Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Pharmaceutical and Biomedical Analysis 41 (2006) 26–35
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Simple assessment of homogeneity in pharmaceutical mixing processes using a near-infrared reflectance probe and control charts

E.T.S. Skibsted[a,b], H.F.M. Boelens[a], J.A. Westerhuis[a,*], D.T. Witte[c], A.K. Smilde[a]

[a] *Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*
[b] *Pharmaceutical Site Maaloev, Novo Nordisk A/S, Novo Nordisk Park, 2760 Maaloev, Denmark*
[c] *N.V. Organon, Molenstraat 10, P.O. Box 20, 5340 BH Oss, The Netherlands*

Received 16 June 2005; received in revised form 6 October 2005; accepted 6 October 2005
Available online 9 November 2005

Abstract

Determination of homogeneous mixing of the active pharmaceutical ingredient (API) is an important in-process control within the manufacturing of solid dosage forms. In this paper two new near-infrared (NIR) based methods were presented; a qualitative and a quantitative method. Both methods are based on the calculation of net analyte signal (NAS) models which were very easy to develop, specific with respect to the API and required no additional reference analysis. Using a well-mixed batch as a 'golden standard' batch, control charts were developed and used for monitoring the homogeneity of other batches with NIR. The methods were fast, easy to use, non-destructive and provided statistical tests of homogeneity. A mixing study was characterized with the two methods and the methods were validated by comparison with traditional HPLC analysis.

Keywords: Net analyte signal; Process analytical technology; Powder homogeneity; Near-infrared spectroscopy; Process control charts

1. Introduction

An important unit operation in the production of a pharmaceutical solid dosage form is the mixing of the powder constituents. One of the primary objectives of the mixing process is to obtain a homogeneous mixing of the active pharmaceutical ingredient (API) in the formulation. The homogeneity is normally determined by removing a small number of powder samples from the powder mixer with a probe thief and sending the samples to a quality control laboratory distant from the process line for, e.g. high pressure liquid chromatography (HPLC) analysis. The API content in the samples and the relative standard deviation (R.S.D.) of the API content between the samples is then obtained and compared to specified values in order to decide wherever the batch is sufficiently homogeneous [1]. This procedure is slow and is not suited for process control of the mixing process to obtain the highest possible homogeneity. Also a limited number of samples are used which makes a full characterization of homogeneity difficult.

Therefore, near-infrared (NIR) based spectroscopic methods which are non-destructive, fast and require little resources have been proposed [2–10]. NIR has been applied in many different manners. Samples were removed from the system and measured with NIR (off-line) [2,4]. This method faces some of the classical powder sampling problems where large variance is introduced when inserting a sample thief into the powder bed and removing a sample. In other papers, diffuse reflectance probes have been inserted directly into the powder bed at a fixed position (in-line) [5,6] or the mixer was fitted with quartz windows through which monitoring could be applied from outside the mixer (semi[1] non-invasive) [7]. These types of methods only monitor the powder

* Corresponding author.
E-mail address: westerhuis@science.uva.nl (J.A. Westerhuis).

[1] The application was not truly *in-line* because the mixer had to be stopped to perform monitoring with a diffuse reflectance probe that was moved around from window to window.

0731-7085/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2005.10.009

bed at one or a few points and only provide information about the homogeneity in these particular positions assuming that the point or points are representative for the entire powder bed. In cases with 'dead spots' or de-mixing phenomena this assumption would not hold.

Many different data analysis methods have been applied to the spectral data in order to derive the optimal mixing time. The span of methodologies are from simple qualitative methods looking for absence of changes in the spectra, i.e. moving block [5], dissimilarity indexing [2], principal component analysis (PCA) [4], soft independent model class analogy (SIMCA) [7] and bootstrapping techniques (BEST) to quantitative calibration models using partial least squares (PLS). The qualitative approaches evaluate only spectral homogeneity, i.e. variation between spectra in time or position but are not specific for the analyte. The PLS method on the other hand, directly evaluates the concentration homogeneity of the analyte of interest. Some of the qualitative methods compare the measured spectra with a set of "target" spectra of a homogeneous mixture.

In the present paper we introduce a new method for monitoring homogeneity of powder mixtures by testing the variation and the level of the API content over the mixer. The API content is based on the net analyte signal (NAS) value of NIR spectra. The NIR spectra are recorded directly in the mixer using a fibre based handheld diffuse reflectance probe. Thus, by just sticking the probe in the mixer a number of times, a measure of homogeneity is obtained that is statistically validated.

A qualitative approach of the method is suited for early development studies when only a few batches are available, and further on in the development stage, calibration can be used to make the method quantitative. In both methods a 'golden batch' is chosen and two control charts are developed based on measurements of the 'golden batch'. In the two control charts the variance and the mean are observed. With the proposed method, a mixing process is examined. In order to validate the proposed NIR methods control charts are also developed using values obtained from traditional HPLC analysis of powder samples. Results from four independent batches are then plotted in the control charts and the results are compared to NIR. Variance components in the measured signal are described and finally we show how many spectra need to be recorded in order to not make false conclusions about homogeneity.

2. Theory

In this theory section we will first discuss the mixing of pharmaceutical powders and the idea of a random and homogeneous mixture. Then the different sources of variation in analytical measurements of homogeneity are discussed. The net analyte signal approach used to relate a multivariate NIR measurement to the API content is discussed shortly. Afterwards we first introduce the qualitative approach for monitoring homogeneity. Finally we will also show how a quantitative model can be developed in which actual concentrations of the API are monitored.

2.1. Mixing and determination of a random mixture

> Mixing is the treatment of two or more components in such a way, that the individual particles of the different components in the mixture are evenly distributed and lie adjacent to each other within the highest possible probability [11].

A perfect mixture between two or more components would be one in which each sample contains exactly the proper amount of each of the components. If, e.g. two components were mixed (A and B) in a 1:1 relationship a perfect mixture would be when every second particle would be A. Such a perfect mixture is not achievable but a random mixture is [11] and the aim of a pharmaceutical mixing process is to achieve a random mixture which then would stand for a homogeneous mixture. An example of a random mixture was presented by Muzzio et al. [12]. A random mixture can be considered as a mixture where all samples removed from the mixture would be normally distributed with same mean (μ) and standard deviation (σ) regardless from where in the mixture the sample was removed.

