

PAT: THE EXTRACTION OF MAXIMUM INFORMATION FROM MESSY SPECTRAL DATA

Zeng-Ping Chen and Julian Morris

*Centre for Process Analytics and Control Technology
School of Chemical Engineering and Advanced Materials
University of Newcastle upon Tyne, NE1 7RU, UK*

Abstract: With the increasing take-up of PAT¹ by the pharma- and bio- industries there is a critical need for robust spectral calibrations for processes which are subject to the variations in physical properties such as sample compactness, surface topology, etc. The variation in the optical path-length materializing from the physical differences between samples may result in multiplicative light scattering influencing spectra in a nonlinear manner leading to the poor calibration performance. A new approach “Optical Path Length Estimation and Correction” overcomes the limitations of existing light scattering correction methods. *Copyright 2007 IFAC*

Keywords: Chemical Processes, Multiplicative Light Scattering, Near Infrared Spectroscopy

1. INTRODUCTION

Bio- and pharmaceutical development and production are now being heavily influenced through the recent FDA PAT¹ initiative and spectroscopic instrumentation is now being widely applied to or at the very least explored for on-line real-time applications. This drives the urgent need to incorporate and integrate the detailed spectral information into process performance monitoring schemes. The enhancement of spectroscopic data analysis and calibration algorithms and software thus becomes even more important if PAT is to be widely applied and accepted. Spectroscopic technologies such as near infrared spectroscopy has been widely applied in areas of food technology, agriculture, pharmaceuticals, etc (Siesler, et al. 2002), due to their high measuring speed and fewer or no sample preparation requirements which make them highly suitable for in-line process monitoring. Generally, it is the chemical information (in most cases the concentrations of the chemical or biological

compounds) inherent within the spectroscopic measurements, rather than the spectroscopic measurements themselves, that are used for efficient management and optimization as well as for quality control. Calibration models are, therefore, needed to extract the desired concentration information from abundant spectroscopic measurements. To achieve these aims two areas of complexity need to be addressed.

The first relates to the variations in external process variables which can have different impact on different chemical species in mixture samples. For example, fluctuations in temperature will provoke non-linear shift and broadening in spectral bands of absorptivity spectra of constituents in mixture samples via the changes in intermolecular forces. Such temperature-induced non-linear spectral variations will have detrimental effects on predictive performance of multivariate calibration model, if not being properly taken into account when developing the model. Chen et al (2004) proposed a new approach termed Individual Contribution Standardization (ICS) to eliminate the temperature effects on the predictive abilities of calibration

¹ FDA, PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, <http://www.fda.gov/cder/guidance>, 2004.

models for white chemical system. In ICS, it was assumed that the absorbance of each chemical species in every wavelength follows simple monotonic smooth nonlinear function with respect to temperature. ICS had the advantages of providing an enhanced performance with ease of implementation. Furthermore, it did not require the training samples measured at all training temperatures to be exactly the same. Unfortunately, since ICS was specially designed for white chemical systems, it could not be applied to grey chemical systems. In order to overcome these limitations, Chen et al (2005) generalized the ideas behind ICS and designed a Loading Space Standardization (LSS) approach to correct temperature-induced spectral variations for grey chemical systems.

Another area of complexity concerns the issues arising when analyzing more or less intact complex samples by spectroscopic instruments where the uncontrolled variations in optical path length due to the physical variations of samples such as particle size and shape, sample packing and sample surface may cause dominant multiplicative light scattering perturbations which will mask the spectral variations related to the content differences of chemical compounds in samples. The effects of multiplicative light scattering are difficult to approximate with linear factor combinations during calibration process. Hence, without being properly pre-processed, spectral data dominated by multiplicative light scattering effects cannot be explicitly modelled by any of the current popular bilinear calibration methods such as PCA (Cowe, et al. 1985) and PLS (Martens, et al. 2001). Though there are several methods available for correcting the effects of Multiplicative Light Scattering - MSC (Geladi, et al. 1985), ISC (Helland, et al. 1995), EMSC1 (Martens, et al. 1991), EISC1 (Pedersen, et al. 2002), EMSC2 and EISC2 (Martens, et al. 2003), they all have various stringent underlying requirements on the spectral data which are quite difficult to satisfy in practice. For example, pre-processing spectral data by MSC, ISC and EISC1 is safe only if the effects of chemical variations among samples are negligible. The success of EMSC1, EMSC2 and EISC2 mainly depends on the availability of the pure spectra of every chemical constituent in samples and the consistency of spectral contributions from constituents in mixtures with those of isolated constituents in pure state. Therefore, methods which can correct the effects of multiplicative light scattering for systems with little or even no prior chemical knowledge are therefore highly desirable.

