UNIVERSIDADE DE LISBOA

FACULDADE DE FARMÁCIA

DOSING THE ACTIVE INGREDIENT IN PHARMACEUTICAL POWDER MIXTURES, USING NEAR-INFRARED SPECTROSCOPY (NIRS)

Ana Rita de Paiva Canelas

Thesis presented in fulfillment for the degree of

Master in Pharmaceutical Engineering

2014

UNIVERSIDADE DE LISBOA

FACULDADE DE FARMÁCIA

DOSING THE ACTIVE INGREDIENT IN PHARMACEUTICAL POWDER MIXTURES, USING NEAR-INFRARED SPECTROSCOPY (NIRS)

Ana Rita de Paiva Canelas

Thesis presented in fulfillment for the degree of

Master in Pharmaceutical Engineering

Supervision: Prof. Doutor José Monteiro Cardoso Menezes (DEQB)

Doutor Nuno Moreira (OM Pharma)

2014

AGRADECIMENTOS

O trabalho apresentado nesta dissertação é fruto de horas de dedicação para alcançar o fim a que me propus. A realização e a obtenção do trabalho e resultados aqui sumarizados e discutidos não seriam de todo possíveis sem ajuda de pessoas que tiveram um papel fundamental durante estes meses de reflexão e estudo estando sempre a meu lado, seja pelos seus ensinamentos, sugestões de melhoria, apoio e dedicação. Por esta razão, desejo expressar os meus sinceros agradecimentos:

Ao Professor Doutor José Cardoso Menezes, meu orientador, pela competência científica, pelo apoio, conhecimentos transmitidos ao longo da elaboração desta dissertação e pela prontidão na disponibilidade para me auxiliar sempre que necessário.

Ao Doutor Nuno Moreira por ter aceite ser o co-orientador desta dissertação e ter participado de forma inconfundível na execução do trabalho através dos seus conhecimentos, sugestões, ensinamentos e por ter permitido que este trabalho pudesse ter uma visão mais abrangente comparativamente ao inicialmente proposto.

Á OM Pharma, em especial à Doutora Isilda Serra, Doutora Mafalda Quinto, Doutor António Jordão e Doutor Nuno Moreira por permitirem a realização deste trabalho.

À Doutora Sandra Vilela, Responsável de Controlo de Qualidade da OM Pharma, por prontamente ter aceite a realização deste trabalho no Laboratório de Controlo de Qualidade. O seu apoio, compreensão e dedicação incondicionais foram de elevada importância durante todo o percurso de desenvolvimento deste trabalho.

Ao Pedro Ferreira que teve um papel fundamental em todas as etapas de desenvolvimento deste trabalho estando presente nos bons e maus momentos e participando sempre em todas as conquistas alcançadas para hoje este trabalho ter sido possível. Pelos conselhos, dedicação, ajuda, encorajamento e positivismo, o meu sincero e profundo obrigado (*thank you for everything that you did for me*).

Às minhas queridas Lúcia Sousa e Daniela Cardoso o meu muito obrigada, sem palavras para vos agradecer.

Ao Duarte Mendes, não apenas pela sua competência e tacto na revisão do meu trabalho, mas também pelo seu entusiasmo e encorajamento.

Seria impossível descrever o carinho que tenho pelos meus pais e irmão que tanto me apoiaram e compreenderam durante este atribulado período de desenvolvimento e graças a eles este projecto foi possível, a eles dedico este trabalho.

A todos os meus colegas de Laboratório que de uma maneira ou de outra, seja por um olhar, por uma palavra amiga, por um gesto, pela participação directa neste trabalho me ajudaram (ordem alfabética): Amândio Araújo, Ema Marques, Isabel Caldeira, Isabel Sarmento, Sara Fernandes, Natércia Pires.

À Margarida Martins, Cláudia Matos e Carla Louro amigas inconfundíveis que me apoiaram e aconselharam sempre durante este atribulado percurso. A compreensão da minha indisponibilidade foi muito importante para mim.

A todos os outros que por esquecimento ou lapso não refiro mas todos tiveram um papel muito importante no desenvolvimento desta dissertação, o meu muito obrigado.

RESUMO

O primeiro objectivo desta tese foi o desenvolvimento de um método de quantificação por espectroscopia de infravermelho próximo de modo a optimizar um processo de fabrico da OM Pharma numa fase crítica de produção: fase final da mistura. Em seguida avaliou-se a aplicabilidade do mesmo tipo de estratégia ao controlo (identificação, qualificação e quantificação) de outros pontos críticos do processo. A necessidade da implementação desta metodologia incide no facto de a este processo estar associado um volume de trabalho significativo de trabalho do Controlo de Qualidade.

O modelo de quantificação desenvolvido permite a determinação do parâmetro de qualidade crítico do processo de fabrico do produto: a conformidade do teor em API na mistura; após a verificação da conformidade processo produtivo evolui para outra fase – produto final. A aplicação desta técnica desenvolvida permite, em rotina, a redução do tempo despendido em análise pelo Controlo de Qualidade. O modelo obtido foi validado de acordo com as *Guidelines* em vigor.

