MONITORING FORM CONVERSION THROUGH BATCH LEVEL PROCESS ANALYTICAL DATA

Taniya Mandal[1], Jo Ann Coleman[1] and James Rydzak[2]

[1] Statistical Sciences of North America
GlaxoSmithKline
709 Swedeland Road, King Of Prussia, PA 19406, USA
also
Department of Statistics
University of Georgia, Athens, GA, USA
(e-mail: taniya.2.mandal@gsk.com)

[2] Statistical Sciences of North America
GlaxoSmithKline
709 Swedeland Road, King Of Prussia, PA 19406, USA
(e-mail: jo.ann.n.coleman@gsk.com)

[3] Chemical Development, PPSE
GlaxosmithKline
709 Swedeland Road, King of Prussia, PA 19406, USA
(e-mail: james.rydzak@gsk.com)

ABSTRACT. Near infrared (NIR) spectroscopy can be used for monitoring of the form conversion of a compound. A range of wavelengths is required to characterize the conversion. Multivariate data analysis techniques such as Principal Components Analysis, (PCA), compress high-dimensional, correlated variables to lower dimension while retaining the important information. Additionally, Partial Least Squares (PLS) modeling on the time course data for each batch and PCA, on batch level scores, are used to visualize differences as the material converts from monohydrate to anhydrate over time and across batches. The advantage of these techniques is in being able to capture the complexity of the chemical changes through multiple wavelengths while still reducing the monitoring task to a few simple graphs.

1 METHODS

A NIR spectrometer equipped with a diffuse reflectance probe was interfaced to a process vessel to monitor the form conversion of 14 batches manufactured during a campaign. Batch level analysis of the NIR spectral data using the SIMCA suite of software provides a tool to follow the batch evolution, and look at the overall data to see the consistency of the batch behavior. Data was time aligned relative to the start of the process in SIMCA P+. A series of spectral pre-treatments were applied to these data. The data presented here are 1st derivative (to account for changes in baseline), quadratic, 25 point (for smoothing purposes) processed using only the regions from 5150-4800 cm-1 and 6200-6000 cm-1.

2 OBJECTIVES

This data was collected for proof of concept as to whether form conversion could be detected for the conversion. In addition it was of interest to test whether NIR derived process signatures could detect differences between batches. Other measures of conversion and process performance displayed considerable variability. Use of spectral data was proposed as potentially more reliable. Models were constructed to look at both the evolution of the batches over time and to compare batches at completion.

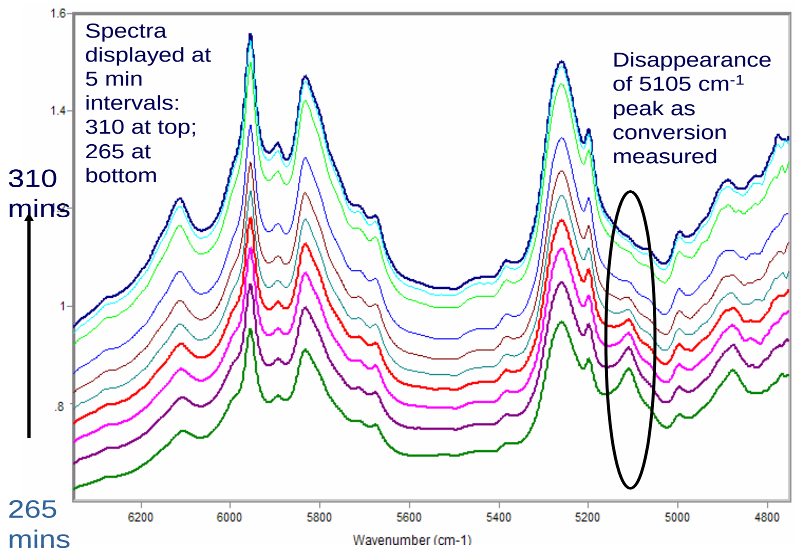

Figure 1. NIR spectrum of the compound with Form Conversion.It captures the full NIR spectrum of the product as it evolves over time for a typical batch. The series of 10 spectra are separated by 5 minute intervals.

3 Data Acquisition

NIR spectra were recorded working in diffuse reflectance mode. 14 independent batches were measured 'in situ' in real time. The purpose was to detect the conversion time from Form A to Form B. Multiple time points were acquired for each batch to monitor the conversion from Form A to Form B. Data was collected at 1 minute intervals. NIR data was normalized by means of first derivative in order to correct variability of the baselines due to scattering phenomena using SIMCA P+. There are 2077 variables i.e. total derivatized wavelengths and 10890 rows of data which is real time observations. Hence is the need for dimension reduction and exploratory analysis in variable space and observation space using PCA.

