University of
Strathclyde
Glasgow

Strathprints Institutional Repository

http://strathprints.strath.ac.uk/

Full Correction of Scattering Effects Using Radiative Transfer Theory for Improved Quantitative Analysis of Absorbing Species in Suspensions

Raimundas Steponavičius,[1,3] *Suresh N. Thennadil*[2*]

[1]Merz Court, School of Chemical Engineering and Advanced Materials, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom; [2]James Weir Building, 75 Montrose Street, Department of Chemical and Process Engineering, University of Strathclyde, Glasgow, G1 1XJ, United Kingdom; [3]Bijūnų g. 17, Department of Informatics Engineering, Klaipėda University, Lithuania.

[*]To whom correspondence should be addressed: suresh.thennadil@strath.ac.uk.

ABSTRACT

Sample-to-sample photon path length variations that arise due to multiple scattering can be removed by decoupling absorption and scattering effects using radiative transfer theory with a suitable set of measurements. For samples where particles both scatter and absorb light the extracted bulk absorption spectrum is not completely free from nonlinear particle effects since it is related to the absorption cross section of particles which changes nonlinearly with particle size and shape. For the quantitative analysis of absorbing only (i.e. non-scattering) species present in a matrix that contains a particulate species which absorbs and scatters light, a method to eliminate particle effects completely is proposed which utilizes the particle size information contained in the bulk scattering coefficient extracted using Mie theory to carry out an additional correction step to remove particle effects from bulk absorption spectra.

1

This would result in spectra which are equivalent to spectra collected using only the liquid species in the mixture. Such an approach has the potential to significantly reduce the number of calibration samples as well as improving calibration performance. The proposed method was tested using both simulated and experimental data from a 4 component model system.

KEYWORDS: Scatter correction, multivariate calibration, Near-infrared spectroscopy, multiple light scattering, radiative transfer equation, adding-doubling method.

INTRODUCTION

Acquisition of accurate and robust calibration models and thus applicability of spectroscopic methods in process analytics for quantitative NIR analysis of turbid samples is hindered by nonlinear multiple light scattering effects that degrade conventional multivariate linear calibration models and make extraction of chemical information from such samples challenging. There are essentially two ways to deal with undesirable scattering effects in NIR measurements: remove/minimize them by means of empirical pre-processing or separate scattering effects from absorption using first principles i.e. by invoking light propagation theory such as the radiative transfer theory. In either case, the goal is to obtain a measure of absorption per unit length, which is independent from variations in path length of photons that occur due to multiple scattering and linearly proportional to concentrations of constituents. Although considerable progress has been made in the empirical scatter correction techniques [1-6] they are not expected to solve the problem of multiple light scattering completely because they are based on assumptions which over-simplify the scattering problem and thus are not valid for many situations which involve large differences in the scattering profile between samples in the data set. Therefore, approaches based on separating absorption and scattering effects using fundamentals of light propagation are gaining more and more impetus in the field of spectroscopic quantitative analysis of highly scattering samples. [7-12] It should be noted however that these approaches also involve errors such as measurement errors and computational errors that have a direct impact on the prediction performance.

One of such approaches has been experimentally tested in two recent works.[11,12] The methodology presented there was based on building multivariate calibration models using partial least squares (PLS) regression on the bulk absorption coefficient spectra $\boldsymbol{\mu_a}^*$ extracted from a set of different spectroscopic measurements using the inversion of the radiative transfer equation. However, although a significant improvement in the prediction performance was achieved by decoupling absorption from scattering in comparison with using empirical pre-processing techniques, it was still not up to the level that can be achieved with the transparent (non-scattering) media where the measured absorption is linearly proportional to concentrations of chemical species. This is because in the cases where particles both scatter and absorb light the extracted bulk absorption coefficient is not completely free from nonlinear particle effects since it is related to the absorption cross section of particles $\sigma_{ap}(\lambda)$, which changes nonlinearly with particle size and shape. For the quantitative analysis of absorbing only (i.e. non-scattering) species present in a matrix that contains a particulate species which absorbs and scatters light, it is possible to eliminate particle effects completely and further improve the performance of calibration models by carrying out an additional correction step and removing particle effects from $\boldsymbol{\mu_a}$.

There is a further advantage that has the potential to significantly impact in terms of reduction in the size of calibration dataset (since it would not be necessary for the data set to contain samples with a range of particle sizes) as well as significantly improving the robustness of the calibration model and the possibility of effective calibration transfer. The ability to mathematically remove all effects of particulate component from the spectrum implies that in principle, the calibration model will effectively be the same as it would be if the model was built using a calibration set consisting of only the liquid components in the mixture. Thus an effective method which can be used to remove the effect of particulate species on the spectra of samples, will enable the application of a model to batches where the particle size/shape differ significantly from that encompassed in the calibration set. Hence, the aim of this work was to propose and test the feasibility of a full correction approach for prediction of

---

[*] Bold symbols used for vectors of values representing a range of wavelengths λ and also matrices containing values for a range of wavelengths and multiple samples.

absorbing-only chemical species. The approach was tested on both simulated and the experimental datasets. A simulation study is presented to show the maximum theoretical improvement in the prediction accuracy possible using the methodology described in this paper. The method was then applied to an experimental data set to examine how much improvement can be obtained when this approach is used in conjunction with measurements from an integrating sphere set up.

MATERIALS AND METHODS

Methodology for full correction of multiple scattering effects

In the case of quantitative analysis of non-scattering samples total absorbance is a suitable quantity for multivariate calibration since it varies linearly with concentrations of absorbing species as indicated by the Beer-Lambert law. Absorptivities (i.e. absorption cross-sections) of chemical species and the path length of light are constant in this case, the path length travelled by the light being equal to the sample thickness. However, in turbid samples the path length of photons is not constant and the absorption cross-section of particles may also not be constant depending on whether particles absorb or not. If particles only scatter light then the path length normalized μ_a is a suitable quantity for multivariate calibration and therefore the scatter correction approach involving the extraction of the bulk absorption coefficient using the radiative transfer theory[11,12] is sufficient in this case. When particles not only scatter but also absorb light, the extracted μ_a is still nonlinearly related to the physical properties of particles such as size through the absorption coefficient of particles $\mu_{ap}(\lambda)$. This can be seen from the following equation (eq. 4 in Reference 10):

$$\mu_a = \sum_{i=1}^{n} \mu_{a_i} = \sum_{j=1}^{n_p} \sigma_{ap_j} \cdot c_{p_j} + \sum_{k=1}^{n_a} \sigma_{a_k} \cdot c_k \qquad (1)$$

where σ_{ap} is the spectrum of absorption cross-section (cm^2) of the particulate species j, c_{p_j} is the concentration of the particulate species j expressed as number density i.e. number of particles per unit volume (cm^{-3}) and n_p is the number of different particulate species present in the sample. σ_{a_k} represents