Um segundo objectivo, foi o de generalizar a aprendizagem anterior e desenvolver uma biblioteca para diversas matérias-primas (princípios activos farmacêuticos) que permitisse a sua identificação e no futuro possivelmente a sua qualificação; esta necessidade surge devido à elevada quantidade de lotes de matéria-prima recepcionada periodicamente na OM Pharma.

A criação da biblioteca consiste no desenvolvimento de um método que permita identificar o princípio activo referido no modelo de quantificação, o que acarreta a construção de uma Biblioteca de Princípios Activos (API) obtida pela aquisição de espectros NIR de todos os API´s da OM Pharma. A biblioteca desenvolvida foi sujeita a validação interna e externa de acordo com os requisitos das *Guidelines* em vigor.

Concluiu-se que a espectroscopia de infravermelho próximo é um método preciso e benéfico para a análise e controlo de qualidade no controlo da fase final de produção e na identificação de matérias-primas na Indústria Farmacêutica. Associada à utilização desta técnica, o aumento da produtividade através da redução do tempo de análise e, consequentemente, a redução dos custos operativos é sem dúvida um factor muito positivo.

Palavras-chave: Espectroscopia NIR; Método de Quantificação de API em misturas; Identificação de matérias-primas; Industria Farmacêutica; modelos PLS

ABSTRACT

The first goal of this thesis was the development of a near infrared quantification method in order to optimize the mixing process in OM Pharma's production phase, followed by the application of this method in controlling other critical processes such as identification, qualification and quantification. The fact that this process is associated with a significant part in Quality Control's work volume justified the implementation of this methodology.

The developed quantification PLS model allows the determination of a product manufacturing process critical quality parameter: the compliance of the API content; after checking this production phase conformity, the process evolves into another phase - final product. The application of this technique allows the reduction of the spent time on routine analysis. The model was validated according to the guidelines.

A second goal consisted in developing a library allowing to identify several raw materials (pharmaceutical active ingredients) and in the future it's possible qualification. This need arose due to the high amount of raw material batches periodically received in OM Pharma.

The library development is based on a method developed for identification of the active principle in said quantization model which leads to the construction of a library of active ingredients (API) obtained by the acquisition of NIR spectrum. The developed library was subjected to internal and external validation according to the requirements of the Guidelines in effect.

The near infrared spectroscopy method proved itself as an accurate and beneficial method for the analysis and quality control in controlling the final stages of production and raw materials identification in the pharmaceutical industry. Associated to the use of this technique, increased productivity by reducing the analysis time and, consequently, the reduction of operating costs is without a doubt very positive.

Keywords: NIR Spectroscopy; API in blends quantification method; Raw materials identification; Pharmaceutical Industry, PLS models

RESUMO E ENQUADRAMENTO DO TRABALHO

A Espectroscopia de Infravermelho Próximo (NIRs) é uma técnica de análise rápida e não destrutiva que oferece muitas vantagens para uma variedade de aplicações industriais. A implementação de uma nova metodologia de trabalho requer o desenvolvimento e a utilização de técnicas de análise simples, rápidas e eficientes para adquirir num curto espeço de tempo a informação necessária; as técnicas espectroscópicas são especialmente susceptíveis a esta utilização e permitem obter respostas no local de produção com níveis de exactidão e precisão comparáveis aos métodos de referência primários.

A tecnologia de Infravermelho Próximo (NIR) permite análises qualitativas e quantitativas. A análise qualitativa envolve, através do tratamento espectral, o agrupamento de amostras similares criando, assim, grupos de espectros padrão para gerar uma biblioteca de referência. A aplicação desta técnica na previsão de parâmetros quantitativos envolve técnicas de quimiometria e análise multivariada necessárias para construir os modelos de calibração relacionando cada espectro com os valores de referência obtidos para os parâmetros em estudo.

A NIRs tornou-se, nas últimas décadas, uma ferramenta de grande aplicabilidade e utilidade na Industria Farmacêutica, em particular, associado ao Controlo de Qualidade e, também, à monitorização de processos de fabrico. O Controlo de Qualidade é um elemento essencial dos processos de produção farmacêutica em virtude da alta segurança e dos níveis exigidos pelas Entidades associado ao controlo das formulações disponíveis no mercado. O impacto da aplicação desta tecnologia foca-se essencialmente na celeridade atribuída aos processos que por vezes é afetada pelos tempos de espera necessários para a análise dos produtos pelo Controlo de Qualidade.

Através de uma análise ao impacto desta formulação na OM Pharma surge a necessidade de implementação da nova tecnologia NIR para optimizar o fluxo do processo de fabrico em termos de celeridade de respostas e tempo dispensado. A formulação estudada garante uma parte significativa do volume de trabalho do Controlo de Qualidade da OM Pharma.

