Universitat Autònoma de Barcelona

Development of new analytical methods using Near Infrared and Chemometrics for pharmaceutical quality control: Enhancement of modelling strategies towards a better product understanding

Vanessa Cárdenas Espitia

Doctoral Thesis

Doctoral Studies in Chemistry

Supervisors: Marcelo Blanco Romía

Manel Alcalà Bernàrdez

Departament de Química

Facultat de Ciències

2015

Thesis submitted to aspire for the Doctor Degree

Vanessa Cárdenas Espitia

Supervisor's Approval:

Prof. Marcelo Blanco Romía **Dr. Manel Alcalà Bernàrdez**

Bellaterra (Cerdanyola del Vallès), 12th November 2015

CONTENTS

ACKNOWLEDGEMENTS

This thesis has been developed with the assistance of the following institutions:

Universitat Autònoma de Barcelona

Pre-doctoral fellowship PIF

Ministerio de Economía y Competitividad (MINECO)

Project CTQ2012-34392 "Productos y procesos mediante técnicas espectroscópicas moleculares combinadas con tratamiento de datos multivariable"

Applied Chemometrics Research Group

Departament de química, Universitat Autónoma de Barcelona

Bellatera, Spain

Laboratorios Menarini S.A

Badalona, Spain

Büchi Labortechnik AG

Flawil, Switzerland

Massó Analitica S.A

El Prat de Llobregat, Spain

ABBREVIATIONS

NIR	Near Infrared Spectroscopy
AOTF	Acousto-Optical Tunable Filters
API	Active Pharmaceutical Ingredient
CCD	Charged Coupled Devices
CLS	Classical Least Square
CPAC	Center of Process Analysis and Control
CPP	Critical Process Parameters
CQA	Critical Quality Attributes
EMA	European Medicine Agency
EMSC	Extended Multiplicative Scatter Correction
FDA	Food and Drug Administration
FIR	Far-Infrared
FT	Fourier Transform
GMP	Good Manufacturing Practices
HPLC	High Performance Liquid Chromatography
ICH	International Conference of Harmonization
LVF	Linear Variable Filters
MEMS	Micro electromechanical Systems
MIR	Mid-Infrared
MLR	Multiple Linear Regression
MCR-ALS	Multivariate Curve Resolution-Alternating Least Squares
MSC	Multiplicative Scatter Correction
PAC	Process Analytical Chemistry
PAT	Process Analytical Technologies
PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares
PRM	Pattern Recognition Methods
QA	Quality Assurance
QbD	Quality by Design
QC	Quality Control
RCA	Rapid Content Analyzer
RMSE	Root Mean Square Error
RSE	Relative Standard Error
RTRT	Real Time Release Testing
S.G	Savitzky-Golay
SNV	Standard Normal Variate
Sp	Process Spectrum
UV-Vis	Ultraviolet-Visible

ABSTRACT

Accomplish high quality in the pharmaceutical industry is a constant challenge that require strict control and supervision not only from final products but also from all manufacturing steps according to process analytic technologies (PAT) initiative.

In recent years, the simplicity and expeditiousness of near infrared spectroscopy (NIRS) together with chemometrics data analysis have substantially fostered its use for the determination of pharmaceutical´s physical and chemical properties.

The subject of this thesis is the development of NIR methodologies useful for the quality control for the pharmaceutical industry, proposing and optimizing mainly new strategies for the selection and design of the calibration set and improving NIR methodologies based on product knowledge and process understanding.
The thesis consists in three cases of study where different solid pharmaceutical formulations were evaluated. In the first study a new methodology for constructing the calibration set is proposed and its suitability for quantifying an API present in low concentration (10 $mg.g^{-1}$) was evaluated by comparison with two methodologies broadly known. The main results showed the efficiency and suitability of the calculation ad addition of process spectra methodology for the quality control of a pharmaceutical granulate (final product). This methodology showed an outstanding performance in terms of robustness and operational simplicity in comparison with the other evaluated strategies.

The second study consists in the optimization of a novel methodology through the incorporation of statistical parameters for the selection of important factors used, and three calibration models were calculated for the quality control of a formulation in its three manufacturing steps: powder, cores and coated tablets. Based on the results obtained with this study the model space concept was established defined by Hotelling's $T^2$and Q-residuals statistics for outlier identification – inside/outside the defined space – in order to select objectively the factors to be used in the calibration set construction. Also the efficacy of the

proposed methodology for stepwise pharmaceutical quality control was confirmed. This work represented a contribution for the field as a guideline for the implementation of this easy and fast methodology in the pharma industry.

A third study case addresses a concern in the pharmaceutical industry regarding samples with tendency of segregation and its analysis by using NIR spectroscopy. In this study the effective scanned area and the sample representativeness is evaluated, and an effective alternative based on spectra acquisition on moving samples is proposed. The results obtained in this work confirmed the influence of the scanned area with the representativeness of the analyzed sample and this, in turn, the performance of the calculated calibration models. This study highlights the importance of the optimization of the surface scanned area since the quality of the methodologies to be developed depends strongly on this factor. This work presents an effective alternative for quality control on samples with tendency to segregate.

The proposed methods were evaluated according to the European and international guidelines and represent a contribution to the PAT initiative and the development of NIR methodologies based on its improvement by a better product knowledge and process understanding.

RESUMEN

Obtener una alta calidad en la industria farmacéutica es un reto constante que requiere un estricto control y supervisión de los productos manufacturados. La tecnología analítica de procesos (PAT, *Process Analytical Technology*) propone que esto puede lograrse de una manera óptima y sistematizada mediante el control de calidad en diferentes etapas de manufactura (materias primas, intermedios y producto acabado).

La simplicidad y rapidez de la espectroscopía de infrarrojo cercano (NIRS, *Near Infrared Spectroscopy*) junto con el análisis quimiométrico de datos multivariables ha demostrado su eficacia para la determinación de propiedades físicas y composición química de productos farmacéuticos, y su uso en esta industria se ha incrementado considerablemente en los últimos años.

El objetivo de esta Tesis es el desarrollo de métodos NIR para el control de calidad en la industria farmacéutica. Para ello, se propusieron y optimizaron nuevas estrategias de diseño y selección del conjunto de calibración y la mejora de las estrategias de modelado NIR, basado en conceptos como el conocimiento del producto y comprensión del proceso de acuerdo con la filosofía de trabajo PAT.

La Tesis consiste en tres casos de estudio de diferentes formulaciones farmacéuticas sólidas. En el primer estudio se evalúan diferentes metodologías para la construcción del conjunto de calibración, y también se propone un nuevo modo de cálculo y adición del espectro. Se evaluó la idoneidad de dichas estrategias para la cuantificación de un principio activo (API, *active pharmaceutical ingredient*) presente a baja concentración (10 miligramos por gramo). Los principales resultados mostraron la eficacia de esta nueva metodología para el control de calidad de un granulado farmacéutico (producto acabado), y demostró una sobresaliente capacidad en términos de robustez y simplicidad experimental en comparación con otras estrategias evaluadas.

El segundo estudio consiste en la optimización del proceso de selección de factores relevantes en la implementación de una metodología novedosa para la preparación

del conjunto de calibración. Para ello se calcularon modelos para el control de calidad de una formulación en sus tres etapas de fabricación: polvo, núcleos y comprimidos recubiertos. De los resultados obtenidos en este estudio se confirmó la idoneidad del concepto de espacio del modelo definido por los estadísticos T^2 Hotelling y residuales Q para la identificación de muestra anómalas (dentro/fuera del espacio), y para la selección objetiva de los factores a utilizar en la construcción conjunto de calibración. También se corroboró la eficacia del método NIR propuesto para el control de calidad en variadas etapas del proceso.

El tercer caso de estudio aborda una problemática en la industria farmacéutica que concierne al análisis NIR de muestras sólidas con tendencia a la segregación de sus componentes. En este estudio se evalúa el área efectiva de escaneado espectral y la representatividad de la porción de muestra analizada. También se propone una alternativa eficaz de análisis por NIR para este tipo de muestras basada en la adquisición de espectros de muestras en movimiento. Los resultados obtenidos en este trabajo confirman la influencia del área escaneada con la representatividad de la muestra analizada y esto, a su vez, a la capacidad predictiva de los modelos de calibración. Este estudio resalta la importancia de la optimización del área de escaneado de superficie previo cálculo del modelo, ya que se confirmó como estos factores influyen en los estadísticos de predicción de los métodos NIR propuestos. Este trabajo presenta una alternativa eficaz para el control de calidad de muestras heterogéneas o con tendencia a segregación.

Los métodos propuestos en cada uno de los casos de estudios fueron evaluados de acuerdo con las directrices internacionales y europeas, y representan una contribución a la iniciativa PAT y el desarrollo de metodologías NIR en base a su mejora mediante un mejor conocimiento del producto y una mayor comprensión del proceso.

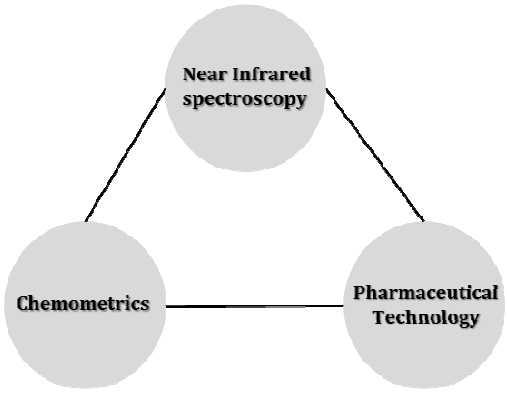
PREFACE

The following thesis has been developed in the Applied Chemometrics Research Group at the Universitat Autònoma de Barcelona in partnership with the pharmaceutical industrial sector (Laboratorios Menarini SA) during the academic years 2013-2015.

From a broad perspective, the general motivation of this thesis was the development of new methodologies useful for the pharmaceuticals control quality, and the optimization/implementation of multivariate calibration strategies and statistics parameters for molecular spectroscopy modelling enhancement.

All the approaches presented in this thesis were developed within a work framework of Process Analytical Technologies guidelines (PAT), which is defined as a system for design, analyzing, and controlling the process through monitoring each manufacture step. In general, the studies presented in this thesis show alternatives for improving productivity by implementing new scientific knowledge to industrial processes based on product knowledge and process understanding.

The scope of the work presented in the following sections are of interest to different sectors such as analytical chemists, pharmaceutical scientists and process engineers, since this is an interdisciplinary project which involves near infrared spectroscopy, chemometrics data analysis and pharmaceutical technology.



This thesis is divided in five main parts, one introduction which contains the general basics of the used disciplines, one section of objectives which showed in detail the aim of the work performed, followed by three study cases in which

different industrial problematics were solved using NIR spectroscopy-chemometrics data analysis.

Three solid pharmaceutical formulations were evaluated for the study cases as final and intermediate products, therefore the suitability of the proposed methodologies was broadly evaluate in different products.

The work presented in this thesis was communicated in different conferences and seminars, and also three articles were written (two published and one in preparation), representing a relevant contribution for both industrial and academic sectors.

INTRODUCTION

1. The Pharmaceutical Industry

1. THE PHARMACEUTICAL INDUSTRY

1.1 Overview

The target of the pharmaceutical industry is the development of pharmaceutical formulations capable of cure diseases and improves people's life. In general these substances contain one or more active pharmaceutical ingredients (API), which are molecules that act against a health malfunction or a pathogenic entity. The pharmaceutical formulations also contain another substances named excipients, which are inactive but provide important properties as stabilization, improvement of odor/flavor, drug delivery, pharmaceutical dissolution, among others[1], [2]

A pharmaceutical product can be freely launched for consumption after the strict evaluation of three dimensions -constituted by critical factors- that strongly intervene in drug performance. These dimensions can be summarized in three concepts: *safety*, *medical utility* and *industrialization* (**Fig 1**)

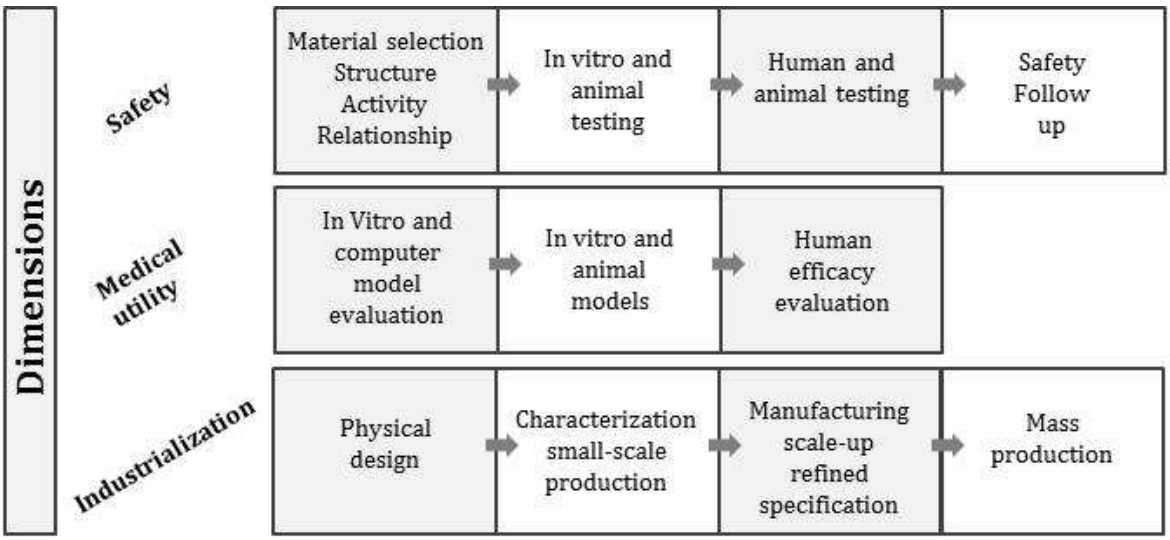


Fig1. Three dimensions of pharmaceuticals development

Since pharmaceuticals affect human's body and body functions, the health risks are an unavoidable component when the consumers go under pharmacological therapies. Therefore safety and medical utility must be ensuring before any further

use of the pharmaceutical product.

These two dimensions are very important and highly correlate in which all the concerning regarding molecular interaction, biological activity, toxicity and optimal dose must be completely evaluate and under control.

In the industrialization dimension aspects as optimization of the drug obtainment, technology transfer and method validation takes place. A careful planning of the stepwise manufacturing and plant design are required to assure that the chemical process works in an industrial scale. Moreover, the process must be carefully tuned, and a quality control system must be established to accomplish the requirements demanded by the regulatory agencies to ensure customer safety and high quality.

For obtaining successfully pharmaceutical formulations, these three dimensions must be carefully evaluated and work perfectly together towards the accomplishment of regulatory agencies requirements; in this way the three of them could converge in the generation of a product with high quality using a optimized system in terms of expenses and use of resources.

Since the scope of these thesis is focused in the pharma process, especially and its quality control methodologies, the further content is focus in the industrial dimension.

1.2 The manufacturing process

The pharmaceutical products can be manufactured in different physical forms according with their proposed action and targets in the human body. However, tablets and capsules comprise the 80% of the products. A typical flowsheet of their manufacture is shown in the **Fig 2** [2], [3]**.** This process can be summarized in several units operations which start by obtaining the pure compounds of the pharmaceutical formulation –API and excipients-, following for a sequence of blending, milling granulating, and ending by turning the mixture into structured products and packing.

Nevertheless these operations are also used for the manufacture of other pharmaceutical products including aerosols, injectables, suspensions and topic products (creams, pomades i.a).

Considering this fact, the pharmaceutical manufacturing process can be consider in a big picture as a "powder technology", in which under a variety of manners is dedicate basically to make particles, modify their properties and turning them into solid structured products [3].

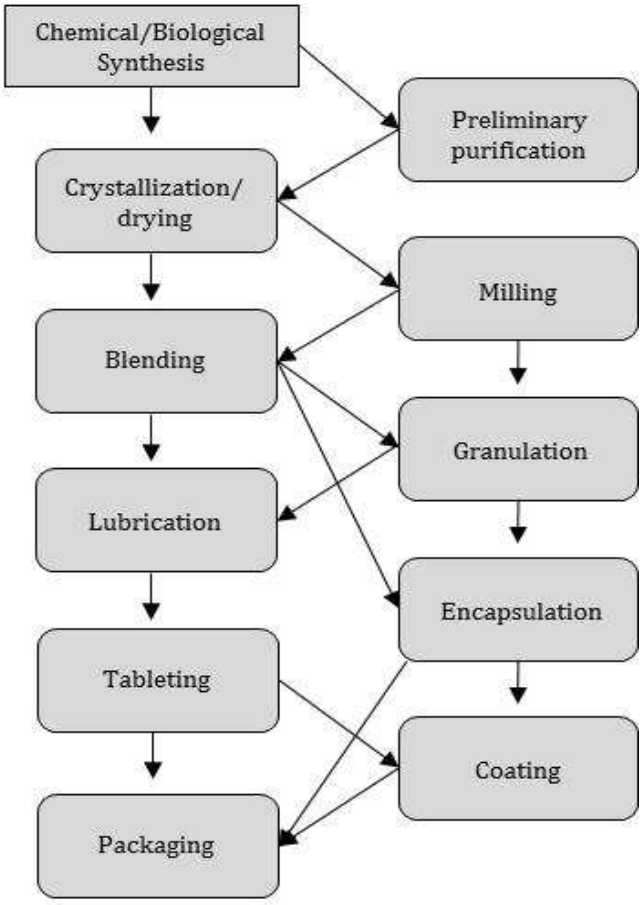


Fig2. Typical flowsheet for tablet and capsule manufacturing

Each individual manufacture step in the process is called a unit operation, and the pharmaceutical's manufacture requires many of them to deliver the final product.

There are three principles methods for the production of a tablet: ***direct compression, wet granulation*** and ***dry granulation***. The manufactures choose the way in which the tablet must be produced based on the characteristics of the formulation's compounds [4].

Tablets can be made by 1) blending the API and excipients and 2) compacting and giving the form of a tablet to the blended mixture, this process is known as ***direct compression***. This method can be chosen when all the components of the formulation after blending show proper homogeneity –uniform distribution of all the components in a mixture- and good flowability -the characteristic of a powder to flow, related to the sliding scale of its constituents-.

It must be point out that the obtained tablets must have *optimal hardness*, *optimal friability* and *fast dissolution* [2], [4]. The *hardness* is related to the breaking point and structural integrity of a tablet under storage, transportation and handling before usage; while the *friability* is considered the condition of being crumbled and to be reduced to smaller pieces with little effort. Furthermore the pharmacokinetics of the pharmaceutical product and its ability to be released has to do with the speed of *dissolution*. These three characteristics influence strongly the product, therefore ensuing that tablets meet the optimal point of each of them is essential for obtaining high quality.
Otherwise, if the components of the formulation do not compress well, do not have good flow ability, are too fluffy, or segregate (heterogeneity) after blending a granulation step is required.
One needs to have in mind that the factors mentioned above affect the quality of the tablets and also can occasion problems that not only affect the manufacture but also the machinery for further use, if the components get stacked in the drums.

In ***Granulation*** the small particles are grouped into larger ones called granules. Each granule should contain a proper mixture of all the components of the formulation and those attached to each other either by mechanical forces -that densify and compacts all the powders together- or by the use of an aqueous

solution binder [2], [4], [5].

As a summary the reasons to granulate are mentioned bellow:

- To improve powder flow
- To improve compressibility
- To reduce fines
- To control the tendency of powders to segregate
- To control density
- To capture and fuse small quantities of active material

The granulation methods can be divided into two major types: *dry granulation* or *wet granulation* [6].

- **Dry Granulation**

This type of granulation must be carried out when the pharmaceutical compounds are sensitive to moisture and/or unable to withstand elevated temperature during drying. This method is also called *slugging*, *chilsonating* or *roller compaction* and it is perform through the reground of the mixed powders into a precise powder. This action increases particle density, improves powder flow and captures fines.
The normal procedure starts by powder compaction using a tablet press followed by a *milling* procedure. In order to densify the powders, the manufacturers usually use a low shear producing fine particles.

- **Wet Granulation**

Along this method several unit operations are perform in order to get the granule. Firstly a *pre-mix stage* is performed where API and excipients are blended prior addition of the binder substance. Once uniformity is achieved, during the *wet massing step* the binder is added to the mixture and the components are massed to a predetermine point. Then a *drying* procedure is followed until full elimination of the moisture and tested by a "LOD" or loss on drying test. The obtained granulation is then *milled* to reduce the size of any caked material into a standardized particle size distribution through a *sieving* step*.* In the *final blend,* post-granulation

excipients are eventually added to improve the properties of the final granule mixture.

Even though wet granulation is the most common processing method used in pharmaceutical manufacturing, it is an expensive, time/energy-consuming procedure in comparison with direct compression or dry granulation. Moreover since many unit operations take place in order to obtain granulates, it is indispensable to consider the complex interaction of several variables in each step, and a deep knowledge of the formulation it is required to control the granulate formation.

As it was mentioned, the pharmaceutical compounds go through several unit operations for the modification of their properties towards obtaining the final product. Each operation is very important and influences strongly the characteristics of the products and its quality. These procedures are described briefly below.

Milling is done when it is necessary to reduce particle size from mixtures, or there is need for de-lump them- in case they form lumps as a result of material's moisture-. It takes place in the three types of methods: *Dry compression* and *dry and wet granulation*. Basically the reasons to milled pharmaceutical compounds are to reduce segregation, improve flow, enhance drying and limit wide particle size distribution. The machinery used to mill the pharmaceutical mixtures is choose based on the desired particle size, and these are categorized according to the force they impart on the powders – shear force-.

Blending is one of the first steps in the pharmaceutical manufacturing, but is also used along the whole process. At the beginning of the manufacture, once the raw materials are rigorously checked, the API and the excipients –all in powder form- go under a mixing procedure in order to get a homogeneous distribution of all the components in the drug. During manufacture granulated powders or extruded pellets need to be also mixed to ensure homogeneity. Once the products are blended these are discharge into a drum, emptied into a hopper of a press or

encapsulator, and divided into the final dosage form. Inadequate mixing somewhere along the production sequence can result in rejection of finished product due to poor quality [3].

Blend studies determine that there is an optimum endpoint for each mixture, and every pharmaceutical formulation have a unique pathway to their optimum state of uniformity which is affected by factors as particle size, density between components and based on those, a mixing times can be predetermine.

Time is a very important factor because if the mixture is either under blended or over blended problems like segregation, weight and hard variations are likely to appear [7].

Regarding the machinery to be used, there are a variety of equipment for mixing materials, which operate in different forms: *mobile blenders*, in which the bin have different forms and rotate about an specific axis, and the *fixed blenders* that are equipped with spiral bands, propellers or inside paddles, who are responsible for shearing the mixture, breaking the possible agglomerates, and drag some of it from one part of the container to the another in order to homogenize the whole mixture.

However, the most common blenders used for final blending are the v blender, the double cone blender and the tote blender. All of these use low shear tumble blending as the most effective way to achieve good mixing with a variety of powders and granules[7].

Once the powders are properly mixed and have all the desire characteristics –good flowability and uniform content- they go to the latest unit operations in order to create the tablet. It should be pointed out that the prior operations influence strongly quality parameters of the tablet and those are not merely subject to the tablet formation.

Tablet compression is the process in which a small quantity of the powder pharmaceutical mixture is pressed to obtain a define form. Technical innovations to tablet compression machinery have improved production rates to the point where more than 500.000 tablets per hour are obtained.

This unit operation consists in four steps: *filling*, *metering*, *compression* and *ejection* [8].

The *filling* step involves the transfer of the blended/granulated material into position to the compressing station. Then, during *metering* the excess of material is removed and the exact weight (volume) of the granulation is tuned to be compressed into tablets. The control of the weight of the mixture is controlled by the lower punch in the mold.

As the name indicates, during the *compression* step the mixture is compress to form a tablet; for this, the upper and lower punches in a predetermine pressure get together within the mold. The distances between the punches determine the thickness and the hardness of the tablet, and these two characteristics are very important in the quality of the product.
The final step involves *ejection* in which the tablet is removing from the lower punch-mold station, and the tablet is ready to go through the coating unit operation.

Tablet coating is important for several reasons, like making the tablet stronger and tougher, improving of taste, addition of color and makes them easy to transport and pack. Also coatings concede specific characteristics to tablets, which is the case of *sustained released coatings* in which the tablet is released slower and steadier into the stomach while having the advantage of being taken at less frequent intervals than immediate release formulation of the same drug.
The coating used could be a thick sugar based layer or a very thin film, but most pharmaceutical tablets are cover with the second option.

The coating process is perform by spreading a solution-which contains the coating substance and the solvent- to the tablet; for several years the manufactures used alcohol as the solvent making the dry process easier. The use of such a solvent shows problems in handling operator safety, solvent recovery and the odor of the tablet. Therefore the improvements in tablet coating equipment have evolved to the use of water and the drying system with constant flow of hot air.

The final goal is obtaining high quality tablets with the following attributes:

- Good weight, thickness and hardness control
- Good ejection
- No capping, lamination or sticking
- Good friability, disintegration and dissolution.
-

It is important to remark that obtaining successfully pharmaceutical products with good quality it is only possible if the right materials, the appropriate set of unit operations and an optimal quality system is chosen.
To sum up, the manufacture process in the pharmaceutical industry basically blend the formulation compounds, and carried out several physical transformations through its unit operations with the aim of obtaining homogenous unit doses that meet all the quality requirements defined in the guidelines.

1.3 Quality management in the pharmaceutical industry: quality assurance and quality control

Since obtaining pharmaceuticals with high quality it is not only an issue that involves industrial productivity but also public health, manufactures, regulatory authorities and governments work together to ensure that pharmaceutical products meet acceptably all the standards of quality, safety and efficacy[9].

The quality management in the pharmaceutical industry is usually defines as the function that conducts, determines and implements the quality policy, in which two concepts are crucial: *quality assurance* (**QA**) and *quality control* (**QC**).
QA is a "wide ranging concept covering all matters that individually or collectively influence the quality of a product; It is totally the arrangements made with the object of ensuring that pharmaceuticals are of the quality required for in their intended use"[10].
According with this definition, **QA** could be consider as the planning, monitoring and tuning system that brings together control quality with the good practice of

manufacture (GMP's) and other guidelines of the regulatory authorities towards one goal: *accomplish quality* [11].

By definition **QC** is a procedure or a set of procedures intended to ensure that a manufactured product accomplish a define set of quality criteria, and specially concerns sampling, specifications and testing; all these together with organization, documentation and release procedures, ensure that the necessary and relevant tests are carried out, and the materials and the products are not release for use or sale until their quality has been judged to be satisfactory [10].

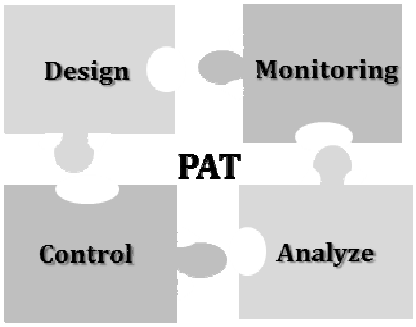
Along the pharmaceuticals manufacture physical and chemical characteristics of the product are evaluated, starting by the identification of raw materials, followed by a batch control within the process, and ending up by tests that confirm quality achievement of the final product.

