Suppression of inter-device variation for component analysis of turbid liquids based on spatially resolved diffuse reflectance spectroscopy

Shengzhao Zhang, Linna Zhang, Gang Li, and Ling Lin

Citation: Review of Scientific Instruments 88, 033104 (2017); doi: 10.1063/1.4977788

View online: http://dx.doi.org/10.1063/1.4977788

View Table of Contents: http://aip.scitation.org/toc/rsi/88/3

Published by the American Institute of Physics

Articles you may be interested in

Fast, precise, and widely tunable frequency control of an optical parametric oscillator referenced to a frequency comb
Review of Scientific Instruments 88, 033101 (2017); 10.1063/1.4977049

Laboratory-based high pressure X-ray photoelectron spectroscopy: A novel and flexible reaction cell approach
Review of Scientific Instruments 88, 033102 (2017); 10.1063/1.4975096

A compact atomic beam based system for Doppler-free laser spectroscopy of strontium atoms
Review of Scientific Instruments 88, 033103 (2017); 10.1063/1.4977593

High stable remote photoelectric receiver for interferometry
Review of Scientific Instruments 88, 033105 (2017); 10.1063/1.4978341

Space-resolved visible spectroscopy for two-dimensional measurement of hydrogen and impurity emission spectra and of plasma flow in the edge stochastic layer of LHD
Review of Scientific Instruments 88, 033501 (2017); 10.1063/1.4976963

Development and experimental study of oil-free capacitor module for plasma focus device
Review of Scientific Instruments 88, 033502 (2017); 10.1063/1.4977218
Suppression of inter-device variation for component analysis of turbid liquids based on spatially resolved diffuse reflectance spectroscopy

Shengzhao Zhang,1,2 Linna Zhang,1,2 Gang Li,1,2 and Ling Lin1,2,a)
1State Key Laboratory of Precision Measurement Technology and Instrument, Tianjin 300072, China
2Tianjin Key Laboratory of Biomedical Detection Techniques & Instruments, Tianjin University, Tianjin 300072, China

(Received 22 September 2016; accepted 16 February 2017; published online 10 March 2017)

Diffuse reflectance spectroscopy is a useful tool for obtaining quantitative information in turbid media, which is always achieved by developing a multivariate regression model that links the spectral signal to the component concentrations. However, in most cases, variations between the actual measurement and the modeling process of the device may cause errors in predicting a component’s concentration. In this paper, we propose a data-processing method to resist these variations. The method involves performing a curve fitting of the multiple-position diffuse reflectance spectral data. One of the parameters in the fitting function was found to be insensitive to inter-device variations and sensitive to the component concentrations. The parameter of the fitted equation was used in the modeling instead of directly using the spectral signal. Experiments demonstrate the feasibility of the proposed method and its resistance to errors induced by inter-device variations. Published by AIP Publishing.

I. INTRODUCTION

Diffuse reflectance spectroscopy is an important tool for the composition analysis of scattering materials. Compared to conventional chemical analysis methods, spectroscopy is a convenient technique that does not require any chemical reagents. Spectroscopy has been widely used in agriculture, environmental monitoring, food safety, medicine, and other fields.1–6 Among those applications, the steady-state spatially resolved reflectance spectroscopy of the surface of a medium is commonly used to obtain quantitative information on the components in a homogenously turbid medium.

To quantitatively analyze a material’s composition, one approach is to develop a multivariate regression model of the diffuse reflectance spectra versus the component concentrations. In the modeling methodology, it is essential that the device used for the actual measurement should have the same instrument response as the device used for the modeling process. In spatially resolved reflectance spectroscopy measurements, the diffuse reflectance lights at multiple positions are often measured by a hyperspectral image CCD camera or a spectrometer with an optical fiber scanning on the surface of the sample.3,6–8 The data from the instruments corresponding to the light intensities not only depend on the components in the sample but also on the acquisition system. When the light is collected by an optical fiber spectrometer, the size of the light beam from the light source, the distance from the optical fiber to the sample surface, the numerical aperture of the optical fiber, and the degree of fiber bending3,9,10 will have impacts on the measured data corresponding to the light intensity. However, it is not possible to keep all those conditions constant in actual measurement. Although researchers can calibrate the instrument before the actual measurements, this may not be sufficient to remove the error introduced by inter-device variations. A method to establish a model that is resistant to inter-device variations may thus be helpful.