4

the spectrum of absorptivity (cm^2/g) of the purely absorbing species k, c_k is the concentration (g/cm^3) of the absorbing species k and n_a is the number of absorbing only species present in the sample. It should be noted that the bulk absorption and scattering coefficients as well as the absorption and scattering cross-sections of the particles and the absorptivity of the absorbing only species are all wavelength dependent. In equation 1 μ_a has been split into two terms. The first summation represents the contribution from the particulate species and the second summation represents the contribution from the purely absorbing species. The first term will vary both due to the concentration of the particulate species as well as its particle size because the absorption cross-section $\sigma_{ap}(\lambda)$ is dependent on the particle size and shape. The second term varies only with the concentration of the purely absorbing species. In such cases, the bulk absorption coefficient of the absorbing only species, represented by the term $\sum\limits_{k=1}^{n_a}\sigma_{a_k}\cdot c_k$ is theoretically the most suitable quantity for multivariate calibration. To obtain this quantity two pre-processing steps are required, first to extract the bulk absorption coefficient spectrum μ_a and second to subtract μ_{ap} from it. In the first step (extraction of μ_a) we eliminate photon path length variations that predominate in all measurements in highly scattering media and in the second step (subtraction of μ_{ap} from μ_a) we remove nonlinear variations in $\mu_a(\lambda)$ because of $\mu_{ap}(\lambda)$. This is referred to as the full correction of multiple scattering effects in this work. Conceptually, the full correction is the same as taking particles away from the medium, but using calculations based on light propagation theory rather than doing that physically. The proposed concept of full correction of nonlinear scattering effects for estimation of concentrations of absorbing only species in suspensions is presented in figure 1. The most difficult part here is to get an estimate of $\mu_{ap}(\lambda)$. The two other optical properties of particles namely $\mu_s(\lambda)$ and $g(\lambda)$ are directly obtained using inverse adding doubling (IAD) (if we have just one scattering species), but $\mu_{ap}(\lambda)$ is an implicit part of the bulk absorption coefficient and so cannot be extracted directly. For spherical particles, $\mu_{ap}(\lambda)$ as well as $\mu_s(\lambda)$ and $g(\lambda)$ can be computed using Mie solution

given the particle size distribution (PSD) and number density N_p (concentration) and complex refractive index. Since PSD and N_p are unknown one has to find them first, by inverting Mie solution using the known optical properties, to be able to compute $\mu_{ap}(\lambda)$. Inversion of Mie solution is a well-known problem. [13-19] The fact which simplifies this problem in our case is that the goal is not to obtain an accurate particle size distribution. Even an approximate estimate of $\boldsymbol{\mu}_{ap}$ and a subsequent correction may partly linearize $\boldsymbol{\mu}_a$ and that may improve calibration results, which is our ultimate goal.

Conventionally, values of $\mu_s(\lambda)$ at multiple wavelengths have been used for the inversion of Mie solution and acquisition of PSD.[13,14] The objective function used in the nonlinear optimization being the length of the error vector of $\boldsymbol{\mu}_s$:

$$f = \sqrt{\sum_{i=1}^{n}\left(\mu_s(\lambda_i) - \hat{\mu}_s(\lambda_i)\right)^2} \qquad (2)$$

Where $\mu_s(\lambda_i)$ is the value of bulk scattering coefficient at wavelength λ_i which is extracted from the measurements, $\hat{\mu}_s(\lambda_i)$ is the value calculated using Mie theory and n is the number of wavelengths used for the extraction of PSD and N_p. In our case, the multiple measurements allow us to extract all three optical properties and they all can be included in the objective function. Thus, we can constrain this least squares fitting problem better by giving more points that have to be fitted, e.g. instead of minimizing the error only in $\mu_s(\lambda)$ one can minimize the error in both $\mu_s(\lambda)$ and $g(\lambda)$. The more points there are the better the estimates of PSD, N_p and $\mu_{ap}(\lambda)$ are likely to be. Inclusion of $\mu_a(\lambda)$ however is not straightforward. The way it can be included into the objective function is described here. This method is applicable if the absorbing-only species that constitute the medium are known. Then, one can form a matrix **A** made up of the spectra of pure components (i.e. absorbing only species) or mixtures of these components:

$$\mathbf{A} = \begin{bmatrix} \mu_{a,1} & \mu_{c,2} & \mu_{3} & \cdots & \mu_{n} \\ \mu_{a} & \mu_{c} & \mu_{} & \cdots & \mu_{} \\ \mu_{a,} & \mu_{c,} & \mu_{} & \cdots & \mu_{} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_{a,m1} & \mu_{a,m2} & \mu_{a,m3} & \cdots & \mu_{a,mn} \end{bmatrix} \tag{3}$$

Every spectrum of a mixture of the pure components must be in the column space of $\mathbf{A}$ since they are all some linear combination of the columns of $\mathbf{A}$ that form the basis of that space. For a binary mixture, the space that the two column vectors of $\mathbf{A}$ form is a plane and the spectra of all possible mixtures of the two pure components would lie in this plane too. Now, if some particles are added into that mixture the dimensionality of the bulk absorption spectra would increase due to the nonlinear $\boldsymbol{\mu}_{ap}$ component and so their vectors would be sticking out of the plane defined by the two absorbing only components (see illustration in figure 2). The distance from the vector $\boldsymbol{\mu}_a$ to the plane represents the error due to uncorrected $\boldsymbol{\mu}_{ap}$. It is equal to the length of the error vector $\mathbf{e}$, which can be calculated as:

$$\mathbf{e} = \boldsymbol{\mu}_a - \boldsymbol{\mu}_{a,\text{corr}} \tag{4}$$

where, the vector $\boldsymbol{\mu}_{a,\text{corr}}$ is the projection of the vector $\boldsymbol{\mu}_a$ onto the plane. The projection of the vector $\boldsymbol{\mu}_a$ onto the column space can be found using the projection matrix $\mathbf{P}_r$:

$$\boldsymbol{\mu}_{a,\text{corr}} = \mathbf{P}_r \cdot \boldsymbol{\mu}_a \tag{5}$$

$$\mathbf{P}_r = \mathbf{A} \cdot \left(\mathbf{A}^T \cdot \mathbf{A}\right)^{-1} \cdot \mathbf{A}^T \tag{6}$$