The aim of this study is to introduce a new multiplicative light scattering correction method, called Optical Path Length Estimation and Correction [OPLEC] (Chen, et al. 2006), without any requirement on prior chemical knowledge, and to compare its performance with that of EISC1 and EMSC2 for near infrared transmittance spectra of mixtures of powders of wheat gluten and starch. Powder mixing, along with grinding and blending, are critical unit operations in pharmaceuticals

manufacturing in order to ensure ideal particle size distribution of the active ingredient.

2. MULTIPLICATIVE LIGHT SCATTERING

For I transparent solutions comprising J absorbing chemical components, when the cuvette width is kept constant during each measurement, according to Beer-Lambert law, the theoretical absorbance spectrum ($\mathbf{x}_{i,Chem}$, row vector) of sample i is a linear combination of the absorbance contributions of all the J constituents.

$$\mathbf{x}_{i,Chem} = \sum_{j=1}^J c_{i,j} \mathbf{s}_j, \quad i = 1, 2, \dots, I \quad (1)$$

Where c_{ij} is the concentration, and row vector \mathbf{s}_j , the absorptivity spectrum of j th constituent in sample i . Assume \mathbf{s}_j ($j = 1, 2, \dots, J$) are linearly independent of each other, the multivariate linear calibration model build between $\mathbf{x}_{i,Chem}$ and c_{ij} ($i = 1, 2, \dots, I$) can provide satisfactory predictions for concentrations of constituent j in future solution samples.

If the samples to be analyzed are solid samples (powder, granules) as well as emulsions and dispersions, it is practically quite difficult to make the optical path length constant over samples. For relatively simple systems, the effects of light scattering caused by the changes of optical path length due to the physical variations of samples can be approximated by the following EMSC2 model (Martens, et al. 2003).

$$\mathbf{x}_i = a_i \mathbf{1}_r + \sum_{j=1}^J b_j c_{i,j} \mathbf{s}_j + d_i \lambda + e_i \lambda^2 + \boldsymbol{\varepsilon}_i \quad (2)$$

Where \mathbf{x}_i is the measured absorbance spectra of sample i . $\mathbf{1}_r$ is a row vector with all its elements equal to one. Coefficients a_i and b_j stand for additive and multiplicative effects of light scattering due to the physical variations of sample i relative to a reference sample. Coefficients d_i and e_i are introduced to account for smoothly wavelength dependent spectral variations from sample to sample. $\boldsymbol{\varepsilon}_i$ represents unknown and irrelevant types of spectral variations. Without being pre-processed by appropriate methods, the relationship between the measured absorbance spectra \mathbf{x}_i and c_{ij} ($i = 1, 2, \dots, I$) can not be fully explicitly modeled by any of the current popular multivariate linear calibration methods such as PCA and PLS.

2.1 Optical Path Length Estimation and Correction

Though the models used in existing multiplicative light scattering correction methods are slightly different from each other (Helland, et al. 1995), most of them can be seen as simplification or modification versions of the above EMSC2 model. They generally share the same parameter estimation and spectral correction procedures:

- (i) construct a regressor matrix based on the model employed

- (ii) estimate model coefficients (such as a_i , b_i , d_i and e_i in Eq.2) by least squares regression and regress onto the regressor matrix
- (iii) insert the estimated coefficients into the model and remove the effects of multiplicative light scattering to yield corrected spectrum $x_{i,Corrected}$ ($x_{i,Corrected} \approx x_{i,Chem}$) with only absorbance contributions from chemical variations.

