Proceedings of the 11th International Symposium on
Computer Applications in Biotechnology (CAB 2010),
Leuven, Belgium, July 7-9, 2010
Julio R. Banga, Philippe Bogaerts, Jan Van Impe,
Denis Dochain, Ilse Smets (Eds.)

Towards the Robust Application of PAT in Real Time Control

Julian Morris*, Zengping Chen and David Lovett*****

* Centre for Process Analytics & Control Technology, Newcastle University. UK
**State Key Laboratory of Chemo/Biosensing & Chemometrics, Hunan University, China
***Perceptive Engineering, Daresbury Innovation Centre, Daresbury, Cheshire, UK

Abstract: Process Analytical Technologies (PAT) are increasingly being explored and adopted by pharma-chem and bio-pharma companies for enhanced process understanding, Quality by Design (QbD) and Real-time-Release (RTR). To achieve these aspirations there is a critical need to extract the most information, and hence understanding, from complex and often very 'messy' spectroscopic data. A number of new approaches will be shown to overcome the limitations of existing calibration/modelling methodologies and algorithms and their use in some industrial applications will be presented.

Keywords: Process Analytical Technologies, Spectroscopic Data, Chemometrics, Calibration, Process Control

1. INTRODUCTION

In 2004 the FDA published its Process Analytical Technology (PAT) guidance and the its cGMPs for the 21st Century which calls for the design of effective and efficient manufacturing processes to assure product quality and performance; product specifications based on a mechanistic understanding of how different formulations and processes affect product performance and continuous real-time assurance of quality and European Medicines Agency (EMEA) published its Road Map to 2010 'Preparing the Ground for the Future'. These publications released the potential for significant changes in the development and manufacturing of pharmaceuticals. Folestad (1999) stated that "*PAT* is a system for designing, analyzing, and controlling manufacturing through timely measurements (i.e., during processing) of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality. *QbD* is "a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding based on sound science and quality risk management (definition in ICH Q8R, Annex to ICH Q8: 'Pharmaceutical Development). *Design Space:* The multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality. Working within the design space is not considered as a change (definition in ICH Q8: 'Pharmaceutical Development). *Quality:* The suitability of either a drug substance or drug product for its intended use. This term includes such attributes as the identity, strength, and purity. *Real-Time-Release:* A system which ensures that a product is of the intended quality, while reducing or (in some cases) making end-product testing redundant, by utilising an appropriate combination of the following: formulation design, process design, process validation, qualification of raw materials, and in-line/at-line control of key process parameters; at-line/in-line measurement of appropriate product attributes, and compliance with specific cGMP requirements (EFPIA, 2002); Ability to evaluate and assure acceptable quality of finished product based on process data, typically a valid combination of material attributes and process controls (process parameters) (FDA, PAT Guidance 2004 Real-time quality control, leading to a reduction of end-product release testing, ICH Q8)".

It is interesting to observe that the industry has always used QbD to some extent to design and build quality into product and manufacturing process quality - don't test for quality (Edwards Deming); monitor and improve processes to reduce variability; and use quality risk management to focus resources into areas critical to the patient. This approach aligns with continual process improvement (e.g. lean six sigma); and enhanced innovation by reducing regulatory burden associated with changes. Although presently under current review by the American Society for Testing and Materials (ASTM) committee E55 (Manufacture of Pharmaceutical Products) there is no standard PAT instrument verification system in place, although there are numerous currently non-standard techniques available. This paper considers the development and application of recent multivariate methods to verify the integrity of the PAT "system", including sensor, calibration model, and fault diagnostics. Of particular interest in meeting these challenges is the on-line real-time use of process analytics (e.g. spectroscopic instrumentation) for process monitoring and control applications. These drive the need to incorporate and integrate the detailed spectral information into process performance monitoring and predictive control schemes for real-time-release.

In other process industries, on-line real-time spectroscopic technologies such as near infrared, mid infrared, UV-Vis, Raman, X-ray diffraction, etc., have been widely applied but to a lesser extent in pharmaceuticals and virtually none in closed process control and optimisation. Although a number of methodologies are available for addressing, in one way or another, such challenges, much more needs to be done.

Compounding these issues is the need for transferable and fit-for-purpose (across scales and equipment differences) spectroscopic calibration models for in on-line/in-line monitoring – still a major concern today. Two new methods, a modified *Loading Space Standardization (LSS)* algorithm and *Systematic Prediction Error Correction (SPEC)*, have been developed to maintain the predictive abilities of multivariate calibration models when e.g. the spectrometer or measurement conditions change. The performance of the methods has been evaluated on two NIR data sets (one with changes in instrumental responses, the other with temperature induced spectral variations) and compared with that of two commonly used methods - Direct Standardization (DS) and Piecewise Direct Standardization (PDS).