In most cases, the properties of a data set containing light intensities at multiple positions can be described with a few parameters by fitting the data set to a mathematical function. Instead of directly using the raw spectral data, the parameters obtained by curve fitting can be used in the model development process as the independent variables. This idea was successfully used in the application of Rady and that of Peng and Lu.3,11,12 In Rady’s application, all the parameters were used to extract quantitative information of the samples. In our opinion, it is not necessary that all the parameters obtained by curve fitting should be used in the modeling process. There is the possibility that some of the parameters may be sensitive to the concentration of a component while others are not. It is possible to find a parameter in the fitted function that is sensitive to the component concentration and insensitive to the inter-device variations. The relationship model between the selected parameter and the concentration of the component may have the potential to resist inter-device variations.

In this study, we extracted quantitative information from multiple-position diffuse reflectance spectra. A data-processing method based on curve fitting was proposed that would be performed before establishing a multivariate regression. At each wavelength, the data at all positions were described by parameters through curve fitting. We assessed to what degree those parameters are sensible to the component concentration and the inter-device variation through simulation data. A mathematical function was used in the curve fitting and a model of the relationship between the parameters and the concentration of the component was developed. To make a comparison, the model with conventional methodology was also developed. Compared with the conventional modeling
method, the model built with fitted parameters could better suppress the error caused by inter-device variations.

II. SIMULATION DATA ANALYSIS FOR CURVE FITTING

As is mentioned in Section I, using curve fitting, the properties of a data set can be described with the parameters in the fitted mathematical function. In this section, simulation data were obtained for analyzing the relationship between the parameters and the optical properties. We also assessed how those parameters were influenced by the measurement conditions.

We prepared simulation data to study the spatially resolved diffuse reflectance light measurement. A schematic of the spatially resolved diffuse reflectance light measurement is shown in Fig. 1. A light beam of finite size is perpendicularly incident to the sample surface, and the diffuse reflectance light is collected by an optical fiber that is placed over the sample surface at a certain height. During the measurement, the optical fiber is scanning along the horizontal direction to detect the radial distribution of the diffuse reflected light.

The simulation experiment was performed using the Monte Carlo method. Our simulation program was written on the basis of Erik Alerstam’s CUDAMCML program and Lihong Wang’s CONV program. The number of photons in each simulation was set to $10^9$. The power of the light beam was set to 1000 µW and the diameter of the optical fiber was 400 µm. Five groups of data were generated from the simulations with different characteristics of light beams and fiber probes. The characteristic of the light beam, the numerical aperture of the fiber, and the fiber’s distance to the sample surface in the different data groups are listed in Table I. For flat beam, the parameter D denotes the diameter of the light beam. While for Gaussian beam, D denotes the diameter of a circle on which the light intensity is $1/e^2$ of the light intensity in the beam center. For data groups 1-4, with the constraints of the fiber diameter and numerical aperture, the light collected by the fiber was calculated at 8 positions with the radial distances from the light source of 10.0 mm, 11.0 mm, 12.0 mm, 13.0 mm, 14.0 mm, 15.0 mm, 16.0 mm, and 17.0 mm. For each group, 1200 samples with different optical properties were tested and numbered 1 to 1200. The absorption coefficient and scattering coefficient of a sample numbered k are described in Eqs. (1) and (2), where $k = 1, 2, 3, \ldots, 1200$,

\[ \mu_a^k = \text{mod}(k-1, 60) \times 0.1 + 0.1 \left( \text{cm}^{-1} \right), \quad (1) \]

\[ \mu_s^k = \left\lfloor \frac{(k-1)}{60} \right\rfloor + 1 \left( \text{cm}^{-1} \right). \quad (2) \]

In Eq. (2), $\lfloor (k-1)/60 \rfloor$ is rounded down to the nearest integer. According to the analysis by Cui and Feng, the incident photons migrate to the detector through the paths distributed in a region shaped as a “banana,” with its two ends connecting the sources and the detector and its mid portion reaching the deepest. According to the equations given in the paper of Cui and Feng, the minimum depth should exceed 12.02 mm when the radial distance between the light beam and the fiber probe is 18 mm. In these simulations, the depths of all the samples are set to 50 mm.

Fig. 2 shows the calculated light intensity when $\mu_a = 3.0 \text{ cm}^{-1}$, $\mu_s = 10.0 \text{ cm}^{-1}$ at 8 different positions under 5 different measurement conditions. Fig. 2 also shows that under different light beams or fibers, the calculated values exhibit large differences.