O objectivo proposto para este trabalho incide no desenvolvimento de um método quantitativo NIR que permita controlar a fase final de um processo de fabrico – mistura de pós, fase anterior à finalização do processo de produtivo. Dado o significativo impacto em termos de número de lotes fabricados associado a este produto, em particular, na OM Pharma surge a necessidade de controlar não apenas o processo de fabrico numa fase final de produção, mas também o controlo desde a entrada do princípio activo (API) no Armazém da OM Pharma através de uma biblioteca de identificação até à sua utilização como parte integrante na formulação. A avaliação do impacto associado ao processo produtivo abordado neste trabalho foi um factor crítico na aplicabilidade da nova tecnologia.

Todas as amostras envolvidas na construção de ambos os modelos (modelo quantitativo e biblioteca de espectros) foram devidamente analisadas pelas respectivas monografias e todas cumprem os critérios exigidos pelas mesmas. O armazenamento das amostras foi efectuado em local apropriado sob condições de humidade e temperatura controladas. A aquisição de espectros foi efectuada utilizando o equipamento microPHAZIR^{TM} Handheld NIR *Analyzer spectrometer*.

A recente Guideline *da European Medicines Agency* (EMA) que foca exclusivamente a utilização e implementação de métodos NIRs na Industria Farmacêutica (*Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations*) foi guia para o desenvolvimento e validação dos métodos NIR envolvidos neste trabalho.

A aquisição de espectros para a construção de modelos é um factor muito importante e crítico. Um bom conhecimento das possíveis interferências na aquisição espectro da amostra assegura um bom resultado no desenvolvimento e construção de um modelo. Para a avaliação das possíveis interferências foi efectuada uma análise de risco aos factores que poderiam ter um impacto sobre os atributos de qualidade desejados para os métodos NIR desenvolvidos.

O modelo de quantificação desenvolvido permite a determinação do parâmetro de qualidade crítico do processo de fabrico do produto: a conformidade do teor em API na mistura final. Para a construção do modelo foram utilizadas 70 amostras o que equivale a 700 espectros. As amostras reais abrangem, regra geral, a gama de teor em activo de 97% (m/m) – 102% (m/m) pelo que surgiu a necessidade de produção amostras contaminadas (amostras de laboratório). Face ao exposto, as amostras utilizadas na construção do modelo correspondem a amostras reais e amostras contaminadas por adição de placebo ou activo com o objectivo de abranger a gama pretendida: 80% (m/m) – 95% (m/m) (adição de placebo) e 105% (m/m) - 110% (m/m) (adição de API). Todas as amostras envolvidas no modelo de calibração foram devidamente analisadas pelo método de referência de modo a obter os valores reais de percentagem em API. O modelo desenvolvido por NIR permitiu obter um método robusto que permite a quantificação em API numa formulação na gama de 80% (m/m) – 110% (m/m). Fontes de variabilidade como campanhas de lotes de produção, concentrações de API, operadores, temperatura da amostra e temperatura ambiente foram introduzidas no modelo de quantificação para aumentar a robustez do mesmo. Com base na validação cruzada foram obtidos quatro factores *Partial Least Squares* (PLS) para o modelo de previsão, *Root Mean Square Error of Cross Validation* (RMSECV) (%, m/m) de 1,73 e coeficiente de determinação (R^2): 0,957. O método foi validado para uma gama de concentração de 80% (m/m) – 110% (m/m) em API utilizando um conjunto de validação externa. A avaliação da capacidade preditiva obtida a partir dos resultados de validação (*Root Mean Square Error of Prediction* (RMSEP) (%, m/m): 1,67, coeficiente de determinação (R^2): 0,961) confirmou a precisão adequada dos resultados obtidos pelo método NIR na gama de concentração em API estudada. Após a finalização da construção do modelo quantitativo, amostras reais não envolvidas na validação e construção do mesmo foram submetidas ao novo método. Os resultados obtidos pelo método NIR comparativamente aos obtidos pelo método referência para as mesmas amostras demonstram uma boa concordância.

O método de identificação acarreta a construção de uma Biblioteca obtida pela aquisição de espectros NIR de todos os API´s da OM Pharma com a finalidade de permitir a identificação directa e rápida aquando da recepção da elevada quantidade de lotes de API envolvido neste projecto. Este método implicou a criação de uma biblioteca de referência para cada material; a média de cinco espectros constitui um padrão incluído na biblioteca espectral. O desenvolvimento da biblioteca de identificação de API foi efectuada utilizando o algoritmo *de K-Nearest Neighbour* (KNN). A selecção dos lotes evolvidos teve como base uma avaliação dos parâmetros críticos de qualidade química e física de cada um dos lotes de API envolvidos. A variabilidade associada a cada um dos materiais é um factor de elevada relevância para a preparação da biblioteca no reconhecimento dos materiais submetidos a este método. Após a construção e validação (interna e externa) da biblioteca é possível identificar correctamente os API´s utilizados na OM Pharma.

