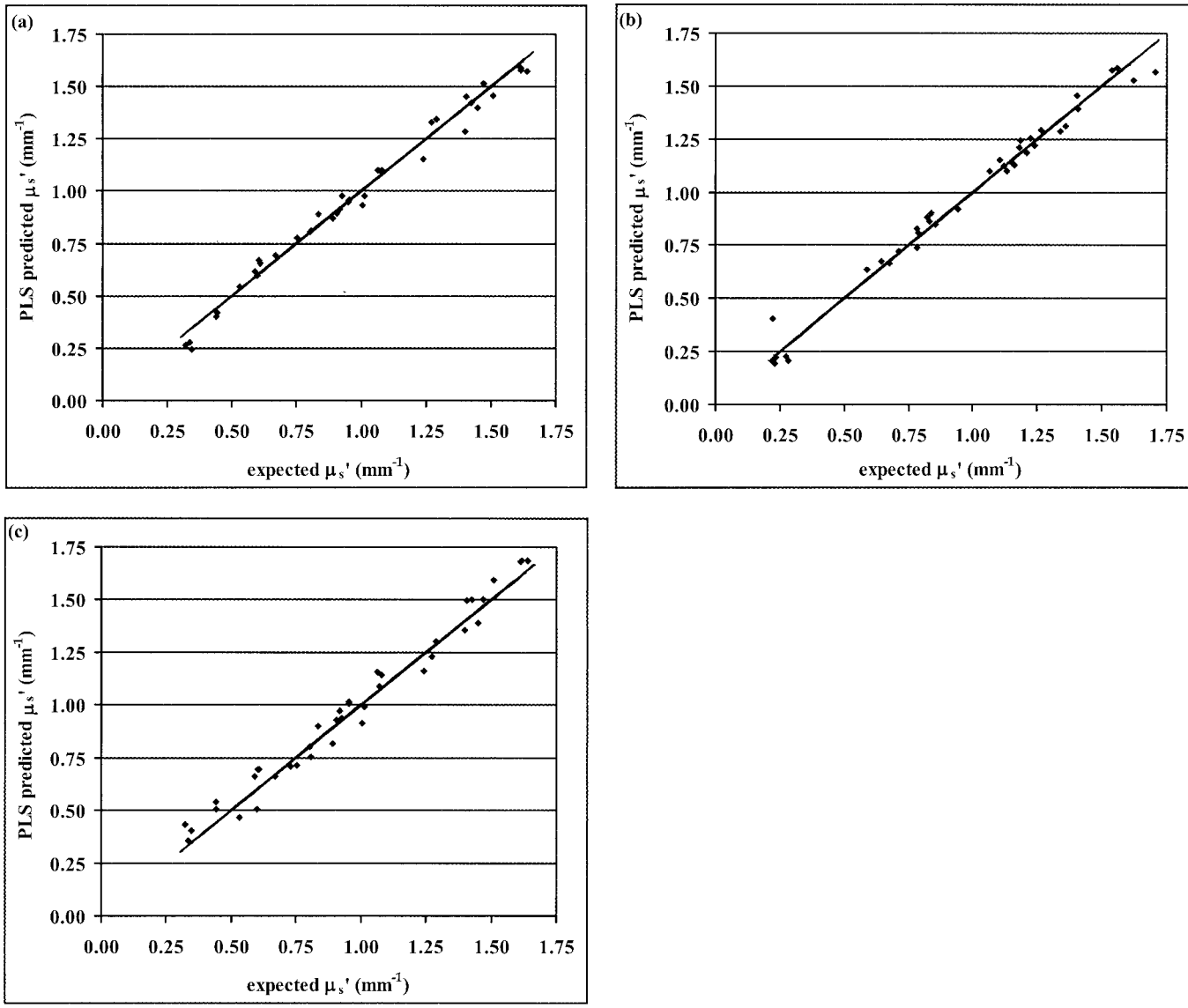



FIG. 6. PLS prediction of $\mu_s'$ on all 40 phantoms is shown for the three PLS models at 640 nm wavelength. (**a**) PLS regression and full-cross validation were performed on the turbid phantom data set (trainTPD–valTPD). The slope, correlation coefficient, and RMSEP of $\mu_s'$ prediction were 0.974, 0.978, and 0.084 $mm^{-1}$, respectively, for a 5-component model. (**b**) PLS regression and full-cross validation were performed on the MCSI⊗IR data set (trainMCSI–valMCSI). The slope, correlation coefficient, and RMSEP of $\mu_s'$ prediction were 0.971, 0.992, and 0.040 $mm^{-1}$, respectively, for a 5-component model. (**c**) PLS regression was first performed on the MCSI⊗IR training set; the resulting PLS model was then used to predict $\mu_s'$ from the turbid phantom data set (trainMCSI–predTPD). The slope, correlation coefficient, and RMSEP of $\mu_s'$ prediction were 0.956, 0.980, and 0.085 $mm^{-1}$, respectively, for a 5-component model.