2. TEMPRATURE INDUCED SPECTRAL VARIATIONS

When spectral measurements are subject to such changes and variations, methods for calibration model maintenance (e.g. Forina et al, 1995; Feudale et al, 2002; Bouveresse, et al, 1995; Despagne, et al, 1998) are needed to prevent degradation in the accuracy and reliability of multivariate calibration models and avoid time-consuming full recalibration procedures. Calibration model maintenance methods can be roughly classified into three categories, i.e. calibration model coefficients updating methods (e.g. Bouveresse et al, 1996; Greensill et al, 2001) prediction correction methods (e.g. Bouveresse, et al, 1994) and spectral responses standardization methods (Bouveresse et al, 1995; Despagne et al, 1998). For complex situations, spectral responses standardization methods are generally accepted to be more suitable than prediction correction methods. The idea of spectral responses standardization is to find a transformation matrix, which transforms the spectra of future test samples into the corresponding spectra as if they were measured under the same conditions or on the same instrument as the calibration samples used to build the original calibration model. Therefore, the original calibration model can be used for prediction without having to update the model coefficients. The transformation matrix can be obtained by regressing the spectra of a subset of samples (often called standardization samples) measured on the primary instrument (or initial calibration conditions) against the spectra of the same subset measured on the secondary instrument (or under the modified test conditions). *Direct Standardization (DS)*, which directly relates the response of a sample measured with one instrument to its response obtained on another instrument, e.g. Wang, Veltkamp and Kowalski, 1991; and *Piecewise Direct Standardization (PDS),* e.g. Wang et al, 1991; Wang & Kowalski, 1993; Wang, Dean and Kowalski, 1995; are well-established methods for the correction of complex nonlinear spectral variations between measurements. However, the main limitation of these methods is the inability to handle the continuous nature of temperature. To overcome this limitation, PDS was generalized to continuous variables - *continuous* piecewise direct standardization (Wűlfert et al, 2000; Wülfert, et al., 2000a,b; Barring et al, 2001). Both DS and PDS can handle complex situations due to their multivariate character. The main drawback of DS is the over-fitting problem resulting from the fact that the number of standardization samples is much smaller than the number of variables (e.g. wavelengths). So the transformation matrix is typically estimated by means of PCA and PLS. PDS adopts the same linear model as DS does. PDS differs from DS only in the way that the transformation matrix is computed. In PDS, the transformation matrix is estimated by the moving window procedure, which can to some extent alleviate the over-fitting problem and also enable better modeling of possible non-linearity's. However, the use of the moving window procedure also compounds a practical drawback of PDS, viz. a relatively long computation time owing to the numerous local multivariate regression models. Also, the window size has a significant effect on the performance of PDS and needs to be carefully determined. Moreover, poor estimation of the local rank of each local multivariate regression model may lead to discontinuities in the PDS transformed spectra (Gemperline et al, 1996; Chen, Morris and Martin, 2005). Furthermore, with PDS, the spectra of the standardization samples must be measured on both instruments or under both sets of measurement conditions, which makes the procedure inapplicable to on-line/in-line monitoring of chemical and bio-chemical processes where such a requirement is difficult to satisfy in practice. Due to the various theoretical and practical limitations of the existing calibration model maintenance methods, there is still a need for methods that are easier to implement (i.e. require fewer or even no meta-parameters) and at the same time provide better performance. In this paper, two new methods for calibration model maintenance are described and their performance compared to that of DS and PDS using industrial NIR data.

3. THEORY

Suppose a multivariate calibration model, $y = \mathrm{f}(\mathbf{x})$ has been established on the spectra of a set of calibration samples measured at the selected calibration conditions or on the primary instrument ($\mathbf{x}$ and y represent the spectrum of a sample and the concentration of the target constituent in the sample, respectively). The task is now to enable the calibration model to give correct quantitative predictions for the target constituent in test samples, based on their spectra measured at the test conditions or on a secondary instrument.

3.1 Loading Space Standardization (LSS)

Assume the rows of spectral matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ are the corresponding spectra of the same subset of standardization samples measured at the calibration and test conditions (or on the primary and secondary instruments), respectively. According to Beer-Lambert law, $\mathbf{X}_1$ and $\mathbf{X}_2$ can be decomposed as follows:

$$\mathbf{X}_1 = \mathbf{C}\mathbf{S}_1^{\mathrm{T}} + \mathbf{E}_1 \,;\; \mathbf{X}_2 = \mathbf{C}\mathbf{S}_2^{\mathrm{T}} + \mathbf{E}_2 \qquad (1)$$