INDEX

TABLE AND GRAPHS INDEX

FIGURES INDEX

ABBREVIATIONS

API: Active Pharmaceutical Ingredient/Principio activo
CC: Correlation coefficient
EMA: European Medicines Agency
QC: Quality Control
GMP: Good Manufacture Practices
IQ: Installation Qualification
ICH: International Conference on Harmonization
InGaAs: Indium Gallium Arsenide
KNN: K-Nearest Neighbour algorithm
MSC: Multiplicative Scatter Correction
NIR: Near-infrared
NIRs: Near-infrared spectrometry
OQ: Operational Qualification
PAT: Process Analytical Technology
PQ: Performance Qualification
PCA: Principal Components Analysis
PMG: Polychromix Method Generation
PLS: Partial Least Squares
PC: Principal Component
PbS: Lead Sulfite
QC: Quality Control
R^2: Coefficient of determination
$RSD_{\%}$: Standard deviation
RM: Reference Method
RMSEC: Root Mean Square Error of Calibration
RMSECV: Root Mean Square Error of Cross Validation
RMSEP: Root Mean Square Error of Prediction
S. G.: Savitzky–Golay
SNV: Standard Normal Variate
USP: United States Pharmacopeia

1. GOALS

The initial goal for this master thesis was the evaluation of the final phase of a mixing process by developing a quantification method, based on NIR spectrum, so a critical parameter could be estimated, the API content.

A total of 70 samples (37 real samples and 33 placebo/API spiked samples) were used to develop the quantification method.

A risk analysis was performed in order to evaluate possible interferences by testing critical risk factors through a Principal Components Analysis and Partial Least Squares regression analysis.

When developing this work, the need to develop an API library came up, in order to facilitate the whole process, from its reception as a raw material to its quality control in a formulation. From this need, a second goal was established. To build the API library, every active ingredient, a total of 10, used OM Pharma were incorporated. The library was build with the K-Nearest Neighbour (KNN) algorithm.

Thus, this thesis is divided in two distinct parts:

A. Quantification Model (API quantification in the final phase of mixing)

- Quantification Model Development
- Quantification Model Validation

B. Qualification Model – API Library (stored API identification)

- Library Development
- Library Validation

Both quantification and identification methods were validated according to the European Medicines Agency (EMA) Guideline: Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations.

2. INTRODUCTION

Similarly to all pharmaceutical industries, OM Pharma is governed by the Good Manufacturing Practice (GMP) guidelines, focusing on regulating the manufacturing and laboratory testing environment to ensure that the manufacturing of pharmaceutical products is controlled and managed. A key component of GMP is the documentation of all stages of the manufacturing process. The documentation enables traceability of the product throughout the supply chain. Additionally, GMP requires that all manufacturing and testing equipment has been qualified as suitable for use, and that all operational methodologies and procedures (such as manufacturing, cleaning, and analytical testing) used in the drug manufacturing process have been validated to demonstrate that they can perform their functions.

Quality Control is the part of GMP which its main activities refer to sampling, specifications, tests, procedures, organization and release documents to ensure that the necessary and relevant tests are carried out and that products are not released for sale or supply, until their quality is considered satisfactory, i.e., meeting the requirements.

NIR technology allows a qualitative and quantitative analysis. Qualitative application is seen as a fingerprint of the sample and is based on the processing of the entire spectrum, allowing the grouping of similar samples. For the quantitative application, chemometrics techniques are required to build the calibration models. These techniques allow chemometrics, multivariate analysis, relating the data obtained in each spectrum with the reference values for each parameter, so that new samples can be estimated by the system.

This work demonstrates the development of models using chemometric techniques, capable of generating calibration models that predict the above direct analysis of samples quality parameters. To achieve this, a strong analytical component for determining parameters by a reference method is needed, in order to obtain analytical results for a sufficient amount of samples allowing the preparation of the calibration model.

The construction of the API library involved a careful selection of raw materials. This selection was performed by picking batches which critical parameters are associated, encompassing them in the construction of the Library.

This interest in pharmaceutical industry is due to NIR's advantages over other analytical techniques, namely, its capability for non-destructive analyses (a sample can be reused after the measurement), an easy sample preparation without any pre-treatments or reagents consumption, it's speed (a spectrum can be recorded in only a few seconds), economy, accuracy, precision and the prediction of chemical and physical sample parameters from one single spectrum.

In this work, the development and validation of a quantitative method and an identification method for raw materials (API) will be described.

3. LITERATURE REVIEW

3.1. OM Pharma

OM Pharma is a division of Vifor Pharma, a company controlled by Galenica Group since 2009. It is an internationally active company, investing in research and development of new products. In partnership with Vifor Pharma, OM Pharma produces pharmaceuticals, owning several production facilities across the world.