This error can be included into the objective function for the inverse Mie solution as a measure of how accurate the estimated $\boldsymbol{\mu}_{ap}$ and consequently the corrected $\boldsymbol{\mu}_a$ are at the end of each iteration of nonlinear optimization. The more accurate the estimated $\boldsymbol{\mu}_{ap}$ the more accurate the corrected $\boldsymbol{\mu}_a$ and the smaller the

error. When $\boldsymbol{\mu}_{ap}$ is determined accurately it would be completely removed from $\boldsymbol{\mu}_a$ and the error should be equal to zero. Hence the error in the corrected $\boldsymbol{\mu}_a$ can be calculated as follows:

$$\mathbf{e} = \hat{\boldsymbol{\mu}}_{a,corr} - \mathbf{P}_r \cdot \hat{\boldsymbol{\mu}}_{a,corr} \qquad\qquad (7)$$

where $\hat{\boldsymbol{\mu}}_{a,corr}$ is the estimated value of "particle free" bulk absorption coefficient spectrum. Spectra that are already in the column space of $\mathbf{A}$ will stay unchanged by the projection matrix $\mathbf{P}_r$, that is $\mathbf{P}_r \cdot \boldsymbol{\mu}_{a,corr} = \boldsymbol{\mu}_{a,corr}$. Note, that not only the spectra of pure components can be used in the matrix $\mathbf{A}$, but spectra of mixtures of pure components too. The subspace of the pure components in both cases will be the same. We might well have the case when the absorption coefficients of pure species are not linearly additive at some absorption bands as is the case in the samples considered in this study.[12] Then the matrix of just pure components is not sufficient, because they do not span/define the whole subspace of spectra of all possible mixtures. In such cases, we can use methods such as Principal Component Analysis (PCA) to model approximately the variation in absorption spectra. But for that we need to collect absorption spectra of different mixtures. The extracted principal components can then be used as columns of the matrix $\mathbf{A}$. Therefore, the following objective function was used in the proposed methodology:

$$f = w_1 \sqrt{\sum_{i=1}^{n}\left(\mu_s(\lambda_i) - \hat{\mu}_s(\lambda_i)\right)^2} + w_2 \sqrt{\sum_{i=1}^{n}\left(g(\lambda_i) - \hat{g}(\lambda_i)\right)^2} + w_3 \sqrt{\sum_{i=1}^{n} e_i^2} \qquad\qquad (8)$$

where w_1, w_2 and w_3 are the weights and n is the number of wavelengths. In this way, all available information is used and so there is a higher likelihood to get better estimates of PSD, N_p and $\boldsymbol{\mu}_{ap}$ than using just $\boldsymbol{\mu}_s$, because the nonlinear optimization will now have to fit all three optical properties of particles namely $\boldsymbol{\mu}_s$, $\mathbf{g}$ and $\boldsymbol{\mu}_{ap}$ (indirectly). Values of $\boldsymbol{\mu}_s$ and $\mathbf{g}$ may be in a very different scale, therefore they were normalized to unit length so that they had equal weights in the objective function. The length of the error vector $\mathbf{e}$ was about the same scale as of $\boldsymbol{\mu}_s$ and $\mathbf{g}$. If needed the weight of each term in the objective function can be adjusted using additional multipliers w_1, w_2 and w_3. It is particularly useful

when the extracted optical properties are of different accuracy. For example in the given experiment, the accuracy of the extracted **g** was significantly lower than the accuracy of μ_s therefore a lower weight was used on **g**.

To further simplify the inversion of Mie solution we assumed that the form of the particle size distribution was also known. In many practical situations this assumption may be reasonable because approximate distributions are often known. Commercial particle sizing instruments based on light scattering measurements usually make this assumption. The size of the latex particles that were used in the experiments followed the normal distribution. Thus, there were three unknown variables in the inversion, namely, mean and standard deviation of PSD and N_p.

The inversion of the radiative transfer equation (RTE) for extraction of the optical properties and the calibration steps are described in the references given.[11,12]

Simulation

The simulated dataset was similar to the actual four-component system used in the experimental part. The refractive indices of the same four chemical components as those used in the experiment, namely: water, deuterium oxide, ethanol and polystyrene particles, were taken to model the optical properties of the samples. Optical properties of particles were simulated using Mie code for poly-disperse particles (the code was based on Bohren's and Huffman's code,[20] but extended for poly-disperse particles). The concentrations of ethanol (c_{eth}), polystyrene (c_{pst}) and deuterium oxide (c_{hw}) were varied from 0 to 0.1, from 0 to 0.2 and from 0 to 0.5 in fractions of mass respectively. The concentration of water (c_w) was taken such that the closure condition was satisfied i.e. $c_w + c_{hw} + c_{eth} + c_{pst} = 1$ The mean of the radii of particles was varied from 50 to 2000 nm and the standard deviation from 1 to 50 nm. The values of the mean and standard deviation were drawn randomly in the predefined ranges. The wavelength range used was 400-1880 nm. The dataset comprised 400 samples in total.

The projection matrix required for acquisition of μ_{ap} estimates and subsequent correction of μ_a was made out of the pure components in this case. The dataset comprising 400 samples in total was divided into the training set (50 samples) and the validation set (350 samples). The training dataset was used for

building the PLS calibration model whilst the validation dataset was used for obtaining the unbiased statistics of the accuracy of the predictions of ethanol concentrations i.e. root mean square error of prediction (RMSEP). Three calibration models were compared: PLS model built on μ_a, PLS model built on the theoretical values of the corrected (i.e. particle free) bulk absorption coefficient $\mu_{a,corr}$ which was computed using Mie theory and PLS model built on $\hat{\mu}_{a,corr}$ which is extracted using the proposed methodology (i.e. an estimate of $\mu_{a,corr}$). The leave-one-out cross validation method was used to obtain the RMSECV statistics.