![FIG. 1. Schematic of spatially resolved diffuse reflectance light measurement.](image1)

![FIG. 2. Calculated light intensities when $\mu_a = 3.0 \text{ cm}^{-1}$ and $\mu_s = 1.0 \text{ cm}^{-1}$ at 8 positions under different light beam and optical fiber conditions.](image2)
An exponential distribution was applied for fitting the radial distribution of the light intensities and the function was expressed by Eq. (3) as follows:

\[ I = \exp(ax^2 + bx + c), \]  

where \( x \) is the location number of the 8 points from near to far (1, 2, 3, ..., 8); \( I \) is the light intensity of a measurement point corresponding to a certain measurement condition; and \( a \), \( b \), and \( c \) are undetermined parameters. To make the curve fitting easier, Eq. (3) can also be written as

\[ \ln(I) = ax^2 + bx + c. \]  

By fitting function (4) to the simulation data, the values of \( a \), \( b \), and \( c \) can be obtained, and they are shown in Figs. 3–5, respectively. The maximum norm of the residuals of all the fittings was \( 1.75 \times 10^{-2} \). Figure 3 shows that the value of “\( a \)” changes irregularly with the number of the sample.

Figures 4 and 5 show that parameters \( b \) and \( c \) regularly changed with the changing of the absorption coefficients or scattering coefficients. The absorption coefficient or scattering coefficient of a solution has a direct linear relationship with the concentrations of its components; therefore, the quantitative information of material’s composition can be theoretically predicted by the values of \( b \) and \( c \) at different wavelengths. The 5 curves in different colors in Figs. 4 and 5 represent the value of \( b \) and \( c \) under the 5 different measurement conditions listed in Table I. The curves of \( b \) basically overlap while those of \( c \) do not. We defined a difference coefficient (DC) to quantitatively describe the difference between two curves,

\[ DC(P, Q) = \frac{\sum_{i=1}^{n} (p_i - q_i)^2}{\sqrt{\sum_{i=1}^{n} (p_i - \bar{p})^2 \sum_{i=1}^{n} (q_i - \bar{q})^2}}. \]  

The maximum DC between the curves in Fig. 4 is 0.0064 while the maximum DC between the curves in Fig. 5 is 0.1006. Parameter \( b \) does not appear to be sensitive to inter-device variations caused by the light beam changes, numerical aperture changes when the optical fiber is replaced, changes in the fiber’s height from the sample surface, or a certain offset in the distant of the source detector, leading to only very small changes in the value of \( b \). Therefore, if the value of \( b \) is used to predict the material composition, even if disturbances occur in the measurement device, their influence on the measurement results will be very small.

III. MATERIALS AND EXPERIMENTS

A. Acquisition system

The diffuse reflectance spectral acquisition system is shown in Fig. 6 and is mainly composed of a light source, a sample pool, an electronically controlled translation stage, and a spectrometer.

The light source used was a supercontinuum white light laser source (SuperK™ Compact, NKT Photonics, Denmark) with a spectral range of 500–2400 nm and an average output power of 100 mW. The light beam illuminating the sample pool was 4 mm in diameter with a divergence angle of approximately 0.04 rad. An electronically controlled translation stage (TSAx-C, Zolix) with an orientation precision of 0.5 µm and a resolution of 0.625 µm was used to carry a fiber probe scanning on the surface of the sample along the radial direction.
Two fiber probes were prepared for different experiments, each with a fiber diameter of 400 \( \mu \)m and with numerical apertures of 0.22 and 0.37, respectively. The other end of the fiber was coupled into a spectrometer (AvaSpec-HS1024*58TEC-USB2, Avantes, Netherlands). The spectrometer has a spectral range of 299.87–1160 nm, and the resolution of the analog-to-digital converter in the spectrometer was 16-bit. A glass sample pool with a 120 mm diameter and 50 mm height was placed under the fiber probe.

B. Samples

Water dilutions of India ink and Intralipid have widely been used to prepare diffusive phantoms for optical studies. Intralipid is a nutritional product that is composed of soybean oil, lecithin, glycerin, and water. It is a highly scattering medium with an absorption coefficient as small as that of water. India ink is a simple ink that consists of carbon particles suspended in water. The absorption coefficient of India ink is far larger than its scattering coefficients. The optical properties of Intralipid and India ink provide the flexibility to formulate a variety of diffusive liquid phantoms with different optical properties.

Using Intralipid (30% lipid, Huari Pharmaceutical Co., China), India ink (Beijing Solarbio Science & Technology Co., Ltd.), and distilled water, we prepared two groups of samples for the experiments in this paper with compositions of those samples listed in Table II. In the table, IL denotes Intralipid and INK denotes India ink. We prepared 20 different samples in Group 1 with different concentrations of IL. In Group 2, 30 different samples were prepared with same concentration of IL and different volume ratios of INK. In Group 1, all the samples have the same absorption coefficient but different scattering coefficients, and in Group 2, the samples have different absorption coefficients but approximately the same scattering coefficient.

