USE OF NIR SPECTROSCOPY AND MULTIVARIATE PROCESS SPECTRA CALIBRATION METHODOLOGY FOR PHARMACEUTICAL SOLID SAMPLES ANALYSIS

Vanessa Cárdenas Espitia

Prof. Dr Marcel Blanco Romía

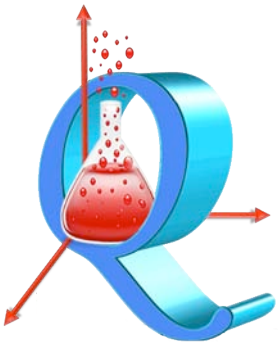
Prof. Dr. Manel Alcalà Bernardez

Applied Chemometrics Group

Chemistry department

Universitat Autònoma de Barcelona

Spain



Memoria presentada con el fin de aprobar el correspondiente módulo de iniciación a la investigación y trabajo de fin de master, correspondiente al Master Oficial en Ciencia y Tecnologías Químicas.

Vanessa Cárdenas Espitia

Visto bueno

Prof. Dr. Marcel Blanco Romía

Prof. Dr. Manel Alcalà Bernardez

Bellaterra, 3 de Septiembre de 2012

AGRADECIMIENTOS

Para empezar quisiera expresar mis mas sinceros agradecimientos al Profesor Marcel Blanco Romia y al Dr. Manel Alcalà Bernardez por su paciencia, enseñanzas e incondicional asesoria. Gracias por orientarme en estos primeros pasos de la quimiometría.

A mis compañeros del grupo de investigación en quimiometria aplicada y todo el personal docente, de cada uno de ellos he tenido la oportunidad de aprender. Gracias por su ayuda y soporte técnico. Especialmente a mis compañeros del día a día (Anna, David, Juan y Sergi) quienes han hecho de lab una muy buena experiencia.

A la Universitat Autonoma de Barcelona, quienes a través de la beca PIF hicieron possible la realización de este master.

Este es el momento tambien para agradecer a Mamá, Papá y a mi hermano Santi quienes de lejos siempre me apoyan y me han enseñado a ser fuerte lejos de casa.

También a todos quienes contribuyeron de una u otra forma para hacer este trabajo possible.

Gracias a todos

Vanessa

ABSTRACT

Accomplish high quality of final products in the pharmaceutical industry is a constant challenge that requires the control and supervision not only of final products but also of all the manufacturing. This request created the necessity of developing fast and accurate analytical methods that can determine important parameters of final product as of the process itself.

Near infrared spectroscopy together with chemometrics data analysis, being one of the most recent methodologies, fulfill this growing demand. The high speed to providing relevant information, the versatility of its application to different types of samples and the quality of the results are some of the most important characteristics, leading this combined techniques as one of the most exact and appropriated in the field.

This study is focused on the development of a calibration model based on synthetic samples (powder laboratory mixtures) able to determine amounts of active pharmaceutical ingredient (API) from industrial granulates using NIR and chemometrics data analysis.

Moreover, in this study the process spectra methodology is used, which is a new method to calculate process variability (such as granulation, compaction or coating) by difference amongst industrial and synthetic samples with the same composition. After the addition of process spectra to the powder laboratory samples, this new set containing the process variability is used to build up a PLS model.

The following chapters describe and discuss the relevant information obtained in this work**.**

TABLE OF CONTENTS

1. INTRODUCTION

The quality control in pharmaceutical industry is strictly regulated by different national and international organizations. These organizations are focused on the monitoring and control of raw materials, final products and the industrial process itself.

Nowadays, the use of different analytical technologies involving labor intensive, time consuming and slow protocols for pharmaceutical quality control has create the need to replace them for fast, accurate and non-invasive analytical methods. Therefore, the use and importance of NIR spectroscopy together with chemometrics data analysis has increased in the last years.

This study shows the advantages of using these combined methodologies for the extraction of relevant information from the analytical data and the subsequent application to industry.

Consequently, the aim of this work is the development of a calibration model based on synthetic samples (laboratory powder mixtures of active pharmaceutical ingredient (API and excipients) able to predict concentration of API in industrial granulates.

The formulation object in this study contains as active the pharmaceutical ingredient dexketoprofen trometamol, which is classified as a non-steroidal anti-inflammatory drug (NSAID). This substance blocks the action of cyclo-oxygenase, which is involved in the production of prostaglandins. Prostaglandins are produced in response to certain diseases or injuries and are the cause of swelling, stiffness, tenderness, increased temperature, inflammation and pain. By blocking the cyclo-oxygenase, dexketoprofen prevents the production of prostaglandins and reduces the pain.

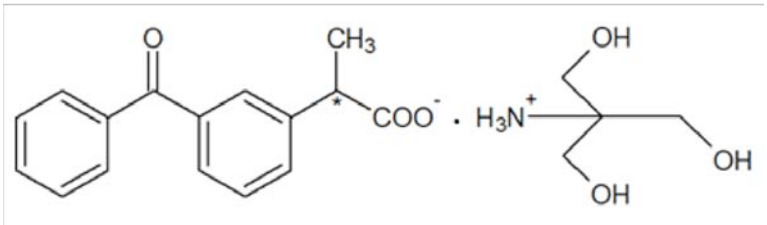


Fig. 1 Dexketoprofen Trometamol molecular structure

1.2 NEAR INFRARED SPECTROSCOPY (NIR)

Although the presence of light in the infrared region was observed in the 19th century by Herschel, the application and research of NIR as a useful analytical technique for the industry just had happened between the 70`s and 80´s.

Problems like absence of relevant structure information, lack of sharp peaks, loss in sensitivity compared to the mid infrared and the difficulty of making band assignments owing to the presence of numerous overtone and combination bands delayed the development of this technique.

Nowadays, a NIR spectrum can be well obtained in less than a minute (unusually fast spectra record compared with other analytical techniques), and advantages like no previous required sample preparation and being a nondestructive method have made of NIR very useful for the industry. Furthermore, if samples contain such bonds as C-H, N-H or O-H, and if the concentration of the analyte exceeds about 0.1% of the total composition, then it is very likely to yield acceptable results even in the hands of untrained personal (1, 2).

1.2.1 BASIC PRINCIPLES OF NIR

The NIR region in the electromagnetic spectrum is located at the wavelength range between 780 and 2565 nm corresponding to the wave number range 12820-3959 cm^{-1}, and it covers the wavelength range adjacent to the mid infrared, extending up to the visible region (3).

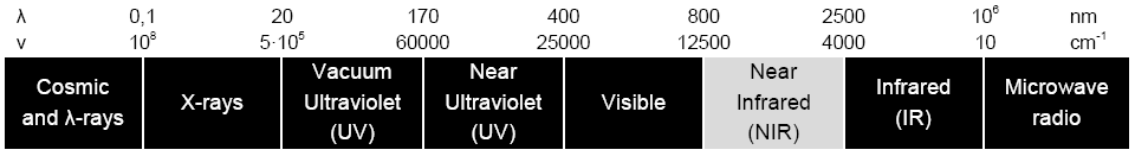


Fig. 2 Electromagnetic spectrum (4)

In NIR spectroscopy, the samples are irradiated with NIR light. This light is just absorbed by molecules when a change of the dipole moment occurs as a consequence of vibrations. R–H groups have a high dipole moment, thus, O–H, N–H, C–H, S–H bonds

are therefore strong NIR absorbers. Diatomic molecules (H_2, O_2, N_2) in another case do not absorb NIR radiation as no change in dipole moment occurs during its vibrations.

