Real-time NIR Monitoring of a Pharmaceutical Blending Process through Multivariate Analysis-derived Models

PIERRE-PHILIPPE LAPOINTE-GARANT [a,b], JEAN-SÉBASTIEN SIMARD [b] and NICOLAS ABATZOGLOU [a,(*)]

[a] Department of Chemical Engineering, Université de Sherbrooke, Sherbrooke, Quebec, CANADA
[b] Technology Department, Wyeth Pharmaceuticals, Saint-Laurent, Quebec, CANADA

(*) Corresponding author

***Abstract*: -** Process analytical technology guidance by the US Food & Drug Administration lately became the major driver of pharmaceutical process optimization. The majority of these processes are complex and consequently multivariate. Although new insights have improved knowledge of the phenomena taking place, it is not usually possible to develop deterministic modeling. Processes involving powder handling, such as multi-component pharmaceutical formulation blending, are common and the real-time monitoring of their physico-chemical attributes is challenging. In this work, we propose multivariate analysis of a V-blender mixing unit operation with an in-line near-infra red (NIR) measurement technique.

The NIR measurement system used in this study consists of a micro-electro-mechanical system (MEMS)-based spectrometer connected to an IP-65-encased optical measuring head (sampling probe) through a 1-meter length umbilical wire cord. It deploys diffuse reflectance sampling technology, providing 40-mm spot size with a spectral range of 1,350 to 1,800 nm. The methodology includes the following steps: (1) modification of a nominal 1-ft³ (30-L) V-blender unit to accommodate Axsun's NIR spectroscopy system; (2) 3 experimental runs, each with a different mixing time, while monitoring powder homogeneity with NIR spectroscopy; (3) acquisition of 10 powder samples after each run from pre-determined locations in the V-blender, evaluated with current Quality Assurance (QA)/Quality Control (QC) lab methods, to determine reference mixing endpoints; and (4) NIR data analysis by SIMPA-P+ and GRAMS chemometrics software and comparison with reference mixing time. Two qualitative algorithms (analysis of spectral variance, distance analysis with Hotelling T^2) for real-time homogeneity determination are developed, and their efficiency evaluated.

The size of the acquired information is not comparable to classical "thief analysis", and the result (prediction of the mixing endpoint) with the recommended methods proved to be equally or more efficient than with actually-employed quality control protocols. In addition, this information can be obtained in real-time using chemometric models. The time savings are important when compared to classical laboratory analysis (such as high pressure liquid chromatography analysis). It is expected that any one of the presented NIR analyses can be beneficial on many aspects of pharmaceutical blending, such as: (1) Real-time quality monitoring of current manufacturing batches; (b) Improved process efficiency and performance by selecting adequate process parameters and blending time; (3) Quality by design initiatives during the development of blending processes for new formulas.

Key-Words: - Process analytical technology (PAT), Multivariate data analysis, Chemometrics, Pharmaceutical processes, Powders, Mixing, Blending, Quality by design, Real-time monitoring

1 Introduction

In a typical oral contraceptive pill, the active pharmaceutical ingredient (API) has a weight concentration between 0.01 and 0.05 wt%. For a batch from a typical 20-ft³ (about 0.5-m^3) V-blender, this represents mixing only a few hundred g of APIs in over 300 kg of excipients. In several vitamin formulations currently available on pharmacy counters, over 30 ingredients are mixed together, each having a concentration ranging from 0.01 to 50 wt%. These formulas are often mixed in tumbling blenders, such as V-blenders, in batches of over 2,500 kg where, ideally, every ingredient must be present at the right concentration in a sample size of approximately 1.2 g.

These 2 examples represent some challenges that the pharmaceutical manufacturing industry faces in the production of large-scale formulations for consumers. In each case, the consumer expects to have the exact claimed concentration in each tablet. With such

challenges in mixing unit operations, why does the pharmaceutical manufacturing industry remain one of the least efficient in in-line or on-line process control? Luckily, the quality of the pharmaceutical tablets or pills sold to the consumer is rarely deficient, thanks to very tight quality procedures and controls, but controls are mostly conducted on the finished product and not during manufacturing steps.

The process analytical technology (PAT) initiative aims at solving this issue by promoting analytical tools located directly on the process to monitor quality attributes or critical product parameters. In the case of mixing unit processes, fundamental understanding of the mixing mechanisms can help in controlling the mixing endpoint, thus ensuring good mixture homogeneity.