Even though in the regulations there is any mandatory request for certain analytical methodologies to carry out the **QC** in this industry, manufactures and regulatory agencies coordinate together to choose the appropriate methods based on their characteristics and the quality of the generated data; most methodologies used in pharmaceutical control has shown a good performance and suitability for the purpose over the time, being ultraviolet spectroscopy UV, chromatographic techniques and Karl Fisher titration broadly used.

The currently used methodologies allow analysis *off line* or *at line* which means that a portion of the sample must be withdrawn from the bulk and take it to the an external analysis station which could be located besides production or transported to an extra laboratory.

Even though these procedures has been allowed manufacturers accomplish high quality in their products until now, some constraints regarding sampling and time consuming are more notable since the last years. However the pharmaceutical industry is quite hesitating in the implementation of new technologies due to the economic risk that these could represent, but new guidelines has been launch by the regulatory agencies suggesting and encouraging manufactures to include new scientific knowledge in process technology to improve productivity and optimize the use of the resources [12].

1.4 Process analytical technology (PAT) in pharmaceutical industry



With the purpose of improve manufacture in the pharmaceutical sector, in 2004 the FDA release an initiative which promotes a work philosophy for the innovation of pharmaceutical development, manufacturing and quality assurance.

In this guideline the concept of quality change substantially, which is no longer addressed as a set of requirements to be achieved, and instead, quality must be build up in each manufacture step through the deep understanding of products and processes. Moreover, the guideline encourages the pharmaceutical sector to embrace and implement innovation from the scientific and engineering field, with the goal of enhance understanding and control of manufacturing [13].

PAT is define in the initiative as a "system for designing, analyzing and controlling manufacturing through timely measurements (i.e during processing) of critical quality and performance attributes of raw and in-process materials, with the goal of ensuring final product quality"[13].

By its own definition, this philosophy suggest that products with high quality can be obtain as the result of an strict control and supervision of parameters that influence quality in each manufacture step; in this sense if quality can be ensure in each unit operation, the unique outcome would be products with high quality [12]. Moreover, this working methodology also allows spotting any problem within the process and allows immediate corrections, which can spare resources and money in case of failing.

For all these to become real there is a need for the development and implementation of analytical methodologies suitable for the task, which allow online, simple and fast analysis [14].

On the other hand although PAT is a novel work philosophy for the pharmaceutical industry its origins date back to the 70´s; this has been used since then for chemical and petrochemical industries under the name of process analytical chemistry (PAC) or process analytics (PA). PAC by definition is the application of "on-field analysis" and chemometrics for monitoring chemical or physical attributes or detection of events that cannot be derived from conventional variables (temperature, pH, pressure, flow i.a.). While PAC is more related to real-time analysis for the solution to production problems, PAT is a broader field encompassing a set of tools and principles to enhance manufacturing, process understanding and control which includes several areas such as process analysis, chemical engineering, chemometrics i.a. So, it can be considered that PAC has evolved in the pharmaceutical field of application into what is known nowadays as PAT [14].
As an additional data of interest, it must be mentioned that the Center of Process Analysis and Control (CPAC) was established in 1984 at the University of Washington and still operating nowadays. This works as a consortium of industrial, national laboratory and government agency sponsor which addresses multidisciplinary challenges in PAT and process control through fundamental and directed academic research [15].

1.5 Regulations in the pharmaceutical industry

Since pharmaceutical products affect strongly consumers health and any failure represent a potential public health risk, the pharmaceutical production cycle is strictly regulate to ensure that all drugs are properly tested and produced, and that the test results available to regulatory authorities are complete and unbiased.

As a public health concern, pharmaceuticals manufacture was an issue that each country addresses in the past with its own regulation, but with the expansion of the markets and economy globalization, the national regulatory agencies turn out to constitute international organisms that control the pharmaceutical production under the same notion of quality, assurance and efficacy [11], [16].

The harmonization of regulatory requirements seems to be feasible in 1980s, when the European community moved towards the development of "international products" and moved to a single market concept. Around 10 years later a project known as *international conference of harmonization ICH* reunited together the regulation authorities and experts from Europe, US and Japan in order to discuss scientific and technical aspects of pharmaceuticals [17].

The ICH aim is to provide guidance to harmonize the interpretation and application of technical guidelines and requirements for pharmaceutical research, development and manufacture; all these optimizing the expenses, avoiding duplication of testing without compromising safety and effectiveness.
The ICH guidelines are divided into four groups: **Q** from quality, **S** from safety, **E** from efficacy and **M** from multidisciplinary.

The **Q** guidelines encompass the quality area recommending how to conduct stability studies, defining relevant thresholds for impurities testing and a more flexible approach to pharmaceutical quality based on Good Manufacturing Practice (GMP) risk management. Moreover this part also conducts the harmonization of the three pharmacopeias (European, US and Japan) through a tripartite pharmacopeial harmonisation program known as the Pharmacopoeial Discussion Group (PDG)[17] . A summary of all the categories of the guidelines is condensed in **Table 1** [11].

Table1. ICH guidelines categories and main topics

Q: Quality topics	**S: Safety topics**
Related to chemical and pharmaceutical quality assurance	
1. Stability	1. in vitro and in vivo pre-clinical studies
2. Analytical validation	2. Carcinogenicity studies
3. Impurities	3. Genotoxicity studies
4. Pharmacopoeias	4. Toxicokinetics and Pharmacokinetics
5. Quality of biotechnological products	5. Toxicity testing
6.Specifications	6.Reproductive toxicology
7.Good manufacturing practice	7.Biotechnological products
8. Pharmaceutical development	8. Pharmacology studies
9. Risk management	9. Immuno-toxicology studies
	10.Joint safety/Efficacy (Multidisciplinary topic)
E: Efficiency topics	**M: Multidisciplinary topics**
Related to clinical studies in human subject	Several topics that do not fit in other categories
1. Clinical safety	M1. Medical terminology (MedDRA)
2. Clinical study reports	M2. Electronic standards for transmission of regulatory information (ESTRI)
3. Dose-response studies	M3. Timing of preclinical studies in relation to clinical trials
4. Ethnic factors	M4.The common technical document (CDT)
5. Good clinical practice	M5. Data elements and standards for drug dictionaries
6.Clinical trials	
7.Guidelines for clinical evaluation by therapeutic category	
8.Clinical evaluation	

However it is important to point out that there are other important regulatory agencies such as the *European Medicine Agency* (**EMA**), the *Food and Drugs Administration* (**FDA**), or the medicine and health agencies of each country. All these publish useful GMP´s and manufacture guidelines regarding qualitative and quantitative drug's requirements and mandatory tests to assure quality, nevertheless all these documents are based on the regulations of the international organisms.

1.6 Analytical technologies in real time

As it was mentioned before, in 2004 the FDA published a final guidance for industry introducing the concept of PAT and redefining pharmaceutical manufacturing and quality assurance for the future.
The basic notion of this work philosophy aim to enhance product and processes through the analysis, design, understanding and monitoring each manufacturing step; in that sense the concept of quality changes to some extent to design and build quality into product and manufacturing process, rather than "testing for quality".

The guideline also addresses the concept known as *real time release* which is defined as "the ability to evaluated and ensure the acceptable quality criteria of in-process and/or final product based on data". This concept of PAT includes according to the guidance, "a valid combination of assessed material attributes and process controls," and based on the 1985 guidance on parametric release, which is used primarily in heat based sterilization of drugs.
Some years later -2009- the parties of the ICH adopted *the ICH Q8(R2) pharmaceutical development*, which used the term *real time release testing* (RTRT). The definition of this term in the ICH Q8(R2) changed the emphasis from the decision to release a batch to the measurements themselves, as follows: "the ability to evaluate and ensure the quality of in-process and/or final product base on data process, which typically include a valid combination or measured material attributes and process controls"

The implementation of RTRT represents big benefits to the industry in terms of economy, productivity and time. Despite of this, the industry still trying to work out the practicalities of implementing the approach for on-line and in line analysis, and therefore is not broadly applied for all the pharmaceutical manufacturers.
Many questions remain about which instrumentation must be use and when or where on the production line the test must be conducted, how to evaluate on-or in-line analyzers during manufacture and what regulatory authorities expect.

In another hand spectroscopic technologies have been broadly use in the las years for process analysis and increasingly for on-line process monitoring in different type of industries: chemicals, food, agriculture and pharmaceutical i.a.
Advantages as being simple, fast, and non-invasive techniques have probed their suitability for quality control in the pharmaceutical industry within the PAT framework.
Some years ago the PAT guideline was published as an advisable methodology to conduct quality control in the pharmaceutical industry, nowadays regulatory agencies required an organized approach based on risk management through the lifecycle of a pharmaceutical product [18], turning PAT as the more suitable approach to implemented.
The ICH Q8(R2) clearly show the importance of monitoring: 1) critical process parameters (CPP) whose variability impact 2) critical quality attributes (CQA) and c)quality target product profile through a careful design of experiments.

To sum up, enough evidence has been presented by several sectors –academy and regulatory agencies- of the efficiency and suitability of the implementation of new on-line/in-line methodologies for pharmaceuticals manufacture, also as the advantages of designing and creating quality in each process step.
The use and implementation of PAT by the industries is widely increasing, as a result the academic sector is in charge to develop and improve all methodologies to make this transition as easy and profitable as possible.
The most recent Appendix 15 of the EUGMPs [18] for the validation of pharmaceutical processes strongly requires the use of PAT and QbD (quality by design) knowledge to provide the highest confidence of the manufactured products. Moreover, this framework allows the chance to manufacture in real time release mode, consequently, efficient instrumental techniques combined with multivariate data analysis are crucial for this end.

2. Near Infrared Spectroscopy

2 NEAR INFRARED SPECTROSCOPY (NIR)

2.1 Overview

The first observations of the presence of light in the near infrared region data back to 1800`s by Sir *William Hershel* by his studies about the heating effect in the spectrum of solar radiation. In his experiments he used a prim to disperse the light into different colors, and their temperature was measured. Surprisingly he found out that beyond the red, at end of the spectrum the temperature appeared at its greatest, and from that point onwards this part of the spectrum got a little interest for the chemist back then [19]–[23].

The lack of knowledge about optical characteristics of the infrared light delayed its use until 1881, when *Abney* and *Festing* recorded the first spectra of organic liquids. This work was significantly important because for the first time a formal NIR measurement was performed, and also the earliest spectra interpretations; also the importance of the hydrogen bond in the NIR spectrum was reported by them [19], [20].

In the earlies 1900`s *Coblentz* build up one of the first spectrometers which was vibrational-thermal sensitive. Although an hour was needed to obtain a spectrum with this instrument, the experiments performed with it allowed *Coblentz* to discover that each compound has a NIR finger print in the spectrum and that the spectra of two compounds are different one form the other. Moreover, he also noticed the spectral patterns in compounds with similar functional groups, especially a remarkable band in those that contain the OH- group [19], [20].

Around 1950´s a growing demand for fast and quantitative methods to determine moisture and protein in food boost the use of NIR -especially by *Norris* at the US department of agriculture-. And by 1970 already 50 works were published showing the performance of the technique in analysis of atmospheric humidity, water, gelatin i.a [19], [21]–[23].

Although the discovery of NIR was long ago the importance and use of the technique was delayed by the state of the knowledge and the technology back then. Moreover the spectroscopist from that time pointed out constraints for the use of

the technique such as absence of relevant structure information, lack of sharp peaks, and the difficulty of making band assignments in comparison with Mid-Infrared. However technological advances, developments in the instrumentation and new algorithms-that allow statistical and mathematical analysis- turned these constraints into strengths, positioning in this moment the NIR spectroscopy as a versatile, fast, simple and robust analytical methodology in different research and industrial fields [19], [21].

2.2 Principles of radiation

The infrared (IR) region is comprised between 780 and 10^6 nm (12800-10 cm^{-1}) between the ultraviolet-visible (UV-Vis) and microwaves in the electromagnetic spectrum. This region corresponds mainly to molecular rotations and vibrations, because IR radiation is not energetic enough to cause electronic transitions, which is the case for UV-Vis or X-rays.

According the mechanism of the incident radiation and the characteristics of its interactions with the matter, the IR is, in turn, constituted by three different zones: Far Infrared (FIR, 4.10^4 -10^6 nm), Mid Infrared (MIR, 2500-4.10^4 nm) and Near Infrared (NIR, 800-2500 nm). Thus, in the FIR zone the absorption bands correspond to molecular rotations, while in MIR to molecular vibrations. Overtones and combination bands of the lower energy fundamental molecular vibrations are the ones observed NIR; these bands are significantly weaker in absorption comparing to the fundamental bands from which they originated, and can be considered as faint echoes of Mid-IR absorption [24], [25].

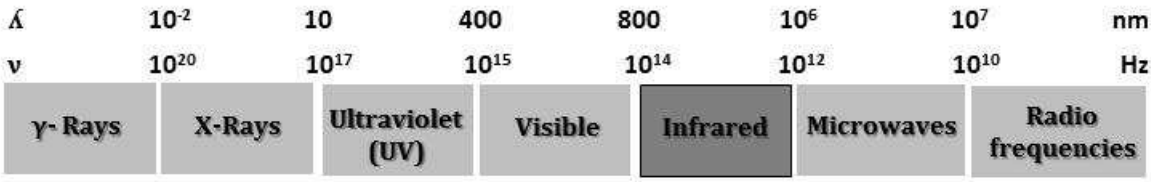


Fig1. Electromagnetic spectrum

A molecule can only show infrared absorptions when the molecules experience a change in the dipole moment during the vibration or rotation. Just under these circumstances, the electrical field of the radiation can interact with the molecule, and provoke changes in the amplitude of its movements, resulting in the radiation absorption [26].

Since the dipole moment is determine by the magnitude of the charge difference and the distance of the charge centers, homonuclear species as (H_2, O_2, N_2) do not absorb in the NIR, because its dipolar moment is not alter during vibrations and

rotations [25].

The interactions of infrared radiation with matter may be understood in terms of changes in molecular dipoles associated with vibrations; vibrations can involve either a change in bond length (stretching) or bond angle (bending). However even if the distance amongst atoms is always affected, not all vibrations modes are active in the IR, and it depends strongly on the structure of the molecule itself.[20], [25].

A molecule can be looked upon as a system of masses –atoms-joined by bonds with spring like properties, which can be explained by the harmonic oscillator model, obeying to Hooke's law. When the masses are at rest in the equilibrium position, its potential energy is zero, whereas when the spring is compressed or extended its potential energy increases. The movement of atoms in the molecule is confined in a potential well, characterized by atoms attraction and repulsion. The energy levels of the atoms confined in the potential are quantized [24].

The energetic levels in this model are equidistant distributed, and transitions are only allow it between neighbors levels ($\Delta n = \pm 1$); this is known as the selection rule **Fig2** [20], [21], [23].

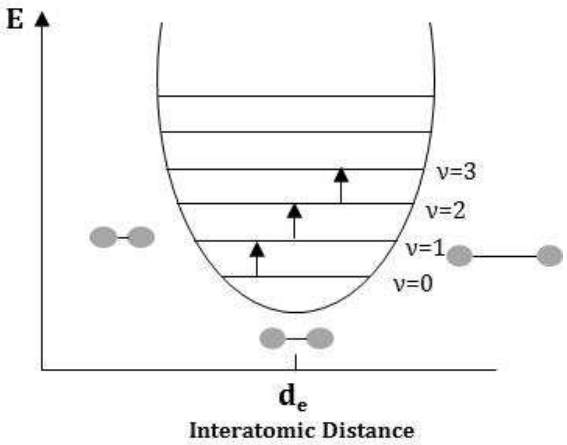


Fig2. Schematic representation of the harmonic model

The potential energy of the bond is defined by a simple harmonic oscillator, given by the following expression [24], [25], [27]:

$$\mathrm{E} = \frac{1}{2}\,\mathrm{k}\,\mathrm{x}^2 \tag{1.1}$$

Where *k* is the bond force constant and *x* the distance amongst atoms

An elastic bond, like a spring, has a certain intrinsic vibrational frequency, dependent on the mass on the system and the force constant. Classically it is simple to show that the oscillation frequency is[24], [25], [27]:

$$v = \sqrt{\frac{1}{2\pi}\left(\frac{k(m_1 + m_2)}{m_1 m_2}\right)} \tag{1.2}$$

Where *v* is the frequency, *k* the force constant and *m1-m2* the masses of each atom.

Vibrational energies like all other molecular energies are quantized, and classic mechanics failed to fully describe the behavior of atoms and molecules. Therefore the allowed vibrational energies for any particular system may be calculated from de Schrödinger equation. For the simple harmonic oscillator[24], [25], [27]:

$$\mathrm{E}_{\mathrm{vib}} = \left(n + \frac{1}{2}\right)\mathrm{hv} \tag{1.3}$$

Where *n* is the vibrational quantum number (0, 1, 2…), *h* the Planck constant and *v* the vibrational frequency.

Combining the equations (1.2) and (1.3) we can describe the expression that better represent the vibrational energy for a molecule taking into account their vibrations and quantized energy[27]:

$$\mathrm{E}_{\mathrm{vib}} = \left(n + \frac{1}{2}\right)\frac{\mathrm{h}}{2\pi}\sqrt{\frac{\mathrm{k}(\mathrm{m}_1 + m_2)}{\mathrm{m}_1 m_2}} \tag{1.4}$$

The harmonic oscillator model cannot explain the behavior of real molecules, as it does not take into account Columbic repulsions between atoms or dissociation of bonds, which affects the potential energy of the molecule. As a result, the behavior of molecules resembles more closely to the model of an anharmonic oscillator **Fig 3** [20], [21], [23]–[25], [27].

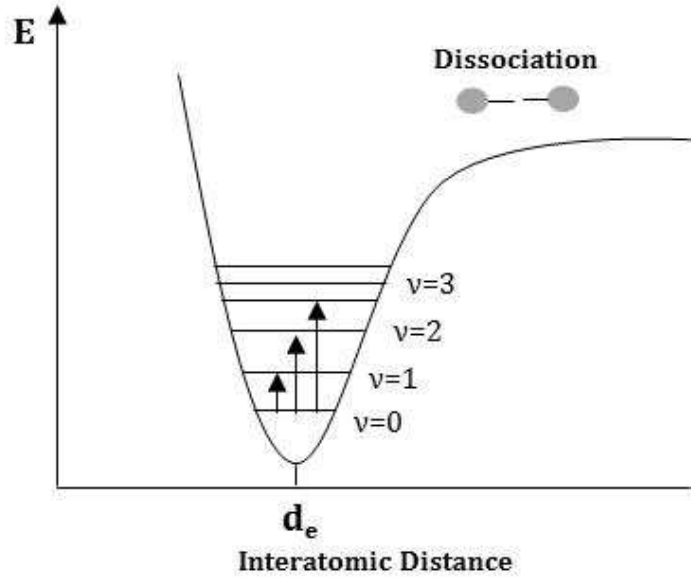


Fig3. Schematic representation of the anharmonic model

Energy curves of harmonic and anharmonic oscillators are very similar to low levels, so that the molecules have only a harmonic behavior around the equilibrium position, but the energy difference decreases with the increasing of the vibrational quantum number *n*.

Therefore the expression of vibrational energy (eq 1.3) can be correct in terms of the anharmocinity of the molecules by the following expression [27]:

$$E_{vib} = \left(n+\frac{1}{2}\right)hv - \left(n+\frac{1}{2}\right)^2 hvy - \left(n+\frac{1}{2}\right)^3 hvyy'... \qquad (1.5)$$

Where *n* is the vibrational quantum number (0, 1, 2…), *h* the Planck constant and *v* the vibrational frequency. And y, y` are anharmonicity constants. As *n* increases the terms can be ignored.

The energetic levels for the anharmonic model are not equidistant distributed; thus at higher energetic levels, the energy between levels is lower. The allowed energetic transitions are not only observed for the fundamental band ($\Delta n = \pm 1$), but for another transitions ($\Delta n = \pm 2, \pm 3, ...$) which correspond to overtones bands that show up at greater wavelengths in the NIR region. The overtones appear between 780 and 2000 nm, depending on the overtone order and the bond nature and strength. Since this transitions are less frequent than MIR fundamental vibrations, its bands intensity decreases between 10- 100 times for the first overtone (depending of the particular bond) [20], [21], [24].

For polyatomic molecules the simultaneous changes in the energy of two or more vibrations modes are named combination bands, the frequencies of which are the sums of multiples of each interacting frequency. Combination bands appearing between 1900-2500 nm [21] [21].

Non fundamental transitions are much less likely than transitions between consecutive energetic levels, so the NIR bands are less intense than those in the IR region. In addition, NIR bands are wider and less well defined as a result of overlap overtones and combination bands. However, the intensity of NIR bands depend on the dipole moment and the anharmonicity of the bond, thus O–H, N–H, C–H, S–H bonds are strong NIR absorbers and present strong bands; By contrast bands like C=O, C-C, C-Cl, C-F are much weaker or even absent [21], [22].

Atomic interactions between molecules (for instance hydrogen bonding or dipole interactions) alter vibrational energy states, thereby shifting existing absorption bands and giving rise to new ones, through difference in their crystal structure. This allow crystal forms to be distinguished one from another and physical properties (such as viscosity, particle size, particle size amongst others) to be determined [21].

2.3 Instrumentation

The NIR instrumentation has evolved dramatically over the years, and the development of its devices has responded the demand of its use for analysis of different types of samples and diverse environments. NIR spectrophotometers have a huge advantage over other analytical methodologies because they have a broad variety of devices which can be adapted to several samples-from gels to grains- and that allow the analysis in different conditions and scenarios. All this together have fostered the use of NIR spectroscopy, as a versatile and flexible technique [19], [21], [28].

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 **Fig4**[20]–[22]**.**

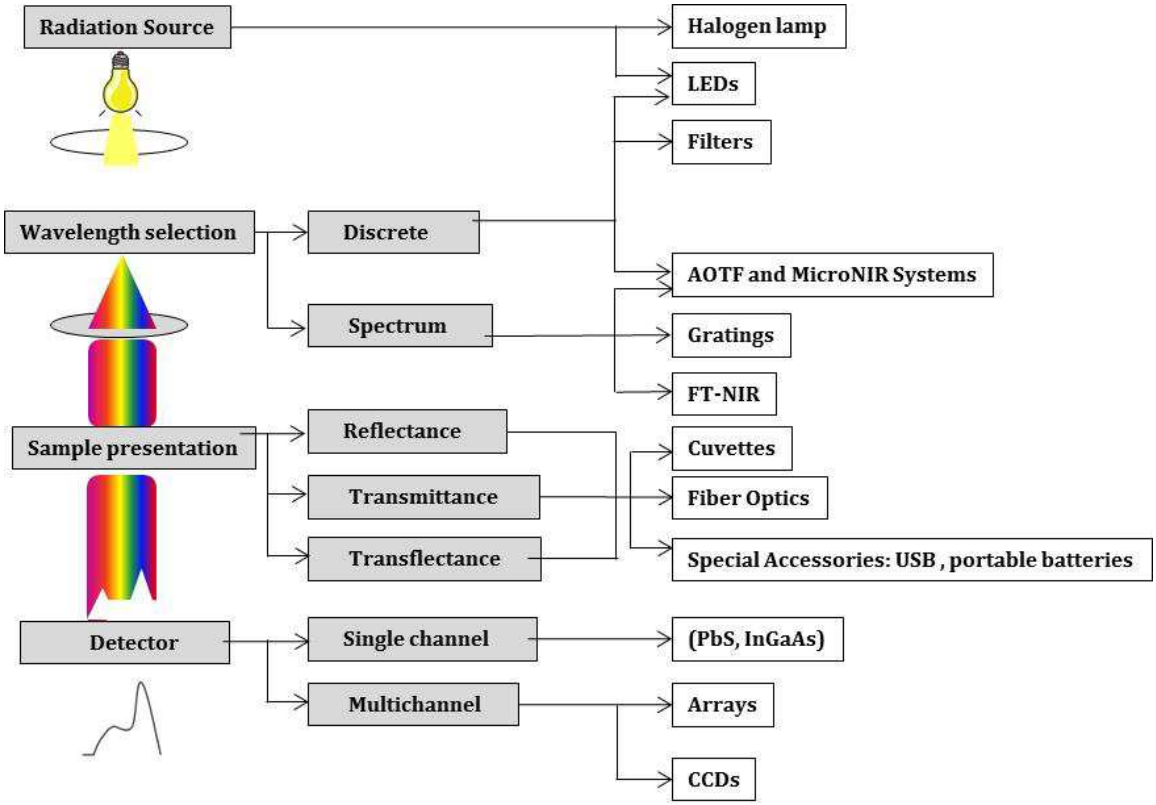


Fig4. Main characteristics of a NIR instrument

- **Radiation source:** The light source generates a beam that can irradiate the samples. The 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. Another light source that can be used is named LEDs (Light Emitting Diodes), that depending of their composition is able to emit up to 1600 nm. The halogen lamps require wavelength selection system, while LEDs does not [21], [28].

- **Wavelength selection system:** It is very important to have an appropriate system for wavelength selection, because the sensitivity of the instrument depends very much of it; a narrow width band increases the sensitivity of the measurements [25].

NIR spectrophotometers can be grouped in two types with respect to wavelength selection***: discrete wavelength*** and ***whole spectrum instruments*** [21].

The *discrete wavelength instruments* are simpler than the others, since they irradiate samples with only few wavelengths; therefore they are useful for those applications that just required analysis in a specific spectral range.

These instruments can selected the wavelengths by using lights sources filter that allow the passage of variably broad wavelengths or light-emitting diodes (LEDs) that emit narrow bands. The no need of moving parts makes LED-spectrophotometers simple and robust, encouraging its use for use in portable equipment [20]–[22].

The instruments based *on Acousto-Optical Tunable Filters (AOTF)* also belong to the discrete wavelength instruments. These devices exploit the properties of the birefringent materials -usually crystals of TeO_2 cut it in a special angle- that have the ability to change its refractive index when is crossed by an acoustic wave. In this way, one of the wavelengths of the incident polychromatic light is diffracted by the material and directed towards the sample, while the remainder pass through the TeO_2crystal, which is transparent to NIR radiation [20]–[22].

The acoustic waves are obtained by transformation of radiofrequency signals by piezoelectric transducers, so that the selection of wavelengths is performed by modulating the initial signal radio. This means that the wavelength scanning is very fast and allows the selection specific wavelengths required for each analysis **Fig5** [29].

In addition, the instruments based on AOTF do not require moving parts, so it is a very robust technology and high repeatability, especially suitable for work in harsh conditions, such as industrial plants.

