



HAL
open science

PAT and compensating for non-linear effects in process spectroscopic data for improved process monitoring and control ((Research Highlight))

Julian Morris, Zeng-Ping Chen

► To cite this version:

Julian Morris, Zeng-Ping Chen. PAT and compensating for non-linear effects in process spectroscopic data for improved process monitoring and control ((Research Highlight)). *Biotechnology Journal*, 2009, 4 (5), pp.610-n/a. 10.1002/biot.200800314 . hal-00484125

HAL Id: hal-00484125

<https://hal.science/hal-00484125>

Submitted on 18 May 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



PAT and compensating for non-linear effects in process spectroscopic data for improved process monitoring and control ((Research Highlight))

Journal:	<i>Biotechnology Journal</i>
Manuscript ID:	BIOT-2008-0314.R1
Wiley - Manuscript type:	Other contribution
Date Submitted by the Author:	18-Apr-2009
Complete List of Authors:	Morris, Julian; Technical Director: Centre for Process Analytics and Control Technology, School of Chemical Engineering and Advanced Materials Chen, Zeng-ping; Newcastle University, Chemical Engineering & Advanced Materials



1
2
3 **PAT and Compensating for Non-linear Effects in Process Spectroscopic Data**
4 **for Improved Process Monitoring and Control**
5
6

7 *Zengping Chen and Julian Morris**
8

9 *Centre for Process Analytics and Control Technology*
10

11 *School of Chemical Engineering and Advanced Materials*
12

13 *Newcastle University, NE1 7RU, England, UK*
14
15
16
17
18
19

- 20 1) **Abstract:** Robust fit-for-purpose multivariate calibration models are of critical importance
21 to on-line/in-line quantitative monitoring of bio-chemicals and pharmaceuticals using
22 spectroscopic instruments. Unlike in off-line assays, the spectroscopic measurements in
23 on-line/in-line real-time applications are almost inevitably subjected to variations in
24 measurement conditions (e.g. temperature) and samples' physical properties (e.g. cell
25 density, particle size, sample compactness), which can invalidate the assumption of a linear
26 relationship between the spectroscopic measurements and the concentrations of the target
27 chemical components. This paper discusses the effects of such variations on spectroscopic
28 measurements, and presents an overview of recent work on modelling and correcting of the
29 detrimental effects of variations in measurement conditions and samples' physical
30 properties. A number of application studies to complex data sets and an industrial plant
31 demonstrate the methodologies and algorithms discussed.
32
33
34
35
36
37
38
39
40
41
42
43
44
45

46 **Keywords:** Temperature-induced spectral variations, Multiplicative Light Scattering, Optical
47 Path Length Estimation and Correction, Near Infrared Spectroscopy, Multivariate linear
48 calibration
49
50
51
52
53
54
55
56
57
58
59
60

* Corresponding Author. julian.morris@ncl.ac.uk

Introduction

Pharmaceutical development and production are now being heavily influenced through the FDA PAT initiative¹ with spectroscopic instrumentation now being much more widely applied to or at the very least explored for on-line real-time applications in pharma-chemistry. Vibrational spectroscopic technologies have been widely applied in areas of chemicals processing, food technology and agriculture, etc, because of their high measuring speed and fewer or no sample preparation requirements which make them highly suitable for in-line and on-line process monitoring. In particular, near infrared spectroscopy, Fourier-transform mid infrared spectroscopy and Fourier-transform Raman spectroscopy are now being much more widely applied in pharmaceutical chemistry and in bio-processing in general. For example in the development and production of an Active Pharmaceutical Ingredients (APIs), reaction based intermediates, fermentation, crystallization, downstream processing and separations, granulation, blending and drying, and in the monitoring of raw materials^[1-10]. Indeed there is much that can be learnt from the large numbers of applications of spectroscopy being used for in-line and on-line real-time monitoring outside the pharma-chemistry and bio-pharma industries.

During on-line quantitative monitoring of processes using spectroscopic techniques, it is by and large the chemical and biological information (in most cases the concentrations of the chemical and biological compounds) inherent within the spectroscopic measurements, rather than the spectroscopic measurements themselves, that are needed for use in process monitoring and management, closed loop process control and quality control and process optimisation. Calibration models are therefore needed to transform abundant spectroscopic measurements into the desired concentration information. In most cases, calibration models are based on the assumption of a linear relationship between the spectroscopic measurements and the concentrations of the target chemical or biological components. However, unlike in off-line assays, the spectroscopic measurements in on-line and in-line real-time applications are almost inevitably subjected to variations in measurement conditions, e.g. temperature, as well as changes in samples' physical properties, e.g. cell density, particle size, sample compactness, which can make the assumption of linear relationships invalid.

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

1
2
3 In contrast in bio-pharmaceuticals development and production PAT is in its infancy²,^[11-14]
4
5 with applications seen using predominantly parametric methods that monitor and measure
6
7 surrogate markers of product quality rather than product quality directly. In a recent IFPAC
8
9 presentation and interview³ the following comments were made “Until PAT for
10
11 bio-manufacturing is really pushed by the FDA and the benefit of PAT is translated (and clearly
12
13 understood) into cost, quality and schedule benefits, it will be slow to take-up in the industry.
14
15 Sensor technology limitations are the other big issue. There are only a handful of sensors
16
17 available now and they are surrogate indicators”. “The key differences in how PAT is applied to
18
19 biologics manufacturing, as opposed to conventional pharmaceutical manufacturing are (i) there
20
21 are approximately 15 quality attributes of a biologic (due to their complex structure and complex
22
23 biological function) compared with typically five or so for small molecules; (ii) the fluids in
24
25 which biologics are produced are more complex and contain more ingredients and therefore
26
27 present more interferences to sensors; (iii) several steps in the biological manufacturing
28
29 processes need to be sterilized with heat and many biologics measurement sensors are too
30
31 delicate to withstand sterilization temperatures. Nevertheless, there are several good PAT
32
33 applications in use at, for example at Eli Lilly, Biogen Idec and Baxter Healthcare³”.

Situations where the assumption of a linear relationship can become invalid

Influence of external variables: Variations in measurement conditions such as temperature have different impact on different chemical species in mixture samples. For instance, fluctuations in temperature can provoke non-linear shift and broadening in spectral bands of absorptivity spectra of constituents in mixture samples (Figure 1) by the way of changes in intermolecular forces^[15,16].

Figure 1: Five NIR spectra for a ternary mixture sample measured at five different temperatures (black solid line: 30°C, red dash line: 40°C, green dash-dot line: 50°C, pink dotted line: 60°C, blue dash-dot-dot line: 70°C)

² Shaw, G., “Biopharma says Bye Bye to Black Magic”, Pharma Quality, September 2008.

<http://www.pharmaquality.com/ME2/Audiences/Segments/Publications/Print>.

³ Galliher, P., PAT for Biopharma Slow to Take Up, Process Analytical Technology, 2005.