The parameter estimation and spectral correction of conventional methods heavily relies on the availability of the regressor matrix such as the pure spectra of every chemical constituent in samples for EMSC2, which hinders their application in practice.

However, if the aim of removing the effects of multiplicative light scattering is to build a robust and accurate calibration model, instead of all the model parameters, only parameter b_i which carries the information about multiplicative effects of optical path length variation is needed. The following section will focus on how to estimate b_i for both calibration samples and how to use it for prediction under the circumstance of no prior information about the pure spectra of constituents in the samples.

The influence of baseline offset (a_i) and smoothly wavelength-dependent spectra variation from sample to sample (d_i and e_i) in Eq.2 can be removed by projecting the measured spectrum \mathbf{x}_i onto the orthogonal complement of the space spanned by the row vectors of $\mathbf{P} = [\mathbf{1}_r; \lambda; \lambda^2]$.

$$\mathbf{z}_i = \mathbf{x}_i(\mathbf{I} - \mathbf{P}^+\mathbf{P}) = \sum_{j=1}^J b_i c_{i,j} \mathbf{k}_j + \boldsymbol{\varepsilon}_i^* \quad (3)$$

$$\mathbf{k}_j = \mathbf{s}_j(\mathbf{I} - \mathbf{P}^+\mathbf{P}), \quad \boldsymbol{\varepsilon}_i^* = \boldsymbol{\varepsilon}_i(\mathbf{I} - \mathbf{P}^+\mathbf{P})$$

Suppose the first component is the target constituent in the mixtures and $\sum_{j=1}^J c_{i,j} = 1$ (which strictly holds for $c_{i,j}$ representing unit-free concentration such as weight fraction and mole fraction), then Eq.3 can also be expressed as:

$$\mathbf{z}_i = b_i c_{i,1} \Delta \mathbf{k}_1 + b_i \mathbf{k}_2 + \sum_{j=3}^J b_i c_{i,j} \Delta \mathbf{k}_j + \boldsymbol{\varepsilon}_i^* \quad (4)$$

$$\Delta \mathbf{k}_j = \mathbf{k}_j - \mathbf{k}_2$$

Given the concentration vector $\mathbf{c}_1 = [c_{1,1}; \dots; c_{i,1}; \dots; c_{I,1}]$ of the target constituent in the calibration samples, column vector $\mathbf{b} = [b_1; \dots; b_i; \dots; b_I]$ can be obtained by the following procedure even if the pure spectra ($\mathbf{s}_j, j = 1, 2, \dots, J$) are unavailable.

Suppose \mathbf{Z}_{base} is a full-row-rank matrix assembled by J spectra (row vector) deliberately selected from $\mathbf{Z} = [\mathbf{z}_1; \dots; \mathbf{z}_i; \dots; \mathbf{z}_I]$ and \mathbf{Z}_{rest} is composed of the rest spectra (rows) in \mathbf{Z} . \mathbf{b}_{base} , \mathbf{b}_{rest} , $\mathbf{c}_{1,base}$ and $\mathbf{c}_{1,rest}$ consist of the corresponding elements of \mathbf{b} and \mathbf{c}_1 with the same values of index i as the spectra in \mathbf{Z}_{base} and \mathbf{Z}_{rest} , respectively. It is obvious that each row in \mathbf{Z}_{rest} can be expressed as a linear combination of the rows in \mathbf{Z}_{base} .

$$\mathbf{Z}_{rest} = \mathbf{A} \mathbf{Z}_{base} \quad (5)$$

$$\mathbf{A} = \mathbf{Z}_{rest} \mathbf{Z}_{base}' (\mathbf{Z}_{base} \mathbf{Z}_{base}')^{-1}$$

According to Eq.4, there are linear relationships between \mathbf{z}_i and b_i , and also between \mathbf{z}_i and $b_i c_{i,1}$.