OM pharma products are exported and sold to several countries, as well as in Portugal. From all the products, some of them stand out in the Portuguese market, such as: AeroOM®, Aero-Bio®, Aero-Bio® Adulto, Bronquial, Doxiproct® pomada, Hylo-Comod®, Hylo-Care®, Lakripos® Gel, Vis-OM®, Doxi-OM®, Equazen™, Luzon, Arteoptic®, Colpotrophine®, Glimepirida Diapiride, Gastrex®, Femsete, Lutenyl, Maltofer®, FerrumHausmann®, FerrumFol, Broncho-Vaxom® Adulto e Infantil, Uro-Vaxom®, Estreva®-Gel, Mictonorm®, Trophoseptine, Viternum-Xarope, Ferinject®, Venofer®.

3.1.1. Quality Control – OM Pharma

The Quality Control is performed in several stages, in the reception of raw materials, during the manufacturing process, bulk product, in finished product and the stability of the product throughout its shelf life.

Quality Control Laboratory is included in Quality Department. The Quality Control Laboratory is divided in five different sections: Physical-chemical control, Microbiological control, Stability studies, Raw Material and Packaging material. Some of quality control activities involved in each section are:

Figure 1 – Quality Control section of OM Pharma

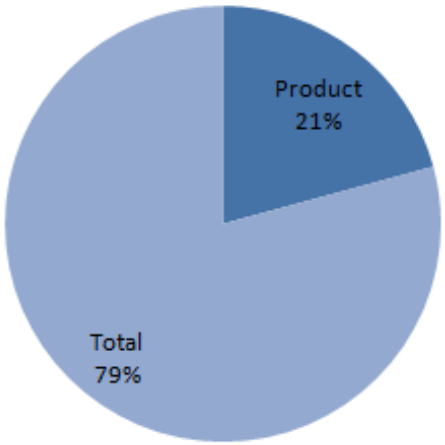
As all pharmaceutical companies, OM Pharma Quality Control spends much of his time analyzing bulks and the results have to be obtained in the shortest possible time because the production depends on the results to proceed with the process.

The raw material analysis is also critical, since as a regulatory requirement, every single container of each batch of each raw material must be identified, which is considerable work in terms of time, cost and productivity. Additionally, raw materials identification has different Pharmacopeia methods and sometimes more than one test is mandatory.

The implementation of NIR in OM Pharma aims to minimize the critical activities of the Quality Control: bulk analysis (the only critical product in terms of number of batches received) and identification of raw materials (all solid actives substances).
Considering these advantages, quality control department decided to implement NIR technology to reduce time and cost associated to these stages and increase productivity, hence the development and validation of

a quantification method and a raw material identification method using NIR spectroscopy was crucial, so the technique could be used in a routine basis.

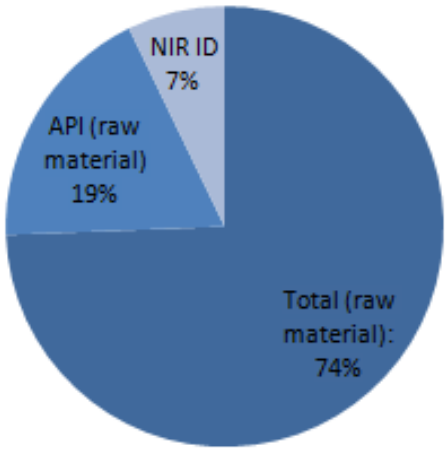
To evaluate the NIR quantification method's impact in analysing blends at OM Pharma, a cost and spent time analysis was performed, taking into account produced batches comparing to the annual production plan.



Graph 1 – Blend (Product) impact in Annual Production Plan (batch numbers per year)

As seen in **Graph 1**, 21% of the annual production arrives from the main product (blend) addressed in this work, so a real opportunity of improvement exists.

Regarding the raw-materials, the number of API's per year impact is significant when comparing to the total number of raw materials tested at the Quality Control – **Graph 2**.



Graph 2 –API *vs* total raw materials impact
(NIR ID: NIR method developed for liquid raw materials; Total (raw material): developing method NIR; API (raw material): NIR method developed in this work for API OM Pharma)

About 19% of all the raw-materials entering OM Pharma are API batches. A NIR identification method was already developed in OM Pharma for identification of liquid raw materials, representing 7% of total raw

materials. Again, this represents an opportunity to reduce analysis cost and time, since about 26% of OM Pharma's raw materials could be identified by NIR rather than the reference method.