This fact has been outlined in the trial of the United States of America versus Barr Laboratories in 1992-93. In this case, Barr Laboratories was accused of adulteration of its product, validation failures and current good manufacturing practice (cGMP) irregularities, particularly for blend uniformity testing: see Muzzio and Robinson's comprehensive review [1]. Since then, many initiatives have been taken to specify correct blend uniformity assessments and controls. One such initiative is the Blend Uniformity Working Group (BUWG) from the Product Quality Research Institute. By studying pharmaceutical blending processes and by conducting a discussion forum with the US Food & Drug Administration and different pharmaceutical manufacturers, the BUWG issued recommendations for blend uniformity testing in accordance with cGMP and United States Pharmacopeia guidelines. The following guidelines were identified for validation batches: at least 10 sampling locations identified in the blender representing potential areas of poor blending and collecting at least 3 replicate samples for each location. The acceptance criteria identified as relative standard deviation (RSD) of all API assays at most 5.0%, and that the RSD of all individual results must be within 10.0% of the mean. Also, it was recommended to have in-process analysis for routine batches.

While many pharmaceutical manufacturers are in compliance with the acceptance criteria recommended by the BUWG when performing validation batches, very few respect routine in-process control since most of them still use the typical sample thief. These tools are time-consuming, require expensive laboratory analyses and increase pharmaceutical exposure of the operators. Moreover, it is widely recognized that the sample thief can bias the results by introducing a perturbation in the powder bed and by suffering from segregation mainly due to particle size when flowing into the sampling aperture. Many studies were conducted to try to alleviate the problems associated with thief sampling techniques [2,3]. However, none has truly solved the issue for routine batch process control.

That is why many pharmaceutical manufacturers are looking for other tools to assess blend homogeneity for production batches. They are now opting for near-infra red (NIR) spectroscopy. This interest is mainly justified by the fact that NIR can provide fast, non-invasive and non-destructive measurements of powder samples. Samples can thus be analyzed in-line without the need for sample preparations. Scans can typically be performed in less than 100 ms, and the instrument itself can be fairly compact and robust. New micro-electro-mechanical systems **(**MEMS)-based spectrometers are small and require no moving parts. The instrument can thus be fitted directly on the shell of typical tumbling blenders. Therefore, the instrument can perform in-line spectral analysis and assess mix quality during mixing.

But why are the pharmaceutical manufacturers only now developing PAT with NIR technology? This is mainly due to the development of computerized mathematical tools called chemometrics. Chemometrics can be defined as: "… *the chemical discipline that uses mathematical, statistical and other methods employing formal logic to design or select optimal measurement procedures and experiments and to provide maximum relevant chemical information by analyzing chemical data"* [4].

When NIR spectroscopy was compared with typical infra red (IR) or even Raman spectroscopy prior to the advent of chemometrics, NIR was a far less efficient instrument to quantify or even qualify any component in a solid state. This fact can be explained by the wide and broad band peaks that reside in NIR regions. NIR has less specificity than many other types of spectroscopy available in a laboratory environment. It is why very few banks of information exist that would identify the absorption regions of NIR molecular vibrations compared to IR spectroscopy. However, with chemometrics, precise and quality data can be obtained by correlating multiple values of absorbance to identify and quantify a component. Instead of employing 1 peak to quantify a component, the variation of many peaks is what correlates the information [5]. NIR standards and references are required to help multivariate analyses link spectral variations with the sought information (i.e. component type and concentration in a complex mixture).

Many types of chemometrics can be developed, and each will yield its own type of information. The key is

to select one that gives the best and most robust prediction model. These chemometric models can be divided into 2 main classes: qualitative and quantitative.

No direct quantities are measured with qualitative chemometrics. Qualitative analysis provides mainly categorical information; this means that it assesses whether the analyzed process passes or fails quality control norms. For example, in a pharmaceutical mixing unit operation, it can determine if the quality of the blend is acceptable before releasing the latter to the next production step. Many of these qualitative analyses rely on the comparison of each spectrum acquired during in-line monitoring to a set of reference samples known to be homogeneous and that have the same formulation. SIMCA or PC-MBEST methods are such types of analysis and have been investigated by El-Hagrasy et al. [6]. Many other kinds of comparison computations have also been developed for certain applications [7,8].

Blanco et al. [9] compared different types of qualitative NIR spectroscopy analyses for pharmaceutical blend monitoring. In their study, 3 types of blend homogeneity indicators were investigated: (a) spectral dissimilarity through the computation of vectorial distances; (b) mean standard deviations of 3 or 6 consecutive spectra; and (c) the mean square of differences. The added advantage of the latter 2 methods resides in the fact that they are independent of the formulation. Indeed, Equations (b) and (c) only calculate consecutive variance of the spectra acquired in-line and do not require a reference set. The downside is that there is no information about the identity or quantity of the product being mixed that a qualitative dissimilarity method would provide.

