Quality by Design in the Pharmaceutical Industry:
Process Modelling, Monitoring and Control using Latent Variable Methods

Theodora Kourti

GlaxoSmithKline & McMaster University

E-mail: kourtit@ mcmaster.ca

Abstract: The pharmaceutical industry has entered a new era. Attention is now being paid to real time process monitoring, real time process control, continuous improvement of processes, and quick product technology transfer. Terms like Quality by Design, Design Space, Control Strategy, Process Analytical technology, Process Signature reflect the current state. Multivariate Statistical Analysis has played an integral part in several industries, enabling process understanding, process monitoring, utilization of real time analysers and real time product release. It is therefore appropriate to see it as an integral part of the pharmaceutical industry effort to address issues like Design Space, Control Strategy, real time process signature monitoring, process understanding and correct technology transfer. In this work it is demonstrated that multivariate, data based statistical methods play a critical role in providing solutions to these issues. From determining the acceptability of raw material entering the plant to ensuring quality of the product that leaves the plant, the multivariate analysis philosophy should govern all the operations that take that raw material and convert it to a final product in a cost efficient way, while meeting safety and environmental constraints, from development to manufacturing to site transfer.

Keywords: Multivariate Process Monitoring, Design Space, process analytical technology, multivariate statistical process control, scale – up, latent variables, process understanding

1. INTRODUCTION

"The pharmaceutical industry has a little secret: Even as it invents futuristic new drugs, its manufacturing techniques lag far behind those of potato-chip and laundry-soap makers" proclaimed the Wall Street Journal in 2003 (Abboud and Hensley, 2003). The article went on to explain that "in other industries, manufacturers constantly fiddle with their production lines to find improvements. But FDA regulations leave drug- manufacturing processes virtually frozen in time. As part of the drug- approval process, a company's detailed manufacturing plan -- and even the factory itself -- must pass FDA muster. After approval, even a tiny change to how a drug is made requires another round of FDA review and authorization, requiring time and paperwork. The process discourages updating by the companies, which worry they will face a production delay that could cost them heavily". The article mentioned FDA as a regulatory agency because it was published in the USA, but similar were the situations with Pharmaceutical companies and other regulatory bodies around the world.

A lot of changes have happened since that article was published and the pharmaceutical industry has entered a new era. The FDA guidance on Process Analytical Technology (PAT) was introduced in 2004 which aims to improve product quality and process performance (manufacturing efficiency) in the pharmaceutical industry; it describes PAT as: systems for the analysis and control of manufacturing processes based on timely measurements during processes of critical quality parameters and performance attributes of raw and in-process materials and processes to assure acceptable end product quality at the completion of the process. (Guidance for Industry: PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. FDA. September 2004)

The introduction of concepts like Quality by Design, Design Space and Control Strategy are also examples of this change. These terms are defined as follows by the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH is a unique project that brings together the regulatory authorities of Europe, Japan and the United States and experts from the pharmaceutical industry in the three regions to discuss scientific and technical aspects of product registration):

Quality-by-Design (QbD) is defined as a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management (ICH, 2008a).

Design Space is the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality (ICH, 2008b).

Control Strategy: a planned set of controls, derived from current product and process understanding that ensures (good) process performance and product quality. The controls can include parameters and attributes related to drug substance and drug product materials and components, facility and equipment operating conditions, in-process controls, finished product specifications, and the associated methods and frequency of monitoring or control (ICH 2008a, 2008b).

The above definitions and actions indicate that the regulatory framework for the Pharmaceutical industry is changing.
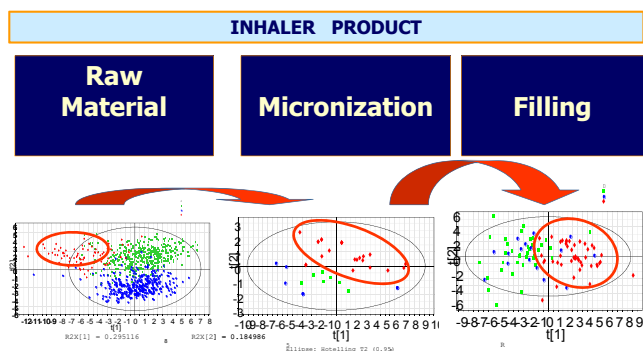