Therefore, the following equations hold.

$$\mathbf{b}_{rest} = \mathbf{A} \mathbf{b}_{base} \quad (6)$$

$$diag(\mathbf{c}_{1,rest}) \mathbf{b}_{rest} = \mathbf{A} diag(\mathbf{c}_{1,base}) \mathbf{b}_{base} \quad (7)$$

Where, $diag(\mathbf{c}_{1,rest})$ denotes the diagonal matrix in which the corresponding diagonal elements are elements of $\mathbf{c}_{1,rest}$. Inserting eq.6 into eq.7 yields:

$$diag(\mathbf{c}_{1,rest}) \mathbf{A} \mathbf{b}_{base} = \mathbf{A} diag(\mathbf{c}_{1,base}) \mathbf{b}_{base} \quad (8)$$

Since there is no need to know the absolute value of b_i , the first element of \mathbf{b}_{base} can be assigned to take the value of 1. The rest elements of \mathbf{b}_{base} can be calculated by nonnegative least squares regression. Given \mathbf{b}_{base} , \mathbf{b}_{rest} can then be estimated out according to the following equation.

$$\mathbf{b}_{rest} = \{diag(\mathbf{c}_{1,rest}) + \mathbf{I}\}^{-1} \mathbf{A} \{diag(\mathbf{c}_{1,base}) + \mathbf{I}\} \mathbf{b}_{base} \quad (9)$$

The estimation of \mathbf{b} is obtained by rearranging the elements of \mathbf{b}_{base} and \mathbf{b}_{rest} in appropriate order.

Due to the existence of noise and possible interferences, different selections of \mathbf{Z}_{base} may produce more or less different estimations of \mathbf{b} . Therefore, in this contribution, a set of matrixes $\mathbf{Z}_{i,base}$ ($i = 1, 2, \dots, I$) are constructed by first selecting the i -th spectrum \mathbf{z}_i from \mathbf{Z} and then sequentially adding new spectrum which carries the most new spectral information not contained in spectra already selected, until the number of spectra in $\mathbf{Z}_{i,base}$ is equal to J (the number of chemical components in samples). Each $\mathbf{Z}_{i,base}$ produces an estimation of \mathbf{b} (\mathbf{b}_j). The average over all \mathbf{b}_j 's ($i = 1, 2, \dots, I$) provides a good estimation of \mathbf{b} .

With \mathbf{b} and \mathbf{c}_1 available the following two calibration models can be built by multivariate linear calibration methods such as PLS.

$$diag(\mathbf{c}_1) \mathbf{b} = [\mathbf{1}_c, \mathbf{Z}] \boldsymbol{\beta}_1, \quad \mathbf{b} = [\mathbf{1}_c, \mathbf{Z}] \boldsymbol{\beta}_2 \quad (10)$$

Where $\mathbf{1}_c$ is a column vector with its elements equal to unity. The two estimated regression vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ can be used to correct the effects of multiplicative light scattering on the concentration predictions of the target constituent in any test samples.

$$\mathbf{z}_{test} = \mathbf{x}_{test} (\mathbf{I} - \mathbf{P}^+\mathbf{P}), \quad b_{test} c_{1,test} = [\mathbf{1}_c, \mathbf{z}_{test}] \boldsymbol{\beta}_1 \quad (11)$$

$$b_{test} = [\mathbf{1}_c, \mathbf{z}_{test}] \boldsymbol{\beta}_2, \quad c_{1,test} = \frac{[\mathbf{1}_c, \mathbf{z}_{test}] \boldsymbol{\beta}_1}{[\mathbf{1}_c, \mathbf{z}_{test}] \boldsymbol{\beta}_2}$$

3. NIR IN POWDER MIXING

The data consists of 100 near-infrared transmittance spectra of five mixtures of gluten and starch powder with different weight ratios (1:0, 0.75:0.25, 0.5:0.5,