<http://www.pharmamanufacturing.com/articles/2005/199.html?page=print>

1
2
3 For a transparent solution comprising K absorbing chemical components, the absorbance
4 spectra $\mathbf{x}(t)$ measured at temperature t can be expressed as linear combinations of the absorbance
5 contributions of all the K constituents.
6
7

$$\mathbf{x}(t) = \sum_{k=1}^K c_k \mathbf{s}_k(t) + \mathbf{e} \quad (1)$$

9
10
11
12

13
14 where c_k is the concentration of k -th chemical component in the mixture, and $\mathbf{s}_k(t)$ represents the
15 absorptivity spectrum of k -th chemical component at t and \mathbf{e} is the measurement error. Since the
16 absorptivity spectrum of k -th chemical component at t and \mathbf{e} is the measurement error. Since the
17 absorptivity spectrum $\mathbf{s}_k(t)$ is temperature dependent, using the spectra with temperature-induced
18 non-linear spectral variations in the building of calibration models will result in the ensuing
19 predictions being poor and potentially not fit-for-purpose. Consequently the correction of
20 temperature effects on the spectra is essential for the building of a robust calibration model, even
21 if the data were generated from a design-of-experiments programme. This becomes critical for
22 assured process analysis, process monitoring and ultimately the elucidation of process
23 understanding and in the application of closed loop process control^[29].
24
25
26
27
28
29
30
31
32
33
34
35

36 *Influence of samples' physical properties:* Fermentations and other bioprocesses are multiphase
37 systems including solution phase(s) and solid phase(s) of microbes, filamentous and mammalian
38 cells, etc. The existence of solid phases in samples will confound the extraction of quantitative
39 information from spectroscopic measurements, since the uncontrolled changes in a samples'
40 physical properties, such as cell number, particle size and density of mixtures can result in
41 variations of optical path length and hence the multiplicative light scattering effects which can
42 'scale' the entire spectral measurement (Figure 2).
43
44
45
46
47
48
49
50
51
52

53 Figure 2: Mean-centred NIR spectra of a powder mixture measured under different sample compactness.
54
55
56

57 For relatively simple systems, multiplicative light scattering effects can be formulated as
58 shown in eq.2^[23].
59
60

$$\mathbf{x} = b \sum_{k=1}^K c_k \mathbf{s}_k + \mathbf{d} + \mathbf{e} \quad (2)$$

where, the coefficient b accounts for the multiplicative light scattering effects due to the variations in optical path length caused by the uncontrolled changes in samples' physical properties. The term \mathbf{d} represents spectral variations that have not been explicitly modeled. It can be seen from eq.2 that multiplicative parameter b confounds with the concentrations of chemical species in the samples. Hence, without being properly pre-processed, spectral data contaminated by multiplicative light scattering effects may not be explicitly modelled by any of the current popular bilinear calibration methods.

When spectral measurements are subject to such changes and variations, using the measured spectra in the development of calibration models without taking these effects into account can result in the calibration predictions being poor and problematical in their application. When building calibration models, the spectral variations caused by changes in variables other than the concentrations of constituents in the samples need to be explicitly modelled, or corrected for, in order to maintain an assured predictive ability of the calibration model.

Strategies for maintaining linear relationships in spectroscopic measurements

Correction of temperature-induced spectral variations

For situations with relatively small temperature variations, multivariate linear calibration methods such as PCA and PLS can to some extent implicitly model the temperature effect by treating it as if it were an additional component or interference^[17]. A synthetic model^[18] utilising constant temperature data that is augmented with a classical least-squares estimate of the spectral effect of temperature obtained from variable-temperature aqueous sample spectra also works well for white chemical systems (i.e. fully characterized chemical systems). However, when temperature varies over a large range, both linear implicit and explicit modelling methods are not sufficient. Wülfert et al^[19]. developed continuous piecewise direct standardization (CPDS) for correcting complex temperature-induced nonlinear spectra variations. The idea of CPDS is to find a transformation matrix for each temperature level, which transforms the spectra measured at an arbitrary temperature into the corresponding spectra as if they were measured at the same

1
2
3 temperature as the calibration samples used to build the original calibration model. With a view
4
5 to compensate for the temperature effects on the spectra, Eilers and Marx^[20] have extended
6
7 penalized signal regression (EPSR) to incorporate temperature as a covariate.. It was shown that
8
9 both CPDS and EPSR can effectively correct temperature effects on spectra and maintain the
10
11 predictive abilities of calibration models despite of the fluctuations of temperature. However,
12
13 both CPDS and EPSR involve several meta-parameters. The setting of these meta-parameters is
14
15 crucial for their performance. The optimization procedures for these meta-parameters are
16
17 complex and time consuming. Chen, et al.^[21] proposed a new approach termed Individual
18
19 Contribution Standardization (ICS) to eliminate the temperature effects on the predictive abilities
20
21 of calibration models for white chemical system. In ICS, it was assumed that the absorbance of
22
23 each chemical species in every wavelength follows simple monotonic smooth nonlinear
24
25 functions with respect to temperature. ICS had the advantages of providing an enhanced
26
27 performance with ease of implementation. Furthermore, it did not require the training samples
28
29 measured at all training temperatures to be exactly the same. Unfortunately, since ICS was
30
31 specially designed for white chemical systems, it could not be applied to grey chemical systems
32
33 (or incompletely characterized chemical systems). In order to overcome these limitations, Chen,
34
35 et al.^[22] generalized the ideas behind ICS and designed a Loading Space Standardization (LSS)
36
37 approach to correct temperature-induced spectral variations for grey chemical systems. This was
38
39 achieved by assuming that the rows of the spectral matrices as $\mathbf{X}(t_1)$, $\mathbf{X}(t_2)$, ..., and $\mathbf{X}(t_K)$
40
41 are the corresponding spectra for the same m training mixture samples measured at training
42
43 temperatures t_1 , t_2 , ..., t_K . $\mathbf{P}(t_1)$, $\mathbf{P}(t_2)$, ..., and $\mathbf{P}(t_K)$ are the corresponding loading
44
45 matrixes of the spectral matrixes calculated by singular value decomposition. The underlying
46
47 assumption of LSS is that the variations of the elements of the loading matrixes with respect to
48
49 temperature follow simple monotonic smooth nonlinear functions. In this way, LSS has the
50
51 advantages of a rather straightforward implementation and good performance over the existing
52
53 methods designed for the same purpose and provides a promising approach for correcting
54
55 temperature effects on spectroscopic measurements.
56
57
58
59
60

1
2
3 An application of LSS to Near-IR Spectral Data is set out below and is taken from Chen, *et*
4 *al.*^{[22],4}. The NIR data set comprises 95 Near-IR spectra of 19 ternary mixtures of ethanol, water
5 and 2-propanol, measured by at five different temperature levels (30°C, 40°C, 50°C, 60°C and
6 70°C). The 19 samples at each temperature were divided into 13 training samples and 6 test
7 samples. In Chen et al ^[22] a plot of the first three loading vectors for spectral data measured at all
8 training temperatures (not included here) shows that they followed simple monotonic smooth
9 nonlinear functions with respect to temperature substantiating the theoretical basis of LSS.
10
11
12
13
14
15
16