The potential energy of these vibrations is dependent on the bond length, and since the energy curve of an oscillating molecule is affected by intermolecular interactions, the equilibrium position is non-symmetric and the spacing between energy levels that the molecule can attain are not identical and the model of an anharmonic oscillator explain the situation for real molecules (3,5).

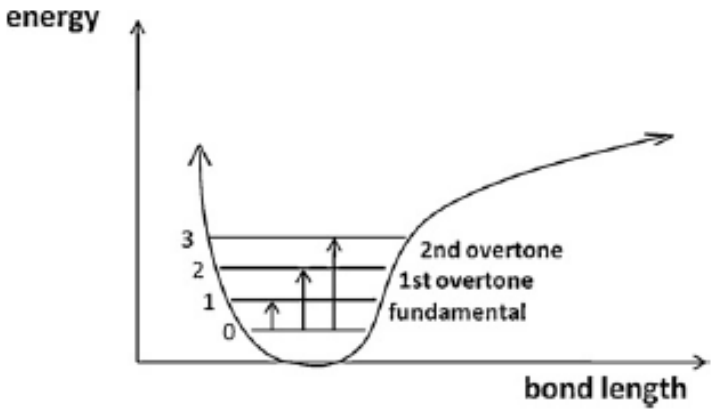


Fig.3 Anharmonic vibration model (5)

The most prominent absorption bands occurring in the NIR region are related to overtones and combinations of fundamental vibrations mainly due to –CH, -NH, -OH functional groups.

Since quantum mechanical selection rules do not rigorously exclude transitions with $\nabla\nu\rangle 1$ for anharmonic systems, transitions between vibrational states of $\nabla\nu=2$ or 3 are possible (Fig3). These multi- level energy transitions are the origin of NIR overtone bands (3, 5).

The combination bands are originated only in polyatomic molecules and these are the result of simultaneous energy changes of two or more vibrational modes.
Combination bands appearing between 1900-2500 nm are the result of vibrational interactions. Vibrations of molecules that absorb near infrared light occur in two modes stretching and bending. Stretching is defined as a continuous change in the interatomic

distance along the axis of the bond between atoms, while bending is a change in the bond angle (5, 6).

1.2.2 INSTRUMENTATION

The evolution of the instrumentation used in NIR spectroscopy has been based on the need of a faster and more flexibility of the analysis for different kind of samples. NIR spectrophotometers have the advantage of incorporating a broad variety of devices depending on the characteristics of the samples and the analytical conditions. This makes NIR spectroscopy a versatile and flexible technique as compared to other instrumental methods (7).

A NIR spectrometer is generally composed of a light source, a wavelength selection system, a sample holder or a sample presentation interface and a detector.

Blanco et al summarize the principal features of a NIR spectrophotometer with different devices in the following figure.

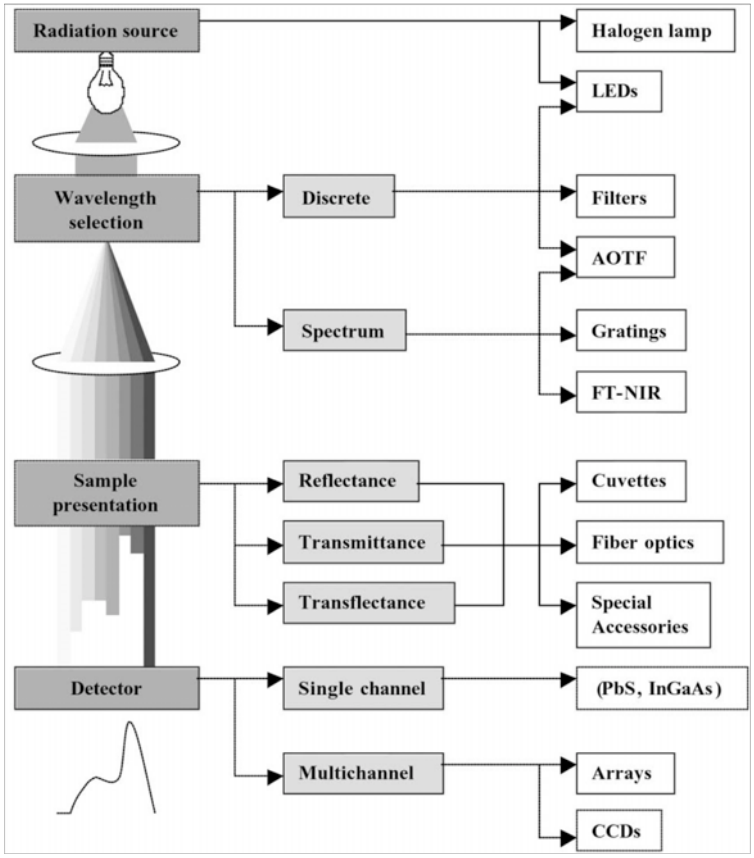


Fig 4.Main features of a NIR spectrophotometer (7)

The light source generates a beam that can irradiate the samples. The most commonly used is a halogen light with tungsten filament and quartz window that is capable to emit a continuous spectrum from 320 nm to 2500nm. Other light source that can be used is named LEDs (Light Emitting Diodes), which depending on their composition is able to emit up to 1600 nm. The halogen lamps require a wavelength selection system, while LEDs do not.

In function of the wavelength selection system, the NIR spectrophotometers can be classified in two types, dispersive and non-dispersive instruments. Within the dispersive instruments the most commonly used wavelength selection systems are the monocromators. The non-dispersive instruments are widely used, and the variety of selection systems is broad. Actually the instruments use different selection devices such as conventional filters, Fourier transform (FT)-NIR type and AOTFs (acusto-optic tunable filters).These chosen wavelengths by using radio-frequency signals to alter the refractive index of a birefringent crystal (usually TeO2) so that it transmits light of a given wavelength or performs a wavelength scan much more rapidly than with the previous designs (7, 8).

Detection in NIR spectroscopy uses devices comprising semiconductors such as PbS or InGaAs, like single channel detectors, in multi-channel detectors, several detection elements are arranged in rows (diode arrays) or planes charged coupled devices (CCDs) in order to record many wavelengths at once, so as to increase the speed at which spectral information can be acquired (7)

Regarding the sample acquisition, there are three modes to take a NIR spectrum (reflectance, transmittance and transflectance), and choosing one of the other rely completely in the nature of the sample (3, 7, 8).

1.3 CHEMOMETRICS

Chemometrics is an interdisciplinary field which uses mathematics and multivariate statistics in order to process, extract and understand relevant information from analytical data.

Analytical chemists are major users of chemometrics, however there are several fields which are supported strongly by the use of chemometrics from physical chemistry such as kinetics and equilibrium studies, to organic chemistry such as reaction optimization etc (9).

1.3.1 SPECTRAL PRETREATMENTS

Pre-processing of spectral data is often of vital importance to obtain as much information is possible from the analytical data. These pretreatments allow the corrections to the spectrum by increasing the signals and minimizing undesired information such as background noise or baseline shift (9,10).

The most common spectral pretreatments are described below:

- **Average spectra:** it involves the calculation of the absorbance average for each wavelength from replicate spectra of the same sample.

- **Standard Normal Variate (SNV):** It corrects multiplicative variations between spectra. These variations often originate from accidental or uncontrolled differences in sample path length, due to variations in sample physical properties (particle size, thickness), sample preparation, sample presentation and perhaps even variations in spectrometer optics. Sometimes such variations can be problematic because they are confused with multiplicative effects from changes in component concentrations, which often model the signal in quantitative applications. Multiplicative variations cannot be removed by derivatives, centering or scaling. The transformation is performed for each spectrum for which it will obtain an absorbance spectrum mean 0 and standard deviation 1. The equation for calculating the absorbance at $X_{i,m}{}^{SNV}$ can be calculated:

$$X_{i,m}^{SNV} = \frac{X_{i,m} - \bar{X}i}{Si}$$

Where $X^{SNV}{}_{i,m}$ is the absorbance value from the row i (or spectrum i) and the column m (or variable m) once the pretreatment is applied. $X_{i,m}$ is the original

absorbance value from the spectrum i and Si is the standard deviation from the row i (spectrum i).