2.2. Variance component model

In order to understand the total variance the following variance component model is proposed for a NIR based method:

$$\sigma^2_{\text{total}} = \sigma^2_{\text{analysis}} + \sigma^2_{\text{sampling}} + \sigma^2_{\text{random mixture}} + \sigma^2_{\text{heterogeneity}} \quad (1)$$

The variance due to the analysis, $\sigma^2_{\text{analysis}}$, e.g. repeatability, is low for modern Fourier-transform FT-NIR instruments with cooled detectors.

The sampling procedure is recognized as a major contributor of error in the final result when samples are withdrawn with a sample thief [12]. When using a NIR probe, the sampling variance component, $\sigma^2_{\text{sampling}}$ is believed to be affected by some of the same phenomena as sample thieves, e.g. transportation of powder from higher placed layers when penetrating the mixture. Some studies recommend sample thieves with front sampling instead of side port sampling. A NIR probe can be compared to a front sampling thieve. When the NIR probe is inserted into the powder mixture at a given position, the probe can be tilted slightly prior to the measurement in order to measure an undisturbed sample. This method has been tested in our laboratory and it appears to be an efficient way to avoid sampling material which has been transported with the probe during insertion.

Another advantage of NIR is that no sample preparation is required which in return minimize sampling error. Finally, using a proper pre-processing and wavelength selection of the NIR spectra would minimize the influence of artifacts in the spectra from varying particle sizes or other physical parameters.

The variance component due to the random mixture, $\sigma^2_{\text{random mixture}}$, is the inherent variance one observes in a random mixture. This is the variation in API concentration between

several measurements inside the random mixture. This variation greatly depends on the sample size [11]. An increase in sample size will show a decrease in this variance component.

The last variance component $\sigma^2_{\text{heterogeneity}}$ (the mixing component) represents the degree of heterogeneity. In a random mixture the variance due to heterogeneity will be zero. Thus in monitoring homogeneity, the actual test is whether the $\sigma^2_{\text{heterogeneity}} = 0$.

2.3. Net analyte signal (NAS)

When the analyte of interest is absorbing in the near-infrared (NIR) wavelength region and the shape of the analyte spectrum and interfering constituents spectra is different, a net analyte signal (NAS) vector, which is unique for the analyte, can be derived. The NAS vector is defined as the part of the sample spectral vector that is orthogonal to a subspace called the interferent space. The NAS vector is unique for the analyte of interest in the given mixture of interferents. The interferent space is spanned by spectral vectors of the interfering constituents in the sample matrix, i.e. all other components except for the analyte of interest. In order to construct a robust NAS vector sufficient variation should be present in the NIR measurements of interferents mixtures. This can be assured by manufacturing batches of interferents mixtures after designed experiments.

2.4. Notation

The following notation is used during the presentation of equations. Boldface capital characters denote matrices, boldface lower-case characters denote vectors and lower-case italic characters denote scalars, superscript T denotes the transposed matrix or vector and the superscript + denotes the Moore–Penrose generalised inverse of a matrix. The matrix $\mathbf{I}_J$ is the $J \times J$ identity matrix.

2.5. Calculation of the NAS value of a spectrum

Using a set of NAS vectors, a NAS regression vector can be calculated. With the NAS regression vector the net analyte signal value or simply the net analyte signal of a sample spectrum can be computed. The NAS is directly proportional to the analyte concentration [13] which makes this value a potential candidate to be used to evaluate the distribution of analyte in a set of samples.

To compute the NAS regression vector, first the interference space needs to be defined. Pure component spectra of the interferents can be used to span the interferent space. In our experience [14–16] it is better to use spectra of blank samples, i.e. mixture samples of the interfering constituents only.

First the anti-projector ($\mathbf{A_R}$) is constructed using a set of blank spectra ($\mathbf{R}_{-k}$)

$$\mathbf{A_R} = (\mathbf{I}_J - \mathbf{R}_{-k}\mathbf{R}^+_{-k}) \tag{2}$$

$\mathbf{R}_{-k}$ is a $J \times I_b$ matrix with I_b blank spectra measured at J channels. The subscript $-k$ indicates that the spectra does not contain NIR response from the analyte k. $\mathbf{R}^+_{-k}$ is the pseudo-inverse matrix of the blank samples, an $I_b \times J$ matrix.

Next the NAS regression vector is computed with a set of model spectra ($\mathbf{R}_{\text{mod}}$). Pure analyte spectra or spectra of samples with the interfering constituents and the analyte can be used. The latter is preferred because the spectral response of the analyte in presence of other constituents might be different than the spectrum of the pure analyte alone. With the anti-projection matrix and the model spectra the orthogonal part to the interferent space can now be computed

$$\mathbf{B} = \mathbf{A_R} \cdot \mathbf{R}_{\text{mod}} \tag{3}$$

The $\mathbf{B}$ is a $J \times I_{\text{mod}}$ matrix with I_{mod} vectors which all are orthogonal to the interference space and point more or less in the same direction. The average vector $\mathbf{b}$ is used to define the unique NAS direction and is called the NAS regression vector

$$\mathbf{b} = \frac{\sum_{i=1}^{I_{\text{mod}}} \mathbf{B}_i}{I_{\text{mod}}} \tag{4}$$

With the NAS regression vector, the NAS value of a sample spectrum ($\mathbf{r}$, a $J \times 1$ vector) can be computed simply as the score of the projected spectral vector onto the NAS regression vector

$$\text{NAS} = \mathbf{r}^{\text{T}}\mathbf{b} \tag{5}$$

2.6. Development of a qualitative model based on NAS values

In the early stages of a drug development program different formulations are assessed. Development of quantitative NIR models using several calibration batches for each formulation would require many resources. A qualitative method that requires a minimum of resources is therefore beneficial. The proposed method requires two batches: (1) a placebo batch and (2) a batch with API that is homogeneous mixed, i.e. a 'golden batch' where the $\sigma^2_{\text{heterogeneity}} = 0$. In the following a stepwise procedure is described how to develop the qualitative method and control charts. Pre-processing method and wavelength selection of the NIR spectra are chosen based on the SE indicator [14].