0.25:0.75 and 0:1)². For each of the five powder mixtures, five samples were randomly taken and filled loosely into five different glass cuvettes. Two consecutive transmittance spectra were recorded for each sample. After that, each loosely packed sample was packed more firmly, and another two consecutive transmittance spectra were measured. Hence there are 100 spectra in total. Each of the 100 transmittance spectra was transformed into absorbance spectra. The 100 absorbances between 850nm and 1050nm formed the basis of the data analysis. 60 spectra from the three mixtures with gluten/starch equal to 1:0, 0.5:0.5 and 0:1 were used to construct the calibration set. The test set was composed of the remaining 40 spectra from the other two mixtures. For more experimental details, readers are referred to the original paper of Martens, et al. (2003).

For the powder mixture data, partial least squares (PLS) regression was used to build the calibration models between the mean-centred concentrations of the target constituent (gluten) in the samples and the corresponding mean-centred raw or pre-processed near-infrared spectra. The root mean square error of prediction for independent test set ($RMSEP_{test}$) is used as performance criterion to assess the predictive power of the PLS models. All pre-processing and multivariate calibrations were carried out on a Pentium PC using MATLAB Version 6.5.

4. RESULTS AND DISCUSSION

The 100 near-infrared absorbance spectra of five mixtures of gluten and starch powder in 20 replicates are displayed in Fig.1. From the standpoint of calibration, it is expected to see five bunches of spectra. However, due to multiplicative light scattering effects caused by the changes in optical path length, the 20 spectra from the same mixture are quite different from each other. PLS was used to build the calibration model between the concentrations of the gluten in powder mixtures and the corresponding raw spectra. Both the leave-one-out cross validation root mean square error of prediction for the calibration set and the root mean square error of prediction for the independent test set suggest that PLS model with 9 components can provide the best predictions for the raw data.

Figure 2 shows the predictive results of the 9-component PLS model for the raw spectra. It can be seen that the predictions of the optimal PLS model are not as good as expected. $RMSEP_{test}$ of the optimal PLS model is 0.024, which is equivalent to a relative predictive error of 6.08%. It clearly demonstrated that even when the number of PLS components used is sufficiently large, PLS still can not fully model raw spectra masked by multiplicative light scattering. With a view to separating the spectral variations caused by multiplicative light scattering due to the changes in physical properties

of samples from those contributed by chemical or biological constituents, OPLEC, EISC1 and EMSC2 were applied to pre-process the raw spectra plotted in Fig.1.

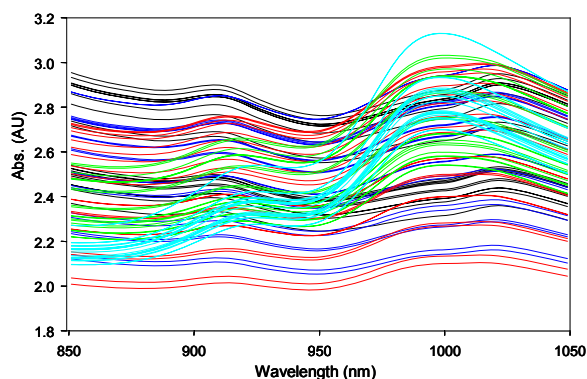


Fig.1. The raw absorbance spectra of the five mixtures of gluten and starch powder with different weight ratios (Black lines: 1:0; blue lines: 0.75:0.25; red lines: 0.5:0.5; green lines: 0.25:0.75; cyan lines: 0:1)

The number of spectra used in constructing the matrices $Z_{i,base}$ in OPLEC is set as two, the number of chemical constituents in the powder mixtures. As suggested in the original paper of Martens, et al (2003), spectra No.3 and No.93 were used in EMSC2 as the pure spectra of gluten and starch, respectively. To investigate the possible influence of the choice of pure spectra on the performance of EMSC2, the mean spectra of 20 replicates of pure gluten samples and pure starch samples were also considered as input pure spectra. For the convenience of comparison, both sets of pure spectra were normalized. Hereafter, EMSC2 with the two different choices of input pure spectra will be simply referred as $EMSC2_{3,93}$ and $EMSC2_{mean}$.