The scheme below shows the NIR implementation and its impact in OM Pharma's Quality Control:

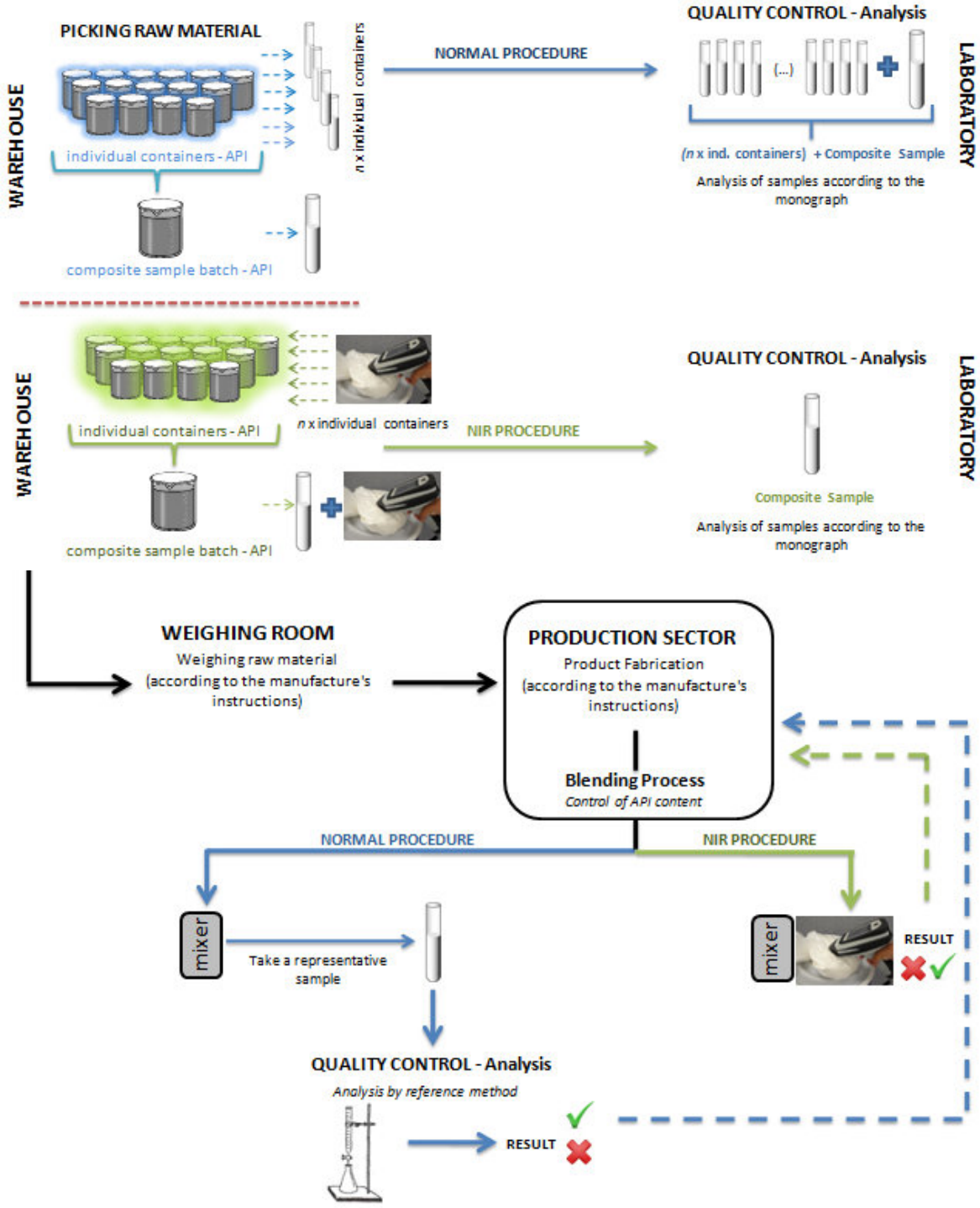


Figure 2 – OM Pharma's NIR implementation impact

By implementing the NIR method, the actual flow associated with both raw-materials (sampling and identification) and final mixing phase (quantification) analysis would be optimized. Looking at **Figure 2** the following is observed:

- **Raw material identification**

At this moment every API container is sampled in the warehouse to be later identified in the CQ lab. With implementation of NIR identification there are practical advantages, as raw materials can be analyzed as they arrive the warehouse, and do not have to be transported for analysis (Quality Control Laboratory), ensuring improved workflow, since all identification would be performed at the warehouse alongside the sampling, and only the composite sample would be analyzed in the CQ lab according to the respective monograph. This fact has a high impact in the number of samples tested in the CQ lab, with the NIR method only 1/6 of the actual identification would be made. If the analysis is carried out in the warehouse, no sampling is required by GMP guidelines and therefore the time per analysis is reduced.

- **API control in the final phase of a mixing process**

The final mixture control is done at the CQ lab. After sampling the mixture at the production facilities, it is transported to the lab for it's analysis to be performed. This fact means that a considerable waiting time is needed, corresponding to at least the time spent with the reference method analysis. By implementing the NIR method this procedure would be carried out at production, obtaining a fast result.

3.2. Spectroscopy

Infrared spectrum is an important characterization record, providing chemical and physical information of a compound. In recent years, NIR spectroscopy has become widespread in process analysis and within pharmaceutical industry for raw material testing, product quality control and process monitoring. [1]

3.2.1. Background

Spectroscopy analytical methods designates itself by the investigation of the interaction of electromagnetic radiation with the molecules. The connection of two atoms in molecules involves different types of energy, such as translational energy, vibrational and electronic. In the case of infrared spectroscopy, it is based on the relative motions of atoms in a molecule, that is, its vibration. Therefore, this spectroscopy detects the radiation that is absorbed by the molecular vibrational links.