4 RESULTS

4.1 OBSERVATION LEVEL MODELING:

In order to accomplish observation level modeling, the batch data is unfolded, preserving variable direction to a two-way data table. This type of unfolding has previously been shown to be successful for detection of abnormal batches (Wold, [1987], [1998], [2001] and ndey et al., 2003). As described above, modeling was performed by projecting observations on the hyperplanes and translating them into latent variables.

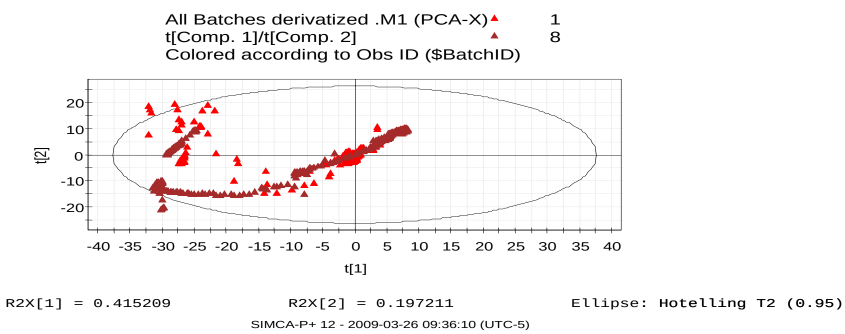

Figure 2. Scores Plot for Batches 1 and 8 showing Evolution in time of all the variables of the two batches. The batches start at the centre of the ellipse and progresses outward and once they convert, they loop back towards the centre again.

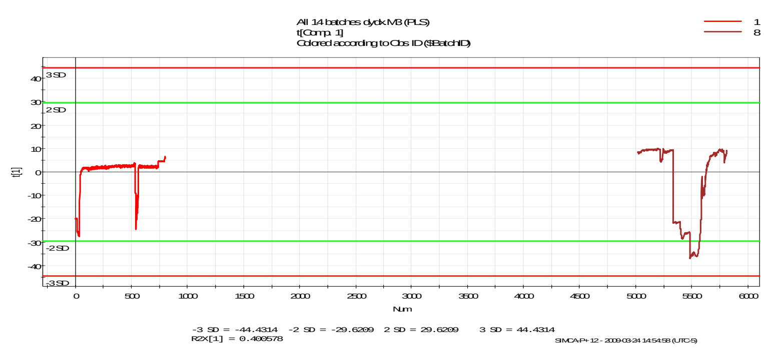

Figure 3. Line Plot of PCA 1 for Batches 1 and 8 indicating Batch Progression Over Time and Form Conversion

Batch 8 has been recorded as an aberrant batch out of all the 14 while Batch 1 is considered the standard. As we can see, Batch 8 has inherent variability and does not plateau after converting its form.

4.2 BATCH LEVEL MODELING:

At the batch level, all available data are used to model the whole batches as units. In contrast to the previous level, where each row corresponded to one time point in one particular batch, each row in a data table at the batch level now represents one whole batch. In creating a batch level model you can choose to use the original variables or the scores derived from the lower observation level model. In order to accomplish batch level modeling the batch data table is unfolded preserving batch direction. These models use the initial conditions plus summaries of the batch evolution as variables. Each batch is summarized by a set of batch level score values. Batches that cluster close to each other in the score space are similar and those far apart are different. For example shifts in raw material quality or step changes in media formulation are summarized as shifts or clusters in the score space.