Figure 1. Projection space. Raw material properties, micronization properties and filling performance are projected on a latent variable space. Product batches produced from similar raw material (red) have similar filling performance.
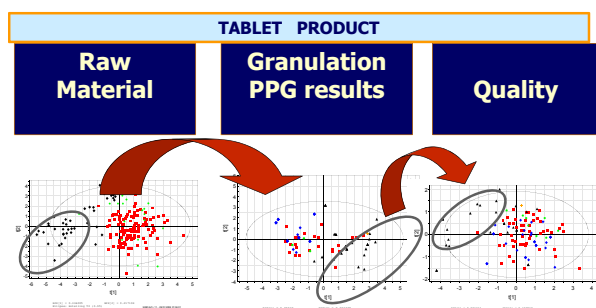

Figure 2. Tablet Product. Batches produced from raw material with similar characteristics, have similar final quality.

The ICH definition of the design space reflects a well known concept, namely that variability in the input of a process will be transferred to the quality of the final product (output) if the process is not controlled to compensate for such variability. Despite the fact that the concept is well known, it requires new ways of thinking in the pharmaceutical industry that was used to dealing with "fixed" processes, as described in the above mentioned Wall Street Journal article.

In this work the role of multivariare statistical methods in modelling, process control and monitoring under this new regulatory framework will be discussed. Multivariate latent

variable methods are shown to be most suitable for process understanding, modelling for Design Space, multivariate statistical process control (MSPC), process control and product transfer. The use of these methods for the development of the Design Space for multi-unit operations will be illustrated in a case where the Tablet Quality is related to API, Excipients, Granulation, Drying and Compression parameters. Examples of how the Control Strategy can be derived from such models will also be shown.

Other topics like Process Signature and MSPC, application of soft sensors, relation of design space to clinical relevance as well as quality by design for analytical methods will be discussed.

2. LATENT VARIABLE METHODS

Latent variables exploit the main characteristic of process databases, namely that although they consist of measurements on a large number of variables (hundreds), these variables are highly correlated and the effective dimension of the space in which they move is very small (usually less than 10 and often as low as 2). Typically only a few process disturbances or independent process changes routinely occur, and the hundreds of measurements on the process variables are only different reflections of these few underlying events. For a historical process dataset consisting of a $(n \times k)$ matrix of process variable measurements $\mathbf{X}$ and a corresponding $(n \times m)$ matrix of product quality data $\mathbf{Y}$, for linear spaces, latent variable models have the following common framework :

$$\mathbf{X} = \mathbf{T}\,\mathbf{P}^T + \mathbf{E} \qquad (1)$$

$$\mathbf{Y} = \mathbf{T}\,\mathbf{Q}^T + \mathbf{F} \qquad (2)$$

where $\mathbf{E}$ and $\mathbf{F}$ are error terms, $\mathbf{T}$ is an $(n \times A)$ matrix of latent variable scores, and $\mathbf{P}$ $(k \times A)$ and $\mathbf{Q}$ $(m \times A)$ are loading matrices that show how the latent variables are related to the original $\mathbf{X}$ and $\mathbf{Y}$ variables. The dimension A of the latent variable space if often quite small and determined by cross-validation or some other procedure.

Latent variable models assume that the data spaces ($\mathbf{X}$, $\mathbf{Y}$) are effectively of very low dimension (i.e., non-full rank) and are observed with error. The dimension of the problem is reduced by these models through a projection of the high-dimensional $\mathbf{X}$ and $\mathbf{Y}$ spaces onto the low-dimensional latent variable space $\mathbf{T}$, which contains most of the important information. By working in this low-dimensional space of the latent variables ($t_1, t_2, \dots t_A$), the problems of process analysis, monitoring, and optimization are greatly simplified.

Multivariate Statistical Process Control is possible utilizing latent variable methods. The following charts are used:

The Hotelling's T^2 for scores (derived either from PCA or PLS models on typical production) is calculated as:

$$T_A^2 = \sum_{i=1}^{A} \frac{t_i^2}{\lambda_i} = \sum_{i=1}^{A} \frac{t_i^2}{s_{t_i}^2} \qquad (3)$$

where $s_{t_i}^2$ is the estimated variance of the corresponding latent variable t_i. This chart essentially checks if a new observation vector of measurements on k process variables projects on the hyper-plane within the limits determined by the reference data.