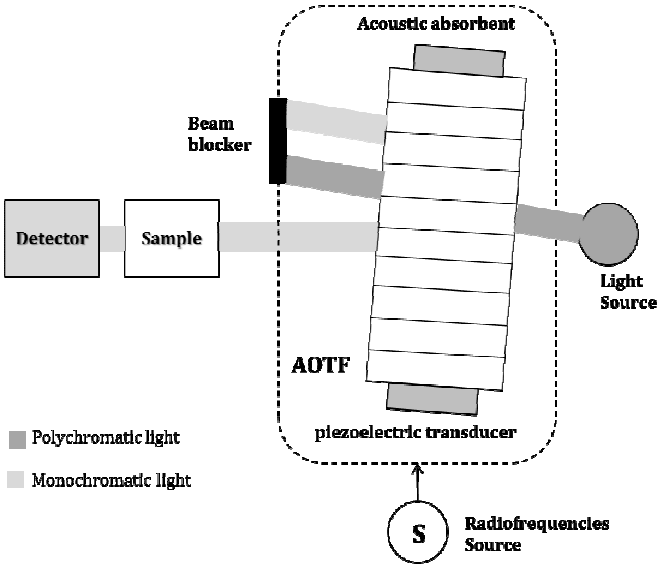


Fig5. Scheme of instruments based on Acousto-Optical Tunable Filters (AOTF)

Whole spectrum instruments are much more flexible than discrete wavelength instruments, so they can be used in wider situations. Usually include ***diffraction grating instruments*** or be of the ***Fourier transform (FT)-NIR type***[21].

The *diffraction grating instruments* use monochromators as wavelength selection systems which are constituted by a set of collimators -that narrows the input beam to an output aligned in a specific direction- together with a dispersing element. This is the fundamental part of the system, allowing the decomposition of the incident beam as a result of constructive and destructive interactions. The most

commonly used dispersants elements are diffraction gratings **Fig6**[20], [21], [28].

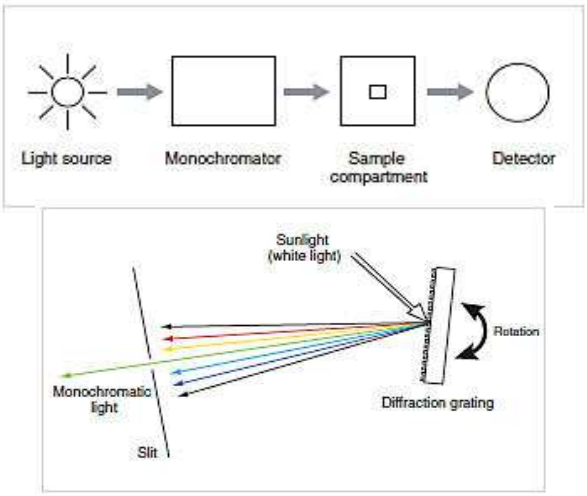


Fig6. Scheme of instruments based on diffraction grating instruments

The instruments based on *Fourier transform (FT)* are based on the division of the incident light into two de-phased beams subsequently recombined. The most commonly used device for the light division and phase shift of the beams is the Michelson interferometer, which is constituted by a beam splitter, a fixed mirror and a movable mirror. The beam splitter divides the polychromatic radiation into two beams, one of which is reflected by the fixed mirror and the other on the movable mirror. The change in the position of the movable mirror causes that both beams are out of phase due to the difference of optical paths and when they recombined again interfere constructively or destructively. The signal obtained is called interferogram and is a representation of signal strength versus time. With the implementation of the FT algorithm, the time domain becomes the frequency, and in this way the NIR spectrum is obtained. FT-NIR instruments are undoubtedly the instruments combining most of the best characteristics in terms of wavelength precision and accuracy, high signal to noise ratio; however they are sensitive to vibrations and slower than AOTF based instruments [20], [21], [28], [29]**Fig7**.

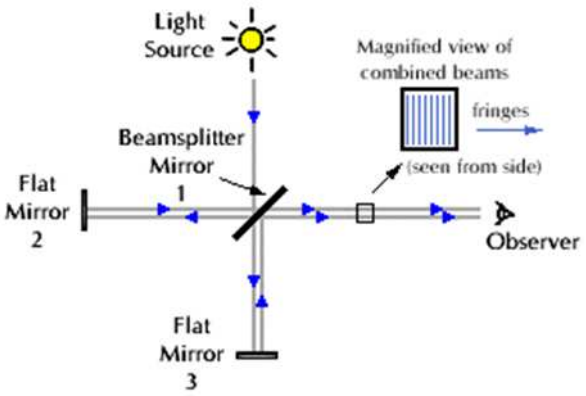


Fig7. Scheme of instruments based on Fourier transform-Michelson interferometer

Micro NIRs are the most modern NIR instruments in the market since they are integrated systems practically with no moving parts build it in very small size – comparable to a sugar cube in some cases-. Their exceptional benefits rely on the adaptation of the NIR optics from traditional instruments to small scale devices, using low cost materials as silicon. Also their innovative design which allows using a simple USB cable as power source and output system, or the use of portable batteries increase the versatility of the technique and enhance its suitability for "on the field" applications.

One of the examples of this successful instrument miniaturization is the *Micro electromechanical Systems* (*MEMs) based on scanning grating spectrometer*, with an integrated InGaAs diode for detection in the near infrared spectrum. The suitability of scanning grating spectrometers for different applications are broadly known, with numerous descriptions and advancements published. Microelectromechanical systems (MEMS) technology is capable of producing and assembling scanning gratings. Combining the two, a scanning grating spectrometer can be fabricated, which benefits from the energy and cost efficiency inherent to MEMS components [30] **Fig8**.

Also the use of Linear Variable Filters (LVF) has shown its successful implementation in the fabrication of micro NIRS. LVF consist in a wedged filter on top of a linear array of photodetectors and enables the transfer of the optical spectrum into a lateral light intensity profile over the array of photodetectors [31] **FIG 8**.

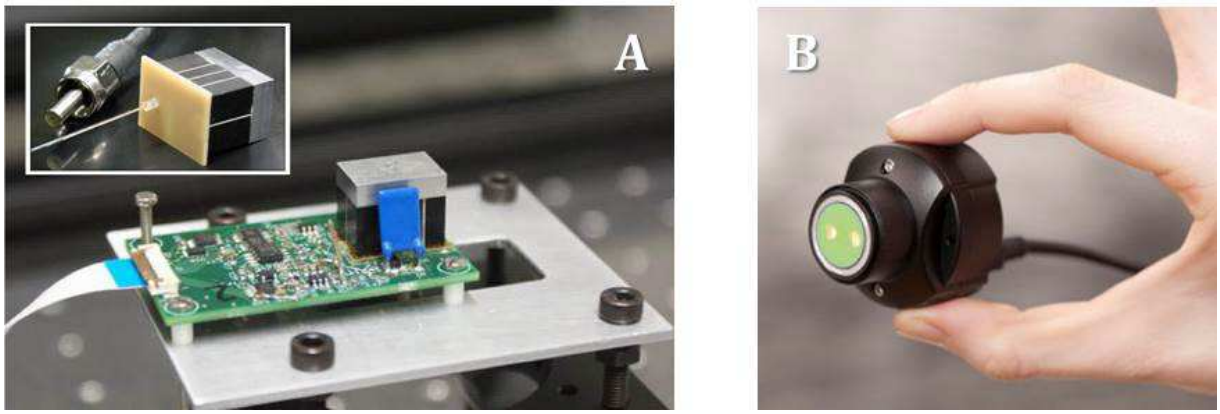


Fig8. Fully integrated micro NIR system featuring A) MEMs system and B) LFV system

- **Detectors:** The detection systems in NIR spectroscopy use devices with semiconductors such as PbS or Ingaes, 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 [21], [22], [28].

2.4 Sample Acquisition Modes

One remarkably advantage of this technique is the versatility of NIR analysis for different kind of samples –gas, liquids, solids and even mix phase materials- without need of sample pretreatment; this has fostered its use for both academia and industrial purposes.

The options of sample acquisition depend strongly on the nature of the sample and the environment of the analysis. There are three modes to take a NIR spectrum (reflectance, transmittance and transflectance) and the main difference is the position of the sample, the detector and the optical path length [20], [22], [26], [32]**Fig 9.**

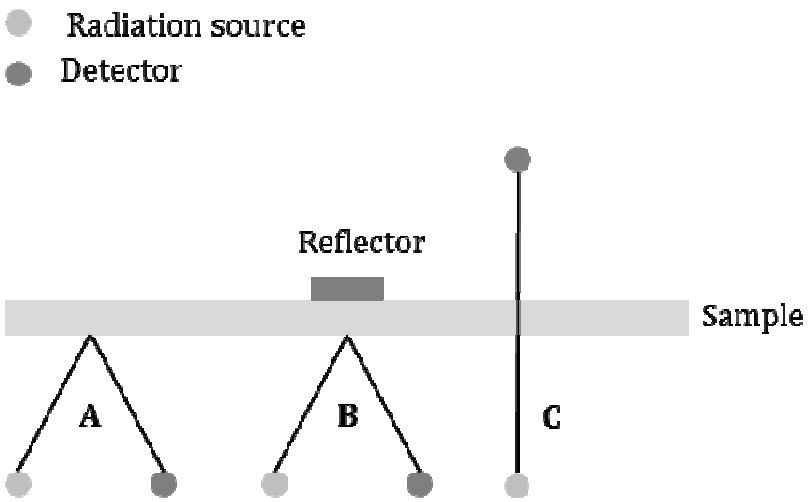


Fig9. Scheme of acquisition sample modes in NIR A) Reflectance, B) Transflectance and c) Transmittance

In the *Transmittance mode* is measured the fraction of transmitted radiation (not absorbed) by the sample, ie, the radiation reaching the detector after passing[22] through the sample[22], [26]. This measurement mode obeys the Beer-Lambert law listed in the following expression[27]:

$$A = \log\frac{1}{T} = \log\frac{P_0}{P} = abc \qquad (1.6)$$

Where A is the absorbance of the sample, T the transmittance, P_0 the intensity of the incident radiation, P the intensity of the transmitted radiation, a the molar absorptivity, b the optical path and c concentration.

Some deviations of this law can occur due to different causes such as: reflection and/or scattering of radiation, very high concentrations, and chemical interactions of the sample, amongst others.

In the *Transmittance mode* liquid or semi-liquid samples are measured. In this mode the sample is placed in a tray together with a one side reflective surface. The beam of the incident radiation enters the transparent face of the tray, passes through the sample and is reflected on the other side of the tray. Thus, it returns back through the sample and the radiation is captured by the detector. The

resulting optical path is double the length between both tray surfaces [22], [26].

The *Reflectance mode* measures the radiation reflected from a surface, where the incident radiation is focused onto the sample and two forms of reflectance can occur, specular and diffuse. The first one is governed by the laws of Fresnel, and it is reflected at the same angle of incidence without penetrating the sample, which contains no information on it. The contribution of the specular reflectance (as noise only) often played with an appropriate relative position of the detectors and sample. On the contrary, the diffuse component result of the partial absorption of the radiation by the sample and scatter in all directions[22], [26].

Kubelka and Munk developed a theory describing the diffuse reflectance process for powdered samples which relates the sample concentration to the scattered radiation intensity. The *Kubelka–Munk* equation is as follows[27]:

$$\frac{(1-R)^2}{2R} = \frac{c}{k} \tag{1.7}$$

Where *R* is the absolute reflectance of the layer, c is the concentration and *k* is the molar absorption coefficient. An alternative relationship between the concentration and the reflected intensity is now widely used in near-infrared diffuse reflectance spectroscopy, namely[20]:

$$A = \log\frac{1}{R} = a'c \tag{1.8}$$

Where *A* is the absorbance, *R* the relative reflectance ($R = R_{sample}$ /-$R_{reference}$), *a* 'the constant of proportionality and *c* concentration.

In the *Transflectance mode* the sample, usually liquid or semi-liquid, is placed in a cuvette with a reflective surface on one side. The incident radiation beam enters the transparent face of the cuvette, passes through the sample and is reflected on the other side. In this way the beam returns back through the sample and the radiation is captured by the detector. The resulting optical path is double the size

between both surfaces of the cuvette. Currently, measurement systems by means of optical fiber transflectance permit adaptation of NIR instruments in-line analysis.

As it can be seen NIR spectroscopy offers to the analyst a broad set of instrumentation and also several modes of spectra acquisition. Choosing the better configuration depends strongly in the sample and the environment of the analysis; basically the NIR adjusts to the sample and not the other way around, which is a big advantage that positioned highly this technique for process analysis.

Since the samples evaluated in this thesis were solid pharmaceuticals in different forms: powders, granulates, cores and tablets the acquisition mode used was diffuse reflectance. Two types of NIR instruments were used: 1) FOSS NIR systems 5000 equipped with a rapid contain analyzer module (RCA) for solid samples and 2) Buchi FT-NIR-Flex 500 spectrophotometer equipped with a module for solids - Petri solid sample holder **Fig10**.

Fig10.Solid sample accessories from the NIR used A) RCA from Foss NIR Systems and B) petri dish sample holder from Buchi

2.5 Near infrared spectroscopy as an analytical tool in the pharmaceutical process

In the last years NIR spectroscopy was found to be very useful for industrial analytical applications due to several characteristics that led NIR as a fast, simple, versatile and robust technique [33].

The current analytical techniques used for pharmaceuticals quality control are more expensive in terms of time and money in comparison to NIR, since they requires sample preparation –in some cases requiring solvents- and its instrumentation at the moment do not allow on line analysis, which increases the time of analysis and this, in turn, decreases the production productivity.

The landscape is quite the opposite for NIR due to its quickness to collect a spectrum -only few seconds- that will contain both physical and chemical information, therefore several parameters can be evaluated in a single measurement. It must be point out that the most interesting advantage of NIR spectroscopy is its nondestructive character of the analysis: a sample can be analyze without previous sample preparation, in this way avoiding important steps responsible for error sources. This also allows in many cases the sample to be reuse.

Quality control by the pharmaceutical industry has traditionally relied on assessment of the raw materials prior to processing and analytical determinations of the end-product. Although this methodology usually allows product quality regulations to be met, errors or unexpected variability arising at some stage of the process may not be detected before reaching the end-product and lead to time and money losses in addition to diminished productivity [34].

In recent years, the US FDA has encouraged the use of process analytical technology (PAT) by the pharmaceutical industry. PAT is intended to assure product quality via careful design, monitoring, control and surveillance of each manufacturing stage. With this methodology, quality in the product and efficiency in the production process result from a deep knowledge of the process and strict control of any physical, chemical and quality-related factors influencing each stage.

Quality in pharmaceutical production processes cannot be assured merely by analysing raw materials and end-products; rather, it requires carefully designing and implementing each production stage [13], [14].

There is ample evidence of the usefulness of near infrared spectroscopy (NIRS) as a pharmaceutical process control analytical methodology. In fact, NIRS is a simple, expeditious, non-destructive instrumental technique [21], [33], [35]–[37] and NIR spectra provide both physical and chemical information about solid samples. As a result, its use combine with chemometrics data processing have turned it into a promising tool for process control within the framework of PAT.

Moreover the recent advances in the development of *Micro NIRs* regarding portability, versatility and cost allow the measurement in different locations in the process line where traditional instruments could not be placed. These big advantages of such a small scale instruments contribute to accomplish a complete embrace of PAT schemes into manufacturing.

3. Chemometrics

3 CHEMOMETRICS

3.1 Overview

Nowadays there is a broad set of analytical techniques which can be used for analysis of compounds; each methodology can provide a specific kind of instrumental signal. Some of them provide as an analysis outcome, a single output variable, some others more than just one. However what is important to point out is that these signals are related to sample characteristics, allowing in this way species identification, characterization, and quantification.

From the instrumental signals the analyst can obtain certain data; the amount and the quality of the information obtained from it depends on how the data is processed, and the relevant information is extracted; all these procedures are aimed to turn data into information, and this information into knowledge.

The data processing in this way enhance to the maximum the advantages of each technique, and represent a useful tool for analysts for predicting unknown sample's properties and optimizing systems, reactions or processes.

Spectra are rich in information. That is why we can often infer chemical or physical information properties of a material from spectra alone. However NIR spectra present broad and overlapped bands which need chemometrics data processing to extract and understand this vast amount of data [38], [39].

Chemometrics according to *Massart* [40] definition is considered as "the chemical discipline that uses mathematics, statistics, and formal logic a) to design or select optimal experimental procedures b) to provide maximum relevant chemical information by analyzing chemical data and c) to obtain knowledge about chemical systems".

In another hand *Miller* defined chemometrics as "*the way of analyzing chemical data, in which both elements Statistical and chemical thinking are combined*"; also he remarks that there are many definitions of chemometrics but there are three consistent elements in all of them:

i) *Empirical modelling* ii) *multivariate model* and iii)*chemical data*

Taking this into account, the easiest definition of chemometrics is "*the application of multivariate, empirical modelling methods to chemical data*" [39].

Nowadays the amount available tools and softwares to apply chemometrics to data analysis are quite big, but the analyst always must keep in mind to use just the ones needed to solve its particular problem. *Miller* also advice three principles to the successful applications of chemometrics: a) Keep the models simple, b) Do your best to include all the relevant responses of your analyzer to your calibration data and c) always contemplate your problem in both ways: statistically and mathematical [39].

This thesis is focused on the use for multivariate data analysis using chemometrics tools for the development of NIR methodologies useful for pharmaceutical quality control based on product knowledge and calibration model optimization.

3.2 Modelling stages

Chemometric modeling can be considered as the process of correlating properties to spectra, using mathematical and statistical procedures over analytical data.
When NIR spectra are used to establish this correlation the amount of outcome variables linked to the properties is quite large, therefore multivariate analysis is required [38].
The difference between univariate and multivariate analysis is that the first one related a single outcome variable with the property of interest, while in multivariate analysis several outcome variables X are the ones to must be related.
The main goal of the multivariate methods is to establish classification or calibration models able to predict unknown properties of the samples of interest.
The protocol to follow is described below **Fig1** [20][41]:

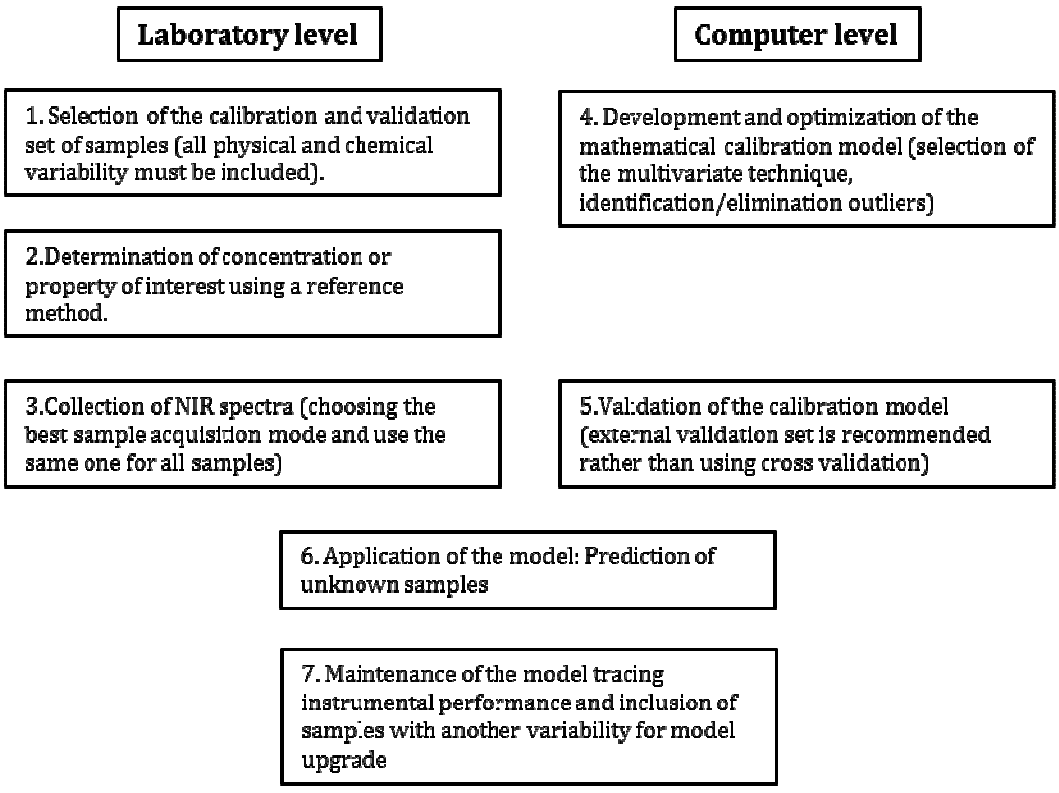


Fig1. Principal steps in the development, evaluation, use and maintenance of quantitative model based on NIR spectroscopy

3.2.1 Selection and preparation of calibration set

The feasibility of the model depends strongly in one fundamental assumption: *the samples of the calibration set must be of the same nature of the samples to predict*. That means that all possible sources of variability due to the manufacturing process must be consider in the preparation of these samples. Moreover the concentration of those must be spanned in a wide enough range of values to determine the target parameter.

As it was mentioned before the NIR spectrum contains physical and chemical information, therefore the variability regarding this both is important to consider. The *physical variability* refers to all the physical characteristic of the sample resulting from manufacturing steps such as: size, form, particle distribution and degree of compression. In addition the *chemical variability* is related to the concentration range in which the samples are spanned, and it is important to point out that this range must be wide enough to facilitate the quantification of the parameter of interest and for outliers detection [41].

Constructing calibration sets using merely production samples is impossible since their concentrations will typically span a too narrow range (usually not greater than ±5% around the nominal API content) for a robust, accurate model to be constructed. A number of strategies have been proposed to develop accurate calibration sets spanning the desired concentration ranges and containing physical variability from the manufacturing process [41], [42]. The following are among the most salient proposals.

(a) *Using pilot plant samples*. With this strategy, the calibration set is constructed from samples prepared at a pilot plant mimicking the operations of the target industrial process (e.g., granulation, grinding, compacting). The ensuing samples can thus be expected to be physically similar to actual production samples. The API and excipients contents to be used should span the pre-set concentration range. Although this strategy ensures incorporation of most of the variability's source in the samples into the calibration set, it involves a labour-intensive process in addition to setting up a pilot plant to conduct the industrial process at a smaller scale.

(b) *Underdosing and overdosing industrial samples*. This strategy expands the concentration range spanned by samples of industrial origin by adding a small amount of API or excipients to powder or granulate samples in order to obtain new, doped samples with an API content above (overdosed samples) or below the nominal value (underdosed samples). Because a small addition of API or excipients causes no appreciable physical change, the physical variability of overdosed and underdosed samples is intrinsically identical with that of production sample [42], [43]. However the implementation of this strategy must be carefully addressed when any experimental design is used to avoid eventual correlations between API and excipients, which can lead to possible collinearity.

By the addition of API for overdosed samples preparation, its concentration increases while the concentration of the mixture of excipients decreases (correlation r=-1); moreover since the concentration of all excipients simultaneously decreases in the meantime a correlation of r=1 is also presented between excipients. The same happens for underdosed samples preparation.

These tight correlations (collinearity of the concentrations in the sample set) interferes with the modelling criteria that allow the determination of the property of interest, since the influence of each compound for the modelling can be hardly distinguish, leading in a poor selectivity of the model.

(c) *Laboratory samples*. Laboratory samples are obtained by weighing and homogenization of appropriate amounts of API and excipients powders close to the nominal values and spanning the desired concentration range. However, a calibration set obtained from laboratory samples alone cannot represent the whole physical variability of production samples and must thus be completed with samples from the industrial manufacturing process. Several works have showed the suitability of this methodology for the preparation of the trainings set [41]–[43].

However, spectral differences between laboratory-made samples and production samples can be quite large that ensuing models are rather complex and scarcely

robust. This strategy is suitable for modelling when there are no spectral differences between laboratory and production samples; when these differences exist, this is not the most recommendable.

(d) *Calculation and addition of the process spectrum*. This strategy, proposed by *Blanco et al*., [44], [45]relies on the fact that physical variability in industrial samples can be mathematically added to the spectra for powder laboratory samples in order to enable incorporation of all sources of variability into the calibration set. The variability in the production process is incorporated by calculating a virtual spectrum called the process spectrum (Sp) and adding it to a calibration set consisting of powder samples obtained by weighing of the different components. The concentration of each mixture component in the samples is previously established in order to encompass the desired API concentration range while reducing collinearity between concentrations. This procedure requires no reference method to determine the concentration of the target species, and has showed to be an easy, robust and accurate methodology suitable for the purpose.

3.2.2 Determination of reference values

Constructing the calibration model requires a previous determination of the variables to be determined. For this, a reference method must be chosen to provide precise and accurate values since from these depend strongly the quality of the model to be developed. NIR methods have shown indeed a better performance in terms of precision –since there is no need for sample preprocessing- in comparison with other methodologies, but its accuracy depends strongly on the reference method; hence the need of carefully chosen of the reference method, therefore in this thesis was used high performance liquid chromatography (HPLC) because is a broadly used methodology in the pharmaceuticals manufacture and its suitability for the purpose is widely known. Also the analytic balance was used as reference method.

3.2.3 Spectra acquisition

Based on the fact that samples of the calibration set must be of the same nature of the samples to be predicted, the spectra acquisition must be carried out carefully assuring that the same instrument and computer are used, and that all the samples are recorded under the same conditions.
Since the main goal of calibration is to predict properties of unknown samples, it is very important to assure that the variability of the instrument is very low, and that the main spectral contributions are due to the properties of interest is rather than the noise [41].

In some cases certain characteristics of the samples can be pretty notable in the spectra but they are not related to the parameter of interest, for instance, physical characteristics of the sample when a model is meant to quantify the API in a formulation. When this hindrance appears, the use of spectral pretreatments is recommended to decrease or cancel these contributions, and enhance the convenient signals for further modelling.

In another hand, it must be carefully considered the type of sample and the environment of the analysis in order to select the best spectral acquisition mode and instrumentation-since the technique offers a wide range of analysis possibilities-; these with the aim of the implementation of the model in routine, and assure all the factors that can influence its robustness prior model installation.

3.2.4 Spectral pretreatments

Preprocessing NIR spectra data has become an essential part of chemometrics, and there are three aims for this preprocessing stage in data analysis [46]:
a) To compress the amount of data and eliminate data that is irrelevant to the study that is being undertaken.
b) To preserve or enhance sufficient information within the data in order to achieve the desire goal, reduce noise, increase resolution.

c) To extract the information in, or transform the data to, a form suitable for further analysis.

Moreover, when chemical determination is needed preprocessing spectra is important to remove physical phenomena in the spectra in order to improve the subsequent multivariate regression, classification model or exploratory analysis.
The most widely used approaches for preprocessing can be divided in two groups: *scatter correction methods* and *spectral derivatives* [47].
However the *average of the spectra* is also broadly used, because it allows noise reduction of the data.

3.2.4.1 Average of the spectra

This is a common preprocessing technique that has been automatized for almost all modern instruments, where several spectra of the same sample are recorded and once the signal has been accumulated, each wavelength is averaged by dividing the sum by the number of scans performed.
The method is based on the assumption that noise is random, whereas the signal is not.