Experiment

The dataset of the previous four-component experiment[12] was used in this study. The model multi-component system comprised water, deuterium oxide, ethanol and polystyrene particles. Five particle sizes of mean diameters 100, 200, 300, 430 and 500 nm, five particle concentrations centred around 1, 2, 3, 4 and 5 %wt. and five concentrations of ethanol centred around 2, 4, 6, 8 and 10 %wt. were used to build a dataset consisting of a total of 44 samples. These were prepared using various combinations of the concentrations of the components and particle sizes in a manner to ensure that in the resulting dataset the correlation between ethanol concentration and the other components was negligible. For a detailed design of it (i.e. design of experiments), the measurement setup, the materials and the method of extraction of optical properties refer to the previous work.[12] Calibration was carried out as described in ref. 11.

Because the absorption of the mixtures of the absorbing-only species used in the four-component experiment was not equal to a linear sum of absorption coefficients of the pure components,[12] i.e. the spectra of the mixtures could not be modelled as some linear combination of the spectra of the pure species in this case, Principal Component Analysis had to be used for building **A**, which is required for finding the projection matrix **P**$_r$. The dataset consisting of spectra of mixtures of absorbing-only species (water, deuterium oxide and ethanol) required for the Principal Component decomposition was collected from the samples prepared according to the same design of experiment but without polystyrene particles

10

i.e. without the scattering species in them. The absorption was measured by measuring the collimated transmittance using the same setup. The number of principal components for finding the projection matrix was determined from cross validation. The cross validation was carried out using the leave-one-out method.

RESULTS AND DISCUSSION

Simulation

By carrying out the first correction step i.e. decoupling of absorption and scattering, which is referred to as the partial correction here, we remove the main part of undesirable particle effects in highly scattering media i.e. the nonlinear multiple scattering effects. The feasibility of getting any actual improvement by doing the second correction step i.e. applying the full correction, in the presence of signal noise and inversion error, is the primary goal of the simulation.

The simulated dataset closely resembles the four component experimental dataset but the size parameter and the concentration of particles vary more in the simulated dataset. The two optical properties (μ_s and g) related to scattering are shown in figure 3. As one can notice the values of μ_s are very high due to high concentrations of particles and large size parameters. In real life to extract the optical properties from highly concentrated suspensions using the inverse adding doubling method we would have to either dilute it if we are to use total diffuse transmittance (T_t), total diffuse reflectance (R_t) and collimated transmittance (T_c) measurements or use alternative measurements such as diffuse reflectance at several angles. The bulk absorption coefficient and the bulk absorption coefficient corrected using the proposed methodology are shown in figure 4 (only 1500-1880 nm wavelength range is shown in this figure, because the peaks of ethanol and polystyrene that we are interested in appear in this region). The peak in the absorption band 1650-1720 nm, which can be clearly seen in μ_a, is the nonlinearly varying peak of μ_{ap} i.e. polystyrene. As one can notice this peak disappeared in $\mu_{a,corr}$ i.e. it has been removed in the second correction step.

The comparison of the RMSECV curves of the three calibration models namely: PLS model built on μ_a, PLS model built on the theoretical values of the corrected bulk absorption coefficient spectra $\mu_{a,corr}$ and PLS model built on the $\mu_{a,corr}$ estimate ($\hat{\mu}_{a,corr}$) extracted using the proposed methodology is presented in figure 5. The cross validation results show that the performance of PLS calibration model built on $\hat{\mu}_{a,corr}$ is slightly better than the performance of PLS model built on uncorrected μ_a. Approximately the same RMSECV values were achieved with one less latent variable. The calculated RMSEP statistics, given in the table 1, confirmed the same. The conclusion can be drawn that some improvement can be achieved using the proposed full correction approach at the noise levels in the extracted optical properties similar to the one used in the simulation (the absolute error was generated using normal distribution with 0 mean and standard deviation of 0.005). It is interesting to note that if the peak of μ_{ap} happens not to overlap with the peaks of the component of interest, in this case ethanol, than we could just throw away the wavelength region where the nonlinearly varying μ_{ap} peak is, from the calibration. But, if they do overlap, as it is in this case, this approach is not adequate.

Experiment

The full correction methodology was tested on the four-component experimental dataset collected in the previous work.[12] The raw measurements of total transmittance ($\mathbf{T_t}$), total reflectance ($\mathbf{R_t}$) and collimated transmittance ($\mathbf{T_c}$) and the extracted bulk absorption coefficient of the samples have been given in ref. 11. The other two optical properties of the samples namely, μ_s and $\mathbf{g}$, are given here in figure 6. As one can notice from the $\mathbf{g}$ values the inversion of the RTE was not converging to the solution for few samples. The reason was the error in the $\mathbf{T_t}$ and $\mathbf{R_t}$ measurements at the lowest concentrations of particles due to light losses through the sides of the measurement cell and the error in the measurement of the optical depth (i.e. $\mathbf{T_c}$) at the highest concentration of particles due to strong multiple light scattering. Because the extracted values of anisotropy factor $\mathbf{g}$ had a relatively large error/noise a smaller weight (0.8) was used for it in the objective function eq. (7), whereas the weight for μ_a was increased (5) since its values were presumably most accurate i.e. least sensitive to the errors in

12

the measurements. The collected spectral data of mixtures of the absorbing only species used for obtaining the projection matrix $\mathbf{P}_r$ is shown in figure 7. It actually represents the absorption spectra that we should get if particle effects were completely removed i.e. the fully corrected bulk absorption coefficient $\boldsymbol{\mu}_{a,corr}$ that was obtained using Mie theory. This dataset should presumably give the best prediction performance since it is free from nonlinear particle effects and can be used as a benchmark. Based on the cross-validation results it was decided that five principle components should be enough (they explain 99% of variation in the data) to model any spectrum of a mixture of absorbing only species quite accurately.

The spectra of the extracted bulk absorption coefficient, the corrected bulk absorption coefficient and the estimated absorption coefficient of polystyrene particles obtained using the inverse Mie routine with the objective function (8) are illustrated in the figure 8. The corrected spectra of $\boldsymbol{\mu}_a$ for each sample were obtained by subtracting the estimate of $\boldsymbol{\mu}_{ap}$ from the corresponding spectra of the bulk absorption coefficient of the samples. Four calibration models were compared: PLS model built on the total diffuse reflectance pre-processed using EMSCL, PLS model built on $\boldsymbol{\mu}_a$, PLS model built on the non-scattering dataset, which represents the fully corrected $\boldsymbol{\mu}_a$ i.e. $\boldsymbol{\mu}_{a,corr}$ and PLS model built on $\hat{\boldsymbol{\mu}}_{a,corr}$ extracted using the proposed methodology. The leave-one-out cross validation method was used to obtain the RMSECV statistics. RMSECV curves for all four cases are presented in figure 9. The cross validation results show that despite the errors associated with the measurement set up and those in the PSD obtained by inversion of $\boldsymbol{\mu}_s$ using Mie theory, the performance of PLS calibration model built on $\boldsymbol{\mu}_{a,corr}$ is slightly better than the performance of PLS model built on uncorrected $\boldsymbol{\mu}_a$. The black line marks the best possible prediction performance which would be achieved if the full correction was without error. The difference between the blue line and the black line represents the potential for further improvement of the full correction method. Cross validation results are summarized in table 2. It is important to note that the availability of accurate values of the complex refractive index of particles is crucial for the full correction of multiple scattering effects.