Such vibrations lead to overlapping bands which contain both physical and chemical information. In the region of the electromagnetic spectrum, infrared ranges between the visible and microwave and is divided into the near infrared (4000 - 12500 cm^{-1}), medium (400 - 4000 cm^{-1}) and distant (24 - 400 cm^{-1}). The near

infrared spectrum region comprises the range between 800 and 2500 nm, equivalent to 4000 - 12500 cm^{-1} for measurement in wave numbers. The IR region is commonly divided into three smaller areas: far-IR (FIR), mid-IR (MIR) and near IR (NIR) [2].

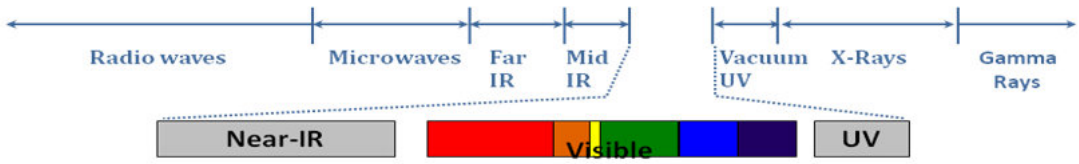


Figure 3 - Electromagnetic spectrum *(Thermo Fisher Scientific material support)*

The most prominent absorption bands occurring in the NIR region are related to the overtone and combination bands of the fundamental molecular vibrations of C-N, N-H, O-H and S-H functional groups with hydrogen. [3, 4]. Thus, most chemical and biochemical species exhibit unique absorption bands in the NIR spectral region that can be used for both qualitative and quantitative purposes.

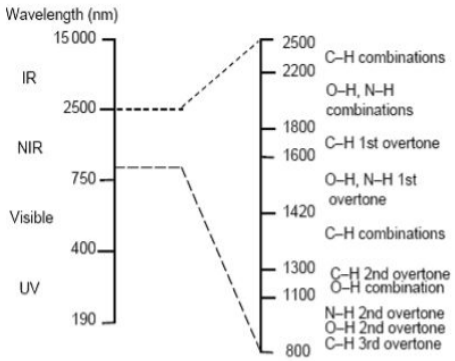


Figure 4 – Characterization of group absorption region [1]

This technique has many advantages, offering rapid spectrum acquisition, non-invasive and non-destructive sample analysis, requiring little or no sample preparation, [2, 5]. NIR spectrum of powder contains both chemical and physical properties and the response is quick, so NIR spectroscopic technique can be used for on-line monitoring of powder from point of view of powder homogeneity, powder flow properties and particle size distribution, in order to ensure consistent uniformity of dosage units. Based on these advantages, infrared spectroscopy matches the PAT (Process Analytical Technology) requirements. A major advantage of NIR spectroscopy is, obviously, immediate (real time) delivery of results. But this technique has some drawbacks, such as, the influence of physical properties on spectrum and the need for calibration using a reference technique [6, 7].

The unique properties of the NIR spectroscopy method makes it suitable for monitoring many manufacturing processes, such as fluidized-bed granulation (wet, in-line), uniformity of mixing, control of spraying and drying in order to determine the end-point of those processes [8]. Thus, chemometric tools such as spectrum pre-treatments and regression methods are needed to extract the significant information [9]. Many pharmaceuticals applications of NIR methods on powders, granulates, liquids and semi-liquid were reported: identification of raw material, powder homogeneity, moisture determination, particle size and characterization of polymorphs. [10]

There are several comprehensive reviews demonstrating that NIR spectroscopy is a robust method for the qualitative and quantitative analysis of the pharmaceutical solids. Briefly, NIR spectroscopy has been widely used off-line in raw material testing and product quality control. [9]

In the manufacturing process, the uniform mixing of drug and excipients is an essential step before proceeding to other operations, since it is well know that inert excipients can affect the characteristic, quality, stability and even the performance of the final product [11, 12].

This technology allows the usage of qualitative but also quantitative applications. Qualitative application is seen as a fingerprint of the sample and is based on the processing of the entire spectrum, allowing the grouping of similar samples [13]. For the quantitative application, chemometrics techniques are required for building the calibration models. As the intensity of an absorption band is proportional to the concentration of the component that causes this band, the amount of a compound existing in a sample can be determined by a calibration curve (concentration versus band intensity) constructed from samples with known compound in question through a multivariate analysis of the large number of variables obtained from a NIR spectrum.

During the calibration process, reference values and corresponding chemometrics techniques are usually required [14]. These techniques allow chemometrics, multivariate analysis, relating the data obtained in each spectrum with the reference values for each parameter, so that it can estimate new samples in the system.