- **Derivates:** The spectra derivation is also one of the most widely used pretreatment in NIR spectroscopy. It is use to minimize the problems of overlapping bands and baseline variations. One of the most used methods is that proposed by Savitzky-Golay. The first derivative can correct baseline shifts and the second derivative corrects deviations which vary linearly with wavelength.

1.3.2 REDUCTION OF VARIABLES BY ANALYSIS OF PRINCIPAL COMPONENTS (PCA)

Since multivariate NIR spectral data contain a huge number of correlated variables, there is a need for reduction of variables, i.e. to describe data variability by a few uncorrelated variables including the relevant information for calibration modeling. The best known and most widely used variable reduction

Method is principal component analysis (PCA). It is a mathematical procedure that resolves the spectral data into orthogonal components whose linear combinations approximate the original data. The new variables, called principal components (PC), Eigen-vectors or factors, correspond to the largest eigenvalues of the covariance matrix, thus, accounting for the largest possible variance in the data set. The first PC represents the maximum variance amongst all linear combinations and each successive variable accounts for as much of the remaining variability as possible (3).

PCA, however, results in an abstract mathematical transformation of the original data matrix, which takes the form

$$\boldsymbol{X = T.\ P + E}$$

Where *T* is called the scores, and has as many rows as the original data matrix; *P* is the loadings, and has as many columns as the original data matrix; the number of columns in the matrix *T* equals the number of rows in the matrix *P* (9).

A good example of a matrix transformation is graphically illustrate by Reich et al in the following graphic

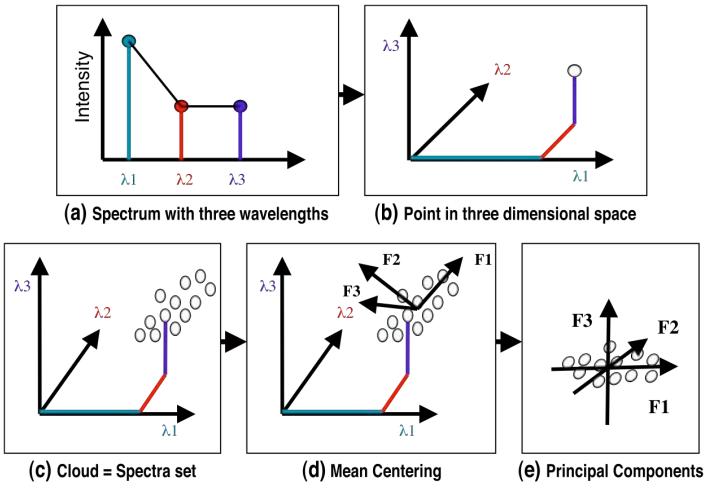


Fig.5. Transformation of a spectrum with three variables, i.e. wavelengths (a) to a new coordinate system with one axis for each wavelength thereby converting the spectrum to a single point in a three-dimensional space (b), cloud formation of several spectra (c), mean centering (d), and determination of principal components F1, F2 and F3 (e) (3).

1.3.3 MULTIVARIATE CALIBRATION FOR QUANTITATIVE ANALYSIS BY PARTIAL LEAST SQUARES (PLS)

Before performing any quantitative analysis in a NIR spectrometer, it has to be calibrated using multivariate methods. The calibration process basically involves the following steps:

1. Selection of a representative calibration sample set.

2. Spectra acquisition and determination of reference values.

3. Multivariate modeling to relate the spectral variations to the reference values of the analytical target property.

4. Validation of the model by cross validation, set validation or external validation.

One of the multivariate regression method most frequently used in quantitative NIR analysis is the partial least-squares (PLS) regression, which will be briefly describe in this study (9).

PLS is a method that generalizes and combines features from PCA and multiple regressions. It is particularly useful when a set of dependent variables from a large set of independent variables has to be predicted. The goal of PLS is to predict the regression coefficients in a linear model with a large number of x- variables that are highly correlated.

The PLS algorithm uses the information contained in both the spectroscopic data matrix, **X**, and the concentration matrix, **Y**, during calibration and compresses data in such a way that the most variance in both **X** and **Y** is explained. In this way, PLS reduces the potential impact of large, though irrelevant, variations in **X** during calibration. In PLS, each component is obtained by maximizing the covariance between **Y** and every possible linear function of **X**.

This regression controls two blocks of variables: predictors (X) and responses (Y). The two data sets can be decomposed separately through PCA.

This gives the outer relations:

$$\mathbf{X = T \times P' + RX}$$

$$\mathbf{Y = U \times Q' + RY}$$

The residuals (RX and RY) are minimized with these calculations and without making any effort to correlate the data sets. A correlation between the two data sets can be found by forming a linear inner relation between the scores for each PC:

$$\mathbf{\hat{u}h = bh \times th}$$

Where **h** is the number of the specific PC and **b** is the regression coefficient. This model shows only a weak relation between the data sets. To improve the model, information from the decomposition of one of the two blocks of variables exchanges information to the other and vice versa (11, 12).

1.3.4 MODEL VALIDATION

The best way to evaluate the predictive capacity of the model is running an external prediction test which will predict known Y values by the PLS model and will compare

them with the known values. Different global statistics values can be evaluated as average of residuals or standard deviation. However, the root mean square error (RMSE) and the relative standard error (RSE) are the standard values to use for evaluating a PLS model. These values evaluate the residuals with the reference values (13).

$$RSE(\%) = \frac{\sqrt{\sum_{i=1}^{n}\left(Yi^{nir} - Yi^{ref}\right)^2}}{\sum_{i=1}^{n}(Yi^{ref})^2}.100 \qquad RMSE(mg/g) = \frac{\sqrt{\sum_{i=1}^{n}(Yi^{nir} - Yi^{ref})^2}}{n}$$

Where n = number of samples, Yi^{nir} and Yi^{ref} are magnitudes of the determined property by NIR or reference method (Ref).

2. EXPERIMENTAL PART

This study shows the data extraction from NIR spectra through chemometrics tools that substantiate its use for the quality control in the industry, basically developing calibration models able to determine the amount of Active Pharmaceutical Ingredient (API) in pharmaceutical formulations.

Therefore, the aim of this study is the development of a calibration model (based on synthetic samples) able to predict amount of API in industrial granulates. The variability of the process is included through process spectra methodology (13).

The advantages of these combined techniques facilitate the development of this study and it will be mentioned below.

2.1 DESCRIPTION AND SAMPLE PREPARATION

2.1.1 Synthetic samples (powder laboratory mixtures)

The pharmaceutical formulation studied in this work is granulated dexketoprofen trometamol (DKP-Trom). The composition of API and excipients in the laboratory

samples corresponds to the proportions of industrial granulates manufactured by Menarini laboratories.

Composition of Formulation	**mg/g**
API (DKP.Trom)	14.7
Lemon flavor (Exc 1)	16
Yellow colorant (Exc 2)	0.64
A.glycirrinate (Exc 3)	0.8
Neohespiridine Dhchalco (Exc 4)	0.6
Sucrose (Exc 5)	967.2

Table1. Pharmaceutical formulation for granulated DKP.Trom

A total of 31 powder samples were prepared using an analytical balance and following the strategy API + Placebo. This strategy consists on the preparation of placebos (samples not containing API) with the different excipients of the formulation. By the subsequent weighting of placebos + API the formulation was considered as a mixture of two components.