In this study, PLS regression was used to predict optical properties as well as concentrations from the hyperspectral diffuse reflectance data. The study results show that PLS regression is an effective approach to rapidly quantifying optical properties of turbid media from the hyperspectral diffuse reflectance data. The accuracy of using a PLS model to predict $\mu_a$ and $\mu_s'$ is comparable to the direct fitting approach.[28] For instance, PLS prediction of $\mu_a$ is accurate to within $\pm 8\%$ while the accuracy for direct fitting is $\pm 7\%$. PLS prediction of $\mu_s'$ has an accuracy of $\pm 5\%$ while direct fitting is $\pm 3\%$. Note that the accuracy values are applicable for homogenous media with optical properties that are within the specified range. However, the approach is applicable for media with optical properties outside the range investigated in the study.

Not surprisingly, the best PLS prediction was obtained when PLS regression and validation were used with MCSI⊗IR data. MCSI⊗IR data were not subject to experimental errors, such as the measurement uncertainties associated with phantom preparation. Additionally, a calibration procedure was used to explicitly determine the instrument response in the case of MCSI⊗IR data. With regard to the turbid phantom data, we relied on the PLS model to implicitly account for the instrument response. The calibration procedure may be more accurate in accounting for the instrument response than PLS regression. Consequently, prediction results are optimal. Because the Monte Carlo simulations and IR contain noise, prediction of MCSI⊗IR data has small but finite errors. However, noise in the Monte Carlo data can be reduced by using more photons in the simulation.

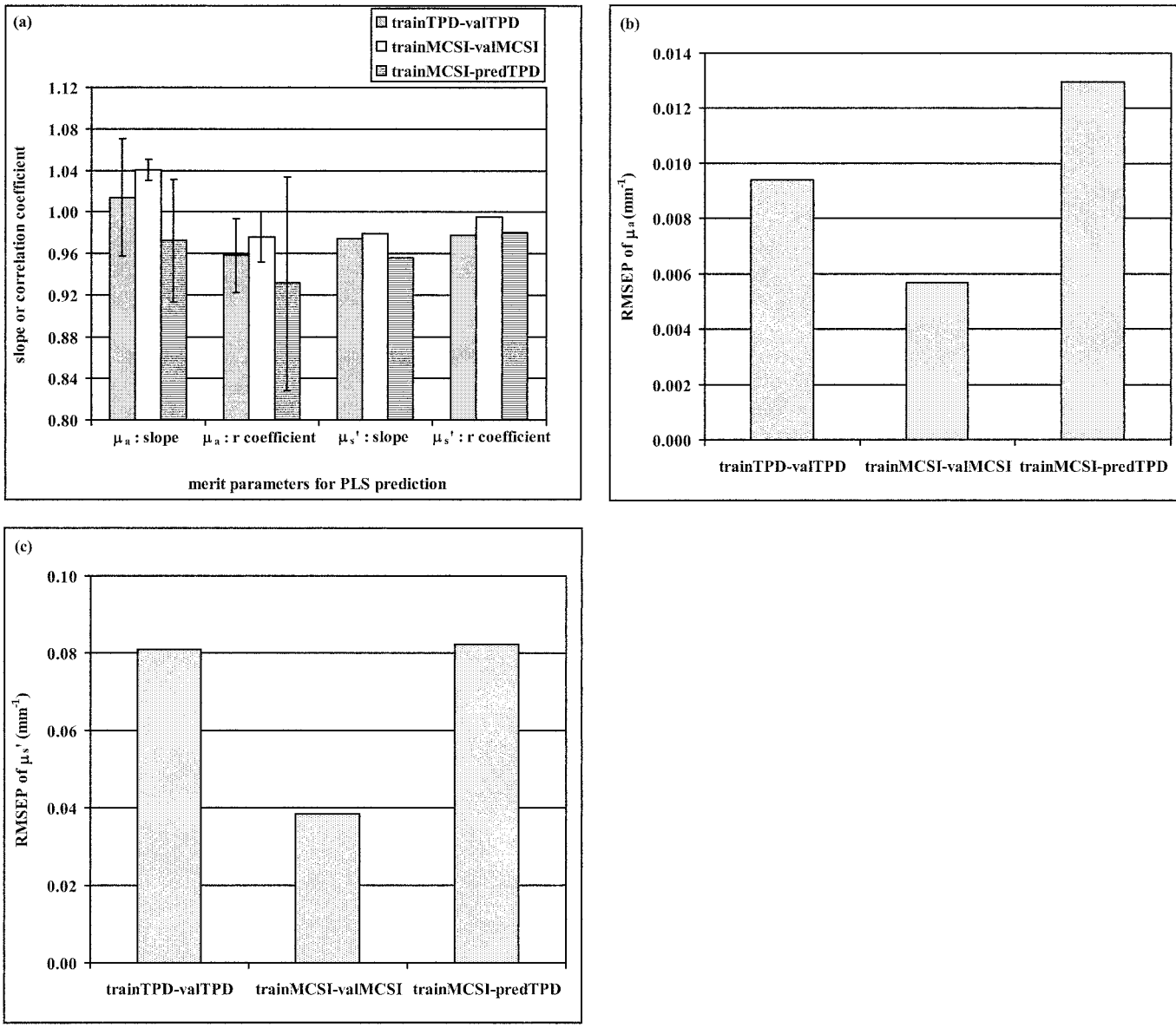
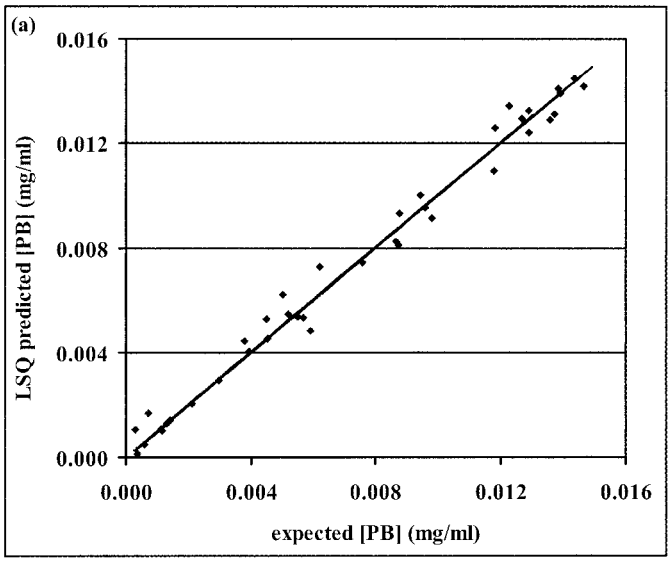