- *Step 1*: A placebo batch is measured with the NIR probe. The spectra are used to calculate the anti-projector (Eq. (2)).
- *Step 2*: A batch with 100% of the concentration in the final drug product is prepared and mixed until homogeneity and then measured extensively with the NIR probe. This batch is named 'golden batch' in the text.
- *Step 3*: A few spectra from the 'golden batch' are used as model spectra ($\mathbf{R}_{\text{mod}}$) and the NAS regression vector is calculated (Eqs. (3) and (4)).
- *Step 4*: The NAS values of the remaining spectra from the 'golden batch' are calculated with Eq. (5).
- *Step 5*: The mean NAS value $\overline{\text{NAS}}_{\text{gb}}$ and variance $s^2_{\text{total NAS,gb}}$ of the NAS values are determined.
- *Step 6*: Finally two control charts are constructed; a variance and mean chart.

2.7. Development of control charts for qualitative NAS model

In this section are the assumptions and equations presented for limits in the control charts. Note that since it is a qualitative approach, it is not possible to test whether a new batch is on target with respect to concentration of API; it is only possible to test whether the batch is equal to a 'golden batch'. Two charts are developed; a variance chart and a mean chart.

2.7.1. Variance chart

In the variance chart the variance for a new batch is compared statistically to the variance of the 'golden batch'. Comparison is done using the ratio between the variance of the new batch and the variance of the 'golden batch'. The ratio of variances follows an F-statistic:

$$\frac{s^2_{\text{total NAS,new}}}{s^2_{\text{total NAS,gb}}} \propto F_{(\alpha, N_{\text{new}}-1, N_{\text{gb}}-1)}$$

where N_{gb} is the number of spectra used to calculate the 'golden batch' variance.

A one-sided F-test is used because it is assumed that in monitoring homogeneity, the actual test is whether $\sigma^2_{\text{heterogeneity}} = 0$. The assumption is that the 'golden batch' has zero heterogeneity and if the new observation would deviate from the 'golden batch' then $\sigma^2_{\text{heterogeneity,new}} > 0$. The upper control limit is then calculated using the following equation:

$$\text{upper control limit} = s^2_{\text{total NAS,gb}} \times F_{(\alpha, N_{\text{new}}-1, N_{\text{gb}}-1)} \quad (6)$$

where the critical value can be found in an F table.

Although the variance of a new batch is expected to be equal to or higher than that of the 'golden batch', in some cases, a lower variance could be found. A lower variance does not signal for heterogeneity, but for model incompetence or probe fouling. Therefore, do we also calculate a lower limit which is called the warning limit:

$$\text{warning limit} = s^2_{\text{total NAS,gb}} \times F_{(1-\alpha, N_{\text{new}}-1, N_{\text{gb}}-1)} \quad (7)$$

2.7.2. Mean chart

If the two variance of the new batch is equal to the 'golden batch', the mean of the new batch can be compared to the mean of the 'golden batch'. This is done by computing the t-value for the new batch and then plot the t-value in the mean chart and compare to the critical value obtained from a table of critical values of t:

$$t\text{-value} = \frac{\overline{\text{NAS}}_{\text{new}} - \overline{\text{NAS}}_{\text{gb}}}{s\sqrt{\frac{1}{N_{\text{new}}} + \frac{1}{N_{\text{gb}}}}} \quad (8)$$

where s is calculated from:

$$s^2 = \frac{(N_{\text{new}} - 1)s^2_{\text{total NAS,new}} + (N_{\text{gb}} - 1)s^2_{\text{total NAS,gb}}}{N_{\text{new}} + N_{\text{gb}} - 2}$$

The critical t-values can be found for a α confidence level and $(N_{\text{new}} + N_{\text{gb}} - 2)$ degrees of freedom. A two-sided test will be used for the t-test to detect whether the operator put too much or too little API in the mixer (compared to the 'golden batch'). The control limits for the mean chart are then:

$$\text{control limits} = 0 \pm t\left(\frac{\alpha}{2}, N_{\text{new}} + N_{\text{gb}} - 2\right) \quad (9)$$

The critical step in this method is in Step 2. It is stated that a batch is mixed to homogeneity ('golden batch') and then measured extensively with NIR. One approach to identify the point of homogeneity could be to make a temporary NAS model. First the anti-projector is calculated from the placebo batch spectra. Carry out a few minutes of mixing of the calibration batch and measure a few NIR spectra. These spectra are used to calculate a temporary NAS regression vector. Then continue mixing of the calibration batch. At different time points the mixer is stopped and high number (e.g. 50) NIR spectra are recorded. The NAS values of the spectra are calculated with the temporary NAS regression vector. By simply plotting the variance over time the point of homogeneity can be identified as time point with lowest variance.

2.8. Development of a quantitative model based on NAS values

When a final formulation has been chosen during the R&D development phase more calibration work can be added and a quantitative model between the NIR spectra and the API concentration developed. A major benefit of a quantitative model is that numbers are now expressed in concentration of the API which is comparable to the standard regulatory methodologies. Also now the mean can be compared to the target content of API instead of mean of 'golden batch'.

The development of the quantitative method is described by the following steps:

- *Step 1*: First a placebo batch is prepared and measured with the NIR probe. The spectra are used to calculate the anti-projector (Eq. (2)).
- *Step 2*: A set of calibration batches spanning a fair API concentration range are prepared, e.g. 70, 85, 100, 115 and 130% of the concentration in the final drug product. The batches are mixed until they are homogeneous and measured extensively with the NIR probe. The spectra are split into a calibration set and a test set.
- *Step 3*: The calibration spectra from each batch are averaged to one calibration spectrum for each calibration batch. Each calibration spectrum is assigned a reference value which is the average API concentration in the particular calibration batch. This value is known from the preparation of the calibration batch [8].
- *Step 4*: A few spectra from one of the calibration batches are used as model spectra ($\mathbf{R}_{\text{mod}}$) and the NAS regression vector is calculated (Eqs. (3) and (4)).
- *Step 5*: Now the NAS value of the average calibration spectra and some extra placebo spectra is calculated (Eq. (5)). These extra placebo spectra should not have been used to construct the anti-projector in Step 1.

