Enhanced Process Performance Monitoring through the Combination of Spectroscopic and Process Data

Julian Morris
Centre for Process Analytics and Control Technology
School of Chemical Engineering and Advanced Materials
Newcastle University, UK
www.cpact.com julian.morris@ncl.ac.uk
CAMO PAT Seminar 27th January 2009, London
"Innovative Approaches to Pharmaceutical & Biopharmaceutical Development and Manufacturing"

Overview of Presentation

- Process Analytical Technologies:
  - Where have we come from? Where are we now? Where are we going?
  - Some challenges in spectroscopic data pre-processing and modelling
- Multivariate Process Performance Monitoring
- Variability - Different Product Formulations, Recipes, Processing Units, Production Sites, Spectroscopic Probe Locations, …
- Integrating (Fusion) Spectroscopic and Process Data
- Closing the Analytical Control Loop
- Closure

The EU provides 32% of the world’s chemicals manufacturing through some 25,000 enterprises of which 98% are SMEs which account for 45% of the sectors ‘added value’, and 46% of all employees are in SME

What does PAT and QbD mean to an SME?
Where are we in Process Analytics and Control Technologies?

Impurities and Polymorphism (Where we were in 1998)

Impurities and Polymorphism

- Impurities effect nucleation and growth processes, and hence can stabilise meta-stable polymorphic forms.
- As product purity improves during process chemistry work-up, the “stable” polymorphic form can change!
- e.g. RITONOVIR aids drug which changed from anhydrous to hydrate crystal after launch:
  - with lower solubility and hence bio-availability.
  - product was withdrawn for a year and reformulated.
  - new FDA approval needed – mega cost implication!

Capsules are available from well-sugar

The problem relates to “undeniable” crystal formations. Abbott says that a series of random production batches of NORVIR capsules failed the approval test for dissolution. The researchers in Abbott’s research labs investigated the production process. It turns out a certain lot of raw material was used, which affects the way it dissolves and possibly its absorption. Following were some points from a number of random batches of capsules were examined and there is no evidence that the dissolution issues have been resolved.

Much involved in managing the situation.

Where are we in Process Analytics and Control Technologies?
Process Analytics – An Observation

Costs the same as

A FTIR – €100,000

A Mercedes S Class - €100,000

While

Lambda Sensor

Mercedes S Class costs €100!

Process Analysis is restricted to large companies while 98% of chemicals manufactured in Europe is by SME’s, primarily in Batch Plants.
Advanced Chemometric Methods

In-Process Analytics & Process Control

CBBN2 Project: collaboration with Leeds, Heriot-Watt and Newcastle University

Partners
- AEA Technology
- AstraZeneca
- Bede Scientific Instruments
- BNFL
- Clairet Scientific
- DTI
- EPSRC
- GlaxoSmithKline
- HEL
- Malvern Instruments
- Pfizer
- Syngenta

- FTIR
- Supersa
- Turation
- Size
- Growth
- Kinetics
- USS
- Batch
- Process
- Monitoring & Control
- Process Conditions
- Heat Transfer

Spectral Calibration Issues

- Spectroscopic measurements in chemical and pharmaceutical processes are always liable to fluctuations in both control and external process variables.
- This can result in noisy spectra, non-linear shifts, broadening in spectral bands and multiplicative light scattering perturbations.

X-ray diffraction profiles of mannitol-methanol suspensions with the content of mannitol varying from 0.0% to 5.0% g/ml.

Five NIR spectra for a ternary mixture sample measured at five different temperatures.

20 NIR spectra of a powder mixture measured under different sample compactness.
Smoothed Principal Component Analysis (SPCA)

- Enhancing signal-to-noise ratio

\[ X_i^T X_r = \lambda_i \times (1 + k \times Q^T Q) \]

\[ F = [r_1, r_2, \cdots, r_m] \times [r_1, r_2, \cdots, r_m] \]

\[ x_r = F x_s + F x_s \]

Raw (a) and Pre-processed (b) XRD profiles of alpha form GA-methanol slurries with concentration varying from 0.02% to 10.00%; The relationships are shown between Raw (a) and Pre-processed (b) XRD profiles of alpha form GA-methanol slurries with concentration varying from 0.02% to 8.00%; The raw (a) and Pre-processed (b) XRD profiles of the beta form of GA-mannitol-methanol suspensions with the contents of mannitol equaling to 0.0%, 0.178%, 0.389%, 0.533%, 0.800% and 1.000% g/ml, respectively. Raw (a) and Processed (b) XRD profiles (by SPCA) for 6 XRD data sets of mannitol-methanol slurries with the contents of mannitol equaling to 0.0%, 0.178%, 0.389%, 0.533%, 0.800% and 1.000% g/ml, respectively.

