Extracting physical information of turbid suspensions using multiplicative scatter correction

Yi-chieh Chen* and Suresh N. Thennadil

Department of Chemical and Process Engineering, University of Strathclyde, Glasgow G1 1XJ, UK.
*Corresponding author: yichieh.chen@strath.ac.uk

Introduction

Multiplicative scatter correction (MSC) has been applied extensively to near infrared (NIR) spectroscopy on particulate systems in which large spectral variation is usually observed as a result of multiple light scattering effects.[1] The MSC equation employs an additive (a) and multiplicative (b) term to account for the scatter contribution on a spectrum with respect to a reference spectrum usually taken from the mean spectra of the calibration data set.[2] The corrected spectrum is considered to differ from the reference spectrum by the difference in the chemical contribution in the spectra. This correction is expected to improve the performance of multivariate calibration models such as Partial Least Squares (PLS) regression which are built on the corrected spectra to predict concentration of unknown samples. The scatter correction parameters (such as a and b in MSC) are normally discarded. These can be potentially valuable as they contain information related to the physical nature (diameter or shape) of the particulate species.

Recently, we proposed a method that utilises full information from scatter correction methods to extract physical information (particle diameter) of a particulate system in addition to the chemical information (concentration) obtained through the use of a PLS calibration model built on the corrected data set.[3] The approach was demonstrated using a two-component system of polystyrene (PS) particles suspension and cross-validated using a leave-one-out method. In the method proposed, the predicted concentration is used as an input for estimating particle diameter. PLS calibration models built on two-component systems tend to enhance the prediction due to the correlation of the concentrations between water and particles. In most applications, the concentration of each analyte varies independently and it is therefore important to investigate the performance of the method in multiple component systems. In this work, the method proposed is examined by application to a data set of a multi-component system which consists of PS particles, ethanol, water and heavy water. The same scatter correction method (MSC) and type of empirical expressions to relate MSC parameters to particle diameters and concentrations are employed so that the effectiveness of the method can be evaluated with minimal interference from the use of different scatter correction approaches and expressions.

Materials and Methods

Samples and NIR Measurements

Diffuse transmittance data of a four component system (PS particles, ethanol, water and deuterium oxide) previously described[4] was used in this study. The data set contained a total of 45 samples with the ethanol concentration, particle concentration and particle size varying for each sample. Ethanol concentration spanned a range between 2 and 10 wt% in 2 wt% intervals while the particle concentrations (y) ranged from 1 to 5 wt% at 1 wt% intervals. Five particle sizes of mean diameters (d_p) 100, 200, 300, 430 and 500nm were used. This data set was chosen as it fitted the purpose of this study i.e. that the correlations between PS particle concentrations and the other components were negligible.

Total diffuse transmittance spectra were taken from the samples using a Cary 5000 spectrometer equipped with an external diffuse reflectance accessory (DRA-2500). For each spectrum, a 10 s integrating time was used and a wavelength range (1550 to1850 nm at 2 nm intervals) was chosen, resulting in 150 discrete wavelengths per spectrum. The data collected was converted into absorbance units.

Methodology

Details of the methodology were previously described[3]. In brief, MSC was applied to the mean centred diffuse transmittance spectra with the mean spectrum of the data set taken as the reference spectrum . MSC corrected spectra ($\boldsymbol{x}_{corr}$) were then subjected to PLS regression to build a calibration model for predicting PS particle concentration ($\hat{y}$). The model was cross-validated using a leave-one-out method; one of the 45 samples was chosen in turns as the test sample to predict $\hat{y}$ from the calibration model built on the rest of the data set. Relationships between MSC parameters (a and b) and particle concentrations and diameters were investigated in order to develop empirical expressions to predict particle diameter ($\hat{d}_p$).To demonstrate the

Reference paper as:
Y. Chen and S.N. Thennadil (2012).Extracting physical information of turbid suspensions using multiplicative scatter correction,
in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 127-129.

effectiveness of the method in a multiple component system, only the investigation of MSC parameter *b* will be presented in this paper. MSC parameter *b* was related to particle diameters using a third order polynomial while the expressions for relating MSC parameters to particle concentration were chosen using a second order polynomial. The overall expression was then chosen in a format similar to that used in the previous study for the two component system[3] and is given as:

$$b = \xi_b + \kappa_1 y + \kappa_2 y^2 \quad 1 + \eta_1 d_p + \eta_2 d_p^2 + \eta_3 d_p^3 \tag{1}$$

MSC parameter *b* for one of the 45 samples was chosen in turns as the test data and the rest of *b* were used to estimate coefficients in (1). These estimated coefficients were then combined with *b* of the left-out sample and its $\hat{y}$ predicted from the PLS calibration model to predict the particle diameter ($\hat{d}_p$).Eq. (1) suggests three possible solutions for d_p but only those with positive real numbers will be taken. The three roots were obtained using the 'roots' function in MATLAB.