17 The effectiveness of LSS for correction of temperature induced spectral variations can be
18 illustrated by the following typical example. Figure 3 shows the 12 spectra of 6 test samples
19 measured at 60°C and 70°C, respectively. Significant spectral differences introduced by
20 temperature variation can be observed. If the calibration model built on the spectra of training
21 samples measured at 60°C is used to predict the concentrations of the three constituents in 6 test
22 samples from their corresponding spectra measured at 70°C, the temperature-induced spectral
23 variations will cause significant systematic errors in the predictions. When the LSS correction
24 model with three significant components built on the spectra of training samples measured at
25 training temperatures other than 70°C was applied, the temperature effects can be effectively
26 removed (Figure 4). After the correction of temperature-induced spectral variations, the
27 calibration model built on the spectra of training samples measured at 60°C can provide accurate
28 concentration predictions for the three chemical components in 6 test samples from the
29 corresponding corrected spectra. Additional details and comparisons with continuous piecewise
30 direct standardization (CPDS) are given in Chen, et al.^[22]. It was concluded that the simplicity in
31 setting parameters in LSS is an appealing feature making LSS highly competitive with CPDS.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 Figure 3. Spectra of 6 test samples measured at 60°C (solid line) and 70°C (dotted line)
50
51

52 Figure 4. Spectra of 6 test samples measured at 60°C (solid line) and those after corrections (dotted line) from
53 70°C to 60°C by LSS with three significant factors using the spectra of training samples measured at other
54 temperatures (except 70°C)
55
56
57
58
59
60

⁴ Reproduced with permission from *Anal. Chem.* 2005, 77, 1376-1384. Copyright 2005 American Chemical Society.

Correction of multiplicative effects

A number of chemometric pre-processing methods^[23-26] have been proposed to explicitly model the multiplicative effects caused by the variations in samples' physical properties. One of the most frequently reported techniques in the literature is that of Multiplicative Signal Correction (MSC)^[23]. However MSC and its modified versions, Inverted Signal Correction (ISC)^[24] and Extended Inverted Signal Correction (EISC)^[25], can only be applied to the spectrum that has wavelength regions containing no chemical information, i.e. influenced only by the multiplicative effects. Otherwise, it will produce dramatically bad results. Martens et al^[26] extended MSC (EMSC) by introducing chemical terms (the pure spectra of the chemical components in the samples) into the model to obtain a more effective separation of the physical effects from the chemical light absorbance effects in vibrational spectra. However, the success of the EMSC is strongly dependent on the availability of the pure spectra for all the chemical components present in the samples and the consistency of the spectral contributions from the components in the mixtures with the components isolated in the pure state. In practice, the applicability of EMSC is limited as a consequence of the difficulties in satisfying these two requirements.

The development of the Optical Path Length Estimation and Correction method (OPLEC)^[27] by Chen, *et al.* provided a major contribution to the solution to the issues discussed above. Without using any prior spectroscopic knowledge, OPLEC was shown to be able to accurately estimate the multiplicative parameter, efficiently separate the multiplicative effects of samples' physical properties from the spectral variations related to the chemical components and hence significantly enhance the prediction accuracy of the calibration models. Compared with other existing multiplicative effects correction methods, there are no additional information requirements with respect to the spectral data for its application. Consequently OPLEC potentially has wider applicability than other methods reported in the literature.

The effectiveness of the OPLEC algorithm with respect to its ability to remove the spectral variations related to multiplicative light scattering was evaluated through its application to two sets of near infrared spectroscopic data^[27]. Only the first study is reported here where in order to demonstrate the potential of the new approach a powder benchmark test data set was initially

1
2
3 used⁵. This data set consisted of 100 near-infrared transmittance spectra of five mixtures of
4
5 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).
6
7 For each of the five powder mixtures, five samples were randomly taken and loosely packed into
8
9 five different glass cuvettes. Two consecutive transmittance spectra were recorded for each
10
11 sample. Following this, each sample was packed more firmly, and a further two consecutive
12
13 transmittance spectra were recorded resulting in a total of 100 spectra. Each of the 100
14
15 transmittance spectra was transformed into absorbance spectra. 60 spectra from the three
16
17 mixtures with the ratio of gluten/starch equal to 1:0, 0.5:0.5 and 0:1 formed the calibration data
18
19 set. The test set comprised the remaining 40 spectra from the other two mixtures. More
20
21 experimental details are given in the original paper of Martens, *et al.*^[26] The 20 replicates of the
22
23 near-infrared absorbance spectra for each of the five mixtures of gluten and starch powder are
24
25 shown in Figure 5. Due to the presence of multiplicative light scattering caused by the changes in
26
27 the optical path-length, it can be observed that the 20 spectra from the same mixture differ
28
29 significantly. Figure 6 shows the spectra pre-processed by OPLEC. The pre-processed spectra
30
31 exhibit distinct spectral patterns for each of the five powder mixtures. The 20 replicates for each
32
33 mixture are more or less indistinguishable. Furthermore, the pre-processed spectra maintain the
34
35 expected equal spacing between the neighbouring spectral patterns, which correctly reflect the
36
37 variations in the concentrations of the components within the five powder mixtures.
38
39
40
41
42

43 Figure 5. The raw absorbance spectra of the five mixtures of gluten and starch powder with different weight
44 ratios (Black solid lines: 1:0; blue small dashed lines: 0.75:0.25; red large dashed lines: 0.5:0.5; green dotted
45 lines: 0.25:0.75; cyan dashed-dotted lines: 0:1)
46
47
48

49 Figure 6: The calibration (blue solid lines) and test (red dotted lines) spectra of powder mixtures
50 pre-processed by OPLEC
51
52
53

54
55 Figure 7 compares the Predictive performance of the PLS models built on calibration spectra
56 of powder mixtures pre-processed by OPLEC, EMSC_{3,93}, EMSC_{mean} and EISC. Where EMSC_{3,93}
57
58
59
60

⁵ Reproduced with permission from *Anal. Chem.* 2006, 78, 7674-7681. Copyright 2006 American Chemical Society

1
2
3 and $EMSC_{\text{mean}}$ refer to EMSC with two different selections of input pure spectra⁶ for full
4 details). It can be seen that the PLS models built on the calibration spectra pre-processed by
5 EISC resulted in unacceptable predictions with errors even larger than those achieved by the PLS
6 model generated from the raw calibration spectra. The failure of EISC on this powder mixture
7 data suggests that it is not suitable for samples with significant spectral variations resulting from
8 changes in the chemical composition. EMSC can greatly improve the predictive accuracy of the
9 PLS models. However, the significant difference between the predictions provided by the PLS
10 models on the pre-processed calibration spectra of $EMSC_{\text{mean}}$ and $EMSC_{3,93}$ demonstrate that the
11 choice of input pure spectra is crucial with respect to defining performance. The application of
12 OPLEC provides almost the same (if not better) improvement in terms of predictive ability of the
13 PLS models as does $EMSC_{3,93}$. Both methods attained the same minimal value of
14 root-mean-square error of prediction for test samples ($RMSEP_{\text{test}} = 0.005$) but with slightly
15 different numbers of PLS components. It is worth noting that after being pre-processed by
16 OPLEC, a 2-component PLS model provided excellent predictive results with a $RMSEP_{\text{test}}$ equal
17 to 0.008 which is equivalent to a relative error with a magnitude of 1.7% whilst the
18 corresponding $RMSEP_{\text{test}}$ of $EMSC_{3,93}$ is 0.013, i.e. 2.5% in terms of the relative error.
19 Considering the fact that OPLEC does not use the pure spectra of the chemical components in
20 the mixture samples as does $EMSC_{2,93}$, such results are encouraging.