As mentioned above the A principal components explain the main variability of the system. The variability that cannot be explained forms the residuals (Squared Prediction Error, SPE). This residual variability is also monitored and a control limit for typical operation is being established. By monitoring the residuals we test that the unexplained disturbances of the system remain similar to the ones observed when we derived the model. When the residual variability is out of limit, it is usually an indication that a new set of disturbances have entered the system; it is necessary to identify the reason for the deviation and it may become necessary to change the model.

SPE_X is calculated as :

$$SPE_X = \sum_{i=1}^{k} (x_{new,i} - \hat{x}_{new,i})^2 \qquad (4)$$

where $\hat{\mathbf{x}}_{new}$ is computed from the reference PLS or PCA model. Notice that SPE_x is the sum over the squared elements of a row in matrix $\mathbf{E}$ in equation (1). This latter plot will detect the occurrence of any new events that cause the process to move away from the hyperplane defined by the refence model. The calculation of the limits for the charts is discussed in Kourti (2009).
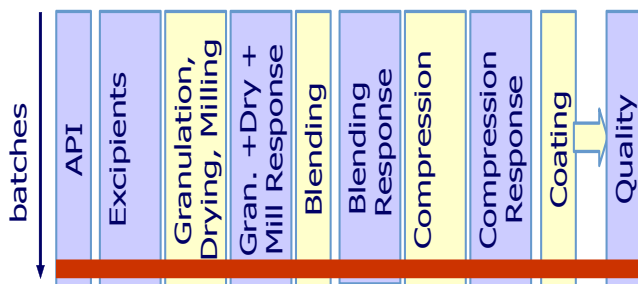

Figure 3. The quality can be modelled as a function of input material and process parameters

These two charts (T^2 and SPE) are two complementary indices; together they can give a picture of the state of the system at a glance. With this methodology, the hundreds of measurements collected from the process variables at each instant in real time are translated into one point for the T^2 chart and one point for the SPE chart (these two points summarize the process at that instant). As long as the points are within their respective limits everything is in order. Once a point is detected out of limit, then the so called *contribution plots* can be utilized that give us a list of all the *process* variables that mainly contribute to the out of limit point, and hence allow us to diagnose the process problem immediately. Contribution plots can be derived for out of limit points in both charts.

A detailed discussion on latent variable methodology for modelling and process monitoring can be found in Kourti (2002, 2005, 2009). Experiences from industrial practitioners can be found in Miletic et al (2004, 2008).

3. PROCESS UNDERSTANDING - EFFECT OF RAW MATERIAL ON FINAL QUALITY

The effect of raw material characteristics in the process performance, if the process operating conditions remain fixed, is demonstrated for an inhaler product utilizing multivariate projection space in Figure 1. The raw material is characterized by several physical and chemical properties. Raw material is produced at three supplier locations and depending on its origin, the data are coloured red, green and blue. The raw material properties are within univariate specifications, at all locations. Projected on a multivariate space, however, they form three clusters, indicating that in a multivariate sense the material possesses slightly different characteristics depending on the location it was produced (covariance structure changes with location). The material properties after micronization are projected on principal components and it can be observed that the material with red coloured origin projects on a different location than the green and blue. The filling performance of the material originating from the red location is different than the rest of the material. A note here that although the control ellipses shown are set by default in the vendor software, they are not interpretable when there is clustering; the assumptions for the calculation of these ellipses are for process monitoring and not for process exploration where there is intentional variation such that introduced by design of experiments.

4. DESIGN SPACE MODELLING

The effect of raw material on the quality as it propagates through different unit operations is shown for a tableting process in Figure 2. When the raw material properties have certain characteristics (marked black) the material projects on a different area. The properties of granules produced from raw material with such characteristics (black) are different from the rest, and the final quality also shows differences.

The difference in the quality can be theoretically explained based on the physical phenomena that govern the whole process. The idea of the design space is to express these phenomena by a model.