- *Step 6*: The NAS values of the average calibration spectra are plotted against their reference value in a NAS value versus concentration plot. A straight line is fitted to the points in a least square sense. With the equation for the line (calibration model) the API concentration can now be calculated for a given spectrum [17].
- *Step 7*: With the calibration model the API concentration c is predicted in all test set spectra.
- *Step 8*: The variance $s^2_{\text{total c,gb}}$ is determined of all concentration predictions as the pooled variance from all test set data.

  Because all calibration batches are assumed to be homogeneous all concentration predictions from the test set can be used to calculated a pooled variance which provides more degrees of freedom:

$$s^2_{\text{total c,gb}} = s^2_{\text{pooled}} = \frac{\sum_{i=1}^{N_1}(c_{1,i} - \bar{c}_1)^2 + \sum_{i=1}^{N_2}(c_{2,i} - \bar{c}_2)^2 + \cdots}{(N_1 - 1) + (N_2 - 1) + \cdots} \quad (10)$$

  where $\bar{c}_1$ is the average concentration of N_1 predictions from calibration batch 1.
- *Step 9*: A variance and a mean chart is developed and used for monitoring of future batches.

2.9. Development of control charts for quantitative NAS model

2.9.1. Variance chart

For the quantitative NAS model, also the spread in the measurements is compared to the spread found for the 'golden batches' in the same way as was presented for the qualitative NAS model. The limits are calculated in the same way substituting variance of NAS values with the pooled variance of the concentration predictions

$$\frac{s^2_{\text{total c,new}}}{s^2_{\text{total c,gb}}} \sim F_{(\alpha, N_{\text{new}}-1, N_{\text{gb}}-1)}$$

2.9.2. Mean chart

With the quantitative model the mean of a new batch can be compared to the target concentration of the product (instead of comparing to the mean a 'golden batch' of which the exact API concentration is unknown). This implies that the mean chart now compares whether the average concentration is within target limits.

The mean chart is constructed like a Shewart chart [18] and the control limits are then

$$\text{control limits} = \text{target} \pm t_{(\alpha/2, N_{\text{gb}}-1)} \frac{s_{\text{total c,gb}}}{\sqrt{N_{\text{new}}}} \quad (11)$$

3. Experimental

3.1. NIR measuring techniques

All NIR measurements were measured with the same Bruker FT-NIR spectrometer, MPA (Multi Purpose Analyzer) [19].

Table 1
Mixture composition

Component	g	% (w/w)
API	100	10
Microcrystalline cellulose	200	20
Tablettose	677.5	67.75
Magnesium stearate	7.5	0.75
Talc	15	1.50
Sum	1000	100

An extended length fibre optic sampling probe with pistol grip was used. The probe was fitted to the MPA with a 1.5 m optical fibre. The probe head diameter is 14 mm and the optical window has a diameter of 4 mm. The length of the probe is 325 mm. A resolution of 8 cm^{-1} was used. The spectral range from 4000 to 12 794 cm^{-1} was scanned. As background, a Spectralon disc was used with 32 co-added scans per spectrum. For sample spectra 16 co-added scans per spectrum were used. The acquisition time per sample spectrum was approximately 8 s which is a practical time for a steady handheld measurement. The reflectance signal from the sample in front of the probe was detected with an InGaAs detector.

3.2. Reference analysis

Powder samples that were removed from the batches were of approximately 300 mg of size. The samples were analyzed with a HPLC method. First the samples were dissolved followed by injection into the HPLC system which used UV-detection at 275 nm, a C18 column and a mobile phase consisting of a salt buffer and acetonitril.

3.3. Composition of pharmaceutical mixture

The analyte of interest was the active pharmaceutical ingredient (API). For reasons of secrecy the chemical identify of the API cannot be revealed. All batch sizes were 1000 g and the composition is listed in Table 1. All mixing was performed in a drum mixer.

3.4. Experiments

In Table 2 is a list of the batches that were used. In the calibration batches the API was interchanged with microcrystalline cellulose and tablettose while maintaining the weight ratio between the microcrystalline cellulose and tablettose.

3.5. NIR measurements and samples analyzed with reference method

Ninety spectra were recorded in the placebo batch. From each of the calibration batches, 50 spectra were recorded. In the mixing experiment the mixer was stopped at various time points, the lid was removed and 50 spectra were recorded at each time point. In all experiments an evenly spatial placement of the probe in the powder mixture was attempted. Samples were also removed for

Table 2
Batch overview

Name	Description and sampling	Use
Placebo batch	No API is added	Calculating of the anti-projector
70, 85, 100, 115 and 130% calibration batches	Varying amount of API from 7 to 13% (w/w). Fifty NIR spectra/batch. Thirty samples for HPLC analysis/batch	Calibration and calculating control limits
Mixing batch	A batch with 10% (w/w) API. Fifty spectra/time point. Several time points measured during mixing	Demonstrate how the methods could be used to monitor a mixing process
Validation batches (four batches)	Each was with 10% (w/w) API concentration. The batch differences were mixing time and mixing order. Thirty NIR spectra/batch. Twelve samples for HPLC analysis/batch	Used to validate the NIR methods

HPLC analysis in the calibration batches and in the validation batches.

3.6. Data acquisition and analysis

All spectra were collected using the OPUS 4.2 software for the NIR instrument [20]. The spectra were converted from the OPUS format to Jcamp and imported into MatLab 6.5 [21] using in-house written routines. All data analysis was performed in MatLab using in-house developed algorithms.

4. Results and discussion

In this section first a qualitative model and control charts are developed. The mixing data are applied to the qualitative model. Then a quantitative model is developed and the mixing data assessed. The NIR models are validated by comparison with HPLC. Finally, the influence of number of spectra on the method is evaluated.