21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 7: Predictive performance of the PLS models built on calibration spectra of powder mixtures pre-processed by different methods (black circle: the raw spectra, blue upwards triangle: OPLEC, yellow diamond: $EMSC_{3,93}$, green square: $EMSC_{\text{mean}}$, red downwards triangle: EISC)

Correction of combined temperature and multiplicative effects

In a recent article by Chen and Morris^[28], a new approach termed Extended Loading Space Standardization (ELSS) was presented and which aims to simultaneously correct both the temperature and multiplicative effects. One study is taken from the article which addresses the

⁶ Martens et al.^[26] used spectra x_3 and x_{93} in EMSC as the pure spectra of gluten and starch, respectively, and to investigate the influence of the choice of pure spectra on the performance of EMSC, the mean spectra of the 20 replicates of pure gluten samples and pure starch samples were also considered as the input pure spectra (normalized). Here, EMSC with the two different selections of input pure spectra is referred to as $EMSC_{3,93}$ and $EMSC_{\text{mean}}$.

1
2
3 calibrations issues associated with monitoring the Vapour Side Draw of a pharmaceuticals
4 manufacturing facility down-stream processing column⁷. Full details of the study including the
5 theoretical developments are provided in reference 28. Down stream processing in
6 pharmaceuticals is an often neglected, but extremely critical, part of the manufacture of
7 pharmaceutical products and increasingly PAT applications are being considered in order to
8 improve processing efficiencies and reduce overall production times. The problem concerned the
9 calibration of an NIR analyser for the on-line prediction of water concentration in vapour side
10 draw stream of a downstream recovery distillation column. 13 mixture samples (9 calibration
11 samples and 4 test samples) consisting of water and other two compounds were prepared. To
12 mimic the real situations in the vapour side draw stream, water is in the range of 0.051% and
13 24.69%. Each sample was heated to a temperature between 28°C and 48°C and pumped into an
14 in-line flow cell. NIR transmittance spectra (between 1858nm and 2058nm) and the
15 corresponding temperature readings were taken as the sample cooled down to a temperature
16 between 11°C and 20°C. All the transmittance spectra were transformed into absorbance spectra.
17 The raw spectra of calibration and test vapour side draw samples are shown in Figure 8a and 8b,
18 respectively. It can be observed that the temperature fluctuations cause perceptible spectral
19 variations on vapour side draw samples, and it was these that caused concern in the industrial
20 application.

21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39 The possibility of global PLS model to model this data was first tested. Cross-validated
40 root-mean-square error of prediction reaches its minimum (0.0073) when seventeen PLS
41 components were included in the calibration model. The results of the global PLS calibration
42 model are quite poor (Figure 9). The RMSEP value is 0.0071 which is equivalent to a relative
43 error of 96.13%. It might be conjectured that the assumption of a linear relationship between the
44 water concentrations and the raw spectra is inappropriate.

45
46
47
48
49
50
51
52
53 Figure 8: The raw training (a) and test (b) spectra of vapour side draw data

54
55
56
57
58
59
60
Figure 9: Water concentrations in test samples predicted by a global PLS model on the raw training spectra of vapour side draw

⁷ Reproduced with permission from *The Analyst*, 2008, 133, 914–922. Copyright The Royal Society of Chemistry 2008

1
2
3 Loading Space Standardisation (LSS) was initially employed to remove the temperature
4 variation effects. Figure 10 shows that after the application of LSS, the spectra of the same
5 sample measured at different temperatures are now almost coincidental with each other,
6 demonstrating the effectiveness of LSS in eliminating temperature-induced spectral variations.
7 Cross-validation suggested that a ten-component PLS model on the training spectra
8 pre-processed by LSS may provide good predictions for the test samples. The decrease of the
9 optimal number of PLS components from seventeen to ten again suggests that LSS
10 pre-processing reduces the complexity of the raw spectra. However, the reduction in complexity
11 by LSS brings about almost no improvement in the RMSEP value (0.0068) over that of the
12 global PLS model on raw spectra. The predictions are still observed to be dispersed and
13 significantly deviate from the actual values (Figure 11, red circles). The large dispersion of the
14 predictions cannot be caused by the 'left-over' temperature effects since the difference among
15 the standardized spectra of the same sample is negligible. A reasonable explanation for this
16 phenomenon is the existence of other nonlinear effects, such as multiplicative effects in this
17 particular data set. The excess number of PLS components used to account for the remaining
18 nonlinear effects amplifies small spectral differences and results in the abnormal dispersion
19 error.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

37 The ELSS algorithm was then used to address the remaining nonlinear effects in the spectra
38 standardized by LSS. As a result of the application of ELSS, a parsimonious three-component
39 PLS is suggested by cross-validation. The predictions of this PLS calibration model for test
40 samples are shown along with the results of LSS shown in Figure 11. In comparison with LSS,
41 ELSS offers far more accurate results. The RMSEP value (0.0026) is 2.6 times smaller than that
42 of LSS. Moreover, the water concentrations predicted from standardized spectra of the same test
43 sample also are quite close together (not shown, see reference 28). The small dispersion error
44 now truly reflects the almost insignificant differences in the standardized spectra of the same test
45 sample. Such encouraging results, especially on this industrial data, demonstrate the capability of
46 ELSS in correcting for both temperature-induced spectral variations and multiplicative effects.
47
48
49
50
51
52
53
54
55
56

57
58
59 Figure 10: LSS pre-processed training (a) and test (b) spectra of vapour side draw data
60

Figure 11: Water concentrations in vapour side draw test samples predicted by PLS model on training spectra

pre-processed by LSS (red circles) and ELSS (blue triangles) respectively

Most recently ELSS has been applied to a complex crystallization process monitored with Raman Scattering^[29]. The aim of the study were (i) to derive a model to explicitly account for the confounding effects of particle size and density of solid phases on Raman intensities, (ii) to develop a robust calibration strategy to separate the Raman contributions due to the changes of concentration of individual polymorphic forms from the confounding effects of samples' physical properties, and (iii) to achieve accurate quantitative determination of the concentration of individual polymorphic forms during solvent-mediated phase transition processes. The study demonstrated that process analytical technologies alone, without smart chemometrics and signal processing do not necessarily provide the underlying understanding needed.