**Case Study 1 – SPCA in Morphology Monitoring**

Raw (a) and Pre-processed (b) XRD profiles of alpha form GA-methanol slurries with concentration varying from 0.02% to 10.00%. The relationships are shown between concentrations and peak heights at peak A1 of the (c) raw and (d) processed spectra.

The raw (a) and Pre-processed (b) XRD profiles of the beta form of GA-methanol slurries with concentration varying from 0.02% to 6.00%. The relationships between concentrations and peak heights at peaks B3 of the raw (c) and processed (d) spectra.
System provides capability to monitor polymorphic form “in-process”, i.e. that unaffected by product separation prior to analysis. Typically circa 1 wt % detectable via in-process XRD, much lower with advanced chemometric analysis (Smoothed PCA).

250L Pilot Plant Batch Agitated Vessel

Münchwilen Foxboro Control System as Set-Up for CBBII Trial on 250 Litre Reactor R-122
Supersaturation Control System Upgrade to PI Capability

Supersaturation Control of L-Glutamic Acid
250 litre Plant Crystalliser

L-Glutamic Acid Crystals
(a) Seeds (b) Product
Multivariate Statistical Process Control (MSPC) or Multivariate Process Performance Monitoring

Variability (or PAT) by Edwards Deming

Cease reliance on mass inspection to achieve quality.

Eliminate the need for mass inspection by building quality into the product in the first place.

Dr W. Edwards Deming
(Circa 1980s)

“Learning is not compulsory, … Neither is survival”

Statistical Process Control (SPC) and MSPC

- A process is said to be in "statistical control" if certain process or product variables remain close to their desired values and the only source of variation is "common cause" variation.

Monitor for ‘Special Causes’

- Changes in raw materials – variability in process inputs.
- Equipment degradation, impurities, fouling, etc.
- Process changes and unmeasured disturbances.
- Control loop and instrumentation problems.
Feedback Process Control and Statistical Process Control

Process Control is NOT the same as Statistical Process Control

- Process Control moves variability from the product to part of a process where it can be tolerated.
- Process Control involves process measurements and feedback.
- (M)SPC assumes a good process control system.
- (M)SPC compensates for process and equipment disturbances.
- (M)SPC provides for infrequent, operator driven, supervision and feedback.

Monitoring with Score and Squared Prediction Error (Q-Statistic) Plots

- A very effective set of multivariate control charts is the Shewart chart on the dominant orthogonal PC’s (t1, t2, ..., tA) plus the SPE (Q-Statistic) Plots.
- The structures of the SCORE and the SPE (Q-Statistic) plots reflect two ways in which non-conforming process behaviour can be identified:
  - If the process change is caused by a larger than normal shift in one or more of the process variables, but the basic relationship between the quality and/or process variables remains unchanged, then a translation in the score plane will result, with the SPE (Q) remaining at an acceptable level.
  - If the process change enters through a new event not captured in the reference data set, it will change the nature and possibly the dimension of the relationship between the process and/or quality variables, then the SPE (Q) will increase.
Plant Signatures: Cusum Scores Plot  
(IEE Proceedings, 1996)

- Fault 'directions', 'profiles' or 'signatures
- Nominal Operations
- Solvent Malfunction
- Reactor Fouling
- Impurity Fault
- Combined Impurity and Fouling

Variable Contribution Plots

\[ t_1 = p_{11}, \text{temperature} + p_{12}, \text{pressure} \]

Contribution of Variable 1 to Principal Component 1
Contribution of Variable 2 to Principal Component 1

\[ t_2 = p_{21}, \text{temperature} + p_{22}, \text{pressure} \]

Contribution of Variable 1 to Principal Component 2
Contribution of Variable 2 to Principal Component 2

Scores Plot of PC1 versus PC2 for 21 Different Operating regions
On-line Process Performance Monitoring

Contribution Plot

Scores Plot of PC7 versus PC8

The identification of Cluster 20 provided savings of circa £1M per year
Case Study - Intelligent MSPC

- Intelligent MSPC resulted from a 12 month joint EU supported ESPRIT project with Computas of Norway, British Steel (now Corus) and CPACT.

- The aim was to integrate into the G2 Real-Time Expert System the CPACT MSPC (PCA and PLS) developments.