3.2.4.2 Scatter correction methods

The scattering effect can generate multiplicative variations between the spectra. These variations are 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 for further modelling when the parameter of interest is the concentration. The most used methods that can correct these variations are: *standard normal variate* (SNV), *multiplicative scatter correction* (MSC), *extended MSC* (EMSC) i.a [48].
The scattering effects due to physical characteristics of the samples were successfully corrected in this thesis using SNV[49], therefore a deeper explanation of how this preprocessing technique works is describes below:

3.2.4.3 The standard normal variate (SNV)

This is a row oriented transformation in which the scattering effects are removed by centering the mean Absorbance of the spectrum to zero and followed by scaling to unit variance.

It operates individually on each spectrum centering at an average intensity of zero and standard deviation equal to one. To achieve this, each wavelength absorbance value (Abs_j^{SNV}) is transformed according to Equation 3.1, where Abs_{avg} is the mean absorbance for each spectrum and S is its standard deviation.

$$Abs_j^{SNV} = \frac{Abs_j - Abs_{avg}}{s} \tag{3.1}$$

3.2.4.4 Spectral derivatives

The spectral derivatives are the most used preprocessing techniques in analytical spectroscopy. Since NIR spectra are characterized by its broad and overlapped bands, they require preprocessing techniques that allows the enhancement and differentiation of the analytical signal of interest; moreover, these spectral pretreatments also eliminate constant and linear baseline spectral drifts. [49].

First and second derivatives are more common in practice than higher-order ones. The first derivative removes only horizontal baseline of varying levels effects, whilst second derivative removes both baseline and linear trends.

The spectral derivation is mainly done for two methods: *Norris and Williams derivation* and *Savitzky-Golay derivation*. Both derivation techniques use smoothing in order to reduce the noise in the corrected spectra[46], [48].

The basic method of derivation is based on finite differences: For the first derivative is calculated as the difference between two subsequent spectral measurements points (wavelengths) [48].

$$x_i' = x_i - x_{i-1} \tag{3.2}$$

The second order derivative is estimated then by calculating the difference between two successive points of the first-order derivative spectra:

$$x_i^{''} = x_i^{'} - x^{'}_{i-1} = x_{i-1} - 2 \cdot xi + x_{i+1} \tag{3.3}$$

Where x' denotes the first derivative and x'' the second derivative at point (wavelength) i.

The Norris-Wiliams derivation is a basic method proposed and elaborate by Norris and Williams in 1984. It consists in two steps [48]:

1) Smoothing of the spectra, where an average is made over a number of predetermined numbers of points:

$$x_{smooth,i} = \frac{\sum_{j=-m}^{m} x_{org,i+j}}{2m+1} \tag{3.4}$$

Where m is the number of points in the smoothing window centered around the current measurement i.

2) For first-order derivation the finite difference between each point is calculated with a given gap size (larger than zero), whilst for the second order derivation, take twice the smoothed value at point i and the smoothed value at a gap distance in both sides:

$$x_i^{'} = x_{smooth,i+gap} - x_{smooth,i-gap} \tag{3.5}$$

$$x_i^{''} = x_{smooth,i-gap} - 2 \cdot x_{smooth,i} + x_{smooth,i+gap} \tag{3.6}$$

By applying the smoothing prior the derivate, the noise problem is decrease; however, Norris and Williams proposed to normalize the spectra to equal intensity at a single selected wavelength after derivation **Fig2**.

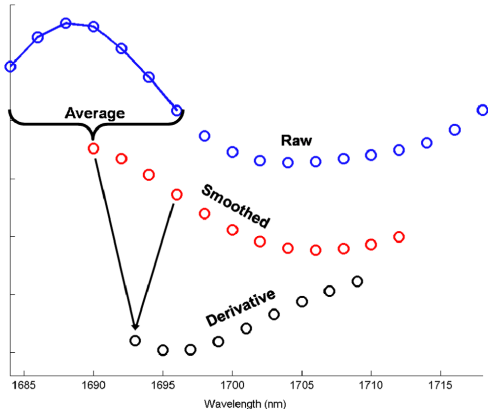


Fig2. Estimation of the first derivative by Norris-Williams. A 7-point window is used for smoothing, and a gap size of 3 is applied in derivation

In the other hand, ***the Savitzky-Golay*** derivation is based on numerical derivation of a vector that also includes a smooth step. In order to find the derivative at center point *i*, a polynomial is fitted in a symmetric window on the raw data. When the parameters for this polynomial are calculated, these values are subsequently used as the derivative estimate -a center point-. This operation is applied to all points in the spectra sequentially. The number of points to calculate the polynomial (window size) and the degree of the fitted polynomial are both decisions that need to be made **Fig3** [48].

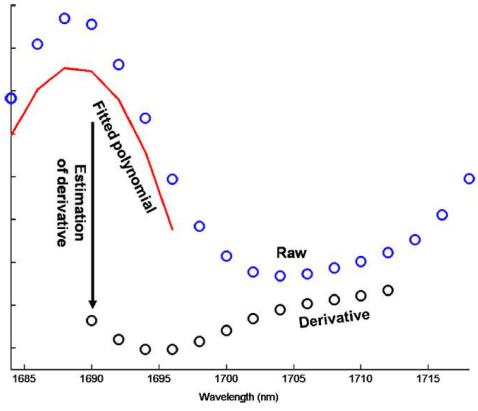


Fig3. Estimation of the first derivative by Savitzky-Golay. A 7-point window and a second-order polynomial is used for smoothing.

As a summary, it must be said that *Norris and Williams derivation* is similar to finite differences, but introduces smoothing and gap-size as counteractions in the estimated derivate spectra, to preserve the signal/noise ratio. However, *Savitzky-Golay derivation* uses more common filtering techniques to estimate the derivative spectra, and, instead of using the finite difference approach, fits a polynomial through a number of points to maintain an acceptable signal/noise ratio. In general, the *Norris and Williams derivation* and *Savitzky-Golay derivation* do not give the same estimates, therefore choosing which one must be used is merely a trial and error task.

3.2.4.5 Reduction of variables by principal component analysis (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 containing the relevant information for calibration modeling. The best known and most widely used variable reduction method is principal component analysis (PCA). This 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), Eigenvectors or factors, correspond to the largest eigenvalues accounting for the largest possible variance in the data set. The first PC represents maximum variance amongst all linear combinations and each successive variable accounts for as much of the remaining variability as possible [22].
This chemometric tool has been broadly used in this thesis. The mathematical bases are described in detail in the following sections.

3.2.5 Qualitative Analysis

3.2.5.1 Overview

The qualitative analysis in pharmaceutical industry is related to the identification or classification of a product based on their chemical or physical properties. In NIR spectroscopy these analysis can be performed by extracting the information from

the spectra using different chemometric techniques based on pattern recognition methods (PRM) [35], [50].

These PRM are based in the evaluation of the similarity of an object to a reference one or a specific class.

Mathematically the comparison is made by the calculation of similarity indicators values, which normally refers to correlation or distance criteria.
The PRMS can be classified in general terms as supervised or non-supervised methods, depending on if there is any prior knowledge of the samples and its nature.

3.2.5.2 Principal component analysis (PCA)

The large amount of experimental data in multivariate data analysis presents logistical and mathematical issues when it comes to process information for further analysis. Data compression is the process of reducing data into a representation that uses fewer variables, but still expressing the most important information.

From a logistical point of view, reducing variables is a more convenient way to storage and transport the information; form a mathematical point of view, the compression of the data allows the reduction of redundant and irrelevant information, facilitating the subsequent modeling model techniques to perform more efficiently. The PCA is not doubt the most used chemometric tool for reduction of variables [22].

The basis of PCA can be explained by a transformation of a bidimesional data matrix
X (NxM) constituted by N samples (recorded spectra) and M number of variables (wavelengths). The PCA aim is to find the directions of maximum variability in which the N points in the space of dimension M are grouped. The reduction of the X

dimensionality is performed by calculating the lower number of new axes called principal components (PC) that are able to explain the maximum variability the samples. The first component (PC1) is the linear combination of the *M* variables that explains the maximum sample variability; the second component (PC2) will be orthogonal to PC1 and collect less variability than the first one. While the number of PC's are increasing, the explained variability between the components decreases until the total variability is explained [50].

Mathematically speaking, the matrix spectral data ***X*** is decomposed in a new scores matrix ***T*** and loadings ***P***, and the residual matrix ***E***. This matrix transformation is described in the **Equation 3.7**

$$X=T.P^t + E \quad (3.7)$$

The loadings in geometric terms correspond to the cosines of the angles formed by the new axes with the original, and the scores are the coordinates of the samples in these new axes.

Due to the orthogonality, all the PCs contain different information. The first PCs normally described the most relevant variability of the samples, while the latter described variations due to noise. The matrix *E* contains the information is not collected in any of the PC's and it is known as the residual.

3.2.5.3 Projection methods

Projection methods are the most used techniques for exploratory analysis, since the results can be easily interpreted. The projection techniques are based also on the dimensionality data reduction, which highlight structure in the data (e.g clusters of samples or variables) although the main aim is not to identify them. Projection methods project samples into a low dimensional space using a specified criterion (e.g., variance in PCA). If data clustering is related to this criterion, clusters of samples may be visualized. The absence of meaningful sample

associations does not necessarily mean the information sought is not present. Modifications to the criterion have strong influence in the results, such as by using a different method [e.g., Independent Component Analysis (ICA) instead of PCA], or preprocessing the data by, e.g., scaling in PCA. [51].

Moreover, some quantitative thresholds can be established in order to delimit the space that belongs to the population encompassed in each cluster of samples, for instance Hotelling statistics, F-Residuals i.a. These statistics were evaluated in one of the study cases presented in this thesis, and they will be described in detail in the following sections.

3.2.5.4 Correlation coefficient

The correlation coefficient is an indicator of how similar are the spectra of an unknown sample and other that belongs to a predefine class.
Generally, the comparison is made between a pure component and an unknown sample, although the class could be predefined before comparison, and a threshold established for the determination that the object belong to the class.
The calculation of the coefficient correlation is given by:

$$r = \frac{\sum_i xi.yi}{\sqrt{\sum xi^2 . \sum yi^2}} \tag{3.8}$$

Where *xi* is the absorbance at wavelength *i* for the spectrum *x* (pure spectrum), and *yi* is the wavelength *i* for spectrum (spectrum of the unknown sample). The maximum similarity value for two spectra would be r = 1. The correlation coefficient is essentially a qualitative parameter for analysis and a measure of the collinearity of two vectors.

3.2.6 Quantitative Analysis

3.2.6.1 Overview

The aim to construct a calibration model is to establish a relation between the analytical signal and the property of interest. Since NIR spectra provide a large number of variables to be related, approaches based on multivariate calibrations must be used.

In pharmaceutical analysis, most of the calibrations models are calculate for the quantification of major constituents in the sample. In general, the detection limit is about 0.1% (w/w), although for certain samples under certain matrix characteristics the NIR can go even to lower detection limit values.
The quantitative methods can be divided in two groups: lineal and non-lineal summarized representation of some of them can be seen in the Fig4 [20], [39]:

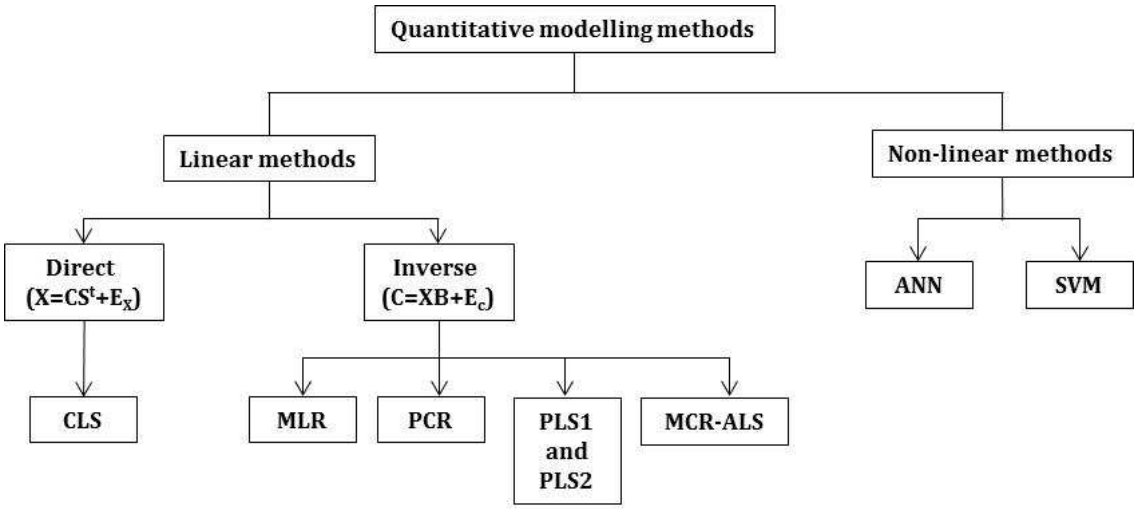


Fig4. Flowchart of various quantitative modelling methods

The *linear methods* are typically used to relate an independent variable to a set of depended variable in the case of this thesis to the spectral data. This technique is very useful when there is a certainty the response of the analyzer contained all the information regarding the property of interest.

The linear methods are in turn subdivided in two types: *Direct* and *Inverse* and its general criteria to distinguish between both of them are based in the relation between the signals and the property of interest. The criteria to distinguish both of them rely on the general form of the model:

$$\textbf{Direct}: X = CS^t + E_x \quad (3.9)$$

$$\textbf{Inverse:}\ C = XB + E_c \quad (3.10)$$

Where C is the concentration matrix or the sample properties of interest, S is the spectral data and E are the model residuals. Moreover the *Direct methods* follow the classical expression of Beer-Lambert law, and express the analyzer response as a function of the concentration, whereas the Inverse methods express concentration as a function of the analyzer responses [39]. The **Table 1** summarized from a broad perspective the basis of the linear methods exposed:

Table1. Summarized characteristics of linear quantitative modelling methods

Method	Type of method	Calibration set/or reference	Suitability for NIR	Characteristic of interest
Multiple linear regression (MLR)	Inverse	Yes	Low. Due to the sharp selection of variables	Limited No of variables Correlation between signals can complicate models 'calculation
Classical least squares (CLS)	Direct	Yes From the pure components	Just suitable for concentration determination	Reflection of the classical expression of the Beer-Lambert law The spectra of the pure components is required (experimental or estimated) Since the spectrum is taken as a linear combination of the spectra of the pure components. Nonlinear responses of the analyzer and strong spectral interaction effects can alter the model
Principal component regression (PCR)	Inverse	Yes and should contain information about the property to be determined	Suitable	Accounts covariance between X variables No problem with correlation between signals The entire spectrum can be used Compression of data based solely in explained variance in *X*, subsequent regression of PC´s
Partial least squares (PLS)	Inverse	Yes and should contain information about the property to be determined	Suitable	Simultaneous decomposition of X and Y matrix. Compressed data contain the most variance from both X and Y No problem with correlation between signals Can quantify one property of interest (PLS1) or multiple (PLS2)
Multivariate curve resolution alternative least squares (MCR-ALS)	Direct	Not mandatory Profiles From the pure components	Suitable	Intend the recovery of pure responses profiles of the chemical constituents of an unresolved mixture Can be used without knowing all the constituents on a mixture Consist of iterative process that not always converge to a useful solution

From a broad perspective it can be said that many of the *non-linear methods* are based on machine learning, which uses algorithms that can learn through the inputs-provide by the analyst -and several iterations; in this way the algorithm make predictions or decision on data. Those methods are very useful when there is not a linear relation between the property of interest and the response of the instrument; however their complexity and the number of samples required for modelling make its implementation difficult on routine models.

Since in this thesis the calibration models were calculate by means of partial least square (PLS), a broader description is presented below:

3.2.6.2 Partial Least Squares (PLS)

One of the multivariate regression method most frequently used in quantitative NIR analysis is partial least squares (PLS) regression.

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 [35], [50], [52].

The PLS algorithm uses the information contained in both the spectroscopic data matrix, **X**, and the property of interest 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 are simultaneously decomposed, giving the outer relations:

$$X = TP^T + E \quad (3.8)$$

$$Y = UQ^T + F \quad (3.9)$$

Where U and T represent the scores, P^T and Q^T represent the loadings and E and F

the residuals for the *X* and *Y* matrixes respectively.

In PLS –like in PCA- a number of appropriate components must be selected, which assure that all the quantitative information is collected with the less amount of possible noise or other spectral information that do not concern to the parameters of interest.

As it was mentioned before the decomposition of the two matrices is performed simultaneously, and the main feature of this decomposition is seeking maximum correlation between the spectra and the property to be determined.

Once the model correct calibration is established, it is possible to predict the outcome for a new sample or a set of external calibration samples. The correct prediction of new samples will depend on the good predictive ability of the model calibration.

3.2.7 Model evaluation

The best way to evaluate the predictive ability of the model is running an external prediction test. This test will predict known Y values by the PLS model and will compare them with the known values. Different global statistics parameters can be evaluate such 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 PLS model testing. These values evaluate the residuals with the reference values [20].

$$RSE(\%) = \frac{\sqrt{\sum_{i=1}^{n}\left(Yi^{nir} - Yi^{ref}\right)^2}}{\sum_{i=1}^{n}(Yi^{ref})^2}.100 \qquad (3.10)$$

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n}(Yi^{nir} - Yi^{ref})^2}}{n} \qquad (3.11)$$

Where n = number of samples, Yi^{nir} and Yi^{ref} are magnitudes of determine property

by NIR or reference method.
It is important to point out that RSE is always expressed as a relative value (%), while the RMSE as an absolute one with the unities of the property to be determined.
Regarding the set of samples that is being validated the RSE and RMSE would be renamed as RSEC/RMSEC (for calibration) and RSEP/RMSEP (for prediction).

3.2.8 Model validation

Validation is the final stage of development of an analytical method. The objective of validation of an analytical procedure is to demonstrate that it is suitable for its intended purpose [53].
Prior validation, the analytical procedures and aim of the proposed method should be clearly defined and understood. This understanding should be obtained from scientific knowledge based on method development and optimization.
The typical parameters tested during validation are listed below:

3 Selectivity
4 Linearity
5 Range
6 Accuracy
7 Precision (repeatability, intermediate precision, and reproducibility)
8 Quantitation limit
9 Detection limit
10 Robustness

However depending on the purpose of the analytical methods some parameters are mandatory while other remain optional Since the methods developed in this thesis are NIR spectroscopy based-methods, some modification in the validation are considered by the regulation guidelines.
Each parameter is described briefly below [53], [54]:

Specificity is the ability to assess unequivocally the analyte in the presence of

components which may be expected to be present. In the case of NIR methodologies the selectivity is tested through the construction of spectral libraries enabling unambiguous identification of the presence of the compound(s) of interest [54].

Linearity implies that the response between the signal and the property is proportional across the range. For univariate calibration this parameter is evaluated though the relation between signal/concentration. For multivariate calibrations the lineal relation between the reference value (obtained with a reference method) and the estimate value (obtained by the proposed method) must be related [53], [54].

Range is related to the interval between the upper and lower concentration of analyte in the sample to be determined. In other words, is the calibration interval which is recommended for pharmaceuticals to encompass ±20% with respect to the nominal concentration [53], [54].

Accuracy expresses the closeness of agreement between the value which is accepted either as a conventional true value or an accepted reference value and the value found. This should be established across the specified range of the NIRS procedure and should be appropriate for its intended use [53], [54].
The regulations recommend to perform at least 9 measurements in at least three different levels (3 levels x 3 replicates). Also a test that allows the evaluation of differences between obtained values and reference values is advised.

Precision expresses the closeness between a series of measurements of the sample.
Precision may be considered at three levels: *repeatability, intermediate precision* and *reproducibility*[53], [54] .
Repeatability expresses the precision under the same operating conditions over a short interval of time. Repeatability is also termed intra-assay precision. It is evaluated using a minimum of 9 determinations at three concentration levels, or

with a minimum of 6 determinations at 100% -nominal value- [53], [54].

Intermediate precision expresses the degree of reproducibility of the results making slightly variations of the normal work manner such as: different days, different analysts, and different equipment i.a.

Reproducibility implies precision between laboratories; it broadens the intermediate precision with this extra factor.

Quantitation limit can be defined as the lowest amount of an analyte that can be detected by the proposed methodology with suitable precision and accuracy.

The quantitation limit is a parameter of quantitative assays for low levels of compounds in sample matrices, and is used particularly for the determination of impurities and/or degradation products [53].

Detection limit is the lowest amount of analyte in a sample which can be detected but not necessarily quantitated as an exact value [53].

For NIR methodologies limits of detection and quantification only need to be demonstrated when relevant and where the analyte is considered to be an impurity [54].

Robustness is the parameter that the reliability of the analytical method under the influence of changes in the standard test conditions [53], [54].

REFERENCES

[1] S. H. Nusim and F. Group, *Active Pharmaceutical Ingredients. Development, Manufacturing, and Regulation*. Taylor & Francis group, 2005.

[2] K. Harbir, "Review Article PROCESSING TECHNOLOGIES FOR PHARMACEUTICAL TABLETS : A REVIEW," vol. 3, no. 7, pp. 20–23, 2012.

[3] F. J. Muzzio, T. Shinbrot, and B. J. Glasser, "Powder technology in the pharmaceutical industry: the need to catch up fast," *Powder Technol.*, vol. 124, no. 1–2, pp. 1–7, Apr. 2002.

[4] M. Tousey, "The manufacturing process: Tablet and capsule manufacturing," 2015.

[5] D. M. Parikh, "Handbook of Pharmaceutical Granulation Technology," p. 532, 1997.

[6] H. Stahl, "Comparing granulation methods," *Pharm. Manuf. Pack. Sourcer*, no. AUTUMN, pp. 58–61, 2010.

[7] J. . Ciurczak, E.W., Drennen, *Pharmaceutical and medical applications of near-infrared spectroscopy*. New York, 2002.

[8] A. H. Gerhardt, "Fundamentals of Tablet Compression," *J. GXP compliance*, vol. 14, no. 1, pp. 70–79, 2010.

[9] L. Levi, G. C. Walker, and L. I. Pugsley, "Quality Control of Pharmaceuticals.," *Can. Med. Assoc. J.*, vol. 91, no. 15, pp. 781–785, 1964.

[10] World Health Organization, "Quality Assurance of Pharmaceuticals. A compendium of guidelines and related materials," vol. 2, 2011.

[11] R. M. Haleem, M. Y. Salem, F. a. Fatahallah, and L. E. Abdelfattah, "Quality in the pharmaceutical industry - A literature review," *Saudi Pharm. J.*, 2014.

[12] Heather E. Fraser, "From art to science The metamorphosis of manufacturing," *IBM Inst. Bus. Value Life Sci.*, 2005.

[13] P. A. T. A. Framework, "Guidance for Industry Guidance for Industry PAT — A Framework for Innovative Pharmaceutical," no. September, 2004.

[14] D. J.E, "Overview of Process Analysis and PAT," in *Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*, Second Edi., John Wiley & Sons, Ltd, 2010, pp. 1–14.

[15] University of Washington, “Center for Process Analysis & Control.” [Online]. Available: http://cpac.apl.washington.edu/.

[16] J. G. Grazal and D. S. Earl, “EU and FDA GMP Regulations :,” vol. 2, pp. 55–60, 1997.

[17] I. C. on Harmonisation, “http://www.ich.org/home.html.” [Online]. Available: http://www.ich.org/home.html.

[18] C. European and health and consumers Directorate, “EU Guidelines for Good Manufacturing Practice for Medicinal Products for Human and Veterinary Use,” 2014.

[19] B. Donald A. and C. Emil W., “Handbook NIR_Ciurczak.pdf.” .

[20] C. Pasquini, “Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications,” *J. Braz. Chem. Soc.*, vol. 14, no. 2, pp. 198–219, 2003.

[21] M. Blanco and I. Villarroya, “NIR spectroscopy: a rapid-response analytical tool,” *TrAC Trends Anal. Chem.*, vol. 21, no. 4, pp. 240–250, Apr. 2002.

[22] G. Reich, “Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications.,” *Adv. Drug Deliv. Rev.*, vol. 57, no. 8, pp. 1109–43, Jun. 2005.

[23] I. Barton, “FE Theory and principles of near infrared spectroscopy,” *Spectrosc. Eur.*, vol. 14, no. 1, pp. 12–18, 2002.

[24] C. N. Banwell, E. M. McCash, and H. K. Choudhury, *Fundamentals of Molecular Spectroscopy*, Fifth Edit. McGraw Hill Education, 2013.

[25] D. A. Skoog, F. J. Holler, and T. A. Nieman, “PRINCIPIOS_ DEL_ ANALISIS_ INSTRUMENTAL_ Apendices.pdf.” .

[26] B. H. Stuart, *Infrared Spectroscopy: Fundamentals and Applications*, vol. 8. 2004.

[27] M. Alcalà, “Utilización de la espectroscopia NIR en el control analítico de la industria farmacéutica. Desarrollos iniciales en PAT,” Universitat Autònoma de Barcelona, 2006.

[28] J. P. Coates, *Process Analytical Technology: Infrared Spectroscopy for Process Analytical Applications*. Chichester, UK: John Wiley & Sons, Ltd, 2010.

[29] A. Palou, “Desarrollo de nuevas tecnologías espectrales para el control analitico de productos y procesos petroquimicos y farmacéuticos,” Universitat Auntónoma de Barcelona, 2014.

[30] S. I. Hintschich, T. Pügner, J. Knobbe, J. Schröder, P. Reinig, H. Grüger, and H. Schenk, “MEMS-based miniature near-infrared spectrometer for application in environmental and food monitoring,” pp. 2–4, 2014.

[31] J. Coates, “A Review of New Small-Scale Technologies for Near Infrared Measurements,” *American pharmaceutical review*, 2014. [Online]. Available: http://www.americanpharmaceuticalreview.com/Featured-Articles/163573-A-Review-of-New-Small-Scale-Technologies-for-Near-Infrared-Measurements/.

[32] K. . Hildrum, T. Isaksson, and T. Naes, *Near Infrared Spectroscopy bridging the gap between analysis and applications*. Great Britain: Ellis Horwood Limited, 1992.

[33] T. De Beer, A. Burggraeve, M. Fonteyne, L. Saerens, J. P. Remon, and C. Vervaet, “Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes.,” *Int. J. Pharm.*, vol. 417, no. 1–2, pp. 32–47, Sep. 2011.

[34] V. Cárdenas, M. Cordobés, M. Blanco, and M. Alcalà, “Strategy for design NIR calibration sets based on process spectrum and model space: An innovative approach for process analytical technology,” *J. Pharm. Biomed. Anal.*, vol. 114, pp. 28–33, 2015.

[35] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, and N. Jent, “A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies.,” *J. Pharm. Biomed. Anal.*, vol. 44, no. 3, pp. 683–700, Jul. 2007.

[36] M. Jamrógiewicz, “Application of the near-infrared spectroscopy in the pharmaceutical technology.,” *J. Pharm. Biomed. Anal.*, vol. 66, pp. 1–10, Jul. 2012.

[37] M. Blanco, M. Bautista, and M. Alcalà, “API determination by NIR spectroscopy across pharmaceutical production process.,” *AAPS PharmSciTech*, vol. 9, no. 4, pp. 1130–5, Jan. 2008.