The design space can be established as a model that relates input material and process parameters to quality. The model may be theoretical (based on first principles) or empirical, derived from design of experiments or, a hybrid. Together with the model one has to specify the range of parameters for

which the model has been verified. The model may cover one unit operation or a series of unit operations.

The design space for the entire tableting process can be derived by relating quality to the raw material properties as well as to the process parameters of the unit operations (Figure 3). One row in the database depicted in Figure 3 would include the process conditions and quality experienced by the material as it is processed through the units. Multivariate projection methods can be used for the empirical modelling.

It should be emphasised here that the Design Space is a collection of models that relate 1) the final quality to all previous units, raw material and intermediate quality 2) intermediate quality to previous unit operations and raw material.

The empirical models derived are causal and based on carefully designed experiments (DOE). Some DOE's will also be necessary to estimate parameters even if mechanistic models are used.

Batch processes are very common in the pharmaceutical industry. Empirical methods for modelling and monitoring batch processes are discussed in Nomikos and MacGregor (1994) and in Kourti (2003).

Foundations for multiblock analysis necessary for multi unit operation systems can be found in Westerhuis et al (1998).

The level of detail in the models varies depending on the depth of process understanding one wishes to achieve. For example the variable trajectories of a granulation may be described by summary data (min, max, slopes, etc) or by the full variable trajectories aligned against time or another indicator variable (Kourti, 2003).
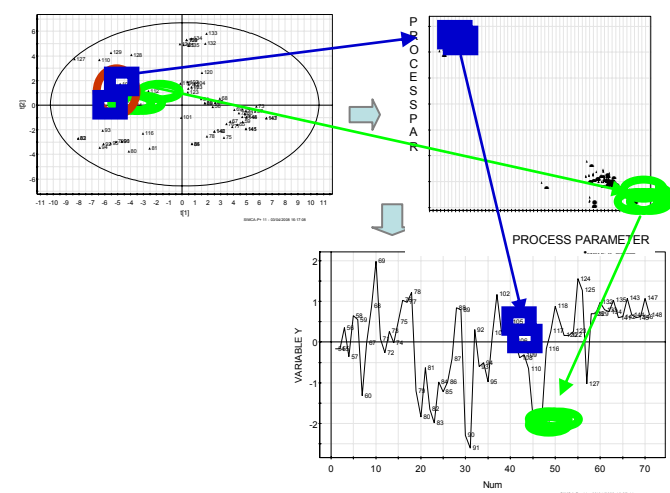

Fig. 4 Control Strategy using Projection Space.

5. CONTROL STRATEGY

Based on the process understanding derived from the design space, control strategy can be derived to assure final quality.

An example in Figure 4 is used to illustrate the new concepts. Control Strategy is devised once the Design Space is established. The example here illustrates a feed forward control scheme for Unit N+1 based on input information on the "state-of-the-intermediate product" from unit N. The settings are calculated and adjusted such that the target value for Quality Y is met.

A multivariate model was built (from batch data) to relate product quality to the process parameters of unit N+1 and the "state-of-the-intermediate product" from Unit N. From this model, a quantitative understanding was developed showing how process parameters in N+1 and the state-of-the-intermediate product from N interact to affect Quality.

Control of batch processes on multivariate space is discussed by Flores-Cerrillo and MacGregor (2004), while product transfer is discussed by García-Muñoz et al (2005).

6. PROCESS SIGNATURE AND MSPC.

It is known from other industries, that sometimes it is not sufficient to characterize a product with "end point quality measurements". The reason is that for some products we do not measure all the possible quality properties (example, downstream processability). The same "measured" quality properties may sometimes be achieved by taking different process paths. In these situations, these different paths may affect the properties that are not measured (i.e. processability). To achieve consistency in all the product properties (measured quality and ability to process down the stream) the process conditions (path to end point) must also be kept in statistical control. When this is not the case, although the measured product properties are on target, the properties that determine other characteristics (i.e., the processability of the product) may not be within acceptable limits. Therefore the "process path to the end point" must also be examined. This "process path to the end point" is also discussed in the European Regulatory Perspective (Graffner, 2005) where it is reported that "during discussions within the industry, the term process signature has been mentioned regularly". To get a common understanding of this, the EU PAT Team had invited public comments on the following definition: "A collection of batch specific information that shows that a batch has been produced within a design space of the product." The EU PAT team mentions as examples of process signatures the amount of water added in relation to time (wet massing), air flow rate, and bed temperature during fall rate drying (fluidized bed drying). They concluded that their understanding is that there is no unique process signature, but instead a family of process signatures with common characteristics (salient features).