13

CONCLUSIONS

The application of the proposed methodology of full correction of multiple scattering effects on the simulated data showed that the performance of PLS calibration model built on the corrected bulk absorption coefficient $\hat{\mu}_{a,corr}$ is slightly better than the performance of PLS model built on uncorrected μ_a. Approximately the same RMSECV values were achieved with one less latent variable. The calculated RMSEP statistics (for the simulated dataset) confirmed the same. Thus, results indicated that improvement can be achieved using the full correction approach as long as the noise levels in the extracted optical properties are not high. The application of the full correction methodology on the experimental data showed that despite measurement noise and inversion errors the performance of PLS calibration model built on $\hat{\mu}_{a,corr}$ was slightly better than the performance of PLS model built on uncorrected μ_a. Finally, the benchmarking analysis revealed that there is still a significant potential for an improvement in the prediction performance in the quantitative analysis of turbid samples. This approach has the potential to significantly reduce calibration efforts as well as leading to more robust models when one of the species is particulate in nature.

REFERENCES

1. S. N. Thennadil, E. B. Martin. "Empirical preprocessing methods and their impact on NIR calibrations: a simulation study". J. Chemom. 2005. 19(2): 77.

2. Z. P. Chen, J. Morris, E. Martin. "Extracting chemical information from spectral data with multiplicative light scattering effects by optical path-length estimation and correction". Anal. Chem. 2006. 78(22): 7674.

3. S. N. Thennadil, H. Martens, A. Kohler. "Physics-based multiplicative scatter correction approaches for improving the performance of calibration models". Appl. Spectrosc. 2006. 60(3): 315.

4. A. Rinnan, F. van den Berg, S. B. Engelsen. "Review of the most common pre-processing techniques for near-infrared spectra". TrAc, Trends Anal. Chem. 2009. 28(10): 1201.

5. S. Ottestad, T. Isaksson, W. Saeys, J. P. Wold. "Scattering Correction by Use of a Priori Information". Appl. Spectrosc. 2010. 64(7): 795.

6. J.-W. Jin, Z.-P. Chen, L.-M. Li, R. Steponavicius, S. N. Thennadil, J. Yang, R.-Q. Yu. "Quantitative Spectroscopic Analysis of Heterogeneous Mixtures: The Correction of Multiplicative Effects Caused by Variations in Physical Properties of Samples". Anal. Chem. 2012. 84(1): 320.

7. L. Leonardi, D. H. Burns. "Quantitative measurements in scattering media: Photon time-of-flight analysis with analytical descriptors". Appl. Spectrosc. 1999. 53(6): 628.

8. C. Abrahamsson, J. Johansson, S. Andersson-Engels, S. Svanberg, S. Folestad. "Time-resolved NIR spectroscopy for quantitative analysis of intact pharmaceutical tablets". Anal. Chem. 2005. 77(4): 1055.

9. J. Chaiken, J. Goodisman. "On probing human fingertips in vivo using near-infrared light: model calculations". J. Biomed. Opt. 2010. 15(3): 037007.

10. M. N. Leger. "Alleviating the Effects of Light Scattering in Multivariate Calibration of Near-Infrared Spectra by Path Length Distribution Correction". Appl. Spectrosc. 2010. 64(3): 245.

11. R. Steponavicius, S. N. Thennadil. "Extraction of Chemical Information of Suspensions Using Radiative Transfer Theory to Remove Multiple Scattering Effects: Application to a Model Two-Component System". Anal. Chem. 2009. 81(18): 7713.

12. R. Steponavicius, S. N. Thennadil. "Extraction of Chemical Information of Suspensions Using Radiative Transfer Theory To Remove Multiple Scattering Effects: Application to a Model Multicomponent System". Anal. Chem. 2011. 83(6): 1931.

13. C. Pariset, S. N. Thennadil. "Evaluation of cross-validation formulas for choosing the regularization parameter for inversion of Fredholm integral of the first kind with non-negativity constraints". The 3rd International Conference on Sensing Technology, 2008.

14. C. Pariset, S. N. Thennadil. "Inversion of spectroscopic measurements for extracting particle size distribution". The 6th International Multi-Conference on Systems, Signals and Devices, 2009.

15. A. Brandolin, L. H. Garcia-Rubio, T. Provder, M. E. Koehler, C. Kuo. "Latex particle size distribution from turbidimetry using inversion techniques; experimental validation". Polymeric Materials Science and Engineering, Proceedings of the ACS Division of Polymeric Materials Science and Engineering, 1990.

16. J. He, S. M. Wang, J. C. Cheng, S. Y. Zhang. "Inversion of particle size distribution from light scattering spectrum". Inverse Probl. 1996. 12(5): 633.

17. G. L. Frontini, E. M. Fernández Berdaguer. "Inversion of elastic light scattering measurements to determine refractive index and particle size distribution of polymeric emulsions". Inverse Probl. Eng. 2003. 11(4): 329.

18. D. M. Scott. "Characterizing particle characterization". Part. Part. Syst. Charact. 2003. 20(5): 305.

19. M. F. Carfora, F. Esposito, and C. Serio. "Numerical methods for retrieving aerosol size distributions from optical measurements of solar radiation". J. Aerosol Sci. 1998. 29(10): 1225.

20. C. F. Bohren, D. R. Huffman. Absorption and scattering of light by small particles. New York, USA: Wiley, 1983.

LIST OF FIGURES

Table 1. Performance of calibration models for estimating the concentration of absorbing only species (i.e. ethanol) in the simulated data-set of the four-component system.

