Analysis of a pharmaceutical process powder mixture with NIR spectroscopy and chemometric tools

Rodrigues, A.[1]; Almeida, J.[2]; Lopes, L.[3]; Menezes, J.[1]

[1] Instituto Superior Técnico da Universidade de Lisboa

[2] Faculdade de Farmácia da Universidade de Lisboa

[3] Generis Farmacêutica, S.A.

ABSTRACT: This work aimed at monitoring a pharmaceutical powder mixture process, using near-infrared (NIR) spectroscopy, which is a technique that has growing acceptance in terms of industrial application in such processes. The most commonly used approach consists in monitoring the mixing process by analyzing spectra variability. Evaluating the spectral variation over time, allows the identification of the end point of the mixture but does not ensures that the product is within specification. Herein, are proposed two alternative strategies, designated as qualitative or quantitative, in order to evaluate different methods for monitoring and controlling mixing processes, capable of being applied in production situations, where of course the relevant compounds are active in the NIR. This work was based on a commercial pharmaceutical product but, due to confidentiality reasons, its name, nature and description of the compounds involved in the formulation are not mentioned.

The proposed strategies are based on samples produced at laboratory scale according to two experimental designs. While in the qualitative case, the spectra were evaluated using principal component analysis (PCA), in the quantitative approach were developed several calibrations based on partial least squares (PLS) between the spectra and the contents of the various components.

Both strategies have demonstrated viability to accompany this powder mixing process, constituting feasible and efficient alternatives to the currently adopted methodology.

Keywords: Powder mixing; chemometrics; near infrared spectroscopy; principal component analysis; partial leasts squares regression

INTRODUCTION

With increasing competition in recent years among pharmaceutical industry, different companies and pharmaceutical corporations have been increasingly need to differentiate themselves from their main competitors, giving increasing importance to satisfaction and consumer confidence.(*Comissão Europeia, 2008*). Often, this effort is also reflected by the enhanced process monitoring capability that requires the adoption of new methodologies and different analytical equipment. These new techniques are related to all unitary operations involved in the manufacture of pharmaceutical products, particularly in the monitoring of mixing, and when applied allow rapid and effective control over the quality attributes of the products *(El-Hagrasy, et al., 2001).*

The near-infrared spectroscopy (NIR) is an analytical technique that has been implemented when it is necessary to provide an industrial process for the manufacture of medicaments monitoring system in real time of the mixing steps. The NIR spectroscopy studies the interaction of infrared radiation with matter and has the advantage of being a non-destructive technique, fast, mechanically simple (possibly without the use of moving parts), precise and with good relation sign/noise. This instrumental technique generally requires the adoption of multivariate

analysis strategies and these may be qualitative or quantitative (*Morisseau, et al., 1995*).

This study proposes the development of models with principal component analysis (PCA) - qualitative methodology and partial least squares (PLS) regression - quantitative methodology. The most commonly used approach to monitor mixing processes by NIR consists in monitoring spectroscopy by evaluating the spectral variability versus time. This allows us to identify the end point of the mixture but does not ensures that the product is within specification. (*Naes et al., 2002).* This study proposes two alternative strategies: one qualitative, which is based on processing the whole spectrum, allowing the grouping of similar samples and other quantitative that permits association of the spectra with the reference values for each analyzed parameter (e.g. a principle content active), so that can be estimated production samples. These approaches aim to evaluate monitoring and controlling mixing processes by two different methods, capable of being applied in production situations, where of course the relevant compounds are active in the NIR. This work was based on a commercial pharmaceutical product but, due to confidentiality reasons, its name, nature and description of the compounds involved in the formulation are not mentioned.

MATERIALS AND METHODS

Commercial product and industrial manufacturing process

The commercial product used in this work is produced in Generis Farmacêutica, S.A.. It is composed of seven components, which for reasons of confidentiality will be referred to as components A, B, C, D, F and G, and the respective proportions w/w (%) shown in the table below. The product will be referred to as P.