Conclusions

During quantitative online/in-line monitoring of complex processes using spectroscopic instruments, calibration models are generally built on the assumption of a linear relationship between the spectroscopic measurements and the concentrations of the target chemical components. However, spectroscopic measurements made during quantitative online/in-line monitoring of such complex processes are inevitably subject to variations in measurement conditions (e.g. temperature) and samples' physical properties which make the assumption of linear relationship invalid. Thus, these detrimental effects on the spectroscopic measurements need to be explicitly modeled and corrected in order to achieve successful quantitative on-line/in-line monitoring. The methods reviewed and discussed in this paper can to some extent correct the effects of changes in temperature and samples' physical properties and improve the predictive abilities of calibration models. Most calibration methods have their limitations and various stringent requirements on the spectroscopic measurements. Once the spectroscopic measurements do not satisfy these requirements, these methods can result in poor and non-robust calibrations. LSS appears to be a very promising solution for removal of temperature induced spectral variations to maintain the predictive abilities of calibration models. OPLEC also appears to be a very promising methodology for modeling and correcting multiplicative effects caused by variations in samples' physical properties. ELSS has been shown to be able to successfully

1
2
3 model both the temperature-induced spectral variations and multiplicative effects. Its capability
4 in correcting external nonlinear effects as well as its simplicity in implementation makes ELSS a
5 promising chemometric tool for in-line/on-line process monitoring as well as in closed loop
6 process control applications where the spectroscopic measurements are not recorded under as
7 well-controlled conditions as they are in a laboratory environment. This has major implications
8 for the assured and robust application of PAT in closed loop control applications, especially in
9 pharmaceutical-chemistry manufacturing^[30].

10
11
12
13
14
15
16
17 In contrast, in bio-pharmaceuticals development and manufacturing the potential benefits are
18 still to be realized. The message, however, from pharmaceutical-chemistry applications of PAT
19 is that process analytical technologies alone, without smart chemometrics and signal processing,
20 will not provide the overall time-to-market reductions and manufacturing improvements and
21 efficiencies desired. Bio-pharma PAT technologies remain very challenging but enabling
22 analysers and sensors are now being developed that will provide enhanced science-based
23 management of product and process development and manufacturing and the potential to lower
24 manufacturing risk.

34 35 **Acknowledgements**

36
37 The authors acknowledge the financial support of the EPSRC awards GR/R19366/01
38 (KNOW-HOW), GR/R43853/01 (Chemicals Behaving Badly II), the University of Strathclyde
39 Research Development Fund and CFACT. The authors also gratefully acknowledge the
40 permission of Analytical Chemistry (References 22 and 27) and The Analyst (Reference 28) for
41 allowing reproduction of selected figures and related text. A version of this paper was presented
42 at 7th European Symposium on Biochemical Engineering Science (ESBES-7), Faro/Portugal,
43 September 7-10, 2008^[31]

References:

- 1) Siesler, H.W., Ozaki, Y., Kawata, S., and Heise, H. M. Near-infrared spectroscopy: principal, instruments, applications, **2002**, WILEY-VCH, Weinheim.
- 2) Fayolle, P.; Picque, D.; Corrieu, G. Monitoring of fermentation processes producing lactic acid bacteria by mid-infrared spectroscopy, *Vibrational Spectroscopy*, **1997**, 14, 247-252.
- 3) Schmidt, S.; Kircher, M.; Kasala, J.; Locaj, J. Near infrared spectroscopy in fermentation and quality control for amino acid production, *Bioprocess Engineering*, **1998**, 19, 67-70.
- 4) Lopes, J.A.; Menezes, J.C. Industrial fermentation end-product modelling with multilinear PLS, *Chemom. Intell. Lab. Syst.*, **2003**, 68, 75- 81.
- 5) Roggo, Y.; Roeseler, C.; Ulmschneider, M. Near infrared spectroscopy for qualitative comparison of pharmaceutical batches, *Journal of Pharmaceutical and Biomedical Analysis*, 36, **2004**, 777–786.
- 6) Jørgensen, P.; Pedersen, J.G.; Jensen, E.P.; Esbensen, K.H. On-line batch fermentation process monitoring (NIR)-introducing ‘biological process time’, *J. Chemom.* **2004**, 18, 81–91.
- 7) Birch, M.; Fussell, S. J.; Higginson, P. D.; McDowall, N.; Marziano, I. Towards a PAT-Based Strategy for Crystallization Development, *Organic Process Research & Development*, 9, **2005**, 360-364.
- 8) Parris, J.; Airiau, C.; Escott, R.; Rydzak, J.; Crocombe, R. Monitoring API Drying Operations with NIR, *Spectroscopy*, 20(2), February **2005**
- 9) Maes, I.; Van Liedekerke, B. The need for a broader perspective if process analytical technology implementation is to be successful in the pharmaceutical sector, *Journal of Pharmaceutical Innovation*, September/October, **2006**, 19-21.
- 10) Ferreira, A.P.; Menezes, J.C. Monitoring a Complex Medium Fermentation with Sample-Sample Two-Dimensional FT-NIR Correlation Spectroscopy, *Biotechnol. Prog.* **2006**, 22, 866-872.
- 11) Roggo, Y.; Chalus, P.; Maurer, L.; Lema-Martinez, C.; Edmond, A.; Jent, N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, *Journal of Pharmaceutical and Biomedical Analysis*, 44, **2007**, 683-700.

- 12) Larson T. M.; Davis J; Lam H; Cacia J. Use of process data to assess chromatographic performance in production-scale protein purification columns, *Biotechnol. Prog.*, 19, **2003**, 485-492.
- 13) Larson T. M.; Lam H. Process analytical technology in biopharmaceutical production: Past successes and future challenges. *The Journal of Process Analytical Technology*, 1, **2004**, 20-22.
- 14) Clementschitsch, F.; Bayer, K. Improvement of bioprocess monitoring: development of novel concepts, *Microbial Cell Factories*, 5 (19), **2006**, doi:10.1186/1475-2859-5-19
- 15) Iwata, T.; Koshoubu, J.; Jin, Okubo, C. Temperature dependence of the mid-infrared OH spectral band in liquid water, *Y. Appl. Spectrosc.* **1997**, 51, 1269-1275.
- 16) Ozaki, Y.; Liu, Y.; Noda, I. Two-dimensional infrared and near-infrared correlation spectroscopy: applications to studies of temperature-dependent spectral variations of self-associated molecules, *Appl. Spectrosc.* **1997**, 51, 526-535.
- 17) Wülfert, F.; Kok, W. Th.; Smilde, A. K. Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models, *Anal. Chem.* **1998**, 70, 1761-1767.
- 18) Haaland, D. M. Synthetic multivariate models to accommodate unmodeled interfering spectral components during quantitative spectral analyses, *Appl. Spectrosc.* **2000**, 54, 246-254.
- 19) Wülfert, F.; Kok, W. Th.; Noord, O. E. de; Smilde, A. K. Correction of temperature-induced spectral variation by continuous piecewise direct standardization, *Anal. Chem.* **2000**, 72, 1639-1644.
- 20) Eilers, P. H. C.; Marx, B. D. Multivariate calibration with temperature interaction using two-dimensional penalized signal regression, *Chemom. Intell. Lab. Syst.* **2003**, 66, 159-174.
- 21) Chen, Z. P.; Martin, E.; Morris, J. Modelling temperature-induced spectral variations in chemical process monitoring, *IFAC International Symposium DYCOPS*, Boston, **2004**.
- 22) Chen, Z.P.; Morris, J.; Martin, E., Correction of temperature-induced spectral variations by loading space standardization, *Anal. Chem.* **2005**, 77, 1376-1384.