Where, $\mathbf{C}$ is the concentration matrix with its ith row representing the concentrations of all the chemical components in the i^{th} standardization sample. $\mathbf{S}_1$ and $\mathbf{S}_2$ are pure spectral matrices, whose columns are the pure spectra of chemical components in the standardization samples at the calibration and test conditions (or on the primary and secondary instruments), respectively. $\mathbf{E}_1$ and $\mathbf{E}_2$ denote the corresponding residual matrices. If $\mathbf{S}_1$ and $\mathbf{S}_2$ are known *a*

priori, for a test sample, its spectrum ($\mathbf{x}_{test}$) measured at the test conditions (or on the secondary instrument) can be easily transformed into spectrum ($\mathbf{x}_{trans}$) as if it were measured at the calibration conditions (or on the primary instrument) through the simple calculations in (2). The multivariate calibration model built at the calibration conditions (or on the primary instrument) can then be used to predict the concentration of the target constituent in the test sample from the transformed spectrum.

$$\mathbf{x}_{trans} = \mathbf{x}_{test}(\mathbf{S}_2^{\mathrm{T}})^{+}\mathbf{S}_1^{\mathrm{T}} + \mathbf{x}_{test} - \mathbf{x}_{test}(\mathbf{S}_2^{\mathrm{T}})^{+}\mathbf{S}_2^{\mathrm{T}} \qquad (2)$$

where, '+' symbolizes the Moore–Penrose generalized inverse**.** In most cases, it is difficult, if not impossible, to obtain $\mathbf{S}_1$ and $\mathbf{S}_2$. Since matrix $\mathbf{R}$ is full rank the following equations hold (Chen et al, 2005).

$$(\mathbf{RS}_2^{\mathrm{T}})^{+}\mathbf{RS}_1^{\mathrm{T}} = (\mathbf{S}_2^{\mathrm{T}})^{+}\mathbf{R}^{+}\mathbf{RS}_1^{\mathrm{T}} = (\mathbf{S}_2^{\mathrm{T}})^{+}\mathbf{S}_1^{\mathrm{T}} \qquad (3)$$

$$(\mathbf{RS}_2^{\mathrm{T}})^{+}\mathbf{RS}_2^{\mathrm{T}} = (\mathbf{S}_2^{\mathrm{T}})^{+}\mathbf{R}^{+}\mathbf{RS}_2^{\mathrm{T}} = (\mathbf{S}_2^{\mathrm{T}})^{+}\mathbf{S}_2^{\mathrm{T}} \qquad (4)$$

Therefore, $\mathbf{S}_1^{\mathrm{T}}$ and $\mathbf{S}_2^{\mathrm{T}}$ in (2) can be replaced by $\mathbf{RS}_1^{\mathrm{T}}$ and $\mathbf{RS}_2^{\mathrm{T}}$, respectively, which are readily obtained by singular value decomposition of $\mathbf{X}_{comb}$=[$\mathbf{X}_1$, $\mathbf{X}_2$].

$$\mathbf{X}_{comb} = [\mathbf{U}_s, \mathbf{U}_n]\begin{bmatrix}\boldsymbol{\Lambda}_s^{1/2} & \\ & \boldsymbol{\Lambda}_n^{1/2}\end{bmatrix}[\mathbf{V}_s, \mathbf{V}_n]^{\mathrm{T}} = \mathbf{T}_s\mathbf{P}_s^{\mathrm{T}} + \mathbf{T}_n\mathbf{P}_n^{\mathrm{T}} \qquad (5)$$

$$\mathbf{T}_s = \mathbf{U}_s\boldsymbol{\Lambda}_s^{1/2},\ \mathbf{T}_n = \mathbf{U}_n\boldsymbol{\Lambda}_n^{1/2},\ \mathbf{P}_s = \mathbf{V}_s,\ \mathbf{P}_n = \mathbf{V}_n$$

Subscripts '*s*' and '*n*' signify that the corresponding factors represent spectral information and noise, respectively. Now, partition $\mathbf{P}_s^{\mathrm{T}}$ into two sub-matrices $\mathbf{P}_1^{\mathrm{T}}$ and $\mathbf{P}_2^{\mathrm{T}}$ ($\mathbf{P}_s^{\mathrm{T}} = [\mathbf{P}_1^{\mathrm{T}}, \mathbf{P}_2^{\mathrm{T}}]$), which have the same sizes as $\mathbf{S}_1^{\mathrm{T}}$ and $\mathbf{S}_2^{\mathrm{T}}$, respectively. Since $\mathbf{X}_{comb} = [\mathbf{X}_1, \mathbf{X}_2] = \mathbf{C}[\mathbf{S}_1^{\mathrm{T}}, \mathbf{S}_2^{\mathrm{T}}] + \mathrm{E}$, it can be shown that there exists a full rank matrix $\mathbf{R}$ satisfying the following equations:

$$\mathbf{RS}_1^{\mathrm{T}} = \mathbf{P}_1^{\mathrm{T}},\ \mathbf{RS}_2^{\mathrm{T}} = \mathbf{P}_2^{\mathrm{T}} \qquad (6)$$

Combining (2), (3), (4) and (6), gives:

$$\mathbf{x}_{trans} = \mathbf{x}_{test}(\mathbf{P}_2^{\mathrm{T}})^{+}\mathbf{P}_1^{\mathrm{T}} + \mathbf{x}_{test} - \mathbf{x}_{test}(\mathbf{P}_2^{\mathrm{T}})^{+}\mathbf{P}_2^{\mathrm{T}} \qquad (7)$$

The above transformation method is a special case of the loading space standardization method developed by the authors (2005). For convenience, it is also referred to as *Loading Space Standardization (LSS)* throughout this paper.