The industrial manufacturing process of the product P begins with the previous calibration for each component in an oscillating granulator with mesh opening of 0.8 mm and 0.5 mm diameter light wire.

Table 1 Drug composition P.

Components	Proportion w/w (%)
A	72.285
B	9.543
C	6.091
D	10.761
E	0.183
F	0.122
G	1.015

The F component is ground to fine powder in a mortar before heading to the granulator. Then comes the process of mixing that occurs in a V mixer with capacity of 1000 L, and is divided into three different mixtures, the mixture one with duration of 20 minutes and 30 minutes for the remaining two blends. The components are added at different times as well as the A component quantities are divided in the three mixtures, since this is a major component and expected to find the final homogeneous mixture. At the end, the mixture homogeneous goes the packing phase, where it is wrapped correctly. During the primary packaging, samples for analysis are taken for quality control, designed industrial samples.

Assay and sampling of the industrial process

The assay components of the mixtures was made through certificate methods and in the quality control laboratory of Generis S.A.. Only components A, B, C and D were assayed. The components B and C are very similar in chemical terms, and during the assay is measured an H component which belongs to the components B and C. For the D component is assayed only the corresponding ion.

The reference methods for confidentiality reasons are not presented in detail. Below a table is shown with the specifications due to w/w (%) of the respective assayed excipients.

Table 2 Specifications in the commercial product P of components A, B, C, D, H based on the reference methods.

Components	Proportion w/w (%)
A	68.63 – 75.94
B	5.36 – 5.93
C	3.03 – 3.35
D	7,35 – 8.12
H	8.26 – 9.12

Samples

A total of nine industrial samples were collected. Eight of these were within specification and corresponded to a batch rejected. The eight samples within specification are identified 1-8 and the sample out of specification as R1.

It was also made reproduction of two batches on a laboratory scale with a total mass of 600g, using a V-blender with 2 liters where the mixture 3 was monitored over time. This analysis was performed using NIR, where every three minutes was inserted the probe into the mixture. At the end of the mixture 3, a sample was taken of mixer for analysis in quality control (assays) and NIR spectrophotometer (samples referred to as Batch 1 and Batch 2).

There were two designs of experiments (DoE) different based on the MODDE software version 10.1 (Umetrics, Sweden). They were weighted quantities for the respective sampling glass making a total mass of 50g. This powder mixture contained within the sampling glass was stirred for 5 minutes. In both DoEs only the major components (excipients A, B, C and D) were varied, and in the first DoE-QLT (DoE utilized for the qualitative approach) the range of changes was ±1,5% compared to theoretical value of proportion in the commercial product. In the second DoE, used for the quantitative approach (DoE-QNT), the range was within ±10% of the theoretical value of the ratios in the commercial product.

NIR Spectroscopy

The NIR spectra were measured in diffuse reflectance mode, a spectrophotometer Thermo Nicolet FT-NIR model ANTARIS, with InGaAs detector, spectral range between 10000 and 4000 cm^{-1} (1000 - 2500 nm), resolution $4cm^{-1}$ and considered the average of 36 readings and controlled by RESULT software (Thermo Nicolet, EUA). Readings were performed by intermediate of a diffuse reflectance probe SabIR model (Solvias, Switzerland) with lighting area of 0.2 cm^2. All samples were measured in duplicate and the average of these spectra considered. Before starting the spectral acquisitions and at intervals of one hour a blank was made with Teflon (spectralon®).

All the samples described in the previous point were analyzed.

Chemometric treatment

The chemometric treatment was done using the MATLAB software, version 8.3 (MathWorks, Natick, MA) e PLS Toolbox, version 7.5 (Eigenvector Research Inc., Wenatchee, WA).

RESULTS AND DISCUSSION

Qualitative Analysis- PCA model

To perform the qualitative analysis a PCA model was built based on the experimental design, in which was varied only the four components A, B, C and D, in a very restricted range (±1.5%). The preparation of this DoE was made taking into consideration the range of specification of each component. In addition to the results of the quality control, a NIR analysis was performed with the 24 samples, which allowed the calibration of the PCA model, and through a group of 6 samples PCA model was validated. After model validation the samples industrially collected were projected and the evolution of the mixtures laboratory monitored was accompanied.