FIG. 7. The merit indicators of PLS prediction (i.e., the slope, correlation coefficient, and RMSEP) at all 24 wavelengths were pooled. Mean and standard deviations of the slope, correlation coefficient, and RMSEP were calculated from the pooled data. (**a**) Plots of the mean slope and correlation coefficient of $\mu_a$ and $\mu_s'$ predictions for the three PLS models. Error bars represent standard deviations. Standard deviations for $\mu_s'$ are quite small; consequently, the error bars do not show clearly on the plot. (**b**), (**c**) Respectively, plots of the RMSEP for $\mu_a$ and $\mu_s'$ predictions of the three PLS models.

Although the accuracy of quantifying $\mu_a$ and $\mu_s'$ was not improved with multivariate analysis, PLS regression is a more attractive approach than direct fitting for several important reasons. First, PLS prediction of $\mu_a$ and $\mu_s'$ is computationally efficient and rapid, since prediction is made directly from the linear PLS model rather than relying on a computationally intensive iterative search algorithm. Second, PLS regression offers practicality and flexibility not seen with direct least-squares fitting. Even when the physical underlying processes are not fully understood, it is still possible to construct an empirical PLS model based on an experimental training set. In cases where the physical processes (e.g., photon propagation in turbid media) and instrumental factors are fully characterized, a training set can be generated from simulated data. In fact, the study results demonstrate that a PLS model trained on simulated data (MCSI⊗IR) predicts optical properties from measured diffuse reflectance data almost as well as a model trained on turbid phantom data. This is, of course, only possible when the relationship between the measured quantities (e.g., diffuse reflectance) and the parameters of interest (e.g., optical properties) is fully understood. The option of using a simulated training set has an important practical implication. It is not necessary to generate a training set from real samples every time the instrument settings are changed. This is preferable because sample preparation and data collection are often time-consuming and cumbersome tasks. In essence, a much simpler IR calibration procedure replaces the time-consuming and inefficient process of constructing the training set from real samples. Moreover, the ability to use simulated data to accurately model the physical