Results and Discussion

Fig. 1(a) shows the changes in MSC parameter *b* with polystyrene particle concentration (*y*). For each particle diameter, the relationship between *b* and *y* was fitted using a second order polynomial, as the curves overlaid on the data points. Similarly, the changes in *b* with d_p is given in Fig. 1(b) with curves generated from a third order polynomial fit for each concentration. For some sample conditions, there are multiple samples containing the same particle concentration and diameter with the contents of the other components being different. If MSC is effective in correcting the scattering effects, similar MSC parameters can be expected from the samples of the same particle diameter and concentration since they are the only variables to contribute the scattering effect. It can be seen however that *b* is not always similar for the same PS conditions, such as in the case of *y* of 4 % and d_p of 500 nm in Fig. 1(a). This could suggest that the simple form of MSC is not as effective in removing scattering effects in multiple component systems as in two component systems. This is expected to introduce errors in $\boldsymbol{x}_{corr}$ and MSC parameters as MSC may not decouple physical and chemical information effectively, leading to the degradation of the calibration model for both concentration and diameter predictions.

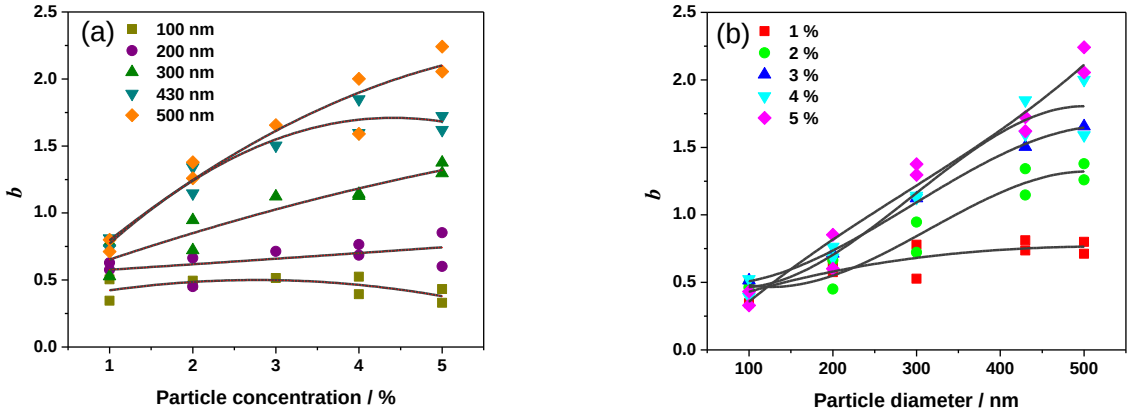

Figure 1. Changes in MSC parameter *b* with (a) particle concentration and (b) particle diameter. Solid curves are generated from the best fits using second order and third order polynomial for (a) and (b), respectively.

The PLS calibration model built on $\boldsymbol{x}_{corr}$ was cross-validated using leave-one-out method as mentioned above; 4 latent variables (LVs) were selected, resulting in values for RMSECV of 0.886 % and R^2 of 0.664. Predicted versus actual polystyrene concentration was plotted in Fig. 2(a). The RMSECV from the model on MSC-corrected spectra is about twice the value from the EMSCL pre-processed spectra but is similar to that of the model built on the data without any pre-processing. However, the number of LVs used for the MSC-corrected data set is less than those reported (LVs = 7).