In the **Table 3** are summarized the obtained parameters to the construction of the PCA model

Table 3 Summary of the parameters involved in the PCA model developed for the qualitative approach.

Parameter	PCA
NIR Spectral Region (cm^{-1})	8670 - 4115 4073
Pre- Treatment	Savitzky-Golay (1st derivative, –21 filter size points, order polynominal 2)
Number of main Components	2
Cumulative variance (%)	99,62

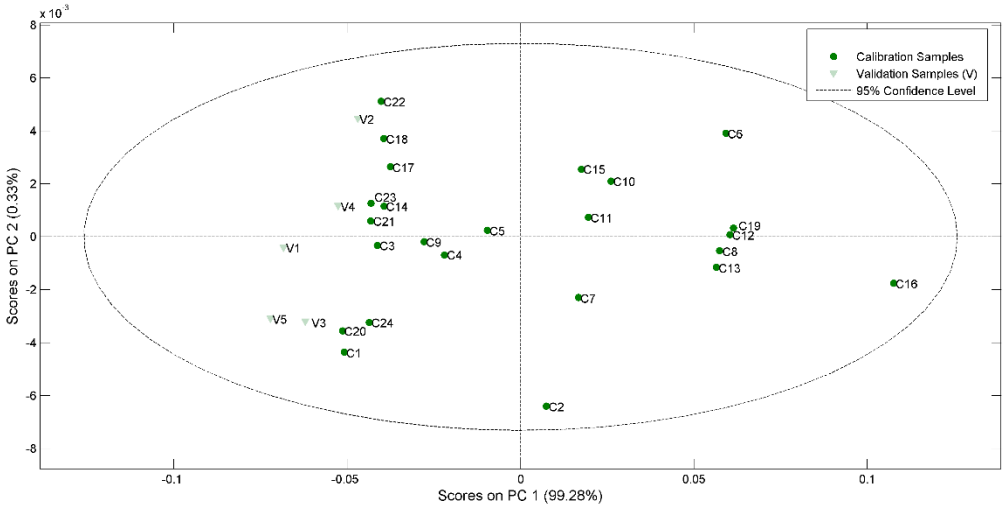

Figure 1 Scores map on the validation of the PCA model, with projection validation samples (●: calibration samples (DoE); ▼: validation samples).

By scores of chart (**Figure 1**) it can be seen that there is some dispersion of the samples without, however exist a sample standard or clusters which would not be expected since these are samples produced by an experimental design. Also, we can see that the DoE-QLT allowed the production of samples that represent various possible scenarios around what would be a mixture containing the theoretical proportions for the product P.

We proceeded to the validation of the PCA model (**Figure 1**), where it was found that all the samples used for model validation are inside the ellipse, as was predicted.

After validating the model, we start to test samples (industrial and laboratory). The results are shown in **Figure 2**. All samples are within the confidence limits, indicating that they are within specification, however, the R1 sample appears outside the ellipse and on the opposite site of the other test samples. This result is consistent with a rejected batch (out of specification).

Concerning the projection of the samples monitored by the mixture process produced under laboratory conditions (Batch 1 and Batch 2), there was found that is possible to follow the evolution of the mixture over time. This evolution begins with the samples outside the ellipse, which indicates that the mixture is still not homogeneous. After approximately 27 minutes, the mixture come into the ellipse unstarting to get homogeneous. The last three minutes are crucial for ending of the mixture, being the final mixture within specification, and this information was confirmed by the PCA model and the reference methods.