C. Experimental procedure

We collected spectral signals using the device described above. To ensure the stability of the light source, the supercontinuum laser was turned on for at least 20 min before the spectrum was obtained. The “shutter” button of the supercontinuum laser could control whether there was a laser output without interrupting the work of the internal light source. The sample was placed under the light source and the optical fiber probe and the spectrometer recorded the data to be saved on the PC. To guarantee that the spectral signals have as high a signal-to-noise ratio as possible, the spectrometer was set at different integration times at different sampling points. The maximum count of the spectrometer was 65535. At each measurement position, the integration time was adjusted to make the spectral peak reach at least 75% (49150) of the maximum count, and the spectral signal \( I(\lambda) \) was recorded at the integration time \( t \). After measuring each position, the “shutter” button of the supercontinuum laser was used to turn off the output of the laser, and the spectral signal \( I(\lambda) \) of the same integration time was recorded by the spectrometer to be regarded as the spectrum of ambient light. When the measurement was completed, the reflection spectrum of each point was calculated using Eq. (6), where \( I^s \) is the actual spectrum,

\[
I^s = \frac{I^A - I^B}{t}.
\]

During the measurement, the electronically controlled translation stage carried the fiber probe to several positions to collect the diffusive reflectance light. Four measurements at different measurement conditions were individually carried out, as is detailed in Table III, which lists the measured samples, the vertical distance from the endpoint of the fiber probe to the sample surface \( H \), the fiber’s NA, and the radial distances from the fiber probe to the center of the light \( R \). In Measurement 1 and Measurement 2, the diffusive reflectance light was collected at 8 positions, where \( R \) equals 10 mm, 11 mm, . . . , 17 mm. In Measurement 3 diffusive reflectance light at 10 positions where \( R \) equals 9.0 mm, 9.5 mm, . . . , 13.5 mm was collected. In Measurement 4, the diffusive reflectance light was collected at 8 positions where \( R \) equals 9.0 mm, 9.5 mm, . . . , 12.5 mm.

D. Data pre-processing

In each measurement, the diffusive reflectance spectrum at each sampling point was obtained and saved. Only the spectra in the wavelength range 500 nm–1000 nm with a total of 550 discrete wavelengths were used in the data processing. For Measurement 1 and Measurement 2, the spectral data of each

<table>
<thead>
<tr>
<th>TABLE II. Composition of samples used in the experiments.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample group 1</td>
</tr>
<tr>
<td>Composition(s)</td>
</tr>
<tr>
<td>Sample group 2</td>
</tr>
<tr>
<td>Composition(s)</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
sample were measured at 8 different positions. The spectral data are arranged by an $8 \times 550$ array in a style described as array (7) in which $I_{jk}^{i}$ denotes the light intensity count measured at wavelength $\lambda k$ at the $i$th position,

$$
\begin{bmatrix}
I_{1}^{11} & I_{1}^{12} & \cdots & I_{1}^{1550} \\
I_{2}^{11} & I_{2}^{12} & \cdots & I_{2}^{1550} \\
\vdots & \vdots & \ddots & \vdots \\
I_{8}^{11} & I_{8}^{12} & \cdots & I_{8}^{1550}
\end{bmatrix}.
$$

(7)

For Measurement 3, the spectral data of each sample were measured at 10 different positions. The spectral data for each sample at the 1-8th positions are arranged in a style described as array (8), and the spectral data at the 3-10th positions are arranged as a style described as array (9),

$$
\begin{bmatrix}
I_{1}^{31} & I_{1}^{32} & \cdots & I_{1}^{3550} \\
I_{1}^{41} & I_{1}^{42} & \cdots & I_{1}^{4550} \\
\vdots & \vdots & \ddots & \vdots \\
I_{8}^{31} & I_{8}^{32} & \cdots & I_{8}^{3550}
\end{bmatrix},
$$

(8)

$$
\begin{bmatrix}
I_{3}^{11} & I_{3}^{12} & \cdots & I_{3}^{1550} \\
I_{4}^{11} & I_{4}^{12} & \cdots & I_{4}^{1550} \\
\vdots & \vdots & \ddots & \vdots \\
I_{10}^{11} & I_{10}^{12} & \cdots & I_{10}^{1550}
\end{bmatrix}.
$$

(9)

The main purpose of the data analysis and processing was to enable us to obtaining quantitative information on the components in the liquids from the diffuse reflectance spectral signal, which is generally achieved by developing a multivariate regression model. Several data preprocessing procedures should be carried out before developing the models. In most studies, the researchers performed normalization to the raw spectral data and then developed a regressive model between the normalized spectral data and the volume ratio or concentration of the samples. In this paper, curve fitting was applied to the spectral data measured at different positions, and then we developed a regressive model between the fitted parameters and the volume or concentration of the samples. In order to make comparison, both methods were carried out.