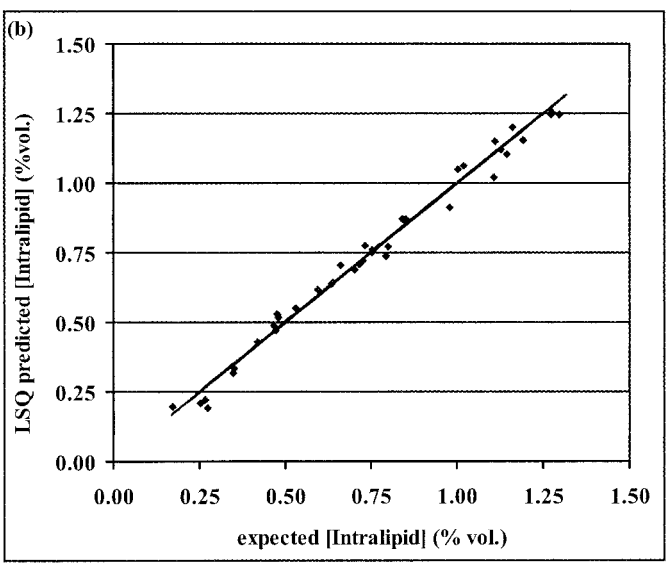




FIG. 8. Least-squares method was used to determine the chromophore and Intralipid concentration. Dye concentrations were determined from $\mu_a$ spectra, while Intralipid percent volume was extracted from $\mu_s'$ spectra. Examples of the least-squares (LSQ) derived concentrations are shown for 40 phantoms: (**a**) Prussian blue concentration and (**b**) Intralipid percent volume. For the data shown, the errors between predicted and expected concentrations are within ±7% and ±0.5% for Prussian blue and Intralipid, respectively. Similar accuracy values were obtained when a second PLS regression was used to predict concentrations from the optical property spectra.

processes is essential for spectroscopy of *in situ* tissues, where it is rarely possible to generate the training set from tissue samples.

In summary, a non-contact FTIIS imaging system in combination with PLS regression can be used to rapidly quantify the optical property spectra of turbid media over a wide spectral range. This integrative approach proves suitable for quantitative spectroscopy of tissue optical properties.

## ACKNOWLEDGMENTS

This work was supported by the Swedish Research Council for Engineering Sciences and the Knut and Alice Wallenberg Foundation. T.H.P. gratefully acknowledges support provided by the Fulbright Fellowship and the Whitaker Foundation Biomedical Graduate Fellowship.