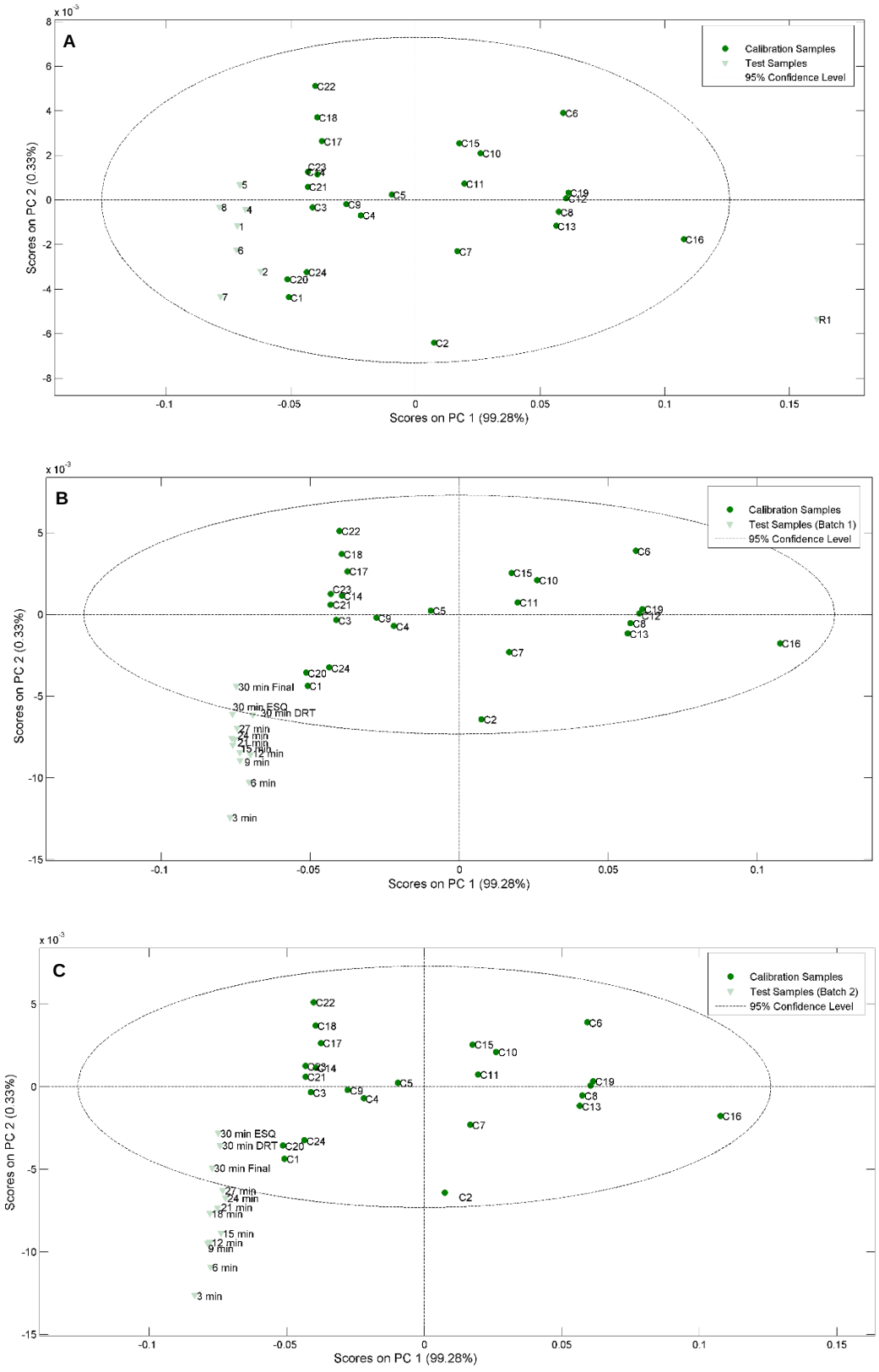

Figure 2 Map of relative model PCA scores with project of the samples obtained during monitoring of the mixtures made under industrial (A) and laboratory (B-Batch1; C- Batch 2) conditions (●: calibration samples (DoE); ▼: test samples).