In each type of analysis, one fact remains constant: the need for pre-process data prior to any analysis. This fact was highlighted by El-Hagrasy et al. [10] in Part 1 of their study. The need comes from phenomena typically occurring with reflectance spectroscopy in PAT NIR applications. Reflectance spectroscopy uses wavelength intensity reflected by the sample to provide an absorbance spectrum. The intensity reflected depends on the intensity emitted, the absorbance of the sample and the light-scattering effect of the sample, which is mostly due to particle size and could be the subject of a modeling study [11]. However, in most NIR analyses, it is an unwanted effect and must be removed to highlight chemical information for subsequent analysis. It requires a baseline correction pre-process, and many algorithms can be tried to perform it: first and second derivative (Savitzky-Golay (SG) [12] or GAP [13]), multivariate scatter correction, standard normal variate or orthogonal signal correction.

This study, conducted by the Université de Sherbrooke within the framework of collaboration with the Technology Services of Wyeth Pharmaceuticals, was aimed at evaluating methods for efficient blend homogeneity determination by both qualitative and quantitative chemometrics. The methods were applied on a real pharmaceutical formulation made of 16 ingredients. The pharmaceutical manufacturing industry produces such complex formulations and faces similar challenges. The methods developed need to be rugged for the production floor but flexible enough for the ever-moving state of formulations. The study presented here compares different kinds of qualitative blend homogeneity analyses in small-scale lab production with a nominal 1-ft^3 (about 30-L) size V-blender.

2 Experimental set-up

We investigated a Wyeth proprietary vitamin formulation containing 16 different ingredients; the exact formulation is proprietary information, but a quick overview of the formulation investigated can be seen in Table 1. Notice that the first 8 ingredients comprise over 80 wt% of the total formulation. Consequently, several ingredients have considerably low concentrations (under 1 wt%) and may not be visible with the sensitivity provided by the NIR system.

The NIR system, a MEMS-based spectrometer, gives spectral coverage of 1,350 to 1,800 nm and an acquisition time of less than 100 ms. The measuring probe provides 40-mm spot size, and penetration of the sample is estimated at about 1 mm. The instrument is installed directly on the V-blender, and a sapphire window is mounted on the cover of the V-blender at the spot where the NIR probe makes the NIR reflectance measurement. The system communicates with a laptop computer through a Wi-Fi wireless connection and can be trigger-operated to automatically acquire a NIR spectrum at a specified point in the rotation time-space. The trigger activates NIR acquisition when the blender is upside down with powder fully covering the V-blender's cover. The powder is considered to be in a static state when powder analysis is performed. Each rotation triggers an acquisition, and each acquisition generates 4 NIR spectra averaged in a single one for a total acquisition time of 400 ms, excluding software integration. This averaging helps in increasing the signal-to-noise ratio.

3 Methodology

To best model a typical production batch, the laboratory 1-ft^3 scale batches in this study underwent the same

manufacturing procedure as production batches except that no side-mixing was done. In other words, entire vitamin formula ingredients were added to the V-blender separately, creating a maximum state of segregation before any mixing.

The experimental runs were performed according to the following methodology: A first batch with a mixing time of 15 min was produced and monitored with NIR technology. A preliminary blending curve was established with simple qualitative analysis. It served to set the 2 other blending times at 69 s and 23 s, respectively. The procedure was undertaken to ensure 3 batches with a different homogeneity state. Since the 1-ft³ V-blender rotates at a speed of ± 26 rpm, the number of rotations for each batch was 380, 30 and 10, respectively. The same number of spectra can be expected for each run. In the following pages, the 3 batches are referred to and reported by their run number:

Run #1 = 15-min blending time = 380 rotations;

Run #2 = 69-s blending time = 30 rotations;

Run #3 = 23-s blending time = 10 rotations.

After the completion of each batch, a typical sample thief (see Figure 1) was disposed to acquire 10 samples at different locations of the V-blender's volume. It is known that a thief sampling method can induce a bias in the results, but this set-up was selected as a reference because it is still generally accepted in production for blend homogeneity determination by most pharmaceutical manufacturers. The 10 sampling locations are shown in Figure 2. Each batch was discarded after sampling because it was assumed that the samples might remove considerable amounts of low-concentration APIs in non-homogenized batches.

The samples were then sent to the Quality Assurance Vitamin Laboratory of Wyeth Pharmaceuticals where the 8 targeted ingredients were quantitated by standard lab methods such as high-performance liquid chromatography and titration.