It should be pointed out here that the process signature in the multivariate statistical process control context is nothing else

but the two multivariate indices Hotelling's T^2 and SPE. As a matter of fact, these indices take it to account not one feature (e.g., water addition rate or, drying rate) but the combination of all the variables affecting the process and product and their correlations both at each time interval but also their time correlations for the duration of the process (auto and cross correlations for the entire batch). They are therefore a more powerful tool to describe the "overall process signature".

Furthermore, these indices can be directly related to the concept of the design space, as outlined here. The design space model relates raw material characteristics, process conditions and quality. Given the characteristics of the raw material and the desired quality, the design space model can be solved to determine appropriate operating conditions. Maintaining T^2 and SPE within their good operation limits for these appropriate process conditions is nothing more than ensuring that the operation is within the design space.

7. PAT and SOFT SENSORS

Accurate on-line measurements of quality variables are essential for the successful monitoring and process control. However, due to measurement difficulties, sometimes process variables may be used to "infer" product quality in real time and therefore replace an analyzer. This is the idea of soft sensors. In many monitoring and control situations we are often lacking real time sensors capable of measuring many of the responses of interest, because the measurement equipment for such quality variables may be very expensive, or difficult to put on-line, or costly to maintain. As a result we often try to develop soft sensors or inferential models which use other readily available on-line measurements such as temperatures, and can be used to infer the properties of interest in a real time manner. In a recent paper it was demonstrated through application to a benchmark simulation of a fed-batch fermentation process that mutli-way PLS can provide accurate inference of quality variables, such as biomass concentration, that are often difficult to measure using on-line sensors. It was also demonstrated that the same PLS model can be used to provide early detection and isolation of fault conditions within a fermenter (Zhang and Lennox, 2004).

The soft sensors can either replace the hardware sensor (analyzer) or be used in parallel with it to provide redundancy and verify whether the hardware sensor is drifting or has failed; when used in parallel the soft sensor will either estimate the property and compare its value with that of the analyser, or it will keep track of the correlation between the analyser reading and the process measurements.   An example where a soft sensor is used to assess the reliability of an analyser was presented in Kourti (2005). Latent variable modeling was used for this purpose.

This idea of using process measurements as a safety net to verify analyser reliability but also to monitor an index of wellness for the process, to check for unforeseen disturbances, is a crucial and important issue for real time release (Kourti, 2006a).

8. INTEGRATION OF CLINICAL TRIALS

As more complex structures of data are being generated,   the multivariate analysis offers great opportunities for information integration and analysis.

Manufacturing Data as well as patient histories can be integrated and then incorporate into design space the clinical trial responses. (Kourti 2006b).

Figure 5 shows an example of the possibilities that can be explored. Quality in product Y can be related to past information of raw materials, preprocessing and holding times, the type of the vessel used, the operator that run the process, and other recipe information as well as process measurement trajectories and analyzer information.

The quality Y (and details of manufacturing) as well as the patient medical histories and clinical responses can be used to establish a better understanding of the design space.
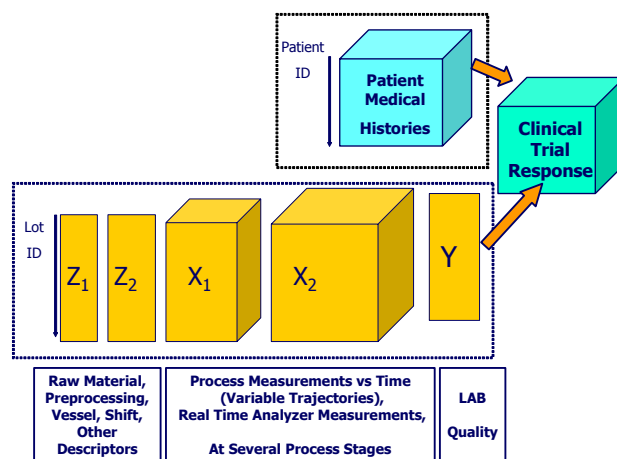

Figure 5 : Examples of complex data structures emerging in industry, that can be mined for a wealth of information. (Kourti, 2006b).