1. Normalization

In conventional data preprocessing, the data are normalized before the modeling process, as was performed in most studies.$^{5,6,18}$ In this process, all of the light intensity counts in the data array were divided by the spectral intensity counts in the first line. After normalization, the data in the first line of the array are discarded and the array is rearranged to be a $1 \times 3850$ array. The normalized and rearranged data for each sample in Measurements 1 and 2 were denoted as $\mathbf{u}_{i}^{IL1}$ and $\mathbf{u}_{i}^{IL2}$ in which subscript $i$ stands for the sample index. For Measurement 3, two data arrays were used to describe the diffusive light. The normalized and rearranged data were denoted as $\mathbf{u}_{i}^{INK1}$ and $\mathbf{u}_{i}^{INK2}$, with respect to arrays (8) and (9), respectively. For Measurement 4, the normalized and rearranged data were denoted as $\mathbf{u}_{i}^{INK3}$.

2. Curve fitting

For the data obtained in Measurement 1, column vector $[I_{1}^{1k}, I_{2}^{1k}, \ldots, I_{8}^{1k}]^{T}$ and column vector $[1, 2, \ldots, 8]^{T}$ were fitted to Eq. (4) to obtain the parameter of “$a$,” “$b$,” and “$c$.“ The values of “$b$” were saved and form a fitted parameter vector as is described by the following equation:

$$
\mathbf{p}_{i}^{IL1} = \begin{bmatrix} b_{i}^{11} & b_{i}^{12} & \cdots & b_{i}^{1550} \end{bmatrix}.
$$

(10)

In the same way, the data in each column obtained in Measurement 2 were also fitted to Eq. (4). The fitted parameter array is denoted as $\mathbf{p}_{i}^{IL2}$. For the data obtained in Measurement 3, the data in arrays (8) and (9) were also fitted and the fitted array are denoted as $\mathbf{p}_{i}^{INK1}$ and $\mathbf{p}_{i}^{INK2}$, respectively, and the fitted parameter array for the data in Measurement 4 is denoted as $\mathbf{p}_{i}^{INK3}$.

To make the description more convenient, some of the symbols used in the data processing are listed in Table IV. The symbol $\phi_{i}^{IL}$ stands for the volume ratio of Intralipid in the $i$th sample in Sample Group 1 and the symbol $\phi_{i}^{INK}$ stands for the volume ratio of India ink in the $i$th sample in Sample Group 2.

### Table IV. Symbols defined for the data array in the data processing.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Normalized data vector</th>
<th>Fitted parameter vector</th>
<th>Volume ratio of the object</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\mathbf{u}_{i}^{IL1}$</td>
<td>$\mathbf{p}_{i}^{IL1}$</td>
<td>$\phi_{i}^{IL}$</td>
</tr>
<tr>
<td>2</td>
<td>$\mathbf{u}_{i}^{IL2}$</td>
<td>$\mathbf{p}_{i}^{IL2}$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\mathbf{u}_{i}^{INK1}$</td>
<td>$\mathbf{p}_{i}^{INK1}$</td>
<td>$\phi_{i}^{INK}$</td>
</tr>
<tr>
<td>4</td>
<td>$\mathbf{u}_{i}^{INK2}$</td>
<td>$\mathbf{p}_{i}^{INK2}$</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>$\mathbf{u}_{i}^{INK3}$</td>
<td>$\mathbf{p}_{i}^{INK3}$</td>
<td></td>
</tr>
</tbody>
</table>
And remaining 4 normalized data vectors were chosen as an independent prediction set. Another calibration model \( M(P_{IL}^{11}, \Phi_{IL}) \) was also built by using the 16 fitted parameter vectors in data set \( \{ p_{IL}^{11} \} \). And remaining 4 fitted parameter vectors were chosen as an independent prediction set.

For Sample Group 2, 24 of the 30 normalized data vectors in data set \( \{ u_{IL}^{1N} \} \) were randomly chosen as the calibration set for developing the calibration model \( M(U_{IL}^{1N}, \Phi_{IL}) \), and remaining 6 normalized data vectors were chosen as an independent prediction set. Another calibration model \( M(P_{INK}^{N}, \Phi_{INK}) \) was also built by using the 24 fitted parameter vectors in data set \( \{ p_{IL}^{N} \} \), and the remaining 6 fitted parameter vectors were chosen as an independent prediction set.