The thief sample locations can be seen in Figure 2. They are numbered from 1A to 5B. This figure illustrates the lab results for Vitamin C. The trend shown in Figure 2 is typical, and it was observed for each quantified component. For each batch and for each component, mass balance closure was 100±5%. Consequently, it can be concluded that the samples taken were representative of the whole batch.

Figure 1: Sample thief deployed in this study

Reference homogeneity analysis was undertaken with the RSDs of concentrations for each component quantified and for each batch. The RSDs were then plotted versus the number of rotations to draw the blending curve: see Figure 2. Regression can be used to approximate the actual time required to achieve a homogeneous state where every component is below 5% RSD; that is the homogeneity criteria recommended by BUWG. However, the type of regression for optimal identification of the point where blending RSD goes below 5% is unknown. With exponential regression, blending time estimated from lab analysis was 126 rotations, or 4 min 50 s, but the blending curve may also be simply linear between run #2 and run #3 and then stable until the RSD values of run #1. In that case, the necessary blending time to reach an acceptable mixing time would be just below 50 rotations.

Table 1: Quantified components and their gravimetric concentration value (8 out of 16 ingredients)

Compound	Range (wt%)
Vitamin C	40-50 wt%
Vitamin E	1-5 wt%
Calcium (Mineral)	1-5 wt%
Vitamin B6	1-5 wt%
Vitamin B3	5-10 wt%
Vitamin B2	1-5 wt%
Vitamin B1	1-5 wt%
Microcrystalline Cellulose (MCC)	10-15 wt%
Total	± 80wt%

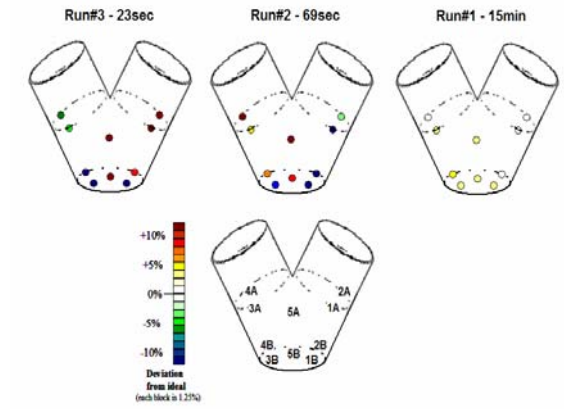

Figure 2: Blend analysis of Vitamin C

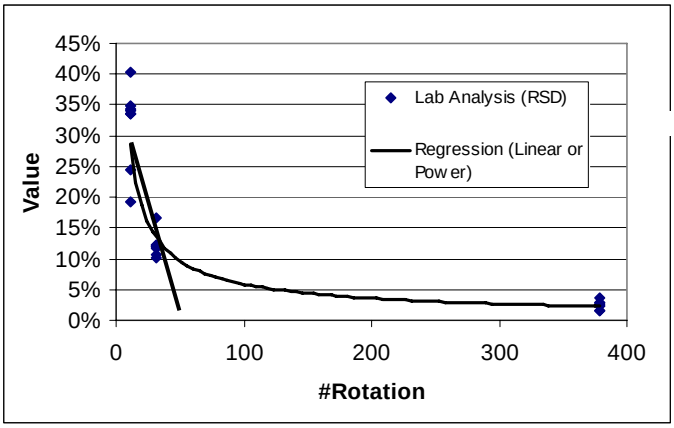

Figure 3: Blending curve from laboratory analysis

4 NIR qualitative results and discussion

Two types of qualitative analyses were investigated in this study. The 2 selected qualitative methods have different advantages and disadvantages, but are similar in a way that no value of weight concentration of any component is ever predicted. The qualitative methods investigated are: spectral variance analysis and distance analysis with Hotelling T^2.

4.1 Spectral variance analysis

Spectral variance analysis has the advantage of being independent of the product formulation, meaning that it can be applied quickly to any product formulation. It eliminates the need for a reference to determine blend homogeneity and thus removes a critical step that may induce significant error with other types of analyses. However, it has the disadvantage of only evaluating the time beyond which the mixture composition does not change. There is no information about the components of the formula or their concentrations. The minimum variance obtained when the process has reached stability should correspond to the repeatability error of the instrument if the formula is perfectly mixed. Many algorithms calculate spectral variance. Some of these algorithms were used in the course of this study, and the results of their application have been compared between each other to investigate which one could most accurately identify the mixing endpoint in accordance with the laboratory analysis.