[38] R. DiFoggio, “Guidelines for applying chemometrics to spectra: feasibility and error propagation,” *Appl. Spectrosc.*, vol. 54, no. 3, pp. 94–112, 2000.

[39] C.E Miller, “Chemometrics in Process Analytical Technology (PAT),” in *Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*, Second Edi., John Wiley & Sons, Ltd, 2010, pp. 353–438.

[40] D. . Massart, S. . Deming, Y.Michotte, L.Kaufman, and B. G. M. Vandeginste, *Chemometrics: A textbook (data handling in science and technology)*. 1988.

[41] B. R. M. A. B. M., “Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries,” in *Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*, Second Edi., John Wiley & Sons, Ltd, 2010, pp. 463–491.

[42] M. Blanco, M. Bautista, and M. Alcala, “Preparing Calibration Sets for Use in Pharmaceutical Analysis by NIR Spectroscopy,” vol. 97, no. 3, pp. 1236–1245, 2008.

[43] M. Blanco, M. Romero, and M. Alcalà, “Strategies for constructing the calibration set for a near infrared spectroscopic quantitation method.,” *Talanta*, vol. 64, no. 3, pp. 597–602, Oct. 2004.

[44] M. Blanco and A. Peguero, “Analysis of pharmaceuticals by NIR spectroscopy without a reference method,” *TrAC Trends Anal. Chem.*, vol. 29, no. 10, pp. 1127–1136, Nov. 2010.

[45] M. Blanco, R. Cueva-Mestanza, and A. Peguero, “NIR analysis of pharmaceutical samples without reference data: improving the calibration.,” *Talanta*, vol. 85, no. 4, pp. 2218–25, Sep. 2011.

[46] M. J. Adams, *Chemometrics in analytical Spectroscopy*. Wolverhampton, UK, 1995.

[47] Å. Rinnan, F. Van Den Berg, and S. B. Engelsen, “Review of the most common pre-processing techniques for near-infrared spectra,” *TrAC Trends Anal. Chem.*, vol. 28, no. 10, pp. 1201–1222, 2009.

[48] Å. Rinnan, F. Van Den Berg, and S. B. Engelsen, “Review of the most common pre-processing techniques for near-infrared spectra,” *TrAC Trends Anal. Chem.*, vol. 28, no. 10, pp. 1201–1222, Nov. 2009.

[49] J. Huang, S. Romero-Torres, and M. Moshbar, “Practical considerations in data Pre-treatment for NIR and Raman Spectroscopy,” 2010.

[50] N. Kumar, A. Bansal, G. S. Sarma, and R. K. Rawal, “Chemometrics tools used in analytical chemistry: An overview,” *Talanta*, vol. 123, pp. 186–199, 2014.

[51] E. Szymańska, J. Gerretzen, J. Engel, B. Geurts, L. Blanchet, and L. M. C. Buydens, “Chemometrics and qualitative analysis have a vibrant relationship,” *TrAC Trends Anal. Chem.*, vol. 69, pp. 34–51, 2015.

[52] R. Kramer, *Chemometric techniques for quantitative analysis*. New ork: Marcel Dekker, Inc, 1998.

[53] International conference on harmonisation of technical requirements for registration of pharmaceuticals, “ICH Guidelines: Q2(R1) Text on validation

of analytical procedures: methodology Q2(R1)," 1996. [Online]. Available: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf.

[54] European Medicines Agency, "Guideline on the use of Near Infrared Spectroscopy (NIRS) by the pharmaceutical industry and the data requeriments for new submission and variations.," 2012. [Online]. Available: EMEA/CVMP/QWP/17760/2009rev2.

OBJECTIVES

The quality in the pharmaceutical industry is strictly controlled and supervised by several regulatory agencies. These regulations ensure the correct and efficient manufacturing of pharmaceuticals, so that in this way the manufactured products meet quality requirements.

Any failure in the production process is not only a concern of the manufacturers, but also represent a public health issue since pharmaceuticals affect strongly the health of the consumers. In the recent years different regulations redefined the concept of quality, which is attempt to be design and build in each manufacture step, instead of testing for quality in the raw materials and final products.
Near infrared spectroscopy has shown its suitability for pharmaceutical quality control due to its simplicity, speed and non-destructive nature. Moreover, recent advances in the instrumentation have enhanced the versatility of the technique allowing the implementation of NIR in different parts of the production line for online analysis.

Based on this, the general aim of this thesis is the development of new NIR methodologies for the quality control of pharmaceuticals through the manufacturing process using multivariate data analysis techniques. These methodologies are aimed to solve real industrial problems though the product understanding and the enhancement of modelling strategies.
For this the following specific objectives were proposed:

- Study of different pharmaceutical formulations –powders, granulates cores and tablets- to evaluate the relation between the NIR spectra and the chemical and physical variability.

- Development of NIR calibration models able to quantify active principle ingredient (API) at different manufacturing steps.

- Enhancements of the multivariate modelling process by the evaluation of different strategies for construct the calibration set and its suitability for the incorporation of physical and chemical variability.

- Study of quality metrics for the optimization of the construction of calibration set strategies.
- Evaluation of the influence of surface scanned area and sample representativeness in the spectral information.

Case of study I

Strategies for selecting the calibration set in pharmaceutical near infrared analysis. A comparative study

In this work, we assessed three different calibration strategies for the quantification of the API in a pharmaceutical granulate in low concentration (10 $mg \cdot g^{-1}$). Such strategies were used to construct calibration models allowing all potential variability in new, unknown samples to be considered.

The models were constructed by PLS using samples of variable origin including laboratory-made powder mixtures and industrial samples; and variability in production samples was incorporated via a mathematical algorithm.

Published article:
V. Càrdenas, M. Blanco, M. Alcalà
Strategies for Selecting the Calibration Set in Pharmaceutical Near Infrared Spectroscopy Analysis. A Comparative Study
Journal of Pharmaceutical Innovation
December 2014, Volume 9, Issue 4, pp 272-281

2.1. INTRODUCTION

In recent years, the expeditiousness, non-destructive nature and high flexibility of near infrared spectroscopy (NIRS), among other favourable features, have fostered its use for the determination of quality-related parameters by the pharmaceutical industry. The quantification of the active pharmaceutical ingredient (API) from the early stages to the end of the process is critical to achieve a high quality of the pharmaceutical products. These advantages have turned NIRS into an effective alternative to more expensive and labour-intensive techniques as HPLC. Also, the greatest disadvantage of NIRS (viz., its limited selectivity due to wide-overlapped bands) can hinder its use for the identification/quantification of components present in a mixture in low concentration. The determination of an API in low dose may present some difficulties depending of the nature of the sample such as homogeneity and API aggregation that could hinder a good spectroscopic measurement etc. Some previous studies have shown the effectiveness of NIR spectroscopy joint to chemometrics to quantify low concentration API´s in tablets, suspension and other forms [1]–[4]. Quantification studies for API´s present in such a small quantity are required for powder mixtures and blends in order to assess and ensure a high quality of the products from the early stages to the end.
The NIR impediments mentioned before can be overcome by using multivariate procedures to extract important information with the aid of an appropriate chemometric algorithm as partial least squares (PLS). At present, all NIRS instruments come with chemometric softwares [5]–[8] intended to facilitate application of the technique to samples of highly diverse nature and easy data collection-further analysis.

Near infrared spectra contain both physical and chemical information about samples. This affords analyses of both the qualitative type (e.g., identification and characterization of raw materials, monitoring of reactions and/or processes, identification of polymorphs, assessment of mixture homogeneity) and the quantitative type (e.g., quantitation of Active Pharmaceutical Ingredients (API), excipients, moisture or particle size). Many of these determinations are critical

with a view to assessing quality in pharmaceutical products [9]–[11].
Developing robust, accurate, precise methodologies for these determinations requires using calibration sets containing as much variability as possible in the nature and properties of the target samples. In addition, the samples to be included in the calibration set for a quantitative determination should span a wide enough concentration range to ensure accurate quantitation of normal samples and easy identification of the abnormals [12], [13].

Constructing calibration sets containing all potential sources of variability in the samples to be predicted is usually difficult and occasionally impossible. Also, it requires exercising due care to avoid developing inaccurate models. Pharmaceutical samples are prepared from very pure high purity raw materials with highly compliant, similar physical properties (reduced variability); also, they are usually subjected to simple, reproducible processes. Such a low variability in the raw materials and manufacturing procedures of pharmaceutical samples can be expected to ensure easy incorporation of their whole variability into a calibration set consisting of a few samples. For accurate determination of unknown samples, calibration samples should exhibit enough chemical variability (viz., API and excipient concentrations ± 20% around their nominal values as per ICH guidelines) and physical variability (particle shape, size and distribution, etc., which are subject to changes arising from the manufacturing process).

Constructing a correct calibration set from production samples alone is impossible since their concentrations will typically span too narrow a range (usually not greater than ±5% around the nominal API content) for a robust, accurate model to be constructed. A number of strategies have been proposed to develop accurate calibration sets spanning the desired concentration ranges and containing physical variability in the manufacturing process [12], [13]. The following are among the most salient proposals.

(a) *Using pilot plant samples*. With this strategy, the calibration set is constructed from samples prepared at a pilot plant mimicking the operations of

the target industrial process (viz., granulation, grinding, compacting). The ensuing samples can thus be expected to be physically similar to actual production samples. The API and excipient contents to be used should span the pre-set range. Although this strategy ensures incorporation of most of the variability's source in the samples into the calibration set, it involves a labour-intensive process in addition to setting up a pilot plant to conduct the industrial process at a smaller scale —which is often impossible. This strategy was omitted here as we had no pilot plant.

(b) Underdosing and overdosing industrial samples. This strategy expands the concentration range spanned by samples of industrial origin by adding a small amount of API or excipients to powder or granulate samples in order to obtain new, doped samples with an API content above (overdosed samples) or below the nominal value (underdosed samples). Because a small addition of API or excipient causes no appreciable physical change, the physical variability of overdosed and underdosed samples is intrinsically identical with that of production samples [13], [14]. However the implementation of this strategy must be carefully addressed because of the eventual collinearity problem (simultaneous increasing/decreasing of the concentration for all the components in the formulation) that will conclude in a poor selectivity of the model

(c) *Laboratory samples*. Laboratory samples are obtained by weighing and homogenization of appropriate amounts of API and excipient powders close to the nominal values and spanning the desired concentration range. However, a calibration set obtained from laboratory samples alone does not represent the whole physical variability of production samples and must thus be expanded with samples from the industrial manufacturing process [13], [14].

This strategy is useful when there are no spectral changes between laboratory powder and production samples –especially when both do not differ on the physical properties-; otherwise this strategy is not suitable for the purpose.

As shown in previous work, expanding the calibration set with about 30% of samples of industrial origin suffices to this end.

(d) *Calculation and addition of the process spectrum*. This strategy, proposed by *Blanco et al*. [15], relies on the fact that physical variability in industrial samples can be mathematically added to the spectra for powder laboratory samples in order to enable incorporation of all sources of variability into the calibration set.

The variability in the production process is incorporated by calculating a virtual spectrum called the process spectrum (Sp) –which only contains physical variability-, and adding it to a calibration set consisting of powder samples obtained by weighing of the different components. The concentration of each mixture component in the samples is previously established in order to encompass the desired API concentration range while reducing collinearity between concentrations. It is important to point out that several process spectra can be calculated and added to the calibration set matrix allowing a fully incorporation of the physical variability of the process (a number between three and four is usually sufficient). Also this variability can be expanded by the use of a factor m that can be multiplied to the Sp according to the predictive ability of the calibration model –which is checked by the PCA score plot obtained from the process spectra projected into the calibration samples-[16].

$$\text{Extended Spectrum} = \text{Laboratory} + \text{Process Spectrum} \times m \ (\text{m}=0.5\text{-}1.5) \qquad (1)$$

This strategy is very useful for pharmaceutical applications since its manufacture process consist in several transformations of the product: granulation, compression and coating i.a.

This procedure requires no reference method to determine the concentration of the target species. The **Fig1** describes in detail the proposed methodology.

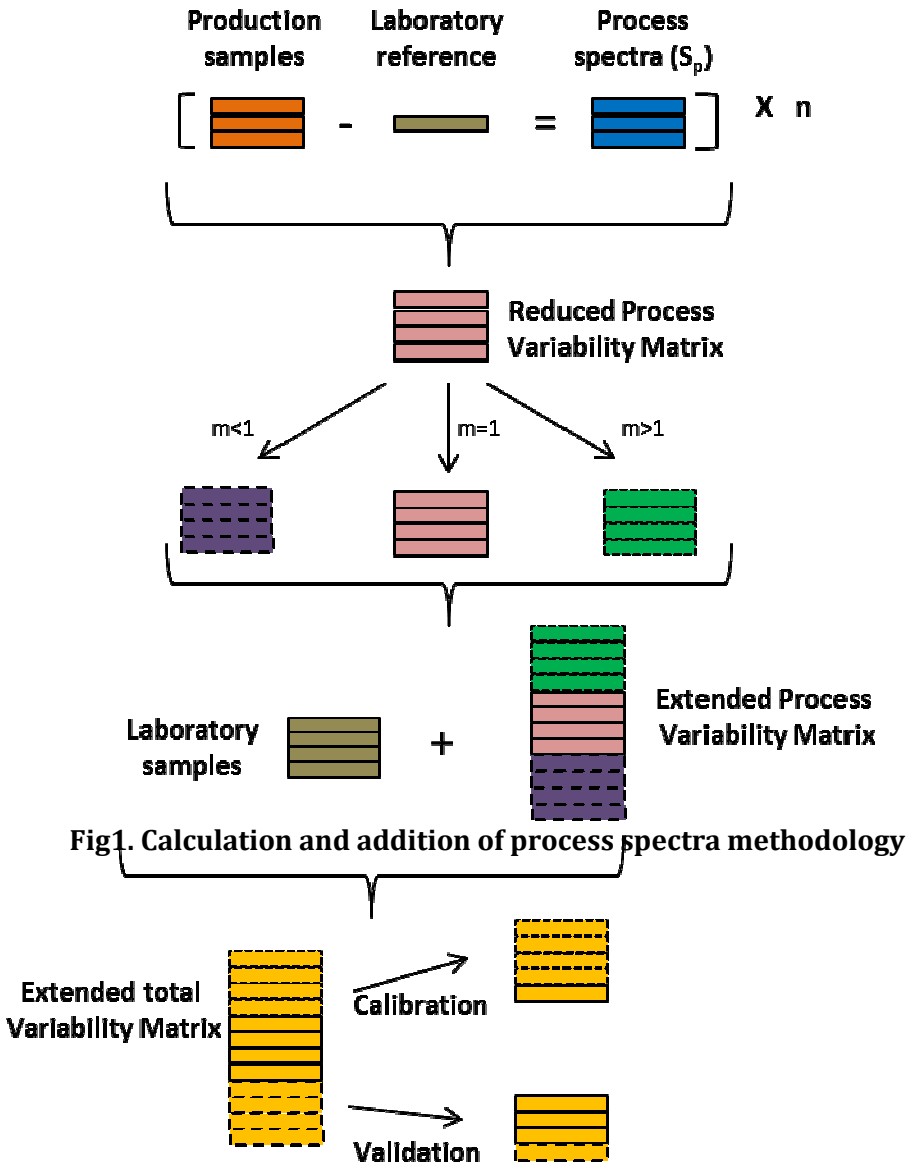


Fig1. Calculation and addition of process spectra methodology

In this work, we compared the efficiency of the last three strategies by using calibration sets constructed from doped samples or laboratory samples, or by calculation and addition of the process spectrum. The calibration sets were constructed using a small number of samples in order to reduce experimental work but still large enough to ensure accurate prediction of industrial production samples.

The strategies were applied to a pharmaceutical preparation with a low content in API and a high content in the major excipient in order to enhance their advantages and disadvantages for easier comparison.

2.2. EXPERIMENTAL SECTION

2.2.1. Production samples

The target pharmaceutical formulation was a granulated solid containing 10 $mg \cdot g^{1}$ Dexketoprofene trometamol as API, sucrose (96% w/w) as major excipient, and lemon flavour, neohesperidine, dihydrochalcone, Quinoline Yellow and ammonium glycyrrhizinate as minor excipients. All pure components and production samples were supplied by Laboratorios Menarini, S.A. (Badalona, Spain). The API content of the samples was determined by HPLC.

2.2.2 Laboratory samples

Powder samples, Laboratory samples in powder form were prepared by mixing appropriate amounts of API, placebo (a mixture of minor excipients jointly accounting for about 3% of the total content) and sucrose to span the API concentration range 8–12 $mg \cdot g^{-1}$ (i.e., ± 20% around the nominal value). Three placebos were used for the samples preparation, in which the concentration of the sucrose vary on a range of ± 5% around the nominal value –since this the major excipient which constituted the 96% of the mixture-. Samples were prepared by randomly supplying each placebo with the required amounts of API and sucrose, and homogenized on a Turbula shaker prior to recording of their near infrared (NIR) spectra. The reference values of these samples were obtained by weighing. The large differences between excipient concentrations precluded the use of sampling design techniques to reduce correlation between the concentrations of API and sucrose (the major excipient).

Doped samples were obtained by adding an appropriate amount of API (overdosing) or placebo (underdosing) to three randomly chosen industrial granulates whose API content was determined by HPLC to obtain a wide enough range of API concentrations. The resulting doped samples were homogenized prior to recording of their NIR spectra. The API reference values for these samples were

obtained by weighing and the concentrations in the industrial samples were determined by using the reference (HPLC) method.

2.2.3. Recording of NIR spectra

The previously prepared laboratory samples were homogenized in a Turbula T2C WAB shaker mixer and their NIR reflectance spectra recorded by using a FOSS NIRSystems 5000 spectrophotometer equipped with a rapid content analyser (RCA) module and governed via the software Vision v. 2.22. Spectra were recorded at 2 nm intervals over the wavelength range 1100–2500 nm. Samples were placed in a glass cell and turned over with a spatula prior to measurement in order to change the surface scanned and sampling another portion of the sample. A blank spectrum was obtained from an empty cell at the start of each working session. A ceramic plate bundled with the instrument was used as reference for measurements.

2.2.4. Preparation of calibration and validation sets

The prepared samples were split into two subsets: one to construct the calibration model and the other to validate it. The number of samples used in the calibration set for the different strategies was similar comprising 18 samples with the doping strategy, 23 with the mixed calibration strategy (powder samples + industrial granulates) and 20 with the process spectrum strategy. For the *Calculation and addition of process spectrum (Sp)* methodology the preparation of the samples involved firstly obtaining the Sp as the difference between the spectrum for an industrial granulate (Sind) and a powder laboratory sample (Slab) containing an identical concentration of API (**Fig. 1**; the calculated difference is expressed as a mathematical vector defining variability in the production process.

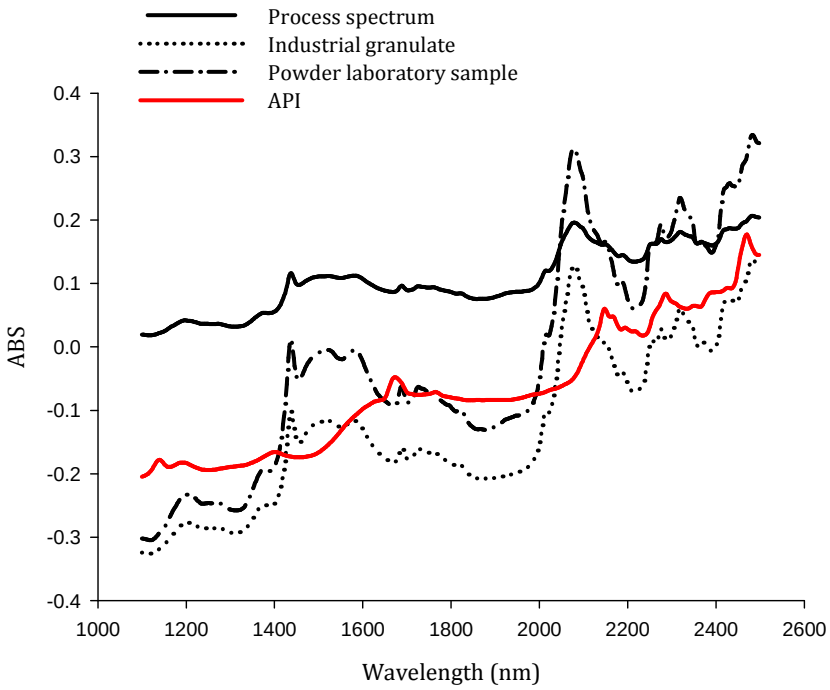


Fig1. NIR spectrum of industrial granulate, powder laboratory and calculated process spectrum

Subsequently, the *reduced process variability matrix* is obtained by the addition of the calculated vector (SP) to the spectral matrix from several powder mixtures (Scon) spanning the desire range of API concentrations. This "new" calculated matrix contains both the physical and chemical variability and was used for both calibration and validation set.

In order to increase or decrease the variability in the process spectrum SP, the calculated vector can be multiplied by a factor m, which is near-unity, and by a simple spectra addition as it was mentioned before an *extended variability matrix* can be obtained.

The confirmation of the incorporated variability was performed by an analysis of a scatter plot of scores obtained from spectra for laboratory samples that were combined with the process spectrum (Slab + Sp) and that for production samples [15], [17].

2.2.5. Construction of calibration models

Spectra were subjected to various pretreatments including the Standard Normal Variate (SNV), and the first and second derivatives. Derivative spectra were obtained by using the Savitzky–Golay and Norris algorithms with a moving window or Gap of variable size in addition to a second-order polynomial. All spectra were processed and multivariate calibration models constructed with the aid of the software Unscrambler v. 9.8 from CAMO (Trondheim, Norway).

Calibration models were constructed by cross-validation (leave-one-out method) using the PLS algorithm. The individual models exhibiting the lowest residual variance in terms of the number of latent variables were selected for refining, using the number of PLS factors leading to the smallest root mean square error of prediction (RMSEP) for an external set consisting of production samples

2.2.6 Validation of proposed calibration models

The potential industrial usefulness of the selected calibration strategies was assessed by validating their results in accordance with ICH and EMA guidelines [18], [19]. The specific parameters assessed included selectivity, linearity, accuracy, precision (repeatability and intermediate precision) and robustness.

2.3 RESULTS AND DISCUSSION

Obtaining effective calibration models for the target pharmaceutical preparation is made difficult by its low API content (10 $mg \cdot g^{-1}$), and the high spectral (0.879) and concentration correlation (0.939) between the API and sucrose —the major excipient, which accounts for 96% of the formulation—. The high correlations between these two components cannot be ignored owing to the high content in sucrose of the mixture and the additional high spectral correlation. In order to circumvent these shortcomings, we used various spectral pretreatments and wavelength ranges to obtain more simple models of adequate predictive ability.

Thus, shifts due to scattering were addressed by using the Standard Normal Variate (SNV) and spectral discrimination was improved by using derivative spectra obtained with the Savitzky–Golay (S.G) or Norris algorithm, which typically lead to simpler, better models. Using the first spectral derivative in combination with SNV in either sequence failed to improve the predictive ability (i.e., to decrease the RMSEP) of the models. On the other hand, a combination of second SNV– derivative (S.G) spectra was highly efficient in improving the calibration sets obtained with two of the three strategies. Also, the spectra pretreatment with SNV followed by derivatives led to slightly better predictions.

The use of second derivative from both Savitzky-Golay and Norris resulted to be the most suitable for processing the NIR spectra for this formulation. Since the nature of the samples is different for each strategy, it is presumed that one algorithm is more effective in one methodology than in the others due to the noise attenuation factor in each derivative. Therefore, the second derivatives were chose for further spectral pretreatment in this study [20].

As can be seen from **Fig 2**, the combination of the previous spectral pretreatments correct the shifts due to scattering effects and hence the differentiation amongst characteristic bands. Additionally, after these pretreatments were applied, the spectral correlation coefficients between the API and sucrose: to 0.080 with SNV + second derivative and 0.077 with 2D Norris + SNV were considerable reduced.. This reduction was expected to facilitate construction of effective calibration models and led us to adopt the two spectral treatments for further testing.

The importance of the order for applying the spectral pretreatments was presented by *Fern* [21], who suggested that when, a combination of SNV and derivatives is suitable for the extraction of spectral information, this must be perform in this order rather in the other way around, since the SNV corrects the scatter by dividing each spectrum by its standard deviation, which not will be the same if the derivative is perform firstly; moreover also this combination is advisable when it is desire to model chemically characteristics of the sample instead of physical. However the selection criterion to choose the most convenient spectral pretreatments was made based on the predictive ability of each calculated model with the different combinations.

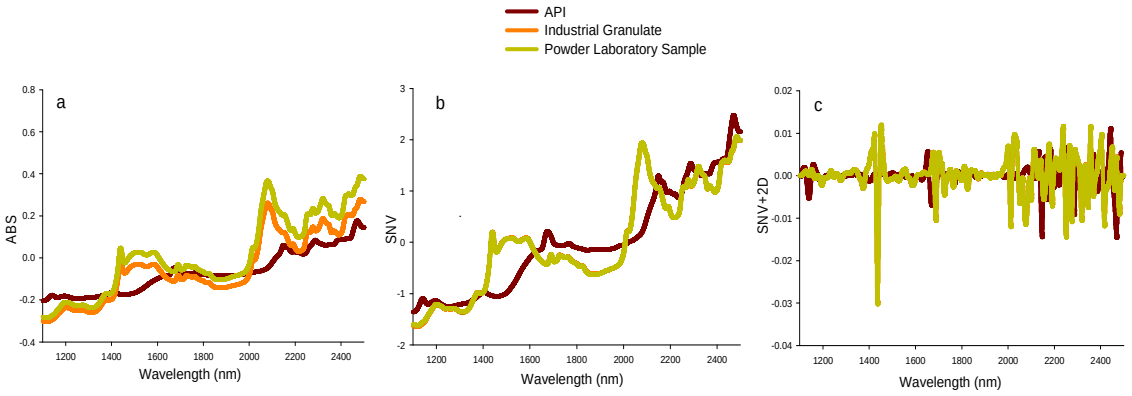


Fig2. NIR spectrum of industrial granulate, powder laboratory and calculated process spectrum a) Absorbance, b)SNV pretreated spectra and c) SNV + 2D (S.G) pretreated spectra

The figure mentioned above also illustrate that the spectrum for the API exhibits no strong characteristic bands and should therefore be used in its entirety. However, the reduced wavelength range from 2114 to 2488 nm, selected in terms of the regression coefficients and loading weights of the models, was used instead for potentially improved model accuracy and predictive ability. Application of the previous spectral treatments over the reduced wavelength range decreased the effects due to scattering and improved discrimination between the spectra for the API and sucrose.

One crucial step in constructing calibration models is selecting appropriate samples for inclusion in the calibration set. In fact, the sample set should contain all chemical and physical variability in production samples since such variability has a strong effect on NIR spectra (see spectral changes due to concentration increased **Fig. 3**). Also, the spectra for all samples should be recorded under identical conditions as those for the samples to be predicted in order to avoid introducing additional sources of variability such as the way spectra are acquired, their noise level or the equipment used. Therefore, all calibration sets should meet two essential requirements, namely: (a) the API contents of the samples should span the present concentration range; and (b) the sample set should contain all potential physical and spectral variability. Both requirements can be met with the three strategies used here.