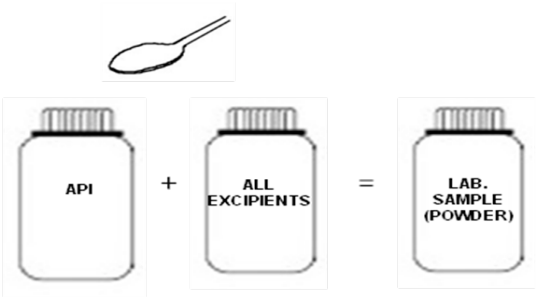


Fig 6. API + Placebo methodology

Five different placebos were prepared, where the quantities of lemon flavor covered ± 5% with respect to the nominal value. For other excipients such as Exc 2, Exc 3 and Exc

4 the quantities remained within the nominal values due to the small amount in the formulation (under 1% w/w).

In the case of sucrose, as major component also no changes in the quantities were performed due to its high composition in the formulation and not being our component of interest, it was decided to avoid any source of variability deriving from this concentration change.

For the API (DKP. Trom), the concentration range covered from 8 mg/g to 12 mg/g (±20% with respect to the nominal value), and in the final mixture the selection of each placebo added to the respective API amount was randomly assigned.

A key point that must be carefully considered in the sample preparation is the value of concentration correlation amongst components in the mixture. The probability of finding two highly correlated components increases as the number of components present in the mixture also increment. For this reason a conscientiously experimental concentration design must be done, mostly if one of the correlated analytes is the one that should be quantified. This high correlation can affect the accuracy and robustness of the calibration model.

2.1.2 Industrial granulates (Industry samples)

A total of 54 industrial samples given by Menarini Laboratories were used also in this study.

These granulated samples were taken over a period of 6 months, and each of which belonging to a different production batch. HPLC chromatography was the reference method mentioned in the analytical industrial reports for determining the concentration of Dexketoprofen in mg/g.

2.2 SPECTRA ACQUISITION

Homogeneity is a paramount factor prior to spectra recording. A lack of this propriety can negatively affect the development of a multivariable calibration model.

In order to overcome this problem, the samples were shacked in a tubular blender for 10 minutes, and before introducing the sample in the quartz cell for the spectra recording, these were hand-shacked once again during 5 minutes.

The samples were considered homogeneous whenever two consecutive spectra are perfectly overlapped.

2.2.1 Instrumentation

The NIR spectra acquisition was carried out using a spectrophotometer FOSS NIRSystems (model 5000) between the spectral range 1100-2500nm and 2nm resolution. Each spectrum was obtained from 32 scans. The instrument was coupled to a Rapid Content Analyzer (RCA), allowing the register of the samples in reflectance mode.

Fig7. Near Infrared spectrophotometer coupled to a RCA module

Previously to the sample measurement, an 80% reflectance reference spectrum was recorded using a ceramic plate coupled to the instrument. Each sample was placed in a quartz cell and recorded three times; between the measurements, the samples were mixed with a plastic spoon, making sure that the spectra contained the information from different particles of the sample. The used spectrum from each sample corresponded to the average of three replicated spectra. The NIR spectra recording for the Industrial samples follow the methodology already described.

2.3 DATA ANALYSIS

2.3.1 Acquisition software

NIR spectrophotometer is controled by the software Vision v2.51 (Foss NIRSystems, silver spring USA). This software allows the user to record and visualize the spectra.

2.3.2 Chemometric spectral treatment

Unscrambler v7.01-9.1(Camo Process SA, Trondheim, Norway) was useful to analyze multivariable data. The program allows the user to work with different chemometrics algorithms towards the development of calibration models. Spectral pretreatments as Standard Normal Variate (SNV), derivates (Savitzky golay) can be performed. This study was mainly focused on the use Principal Components Analysis (PCA) and Partial Least Squares Regression (PLS1). Moreover, after the multivariate calibration is created by PLS, a prediction can be performed in order to obtain the concentration of unknown samples.

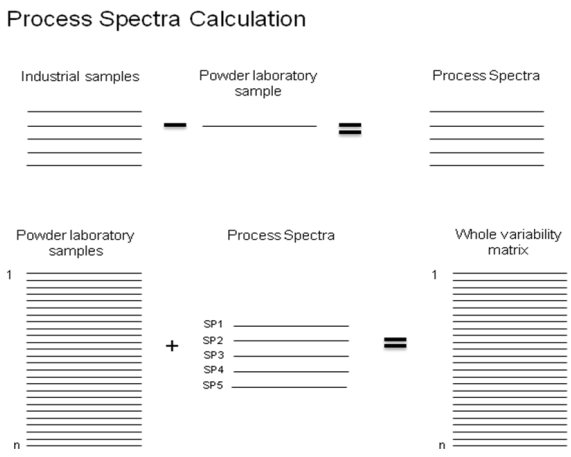
2.4 PROCESS SPECTRA METHOD

This methodology is used in order to construct representative NIR-calibration sets from samples resulting from the production process. It uses a process-variability matrix that incorporates the effects of different production operations (granulation, compaction and coating) on the NIR spectrum. The new matrix thus obtained is used to construct chemometric models that allow all components (API and excipients) to be determined

The first step involved the calculation of the process spectrum (SP), which incorporates process variability into each tablet:

$$\mathbf{SP = S_{ind} - S_{lab}}$$

Where $\mathbf{S_{ind}}$ is the spectrum derived from industrial granulated and $\mathbf{S_{lab}}$ is the spectrum resulted from laboratory powder sample, where both industrial and laboratory samples must have a API concentration close to the nominal value (13). The following scheme summarizes the calculation of process spectra and the creation of the whole variability matrix.



Scheme1. Process spectra calculation methodology

This methodology has been already used, and it shows clearly that the addition of variability with SP allows construction of calibration models for industrial samples based on synthetic samples (13, 14).

3. RESULTS AND DISCUSSION

As it was mentioned above this work is focused in the development of an accurate calibration model based on synthetic laboratory samples in order to predict amounts of API in granulated industrial samples.

The results and relevant discussion are presented below:

3.1 Sample preparation

During the sample preparation step a concentration range of ± 20% was covered, spanning DKP content from 8mg/g to 12 mg/g. This range was used to build up a model able to recognize samples that did not accomplish the quality requirements during the production process.

Five different placebos were prepared by weighting all the excipients of the mixture in an analytical balance. The experimental weight values of each excipient differed from the theoretical values due to errors in the weighted step. The concentration of each excipient varied with respect to the nominal value, although it was not a desired factor taken in account in the experimental design.

These weighting errors did not affect the concentration value for sucrose, which is the major compound in the mixture (98% w/w).

After homogenization of the placebos, a second weighted step was performed with API and placebos. The placebos used for each mixture were distributed randomly, and a total of 10 g of the mixture was prepared.

The following table shows the detailed weighted amounts of each component (g) and the respective proportion of DKP and placebo in the mixture.