- A G2 based system developed at Newcastle in collaboration with a major Pharmaceutical company has also been installed on a fermentation process.
Process Diagnosis – Intelligent MSPC (iMSPC)

Performance Monitoring of a Blast Furnace

Variability with Different Product Formulations / Recipes,
Different Processing Units,
Across Different Production Sites,
Different Spectroscopic Probe Locations, …


The Integration (Fusion) of Spectral and Process Data


Choi, S. W., Morris, J., Lee, I. B. (2008), "Nonlinear multiscale modeling for fault detection and identification", Chemical Engineering Science, 63, 2252 – 2266

Multiblock Approach

Super Block T
Super Scores
Super Level

Process Block Scores
Spectroscopic Block
Other Blocks Scores
Base Level
The Integration of NIR Spectroscopic and Process Data in an Industrial Multi-product Multi-recipe Polymer Resin Manufacturing Process
Process Details
- The process comprised combinations of 5 different reactors processing 4 different viscosity groups.
- Nominal batch data from two viscosity groups is discussed here.
- Thirty batches were used in the development of the nominal models:
  - 16 batches - viscosity group 2
  - 14 batches - viscosity group 4
- All seven process variables were included in the analysis.
- A single, a combined and a common subspace model were investigated.
- A 'non-standard' batch with a temperature control problem was investigated.

Different Production Reactors & Different Viscosity Groups
- Bivariate Scores plot of production batches from five different reactors (PC1 versus PC2)
- Bivariate scores plot of batches from four different viscosity groups (PC1 versus PC2)

Monitoring of Hydroxyl Number
- The path of a batch requiring a corrective addition is monitored.
- When the path leaves an upper confidence limit, an addition is made. This procedure adds about two hours to the total batch time.
- The red curve demonstrates how the corrective addition breaks up long established chains, increasing the hydroxyl number and decreasing the viscosity temporarily.
- The quantity of acid or glycol added is based on a calculation made by the process operator.
- Once the system stabilises, the reaction proceeds to the end zone.
- Samples are taken periodically to ensure that the batch is progressing to plan.
Monitoring of the Hydroxyl Number

Ideal Reaction

Corrective addition

Monitoring of Individual Viscosity Group 2

Process Fault Detection for Individual Viscosity Group 2
Thirty-seven batches were modelled from four viscosity groups.

Analysis of the traditional process data from the thirty-seven available data sets revealed that thirty-two of the batches are considered to exhibit nominal operating conditions.

NIR data was also available for these batches.

A ‘non-standard’ batch was investigated – the duration of the batch was longer than desirable.
The Conjunction of Spectral and Process Data

- A number of techniques have been investigated in a number of industrial studies including crystallisation, polymerisation, fermentation and speciality chemicals.
  - The conjunction of single-source spectral and process data using multi-block techniques.
  - The conjunction of multi-source spectral data (e.g. NIR, MIR, XRD, ultrasonic particle size spectra) with process data.
  - The use of wavelet transformations for the conjunction of spectral and process data.
  - The building of robust calibrations through bootstrap aggregation methods (bagging).
PCA Applied to Raw NIR Data Set

- PCA of the raw, un-scaled spectral data can be used to identify the more significant wavelengths.
- The wavelengths of bond frequencies identified are characteristic of the chemistry that exists within the resins polymerisation reaction.

NIR Wavelet Transformation

- In this particular application, a wavelet transformation of the NIR spectra was used to extract the important features that occurred at different frequencies within each spectrum.
- NIR data was collected over the range 1078nm to 2100nm so that each spectrum now had a dyadic length of 1024 columns and a wavelet transformation was performed using a Symmlet #8 mother wavelet.
- The approximation and detail coefficients were investigated at each level, with the most suitable group of coefficients being identified as the detail coefficients at level 5.
Conjunction of Process and Spectral Data

- The investigation was repeated to include a set of variables from the wavelet transformed NIR data.
- At level 5, the coefficients summarise information from around 22 wavelengths.
- The two areas of interest lie around 1450nm and 1900nm wavelengths.
- To capture all of the information from the transformed spectra, the two detail coefficients from either side of these important wavelengths are extracted for inclusion in the model.
- This leads to the addition of four extra variables to the nominal data set.
Batch Monitoring using Process and Spectral Data

Scores Contribution Plot for PC1 Observation 30

- Variable numbers 7 and 9 lie outside their 99% confidence limits.
- These variables represent the wavelet detail coefficients of the OH and COOH bond frequencies.
- The process variables (variables one to six) lie well inside their respective confidence limits.