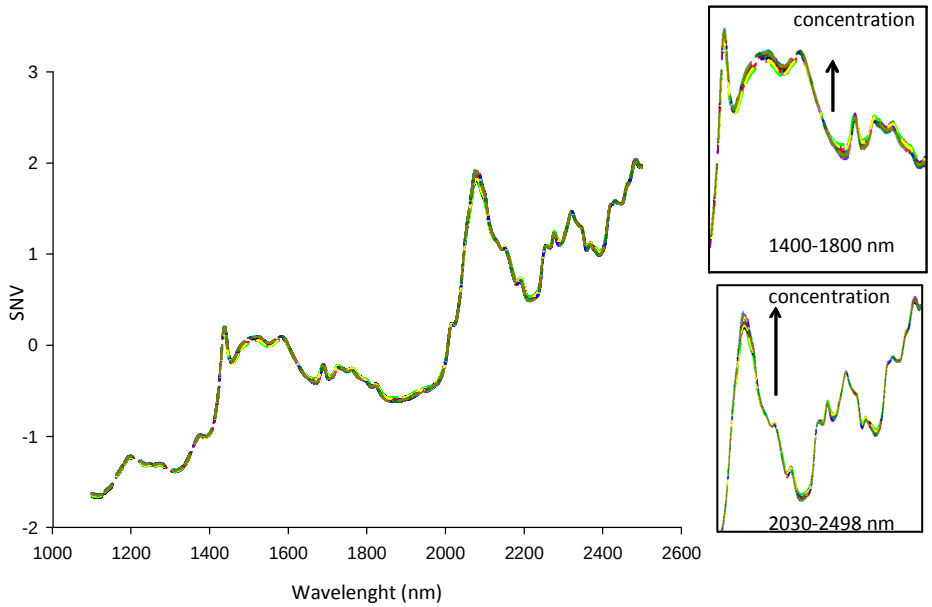


Fig3. Spectral changes due to API concentration increase

One of the aims of this work was to develop calibration models based on a small number of samples in order to reduce experimental work but still large enough to ensure accurate prediction of industrial production samples. We used the three above-described strategies to prepare the samples for inclusion in the calibration set and compared their results. Then, we validated the ensuing models and selected the most suitable among them for use in routine production control analyses. The samples used to construct the models were selected among those described under Experimental and imposed the requirement that their concentrations should span a range ± 20% around the nominal value of the formulation. The number of samples used to construct the models were similar differed between strategies and was suited to the spectral range to be spanned and to physical variability in the manufacturing process. Thus, the calibration set comprised 18 samples with the doping strategy, 23 with the mixed calibration strategy (powder samples + industrial granulates) and 20 with the process spectrum strategy.

Most of the models constructed with the sample doping strategy provided accurate predictions of production samples; the best results were obtained by using SNV in combination with a second derivative treatment based on the Savitzky–Golay

algorithm with a 11-point moving window and a second-order polynomial; however it is important to point out that several combinations of the derivative configuration pretreatment were evaluated: number of points for the moving window-seven, nine and fifteen-and and polynomial order. **Table 1** shows the characteristics of selected models. The best results were provided by two models using the whole wavelength range and the reduced range (2114–2488 nm), with a prediction error (RMSEP) of 0.276 and 0.228 $mg \cdot g^{-1}$ (RSEP%= 2.812, 2.321), respectively. Although the predictive ability for both models goodness of predictions was similarly good, the model using the whole wavelength range required 4 factors for calculation and accounted for 98.7% of the predictive variance, whereas that using the reduced wavelength required 6 factors and accounted for 99.9% of the variance. Besides the Bias for the model with the whole spectral range is lower than the one using the short wavelength (-0.008 and 0.054 respectively) demonstrating the accuracy of it. We selected the former because it was simpler —it used fewer factors— and yet provided accurate predictions.

The calibration models obtained from laboratory samples and industrial granulates (i.e., mixed calibration sets) were constructed from 13 powder samples containing amounts of API and excipients obtained by weighing and 10 granulate samples whose API contents were determined by HPLC. **Table 1** describes the models obtained with different spectral treatments. As can be seen, the model based on SNV in combination with a second-derivative treatment with the Norris algorithm (Gap size = –1, second-order polynomial) provided the most accurate predictions (RMSEP = 0.187; RSEP%=1.902). The model used 5 factors and accounted for 98.7% of the Y-variance. Using this strategy with the reduced wavelength range led to slightly better calibration results; however, the ensuing model was more complex —it required 6 factors—, accounted for 98.9% of the Y-variance and failed to improve predictions (RMSEP-RSEP).

For the strategy based on calculation and addition of the process spectrum involved obtaining the process spectrum the plot revealed that the clusters of production samples and (Slab +Sp) samples were rather distant and hence that the two sample clusters were rather different. This may have resulted from Sp not

being representative of the whole physical variability in the process. Such variability was increased/decreased by using *factor m* (m = 1 or 0.5). Multiplying Sp by *m* provided a cluster of production samples surrounded by the laboratory samples combined with the process spectrum (Slab + mSp). **Fig.4** shows the scores plot of the spectra for calibration and production samples.

Table 1 summarizes the figures of merit of the ensuing models. As can be seen, the model with the highest predictive ability (RMSEP = 0.225; RSEP%=2.295) was that constructed by using powder samples and powder samples + mSp (with m = 1 or 0.5), 6 factors, the whole wavelength range and a Savitzky–Golay second-derivative treatment with a 3-point moving window and a second-

Table1. Figures of merit of calibration model

	Doping strategy				Mixed calibration strategy				Calculation and addition of process spectrum strategy			
*1100-2498 nm **2104-2498 nm	Calibration 18 Samples			Prediction 41 samples	Calibration 23 Samples			Prediction 41 samples	Calibration 20 Samples			Prediction 41 samples
	%Exp Y-var	#Factors	RSEC	RSEP(%)	%Exp Y-var	#Factors	RSEC	RSEP(%)	%Exp Y-var.	#Factors	RSEC	RSEP(%)
SNV+1D (S.G)*	98.7	4	1.332	4.190	97.6	9	1.346	7.280	99.5	7	0.814	16.68
SNV+2D (S.G)*	**98.7**	**4**	**1.326**	**2.812**	98.3	8	1.119	4797	**99.5**	**6**	**0.801**	**2.295**
SNV+2D (S.G)**	99.9	6	0.459	2.321	98.3	8	1.268	6.292	99.7	6	0.783	3.230
2D(Norris)+SNV*	99.1	4	1.120	3.520	**98.7**	**5**	**1.016**	**1.902**	99.3	5	0.979	8.890
2D(Norris)+SNV**	99.4	4	0.922	3.842	98.9	6	0.916	3.294	99.5	5	0.842	4.870

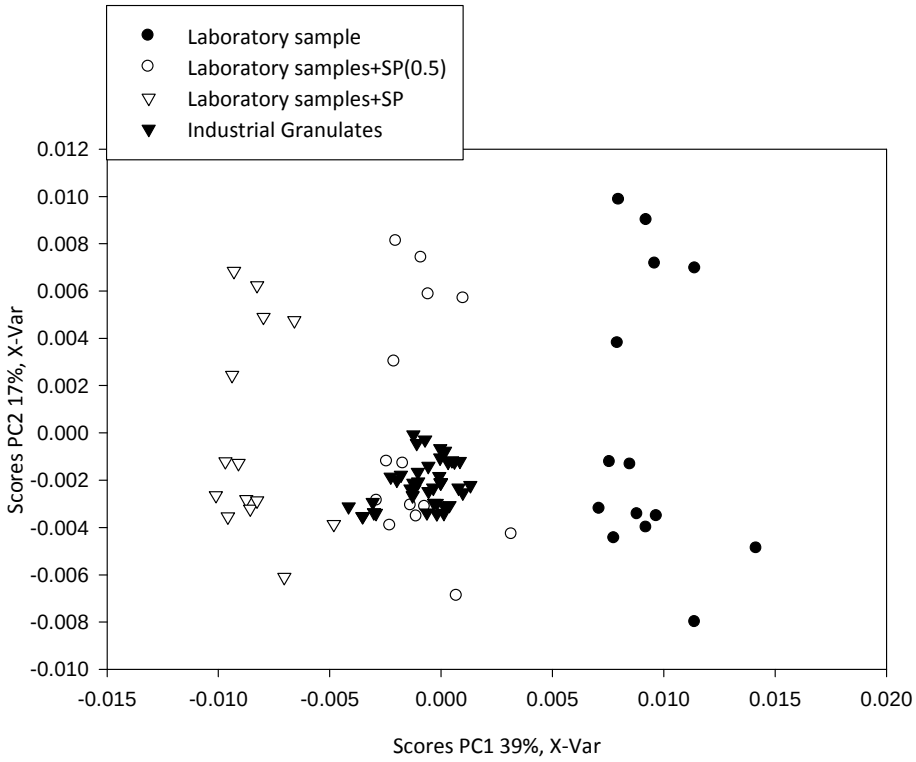


Fig4. Projection of industrial samples in PCA score plot of laboratory samples and laboratory samples+SP

3.1. Validation

The potential industrial usefulness of the selected calibration strategies was assessed by validating their results in accordance with ICH and EMA guidelines [13], [14]. The specific parameters assessed included selectivity, linearity, accuracy, precision (repeatability and intermediate precision) and robustness.

Selectivity of the proposed NIR methods is achieved by identifying the pharmaceutical preparation in a spectral library [22]. The library allows the identification of the pharmaceutical preparation rather than the pure raw materials (active ingredient and excipients) using a supervised pattern recognition method (PRM) criterion. The PRMs rely on similarity measurements, where similarity here is taken to the extent to which an object (spectrum) is identical to one another. While unsupervised methods search

for clustering in a N-dimensional space without knowing the class to which the sample belongs, the supervised methods depend on a previous training of the system using a set of objects belonging to the specific, previously known class. In this work, the applied identification criterion was the residual variance of the principal components, which is primarily based on a spectra PCA calculation defining the model of the known class and establishing a threshold as an indicative value. Then, the spectrum from the sample to be analyzed is reconstructed through the created PC´s and the obtained residuals are used to calculate the sample's probability of belonging to the known class [22]. The library was constructed using 21 granulates belonging to different production batches and the software Vision v2.20-2.51 (FOSS NIRSystem, Silver Spring, USA). The second-derivative spectra (S.G), the wavelength range 1100–2488 nm and a threshold of 0.94 (positive identification for values lower than this value) were used. All 40 samples were successfully identified showing values between 0.645 and 0.933. The values for the pure raw materials were 0.99 and higher.

Linearity was assessed by using 11 samples uniformly spanning the working concentration range (viz., ± 20% around the nominal value) to quantify the API with the three NIR strategies. A plot of responses against reference values had a slope and intercept containing unity and zero, respectively, at the 95% confidence level.

Accuracy was assessed as the degree of agreement between reference and NIR values for 23 samples spanning the working concentration ranges. The sample doping strategy showed a higher Bias (-0.165) while for the mixed calibration and the calculation and addition of process spectrum strategies this value was similar (-0.028 and -0.025 respectively). A t-test on the residuals confirmed the absence of significant differences between methods at the 95% confidence level with each of the three strategies.

Precision was assessed as repeatability and intermediate precision. Repeatability was determined by having the spectrum for an industrial granulate recorded by the same

analyst six times and calculating its relative standard deviation (%RSD) in order to quantify the coefficient of variation for each method. The highest %RSD value was that obtained with the strategy involving the process spectrum. Intermediate precision was determined by having the NIR spectra for an industrial granulate sample recorded by two different analysts on 3 different days. An analysis of variance (ANOVA) with 2 factors (analyst and day) revealed the absence of significant differences from the reference method.

Robustness was assessed by predicting the values for a second set of industrial granulates consisting of 34 samples. The samples were obtained from various production batches manufactured after the first sample set studied and analysed by using the previously developed models a few months after the samples used to construct the models were measured. The results testified to the good predictive ability of the three models for external samples not included in the calibration set and measured after development of the models.

Table 2 summarizes the results for each parameter. Based on them, the proposed strategies meet all validation requirements set in the above-mentioned guidelines.

Table2. Results obtained from calibration model validation

	Parameter	Strategy under and overdosage	Mixing set Calibration strategy	Strategy Calculation and addition of process spectrum
Linearity	*n*	11	11	11
	Concentration range (mg.g^{-1})	7.98-11.80	8.02-11.90	8.27-12.03
	Intercept	0.07±1.80	0.21±1.57	0.05±1.45
	Slope	1.01±0.18	1.04±0.16	1.0±0.15
	R	0.972	0.98	0.982
Accuracy	*n*	23	23	23
	Bias	-0.165	-0.028	-0.025
	S.D	0.266	0.162	0.185
	t. Experimental	0.725	0.204	0.16
	t.Critic	2.074	2.074	2.074
Repeatability	*RSD(%)*	1.099	1.754	3.001
	Mean NIR(mg.g^{-1})	9.531	10.300	10.284
Intermediate precision	*Day*			
	F. Experimental	0.591	1.354	0.192
	F. Critic	19	19	19
	RSD(%)	0.103	0.632	1.704
	Analyst			
	F. Experimental	0.057	1.04	0.015
	F.Critic	18.51	18.51	18.51
	RSD(%)	0.119	0.742	1.508
Robustness	*n*	34	34	34
	RMSEP (mg.g^{-1})	0.254	0.500	0.239
	Bias	-0.017	-0.391	0.008
	S.D	0.257	0.316	0.243
	t. Experimental	0.08	1.473	0.039
	t.Critic	2.035	2.035	2.035

2.3.2. Comparison of calibration models obtained with the three strategies

The three calibration models confirm their ability to predict industrial granulates; therefore the proposed strategies successfully incorporate the whole variability of the industrial samples. Projecting the results for industrial granulates on a scores plot for the calibration samples used with each strategy revealed that the calibration set contained all industrial samples and hence that the proposed strategies allows effective calibration sets to be constructed with a view to developing accurate models for quantitation of the API (see **Fig.5**) [23]. This was particularly so with the strategy based on the process spectrum, which was the most efficient in including the production cluster in the calibration set.

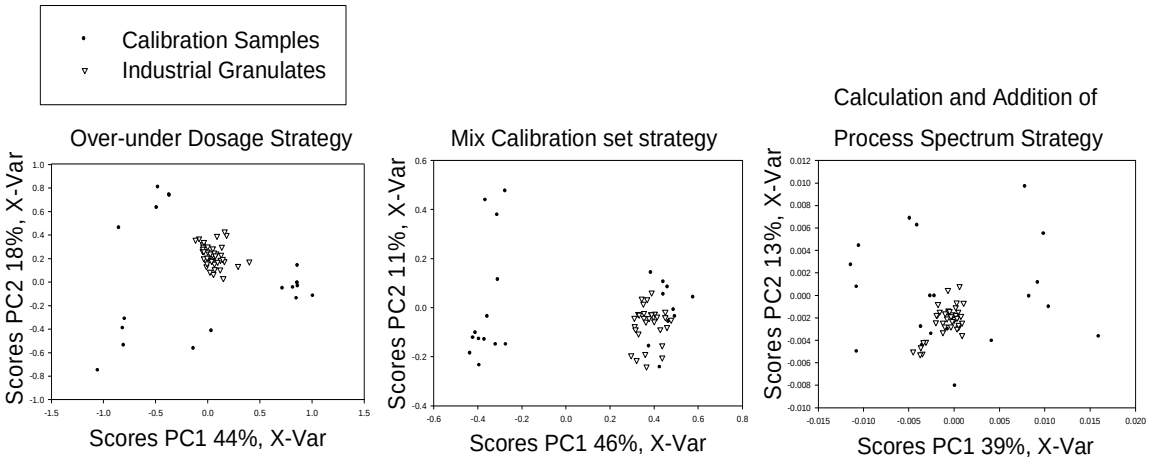


Fig5. Projection of industrial granulates in calibration PCA scores plot of used strategies

The optimum number of PLS factors for constructing the models with the three strategies ranged from 4 to 6; also, the bias and their standard deviations were similar for all models.

The model based on mixed calibration sets and the first set of industrial granulate samples was that providing the most accurate predictions (RMSEP = 0.187; RSEP%=1.902), followed by those based on doped samples (RMSEP = 0.276;

RSEP%=2.812) and addition of the process spectrum (RMSEP = 0.225; RSEP%=2.295), which performed similarly in this respect. However, the analysis of a second set of production samples led to substantially increased prediction statistics with the model using mixed calibration sets (RMSEP = 0.500; RSEP%=5.029), but essentially unchanged statistics with those based on doped samples (RMSEP= 0.254; RSEP%=2.552) and the process spectrum (RMSEP = 0.239; RSEP%=2.407). All RMSEP values, however, were good enough to afford application of the three strategies to the industrial manufacturing process.

One other major factor in choosing a particular methodology is the amount of experimental work needed to prepare samples for inclusion in the different models. Thus, laboratory samples are prepared by weighing of their components, which facilitates obtainment of high-quality reference values; on the other hand, industrial samples must be analysed with a reference method that may be complicated and sluggish, and will certainly be less accurate and precise than weighing.

The mixed model using laboratory and industrial samples ensures incorporation of all physical variability via appropriate industrial samples and all chemical variability via laboratory samples. Although the model requires using an increased number of samples, it can be expected to provide high-quality predictions of industrial samples. However, the obtained experimental data showed a considerable increase of errors amongst the two tested groups of granulates, implicating the robustness of this strategy. The model using doped industrial granulates ensures inclusion of physical and chemical variability; however, this methodology presents some drawbacks as the sample preparation process is time-consuming and reference values are less accurate as they are obtained by application of the reference method and consideration of the amount of API or placebo added —or alternatively, by analysing all samples with the reference method, which means an extra experimental effort. Moreover, homogenizing the small amounts of components added to the samples can be quite difficult.

The strategy involving calculation and addition of the process spectrum requires using no reference method to construct the models as reference values can be obtained simply by weighing —time saving and less experimental work and hence more accurately than with a reference method. This strategy affords optimal incorporation of physical and chemical variability by virtue of sample preparation, and calculation and addition of the process spectrum, being two simple, expeditious processes.

2.4. CONCLUSIONS

An API present in a pharmaceutical granulate in low concentration was successfully quantified using the three proposed methodologies. The chemical and physical variability was incorporated to the calibration sets through different strategies showing a good predictive ability and the developed analytical methodologies were validated according to the normative (EMA& ICH). The calculation and addition of process spectrum methodology was chosen as the most suitable strategy for the purpose due to a higher performance in terms of robustness, easy inclusion of variability in the samples without reference method and less experimental work.

REFERENCES

[1] C. P. Meza, M. a Santos, and R. J. Romañach, "Quantitation of drug content in a low dosage formulation by transmission near infrared spectroscopy.," *AAPS PharmSciTech*, vol. 7, no. 1, p. E29, Jan. 2006.

[2] L. Alvarenga, D. Ferreira, D. Altekruse, J. C. Menezes, and D. Lochmann, "Tablet identification using near-infrared spectroscopy (NIRS) for pharmaceutical quality control.," *J. Pharm. Biomed. Anal.*, vol. 48, no. 1, pp. 62–9, Sep. 2008.

[3] M. Blanco, R. Cueva-Mestanza, and a Peguero, "Controlling individual steps in the production process of paracetamol tablets by use of NIR spectroscopy.," *J. Pharm. Biomed. Anal.*, vol. 51, no. 4, pp. 797–804, Mar. 2010.

[4] A. Porfire, L. Rus, A. L. Vonica, and I. Tomuta, "High-throughput NIR-chemometric methods for determination of drug content and pharmaceutical properties of indapamide powder blends for tabletting.," *J. Pharm. Biomed. Anal.*, vol. 70, pp. 301–9, Nov. 2012.

[5] M. Blanco and I. Villarroya, "NIR spectroscopy: a rapid-response analytical tool," *TrAC Trends Anal. Chem.*, vol. 21, no. 4, pp. 240–250, Apr. 2002.

[6] G. Reich, "Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications.," *Adv. Drug Deliv. Rev.*, vol. 57, no. 8, pp. 1109–43, Jun. 2005.

[7] M. Sarraguça Cruz and J. Lopes Almeida, "Quality control of pharmaceuticals with NIR: From lab to process line," *Vib. Spectrosc.*, vol. 49, no. 2, pp. 204–210, Mar. 2009.

[8] N. Qu, M. Zhu, H. Mi, Y. Dou, and Y. Ren, "Nondestructive determination of compound amoxicillin powder by NIR spectroscopy with the aid of chemometrics.," *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.*, vol. 70, no. 5, pp. 1146–51, Oct. 2008.

[9] T. De Beer, A. Burggraeve, M. Fonteyne, L. Saerens, J. P. Remon, and C. Vervaet, "Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes.," *Int. J. Pharm.*, vol. 417, no. 1–2, pp. 32–47, Sep. 2011.

[10] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, and N. Jent, "A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies.," *J. Pharm. Biomed. Anal.*, vol. 44, no. 3, pp. 683–700, Jul. 2007.

[11] M. Jamrógiewicz, “Application of the near-infrared spectroscopy in the pharmaceutical technology.,” *J. Pharm. Biomed. Anal.*, vol. 66, pp. 1–10, Jul. 2012.

[12] B. R. M. A. B. M., “Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries,” in *Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*, Second Edi., John Wiley & Sons, Ltd, 2010, pp. 463–491.

[13] M. Blanco, M. Bautista, and M. Alcala, “Preparing Calibration Sets for Use in Pharmaceutical Analysis by NIR Spectroscopy,” vol. 97, no. 3, pp. 1236–1245, 2008.

[14] M. Blanco, M. Romero, and M. Alcalà, “Strategies for constructing the calibration set for a near infrared spectroscopic quantitation method.,” *Talanta*, vol. 64, no. 3, pp. 597–602, Oct. 2004.

[15] M. Blanco and A. Peguero, “Analysis of pharmaceuticals by NIR spectroscopy without a reference method,” *TrAC Trends Anal. Chem.*, vol. 29, no. 10, pp. 1127–1136, Nov. 2010.

[16] R. J. Romañach, Roman-Ospino Andrés D., and M. Alcalà, *A Procedure for Developing Quantitative Near Infrared (NIR) Methods for Pharmaceutical Products*. New York, NY: Springer New York, 2016.

[17] M. Blanco, R. Cueva-Mestanza, and A. Peguero, “NIR analysis of pharmaceutical samples without reference data: improving the calibration.,” *Talanta*, vol. 85, no. 4, pp. 2218–25, Sep. 2011.

[18] International conference on harmonisation of technical requirements for registration of pharmaceuticals, “ICH Guidelines: Q2(R1) Text on validation of analytical procedures: methodology Q2(R1),” 1996. [Online]. Available: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf.

[19] European Medicines Agency, “Guideline on the use of Near Infrared Spectroscopy (NIRS) by the pharmaceutical industry and the data requeriments for new submission and variations.,” 2012. [Online]. Available: EMEA/CVMP/QWP/17760/2009rev2.

[20] Å. Rinnan, F. Van Den Berg, and S. B. Engelsen, “Review of the most common pre-processing techniques for near-infrared spectra,” *TrAC Trends Anal. Chem.*, vol. 28, no. 10, pp. 1201–1222, Nov. 2009.

[21] T. Fearn, “The interaction between standard normal variate and derivatives,” *NIR News*, vol. 19, no. 7, p. 16, 2008.

[22] M. Blanco and M. a. Romero, “Near-infrared libraries in the pharmaceutical industry: a solution for identity confirmation,” *Analyst*, vol. 126, no. 12, pp. 2212–2217, Dec. 2001.

[23] C. De Bleye, P.-F. Chavez, J. Mantanus, R. Marini, P. Hubert, E. Rozet, and E. Ziemons, “Critical review of near-infrared spectroscopic methods validations in pharmaceutical applications.,” *J. Pharm. Biomed. Anal.*, vol. 69, pp. 125–32, Oct. 2012.

Case of study II

Strategy for design NIR calibration sets based on process spectrum and model space

An innovative approach for process analytical technology

In this work, we developed calibration models useful in the quality control of a pharmaceutical formulation during its three manufacturing stages, namely: blending (powder), pressing (cores) and coating (tablets). A novel methodology is proposed for selecting the calibration set, the so called "process spectrum" strategy into which physical changes in the samples at each stage are algebraically incorporated.

Also, we established the concept of "model space", which is defined by Hotelling's t^2 and q-residuals. These statistics allow outlier identification inside and outside the model space in order to facilitate more objective selection of the factors to be used in constructing the calibration set.

Published article:
V. Cárdenas, M. Cordobés, M. Blanco, M. Alcalà,
Strategy for design NIR calibration sets based on process spectrum and model space: An innovative approach for process analytical technology
Journal of Pharmaceutical and Biomedical Analysis
October 2015, Volume 114, Issue 10, pp 28-33

3.1 INTRODUCTION

Quality control by the pharmaceutical industry has traditionally relied on assessment of the raw materials prior to processing and analytical determinations of the end-product. Although this methodology usually allows product quality regulations to be met, errors or unexpected variability arising at some stage of the process may not be detected before reaching the end-product and lead to time and money losses in addition to diminished productivity.

In recent years, the US FDA has encouraged the use of process analytical technology (PAT) by the pharmaceutical industry. PAT is intended to assure product quality via careful design, monitoring, control and surveillance of each manufacturing stage. With this methodology, quality in the product and efficiency in the production process result from a deep knowledge of the process and strict control of any physical, chemical and quality-related factors influencing each stage. Quality in pharmaceutical production processes cannot be assured merely by analysing raw materials and end-products; rather, it requires carefully designing and implementing each production stage [1], [2].

There is ample evidence of the usefulness of near infrared spectroscopy (NIRS) as a pharmaceutical process control analytical methodology. In fact, NIRS is a simple, expeditious, non-destructive instrumental technique [3]–[7] and NIR spectra provide both physical and chemical information about solid samples. As a result, its use combine with chemometrics data processing have turned it into a promising tool for process control within the framework of PAT.

More than 80% of all pharmaceutical formulations are available in tablet form. Tablet manufacturing processes are usually complex and involve several steps that can introduce different sources of variability. As a result, assessing tablet quality entails

determining a number of critical attributes at each production stage, in this sense the development and implantation of NIR methodologies in pharmaceutical processes is requested in order to determine critical parameters that affect directly to the quality of products [8], [9].

Tablet production processes involve several stages whereby the raw materials are subject to various —mostly physical— treatments. The first stage is blending of the raw materials, which are usually in powder form. The resulting uniform blend is then pressed to obtain usually oval samples (cores). Finally, cores are coated in order to facilitate preservation under ambient conditions, conceal unpleasant odours or flavours, or ensure appropriate release of the pharmaceutical.