Results for prediction of particle diameters are shown in Fig. 2(b), giving a RMSECV of 80 nm and R^2 of 0.826. For comparison, the calibration model built directly on $\boldsymbol{x}_{corr}$ for the same purpose resulted in values of 79 nm (RMSECV) and 0.716 (R^2). The similarity in model performance could be a result of MSC scatter correction not being effective. Both MSC parameters and $\boldsymbol{x}_{corr}$ contain similar level of the scattering information, leading to similar prediction performance. It is however worth noting that, as the error in $\hat{d}_p$ predicted from the MSC parameter also accounts for the error in $\hat{y}$, $\hat{d}_p$ prediction will benefit from any

Reference paper as:
Y. Chen and S.N. Thennadil (2012).Extracting physical information of turbid suspensions using multiplicative scatter correction,
in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 127-129.

improvement made in the estimation of concentration. This can be achieved by using different multiplicative correction approaches (such as EMSCL) or by applying different pre-processing techniques. Also, the empirical expressions for relating MSC parameters and scattering-related variables can be fine-tuned to minimise errors contained in the coefficients. The method examined in this study provides a generic platform that allows for such a change to be incorporated easily.

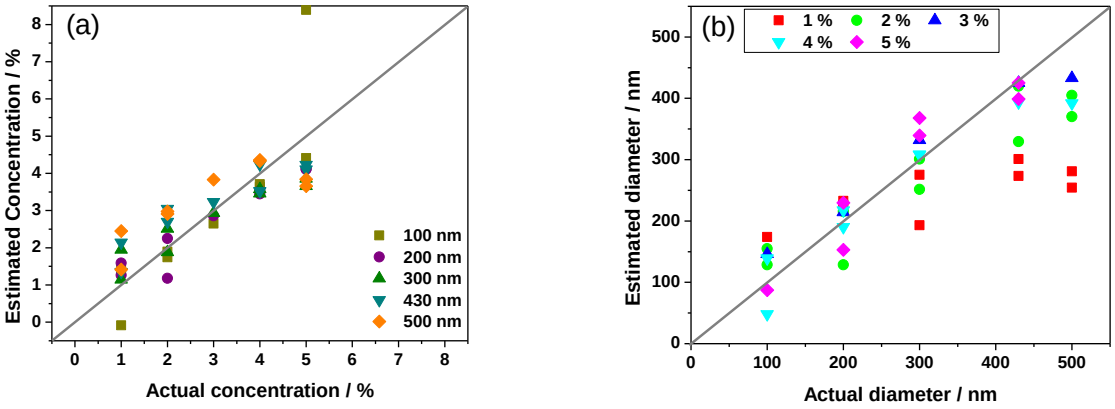

Figure 2. (a) Estimated versus actual concentration from PLS model on $\boldsymbol{x}_{corr}$ for predicting particle concentration. (b) Predicted versus actual particle diameter using the method that combines particle concentration prediction in (a) and MSC parameter *b*.

The results highlight the importance of choosing an effective scatter correction method to separate scattering and absorption information so that the accuracy of predicting both concentration and particle diameter can be improved. For cases in which considerable scattering-related information is contained in $\boldsymbol{x}_{corr}$,as seen in this study, $\hat{d}_p$ estimated using the method utilising MSC parameters only provides similar levels of accuracy as that using the model built directly on $\boldsymbol{x}_{corr}$. As improvement on predicting $\hat{y}$ can be made via the use of effective scatter correction method, a better estimation in $\hat{d}_p$ can be expected as a result of the scatter correction parameters containing more scattering related information and more accurate $\hat{y}$ being used as the input. On the contrary, it would be a trade-off choice to estimate both $\hat{y}$ and $\hat{d}_p$ from $\boldsymbol{x}_{corr}$ since the accuracy of $\hat{d}_p$ could suffer from the improvement made to enhance the absorption information for concentration prediction.

Conclusion

We studied the use of a new methodology for extracting both particle diameter and concentration information from NIR spectra of a multi-component system. The effectiveness of the method was compared to the results reported in the literature and the benchmark method that predicts *y* and d_p from the multivariate calibration models built on the spectra. Similar levels of the performance were found and further refinement to improve the performance of the model for this specific data set is suggested via the use of different scatter correction methods or expressions relating the scatter correction parameters to the sample conditions.

References

1. A. Rinnan, F.v.d. Berg, and S.B. Engelsen, *Trends Anal. Chem.* **28**, 1201-1222 (2009).
2. P. Geladi, D. MacDougall, and H. Martens, *Appl. Spectrosc.* **39**, 491-500 (1985).
3. Y.-C. Chen and S.N. Thennadil, *Anal. Chim. Acta* **746**, 37-46 (2012).
4. R. Steponavicius and S.N. Thennadil, *Anal. Chem.* **83**, 1931-1937 (2011).

Reference paper as:
Y. Chen and S.N. Thennadil (2012).Extracting physical information of turbid suspensions using multiplicative scatter correction,
in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy,
Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 127-129.