Predictions of concentration of absorbing only species (ethanol)				
Dataset which PLS was built on	Pre-processing	LVs	Calibration RMSECV (% vol.)	Test RMSEP (% vol.)
Calibration model built on μ_a	None	4	0.18	0.17
Calibration model built on $\hat{\mu}_{a,corr}$	None	3	0.17	0.17
Calibration model built on $\mu_{a,corr}$	None	3	0.13	0.12

Table 2. Performance of calibration models for estimating the concentration of absorbing only species (i.e. ethanol) in the experimental data-set of the four-component system.

Dataset which PLS was built on	Pre-processing	LVs	RMSECV (% vol.)
$\mathbf{R_t}$	EMSCL	7	0.79
$\boldsymbol{\mu_a}$	None	7	0.61
$\hat{\boldsymbol{\mu}}_{a,corr}$	None	7	0.51
$\boldsymbol{\mu}_{a,corr}$	None	5	0.26

Figure 1. Methodology for full correction of multiple scattering effects.
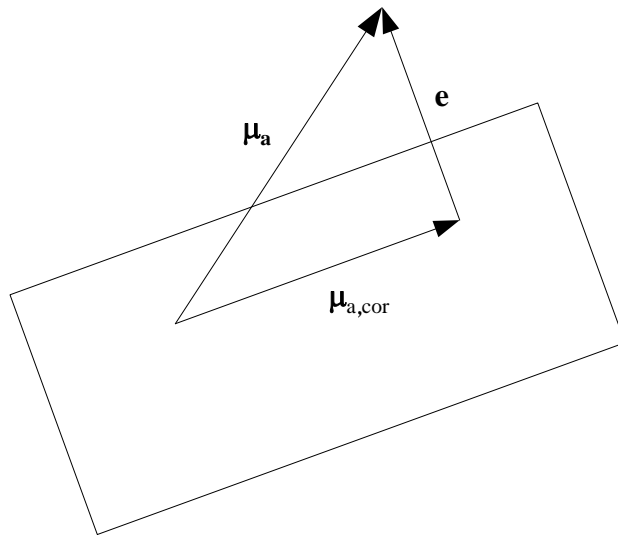
Figure 2. Projection on to the plane defined by two absorbing only components.
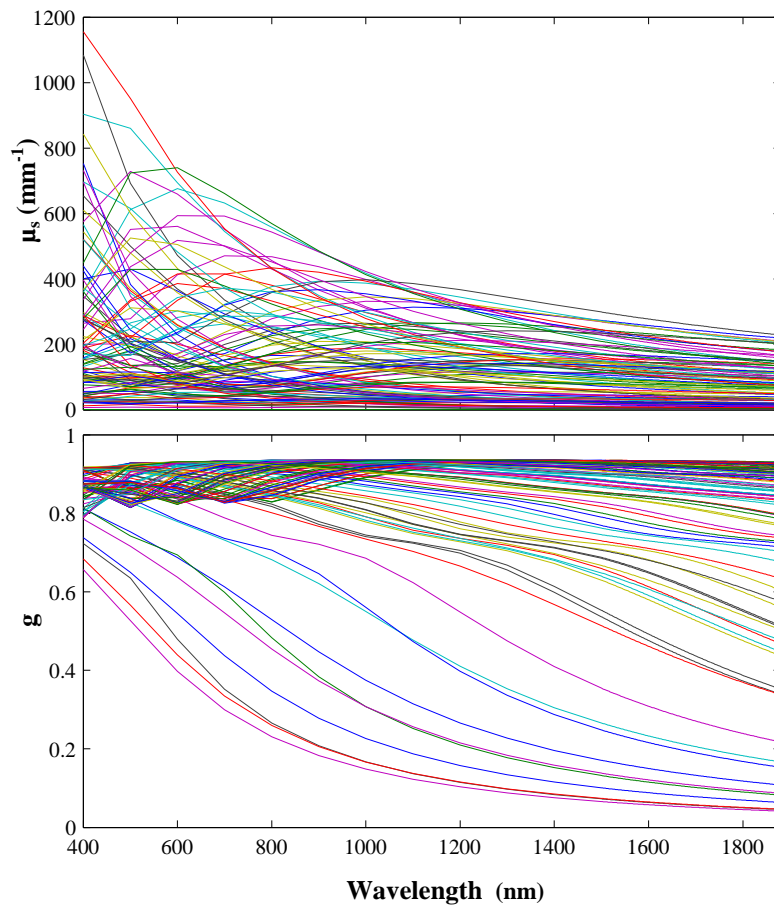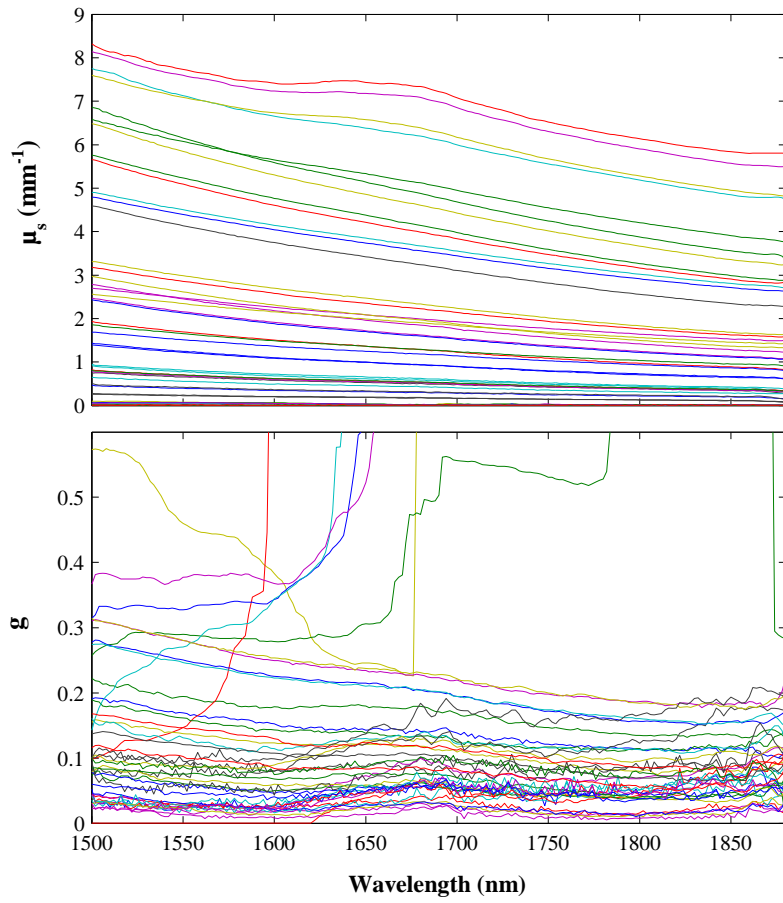