4.1. Qualitative model

4.1.1. Model and control charts

Twenty-three placebo spectra were used to develop the anti-projector $\mathbf{A_R}$ (see Eq. (2)). A calibration batch with 10% (w/w) API was prepared. After mixing, 50 spectra were recorded with the probe in the mixer. Five spectra were used as model spectra to calculate the NAS regression vector. Using the NAS regression vector the NAS values were calculated for the remaining 45 spectra. The limits for the variance and mean charts were calculated.

4.1.2. Mixing data

A batch with a composition of 10% (w/w) API was prepared. During the mixing of the batch the mixer was stopped at time points 16, 20, 25, 30, 35 and 40 min. At each time point 50 spectra were recorded in the mixer. The variances and means were calculated and plotted in the control charts (Fig. 1).

The results showed that at time points 16, 20, 25 and 30 min the variance was comparable to the 'golden batch'. At time points 35 and 40 the variance was higher then in the 'golden batch' showing that de-mixing took place. The means from time points 16, 20, 25 and 30 were compared to the 'golden batch' in the mean chart. They were all within the limits and therefore not different from the mean of the 'golden batch'.

The means of time points 35 and 40 min are indicated with (×) since the equal variance requirement for the *t*-test was not fulfilled anymore and thus the results cannot be trusted.

4.2. Quantitative model

4.2.1. Model and control charts

To develop the quantitative model another four calibration batches were prepared, i.e. with 7, 8.5, 11.5 and 13% (w/w) API. In all five calibration batches (including the 10% (w/w) batch used in the qualitative model) 50 spectra were recorded when they were homogeneously mixed. The spectra from each batch were split into 20 spectra that were used to construct an average calibration spectrum for each calibration batch and 30 spectra called test set spectra. Also after recording of the spectra, 30 powder samples (each approximately 300 mg) were removed from every calibration batch and subject to HPLC analysis. Now the anti-projector was developed. Five spectra from the 130% batch were used to calculate the NAS regression vector. The average calibration spectra from each of the five calibration batches and the NAS regression vector are depicted in Fig. 2. The methodology of using average spectra of powder samples and assigning the nominal batch concentration was first demonstrated by Berntsson et al. [8]. In this study it showed to be a

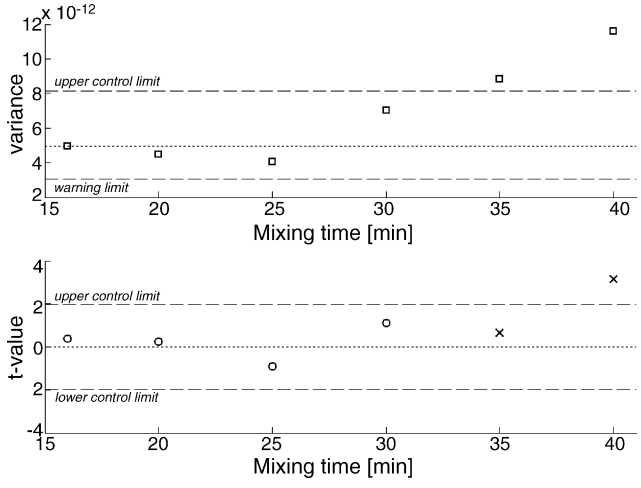

Fig. 1. Mixing experiment data plotted in variance and mean chart using qualitative method. The mean values symbolized with (×) indicate that the assumption for the *t*-test is not fulfilled because the variance is different from the 'golden batch'.

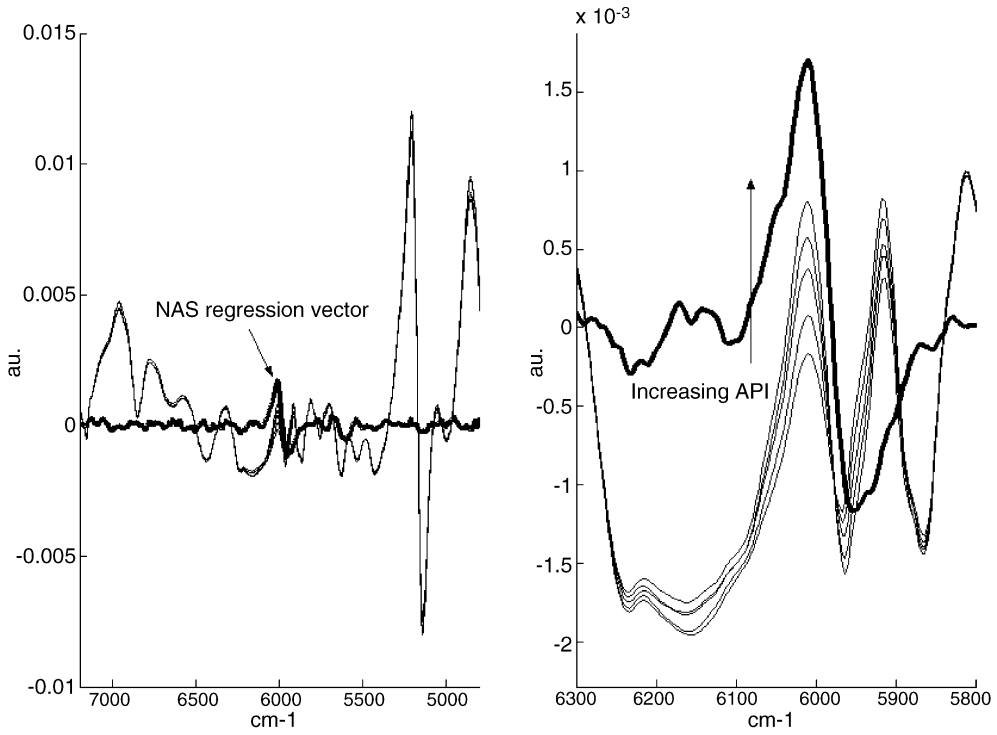

Fig. 2. The five average calibration spectra and the NAS regression vector (bolded line) (left). The API had a characteristic peak near 6000 cm^{-1}. This feature was easily recognized in the calibration spectra and the NAS regression vector (right).

very easy, fast and precise method for making the quantitative calibration models.