- 1
2
3 23) Geladi, P.; McDougall, D.; Martens, H. Linearization and scatter-correction for
4 near-infrared reflectance spectra of meat, *Appl. Spectrosc.* **1985**, 39, 491-500.
5
6
7 24) Helland, I. S.; Næs, T.; Isaksson, T. Related versions of the multiplicative scatter
8 correction method for preprocessing spectroscopic data, *Chemom. Intell. Lab. Syst.* **1995**,
9 29, 233-241.
10
11
12
13 25) Pedersen, D. K.; Martens, H.; Nielsen, J. P.; Engelsen, S. B. Near-infrared absorption and
14 scattering separated by extended inverted signal correction (EISC): analysis of
15 near-infrared transmittance spectra of single wheat seeds, *Appl. Spectrosc.* **2002**, 56,
16 1206-1214.
17
18
19
20
21 26) Martens, H.; Nielsen J. P.; Engelsen, S. B. Light scattering and light absorbance separated
22 by extended multiplicative signal correction. Application to near-infrared transmission
23 analysis of powder mixtures, *Anal. Chem.* **2003**, 75, 394-404.
24
25
26
27 27) Chen, Z.P.; Morris, J.; Martin, E., Extracting chemical information from spectral data with
28 multiplicative light scattering effects by optical path-length estimation and correction,
29 *Anal. Chem.*, **2006**, 78, 7674-7681.
30
31
32
33 28) Chen, Z.P.; Morris, J., Improving the linearity of spectroscopic data subjected to
34 fluctuations in external variables by the extended loading space standardization, *The*
35 *Analyst*, 2008, 133, 914-922.
36
37
38
39 29) Chen, Z.P.; Fevotte, G.; Caillet, A. Littlejohn, D.; Morris, J. Advanced Calibration
40 Strategy for in Situ Quantitative Monitoring of Phase Transition Processes in Suspensions
41 Using FT-Raman Spectroscopy, *Anal. Chem.* **2008**, 80, 6658-6665
42
43
44
45 30) Chen, Z.P., D. Lovett and Julian Morris, 2008, Process Analytical Technologies (PAT) –
46 the Impact for Process Systems Engineering, 18th European Symposium on Computer
47 Aided Process Engineering – ESCAPE 18, June 1-4, Lyon, Eds. Bertrand Braunschweig
48 and Xavier Joulia.
49
50
51
52
53 31) Process Analytical Technologies (PAT) and Process Systems Engineering, 7th European
54 Symposium on Biochemical Engineering Science (ESBES-7), Faro/Portugal, September
55 7-10, 2008
56
57
58
59
60

List of Figures

Figure 1: Five NIR spectra for a ternary mixture sample measured at five different temperatures (black solid line: 30°C, red dash line: 40°C, green dash-dot line: 50°C, pink dotted line: 60°C, blue dash-dot-dot line: 70°C)

Figure 2: Mean-centred NIR spectra of a powder mixture measured under different sample compactness.

Figure 3. Spectra of 6 test samples measured at 60°C (solid lines) and 70°C (dotted lines)

Figure 4. Spectra of 6 test samples measured at 60°C (solid lines) and those after corrections (dotted lines) from 70°C to 60°C by LSS with three significant factors using the spectra of training samples measured at other temperatures except 70°C

Figure 5. The raw absorbance spectra of the five mixtures of gluten and starch powder with different weight ratios (Black solid lines: 1:0; blue small dashed lines: 0.75:0.25; red large dashed lines: 0.5:0.5; green dotted lines: 0.25:0.75; cyan dashed-dotted lines: 0:1)

Figure 6: The calibration (blue solid lines) and test (red dotted lines) spectra of powder mixtures pre-processed by OPLEC

Figure 7: Predictive performance of the PLS models built on calibration spectra of powder mixtures pre-processed by different methods (black circle: the raw spectra, blue upward triangle: OPLEC, yellow diamond: EMSC_{3,93}, green square: EMSC_{mean}, red downward triangle: EISC)

Figure 8: The raw training (a) and test (b) spectra of vapour side draw data

Figure 9: Water concentrations in test samples predicted by a global PLS model on the raw training spectra of a vapour side draw

1
2
3 Figure 10: LSS pre-processed training (a) and test (b) spectra of vapour side draw data
4
5
6

7 Figure 11: Water concentrations in vapour side draw test samples predicted by PLS model on
8
9 training spectra pre-processed by LSS (red circles) and ELSS (blue triangles) respectively
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

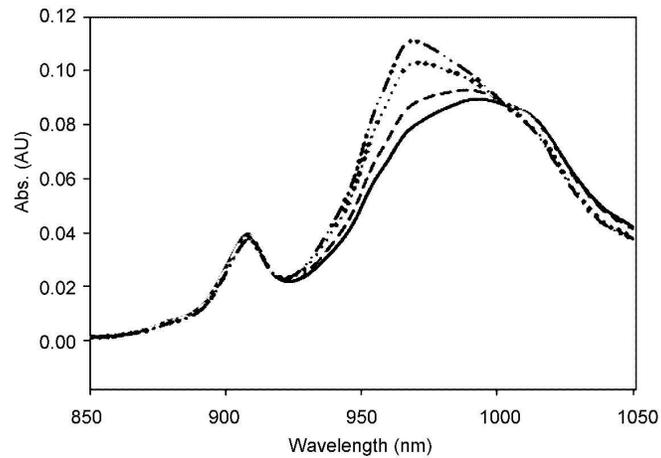


Figure 1

Five NIR spectra for a ternary mixture sample measured at five different temperatures (black solid line: 30°C, red dash line: 40°C, green dash-dot line: 50°C, pink dotted line: 60°C, blue dash-dot-dot line: 70°C)
297x209mm (300 x 300 DPI)

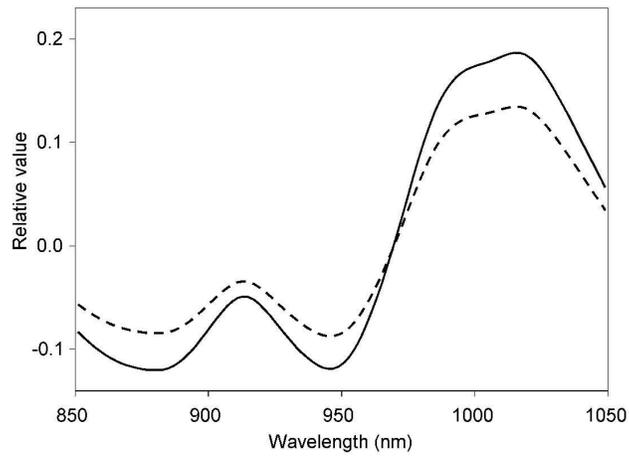


Figure 2

Mean-centred NIR spectra of a powder mixture measured under different sample compactness
297x209mm (300 x 300 DPI)