The following formulas monitor the variance of 2 or 3 consecutives spectra. Equation 1 calculates spectral variance as defined by Blanco et al. [9]. Equation 2 is a slightly modified version of Equation 1, with the exception that the value of Equation 2 will always be higher than that of Equation 1. Equation 3 monitors the standard deviation of 3 consecutive spectra at every wavelength and then computes the average. The results of these 3 equations applied to the analysis of run #1 are shown in Figures 3 to 5.

$$\sigma^2 = \frac{\sum_{i=1}^{N}\left(A_i^{t_i} - A_i^{t_{i-1}}\right)^2}{N}$$

Equation 1: Formula for Variance

$$\delta^2 = \frac{\sum_{i=1}^{N}\left|A_i^{t_1\,2} - A_i^{t_0\,2}\right|}{N}$$

Equation 2: Modified variance formula called Delta2

$$MBSD(3) = \frac{\sum_{i=1}^{N}\sigma_3}{N}$$

Equation 3: Formula for moving block standard deviation

Where:
N = the number of scanned wavelengths
A_i: the absorbance value at wavelength "i"
t: time or iteration of mixing
σ_3: standard deviation of 3 consecutive absorbance values.

To remove the light-scattering effect during the 15-min blending time, pre-treatment was applied to the spectra prior to analysis. A SG first-derivative pre-treatment with 15 points was chosen to remove the light-scattering effect and highlight chemical differences. Note that to maintain the advantage of spectral variance analysis, data pre-treatment, such as unit scaling, normalization and mean centering, could not be performed as they require knowledge of the mean and standard deviation of the whole data set prior to analysis. Note also that in Figures 3 to 5, the values are normalized between 0 and 100 to facilitate comparison between the different types of analysis.

Spectral variance analysis with Equations 1 and 3 seems to underestimate the blend time required to achieve homogeneity when compared to the laboratory results. Indeed, they predict that stability can be reached after only 30 rotations, which may be due to a lack of sensitivity of the formula. Sensitivity may be improved by Equation 2. In this variance calculation formula, the absorbance values are squared before being subtracted. It results in increased sensitivity of the formula to any spectral differences. The formula has been developed during this study and is not found in the open literature. Figure 3 using Equation 2 shows a necessary blending time of at least 50 rotations, but also strong variation at 100 rotations and higher; the latter suggests over-blending.

The distance from a fixed reference can also be an alternative to Equations 1 to 3. In this kind of spectral variance, Hotelling T^2 distance is computed with the scores of principal component analysis (PCA). The reference selected here is simply the scan of 99% 'blank' and would be the same for the analysis of any formulation. The stability of distance is the criterion investigated not the value of distance itself. Therefore, statistical process control rules must determine the stability of the signal to thus determine the mixing endpoint. The result in Figure 6 reveals a similar trend than other spectral variance analyses, that is, a steep increase for the first 50 rotations, followed by a slow decrease of variance for rotations 51 to about 125. This reveals 2 distinct mixing steps, a first fast one followed by a second slower one. See the next section for more information about the computation of Hotelling T^2 distance.