**IV. RESULTS AND DISCUSSION**

The modeling results corresponding to models \( M(U_{IL}^{1}, \Phi_{IL}) \) and \( M(P_{IL}^{1}, \Phi_{IL}) \) are shown in Figs. 7(a) and 7(b), respectively. The calculated volume ratio of IL in both the calibration set and prediction set is shown in the figures. Figures 7(a) and 7(b) show that the calculated IL volume ratios are close to the 1:1 best fit line. The correlation coefficient for calibration (Rc), the correlation coefficient for prediction (Rp), and the root mean square error of prediction (RMSEP) were calculated to evaluate the model, as shown in Table V. The idea of developing a model using the curve fitting parameter to predict the volume ratio of the scattering medium is validated.

The models \( M(U_{IL}^{11}, \Phi_{IL}) \) and \( M(P_{IL}^{11}, \Phi_{IL}) \) are built with the data corresponding to the spectral signals obtained in calibration set and prediction set. Figures 7(a) and 7(b) show that the calculated IL volume ratios are close to the 1:1 best fit line. The correlation coefficient for calibration (Rc), the correlation coefficient for prediction (Rp), and the root mean square error of prediction (RMSEP) were calculated to evaluate the model, as shown in Table V. The idea of developing a model using the curve fitting parameter to predict the volume ratio of the scattering medium is validated.

**TABLE V. Performance of the two models in the volume ratio prediction of IL.**

<table>
<thead>
<tr>
<th></th>
<th>( Model \ M(U_{IL}^{11}, \Phi_{IL}) )</th>
<th>( Model \ M(P_{IL}^{11}, \Phi_{IL}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data set</td>
<td>Rc</td>
<td>Rp</td>
</tr>
<tr>
<td>{( u_{IL}^{N1} )} or {( p_{IL}^{N1} )}</td>
<td>0.9972</td>
<td>0.9844</td>
</tr>
<tr>
<td>{( u_{IL}^{N2} )} or {( p_{IL}^{N2} )}</td>
<td>0.9478</td>
<td>0.9944</td>
</tr>
</tbody>
</table>

**FIG. 7.** Real value of IL volume ratio versus volume ratio calculated by the PLSR model. (a) The IL volume ratio is calculated by applying data set \( \{ u_{IL}^{11} \} \) to model \( M(U_{IL}^{11}, \Phi_{IL}) \). (b) The IL volume ratio is calculated by applying data set \( \{ p_{IL}^{11} \} \) to model \( M(P_{IL}^{11}, \Phi_{IL}) \).

**FIG. 8.** Real value of IL volume ratio versus volume ratio calculated by the PLSR model. (a) The IL volume ratio is calculated by applying data set \( \{ u_{IL}^{12} \} \) to model \( M(U_{IL}^{11}, \Phi_{IL}) \). (b) The IL volume ratio is calculated by applying data set \( \{ p_{IL}^{12} \} \) to model \( M(P_{IL}^{11}, \Phi_{IL}) \).
FIG. 9. Real value of INK volume ratio versus volume ratio calculated by the PLSR model. (a) The INK volume ratio is calculated by applying data set \( \{u_{INK1}^i\} \) to model \( M(U_{INK1}, \Phi_{INK}) \). (b) The INK volume ratio is calculated by applying data set \( \{p_{INK1}^i\} \) to model \( M(P_{INK1}, \Phi_{INK}) \).

Measurement 1. The data sets \( \{u_{IL1}^i\} \) and \( \{p_{IL1}^i\} \) correspond to the spectral data obtained in Measurement 2. The NA of the fibers and the vertical distance from the probe endpoint to the sample surface were different in Measurement 1 and Measurement 2. To show to what extent those changes distort the calculated results of the IL volume ratios, we applied the data set \( \{u_{IL2}^i\} \) or \( \{p_{IL2}^i\} \) to model \( M(U_{IL2}, \Phi_{IL}) \) and applied \( \{p_{IL2}^i\} \) to model \( M(P_{IL2}, \Phi_{IL}) \) to calculate the volume ratio of IL, and the results are shown in Figs. 8(a) and 8(b). The Rp and RMSEP are also provided in Table V. It is shown that changes will cause greater increases of RMSEP and decreases of Rp by using model \( M(U_{IL1}, \Phi_{IL}) \) than using model \( M(P_{IL1}, \Phi_{IL}) \).