Controlling the amount active principal ingredient (API) at all stages of the production process is crucial because it influences not only the quality of the end-product but also consumers' health. In this work, we used NIRS for quality control of a tablet manufacturing process by quantifying the API at the three production stages. To this end, we constructed a different PLS calibration model for each stage by using a simple, novel approach requiring no reference method to select the calibration set. The proposed approach incorporate the variability of production samples to the calibration set via an algebraic procedure involving addition of the *process spectrum*. Also, it uses the *model space* (a new concept based on the statistics Hotelling's T^2 and *Q*-residuals [10]) to optimize the sample selection process and facilitate construction of the calibration model.

Based on the definition that an outlier is considered as "an observation that does not fit to a pattern", these statistics were used precisely to spot those samples that do not have the same spectral characteristics to the samples that want to be predicted. Hotelling's T^2 and *Q*-residuals statistics are very useful to spot X outliers – related to analytical profiles- since they are calculated with the **T** (scores) and **E** (residuals) values obtained by the deconvolution of the X matrix –spectra- by PCA means. In this sense T^2 statistics reflects the extremeness of the samples response within the PCA

model space, while Q values reflects the amount of sample response that is outside of the PCA model space [10], [11].
The use of this two quality-metrics parameters combined allow the creation the model space for a fully evaluation of all the parameters that interfere with the use of the proposed methodology.

3.2 EXPERIMENTAL SECTION

3.2.1. Production samples

The target formulation contained 8.7 wt% API (cetirizine) and four excipients accounting for more than 90 wt% in combination, namely: lactose, microcrystalline cellulose, magnesium stearate and colloidal silica. The oval tablets were 10 × 4 mm. samples of the pure components and of the products processed at each stage [viz., powder, cores and coated tablets (the end-product)] were kindly supplied by Laboratories Menarini, SA (Badalona, Spain).The API contents of the samples were determined by HPLC.

3.2.2. Laboratory samples

A set of 23 laboratory samples was prepared by blending appropriate amounts of API and placebo (the excipient mixture). As per the ICH guidelines [11], the API concentration was expanded ± 20% around its nominal value. Three different placebo blends spanning concentration values ±5% around the nominal amount of each excipient in the formulation were also used. Placebos were prepared by using a d-optimal design in order to minimize collinearity between excipient concentrations.

3.2.3. Recording of NIR spectra

Laboratory samples were blended in a T2C WAB shaker mixer, and their NIR reflectance spectra recorded on a model 5000 spectrophotometer equipped with a rapid content analyser (RCA) and governed via the software vision v. 2.22, all spectra were recorded using Foss NIRsystems. at 2 nm intervals over the wavelength range 1100–2500 nm.

The spectra for the powder samples were recorded in glass cuvettes and samples turned over with a spatula between recordings. Cores and tablets were also analysed in this manner. a blank spectrum for the empty cuvette was recorded at the beginning of each working session. The spectral reference used was the instrument's bundled ceramic plate.

3.2.4. Preparation of the calibration set by calculating and adding process spectra

The process involved calculating the process spectrum (Sp) as the difference between an industrial spectrum (Sind) and that for a laboratory powder sample spectrum (Slab) containing the same concentration API —and also, ideally, of excipients. the difference was expressed mathematically as a vector defining variability in the production process:

$$Sp = Sind - Slab \quad (1)$$

Then, the reduced process variability spectral matrix (Sred) was calculated by adding the process spectrum to the spectral matrix for various powder blends spanning the desired API concentration range (Scon):

$$Sred = Sp + Scon \quad (2)$$

The “new” matrix contained both physical and chemical variability, and was useful as such to selected spectra for the calibration set. Variability in Sp can be increased or decreased by using a multiplying factor m close to unity. Simply adding up the spectra as described above provides an “extended variability matrix” Sred*m [12]–[14].
By using the m factor in this strategy the variability can be increased according to the predictive ability of the calibration set, which is checked by a projection of the samples into a PCA of the calibration set.

3.2.5. Definition of the model space and detection of outliers

In order to determine if the characteristics of the calibration set fall within the variation range of the samples to be predicted, it is required to define a model space allowing suitable samples to be selected for its construction, and any not belonging to the population defining the model to be excluded [10], [15].

The model space was defined from a principal component analysis (PCA) of industrial samples (cores and coated tablets), using the scores T and residuals E of the deconvoluted matrix to calculate Hotelling’s t^2 and q-residuals with $p = 0.05$. Samples for the calibration set were selected and outliers identified from the score scatter plot for the calibration samples provided by the PCA, using Hotelling’s ellipse at the 95% confidence level and a plot of t^2 Hotelling vs q residuals.

3.2.6. Construction of calibration models

Spectra were subjected to various treatments including the standard normal variate (SNV), and the first and second derivative as calculated with the Savitzky–Golay algorithm. spectral treatments were applied and multivariate models constructed by using the software the Unscrambler v. 9.8 from camo (Trondheim, Norway).
The PLS algorithm was used to construct calibration models by cross-validation (leave-one-out method) and the model exhibiting the lowest residual variance with

the number of latent variables selected for further testing.
Calibration models were refined by using the number of PLS factors leading to the lowest relative square error of prediction (RSEP) and root mean square error of prediction (RMSEP) for an external set of production samples.

3.3 RESULTS AND DISCUSSION

Assuring quality in a pharmaceutical end-product and optimizing its production process entails critically controlling and monitoring each potentially influential factor at all manufacturing stages. The API content is one of such factor. In this work, we constructed three different calibration models to quantify the API present in a proportion of 8.7% (w/w)-in the target pharmaceutical formulation in its three forms during the production process, namely: powder, cores and coated tablets.

Ensuring adequate predictive ability obviously required using an appropriate set of calibration samples containing similar variability to the industrial samples aimed to be predicted. This required incorporating chemical variability by using a wide enough range of API concentration to allow accurate quantitation of extreme samples and physical variability by considering spectral changes arising from physical changes in the samples by effect of the different stages of the production process. Monitoring the API content of the target pharmaceutical formation throughout the process required careful consideration of physical changes in the samples owing to their influence on NIR spectra and hence on the ensuing models.

A preliminary spectral analysis of the pharmaceutical formulation was conducted on samples from the three production stages. The samples exhibited substantial differences in their NIR spectra by effect of physical changes due to differences in scatter between each process stage. This variability can be explained to differences in particle size between the powder samples and to the effects of compaction to obtain

cores and lacquering to obtain the end-product: coated tablets. Core compaction cause spectral shifts, whereas lacquering reduced the amount of light reaching the sample through partial absorption in the coating (see **Fig. 1a**). A principal component analysis (PCA) of the results revealed the presence of three distinct clusters of samples despite their having an identical chemical composition (**Fig. 1b**).

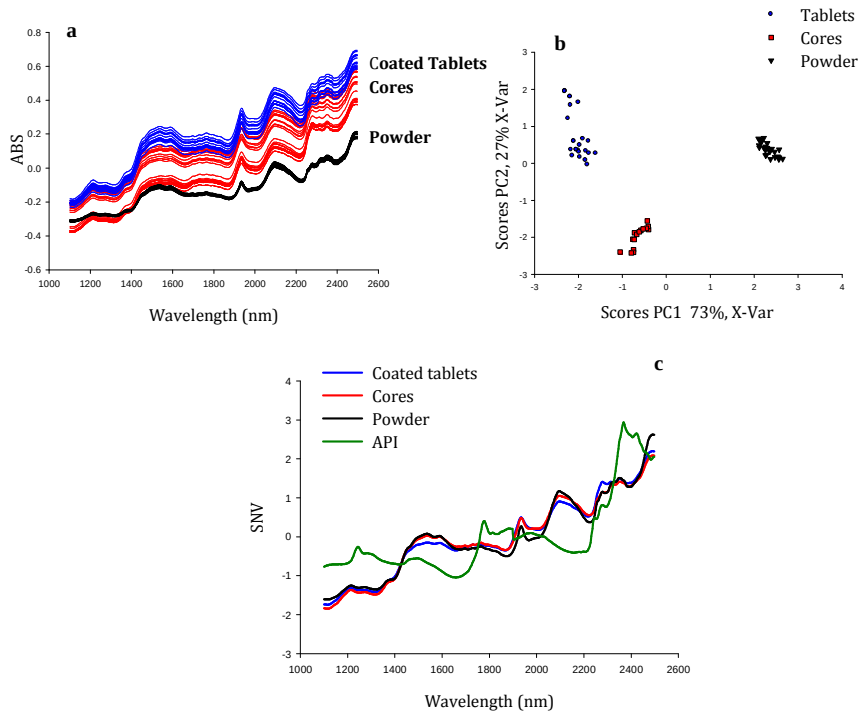


Fig1. (a) NIR Spectra and (b) score scatter plot (SNV) for the target formulation at each production step: powder, cores and coated tablets. (c) SNV spectra from formulation in different forms and API

Initially, the formulation was a powdered blend of the API and excipients. In the absence of chemical interactions between components, the blend was assumed no to require incorporating physical variability into its model, so only chemical variability was considered in the calibration set. This was accomplished by preparing 23 mixtures containing variable amounts of API and excipients that were expanded with others containing concentrations ± 20% and ± 5%, respectively, around the nominal

values in the formulation. The resulting samples were split between a calibration set and a validation set in a ratio about 70/30.

Different spectral pretreatments were tested; Spectral shifts due to scattering were reduced by using a standard normal variate (SNV) treatment and resolution was improved by calculating the second derivative with the Savitzky–Golay algorithm (11 points moving window) although several combinations of the derivative configuration pretreatment was evaluated: number of points for the moving window-seven, nine and fifteen-and the polynomial order. Using second-derivative spectra in combination with the whole spectral range (1100–2500 nm) allowed a simple model consisting of only 4 factors and explaining 99% of the *Y*-variance to be constructed. The model exhibited a good predictive ability: RSEP = 0.812%, RMSEP = 0.071 % w/w and bias = 0.040.

3.3.2. Incorporating physical variability: core and tablet models

Core compaction and tablet coating cause physical changes reflecting in the NIR spectra. Such changes were deemed "physical variability due to the process" and incorporated into the calibration set by calculating and adding the process spectrum. This approach allows one to consider both physical and chemical variability in three steps, namely: (1) calculating the process spectrum as the difference between that for an industrial sample and a powder sample of identical chemical composition; (2) adding the process spectrum (S_p) to a matrix S_{con} obtained from the spectra for powder samples of known API and excipient concentrations spanning the desired range, the resulting matrix, S_{red}, containing both physical and chemical variability; and (3) increasing or decreasing the variability of S_{red} by multiplying S_p by a factor m greater or lesser than unity in order to ensure that the process spectrum will be representative of the whole physical variability in the process.

Previous studies revealed the usefulness and robustness of this approach to

incorporating the whole variability of the process into the calibration set [12] and its ability to provide simple models with a good predictive ability with substantially reduced experimental work and the need for no reference method. However, some critical aspects for the development of this methodology such as selecting the multiplying factor (m) and samples for the calibration set, and confirming that all physical variability was considered by adding S_p, relied exclusively on a PCA of industrial samples and their projections onto the space for the selected calibration set. In order to objectively select those factors governing performance in this calibration methodology, we introduced the concept of *model space*. In this work, the model space was constructed from a PCA of industrial samples (cores or coated tablets). In parallel, critical thresholds were calculated from the statistics Hotelling's T^2 and Q-residuals by relating the scores **T** and residuals **E**, respectively, of the deconvoluted matrix. The information thus obtained was complementary, and the two factors allowed outliers falling inside and outside the space to be identified [8]. A projection of samples onto the PCA model afforded their evaluation in the space defined by the samples to be predicted, and hence confirmation of whether they were suitable for inclusion in the calibration set.

The second stage of the process (core compaction) was modelled by PLS regression on the 23 samples previously used to construct the model for the powder. Their spectra were used to obtain a matrix containing chemical variability (S_{con}). The spectral matrix was projected onto the space defined by the first two PCA scores for the core spectra. As can be seen from **Fig. 2a**, the spectra belonged to two different populations; also, the first factor explained physical differences between the two groups of samples, which suggest that the spectra were unsuitable for constructing an accurate calibration model for the cores. As can clearly be seen from **Fig. 2b**, the powder samples fell beyond the critical thresholds for Hotelling's T^2 and Q-residuals as calculated at the 95% probability level, thereby confirming the previous results for the projection.

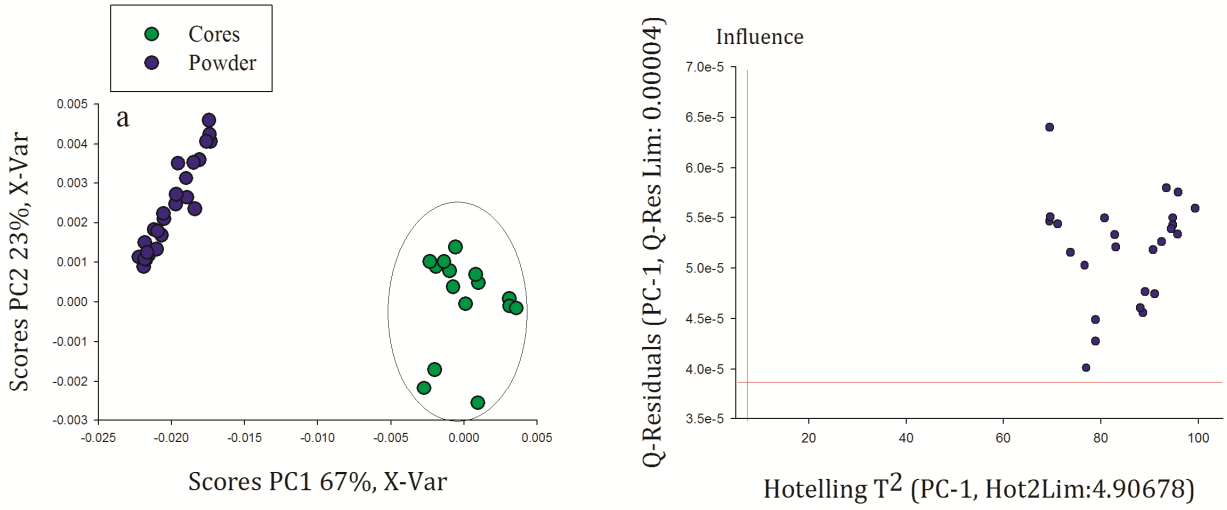


Fig2. (a) Definition of the model space in terms of Hotelling's ellipse at the 95% probability level and projection of powder samples into cores PCA space. (b) Critical thresholds of the model space as calculated from Hotelling's T2 and Q-residuals, and projection of powder samples.

The process spectrum (S_p) was calculated from three samples on the outside on the PCA space in order to include all physical variability in the cores. Subtracting the spectra for such samples from a powder sample of the same concentration gave three process spectra including the physical variability of the cores (see **Fig. 3**). These spectra were randomly added to S_{con} in order to construct a matrix S_{red} including physical and chemical variability.

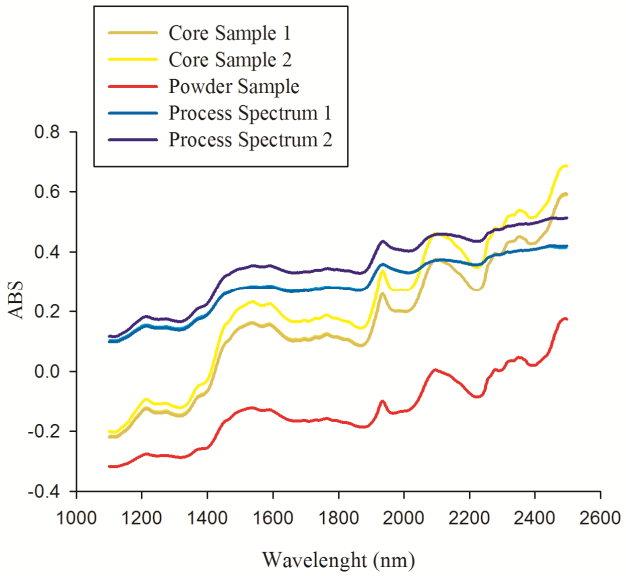


Fig3. NIR spectra for industrial (cores) and powder samples, and calculated process spectra.

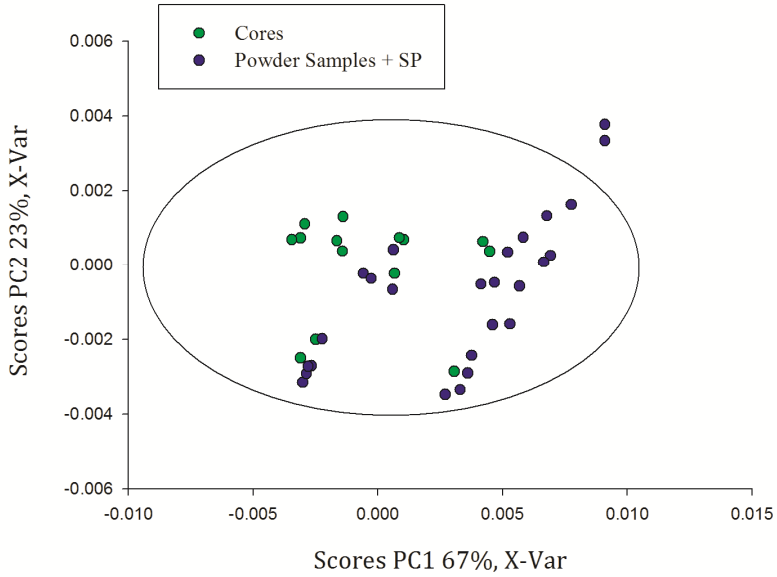


Fig4. PCA Score scatter plot and model space for the cores. Projection of the powder samples + Sp (Sred).

Projecting the scores of S_{red} onto the space defined by the spectra for the cores (**Fig. 4**) revealed that the samples were similar to the cores; however, some fell outside the cluster and were thus insufficiently represented by the model. Because constructing

an accurate model requires ensuring that the samples to be predicted will be included in the same cluster as the calibration samples, we introduced a new treatment involving multiplying the process spectrum by a factor m to obtain new matrix S_{red*m}. The optimum m value for this purpose was found to be 0.7. As can be seen from **Fig. 5**, the scores for the samples in S_{red*m} fell within the model space and comprised all cores. However, some samples fell outside Hotelling's ellipse and beyond the critical thresholds of the T^2 vs Q-residual plot, so they were excluded from the model.

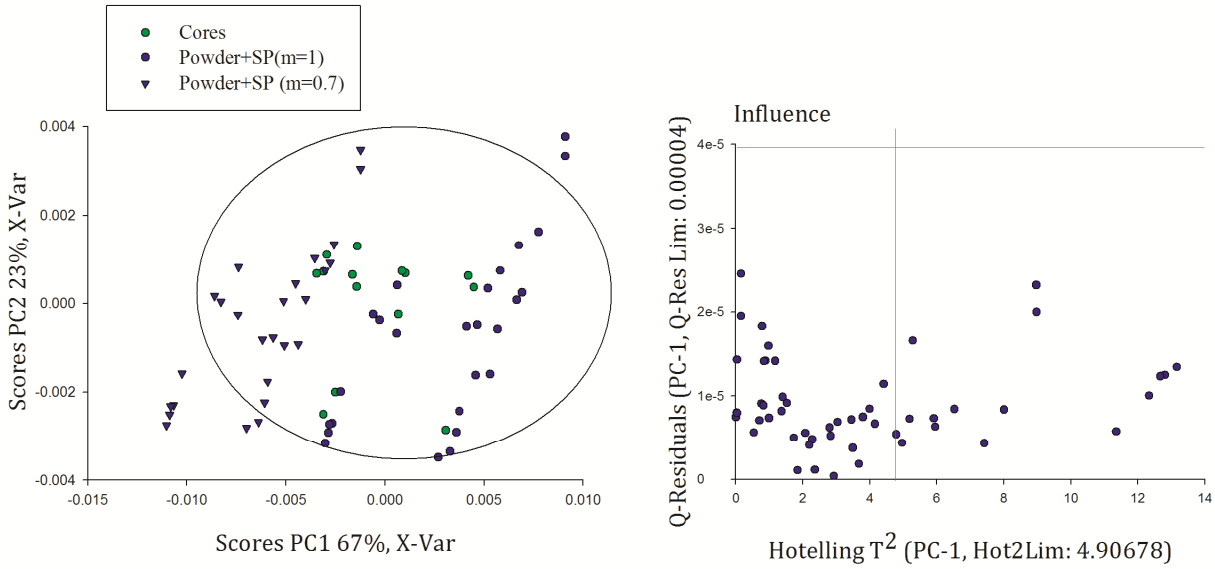


Fig5. Model space for cores (projection of powder samples + Sp, Sred*m0.7, 1). (a) Score scatter plot for the cores. (b) Critical thresholds as calculated from Hotelling's T2 and Q-residuals.

The calibration model was thus constructed from those samples falling within the model space (n = 37), using the first Savitzky-Golay derivative with 11 points, and second-order polynomial. Under these conditions, 7 PLS factors explained 98.3% of the Y-variance and the whole wavelength range. The results thus obtained in the analysis of 15 batches of production cores revealed a good predictive ability in the

model, with RSEP = 2.812%, RMSEP = 0.242 % w/w and bias = 0.060.

The calibration model for the last stage of the process (viz., obtainment of coated tablets) was constructed in the same manner as the previous one (viz., by defining the model space and selecting samples for calculation of S_p). The high variability observed in the distribution of scores for the production samples led us to calculate two process spectra that were added to the spectral matrix for the powder samples. Physical variability was introduced by using a multiplying factor *m* of 0.8. Then, the samples for inclusion in the calibration set were selected as described above. Each step was assessed and samples were selected from the score scatter plot (viz., the space bound by Hotelling's ellipse at the 95% probability level) and the T^2 vs *Q*-residual plot (see **Fig. 6**).

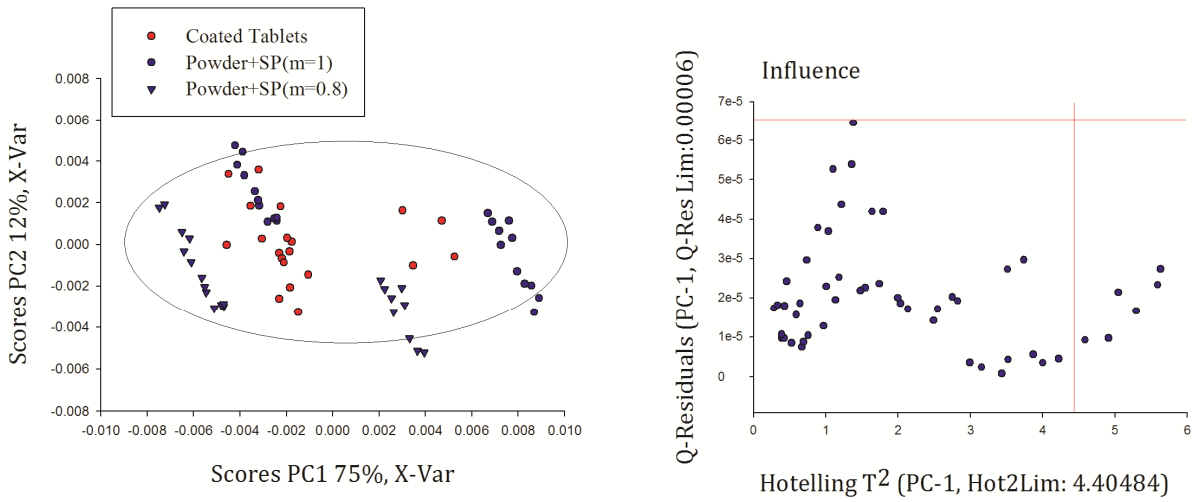


Fig 6. Model space for the tablets (p = 0.05) and projection of powder samples + Sp (Sred*m 0.8, 1).(a) Score scatter plot for coated tablets. Critical thresholds as calculated from Hotelling's T2 and Q-residuals.

The model was constructed from a total of 43 calibration samples, using first Savitzky-Golay derivative spectra consisting of 11 points ans second-order polynomial. 7 PLS factors were found to account for 98% of the Y-variance and the whole wavelength

range. The model was then applied to 25 coated tablets, which were predicted with quite good results: RSEP =
2.682%, RMSEP = 0.223 % w/w and bias = –0.016.

Table 1 summarizes the figures of merit of the models for the three stages of the pharmaceutical production process.

3.4. CONCLUSIONS

A simple, fast NIRS-based methodology for monitoring the API content of a pharmaceutical formulation throughout its production process was developed. Calculation and objective selection of the process spectrum in terms of a model space and the statistics Hotelling's T^2 and Q-residuals allowed optimal calibration sets to be constructed by using the proposed methodology, which should therefore be useful for quality control analyses in the pharmaceutical industry as it requires using no reference method, but only weighing, to quantify the API and excipients

REFERENCES

[1] P. A. T. A. Framework, "Guidance for Industry Guidance for Industry PAT — A Framework for Innovative Pharmaceutical," no. September, 2004.

[2] D. J.E, "Overview of Process Analysis and PAT," in *Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*, Second Edi., John Wiley & Sons, Ltd, 2010, pp. 1–14.

[3] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, and N. Jent, "A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies.," *J. Pharm. Biomed. Anal.*, vol. 44, no. 3, pp. 683–700, Jul. 2007.

[4] M. Blanco and I. Villarroya, "NIR spectroscopy: a rapid-response analytical tool," *TrAC Trends Anal. Chem.*, vol. 21, no. 4, pp. 240–250, Apr. 2002.

[5] T. De Beer, A. Burggraeve, M. Fonteyne, L. Saerens, J. P. Remon, and C. Vervaet, "Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes.," *Int. J. Pharm.*, vol. 417, no. 1–2, pp. 32–47, Sep. 2011.

[6] M. Jamrógiewicz, "Application of the near-infrared spectroscopy in the pharmaceutical technology.," *J. Pharm. Biomed. Anal.*, vol. 66, pp. 1–10, Jul. 2012.

[7] M. Blanco, M. Bautista, and M. Alcalà, "API determination by NIR spectroscopy across pharmaceutical production process.," *AAPS PharmSciTech*, vol. 9, no. 4, pp. 1130–5, Jan. 2008.

[8] B. R. M. A. B. M., "Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries," in *Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*, Second Edi., John Wiley & Sons, Ltd, 2010, pp. 463–491.

[9] M. Blanco and A. Peguero, "Influence of physical factors on the accuracy of calibration models for NIR spectroscopy.," *J. Pharm. Biomed. Anal.*, vol. 52, no. 1, pp. 59–65, May 2010.

[10] C.E Miller, "Chemometrics in Process Analytical Technology (PAT)," in *Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*, Second Edi., John Wiley & Sons, Ltd, 2010, pp. 353–438.

[11] C.E Miller, “Multivariate outlier diagnostics: A critical component of NIR/PAT method QA,” in *IDRC-Chambersburg*, 2014.

[12] V. Càrdenas, M. Blanco, and M. Alcalà, “Strategies for Selecting the Calibration Set in Pharmaceutical Near Infrared Spectroscopy Analysis. A Comparative Study,” *J. Pharm. Innov.*, vol. 9, no. 4, pp. 272–281, Aug. 2014.