Figure 3. Simulated values of μ_s and **g**.
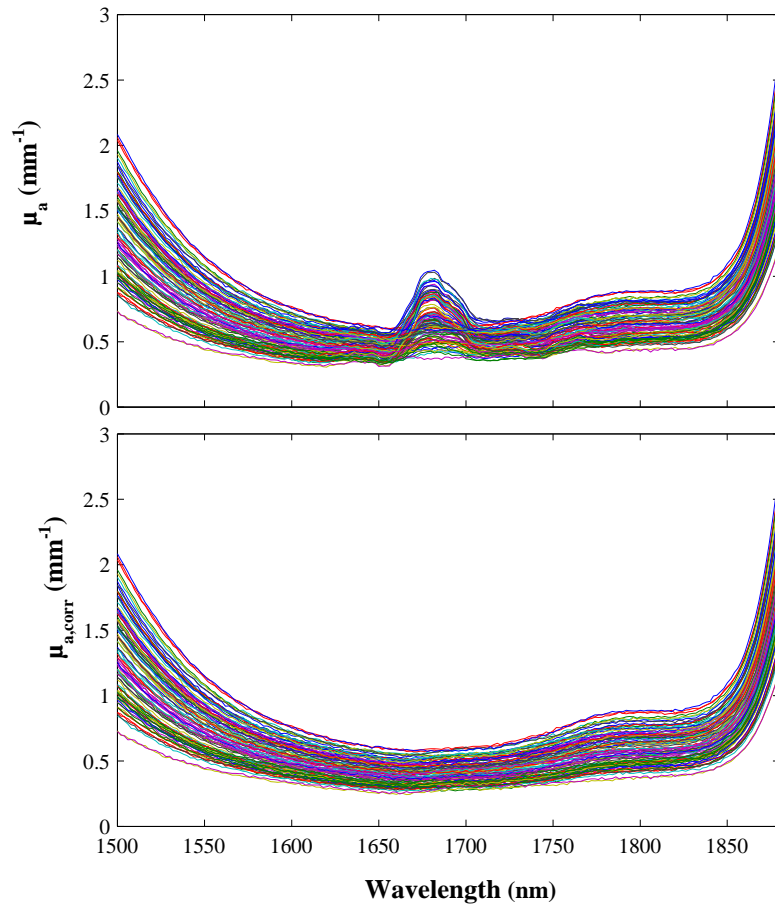
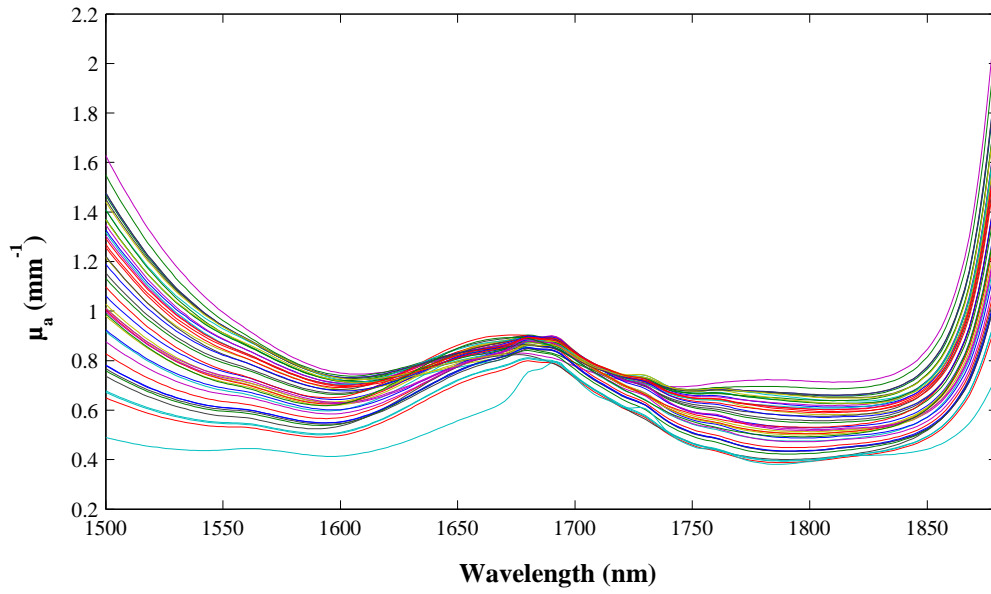**Figure 4.** Uncorrected and corrected bulk absorption coefficients (simulation).
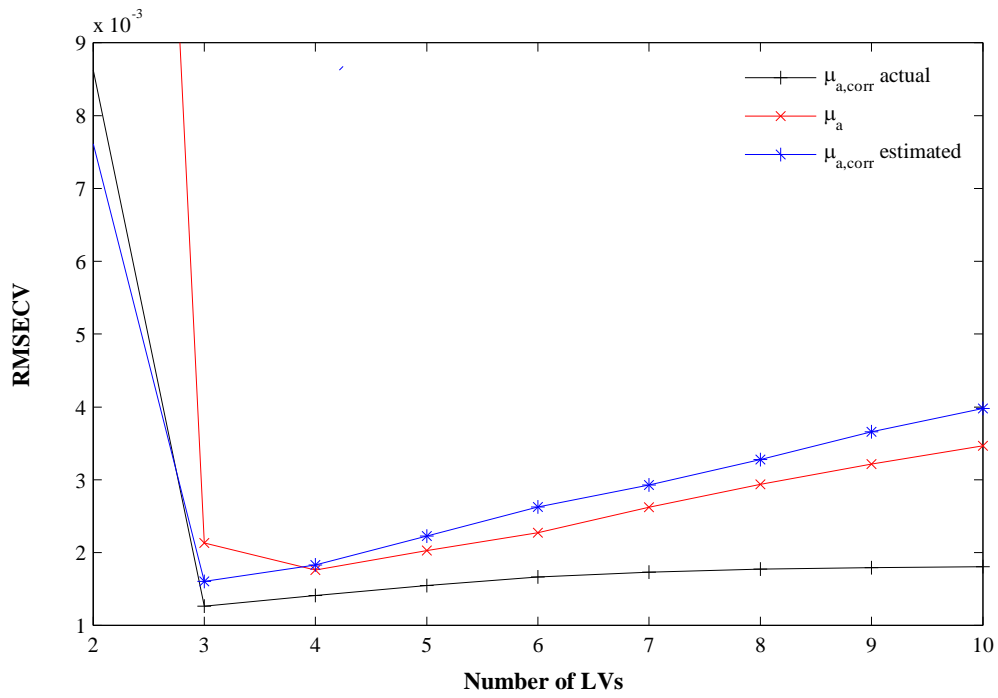
Figure 5. RMSECV curves for bulk absorption coefficients: not corrected μ_a, corrected theoretical (actual) $\mu_{a,corr}$ and the estimated corrected value $\hat{\mu}_{a,corr}$.
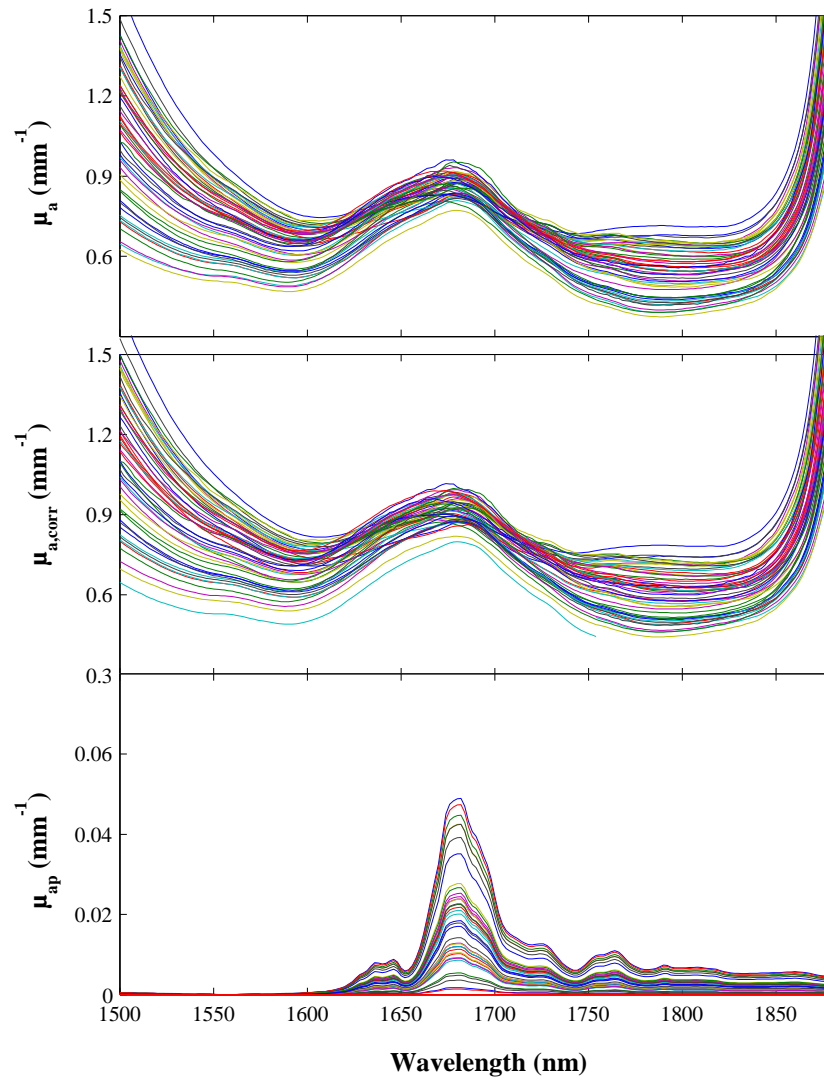
Figure 6. μ_s and **g** extracted from the experimental dataset.