3.2 Systematic Prediction Error Correction (SPEC)

The applicability of LSS lies in the availability of the spectra of a subset of standardization samples recorded at both the calibration and test conditions (or on both the primary and secondary instruments). However, such a requirement is difficult to satisfy in the area of on-line/in-line monitoring of chemical and bio-chemical processes. In such applications, it is relatively easier to measure the spectra of a subset of standardization samples at the test conditions (or on the secondary instrument) and the concentrations of the target constitute in the standardization samples through off-line assay. This section focuses on how to maintain multivariate linear calibration models under this circumstance.

Suppose $\mathbf{X}_2$ is the spectral matrix of a subset of standardization samples measured at the test conditions or on the secondary instrument, and $\mathbf{y}_2$ is a vector containing the concentrations of the target constituent in the standardization samples. For a multivariate linear calibration model, $y = \mathrm{f}(\mathbf{x}) = a\mathbf{1} + \mathbf{xb}$ (where $\mathbf{1}$ is a column vector with its elements equal to unity), the concentrations of the target constitute in the standardization samples can be predicted from their corresponding LSS transformed spectrum ($\mathbf{X}_{trans}$).

$$\begin{aligned}\hat{\mathbf{y}}_2 &= \mathrm{f}(\mathbf{X}_{trans}) \\ &= \mathrm{f}\{\mathbf{X}_2(\mathbf{P}_2^{\mathrm{T}})^{+}\mathbf{P}_1^{\mathrm{T}} + \mathbf{X}_2 - \mathbf{X}_2(\mathbf{P}_2^{\mathrm{T}})^{+}\mathbf{P}_2^{\mathrm{T}}\} \\ &= \mathbf{X}_2(\mathbf{P}_2^{\mathrm{T}})^{+}(\mathbf{P}_1^{\mathrm{T}} - \mathbf{P}_2^{\mathrm{T}})\mathbf{b} + \mathrm{f}(\mathbf{X}_2)\end{aligned} \qquad (8)$$

If we define $\boldsymbol{\beta} = (\mathbf{P}_1^{\mathrm{T}} - \mathbf{P}_2^{\mathrm{T}})\mathbf{b}$ and $\mathbf{T}_2 = \mathbf{X}_2(\mathbf{P}_2^{\mathrm{T}})^{+}$, equation 8 can be rewritten as $\mathbf{T}_2\boldsymbol{\beta} = \hat{\mathbf{y}}_2 - \mathrm{f}(\mathbf{X}_2)$. With a view to obtaining $\boldsymbol{\beta}$, $\hat{\mathbf{y}}_2$ and $\mathbf{T}_2$ can be replaced by $\mathbf{y}_2$ and $\mathbf{T}_{s,2}$ (estimated through the singular value decomposition of $\mathbf{X}_2$ in (9)), respectively.

$$\begin{aligned}&\mathbf{X}_2 = [\mathbf{U}_{s,2}, \mathbf{U}_{n,2}]\begin{bmatrix}\boldsymbol{\Lambda}_{s,2}^{1/2} & \\ & \boldsymbol{\Lambda}_{n,2}^{1/2}\end{bmatrix}[\mathbf{V}_{s,2}, \mathbf{V}_{n,2}]^{\mathrm{T}} = \mathbf{T}_{s,2}\mathbf{P}_{s,2}^{\mathrm{T}} + \mathbf{T}_{n,2}\mathbf{P}_{n,2}^{\mathrm{T}} \\ &\mathbf{T}_{s,2} = \mathbf{U}_{s,2}\boldsymbol{\Lambda}_{s,2}^{1/2},\ \mathbf{T}_{n,2} = \mathbf{U}_{n,2}\boldsymbol{\Lambda}_{n,2}^{1/2},\ \mathbf{P}_{s,2} = \mathbf{V}_{s,2}, \\ &\mathbf{P}_{n,2} = \mathbf{V}_{n,2}\end{aligned} \qquad (9)$$

Therefore, $\boldsymbol{\beta} = \mathbf{T}_{s,2}^{+}[\mathbf{y}_2 - \mathrm{f}(\mathbf{X}_2)]$. For a test sample, the concentration of the target constituent can then be directly calculated from its spectrum ($\mathbf{x}_{test}$) measured at the test conditions or on the secondary instrument without spectral transformation.