# Sample	DKP.Trom Exp (g)	DKP mg/g	Placebo Exp (g)	Placebo mg/g	#Placebo	Total Mix (g)
1	0.1183	8.015	9.882	988.170	1	10.000
2	0.1202	8.143	9.880	987.980	2	10.000
3	0.1222	8.279	9.878	987.780	3	10.000
4	0.1242	8.414	9.876	987.580	4	10.000
5	0.1263	8.557	9.874	987.370	5	10.000
6	0.1284	8.699	9.872	987.160	4	10.000
7	0.1302	8.821	9.870	986.980	2	10.000
8	0.1322	8.957	9.868	986.780	1	10.000
9	0.1343	9.099	9.866	986.570	5	10.000
10	0.1363	9.235	9.863	986.370	3	10.000
11	0.1386	9.390	9.862	986.140	5	10.000
12	0.1405	9.519	9.859	985.950	3	10.000
13	0.1424	9.648	9.857	985.760	2	10.000
14	0.1445	9.790	9.855	985.550	4	10.000
15	0.1465	9.926	9.853	985.350	1	10.000
16	0.1485	10.061	9.851	985.150	3	10.000
17	0.1508	10.216	9.849	984.920	2	10.000
18	0.1529	10.359	9.848	984.711	5	10.001
19	0.1547	10.481	9.845	984.530	1	10.000
20	0.1569	10.630	9.843	984.310	4	10.000
21	0.1588	10.759	9.841	984.120	5	10.000
22	0.1609	10.901	9.839	983.910	4	10.000
23	0.1629	11.037	9.837	983.710	3	10.000
24	0.1649	11.172	9.835	983.510	2	10.000
25	0.1669	11.307	9.833	983.310	1	10.000
26	0.1691	11.457	9.831	983.090	2	10.000
27	0.1712	11.598	9.829	982.881	4	10.000
28	0.1731	11.728	9.827	982.690	3	10.000
29	0.1751	11.863	9.825	982.490	5	10.000
30	0.1772	12.005	9.823	982.280	1	10.000
31	0.1475	10.022	9.824	985.208	5	9.972

Table 6. Experimental weighted amounts of API and excipients, and the respective proportion in the mixture (g)

The analysis with NIR spectroscopy of samples in low API concentration required a special attention in the evaluation of paramount parameters such as the coefficient correlation and Root mean standard error (RMSE), which evaluated the predictive

performance of the model with an independent data set. This both parameters are relevant defining the robustness of the calibration model.
In the case of having high spectral correlation amongst the analytes it is necessary to reduce the concentration correlation in order to avoid that any change in the concentration of any compound affect the accuracy of the model prediction for API.

The following table summarizes detailed the spectral and concentration amongst formulation components

Correlation coefficients		API	Exc1	Exc2	Exc3	Exc4	Exc5
API	concentration	1					
	Spectral	1					
Exc1	concentration	0.040	1				
	Spectral	0.939	1				
Exc2	concentration	0.034	0.215	1			
	Spectral	0.878	0.919	1			
Exc3	concentration	0.017	0.519	0.304	1		
	Spectral	0.969	0.979	0.916	1		
Exc4	concentration	0.001	0.970	0.147	0.422	1	
	Spectral	0.936	0.990	0.950	0.979	1	
Exc5	concentration	0.939	0.301	0.103	0.149	-0.342	1
	Spectral	0.883	0.941	0.778	0.922	0.922	1

Table 5. Spectral and concentration correlation amongst formulation components

It can be clearly seen that there are high values of spectral correlation amongst samples. The reduced contribution of each component to the total sample spectrum demanded a carefully selection of spectral pretreatments and the evaluation of specific wavelength in order to diminish any constraint for the development of the model due to this correlation values.

Also, the table shows that there is a high correlation between the concentration of API and sucrose around 0.94; while the concentration of sucrose decreases the concentration of API increase, and the other way around as well. This fact was expected from the beginnings due to these both components representing the 98% (w/w) of the pharmaceutical formulation.

In the case of high correlation amongst Exc 1 and Exc 4, it was not considered a define factor for the development of the model, due to the low concentration of these components in the formulation (0.06% w/w).

3.2 Spectra Acquisition

The NIR spectra of 31 powder laboratory samples and 54 granulated industrial samples were recorded with NIR spectrophotometer FOSS system. Also the API, lemon flavor and sucrose were measured separately. Their NIR spectra are showed in the following figure.

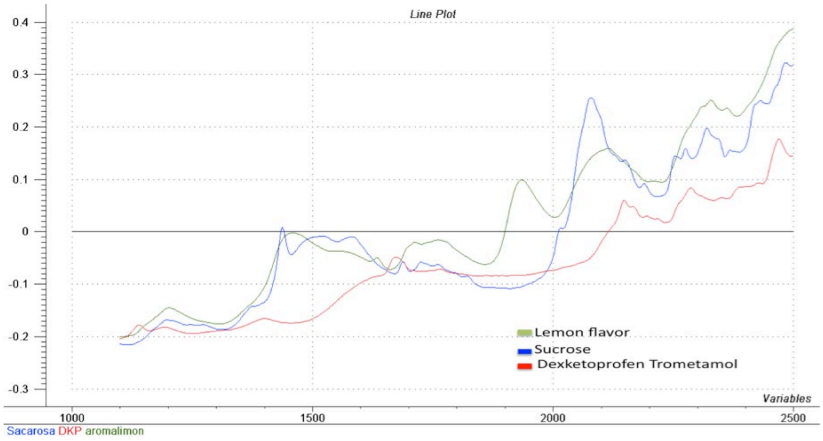


Fig8. NIR spectra from majority components from Dexketoprofen Trometamol formulation

It can be clearly seen that these three compounds present different patterns in the NIR spectrum. However, the API does not show a strong representative band that give a hint of a specific wavelength range that must be used.

After the spectra record, a previous visualization was performed to check any visual outlier in the samples.

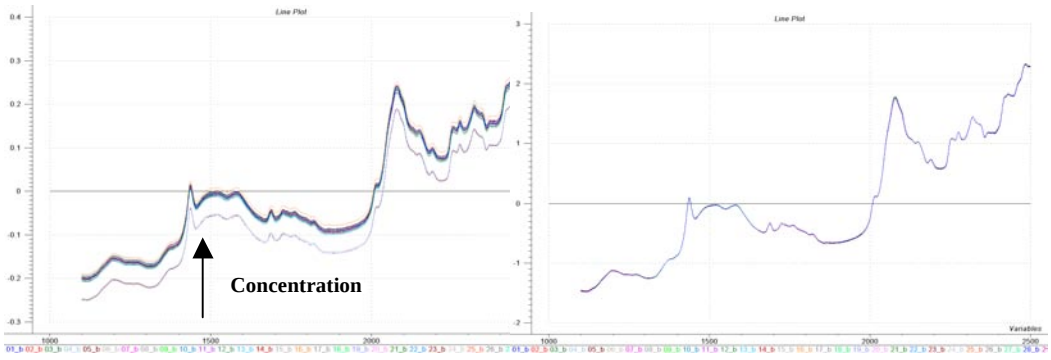


Figure9a. Absorbance NIR spectra from powder laboratory samples.9.b. SNV NIR spectra from powder laboratory samples

The fig 9b.shows that Standard Normal Variate (SNV) spectral pre-treatment allows the correction of the scattering effect by physical properties such as particle size,

subsequently facilitating the visual spectra analysis. Even though the concentration of API and excipients vary amongst samples, this difference is not visually notable. Also it is important to remark that the sucrose spectrum masks the variations from the other compounds.

3.3 Development of calibration model using synthetic samples (powder laboratory samples)

The model was developed with 25 average spectra (from the triplicate of each sample), and it was calculated through partial least squares regression 1 (PLS1) relating the analytical variable (NIR spectra) with the chemical property (concentration).

Some relevant parameters are considered such as: correct spectral pre-treatment, and wavelength range. The determination of the best parameters to develop an accurate calibration model is mainly experimental.

Different characteristics of the model such as Y-explained variance, predictive capacity and residual test are checked in order to determine the best parameters to be used.