The modeling results corresponding to models \( M(U_{INK1}, \Phi_{INK}) \) and \( M(P_{INK1}, \Phi_{INK}) \) are shown in Figs. 9(a) and 9(b), respectively. The calculated volume ratio of INK in both the

<table>
<thead>
<tr>
<th>Data set</th>
<th>Model ( M(U_{INK1}, \Phi_{INK}) )</th>
<th>Model ( M(P_{INK1}, \Phi_{INK}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u_{IL1}^i ) or ( p_{IL1}^i )</td>
<td>0.9930 0.9671 2.31 × 10^{-4}</td>
<td>0.9990 0.9892 1.33 × 10^{-4}</td>
</tr>
<tr>
<td>( u_{IL2}^i ) or ( p_{IL2}^i )</td>
<td>0.7875 5.43 × 10^{-4}</td>
<td>0.9834 1.60 × 10^{-4}</td>
</tr>
<tr>
<td>( u_{IL3}^i ) or ( p_{IL3}^i )</td>
<td>0.4660 7.79 × 10^{-4}</td>
<td>0.7145 6.16 × 10^{-4}</td>
</tr>
</tbody>
</table>
The models $M(U^{\text{INK}1}, \Phi^{\text{INK}})$ and $M(p^{\text{INK}1}, \Phi^{\text{INK}})$ are built with the data corresponding to the spectral signals measured when the radial distance from the fiber probe to the incident light ranges from 9.0 mm to 12.5 mm and the vertical distance from the probe endpoint to the sample surface is 5 mm. The data sets $\{u_i^{\text{INK}2}\}$ and $\{p_i^{\text{INK}2}\}$ correspond to the spectral data obtained when the radial distance from the fiber probe to incident light ranges from 10.0 mm to 13.5 mm and the vertical distance from the probe endpoint to the sample surface is again 5 mm. To show how this change affects the calculated result of the INK volume ratios when the radial distance has a 1 mm offset, we applied data set $\{u_i^{\text{INK}2}\}$ to model $M(U^{\text{INK}1}, \Phi^{\text{INK}})$ and applied $\{p_i^{\text{INK}2}\}$ to model $M(p^{\text{INK}1}, \Phi^{\text{INK}})$, and the results are shown in Figs. 10(a) and 10(b). The calculated INK volume ratios are distorted more severely in Fig. 10(a) than in Fig. 9(a). The values of Rp and RMSEP are also provided in Table VI. It is shown that changes will introduce greater increases of RMSEP and decreases of Rp upon using model $M(U^{\text{INK}1}, \Phi^{\text{INK}})$ than when using model $M(p^{\text{INK}1}, \Phi^{\text{INK}})$.

The data sets $\{u_i^{\text{INK}3}\}$ and $\{p_i^{\text{INK}3}\}$ correspond to the spectral data obtained when the vertical distance from the probe endpoint to the sample surface is 10 mm and the radial distance from the fiber probe to the incident light ranges from 9.0 mm to 12.5 mm. To assess how the change in the vertical distance from the probe to the sample surface affects the result, we applied the data set $\{u_i^{\text{INK}3}\}$ to model $M(U^{\text{INK}1}, \Phi^{\text{INK}})$ and applied $\{p_i^{\text{INK}3}\}$ to model $M(p^{\text{INK}1}, \Phi^{\text{INK}})$ to calculate the INK volume ratio and the results are shown in Figs. 11(a) and 11(b). The calculated results are distorted severely upon using model $M(U^{\text{INK}1}, \Phi^{\text{INK}})$, while the results in Fig. 11(b) show less distortion. The values of RMSEP and Rp show that model $M(p^{\text{INK}1}, \Phi^{\text{INK}})$ is more immune to changes in the probe’s height.

In practical cases, it is not possible that the conditions of the actual measurement and modeling process will be completely identical. There may be minor unexpected offsets of the radial distance between the fiber probe and the incident light. When we put the several liquid samples into the sample pool, it is not possible that the volumes of the samples will be maintained constant, and the consequence is a change in the vertical distance from the sample surface to the endpoint. Those variations may have impacts on the measurement. The experiments involved a scattering medium and an absorption medium. The results show that the model built with the fitted parameters is more immune to the variations than that built with generally used normalized spectral data.

**V. CONCLUSION**

In this paper, spatially resolved VIS-NIR spectroscopy was used to analyze the concentration of scattering liquids. The purpose of the generally used method was to directly develop a relationship model between the spectral signal set and the concentration data set. In this study, the light intensities of multiple positions were fitted to an exponential distribution function with three undetermined parameters. Monte Carlo simulations were performed to calculate the diffuse reflectance light intensities of samples with a variety of optical properties under different measurement conditions. By fitting the light intensity data from the simulations to the mathematical functions, the parameters in the functions were obtained. We analyzed how those parameters in the exponential function changed with the optical properties and also the measurement conditions. The results showed that one of the three parameters was sensitive to the optical properties and insensitive to some measurement conditions. A phantom experiment was designed based on the measurement of the diffuse reflectance spectra of different samples composed of Intralipid or India ink under calibration set and prediction set is also shown in the figures. Figures 9(a) and 9(b) show that the calculated IL volume ratios are close to the 1:1 best fit line. The values of Rc, Rp, and RMSEP were also calculated and are provided in Table VI. It shows that the models can perform well.