$$\mathbf{t}_{test} = \mathbf{x}_{test}\mathbf{P}_{s,2},\ y_{test} = \boldsymbol{f}(\mathbf{x}_{test}) + \mathbf{t}_{test}\boldsymbol{\beta} \qquad (10)$$

In 10, the term $\mathbf{t}_{test}\boldsymbol{\beta}$ can be regarded as the systematic prediction error of the multivariate linear calibration model caused by the spectral differences resulting from the variations in measurement conditions or changes in instrument. This is why the above method for calibration model maintenance is called *Systematic Prediction Error Correction (SPEC).*

4. APPLICATION

The data consisted of 1308 spectra of 654 pharmaceutical tablets measured on two NIR spectrometers in the transmittance mode from 600 to 1898 nm in 2 nm increments. Each individual tablet was subsequently analyzed for assay value of the active ingredient, tablet weight, and tablet hardness. For each of the 1308 absorbance spectra, the 520 absorbance values in the range between 600 and 1638 nm were used for the subsequent data analysis. Absorbance spectra from each instrument were split into a calibration set (155 spectra), a test set (459 spectra) and a validation set (40 spectra). The calibration set includes tablets with a wide range of assay values (151.6~239.1 mg) for developing calibration model. The challenge for this data set is to develop a multivariate linear calibration model for the assay value of the active ingredient on one instrument (the primary instrument), and then to provide the best means of transferring the calibration model to the secondary instrument. Transformation matrices for both DS and PDS were estimated by PLS from the raw absorbance spectra of the

standardization samples. Singular value decomposition of both LSS and SPEC were also carried out on the raw spectra. The root-mean-square error of prediction (RMSEP) is used as the performance criterion.

5. RESULTS AND DISCUSSION

Although the spectra of the same sample measured using two different instruments of the same type may have the same basic shape or profile, subtle differences in instrumental response functions can result in perceptible spectral variations. This is illustrated in Figure 1, where significant spectral variations can be observed in the region between 600 and 720 nm; there are also subtle spectral differences in other regions. These spectral variations can cause large systematic prediction errors when the calibration model built on the primary instrument is applied to the spectra measured on the secondary instrument.

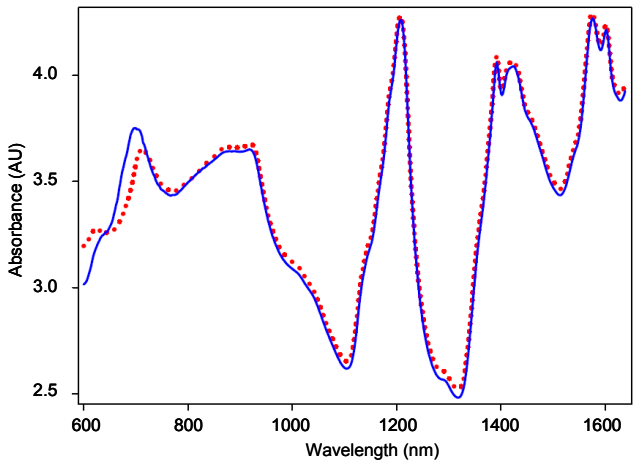

Figure 1. Spectra of the tablet obtained with the primary (red dotted line) and secondary instrument (blue solid line)

As shown in Figure 2, the differences in instrumental responses resulted in significant offsets in the predictions of the PLS calibration model. Besides affecting accuracy, the change in instrument also degraded the predictive precision of the calibration model, which signifies the spectral variations introduced by instrumental differences are not systematic for all samples, but rather differ from sample to sample. So, for the pharmaceutical tablet data, multivariate model maintenance methods rather than univariate methods (such as SBC) are needed.

Four methods (LSS, SPEC, PDS and DS) were used to evaluate the potential elimination of the detrimental effects of instrumental/temperature changes with a view to maintaining the predictive abilities of the calibration models. Before the application of these methods, the influence of some model parameters, i.e. the number of standardization samples, the number of principal components and the window size (exclusively for PDS) on their performances was investigated. In order to obtain reliable results, the number of standardization samples should be at least equal to the number of actual chemical and (or) physical variation sources in the calibration spectra. The larger the number of standardization samples, the higher the probability that good results will be obtained. However, more standardization samples require increased analysis time and effort, with associated higher costs. Therefore, the calibration maintenance method that achieves satisfactory results with fewer standardization samples is preferred in practice.

The effects of the number of standardization samples from the pharmaceutical tablet data, on the performances of the four calibration maintenance methods (LSS, SPEC, PDS and DS) are shown in Figure 3. It is observed that increasing the number of standardization samples does not significantly reduce the RMSEP values for the DS, PDS and LSS methods or for SPEC with 6 or more samples. The main observation is that a lower value of RMSEP was obtained with SPEC or LSS compared to the values for DS and PDS. In PDS and DS, the number of principal components (PCs) is used in the calculation of the transformation matrices; while in LSS and SPEC, they are related to the determination of the number of factors representing spectral information after singular value decomposition.