Figure 7. Absorption spectra of the set of non-scattering samples representing the same samples as in the four component experiment, but without the scattering component (i.e. polystyrene particles).

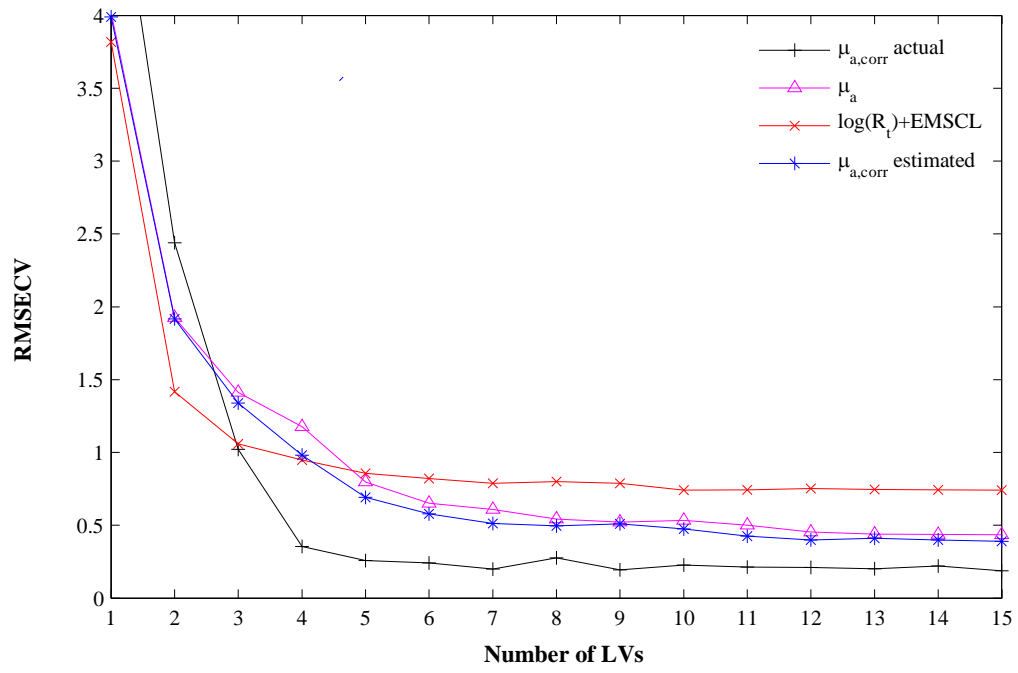**Figure 8.** Spectra of estimated absorption coefficient of polystyrene particles.

Figure 9. RMSECV curves for $\mathbf{R_t}$, $\boldsymbol{\mu_a}$, $\boldsymbol{\mu_{a,corr}}$ and $\hat{\boldsymbol{\mu}}_{a,corr}$.