Analytica Chimica Acta 746 (2012) 37–46

Contents lists available at SciVerse ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Insights into information contained in multiplicative scatter correction parameters and the potential for estimating particle size from these parameters

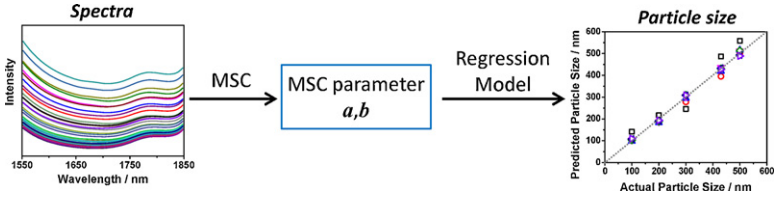
Yi-Chieh Chen, Suresh N. Thennadil*

Department of Chemical and Process Engineering, University of Strathclyde, Glasgow, United Kingdom

HIGHLIGHTS

- Indepth analysis of information contained in MSC parameters.
- MSC parameters vary systematically with particle size and concentration.
- A novel method to extract the mean particle diameter from the MSC parameters.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 23 March 2012
Received in revised form 9 July 2012
Accepted 6 August 2012
Available online 14 August 2012

Keywords:
Multiplicative scatter correction
Light scattering
Particle size estimation
Extended multiplicative scatter correction
Near infrared spectroscopy
Preprocessing

ABSTRACT

This paper investigates the nature of information contained in scatter correction parameters. The study had two objectives. The first objective was to examine the nature and extent of information contained in scatter correction parameters. The second objective is to examine whether this information can be effectively extracted by proposing a method to obtain particularly the mean particle diameter from the scatter correction parameters. By using a combination of experimental data and simulated data generated using fundamental light propagation theory, a deeper and more fundamental insight of what information is removed by the multiplicative scatter correction (MSC) method is obtained. It was found that the MSC parameters are strongly influenced not only by particle size but also by particle concentration as well as refractive index of the medium. The possibility of extracting particle size information in addition to particle concentration was considered by proposing a two-step method which was tested using a 2-component and 4-component data set. This method can in principle, be used in conjunction with any scatter correction technique provided that the scatter correction parameters exhibit a systematic dependence with respect to particle size and concentration. It was found that the approach which uses the MSC parameters gave a better estimate of the particle diameter compared to using partial least squares (PLS) regression for the 2-component data. For the 4 component data it was found that PLS regression gave better results but further examination indicated this was due to chance correlations of the particle diameter with the two of the absorbing species in the mixture.

1. Introduction

Multivariate calibration methods such as partial least squares (PLS) regression have been widely used to build calibration models for predicting the concentrations of chemical components from near-infrared (NIR) spectra. When samples containing particles are encountered, multiple light scattering effects introduce nonlinearities leading to degradation in model performance. Several empirical preprocessing methods such as multiplicative scatter correction (MSC), standard normal variate (SNV), extended multiplicative scatter correction (EMSC), orthogonal signal correction (OSC), and optical path length estimation and correction (OPLEC) have been used to mitigate light scattering effects [1–6]. It is generally assumed that the information removed from the measured spectra by the application of these empirical methods is essentially the manifestation of the underlying physics of light

* Corresponding author at: 75 Montrose Street, James Weir Building, Chemical and Process Engineering, University of Strathclyde, Glasgow, G1 1XJ, United Kingdom. Tel.: +44 141 548 2241; fax: +44 141 548 2539.
E-mail address: suresh.thennadil@strath.ac.uk (S.N. Thennadil).

0003-2670/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.aca.2012.08.006

scattering without significant loss of chemical information, thus improving the performance of the multivariate regression models in estimating chemical information from the corrected spectra.

When these methods are used, the parameters that are applied for correcting the spectra are normally discarded since they are supposed to contain only physical information. If the scatter correction method were effective, the scatter correction parameters would be expected to contain information regarding the particulate species since it is this component which contributes to the light scattering effects. If this information can be extracted then it could provide valuable extra information (particle size) in addition to estimates of concentrations which are obtained from the calibration models built on the scatter-corrected spectra.

Several studies can be found in the literature where scatter correction techniques are applied and compared in terms of the improvement in performance of models built using the corrected spectra. However, the performances of the empirical methods appear to be dependent on the system studied with no single empirical scatter correction method consistently outperforming others across a number of different types of datasets. Among the empirical methods, the more recently developed OPLEC method has been promising [6,7], though it has not yet been applied widely enough to conclude that the method is indeed consistently superior to other available methods. A study based on simulations using a rigorous light propagation model indicated that most of the common scatter correction methods led to similar model performances [8]. In addition, this study also indicated that the effectiveness of a particular scatter correction technique was also dependent on measurement configuration. To-date however, to our knowledge, there have been no in-depth studies that have examined the information contained in the scatter correction parameters themselves. Such a study will be useful for understanding the nature and characteristics of information contained in the parameters of a particular scatter correction method. This could help in identifying situations where they perform the best and could potentially help in modifying the methods to produce more effective scatter correction techniques.