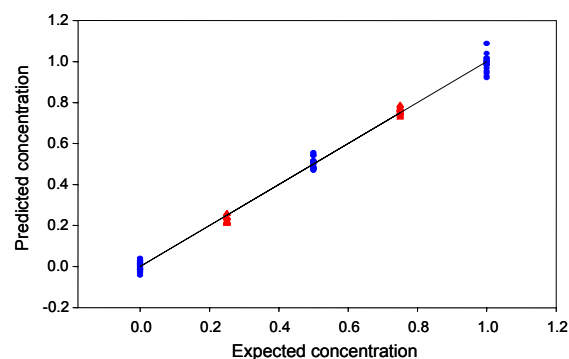


Fig.2. Predictive results of a 9-component PLS model built on raw calibration spectra. Diagonal line: theoretically correct predictions, Blue circle: leave-one-out cross validation predictions for calibration set, Red triangle: predictions for independent test data set

Figure 3 shows the spectra pre-processed by OPLEC, $EMSC2_{3,93}$, $EMSC2_{mean}$ and EISC1, respectively. The spectra pre-processed by all the four methods exhibit five distinct spectral patterns for five powder mixtures. The 20 replicates of each mixture are more or less indistinguishable. However, it does not

² Prof H. Martens, MATFORSK/Norwegian Food Research Institute is acknowledged for providing the powder mixture data

necessarily mean that all the methods provided satisfactory results.