The NAS values of the average calibration spectra and 20 extra placebo spectra were plotted against their reference values (symbolized with circles in Fig. 3) and a straight line was fitted to the points in a least square sense (line in Fig. 3). The correlation coefficient was 0.9998. The NAS values of test set spectra were also plotted in order to demonstrate the variance of the API within each batch (symbolized with dots in Fig. 3).

Using the calibration model concentration predictions were obtained from the test set spectra. The pooled variance was calculated and control charts made.

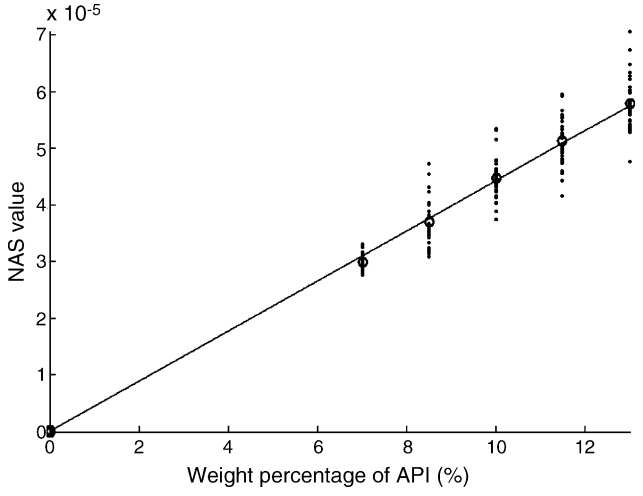

Fig. 3. NAS values vs. concentration plot. The open circle symbolizes the average calibration spectra and the dots symbolize the test set spectra.

4.2.2. Mixing data

The concentration predictions of the mixing data were calculated. Variances and means were plotted in the variance chart and mean chart (Fig. 4).

The variances at time points 16, 20 and 25 min were all within the limits and at 25 min the variance was lowest. The variances at 30, 35 and 40 min were higher then the 'golden batches'. This behaviour of the variances clearly shows a de-mixing behaviour, i.e. optimal mixing time existed around 20–25 min and further mixing worsened homogeneity.

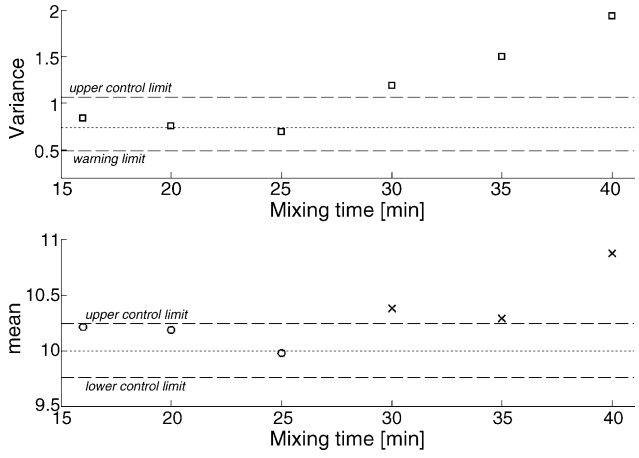

Fig. 4. Mixing experiment data in variance chart and mean chart using quantitative model. The mean values symbolized with (×) indicate that the assumption for the *t*-test is not fulfilled because the variance is different from the 'golden batches'.

The means at 16, 20 and 25 min were also within the target limits. At 25 min the batch was on target with a variance comparable to the 'golden batch'. This optimal mixing time was also identified using the qualitative model.

The results showed that with the qualitative NAS method it was possible to monitor API homogeneity though no calibration model with reference values was developed. The qualitative method required fewer batches and is therefore a strong candidate as method in the early development stages. The advantage of the quantitative model over the qualitative approach was that tighter limits were obtained due to the more degrees of freedom and calculations were in concentration values. Also the mean could be compared to the declared target content of the pharmaceutical product. These differences make adaptation into a highly regulated manufacturing environment easier and the interpretation is straightforward.

Secondly, by making a quantitative calibration line, non-linear behaviour can be detected and more confidence can be put into the pre-processing method and wavelength selection, this is not possible in the qualitative model. With the current formulation no such problems were observed but this is not always the case.

4.3. Validation of proposed methods

To validate the NIR methods comparisons were made between traditional HPLC and NIR using four different validation batches. All four validation batches were prepared with 10% (w/w) API. The differences between the batches were the order of mixing the constituents and the mixing time. When mixing was finalized, 30 NIR spectra were recorded and 12 powder samples were removed for HPLC analysis from each batch. First results from HPLC will be demonstrated and then compared to NIR results.

4.3.1. Validation batch evaluation using HPLC

From the 100% calibration batch ('golden batch') 30 powder samples of 300 mg were removed and subject to HPLC analysis. A variance chart was developed. The limits were calculated using Eqs. (6) and (7), substituting $s^2_{\text{total NAS,gb}}$ with $s^2_{\text{total HPLC,gb}}$ and a mean chart was developed using 10% (w/w) as target and Eq. (9) to calculate the upper and lower control limits, substituting $s^2_{\text{total c,gb}}$ with $s^2_{\text{total HPLC,gb}}$.

Now for each of the four validation batches the following was done; the variance and mean of the 12 HPLC values were calculated and plotted in the variance and the mean chart (Fig. 5). Outliers were identified using a Grubbs test ($\alpha = 0.05$) and removed prior to the calculation of the variance and mean.

The variances of batches 2 and 4 were within the limits and thereby equal to the 'golden batch'. The variance of batches 1 and 3 was below the warning limit signalling model incompetence. The means of the validation batches were all lower then the target value. The means of batches 1 and 3 (symbolized with ×) could not be compared with the 'golden batch' using the *t*-statistic because their variances were different from the 'golden batch'.