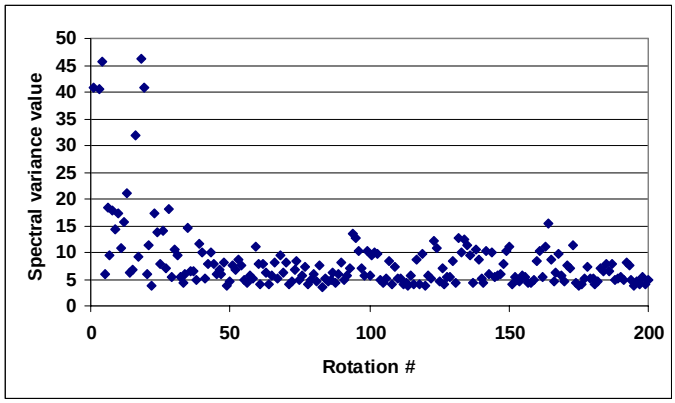

Figure 4: Spectral variance analysis using Equation 2

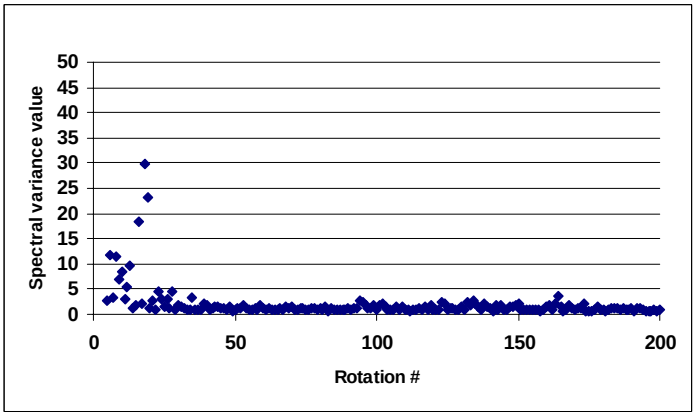

Figure 5: Spectral variance analysis using Equation 1

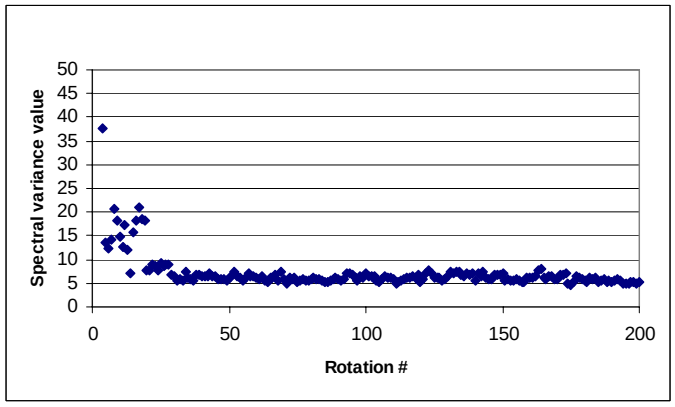

Figure 6: Spectral variance analysis using Equation 3

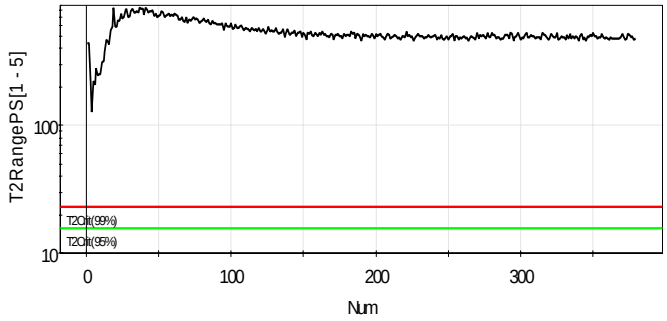

Figure 7: Distance analysis used as a spectral variance method with $T^2_{critical}$(95%, Green Line) = 15.62 and $T^2_{critical}$(99%, Red line) = 23.13

4.2 Distance analysis with Hotelling T^2

The idea behind distance analysis is to compare every newly-acquired spectrum during a batch with a reference group of spectra. This reference group must be carefully chosen to represent an homogeneous state of the studied formula. A dimensionless distance value is then computed between the newly-acquired spectrum and the reference group. When that distance is less than a pre-defined limit value, the blend is homogeneous and the mixing process can be stopped.

Many types of mathematical tools can compute the distance between the reference group and the newly-acquired spectrum. For instance, the Mahalanobis distance is such a mathematical tool; it was first introduced in 1936 [14]. This tool assigns a different weight to every variable; in our case, every wavelength. The reference group is first described in vector space, and variables with the most significant variations are assigned a weight of greater magnitude in the analysis. The Hotelling T^2 distance investigated in this study is similar to Mahalanobis distance but does not use a co-variance matrix. It performs best when applied to scores of the variables after PCA. The distance value is computed by Equations 4 to 7.

This analysis provides additional information when compared to spectral variance since it can also assess, to some extent, whether the mixture contains the measured ingredients at the target concentrations. However, the challenge resides in reference selection. The reference should be able to model the noise present in the production environment and the variation from batch to batch throughout the year. The need for a homogeneous reference is a disadvantage of this analysis when compared to the previous qualitative analysis of spectral variance presented earlier since a new reference will be required for every product formulation targeted. Furthermore, the reference will need to be updated frequently if a raw material supplier changes.

$$\bar{t}_a = \frac{\sum_{1}^{N} t_{n,a}}{N}$$

Equation 4: Average of reference group scores

$$s_{t,a}^2 = \frac{\sum_{1}^{N} \left(t_{n,a} - \bar{t}_a\right)^2}{N-1}$$

Equation 5: Standard deviation of reference group scores

$$T_i^2 = \sum_{a=1}^{A} \frac{t_{i,a}^2 - \bar{t}_a}{s_{t,a}^2}$$

Equation 6: Hotelling T² distance computation

$$T_{critical}^2 = \frac{A(N^2-1)}{N(N-A)} \times F_{critical}$$

Equation 7: Critical value of distance computation

With

- T_i^2: Hotelling T² value of spectrum at time or increment "i"
- i: time or increment of data acquisition
- t_a: score value of principal component "a"
- $\bar{t}_a$: mean value of principal component "a" for all $t_{a,n}$ of the reference group
- N: number of observations in the training set (reference group)
- A: total number of principal components (PCs)
- $F_{critical}$: F distribution critical value with A(v_1) and N-A(v_2) degrees of freedom at "p"
- $T^2_{critical}$: limit value of the confidence interval.