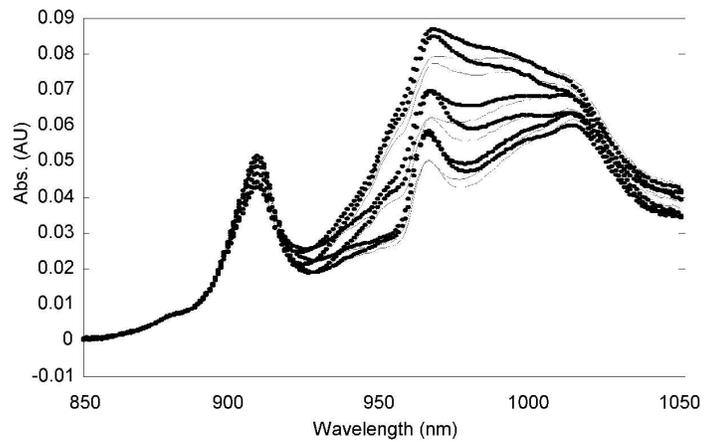


Figure 3

Spectra of 6 test samples measured at 60°C (solid line) and 70°C (dotted line)
297x209mm (300 x 300 DPI)

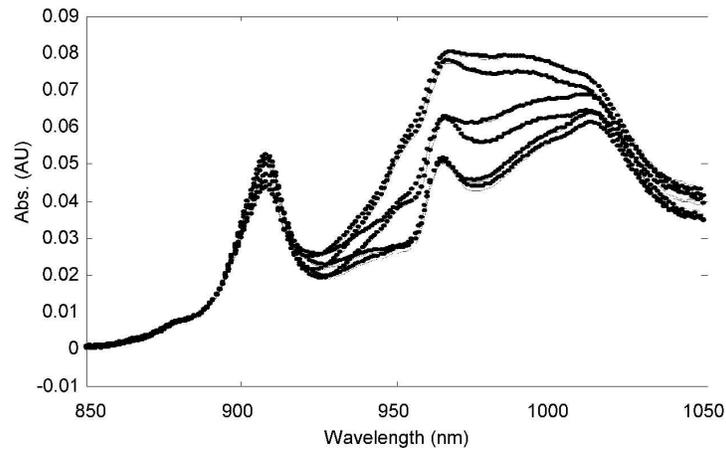


Figure 4

Spectra of 6 test samples measured at 60°C (solid line) and those after corrections (dotted line) from 70°C to 60°C by LSS with three significant factors using the spectra of training samples measured at other temperatures (except 70°C)
297x209mm (300 x 300 DPI)

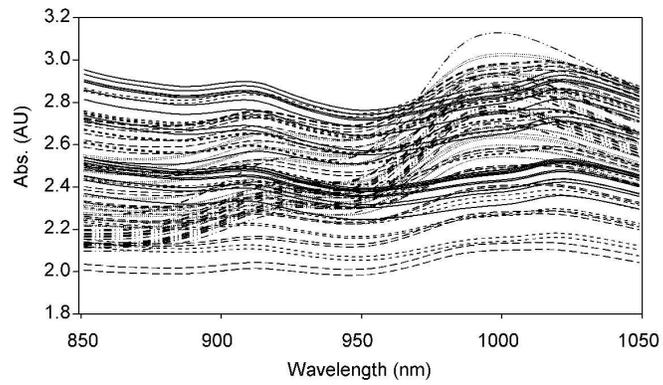


Figure 5

The raw absorbance spectra of the five mixtures of gluten and starch powder with different weight ratios (Black solid lines: 1:0; blue small dashed lines: 0.75:0.25; red large dashed lines: 0.5:0.5; green dotted lines: 0.25:0.75; cyan dashed-dotted lines: 0:1)
297x209mm (300 x 300 DPI)

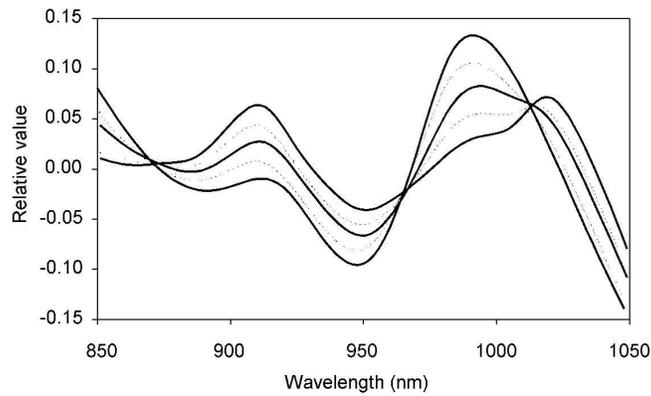


Figure 6

The calibration (blue solid lines) and test (red dotted lines) spectra of powder mixtures pre-processed by OPLEC
297x209mm (300 x 300 DPI)

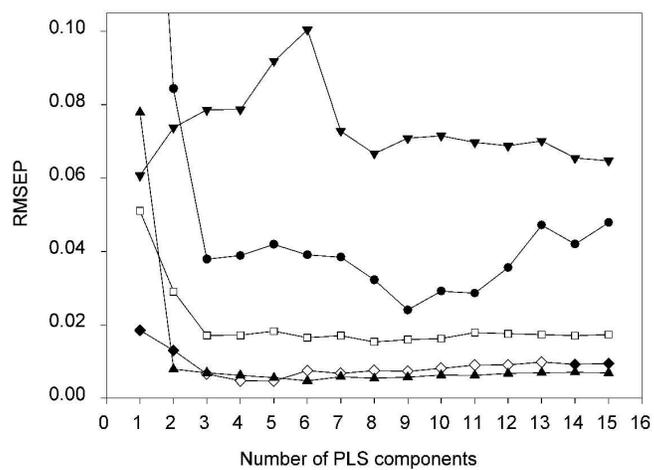


Figure 7

Predictive performance of the PLS models built on calibration spectra of powder mixtures pre-processed by different methods (black circle: the raw spectra, blue upwards triangle: OPLEC, yellow diamond: EMSC3,93, green square: EMSCmean, red downwards triangle: EISC)
297x209mm (300 x 300 DPI)