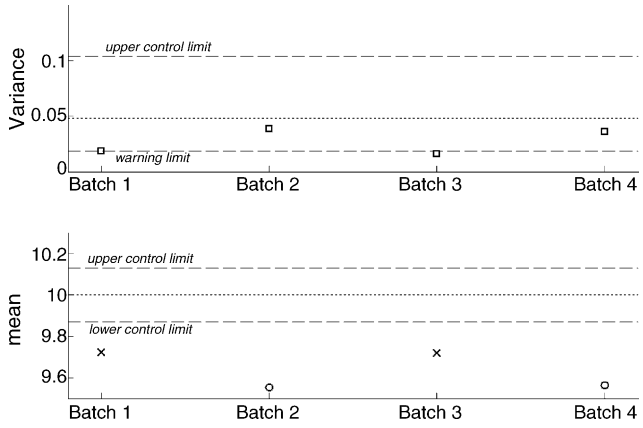

Fig. 5. Variances and means of HPLC values from validation batches in variance and mean charts.

4.3.2. Validation batch evaluation using NIR

Concentration predictions of the NIR spectra from the four validation batches were calculated. Variances and means were calculated and plotted in the control charts (Fig. 6).

The variances for batches 2 and 4 were within the upper control limit and the warning limit. The variances of batches 1 and 3 were below the warning limit signalling model incompetence as in the HPLC variance chart.

The mean of batch 1 was on target but due to the fact that the variance of batch 1 was different from the 'golden batch' the mean could not be compared to the 'golden batch' using *t*-statistics. The means of batches 2, 3 and 4 were below the lower control limit and smaller then target. This was also observed using HPLC.

The conclusions when using NIR or HPLC were comparable, i.e. the variances of batches 2 and 4 were similar to the 'golden batch' and their means below the target value. Also the variances of batches 1 and 3 were below the warning limit signalling model incompetence.

When the control limits for the HPLC values were compared to the control limits for the quantitative NIR method it was clear that there were big differences. The reason was the effective sample size in the two methods which was very different. In the

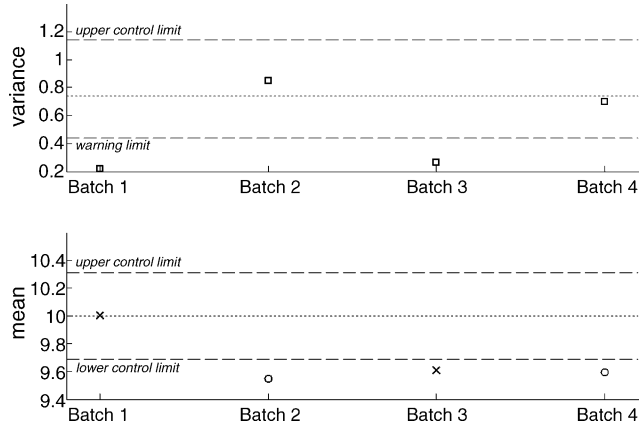

Fig. 6. Variances (upper figure) and means (lower figure) of concentration NIR predictions for the validation batches.

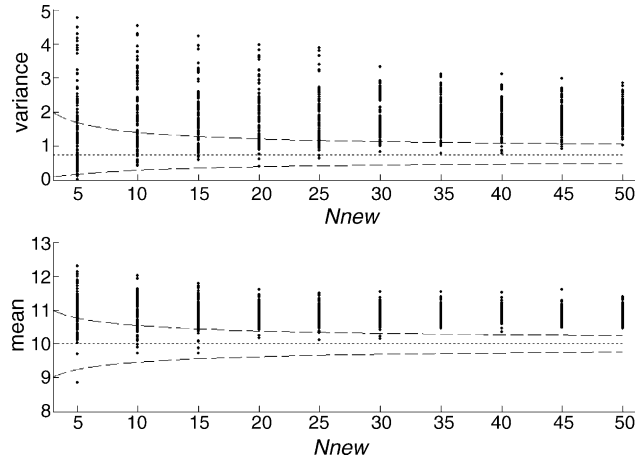

Fig. 7. Control limits as a function of number of samples used to calculate limits.

HPLC method the average sample weight was approximately 300 mg. The effective sample size in a NIR measurement is defined as the amount of sample that contributes to the spectrum. It is difficult to quantify the effective sample size for a NIR spectrum but some methods have been proposed [22]. Effective sample size ranged from 34 to 492 mg/cm^2 illuminated spot area in some typical pharmaceutical powders. The NIR probe used in this study had a spot area of 0.13 cm^2 which would correspond to effective sample sizes from of 5 to 64 mg.

4.4. Control limits dependency on number of samples

The ultimate method to characterize blend homogeneity would be a technique where the entire batch would be sampled. In practice this is not possible and one has to consider how many times it is needed to stick the probe into the powder bed in order to make a good characterization of the homogeneity.

When the 'golden batch' is measured with NIR, many spectra, e.g. 50–100 should be used. If new batches are evaluated at many time points and in a manufacturing environment these numbers of spectra are maybe not practical. Therefore the influence of number of spectra was examined.

In order to show how the number of spectra would influence the control limits and the conclusions one derives from the results, spectra from the mixing experiment were used. In the mixing experiment de-mixing was clearly identified at 40 min (Fig. 4). Because of the clear indication of a heterogeneous mixture at time point 40 min, the 50 concentration predictions from time point 40 min were used for demonstration because they were excellent to demonstrate the danger of using too few measurements to detect de-mixing.

First control limits for the variance and the mean chart were calculated using different values for N_{new} from 3 to 50. The control limits were plotted against N_{new} (Fig. 7). As expected the control limits were wide for small N_{new} and narrowed when N_{new} increased. Then the 50 concentration predictions and variances and means were calculated using different number of concentration predictions (N_{new}). Example wise for $N_{new} = 10$ the calculations were made in the following manner:

(1) Ten values were picked randomly from the original 50 concentration predictions. After a value was picked it was "returned" meaning that each value had the same probability of being picked every time a value was selected, i.e. bootstrap resampling [23].
(2) The mean $\bar{c}$ and variance $s^2_{total,c}$ were calculated from the 10 values.
(3) Steps 1–2 were repeated 100 times.
(4) The 100 mean and variance calculations were plotted in the control charts.

Steps 1–4 were performed using 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50 values using the power calculation method described above. The results are displayed in Fig. 7.

It was observed that when using a small number of samples, e.g. 10 a large proportion of the variances and means were within the control limits (Fig. 7) though they supposed to be above the upper control limits in both charts. When the number of samples was increased to 20, only a few results were below the upper warning limits.