The models $M(U^{\text{INK}1}, \Phi^{\text{INK}})$ and $M(p^{\text{INK}1}, \Phi^{\text{INK}})$ are built with the data corresponding to the spectral signals measured when the radial distance from the fiber probe to the incident light ranges from 9.0 mm to 12.5 mm and the vertical distance from the probe endpoint to the sample surface is 5 mm. The data sets $\{u_i^{\text{INK}2}\}$ and $\{p_i^{\text{INK}2}\}$ correspond to the spectral data obtained when the radial distance from the fiber probe to incident light ranges from 10.0 mm to 13.5 mm and the vertical distance from the probe endpoint to the sample surface is again 5 mm. To show how this change affects the calculated result of the INK volume ratios when the radial distance has a 1 mm offset, we applied data set $\{u_i^{\text{INK}2}\}$ to model $M(U^{\text{INK}1}, \Phi^{\text{INK}})$ and applied $\{p_i^{\text{INK}2}\}$ to model $M(p^{\text{INK}1}, \Phi^{\text{INK}})$, and the results are shown in Figs. 10(a) and 10(b). The calculated INK volume ratios are distorted more severely in Fig. 10(a) than in Fig. 9(a). The values of Rp and RMSEP are also provided in Table VI. It is shown that changes will introduce greater increases of RMSEP and decreases of Rp upon using model $M(U^{\text{INK}1}, \Phi^{\text{INK}})$ than when using model $M(p^{\text{INK}1}, \Phi^{\text{INK}})$.

The data sets $\{u_i^{\text{INK}3}\}$ and $\{p_i^{\text{INK}3}\}$ correspond to the spectral data obtained when the vertical distance from the probe endpoint to the sample surface is 10 mm and the radial distance from the fiber probe to the incident light ranges from 9.0 mm to 12.5 mm. To assess how the change in the vertical distance from the probe to the sample surface affects the result, we applied the data set $\{u_i^{\text{INK}3}\}$ to model $M(U^{\text{INK}1}, \Phi^{\text{INK}})$ and applied $\{p_i^{\text{INK}3}\}$ to model $M(p^{\text{INK}1}, \Phi^{\text{INK}})$ to calculate the INK volume ratio and the results are shown in Figs. 11(a) and 11(b). The calculated results are distorted severely upon using model $M(U^{\text{INK}1}, \Phi^{\text{INK}})$, while the results in Fig. 11(b) show less distortion. The values of RMSEP and Rp show that model $M(p^{\text{INK}1}, \Phi^{\text{INK}})$ is more immune to changes in the probe’s height.

In practical cases, it is not possible that the conditions of the actual measurement and modeling process will be completely identical. There may be minor unexpected offsets of the radial distance between the fiber probe and the incident light. When we put the several liquid samples into the sample pool, it is not possible that the volumes of the samples will be maintained constant, and the consequence is a change in the vertical distance from the sample surface to the endpoint. Those variations may have impacts on the measurement. The experiments involved a scattering medium and an absorption medium. The results show that the model built with the fitted parameters is more immune to the variations than that built with generally used normalized spectral data.

**V. CONCLUSION**

In this paper, spatially resolved VIS-NIR spectroscopy was used to analyze the concentration of scattering liquids. The purpose of the generally used method was to directly develop a relationship model between the spectral signal set and the concentration data set. In this study, the light intensities of multiple positions were fitted to an exponential distribution function with three undetermined parameters. Monte Carlo simulations were performed to calculate the diffuse reflectance light intensities of samples with a variety of optical properties under different measurement conditions. By fitting the light intensity data from the simulations to the mathematical functions, the parameters in the functions were obtained. We analyzed how those parameters in the exponential function changed with the optical properties and also the measurement conditions. The results showed that one of the three parameters was sensitive to the optical properties and insensitive to some measurement conditions. A phantom experiment was designed based on the measurement of the diffuse reflectance spectra of different samples composed of Intralipid or India ink under

![Fig. 11](image-url)
several different measurement conditions. Curve fittings were performed and a model was developed between the curve-fitting parameters and the concentration of the sample. The performances of the models were tested and we also applied spectral data measured under other measurement conditions to test how the models resist to the measurement condition variations. The results show that the model built with certain parameters obtained from the curve fitting had better performance. In conclusion, the data processing method based curve fitting could more efficiently suppress the error induced by the inter-device variations compared with conventional normalization methods.

ACKNOWLEDGMENTS

This project was supported by Tianjin Application Basis and Front Technology Study Programs (No. 14JCZDJC33100).

18 M. Huang et al., “Maize seed variety classification using the integration of spectral and image features combined with feature transformation based on hyperspectral imaging,” Appl. Sci. 6, 183 (2016).