SPE Contribution Plot at Time 30

- The SPE contribution plot for time point thirty identifies variables seven, eight and ten as contributing to the batch trajectory.
- Again, these variables are all associated with the quality information extracted from the NIR data set.
Potential Industrial Impact

- Where modelling traditional forms of process data in a MSPC monitoring scheme offers the potential for detecting faults in operating conditions, the inclusion of chemical information, in terms of wavelength data, allows for faster detection of abnormalities in the chemical make-up of the reactor contents.
- The inclusion of chemical information, even from a range of on-line real-time analysers when they can be justified on a business benefits basis, potentially offers significant advantages to using solely traditional process variables or single analysers.
- Integrating spectroscopic data with process data, properly, is not straightforward.
- Where modelling traditional forms of process data in a MSPC monitoring scheme offers the potential for detecting faults in operating conditions, the inclusion of chemical information allows for detection of abnormalities in the chemical make-up of the reactor contents.

Model Based Multivariate Statistical Process Control

or

Dynamic Model Based Process Performance Monitoring

The Impact of Process Dynamics (Auto-correlated Data)

- In the presence of auto-correlation, the action and warning limits calculated for a process performance monitoring scheme based on the usual assumptions that the observations are Independent and Identically Distributed (IID), are inappropriate making the monitoring charts unreliable, if not useless.
- Developing an MSPC representation for the monitoring of processes, exhibiting time varying characteristics, will result in an increase in the number of false alarms and operational changes not being detected.
- So, how do we deal with auto-correlated data?
Model-based Performance Monitoring

At the heart of a model-based control or monitoring system is a DYNAMIC model.

Impact of Non-linearity – Normal Probability Plots

Impact of Serial Correlation - Partial Auto-Correlation Plots
Closing the Analytical Control Loop

Incorporating PAT Sensors into Real Time Process Control

Data Quality Monitoring

Univariate Data Quality Monitor:
- Individual Signal Validation
  - Logical Checks
  - Statistical checks

Multivariate: Using Robust PCA
- Outlier detection
- Outlier Identification
- Data Quality Monitoring

21 CFR part 11 Records

Data Quality record underpins the validity of the system
– critical in a validated environment

Courtesy Perceptive Engineering

PAT and Advanced Process Control
Why Process Systems Engineering is Important

- PAT can be used as part of a tool box to optimise the way pharmaceuticals are manufactured
- Provide greater understanding of what to control
- Potentially provide a means to control "Critical Attributes" by monitoring and adjusting "Critical Parameters" in real time
- Provides some of the ability, to reduce the risk of process variability, effecting process capability and product quality

Current Model
- Variable Raw Material
  - Input to process
- Fixed Process
- Variable Output

Variable Process Model
- Variable Raw Material
  - Input to process
- Variable Process
- Consistent Output

Continuous Quality Verification
**Real Time Quality Control and Integrated Data Management**

- Real time pre-processed data
- Spectral Data
- Process Data
- Discrete Data
- PLS/PCA Calibration Model
- Dynamic PCA Controller
- Control Space
- Design Space

CtQ Parameters
- Continuously measured OR
- A-periodically measured OR
- Real time value inferred from calibration model OR
- End-point value inferred from calibration model OR
- Scores of calibration model are CtQ parameters

**PAT in Closed Loop Process Control - Some Challenges**

- Real-time management of process and spectroscopic data.
- Real time robust fit-for-purpose 'transferable' calibration models.
- No control system is going to control a spectrum of several hundred simultaneous values; so what is important?
  - Is there a robust fit-for-purpose calibration model to infer specific product properties?
  - Are there particular features / segments of the spectrum of interest?
- What is the impact of process control variables, e.g. temperature on spectroscopic calibrations on control loop performance?
- What is the impact of process conditions on optical light scattering and hence spectroscopic calibrations on control loop performance?

**Acknowledgements**

- The EU and the Consortium Member Companies for providing the research challenges and funding for the Multi-product and Spectral Integration study, in particular Unilever.
- The EPSRC and the Chemicals Behaving Badly (CBBII) consortium for providing the research challenges and funding for the Crystallisation studies.
- Dr Ahmed Al-Alawi for his contributions to the CBBII project.
I will be most happy to answer your questions.

Many thanks to CAMO for their kind invitation and of course you for your kind attention.

APACT09 Glasgow
May 5th to 7th 2009

EUROPACT 2011 Glasgow
April 26th to 29th 2011