The results showed how the risk of committing a Type II error (false positive) increased if N_{new} was low. One has to balance the risk of committing Type II errors with a practical number of spectra in order to choose N_{new}. In the present case 20 spectra seemed sufficient to identify de-mixing.

5. Conclusions

Two NIR models were developed for monitoring blend homogeneity using a NIR probe. First a qualitative model using NAS values from a well-mixed batch were used to generate control charts for future observations. Second a quantitative regression model between NAS values and reference concentration was developed for a set of calibration batches. With the regression model and well mixed batches, control charts were developed for the quantitative model. The qualitative method required less batches to develop and the performance was similar to the quantitative model but some other advantages of the quantitative method were presented. The qualitative model is suitable for initial R&D studies when developing a new solid dosage form drug product. When further development activities have been conducted or prior to implementation in manufacturing supplementary calibration batches can be measured and a quantitative model developed. Generally the models are easily developed and the method and control charts mimics to some degree the current methodologies for determination of blend uniformity which makes implementation or replacement of current methods easier.

In our opinion the proposed methods are better to compare mixing processes and formulation performance then current methods [1]. In current methods two statistics are usually calculated and used for homogeneity evaluation and batch-to-batch comparison, i.e. the mean value ($\bar{c}$) and the relative standard deviation (R.S.D.%). The R.S.D.% is calculated by dividing standard deviation with the mean value (R.S.D.% $= 100 \times s/\bar{c}$). The R.S.D.% is then dependent on two numbers and therefore not a good statistic to do batch-to-batch comparison of

homogeneity, i.e. different information is mixed up in the statistic. With our proposed NIR methods the comparison is performed with two statistics, i.e. mean concentration and variance which are independent and both can be directly compared to the target concentration of the drug product and the variance of a well-mixed batch. Another aspect of the methods is that the control limits are based on actual process measurements and is therefore related to the process capability. In the standard regulatory approaches limits are generic limits and do not relate to the specific formulation or process.

The proposed methods were validated by comparing results obtained with NIR and with traditional HPLC analysis. The comparisons showed that both the qualitative and quantitative NIR models showed similar results as HPLC.

It was shown how the width of the control limits decreased when more samples were used and how the risk of committing a Type II observation error also dropped when the number of measurements were increased. The result was used to provide guidance of picking a practical number of measurements.

The influence of the effective sample size on the variance was demonstrated by comparing NIR variance with HPLC variance. The NIR variance was approximately 10 times higher then the HPLC variance because of the small sample size in NIR.

The NIR models and control charts were used for monitoring of a mixing study and evidence was found of de-mixing after certain duration of mixing. The result showed the necessity and importance of monitoring tools like the proposed NIR methods.

Acknowledgements

Novo Nordisk, Corporate Research Affairs (CORA) who sponsored this work as a part of E.T.S. Skibsted's Ph.D. project. Technician Inga Fuchs who performed most of the mixing experiments and the staff at Analytical Development, CMC Development, Novo Nordisk A/S who performed the HPLC analysis.

References

[1] Uniformity of dosage units, in: U.S. Pharmacopeia (USP), 2003. pp. 2227–2229 (Chapter 905).
[2] F.C. Sánchz, J. Toft, B. Bogaert, D.L. Massart, S.S. Dive, P. Hailey, Fresenius J. Anal. Chem. 352 (1995) 771–778.
[3] S.S. Sekulic, H.W. Ward, D.R. Brannegan, E.D. Stanley, C.L. Evans, S.T. Sciavolino, P.A. Hailey, P.K. Aldridge, Anal. Chem. 68 (1996) 509–513.
[4] D.J. Wargo, J.K. Drennen, J. Pharm. Biomed. Anal. 14 (1996) 1415–1423.
[5] P.A. Hailey, P. Doherty, P. Tapsell, T. Oliver, P.K. Aldridge, J. Pharm. Biomed. Anal. 14 (1996) 551–559.
[6] S.S. Sekulic, J. Wakeman, P. Doherty, P. Hailey, J. Pharm. Biomed. Anal. 17 (1998) 1285–1309.
[7] R.D. Maesschalck, F.C. Sánchz, D.L. Massart, P. Doherty, Appl. Spectrosc. 52 (1998) 725–731.
[8] O. Berntsson, L.-G. Danielsson, M.O. Johansson, S. Folestad, Anal. Chim. Acta 419 (2000) 45–54.
[9] O. Berntsson, Characterization and application of near infrared spectroscopy for quantitative process analysis of powder mixtures, Doctoral Thesis, Kungl Tekniska Högskolan, Department of Chemistry, Stockholm, 2001.
[10] A.S. El-Hagrasy, H.R. Morris, F.D. Amico, R.A. Lodder, J.K. Drennen, J. Pharm. Sci. 90 (2001) 1298–1307.
[11] O. Vaizoglu, Tr. J. Phys. 23 (1999) 97–104.
[12] F.J. Muzzio, P. Robinson, C. Wightman, D. Brone, Int. J. Pharm. 155 (1997) 153–178.
[13] A. Lorber, Anal. Chem. 58 (1986) 1167–1172.
[14] E.T.S. Skibsted, J.A. Westerhuis, H.F. Boelens, A.K. Smilde, D.T. Witte, Appl. Spectrosc. 58 (2004) 264–271.
[15] P.W. Hansen, J. Chemometr. 15 (2001) 123–131.
[16] H.F. Boelens, W.T. Kok, O.E. de Noord, A.K. Smilde, Anal. Chem. 76 (2004) 2656–2663.
[17] N.M. Faber, Chemometr. Intell. Lab. Syst. 50 (2000) 107–114.
[18] J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, 4th ed., Pearson Prentice Hall, Harlow, 2000.
[19] http://www.brukeroptics.com.
[20] Brukeroptics, OPUS, Version 4.2.
[21] MathWorks Inc., MatLab, Version 6.5 R13.
[22] O. Berntsson, L.G. Danielsson, S. Folestad, Anal. Chim. Acta 364 (1998) 243–251.
[23] B. Efron, R.J. Tibshiran, An Introduction to the Bootstrap, Chapman & Hall, New York, 1993.