1. R. Cubeddu, G. Canti, M. Musolino, A. Pifferi, P. Taroni, and G. Valentini, Photochem. Photobiol. **60,** 582 (1994).
2. C. E. Elwell, M. Cope, A. D. Edwards, J. S. Wyatt, D. T. Delpy, and E. O. R. Reynolds, J. Appl. Physiol. **77,** 2753 (1994).
3. B. Chance, Ann. N.Y. Acad. Sci. **838,** 29 (1998).
4. R. A. Weersink, J. E. Hayward, K. R. Diamond, and M. S. Patterson, Photochem. Photobiol. **66,** 326 (1997).
5. J. B. Fishkin, P. T. C. So, A. E. Cerussi, S. Fantini, M. A. Franceschini, and E. Gratton, Appl. Opt. **34,** 1143 (1995).
6. B. J. Tromberg, R. C. Haskell, S. J. Madsen, and L. O. Svaasand, Comments Mol. Cell. Biophy. **8,** 359 (1995).
7. N. Mohandas, Y. R. Kim, D. H. Tycko, J. Orlik, J. Wyatt, and W. Groner, Blood **68,** 506 (1986).
8. J. R. Mourant, J. P. Freyer, A. H. Hielscher, A. A. Eick, D. Shen, and T. M. Johnson, Appl. Opt. **37,** 3586 (1998).
9. A. H. Hielscher, J. R. Mourant, and I. J. Bigio, Appl. Opt. **36,** 125 (1997).
10. F. Bevilacqua, P. Marquet, O. Coquoz, and C. Depeursinge, Appl. Opt. **36,** 44 (1997).
11. I. S. Saidi, S. L. Jacques, and F. K. Tittel, Appl. Opt. **34,** 7410 (1995).
12. R. Bays, G. Wagnières, D. Robert, D. Braichotte, J. F. Savary, P. Monnier, and H. van den Bergh, Appl. Opt. **35,** 1756 (1996).
13. J. R. Mourant, T. M. Johnson, G. Los, and I. J. Bigio, Phys. Med. Biol. **44,** 1397 (1999).
14. S.-P. Lin, L. Wang, S. L. Jacques, and F. K. Tittel, Appl. Opt. **36,** 136 (1997).
15. G. Zonios, L. T. Perelman, V. Backman, R. Manoharan, M. Fitzmaurice, J. van Dam, and M. S. Feld, Appl. Opt. **38,** 6628 (1999).
16. F. Bevilacqua, D. Piguet, P. Marquet, J. D. Gross, B. J. Tromberg, and C. Depeursinge, Appl. Opt. **38,** 4939 (1999).
17. R. Doornbos, R. Lang, M. Aalders, F. Cross, and H. J. C. M. Sterenborg, Phys. Med. Biol. **44,** 967 (1999).
18. T. J. Farrell, M. S. Patterson, and B. Wilson, Med. Phys. **19,** 879 (1992).
19. E. L. Hull, M. G. Nichols, and T. H. Foster, Phys. Med. Biol. **43,** 3381 (1998).
20. S. L. Jacques, A. Gutsche, J. Schwartz, L. Wang, and F. Tittel, "Video reflectometry to specify optical properties of tissue *in vivo*", in *Medical Optical Tomography: Functional Imaging and Monitoring,* G. J. Müller, B. Chance, R. R. Alfano, S. R. Arridge, J. Beuthan, E. Gratton, M. Kaschke, B. R. Masters, S. Svanberg, and P. van der Zee, Eds. (SPIE Optical Engineering Press, Bellingham, Washington, 1993), p. 211.
21. A. Kienle, L. Lilge, M. S. Patterson, R. Hibst, R. Steiner, and B. C. Wilson, Appl. Opt. **35,** 2304 (1996).
22. R. Splinter, G. A. Nanney, L. Littman, C. H. Chuang, R. H. Svenson, J. R. Tuntelder, and G. P. Tatsis, Lasers Life Sci. **6,** 15 (1994).
23. L. Wang and S. L. Jacques, Appl. Opt. **34,** 2362 (1995).
24. P. Marquet, F. Bevilacqua, C. Depeursinge, and E. B. de Haller, Opt. Eng. **34,** 2055 (1995).
25. D. M. Haaland, H. D. T. Jones, and E. V. Thomas, Appl. Spectrosc. **51,** 340 (1997).
26. H. Key, R. E. Davies, P. C. Jackson, and P. N. Wells, Phys. Med. Biol. **36,** 579 (1991).
27. V. G. Peters, D. R. Wyman, M. S. Patterson, and G. L. Frank, Phys. Med. Biol. **35,** 1317 (1990).
28. T. H. Pham, F. Bevilacqua, T. Spott, J. S. Dam, B. J. Tromberg, and S. Andersson–Engels, Appl. Opt. **39,** 6487 (2000).
29. H. Martens and T. Næs, *Multivariate Calibration* (Wiley and Sons, Chichester, 1991).
30. K. Esbensen, S. Schönkopf, T. Midtgaard, and D. Guyot, *Multivariate Analysis in Practice* (CAMO ASA, Norway, 1998).
31. A. J. Berger, V. Venugopalan, A. J. Durkin, T. Pham, and B. J. Tromberg, Appl. Opt. **39,** 1659 (2000).
32. A. J. Durkin and R. Richards–Kortum, Lasers Surg. Med. **19,** 75 (1996).
33. W. H. Steel, *Interferometry* (Cambridge University Press, Cam-

bridge, 1983).
34. F. A. Duck, *Physical Properties of Tissue* (Academic Press Ltd., London, 1990).
35. H. J. van Staveren, C. J. M. Moes, J. van Marle, S. A. Prahl, and M. J. C. van Gemert, Appl. Opt. **30,** 4507 (1991).
36. R. A. J. Groenhuis, H. A. Ferwerda, and J. J. ten Bosch, Appl. Opt. **22,** 2456 (1983).
37. V. Venugopalan, J. S. You, and B. J. Tromberg, Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top. **58,** 2395 (1998).
38. T. Spott, Ph.D. Dissertation, Norwegian University of Science and Technology, Trondheim, Norway (1999).
39. L. G. Henyey and J. L. Greenstein, Astrophys. J. **93,** 70 (1941).
40. S. D. Brown, S. T. Sum, F. Despagne, and B. K. Lavine, Anal. Chem. **68,** R21 (1996).
41. D. M. Haaland and E. V. Thomas, Anal. Chem. **60,** 1193 (1988).
42. E. V. Thomas, Anal. Chem. **66,** A795 (1994).
43. C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Prentice–Hall, New York, 1974).