9. QUALITY BY DESIGN IN ANALYTICAL METHODS

The methodology described for design space can be applied in analytical methods. Chromatography, is a laboratory method but also a Unit operation in Bio – Pharmaceuticals.

Process Transfer ideas can be also applied in method transfer ideas, that is method transfer – and site transfer could be treated with similar principals (García-Muñoz et al., 2003)

ACKNOWLEDGEMENTS

REFERENCES

Abboud, L., and Hensley. S (2003) Factory Shift: New Prescription For Drug Makers: Update the Plants; After Years of Neglect, Industry Focuses on Manufacturing; FDA Acts as a Catalyst; The Three-Story Blender. Wall Street Journal. Eastern edition). New York, N.Y.: Sep 3, 2003. pg. A.1, The Wall Street Journal

Flores-Cerrillo J., and MacGregor J. F. (2004) Control of batch product quality by trajectory manipulation using latent variable models, Journal of Process Control 14, 539 -553.

García-Muñoz, S, MacGregor, J.F., Kourti, T. (2005) Product Transfer between Sites using Joint Y_PLS. Chemometrics and Intelligent Laboratory Systems, 79, 101-114.

Graffner, C (2005). PAT- European Regulatory Perspective, The Journal of Process Analytical Technology, 2 (6) 8-11.

ICH (2008a). International Conference on Harmonization, ICH Draft Step 4, Q8 (R1) Pharmaceutical development revision 1, 11-November-2008.

(ICH, 2008b) International Conference on harmonization of technical requirements for registration of pharmaceuticals for human use, ICH harmonized tripartite guideline, pharmaceutical quality system Q10 Step 4, 9-June-2008.

Kourti, T. (2002) Process analysis and abnormal situation detection: from theory to practice. IEEE control systems, 22 issue 5 : 10-25

Kourti, T. (2003) Multivariate dynamic data modelling for analysis and statistical process control of batch processes, start–ups and grade transitions. J. Chemometrics, 17, 93-109

Kourti, T. (2005) Application of Latent Variable Methods to Process Control and Multivariate Statistical Process Control in Industry. International Journal of Adaptive Control and Signal Processing; 19, 213-246.

Kourti, T., (2006a). Process Analytical Technology and Multivariate Statistical Process Control. Index of wellness of product and process. Part 3. Process Analytical technology, volume 3 (3), 18-24.

Kourti, T. (2006b) Process Analytical Technology Beyond Real-Time Analyzers: The Role of Multivariate Analysis Critical Reviews in Analytical Chemistry, Critical Reviews in Analytical Chemistry, 36:257–278.

T. Kourti (2009) Multivariate Statistical Process Control and Process Control, Using Latent Variables. In: Brown S, Tauler R, Walczak R (eds.) Comprehensive Chemometrics, volume 4, pp. 21-54 Oxford: Elsevier

Miletic, I., S. Quinn, M. Dudzic, V. Vaculik, M. Champagne. (2004) Journal Process Control, 14, 821-836.

Miletic, I.; Boudreau, F.; Dudzic, M.; Kotuza, G.; Ronholm, L.; Vaculik, V.; Zhang, Y. (2008) Experiences in Applying Data-Driven Modelling Technology to Steelmaking Processes. Can. J. Chem. Eng. 86, 937–946.

Nomikos, P. and J.F. MacGregor (1994)"Monitoring of batch processes using multi-way principal component analysis," AIChE Journal, 40 , no. 8: 1361-1375.

Westerhuis, J.A., Kourti, T., and MacGregor, J.F. (1998) Analysis of multiblock and hierarchical PCA and PLS models. Journal of Chemometrics, 12, 301-321

Zhang, H., Lennox, B. (2004) Integrated condition monitoring and control of fed-batch fermentation processes, Journal of Process Control 14, 41–50.