Also, it is important to double check once again the spectral correlation amongst the formulation components in order to see any change with chosen spectral pretreatments, considering that these changes can improve the conditions to build up the calibration model.

However, for practical purposes the data below only shows the best combination of parameters. 2 derivative (2D) combined with Standard normal deviation (SNV) was determined as the optimal spectral pretreatment.

	API (DKP.trom)	**Lemon Flavor**	**Sucrose**
API (DKP.trom)	1		
Lemon Flavor	0.027	1	
Sucrose	0.039	0.076	1

Table 8.Spectral correlation in 2D+SNV of API, lemon flavor and sucrose

The table shows clearly that the correlation values have strongly decreased using this combination of 2D+SNV spectral pretreatments. This low correlation facilitates the development of an accurate calibration model. Likewise, these results show that the use of this spectral pretreatment combination is the ideal for these samples.

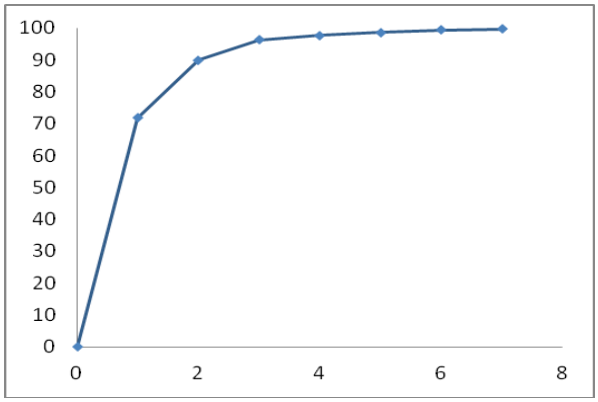
To build up the calibration model, the samples were divided in two sets, one for the calibration and the other one for external prediction in a proportion of 70/30 c.a.

The below detailed table shows the characteristics of the model and the respective predicted values for both powder laboratory samples and industrial granulates.

	CALIBRATION			**PREDICTION**			
					Powder Lab		**Granulated ind.**
Powder Laboratory samples	PLS factors	Y-explained variance	RMSEC (mg/g)	RMSEP (mg/g)	avg Residuals	SD residuals	RMSEP (mg/g)
	3	96	0.24	0.24	0.01	0.257	4.65
	4	98	0.19	0.25	0.08	0.258	4.72
	5	**99**	**0.15**	**0.23**	**0.09**	**0.231**	**4.86**

Table9. ***Characteristics and statistics values for calibration model in the wavelength range 1100-2498 nm and 2 derivative + SNV for the quantification of Dexketoprofen in powder samples.***

It is below also graphically illustrate the Y-explained variance by the different principal components (PCs)



PCs	Y Exp Varian.
1	72
2	90
3	96
4	98
5	99
6	99
7	100

Fig. 10 Y-explained variance graphic through different principal components

As it was mentioned above the values of API concentration obtained using NIR spectroscopy were compared with the theoretical values by partial least square regression. The following figure shows the regression line for both calibration and prediction sets, corresponding of the model above described using 5 factors (99% Y-explained variance) and between the wavelength range 1100-2498nm.

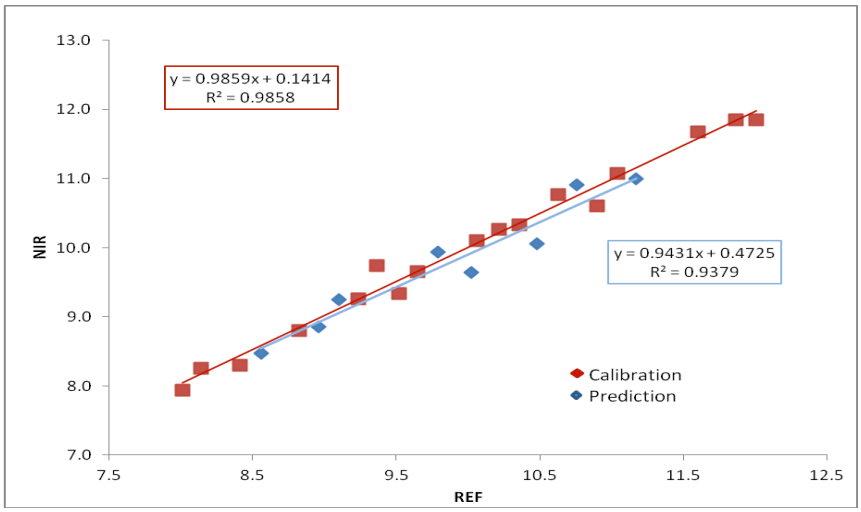


Fig 11.regression line of DKP concentration values obtained with NIR spectroscopy and the theoretical values

As it was mentioned above, just 25 samples were used to build up the model. 6 samples could not be considered since they did not fit into it, either the preparation errors alter their composition or they were not homogeneous and the obtained spectra were different to the others samples, behaving these as outliers.

This detection step is transcendental, because if these samples are not detected and harm the model in the calibration or in the prediction, and even to the point to impede the development of the model. The software Unscrambler 9.8 offers different tools to recognize these kinds of samples.

The predictions for granulated industrial samples were performed to check the predictive capacity of the calibration powder sample set with the industrial granulates. The high RMSEP values show that the powder laboratory samples do not include the process variability leading in low accuracy capacity of this model for this type of samples. For this reason, the process spectra methodology must be used in order to include all this variability and make the model useful for predict API in industrial granulates.

3.4 Calculation of process spectra and development of calibration model for granulated industrial samples through the process spectra methodology

With the purpose of build up a model able to predict industrial granulates, the process variability was included to the powder laboratory samples adding process spectra (SP).

The leading factor to calculate the SP is the selection of the laboratory and industrial samples with the aim of choose those ones that represent and include the most variability. The searched variability in this step of the study regards all the variability included by manufacturing process (In this case, granulation of the powder formulation mixture).

For the first calculation of SP samples with nearest API concentration to the nominal value (10mg/g) were choose.

One synthetic and three industrial samples were selected, with API concentrations of 10.02, 9.98, 9.93 and 9.90 mg/g respectively. The powder laboratory sample was subtracted to the three granulated industrial samples to obtain three resultant spectra that mainly include the information of the process (SP). This SP was subsequently added to the powder laboratory samples in order to obtain a new calibration matrix with the whole variability.

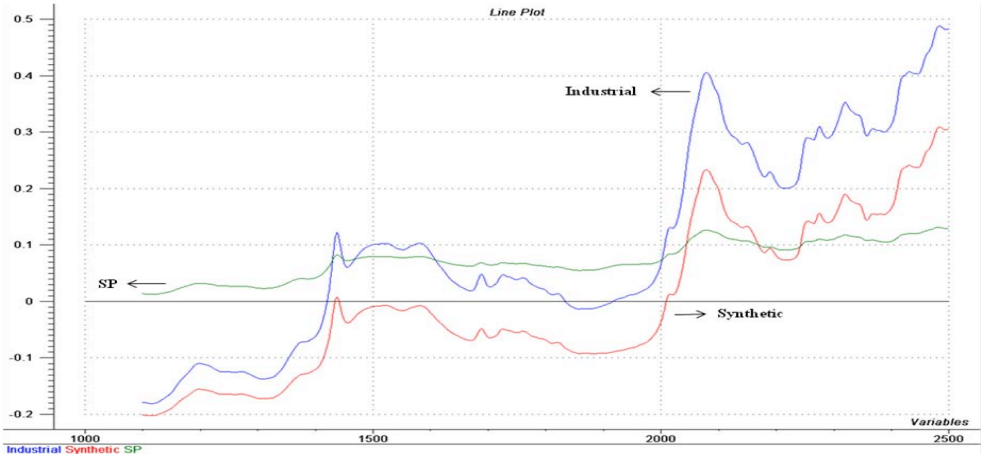


Fig 12. Absorbande spectra for a powder laboratory sample, a granulated industrial sample and a process spectrum

An Analysis of Principal Components (PCA) with this new calibration matrix was performed to project the industrial granulates in order to see if the powder samples + SP

embrace the industrial granulates. In this way it can be check if the selected samples for the SP calculation include the whole process variability.