In our study, the reference selected was scans of the sample thieves from run #1. The reference group was selected because run #1 was proven homogeneous from laboratory reference analyses. Figure 7 illustrates distance analysis with the reference group selected. In this figure, the distance value of each NIR spectrum (1 per turn) is plotted versus its rotation number. The red and green lines are the confidence intervals determined by MVDA software. A spectrum with a distance value higher than 17.01 (red line) should have a 99% chance of being different from the reference. In the same way, a spectrum with a value higher than 12.21 (green line) should have a 95% chance of being different from the reference. Notice that the Y axis is in logarithmic scale.

The data used to obtain Figure 7 have been pre-processed with a SG first-derivative with 15 points. Distance analysis predicts a minimum blending time of 58 rotations to achieve distance values consistently below the 95% confidence value; this is similar to the result obtained by spectral variance analysis with Equation 2. However, as in analysis of variance, it can be seen that the mix seems to shift at the 100th rotation and to stabilize afterwards, suggesting an over-blend or a bias of the reference. Note that this over-blend or bias is mild since the value just barely exits the 99% confidence interval.

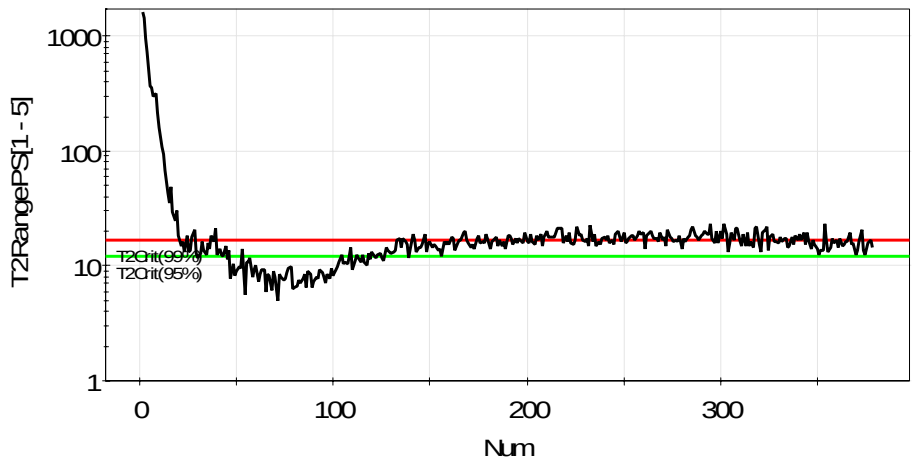

Figure 8: Distance analysis with Hotelling T² distance on PCA scores of run #1 with $T^2_{critical}$(95%, Green line) = 12.21 and $T^2_{critical}$(99%, Red line) = 17.01

5 Conclusions

Two types of blend homogeneity analyses are presented in this study, each having its own strengths and weaknesses. However, in each case, a tremendous amount of information is acquired when compared to current sample thief analysis and, consequently, the level of confidence for the results is greatly improved. In addition, this information can be obtained in real-time once appropriate chemometric models are developed. This represents a huge time saving when compared to laboratory analysis that can take hours to days to perform. It is expected that any one of the presented NIR analyses may be beneficial on many aspects of pharmaceutical blending:

- In real-time quality monitoring of current manufacturing batches.
- In improving process efficiency and performance by selecting adequate process parameters and blending time.
- In quality by design and design space definition initiatives when developing and/or validating blending processes of new formulations.

When comparing the results of the 2 types of analyses with the referenced laboratory methods, it seems that some spectral variance analyses may underestimate the blend time required to achieve optimum homogeneity, Equations 1 and 3 in this study. With Equation 2, the mixing endpoint could be considered at around 50 rotations. This would correspond well with the lab data if a linear regression of the RSD points is assumed.

However, when using distance analysis for spectral variance, stability seems to be achieved at about 120 rotations, which would correspond more closely to the lab data blending curve if an exponential curve is assumed. It is concluded that Equation 2 outperforms Equations 1 and 3, and that it can clearly detect macro-mixing. This indicator can be considered sufficient for simple formulas where micro-mixing is less critical. Distance analysis for spectral variance could be more efficient in detecting micro-mixing. It is also noted that each spectral variance analysis should be investigated in conjunction with a statistical process control method (SPC) to objectively determine the mixing endpoint.