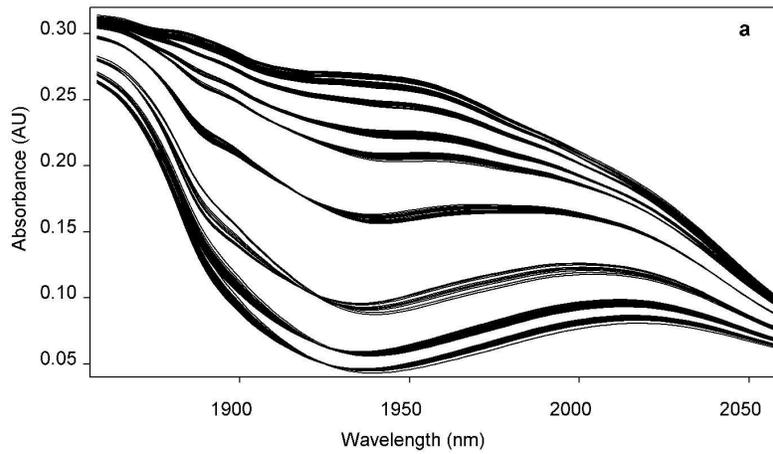


Figure 8a

The raw training (a) and test (b) spectra of the vapour side draw data
297x209mm (300 x 300 DPI)

review

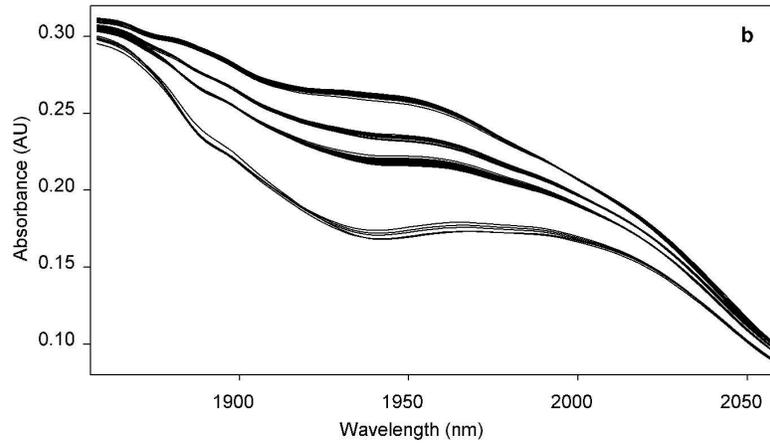


Figure 8b

The raw training (a) and test (b) spectra of the vapour side draw data
297x209mm (300 x 300 DPI)

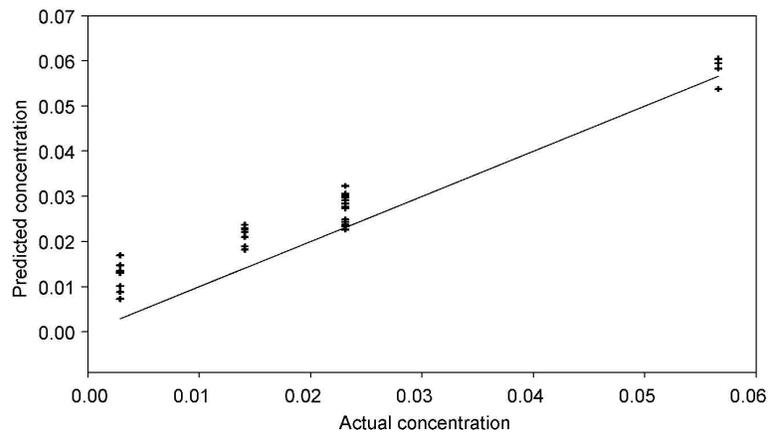


Figure 9

Water concentrations in test samples predicted by a global PLS model on the raw training spectra of vapour side draw
297x209mm (300 x 300 DPI)

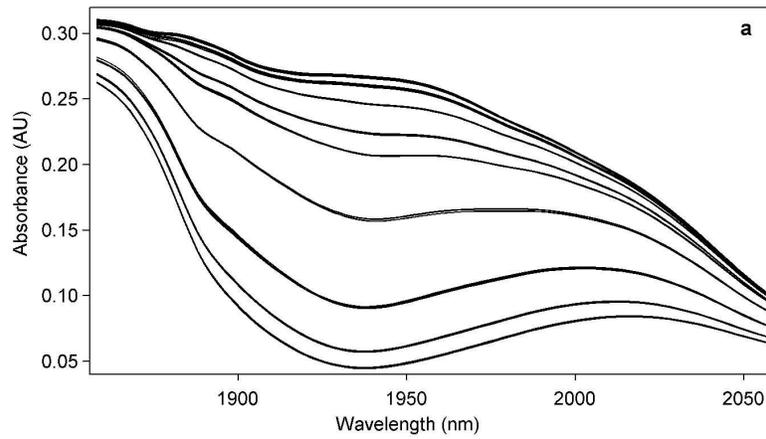


Figure 10a

LSS pre-processed training (a) and test (b) spectra of vapour side draw data
297x209mm (300 x 300 DPI)

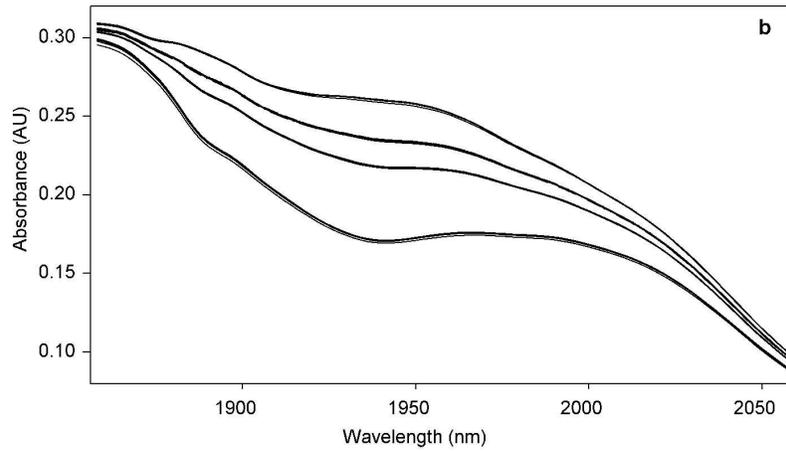


Figure 10b

LSS pre-processed training (a) and test (b) spectra of vapour side draw data
297x209mm (300 x 300 DPI)

review

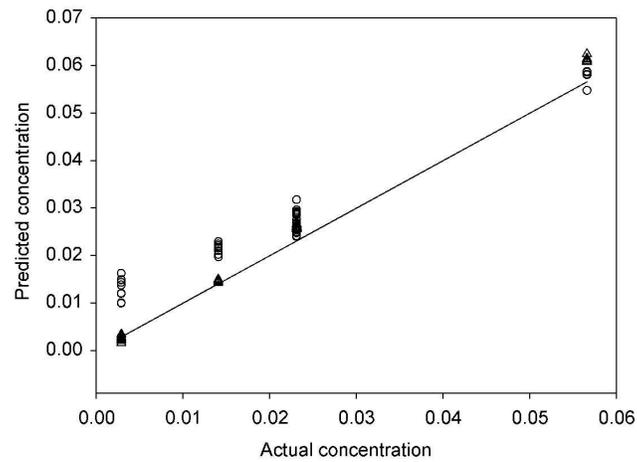


Figure 11

Water concentrations in vapour side draw test samples predicted by PLS model on training spectra pre-processed by LSS (red circles) and ELSS (blue triangles) respectively
297x209mm (300 x 300 DPI)