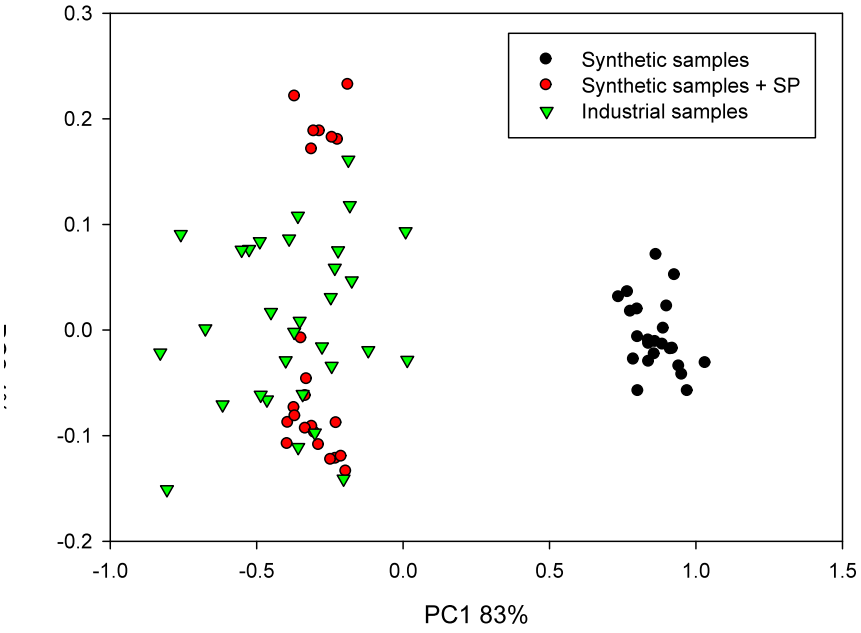


Fig13. Projection Industrial samples in scatter plot of powder samples added process spectra (first calculation).

The scatter plot shows clearly the difference of the powder laboratory samples before and after addition of SP, grouping in different clusters. Also, it is important to note that the samples used to calculate the SP do not contain all the variability of the process and for that reason do not embrace the industrial granulates constraining the development of an accurate calibration model and requiring another SP calculation.

To check this fact a calibration model was develop and a prediction of the industrial granulates performed. The prediction values of the industrial samples corroborate the assumption above describe (RMSEP 1.10 mg/g). Even though the RSEP value was high an improvement of the prediction values is visible compared with the powder laboratory samples model prediction values for the industrial granulates (RMSEP 4.86 mg/g).

A second calculation of SP was performed. For this calculation first a PCA from the industrial granulates was plotted to choose the industrial samples located outmost of this PCA for the subsequent SP calculation.

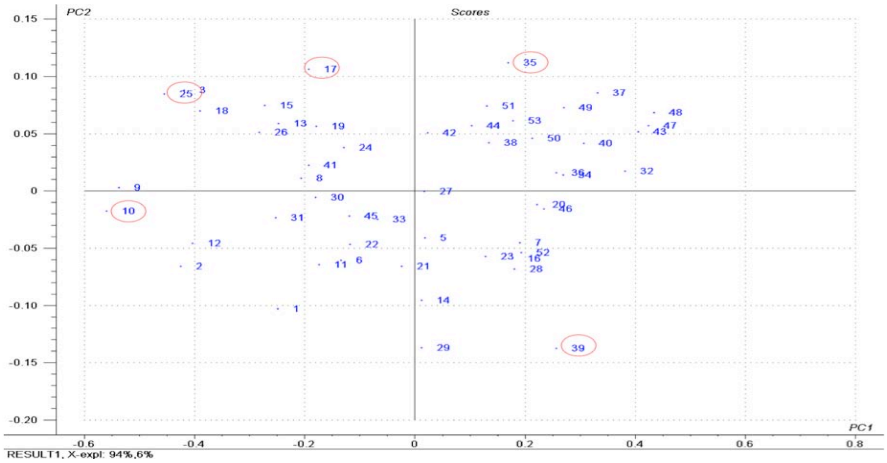


Fig15. Scatter plot of industrial granulates and samples selected for process spectra calculation.

The powder laboratory sample for the calculation was the same used before. After subtraction of the laboratory sample to the industrial granulates the respective five SP were added and a new calibration matrix was obtained. The following projection of the industrial samples in a PCA of the powder laboratory samples + SP shows that with this new calibration matrix most of the industrial samples are encompass.

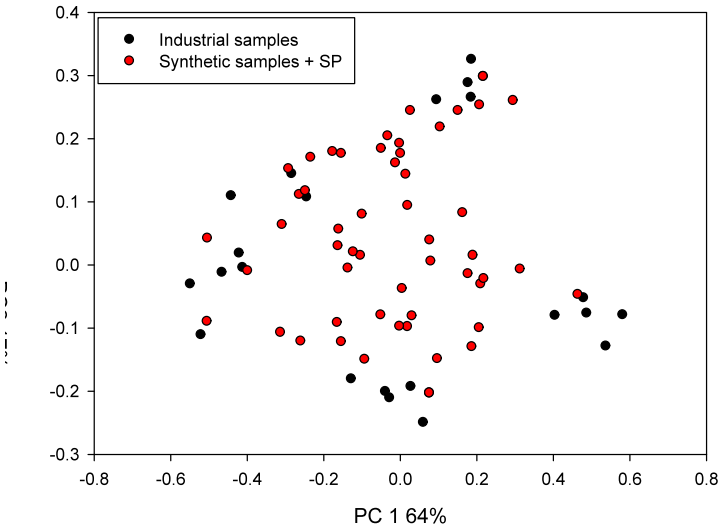


Fig15. Projection Industrial samples in scatter plot of powder samples added process spectra (second calculation).

A calibration model was developed in order to check the predictive ability with this sample set for industrial granulates.

The following table summarizes the characteristics of the developed models in this second SP calculation.

	CALIBRATION			PREDICTION					
				Powder Lab			Granulated ind.		
	PLS factors	Y-explained variance	RMSEC (mg/g)	RMSEP (mg/g)	Residuals Avg	Residuals SD	RMSEP (mg/g)	Residuals Avg	Residuals SD
Powder laboratory samples + SP	4	93	0.31	0.30	0.06	0.302	0.67	0.23	0.62
	5	**97**	**0.21**	**0.18**	**0.04**	**0.183**	**0.76**	**0.30**	**0.71**
	6	98	0.14	0.19	0.08	0.177	0.80	0.34	0.73
	7	99	0.12	0.22	0.07	0.207	0.81	0.33	0.75

Table 10. Characteristics and statistics values for calibration model in the wavelength range (1100-1440)(1630-2498) nm and 2 derivative + SNV for the quantification of Dexketoprofen in granulated industrial samples.

The table shows that with the second SP calculation most variability has included into the new calibration matrix and in this order of ideas, the prediction ability of the developed model for industrial granulates has improved (RMSEP 0.76 mg/g).

The API concentration values obtained using NIR were compared with the theoretical values by partial least square 1. The following figure detailed shows the regression line for both calibration and prediction sets using 5 factors (97% Y- explained variance) and between the wavelength range.