In the distance from a homogeneous reference analysis, it seems that the noise or baseline effect present in the mixing process is not replicated during reference scans of thief samples, thus inducing a small bias. This again highlights the fact that reference selection is a critical step for such analyses and should be done with spectra acquired during the target process over more than 1 batch.

The type of NIR analysis should be selected by the user, depending on what type of information is required and the time available to construct the model. Table 2 presents the pros and cons of each type of NIR analysis investigated. The next step of this investigation is to develop a quantitative model and to apply these methods to commercial and full-scale V-blender lots.

Table 2: Comparative study of NIR blend analysis of this study

Qualitative study	
Variance study	**Distance from homogeneous reference study**
Pros	**Pros**
• Quick implementation for every formula • Diminishes errors due to reference selection • Spectral pre-process less critical	• Precise information about blending state • Can detect wrong components or wrong concentrations • Moderate modeling effort required
Cons	**Cons**
• No information about component concentrations • No information about types of components • Can slightly underestimate blending time • Should be used with statistical process control rules	• No information about component concentrations • Spectral pre-process is critical • Reference is critical

Acknowledgments

The authors are indebted to Wyeth Pharmaceuticals and the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding related to this project. Special thanks are due to Antoine Cournoyer, Éliane Caron and Carlo Benedetti for their technical contribution, along with Ovid Da Silva for reviewing this manuscript.

References:

[1] MUZZIO F., ROBINSON P., Sampling Practices in Powder Blending, *Int. J. Pharm.*, Vol.155, 1997, pp. 153-178.

[2] BERMAN J., PLANCHARD J.A., Blend Uniformity and Unit Dose Sampling, *Drug Dev. Ind. Pharm.*, Vol.21, No.11, 1995, pp. 1257-1283.

[3] PASTOR R., GOMEZ R., BELLAMY M., A Novel Sample Thief Design to Avoid Biased Data, *Pharm. Technol. Yearbook*, Vol.1, 1999, pp. 1-16.

[4] North American Chapter of the International Chemometrics Society (NAmICS): http://www.namics.nysaes.cornell.edu/chem_info.html

[5] Standard Practice for Near Infrared Qualitative Analysis, *ASTM International*, Designation: E 1790-04, Part of Vol. 03.06, 2007.

[6] EL-HAGRASY A.S., DELGADO-LOPEZ M., DRENNEN 3rd J.K., *A Process Analytical Technology Approach to Near-Infrared Process Control of Pharmaceutical Powder Blending. Part II: Qualitative Near-Infrared Models for Prediction of Blend Homogeneity*, J. Pharm. Sci., Vol.95, No.2, 2006, pp. 407-421.

[7] WOLD S., ESBENSEN K., GELADI P., *Principal Component Analysis*, Chemom. Intell. Lab. Syst., Vol.2, 1987, pp. 37-52.

[8] DE HASETH J.A., AZARRAGA L.V., *Interferogram-based Infrared Search System,* Anal. Chem., Vol.53, No.14, 1981, pp. 2292-2296.

[9] BLANCO M., GOZÁLEZ BAÑÓ R., BERTRAN E., *Monitoring Powder Blending in Pharmaceutical Processes by Use of Near Infrared Spectroscopy*, Talanta, Vol.56, 2002, pp. 203-212.

[10] EL-HAGRASY A.S., D'AMICO F., DRENNEN 3rd J.K., *A Process Analytical Technology Approach to Near-Infrared Process Control of Pharmaceutical Powder Blending. Part 1: D-Optimal Design for Characterization of Powder Mixing and Preliminary Spectral Data Evaluation*, J. Pharm. Sci., Vol.95, No.2, 2006, pp. 392-406.

[11] OTSUKA M., *Comparative Particle Size Determination of Phenacetin Bulk Powder by Using Kubelka–Munk Theory and Principal Component Regression Analysis Based on Near-infrared Spectroscopy*, Powder Technol., Vol.141, 2004, pp. 244-250.

[12] SAVITZKY A., GOLAY M.J.E., *Smoothing and Differentiation of Data by Simplified Least Squares Procedures,* Anal Chem., Vol.36, No.8, 1964, pp. 1627-1639.

[13] ESBENSEN K., *Multivariate Data Analysis − In Practice: An Introduction to Multivariate Data Analysis and Experimental Design*, 5th edition, CAMO, Norway, 2004, 598 pages.

[14] MAHALANOBIS P.C., *Proc. Natl. Inst. Sci. India*, Vol.2, 1936, p. 49.