Quantitative Analysis- PLS model

For the quantitative analysis only four components (A, B, C and D) were taken into account, since they are those present in higher concentrations in the mixture, allowing to verify the sensitivity of the PLS model when compared with the proportion in the mixture.

Based on the experimental design was range about 15% of the standard value for each component, to obtain a set of 24 samples with variations in components A, B, C and D. This range is larger when compared with the previous DoE, as in this analysis we intend to create models that reflect quantitative data, according to determinations of the different components.

In accordance to the 24 samples obtained, it was generated an additional set of 6 samples for optimizing the models (called validation samples) and 6 samples for external validation (referred to as test samples). These samples were generated within the range used for calibration of the different models. The proportions of these 12 samples were randomly generated with the only requirement that all components were within the 10% range around the average value for the proportions of components. A total of 36 samples were used to develop this methodology. All samples were analyzed by reference methods as well as NIR

The optimization of the PLS model (preventing the phenomenon of overfitting or underfitting) for each component was made as follows:

1) Choice of a spectral region;
2) Choice of a pre-processing (or combination) of the mentioned;
3) Calibration models given the DoE samples and different numbers of latent variables;
4) Projection of validation samples (V) for each model and obtaining the respective RMSECV;
5) The decision model that produces a smaller to a lower RMSECV.

The spectral region obtained for the model optimization can still be adjusted if it is considered necessary to add more spectral information in the case to add noise to the model. Finally, it is important to observe the straight calibration and the appropriate R^2 for the validation set (V) and verify whether additional adjustments of the model are needed.

After optimization of the models for the various components, the test samples (T) were designed to the model for the test and to confirm the model performance in terms of accuracy (obtaining the RMSEP). Finally the industrial and laboratory samples were designed by calculating the error (RMSEP).

The results obtained for the different models are summarized in **Table 4**. It can be seen that the component A is what presents the best model since it has a high R^2, and a low prediction error (RMSEP), given the range of variation component A. This result was expected because the component A was very active in NIR and it is the major component of the mixture. The components B, C and H have demonstrated a weaker model, which may be due to the low absorbance in the NIR. Although component D also has low concentrations, the model obtained is fairly better that previous ones, because this component is more active in the NIR.

In **Table 5**, are presented the comparisons between industrial and laboratory samples. It can be seen that the component A is that which has smaller difference between the real and predictive value, a result reflected in the obtained error. For minor components (B, C, D and H) error obtained increases with the decreasing proportion of the component in the mixture, while R^2 is directly proportional with the increase of the proportion of the respective components.

Table 4 Summary of parameters obtained for the various models PLS, relative to components A, B, C, D and H.

	Component				
Parameter	**A**	**B**	**C**	**D**	**H**
NIR Region spectral (cm^{-1})	9735-5307	7671 - 4389	7671 - 4389	9735 - 4385	9735 - 4385
Pre- Treatment	Savitzky-Golay (1st derivative, – 19 filter size points, order polynomial 2)	Savitzky-Golay (1st derivative, – 15 filter size points, order polynomial 2)	Savitzky-Golay (1st derivative, – 15 filter size points, order polynomial 2)	Savitzky-Golay (1st derivative, –15 filter size points, order polynomial 2) + SNV	Savitzky-Golay (1st derivative, – 15 filter size points, order polynomial 2)
Number of latent variables	4	5	3	3	3
RMSEC	0,930	0,198	0,584	0,247	0,502
RMSEP	1,70	1,32	0,791	0,869	1,90
R^2	0,920	0,798	0,694	0,954	0,836
Cumulative variance of x captured by the model (%)	90,3	91,3	88,9	80,4	80,4
Cumulative variance of y captured by the model (%)	96,4	99,0	82,2	92,8	92,8

Table 5 RMSEP, R^2 and error obtained in the different PLS models in the prediction of industrial and laboratory samples.