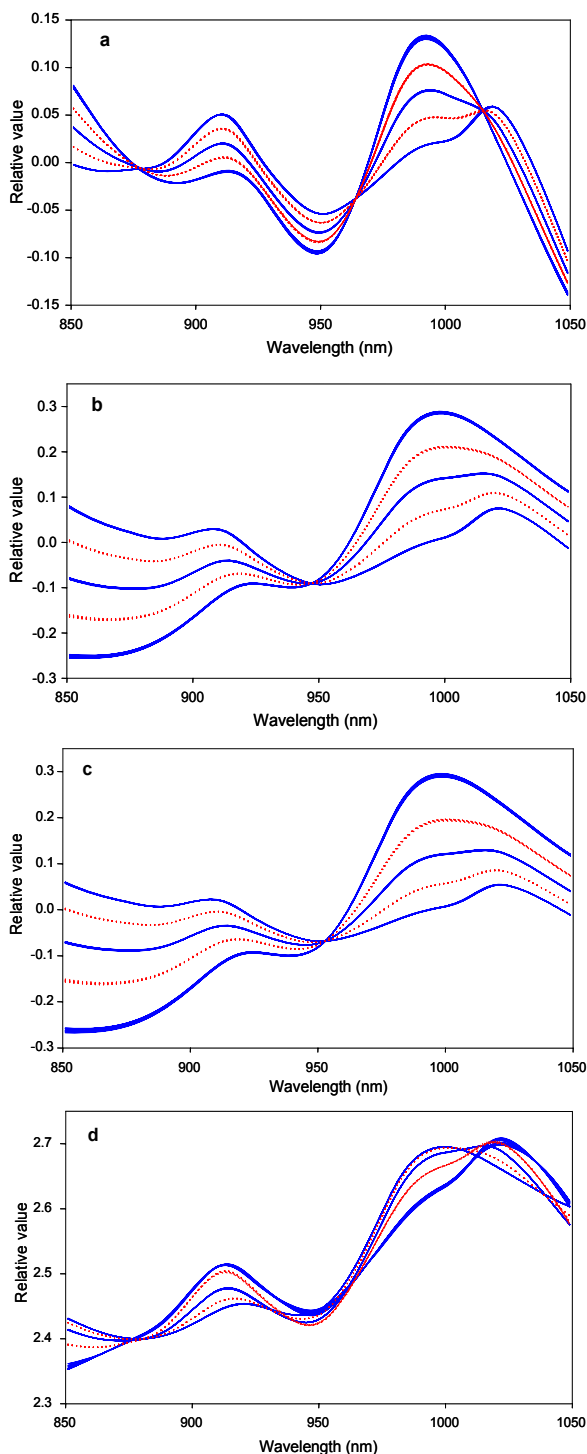


Fig.3. The calibration (blue solid lines) and test (red dotted lines) spectra of powder mixtures pre-processed by different methods: a) OPLEC b) EMSC_{2,3,93}, c) EMSC_{2,mean}, d) EISC1

From the calibration point of view, the variations of the five spectral patterns should also correctly reflect the variations of the constituents' concentrations in the five powder mixtures. From Fig.3, it can be seen that the spectra pre-processed by OPLEC and EMSC_{2,3,93} maintain the expected equal spaces between every two neighbouring spectral patterns. Though the results of EMSC_{2,mean} are quite similar to those of EMSC_{2,3,93}, the more or less unequal spaces

between every two neighbouring spectral patterns reveal that the different choice of input pure spectra affects the performance of EMSC₂. In contrast, the spectra pre-processed by EISC1 do not carry the right information about the constituents' concentrations in the powder mixtures. The significantly unequal spaces between every two neighbour spectral patterns obtained by EISC1 indicate that either some of the chemical information has been wrongly removed along with the effects of multiplicative light scattering or the effects of multiplicative light scattering has not been effectively corrected. Either case will affect the performance of the calibration models built on the pre-processed spectra.

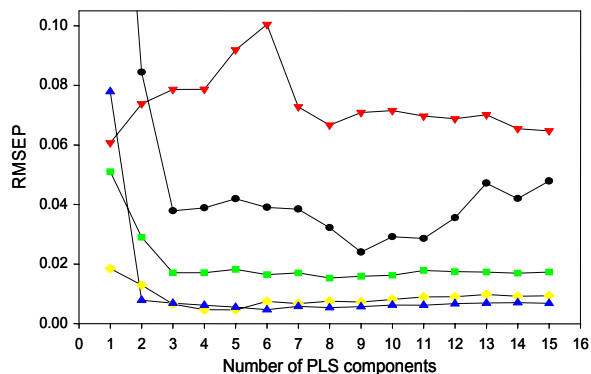


Fig.4. The predictive performance of the PLS models (black dots); calibration spectra pre-processed by OPLEC (blue upward triangle, EMSC_{2,3,93} (yellow diamond), EMSC_{2,mean} (green square) and EISC1 (red downward triangle)

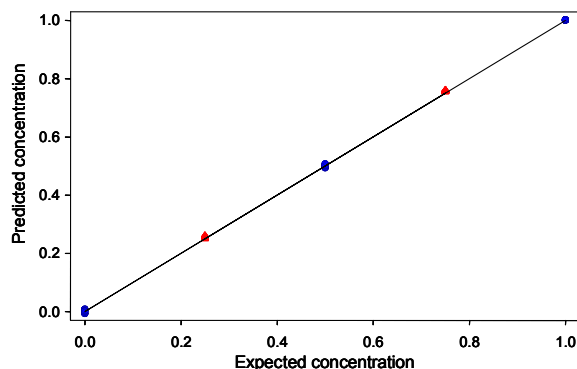


Fig.5. Predictive results of 2-component PLS model pre-processed by OPLEC. Diagonal line: theoretically correct predictions; predictions for calibration set (Blue dots); predictions for independent test set (Red triangle)