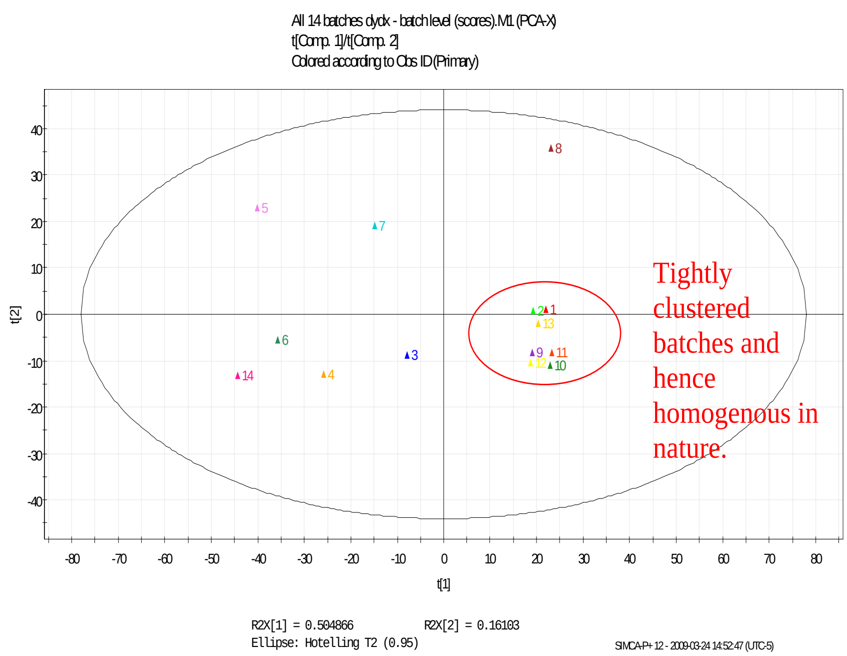

Figure 4. Scores Plot of the first two Principal Component Indicating Batch Similarity when rolled into batch level scores.

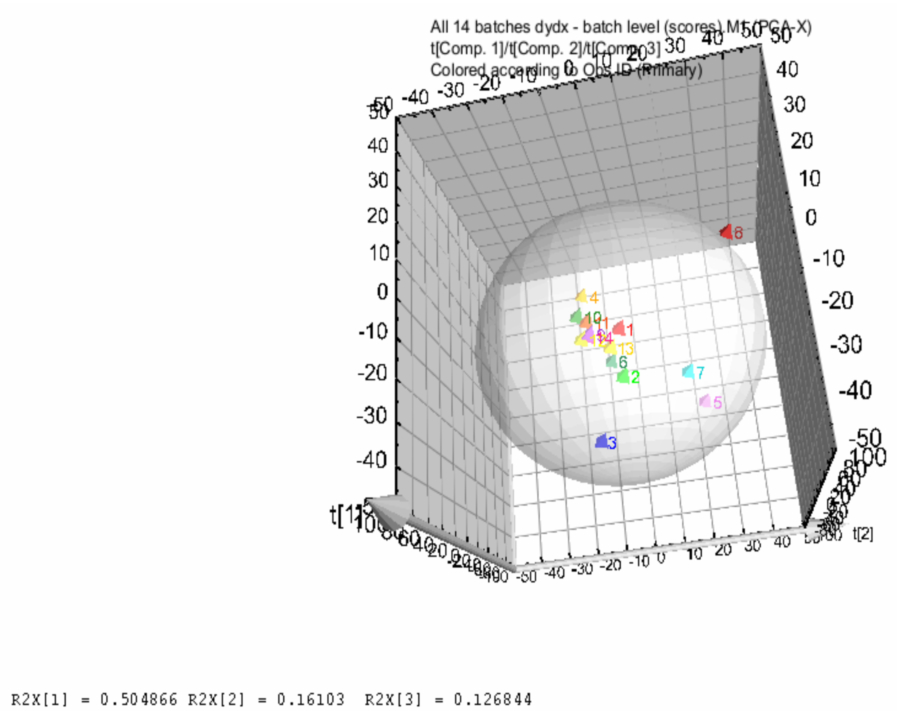

Figure 5. A 3D plot using first three principal components showing that batch 8 does not conform with the rest of the batches and lies outside the sphere.

5 Conclusions

We conclude that:

1. Multivariate Data Analysis enabled the product team to achieve improved extraction of knowledge from data.

2. Through data visualization and trending, we have been able to see the evolution of batches over time.

3. From Figure 2., we have concluded that there is profound amount of variability within batch 8.

4. Through the time trends graphics (Figure 4), we can trace a batch over time and it also leads us to the average conversion time from Form A to Form B for each batch.

5. Figure 5 suggests that the batches vary in correlation with one another. The batches 1 and 12 are tightly clustered suggesting that they are chemically similar which is also expressed via NIR method. Also, we can conclude that there exist common cause variation among the batches but no special causes.

6. Figure 4 and Figure 5 suggests that batch 8 is different from the rest of the batches.

REFERENCES

GLAXOSMITHKLINE RESEARCH and DEVELOPMENT (2007-2008).

Copyright SIMCA P+ version 12:Software tool for Multivariate Modelling.