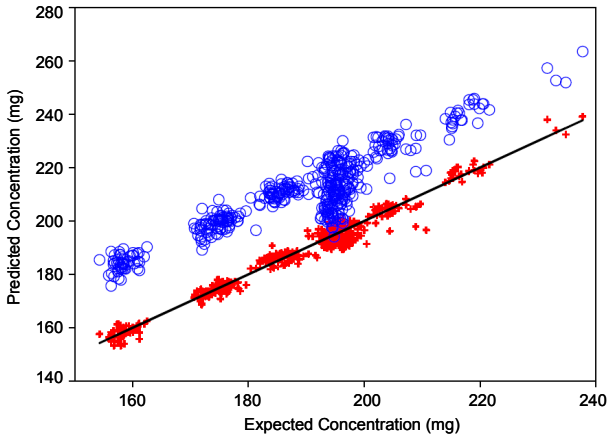

Figure 2. Concentrations of active ingredient in test pharmaceutical tablet samples predicted from their spectra measured with the primary (red cross) and secondary instruments (blue circle) using a PLS calibration model with eleven principal components built from calibration spectra obtained using the primary instrument.

Theoretically, the number of PCs used in all four methods should not be less than the number of chemical components in the system under study. Studies have shown that once the number of PCs is larger than a certain value, a further increase has little effect on the performance of PDS and DS in terms of the RMSEP. Similarly, with SPEC there is only a slight decrease in the RMSEP when the number of PCs is increased from 4 to 12. For LSS, the effect of the number of PCs is different: the RMSEP is similar for 4 to 9 PCs and then increases for 10 to 12 PCs. As a rule of thumb, when the number of standardization samples is small, the number of PCs used in LSS and SPEC can be set to a value equal to the number of standardization samples. The 'window size' parameter (N_{ws}) is only relevant to the PDS method and is one of its weaknesses, e.g. for the pharmaceutical tablet data the RMSEP value varied from 4 to 9 in an unsystematic manner when the window size was changed and is one of the practical problems of PDS.

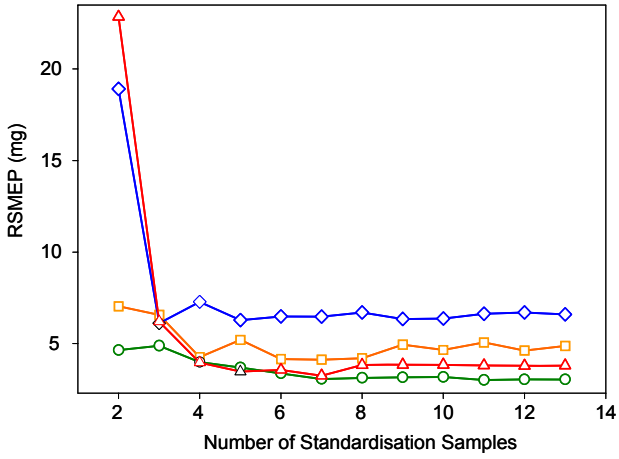

Figure 3. Active ingredient RMSEP values predicted calculated from spectra recorded on the secondary instrument. Effects of number of standardization samples on the performance of the four calibration maintenance methods (green circle-SPEC; red triangle-PDS; orange square-global PLS; blue diamond-SBC).

Table 1 lists the RMSEP values (mg) for the active ingredient concentrations in pharmaceutical tablet samples predicted by different models from their spectra measured on the secondary instrument. Without model maintenance, the PLS calibration model, PLS_1 denotes the PLS calibration model built using the spectra of the calibration samples recorded with the primary instrument. PLS_{sub} denotes the PLS calibration model built using the spectra of six standardization samples recorded with the secondary instrument. PLS_{global} signifies the global PLS model built using the spectra of the calibration samples recorded with the primary instrument and the six standardization samples recorded with the secondary instrument.PLS_2 represents the PLS calibration model built using the spectra of the calibration samples recorded with the secondary instrument.

The RMSEP value for the entire pharmaceutical tablet data set is as high as 22.7 which is equivalent to an average relative predictive error of 11.7% (not shown). The PLS model (PLS_{sub}) built on the spectra of six standardization samples measured on the secondary instrument provided concentration predictions with substantially lower RMSEP values. However, compared with the fully recalibrated PLS_2 model, the predictive errors of PLS_{sub} are still moderately high. The application of calibration model maintenance methods with six standardization samples further reduced the predictive errors. Among the four calibration model maintenance methods, LSS (N_{pc}: 6) achieved the best results with RMSEP values favorably comparable to those of the fully recalibrated PLS_2 model. Though SPEC (N_{pc}: 6) requires less information than the other three methods, its performance is second only to that of LSS. The results of PDS (N_{pc}: 6, N_{ws}: 27) and DS (N_{pc}: 4) are quite similar. For this particular data set, the application of the time-consuming moving window technique in PDS brought no extra benefits to the final results. The two methods of calibration model maintenance, LSS and SPEC, gave significantly lower RMSEP values than DS and PDS when evaluated the tablet data, and which was reflected in another application to ternary data (not presented due to space restrictions), even when fewer standardization samples were used.