Figure 4 confirms the above conclusions. PLS models built on the calibration spectra pre-processed by EISC1 gave unacceptable predictions with errors even larger than those of the PLS models established on the raw calibration spectra. The failure of EISC1 on this powder mixture data suggests that like its predecessor, ISC, it is not suitable for samples with significant spectral variations due to the changes in chemical compositions. It is observed that EMSC₂ can greatly improve the predictive accuracy of the PLS models. However, the significant difference

between the predictions provided by PLS models on pre-processed calibration spectra of EMSC_{2,mean} and EMSC_{2,93} further demonstrates that the choice of input pure spectra is crucial for better performance. The application of OPLEC offers equivalent, if not better, improvement in the predictive ability of the PLS models as EMSC_{2,93}. Both methods attained the same level of minimal RMSEP_{test} (0.005) at a slightly different number of PLS components. It is worth stating that after being pre-processed by OPLEC, a 2-component PLS model can provide excellent predictive results (Fig.5) with RMSEP_{test} equal to 0.008 which is equivalent to relative errors as small as 1.7%, while the corresponding RMSEP_{test} of EMSC_{2,93} is 0.013, i.e., 2.5% in terms of relative error. Considering the fact that OPLEC does not require the pure spectra of the chemical constituents in the samples, as does EMSC_{2,93}, these results are very encouraging.

5. CONCLUSIONS

It has been shown that without using any prior chemical knowledge, the proposed OPLEC algorithm is able to separate the physical light scattering effects from spectral variations related to chemical constituents. This significantly improves the prediction accuracy of calibration models, and in the particular the example presented here for powder mixture data. Compared with other existing multiplicative light scattering correction methods which can be used only when the pure spectra of all chemical constituents in the samples are available, or the effects of chemical variations among the spectra are negligible, OPLEC places no special requirement on spectral data. The new approach has been demonstrated to have significant potential and has a much wider PAT application domain than existing methods. The final paper will also provide applications of the new algorithms to batch cooling crystallization.

6. ACKNOWLEDGEMENTS

The authors acknowledge the financial support of the EPSRC grant GR/R19366/01 (KNOW-HOW) and GR/R43853/01 (Chemicals Behaving Badly II).

REFERENCES

- Chen, Z.P., J. Morris and E. Martin (2004). Modeling temperature-induced spectral variations in chemical process monitoring. *IFAC DYCOPS*, Boston.
- Chen, Z.P., J. Morris and E. Martin (2005). Correction of temperature-induced spectral variations by loading space standardization. *Anal. Chem.*, **77**, pp. 1376-1384.
- Chen, Z.P., J. Morris, and E. Martin (2006). Extracting chemical information from spectral data with multiplicative light scattering effects by optical path-length estimation and correction. *Anal. Chem.*, in press.
- Cowe, I. and J.W. McNicoi (1985). The use of principal components in the analysis of near-infrared spectra. *Appl. Spectrosc.*, **39**, pp. 257-266.
- Geladi, P., D. McDougall and H. Martens (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.*, **39**, pp. 491-500.
- Helland, I.S., T. Næs and T. Isaksson (1995). Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data. *Chemom. Intell. Lab. Syst.*, **29**, pp. 233-241.
- Martens, H. and E. Stark (1991). Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. *J. Pharm Biomed Anal.*, **9**, pp. 625-35.
- Martens, H. and M. Martens (2001). Multivariate analysis of quality: an introduction. John Wiley and Sons, Chichester.
- Martens, H., J.P. Nielsen and S.B. Engelsen (2003). Light scattering and light absorbance separated by extended multiplicative signal correction: application to near-infrared transmission analysis of powder mixtures. *Anal. Chem.*, **75**, pp. 394-404.
- Pedersen, D.K., H. Martens, J.P. Nielsen and S.B. Engelsen (2002). Near-infrared absorption and scattering separated by extended inverted signal correction (EISC): analysis of near-Infrared transmittance spectra of single wheat seeds. *Appl. Spectrosc.*, **56**, pp. 1206-1214.
- Siesler, H.W., Y. Ozaki, S. Kawata and H.M. Heise (2002). Near-infrared spectroscopy: principal, instruments, applications. WILEY-VCH, Weinheim.