4. Case of study III

NIR calibration models for samples with tendency to segregation

In this work, we evaluated two modes for spectral acquisition, namely static and dynamic, in which the effective surface of the scanned area was modified, which is related to the representativeness of the analyzed sample, by using a sample holder accessory that allowed the rotation of the sample while the spectrum was collected.
The collected spectra were subsequently used for the calculation of NIR calibration models in order to study in detail the relation between the scanned areas with the information contained in each spectrum, which in turn also influences the predictive ability of the models.

Article in preparation:
V. Cárdenas, L. García, M. Blanco, M. Alcalà
NIR calibration models for samples with tendency to segregation

4.1 INTRODUCTION

In recent years the use of NIR-diffuse reflectance has increased considerably for the quality control of solid samples in the pharmaceutical industry. This is due to the undoubted advantages that offers such as: being a simple, fast and noninvasive technique; moreover is highly versatile due to the development in its instrumentation, since nowadays the analysis can be performed using several accessories like optical fibers, portable NIR i.a., that facilitate an online monitoring of the manufacturing process [1]–[5].

Eighty percent of the manufactured products by the pharmaceutical industry correspond to tablets and capsules. *Muzzio et al* have defined the manufacturing process in this industry as a "powder technology" which is based primarily on making particles, modifying their properties and finally converting them in structured products [6].

Blending is an essential unit operation for the pharmaceutical industry, in which is the active pharmaceutical ingredient(s) –API(s)- and excipients are mixed until achieving homogeneity; this is done in order to obtain formulations with the correct proportion of all its components. Achieving homogeneous mixtures do not only depend on the unit operation itself, but also on the nature of the components that constitute the mixture and the interaction between them. It is important to evaluate the physical attributes of the granules that compose the solid samples such as particle size, shape, surface properties amongst others, since from these characteristics depends if the mixture can be properly homogenized or if this has a tendency to segregate. Segregation is known as the process in which the components of a powder mixture are separated by effect of an external stimulus, resulting in the spatial heterogeneity [7], [8].

In literature are reported at least ten mechanisms that promote segregation [9], however *Carson and Johanson et al* [10] have simplified them in five: trajectory, sifting,

fluidization, air current and angle of repose. Although the separation of components can occur by combination of several mechanisms, segregation by *sifting* is the most common. This occurs mainly due to differences in particle size between components, and promotes the movement top-to-bottom of smaller particles through a matrix of larger particles, thereby forming a lower layer with the finest particles component and higher layers with larger [11].

It is widely known that the simple vibration of the machinery or the exposition of the material to airflow may cause segregation. Therefore the problems related to this phenomenon can be mitigated by adjustments and adaptation of instrumentation and processes to thereby reduce the adverse effects this may cause this lack of uniformity in the material [7].

For the pharmaceutical industry, segregation represent a big problematic because this affect directly the quality of products and processes. Batch failures related to uniformity content might arise if the components of the pharmaceutical formulation experience segregation; resulting in productivity decrease, waste of resources and an increase in the production costs.

In the analysis of solid samples by diffuse reflectance-NIR there are key factors that strongly influence the information that can be extract from each spectrum: i) the physical presentation of the sample –and the intrinsic characteristic of the compounds-, and ii) the surface scanned area; From these depend that a successful analysis can be performed and, in turn, a suitable monitoring of the process [12]. Moreover it is important that the analyzed samples are representative of the sampled material, that through its analysis all the information related to parameters that influence the quality of the product can be accurately extracted [13], [14].

The physical and chemical information contained in an NIR spectrum corresponds to the number of scans in which the sample has been irradiated in a specific area. This

area is related to the irradiation spot size of the instrument, whereby the information extracted from the analysis depends on the characteristics of the material at that point. Also it can be assumed that the bigger the scanned area, the greater the amount of sample that is analyzed; so this also affects the representativeness of the analyzed sample. On the other hand, *Andersson et al* have demonstrated how the scanned area can be increased through the acquisition of NIR spectra of moving solid samples, using accessories that allow the rotation of the sample during the spectra recording or optical fibers that can be put into tanks in motion [12].

In this study we assessed how the effective surface scanned area -in two modes of NIR spectra acquisition- influences the analysis of a pharmaceutical formulation with tendency to segregation. For this, we used a sample accessory that allows the rotation of the sample while the spectrum was recorded – Petri sample holder-. Spectra were acquired with the accessory static and in motion, and these were used for the subsequent calculation of calibration models; in this way we related the spectral acquisition mode with the quality of the extracted information and in turn, the performance of the calibration models.

4. 2. EXPERIMENTAL SECTION

4.2.1 Production samples

The target formulation contained 166.67 $mg.g^{-1}$ of API (sucralfate) which is present in a proportion of 16.7% w/w in the mixture, and four excipients: vanilla flavor, sorbitol 30/60 GDO, sorbitol GDO P60G and saccharin. The sorbitols represent about 83% w/w of the formulation, while the other excipients are present in low proportion.

The commercial product –production samples- is distributed as fine granules obtained by dry granulation. This formulation has a tendency to segregation due to differences in particle size between its major excipients: 180 microns sorbitol GDO 30/60 and 45 microns for sorbitol GDO P60G.

Samples of the pure components and the end product were kindly supplied by Laboratorios Menarini, SA (Badalona, Spain) and its API content was determined by HPLC.

4.2.2 Laboratory samples

A set of 40 laboratory samples was prepared by blending appropriate amounts of API and placebo (the excipients mixture). The concentration of API was expanded ± 20% around its nominal value. Five placebo blends spanning concentration values ± 7.5% around the nominal amount of each excipient in the formulation were also used. Placebos were prepared by using a fractioned factorial design in order to minimize collinearity between excipient concentrations. The experimental design for the placebos preparation was a factorial fractional with 1 central point in which the major concentration variations were set up for both Sorbitols since those are the major excipients in the mixture.

4.2.3 Recording of NIR spectra

Laboratory samples were blended in a T2C shaker WAB shaker mixer, and their NIR reflectance spectra were recorded on a Buchi FT-NIR-Flex 500 spectrophotometer equipped with a module for solids -Petri solid sample-holder governed by the NIRWare software. The recordings were made using the sample module unmoving and in motion; three spectra of each sample was taken in both recording modes which were acquired in the range 10000 to 4000 cm $^{-1}$ with a spectral resolution of 4 cm^{-1} ; the estimated recording time was 2 scans / sec).

The spectra were recorded by placing the samples in a glass Petri dish, and samples turned over with a spatula between recordings. A blank spectrum for the empty cuvette was recorded at the beginning of each working session. The spectral internal and external reference used was the instrument's bundled ceramic plate.

For the static spectral recording mode, the final spectrum was the result of 32 scans maintaining fixed the irradiation surface of the sample, and the tray was manually turned around 120 degrees during the recording of the three replicates. For the dynamic mode, the 32 scans of each spectrum were accumulated during rotation of the tray.

4.2.4 Determination of the effective surface area scanned

For the determination of the effective surface scanned area, it is important to consider the size of the irradiation spot - for recording spectra unmoving- and also the dimensions of the Petri dish -for the spectral recording mode in motion. Therefore one can be calculated as the area of a circle and the other as an annular area.

4.2.5. Construction of calibration models

Spectra were subjected to various treatments including the standard normal variate (SNV), and the first and second derivative as calculated with the Savitzky–Golay and Norris algorithm. The spectral treatments were applied and multivariate models constructed by using the software the Unscrambler v. 9.8 from Camo (Trondheim, Norway).

The PLS algorithm was used to construct calibration models by cross-validation (leave-one-out method) and the model exhibiting the lowest residual variance with the number of latent variables selected for further testing.

Calibration models were refined by using the number of PLS factors leading to the lowest relative square error of prediction (RSEP%) and root mean square error of prediction (RMSEP $mg \cdot g^{-1}$) for an external set of production samples.

4.2.6 Validation of proposed calibration models

The potential industrial usefulness of the selected calibration strategies was assessed by validating their results in accordance with ICH and EMA guidelines [15], [16]. The specific parameters assessed included selectivity, linearity, accuracy, precision (repeatability and intermediate precision) and robustness.

4.3 RESULTS AND DISCUSSIONS

One of the most common factors that may cause segregation is the difference in the particle size amongst components in a mixture. This is the case of the studied pharmaceutical formulation, in which the most predominant particle size difference is between two major excipients -which constitutes 85%w/w of the mixture- showing a tendency to segregation.

The analysis of parameters that influence quality of pharmaceuticals can present some constrains due to its inhomogeneity. However through the evaluation and optimization of the surface scanned area in NIR analysis it is possible to dispose representative samples that reflect all the characteristics of the sampled material, and at the mean time allow obtaining a successful analysis.

The content of API is one of the quality determining parameters in the manufacture of pharmaceuticals; therefore its monitoring is continuously perform in different process unit operations.

With the aim of evaluate the influence of the surface scanned area on the quality of the information obtained in each NIR spectrum and, in turn, on the performance of the calibration models, two spectra recording modes were studied in which the irradiate surface was set.

4.3.1 Sampling with the spectrophotometer and calculation of the effective surface area

The spectra were recorded using the sample accessory –Petri sample holder- unmoving and in motion; for both modes 32 scans were taken for each spectrum.

For the dynamic recording spectra mode, the spot of irradiation was considered for the calculation of the effective surface scanned area. The diameter of this spot is 1.6 cm, therefore the scanned area calculated was 2.01 cm^2. However considering that three replicates were recorded the total scanned area was 6.03 cm^2.

For the area calculations with the sample accessory in motion, besides of the irradiation spot other factors were taking into account, such as the recording time of the instrument -2 scans/sec- and that for the collection of the 32 scans the sample rotates 360º. Moreover the diameter of the Petri plate is 8.5 cm, and the rotation speed was 1.66 cm/s, therefore the calculated annular area was 34.69 cm^2 **Fig 1**.

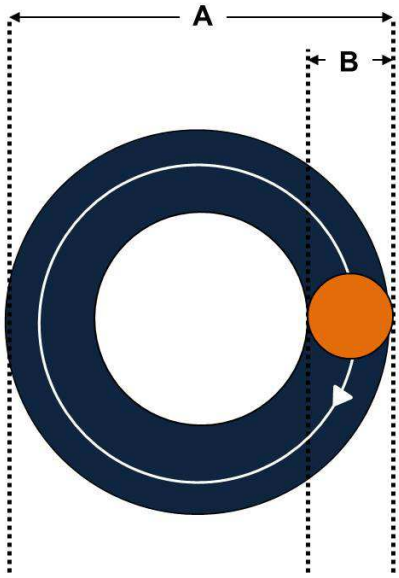


Fig 1Experimental setup for spectra NIR acquisition A)dimension of Petri plate (d= 8.5 cm), in grey the surface area scanned by the spectra recording on moving sample B) dimension of spot radiation (d=1.6 cm), in yellow the surface area scanned in the static spectra recording mode.

With the obtained results from the effective surface scanned area calculations we could confirm that a significantly larger amount of sample could be analysed through the record of the spectra with the samples in motion. For the evaluation of the relation of this amount of sample and the quality of the obtained information, the spectra were used for the subsequent calculation of calibration models.

4.3.2 Construction of NIR calibration models

Once the spectra were recorded in both spectra recording modes and the calculation of the effective surface scanned area performed, we proceeded to construct the calibration models to quantify the API in the pharmaceutical formulation.
A set of 40 samples which contain different quantities of API and excipients were used; their concentration was spanned ± 20% around API nominal value and ± 7.5% from the majoritarian excipients. This set was divided in proportion 70/30 to be used as calibration and validation respectively.
The strategy followed for the construction of the calibration set was a mix calibration, in which both laboratory powder mixtures and production samples were used.

Firstly, a calibration model was calculated using the spectra recorded with the static mode. Prior model calculation different spectral pretreatments were evaluated in order to reduce the scattering effects (by using Standard Normal Variate, SNV) and improve spectra differentiation using first and second derivative (from both, Savitzky Golay and Norris algorithms). The best results were obtained by using Savitzky–Golay algorithm with a 15-point moving window and a second-order polynomial, in combination with SNV in the spectral range 9500-5000 cm^{-1} –where the API showed representative bands- **Fig 2**;moreoverv this model was calculated using 6 factors that explained the 99.5% of Y-Variance. The predictive ability was evaluated by testing 21 samples, 9 of which correspond to production samples and 12 to laboratory powder mixture that constitute the validation set, showing the following prediction errors: ***RSEP(%)= 5.018- RMSEP(mg.g^{-1})=2.830, SD=-1.652.***

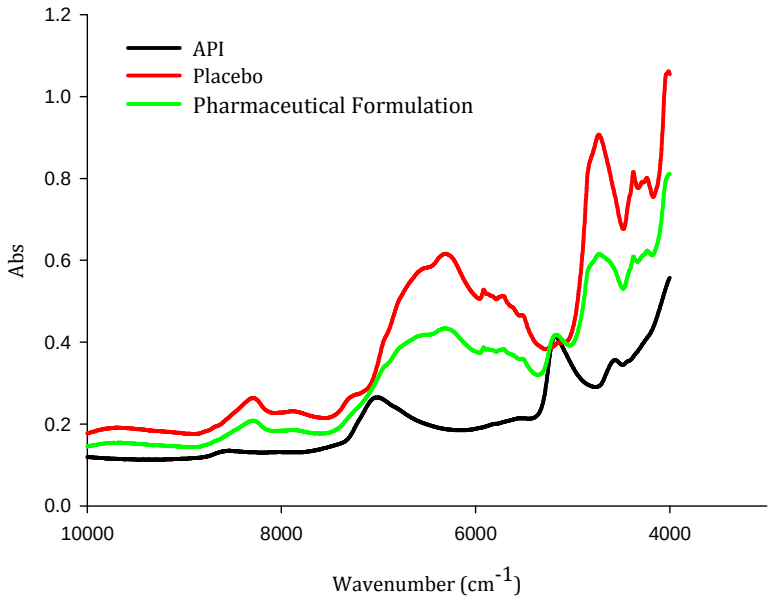


Fig 2. Absorbance spectra of API, pharmaceutical formulation and placebo. Spectral range used for model calculation highlighted

For the construction of the calibration model using the dynamic mode, a previous spectral pretreatment evaluation was performed prior the model calculation. The model based on Norris derivative (second order; Gaps size:9) in combination with SNV in the same spectral using 5 factors -that explained 98% of Y-Variance- provided the most accurate predictions ***RSEP(%)= 2.925- RMSEP(mg.g^{-1})=1.623, SD=2.790*.**

With the results obtained from the calculated models, we confirmed that the spectra recording dynamic mode –in which the surface scanned area is bigger-allow the calculation of simpler and more accurate models. The prediction errors values (**RSEP, RMSEP**) were lower for the model in motion showing a better agreement between the proposed and the reference method. Moreover the values obtained for the standard deviation also showed a better performance of the model in terms of precision.

The comparison between these two models probed the strong influence of the surface scanned area with the information contained in each NIR spectrum, and, in turn, with the performance of the calculated calibration models. Also, with this evaluation we

confirm the importance of sample representativeness for model construction; and how in this way the suitability of the proposed methodologies can be assured giving as an outcome reliable analysis. **Table 1** shows the characteristics of selected models.

Table1. Figures of merit of the calibration model for two spectra recording modes using the solid sample accessory unmoving and in motion.

	CALIBRATION				PREDICTION		
	Spectral Pretreatment	**No Factors**	**%Y-Exp**	**RSEC(%)**	**RSEP%**	**RMSEP**	**SD**
Static mode	2D(S.G)+SNV	6	99.5	0.233	5.018	2.830	2.790
Dynamic mode	2D(Norris)+SNV	5	98	1.215	2.925	1.623	1.652

4.3.3. Validation of the proposed NIR calibration models

The potential industrial usefulness of the selected calibration strategies was assessed by validating their results in accordance with ICH and EMA guidelines [15], [16]. The specific parameters assessed included selectivity, linearity, accuracy, precision (repeatability and intermediate precision) and robustness.

Selectivity of the proposed NIR methods is achieved by identifying the pharmaceutical preparation in a spectral library. The library allows the identification of the pharmaceutical preparation rather than the pure raw materials (active ingredient and excipients) [17].

The applied identification criterion was the correlation coefficient. The library was constructed using 19 granulates belonging to different production batches, the second-derivative spectra (S.G), the whole spectral range 10000-4000 cm^{-1}. All the 19 samples were successfully identify showing identification values between 0.997 and1, while for the API, Sorbitol GDO 30/60 and Sorbitol P60G the values were 0.313, -0.914

and -0.919 respectively. The obtained results showed the suitability of the library for the proper identification of the pharmaceutical formulation and the discrimination of its pure components.

Linearity was assessed by using 12 samples uniformly spanning the working concentration range (± 20% around the nominal value) to quantify the API with both spectra recording modes. A plot of responses against reference values had a slope and intercept containing unity and zero, respectively, at the 95% confidence level. Regarding the correlation coefficient, it can be seen a higher value for the dynamic mode in comparison with the static mode - ***R= 0.974*** y ***R= 0.775*** respectively- showing that the first mentioned model present a better agreement with the reference method.

Accuracy was assessed as the degree of agreement between reference and NIR values for 21 samples spanning the working concentration ranges. The sample predicted with the "in motion model mode" showed a lower Bias (0.368) while for the "unmoving model mode" showed a higher Bias value (-2.306). The same tendency was observed for the Standard Deviation values.
A t-test on the residuals confirmed the absence of significant differences between methods at the 95% confidence level with each of the two models.

Precision was assessed as *repeatability* and *intermediate precision*. *Repeatability* was determined by having the spectrum for an industrial granulate recorded by the same analyst six times and calculating its relative standard deviation (%RSD) in order to quantify the coefficient of variation for each method. The highest %RSD value was that obtained with the "static mode model" **7.91%** while the calculated for the "dynamic mode model" was **2.53%.**

Intermediate precision was determined by having the NIR spectra for an industrial granulate sample recorded by two different analysts on 3 different days. An analysis

of variance (ANOVA) with 2 factors (analyst and day) revealed the absence of significant differences from the reference method.

Robustness was assessed by predicting the values for a second set of industrial granulates and laboratory powder mixtures consisting of 49 samples. The samples were obtained from various production batches manufactured after the first sample set studied and analysed by using the previously developed models a few months after the samples used to construct the models were measured. The results testified to the good predictive ability of the both models for external samples not included in the calibration set and measured after development of the models. However the prediction errors still presenting higher values for the “unmoving mode model”. **Table 2** summarizes the results for each parameter.

Table2. Results obtained from calibration model validation

Parameter		Static mode	Dynamic mode
Linearity	*n*	12	12
	concentration range (mg.g^{-1})	45.3-66.3	43.3-66.3
	Intercept	0.6 ± 23	-1.7±9
	Slope	20 ± 0.4	1.03±0.17
	R	0.775	0.974
Accuracy	*n*	9	9
	Bias	-2.306	0.368
	S.D Residuals	4.172	1.761
	t. Experimental	0.675	0.233
	t. Critic	2.306	2.306
Repeteability	*RSD (%)*	7.91	2.53
Intermediate Precision	*Day*		
	F Experimental	0.6	1.9
	F. Critic	19	19
	RSD (%)	6.82	2.14
	Analyst		
	F Experimental	6.09	0.02
	F. Critic	18.512	18.512
	RSD (%)	3.256	3.45
	RSD global (%)	4.968	3.586
Robustness	*n*	49	49
	RMSEP (mg.g^{-1})	4.95	3.12
	Bias	(-0.145	1.417
	S.D Residuals	5.001	2.809
	t.Experimental	0.035	0.613
	t. Critic	2.01	2.01

4.4. CONCLUSIONS

In this study we evaluated two spectra recording modes for the development of NIR calibration models useful in the quantification of an API in samples with tendency to segregate. With the obtained results we demonstrated the importance of sample representativeness – which is strongly related to the effective surface scanned area- and its influence in the quality of the information obtained in each spectrum. Also we demonstrated how this relationship influences in turn, the development of NIR methods and its performance in terms of simplicity, accuracy, precision and robustness.

This work represents an easy and effective alternative for the analysis of samples with tendency to segregate, and represent a contribution for the optimization of quality control methods based on NIR spectroscopy

REFERENCES

[1] M. Blanco and I. Villarroya, "NIR spectroscopy: a rapid-response analytical tool," *TrAC Trends Anal. Chem.*, vol. 21, no. 4, pp. 240–250, Apr. 2002.

[2] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, and N. Jent, "A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies.," *J. Pharm. Biomed. Anal.*, vol. 44, no. 3, pp. 683–700, Jul. 2007.

[3] T. De Beer, A. Burggraeve, M. Fonteyne, L. Saerens, J. P. Remon, and C. Vervaet, "Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes.," *Int. J. Pharm.*, vol. 417, no. 1–2, pp. 32–47, Sep. 2011.

[4] M. Jamrógiewicz, "Application of the near-infrared spectroscopy in the pharmaceutical technology.," *J. Pharm. Biomed. Anal.*, vol. 66, pp. 1–10, Jul. 2012.

[5] M. Sarraguça Cruz and J. Lopes Almeida, "Quality control of pharmaceuticals with NIR: From lab to process line," *Vib. Spectrosc.*, vol. 49, no. 2, pp. 204–210, Mar. 2009.

[6] F. J. Muzzio, T. Shinbrot, and B. J. Glasser, "Powder technology in the pharmaceutical industry: the need to catch up fast," *Powder Technol.*, vol. 124, no. 1–2, pp. 1–7, Apr. 2002.

[7] A. D. Rosato, D. L. Blackmore, N. Zhang, and Y. Lan, "A perspective on vibration-induced size segregation of granular materials," vol. 57, pp. 265–275, 2002.

[8] X. He, X. Han, N. Ladyzhynsky, and R. Deanne, "Assessing powder segregation potential by near infrared (NIR) spectroscopy and correlating segregation tendency to tabletting performance," *Powder Technol.*, vol. 236, pp. 85–99, 2013.

[9] J. Mosby and S. R. De Silva, "Segregation of Particulate Materials Mechanisms and Testers t," vol. 14, no. 14, pp. 31–43, 1996.

[10] J. W. Carson and J. Johanson, "OVERCOMING PARTICLE SEGREGATION IN THE PHARMACEUTCAL," vol. 14, no. 18, pp. 2749–2758, 1988.

[11] "Why is Understanding Segregation Mechanisms so Important ? By Dr . Kerry Johanson."

[12] M. Andersson, O. Svensson, S. Folestad, M. Josefson, and K.-G. Wahlund, "NIR spectroscopy on moving solids using a scanning grating spectrometer—impact

on multivariate process analysis," *Chemom. Intell. Lab. Syst.*, vol. 75, no. 1, pp. 1–11, 2005.

[13] R. Romañach and K. Esbensen, "Sampling in pharmaceutical manufacturing—Many opportunities to improve today's practice through the Theory of Sampling (TOS)," *TOS forum*, vol. 2015, no. 4, p. 5, 2015.

[14] L. P. J. K. H. Esbensen, "Representative sampling data quality validation-anecessary trinity in chemometrics," *Compr. Chemom.*, vol. 4, pp. Vol. 4, 1–20, 2009.

[15] International conference on harmonisation of technical requirements for registration of pharmaceuticals, "ICH Guidelines: Q2(R1) Text on validation of analytical procedures: methodology Q2(R1)," 1996. [Online]. Available: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf.

[16] European Medicines Agency, "Guideline on the use of Near Infrared Spectroscopy (NIRS) by the pharmaceutical industry and the data requeriments for new submission and variations.," 2012. [Online]. Available: EMEA/CVMP/QWP/17760/2009rev2.

[17] M. Blanco and M. a. Romero, "Near-infrared libraries in the pharmaceutical industry: a solution for identity confirmation," *Analyst*, vol. 126, no. 12, pp. 2212–2217, Dec. 2001.

CONCLUSIONS

Based on the results obtained in each case of study, the following conclusions can be extracted and summarized:

1- New methodologies useful for the quality control of pharmaceuticals have been developed using NIR spectroscopy combined to chemometrics; these methodologies represent a solution for industrial constrains and are suitable for its use in manufacture.

2- The active pharmaceutical ingredient (API) of different pharmaceutical formulations was successfully quantified in different solid forms: powder, granulates, cores and tablets, confirming the suitability of the NIR technique and the chemometric methods for this analysis.

3- Strategies such under/overdosing of industrial samples, mix calibration sets and calculation and addition of process spectra were used for the design and selection of the calibration set and its performance evaluated. The physical and chemical variability were successfully incorporate. The strategy of calculation and addition of process spectrum showed a better performance in terms of robustness, easiness inclusion of sample variability without the need of reference method and simplifying the experimental procedure.

4- The calculation and addition of process spectrum strategy was optimized by using the statistics Hotelling's T^2 and Q-residuals based on the model space for monitoring the API content on a formulation through its manufacturing process: blending, tableting and coating. In this way an objective and successful procedure for the implementation of this strategy was proposed, which showed its suitability for the quality control analysis of pharmaceuticals in different physical forms.

5- The influence of the effective surface scanned area of the spectral acquisition procedure was confirmed by testing two spectral record modes which allows static and dynamic spectral acquisition.

6- The importance of sample representativeness on the quality of the information contained in each spectrum NIR was demonstrated by the association of the surface scanned area and the performance of the calculated models. The results clearly showed that using an spectral acquisition mode which allows a bigger scanned area, enable the construction of simplest, accurate and robust models.

7- Problems in the analysis of samples with tendency to segregation -which can led to batch failures related to uniformity content- can be mitigate by incrementing the effective surface area, and this area can be increased by the use of instrument sampling accessories for dynamic spectral acquisition..

LIST OF PUBLICATIONS AND CONFERENCES

1. Publications

1.1 Published work

- Cárdenas V., Blanco M., Alcalà M., "Strategies for selecting the calibration set in pharmaceutical near infrared spectroscopy analysis. A comparative study". Journal of Pharmaceutical Innovation, 4 (272-281) 2014.

- Cárdenas V., Cordobés M., Blanco M., Alcalà M. "Strategy for design NIR calibration sets based on process spectrum and model space. An innovative approach for process analytical technology". Journal of Pharmaceutical and Biomedical Analysis, 144 (28, 33) 2015.

1.2 In preparation

- Cárdenas V., García L., Blanco M., Alcalà M. "NIR calibration models for samples with tendency to segregation"

2. Communications in scientific meetings

2.1 Oral communications

- Cárdenas V., Alcalà M., Blanco M., "Different strategies for multivariate calibration using NIR data in the pharmaceutical quality control" V Workshop of chemometrics, Badajoz, Spain 2013.

- Cárdenas V., Alcalà M., Blanco M., "Process spectra strategy for NIR calibration set preparation: An innovative tool for pharmaceutical analysis-PAT" 14th Instrumental Analysis Conference". Barcelona, Spain 2014.

- Cárdenas V., Blanco M., Alcalà M., "Efficency of process spectrum methodology in NIR pharmaceuticals quality control analysis. An innovative tool for process analytical technologies". Jornades doctorals, Programa de doctorat en Quimica. Bellatera, Spain 2015.

2.2 Poster communications

- Cárdenas V., Córdobes M., Alcalà M., Blanco M."Multivariate NIR calibration strategies useful for quality control of pharmaceutical formulations", VIII Colloquium Chemiometricum Mediterraneum, Bevagna, Italy 2013.