Although LSS appeared to be superior in maintaining the predictive ability of the PLS models when affected by changes in instrument or measurement conditions, the performance of SPEC was good considering that the method only requires the concentrations of the target constituent in the standardization samples and the corresponding spectra measured at the test conditions, or on the secondary instrument. Consequently, SPEC appears to have a wider applicability than other standardization methods that need the spectra of the standardization samples to be measured under both the calibration and test conditions, or on both the primary and secondary instruments.

Table 1. RMSEP Values (mg)

Models	Calibration Set	Test Set	Validation Set	Entire Set
PLS_1	23.6	22.5	22.1	22.7
PLS_{sub}	8.4	6.6	6.4	7.0
SBC-PLS_1	5.7	6.6	7.6	6.5
PLS_{global}	3.6	4.1	6.1	4.2
PDS-PLS_1	3.4	3.2	6.7	3.6
SPEC-PLS_1	3.6	3.1	5.3	3.4
PLS_2	2.4	2.7	5.2	2.8

6. PAT IN CLOSED LOOP PROCESS CONTROL

With the increasing interest in real time product release based on process analytical and closed loop control technologies there is a critical need for robust data verification, particularly as this information is being included in real-time control or at least advisory feedback applications. *PAT* is driving a paradigm shit across the pharmaceuticals and bio-pharma from a fixed process (variable quality) approach to a new variable process (consistent high quality) approach, signalling the replacement of traditional (lab-centric) production methods to adaptive and flexible production systems that rely on the employment of advanced on-line measurements, advanced control and continuous production process optimization.

Critically, maintaining PAT device integrity in a regulated environment involves the management of real time data, including pre-processing, outlier detection, outlier isolation and record of uncertainty associated with data is vital in a validated environment, to ensure complete traceability of all actions deployed by either a closed loop control system or operator. This management housekeeping, underpins the credibility for any software used for PAT and "Real Time" applications. This is an area that has been considered in depth for safety critical systems, for instance, in the Nuclear Industry, and one that again emphasizes a Process Systems Engineering approach. There is an ongoing development of new ASTM standards (E55) that relate to the Manufacture of Pharmaceutical Products and are forming a detailed practical framework for deployment and management of PAT devices in Pharmaceutical applications. There are existing standards

that include ASTM D6299-09 and D6122-09 that provide a suitable framework for continuous monitoring of analytical instruments in other industries; it is noted, however, that these standards do not employ the multivariate techniques that provide improved fault detection and signal reconstruction capabilities; clearly a major gap.