The implicit assumption when applying scatter correction methods is that light scattering effects manifesting as an additive or multiplicative or more complex (e.g. wavelength dependent) effects in the measured spectra are removed. However, there are other non-chemical effects which can lead to similar manifestations in the spectra as the assumed effect of light scattering (e.g. instrument drift). In other words, the corrections are not necessarily specific to scattering. Hence the terms multiplicative signal correction and extended multiplicative signal correction can sometimes be found in the literature where "signal" is used instead of "scatter" to denote that the techniques are more general in terms of the non-chemical information removed by them [5]. Similarly, the SNV method is clearly a general method which has also been used to correct light scattering effects.

In any dataset consisting of spectroscopic measurements of particulate systems, we can expect the non-chemical variations to be a combination of effects with the light scattering effects usually being the most dominant. There are several possibilities why one scatter correction technique might work better than others: (1) the method removes the most amount of variation due to light scattering compared to others; (2) the method removes the most amount of variation due to all non-chemical effects present in the measurements; (3) the method linearizes the measurements most effectively compared to other methods; (4) the method removes the least amount of relevant chemical information; and (5) the method is the most effective in terms of a combination of the previous four aspects. Therefore the most effective "scatter correction" method will differ from one system to another depending on the dominant type of non-chemical variations in the measurements that form the datasets.

This study had two objectives. The first objective was to examine the nature and extent of information contained in scatter correction parameters. The second objective is to examine whether this information can be effectively extracted by proposing a method to obtain particularly the particle size from the scatter correction parameters. The approach used for this investigation is to examine the scatter correction parameters in terms of the information regarding particle size and particle concentration by using a dataset in which particle size and particle concentration vary significantly and where the values of these parameters have been accurately measured. Since particle concentration and size are the two sample parameters that affect the extent of light scattering by a sample, it follows that any effective correction step will contain information regarding these two sample parameters. Following this logic, if the scatter correction step is effective, then it should be possible to extract information regarding particle size and/or particle concentrations from the scatter correction parameters. This is investigated through an approach for building models to obtain particle size information using the scatter correction parameters. The investigation into the effectiveness of the scatter correction approach to specifically provide information regarding particle size was carried out using two models systems namely, a two component and a four component system both containing polystyrene latex particles as the scattering species.

2. Materials and methods

2.1. Experimental dataset

The two datasets used in this study were obtained from previously published works [9,10]. A brief description of the datasets is given here. Both datasets contain measurements taken using a Cary 5000 spectrometer equipped with an external diffuse reflectance accessory and 1 mm sample thickness was chosen. The first dataset is a polystyrene–water system that consists of a total of 35 samples with 5 particle diameters (d_p = 100, 200, 300, 430 and 500 nm) and 7 particle concentrations (y = 0.1, 0.5, 0.9, 1.23, 1.6, 1.95 and 2.3 in wt.%) for each particle size [9]. Spectra were collected using 0.4 s as integrating time for a wavelength range of λ = 1550–1850 nm with 4 nm interval, resulting in 75 discrete wavelengths per spectrum. The raw spectra were smoothed using Savitsky–Golay filter with window width of 9 and polynomial order of 3 to remove noise in the measurements which was the procedure followed in Ref. 9 where it was found that Savitsky–Golay filter with these parameters were most effective from the point of calibration model performance.

The second dataset is a 4-components system that consists of water (H_2O), deuterium oxide (D_2O), ethanol (C_2H_5OH), and polystyrene particles [10]. The concentration of each component was varied so that the correlation between concentration of polystyrene particles and other components in the sample is negligible. In this dataset there are samples containing the same particle diameter and particle concentration while concentrations for other components vary. 5 particle diameters (d_p = 100, 200, 300, 430 and 500 nm) and 5 concentrations (y = 1–5 in wt.%) were employed to form this dataset of 45 samples. Spectra were collected in the range of λ = 1500–1880 nm with 2 nm intervals and 10 s as the integrating time. The same smoothing conditions applied to the first dataset were also employed for this dataset before subjecting to scatter correction methods. Both datasets contained measurements from three different measurement configurations namely, total reflectance (Rd), total transmittance (Td) and collimated transmittance (Tc). The spectral region covers the first overtone region of organic compounds widely found in industrial processes. For