Designation of the sample	**Error associated with the component (%)**				
	A	**B**	**C**	**D**	**H**
RMSEP	1,58	2,04	1,44	1,08	1,53
R^2	0,956	0,680	0,532	0,930	0,840
1	-0,026	0,059	-0,013	0,052	-0,017
2	0,004	0,002	-0,244	0,104	-0,166
3	0,034	-0,087	-0,136	0,067	-0,129
4	0,016	0,032	-0,332	-0,052	-0,017
5	-0,021	0,332	-0,477	0,010	0,123
6	0,023	0,811	-1,01	-0,123	0,375
7	0,049	0,767	-0,498	-0,293	0,358
8	0,021	0,256	-0,626	0,226	0,042
R1	0,006	-0,169	0,210	0,128	0,014
Batch 1	0,005	0,099	0,069	0,154	-0,059
Batch 2	0,002	0,056	-0,221	-0,001	-0,033

Comparison of the two approaches

Both approaches applied in this study were shown to be applicable to the monitoring of the mixture. Although, it was demonstrated that the obtained PCA model in the qualitative analysis was simple and robust when compared with the PLS models obtained in quantitative analysis. There are several factors related to the low reproducibility of PLS models, including the low absorption of the NIR of some components, respective proportions and the PLS worsening with the decreasing proportion of the component in the mixture. The reason of ranging the four components at the same time, also influences the construction of the PLS models, particularly in terms of introduction of interference that must be compensated by the model. In an attempt to improve the PLS models, it would be relevant to develop a new DoE, made to range only one component at a time, allowing a higher predictive efficiency, particularly in cases of minor components in the mixture.

CONCLUSIONS

It was concluded that NIR spectroscopy associated to chemometrics techniques, permits the construction of models able to monitor the mixture process. Both approaches proved possible implementation.

The created models allowed two different strategies. Qualitative analysis has shown that the samples were in or out of specification. According to the results, it was found that the industrial samples (commercial product) were used within the ellipse that defines the confidence interval, excluding the point corresponding to the lot rejected. Regarding the monitoring of the mixture of the two batches it was found that the last three minutes are critical so that the mixture became homogenous, which may indicate that the mixing time is not sufficient.

With regard to quantitative analysis, it can be considered that there were obtained models with good predictive power for prediction of components A and D and less good models for others. This result can be explained by the fact that the components A and D present enough activity in terms of NIR radiation and due to the selection of a significant spectral region. On the other hand, the remaining components (B, C and H) are less active in the NIR and were found in lower proportions in the mixture, considering that the NIR is also very sensitive to “vestigial” components (below 1% w/w should be used with caution). A strategy for improving the models could be the ranging of only one component at a time, making it a more sensitive and robust model for each of the components.

Comparing both approaches, it appears that the first analysis was able to obtain more reliable results, while the second analysis, due to poor absorption of the components B, C and D, does not allow us to have a good prediction of the same. It is also important to note that the two analyses are important because they allow us to draw different conclusions, although quantitative analysis is more complete on the results.

In general, the obtained models enables the monitoring of the mixture and to understand if any adjustments are needed, so that the final samples can be analysed quickly and inexpensively before proceeding to quality control.

Acknowledgements

The authors would like to thank to MSc Tânia R. Dias for checking grammar and spelling in English.

REFERENCES

Comissão Europeia (2008). EudraLex - Normas que Regulam os Medicamentos na União Europeia. Normas orientadoras sobre Boas Práticas de Fabrico Medicamentos para uso Humano e Veterinário. Vol. 4-Parte I.

El-Hagrasy, A.S., Morris, H.R., D’Amico, F., Lodder, R.A., Drennen, J. K..(2001). Near – Infrared Spectroscopy and imaging for the Monitoring of Powder Blend Homogeneity. J. Pharm. Sci, 90 (9), 1298- 1307.

Morisseau, K.M., Rhodes, C.(1995). Pharmaceutical uses of near-infrared spectroscopy. Drug Dev. Ind. Pharm., 21 (9), 1071-1090.

Naes, T., Isaksson, T., Feam, T., Davies, T.. A User Friendly Guide to Multivariate Calibration and Classification. Sixth Edition NIR Publications (2002)