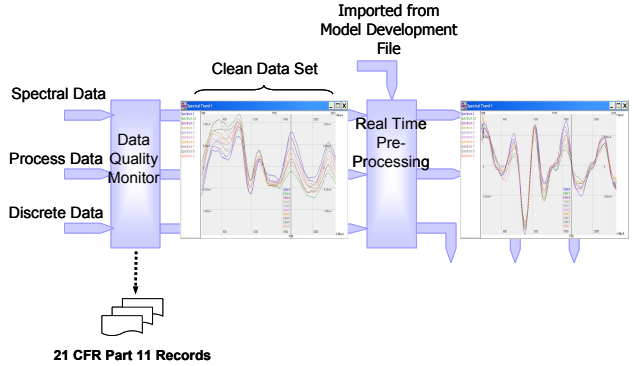

Figure 4. Real-time data management for quality control

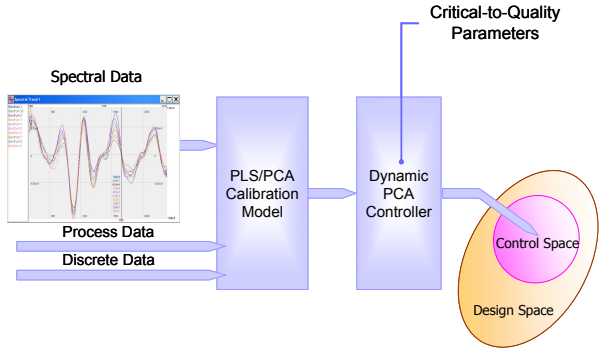

Figure 5. Real Time quality control with spectral data

Major challenges are involved in taking PAT into closed loop process control for Real-time-Release. What data quality monitoring approaches are needed to strengthen the integrity and robustness of on-line models? How will *Critical-to-Quality* parameters be measured – Continuously, or a-periodically, or in real time inferred from calibration model, or from an end-point value inferred from calibration model; or can the scores of a calibration model be interpreted in terms of Q-to-Q parameters?; how will real time data be managed including pre-processing, outlier detection and isolation, the recording of uncertainty associated with monitored data to ensure complete traceability of all actions deployed by either a closed loop control system or by an operator – all within a validated environment? Real-time management of process and spectroscopic data, Figure 4 including robust fit-for-purpose 'transferable' calibration models raise questions such as what will be the impact of controlling temperature using spectroscopic data on control loop performance?; what are the important parts of the spectrum that need to be controlled for reaction monitoring? Such procedures will be essential to underpin the credibility for any algorithms and software used for PAT applications and Real time Release. PAT is part of a tool box to optimise the way pharmaceuticals and biologicals are manufactured, providing greater understanding of the processes involved and what to control, and providing a means to control critical attributes by monitoring and adjusting critical parameters in real time. Figure 5 shows a schematically an industrial application to PAT based control and its relationship to the design and process control space (courtesy Perceptive Engineering).

ACKNOWLEDGEMENTS

ZP Chen acknowledges CPACT and the Strathclyde University Research Development Fund for support.

REFERENCES

FDA (2004), PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, http://www.fda.gov/cder/guidance; "Final Report on Pharmaceutical cGMPs for the 21st Century - A Risk-Based Approach", http://www.fda.gov/cder/gmp/gmp2004/GMPfinalreport2004.htm

Bouveresse, E.; Hartmann, C.; Massart, D.L.; Last, I.R.; Prebblle, K.A. (1996), *Anal. Chem. 68*, 982-990.

Bouveresse, E.; Massart, D.L.; Dardenne, P. (1994), *Anal. Chim. Acta*, *297*, 405-416.

Bouveresse, E.; Massart, D.L.; Dardenne, P. (1995), *Anal. Chem.*, *67*, 1381-1389

Bouveresse, E.; Massart, D.L. (1996), *Chemom. Intell. Lab. Syst.*, *32*, 201-213.

Despagne, F.; Walczak, B.; Massart, D.L. (1998), *Appl. Spectrosc.*, *52*, 732-745.

Chen, Z.P, Morris, J.; Martin, E., (2005), Correction of temperature-induced spectral variations by loading space standardization, Anal. Chem. 77, 1376-1384

Chen, Z.P. and Morris, J., (2008), Improving the linearity of spectroscopic data subjected to fluctuations in external variables by the extended loading space standardization, The Analyst, 133, 914-922

Feudale, R.N.; Woody, N.A.; Tan, H.; Myles, A.J.; Brown, S.D.; Ferré, J. (2002), Transfer of multivariate calibration models: a review, *Chemom. Intell. Lab. Syst. 64*, 181-192

Folestad, S., (1999), "Bridging Science and Regulation - The PAT (R)Evolution in Pharma Industry", *APACT09*, Glasgow.

Forina, M., G. Drava, C. Armanino, R. Boggia, S. Lanteri, R. Leardi, P. Corti, P. Conti, R. Giangiacomo, C. Galliena, R. Bigoni, I. Quartari, C. Serra, D. Ferri, O. Leoni, L. Lazzeri, (1995), Chemom. Intell. Lab. Syst. 27,189-203.

Gemperline, P.J.; Cho, J.H.; Aldridge, P.K.; Sekulic, S.S. (1996), *Anal. Chem 68*, 2913-2915.

Greensill, C.V.; Wolfs, P.J.; Spiegelman, C.H.; Walsh, K.B. (2001), *Appl. Spectrosc.*, *55*, 647-653.

Wang, Y.; Kowalski, B.R. (1993), *Anal. Chem. 65*, 1301-1303.

Wang, Z.Y.; Dean, T.; Kowalski, B.R.(1995), *Anal. Chem.*, *67*, 2379-2385.

Wang, Y, D. Veltkamp, D. J.; Kowalski, B. R. (1991), Anal. Chem., Multivariate instrument standardization, 63, 2750-2756.

Wang, Y,D Y., B. R. Kowalski, (1993) Temperature-compensating calibration transfer for near-infrared filter instruments. Anal. Chem. 65, 1301-1303.

Wülfert, F, Kok, W.T.; Noord, O.E. de; Smilde A.K., (2000a) Correction of Temperature-induced Spectral Variation by Continuous Piecewise Direct Standardization, Anal. Chem. 72, 1639-1644.

Wülfert, F, W. Th. Kok; O . E. de Noord; A. K. Smilde, (2000b) Linear techniques to correct for temperature-induced spectral variation in multivariate calibration. Chemom. Intell. Lab. Syst. 51, 189-200.