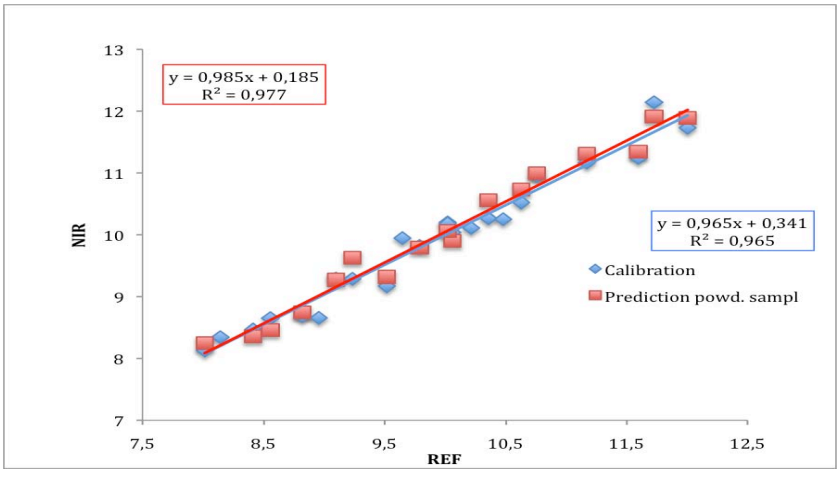


Fig 11.regression line of DKP concentration values obtained with NIR spectroscopy and the theoretical values

Although the RMSEP values for granulated industrial samples still being high, it should take into account that the concentration of API in the formulation is very low, and it hinders the development of the calibration model. Also, as it can be seen that the PCA is useful tool for the selection of the samples involved in the SP calculation.

This study is not complete and the best methodology to obtain the correct process spectra and different parameters to improve this model still under research, with the aim to apply this model into the industry process control.At the moment different spectral pretratments has been studied in order to check any improvement in the prediction ability of the calibration model, but 2D + SNV still being the better combination to treat the data. For the detailed prediction statistics with different pretreatments refer to appendix.

4. CONCLUSIONS

- The use of chemometric tools is essential to obtain relevant information from NIR spectra, and the combination of both methodologies is necessary for the development of calibration models.

- A calibration model for the determination of DKP in powder laboratory samples was successfully developed.

- Process spectra calculation is a useful methodology to add process variability to powder laboratory mixtures and it can be applies to the manufacture of several pharmaceutical products.

- The Analysis of Principal Components (PCA) is useful for the selection of samples involved in the process spectra calculation.

- A calibration model for granulated industrial samples is proposed and different factors must be study for the improvement of its prediction ability.

5. REFERENCES

1. Burns D.A., Ciurczak E.W., *Handbook of near infrared analysis*, Marcel Dekker, United states of America, 1992.

2. Barton II F.E., *Theory and principles of near infrared,* Spectroscopy Europe 14(1), 2002.

3. Reich G., *Near infrared and imaging: basic principles and pharmaceuticals applications*, Advanced drug delivery reviews 57, 1109-1143, 2005.

4. Palou Garcia A., *Estudi de la distribució de components i recobriment en comprimits farmacèutics mitjançant NIR-CI*, Master's Thesis Ciència I Tecnologia Químiques, UAB, Spain, 2011.

5. De Beer T., Burggraeve A., Fonteyne M., Saerens L., Remon J. P., Vervaet C., *Near infrared and raman spectroscopy for the in-process monitoring of pharmaceutical production processes*, International journal of pharmaceutics 417, 32-47, 2011.

6. Skoog, Holler, Nieman, Principios de analisis instrumental, Chapter7, 2001.

7. Blanco M., Villarroya I., *NIR spectroscopy: a rapid-response analytical tool*, Trends in analytical chemistry, 21(4), 240-250, 2002.

8. Peguero Gútierrez Anna., *La espectroscopia NIR en la determinación de propiedades fisicasy composición quimica de intermedios de producción y productos acabados,* Doctoral thesis. Bellaterra, Spain, 2010.

9. Brereton R.G., *Chemometrics, data analysis for the laboratory and chemical plant*, University of Bristol, John Wiley & Sons Ltd, England, 2003.

10. Gemperline P., *Practical guide to chemometrics*, Second Edition, Taylor and Francis, United states, 2006.

11. Da-Wen Sum, *Infrared Spectroscopy for Food Quality Analysis and Control*, Elsevier, United States of America, 2009.

12. Kramer R., *Chemometric techniques for quantitative analysis*, Marcel Dekker Inc, United States, 1998.

13. Alcalá Bernardez M., *Utilización de la espectroscopia NIR en el control analítico de la industria farmacéutica, desarrollos iniciales en PAT*, Doctoral thesis. Bellaterra, Spain, 2006.

14. Blanco M., Peguero A., Analysis of pharmaceuticals by NIR spectroscopy without a references method, Trends in Analytical Chemistry, 29(10), 2010.

15. Rosas J.G., Blanco M., González J. M., Alcalá M., Real-time determination of critical quality attributes using near infrared spectroscopy for process analytical technology (PAT), Talanta, 93(163-170), 2012.

6. APPENDIX

.LABORATORY POWDER + PROCESS SPECTRA				
Spectral Pretratment	**# Factors**	**% Exp y-variance**	**RSEC %**	**RMSEC mg/g**
SNV	8	95		
	9	99		
	10	100		
1D	7	95	2.4	0.24
	8	97	1.92	0.19
	9	98	1.38	0.14
1D + SNV	7	95	2.47	0.25
	8	98	1.56	0.16
	9	99	1.1	0.11
2D	5	95	2.55	0.26
	6	97	1.81	0.18
	7	98	1.46	0.14
2D + SNV	4	93	3.03	0.31
	5	97	2.07	0.21
	6	98	1.43	0.14

Appendix a. Characteristics of the calibration model in different spectral pretreatments

Spectral Pretratment	# Fact.	Laboratory powder + process spectra						Granulates Industrial					
		RSEP %	RMSEP (mg/g)	residual t-test avg res	st DEV	t Crit	t exp	RSEP %	RMSEP (mg/g)	residual t-test avg res	St DEV	t Crit	t exp
1D	7	3.03	0.31	0.07	0.31	2.12	0.27	11.52	1.52	0.58	0.98	2.01	0.72
	8	2.98	0.3	0.05	0.31	2.12	0.19	13.55	1.33	0.93	0.96	2.01	1.18
1D+ SNV	7	3.27	0.33	0.1	0.32	2.12	0.36	13.77	1.38	0.9	1.02	2.01	1.38
	8	3.34	0.34	0.05	0.34	2.12	0.18	14.9	1.46	1.1	0.97	2.01	1.38
2D	5	2.39	0.24	0.06	0.24	2.12	0.3	6.81	0.68	$4x10^{-5}$	0.67	2.01	$7x10^{-5}$
	6	2.09	0.21	-0.05	0.21	2.12	0.28	8.54	0.84	0.38	0.75	2.01	0.61
	7	2.13	0.21	0.07	0.21	2.12	0.41	8.65	0.85	0.38	0.76	2.01	0.61
2D + SNV	4	2.94	0.3	0.06	0.3	2.12	0.23	6.63	0.67	0.23	0.62	2.01	0.4
	5	1.81	0.18	0.04	0.18	2.12	0.28	7.8	0.76	0.3	0.71	2.01	0.52
	6	1.9	0.19	0.07	0.18	2.12	0.47	8.2	0.8	0.34	0.73	2.01	0.56

Appendix B. Statistics for prediction values of the model based on synthetic samples + SP for industrial granulates.