The uniformity is, normally, determined in practice by estimating the distribution of drug, based on its assay in a representative sample of the whole batch. The assay of active content is usually done through conventional methods: high-performance liquid chromatography (HPLC), spectrophotometry, titrations and other analytical techniques [5].

Quality control constitutes an essential element of pharmaceutical production processes by virtue of the high safety levels demanded for commercially available formulations [5]. Conventional techniques such as chromatography, titration and ultraviolet-visible, exist for determining the composition of several samples, employing laborious time-consuming pre-treatments and the use of solvents and reagents. These solvents/reagents are often hazardous, both for the analyst and for the environment and need to be further processed as waste. Instead, NIR analyses are associated with multivariate methods which through statistical methods allow large amount of data treatment, thus, reducing the usage of solvents and reagents and spent time.

This work demonstrates the development of models using chemometric techniques, capable of generating calibration models that predicted the above direct analysis of samples quality parameters. Therefore, an intense analytical component was necessary, that is, determination of parameters using the reference method, in order to obtain test results for a number of sufficient samples to allow the preparation of the calibration model.

NIR is being increasingly considered as an attractive and promising analytical tool for Process Analytical Technology and as any analytical method, the validation of NIRS methods is a mandatory step at the end of the development in order to give enough guarantees that each of the future results during routine use will be close enough to the true value [4]. The success of NIR is also due to the advancement of computing, as many software able to handle a large amount of data that is generated by each sample have been

developed. With this software, one can easily apply several chemometric methods such as principal component analysis and least squares regression processing information of each sample.

3.2.2. Molecular Vibrations [15, 16]

Molecular vibrations can be classified into two types, axial deformation vibration (stretching) vibration and angular deformation (bending). The axial deformation or stretching are radial oscillations of the distances between the nuclei and the angular deformations involving changes of angles between the plane containing the link and a reference plane.

The major types of molecular vibrations are stretching and bending (**Figure 5**). Infrared radiation is absorbed and the associated energy is converted into these types of motions.

The stretch produces a change of bond length. This is a rhythmic movement along the line between the atoms so that the interatomic distance is either increasing or decreasing. Stretching can be either symmetric or asymmetric.

The second type of molecular vibration, bending, results in a change in bond angle. These are also sometimes called scissoring, rocking, or "wig wag" motions.

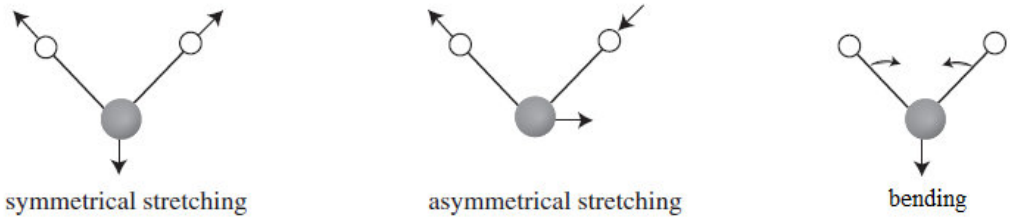


Figure 5 - Example of Stretching and bending vibrational modes for H_2O [16]

Each atom has three degrees of freedom, corresponding to motions along any of the three Cartesian coordinate axes (x, y, z). A polyatomic molecule of *n* atoms has 3*n* total degrees of freedom. In a nonlinear molecule, 3 of these degrees are rotational, other 3 are translational and the remaining correspond to fundamental vibrations; in a linear molecule, 2 degrees are rotational and 3 are translational.

The net number of fundamental vibrations for non-linear and linear molecules is therefore:

- Nonlinear Molecule: 3*n*- 6 degrees of freedom
- Linear Molecule: 3*n*- *5* degrees of freedom

Electromagnetic radiation may be considered as a wave with the properties of simple harmonic motion and may be defined in terms of the frequency of vibrations or the wavelength. The Harmonic oscillator model can be used to describe the vibration of molecules. Molecules consist of atoms combined by covalent bonds sharing electrons between them. Atoms oscillate through their bonds in a molecule like two balls attached to a spring.

The vibration frequency is dependent upon the atomic masses (m_1 and m_2) and the force constant (strength) of the interatomic bond. When the fundamental frequency of a specific vibration (ν), is equal to the frequency of the radiation, and when there is a vibrational change in the dipole moment, there will be a net transfer of energy in the forms of discrete packets, called quanta, from the radiation to the molecule. This causes the radiation to be absorbed, which excites a vibrational change in the molecule. The radiation energy should therefore be high enough to produce vibrational changes in the molecule. Using Hooke's law, defined as an atom shifting from its equilibrium position with strength proportional to the shift, the vibrational frequency (*f*) of a diatomic molecule can be determined:

$$f = \frac{1}{2\pi c}\sqrt{\frac{k}{\mu}}$$

Equation 1

Where, *c* is speed of light; *k* is the force constant of the bond between the two atoms; μ is reduced mass.

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

Equation 2

The variation of the potential energy with bond distance is a parabola centred about the equilibrium distance with evenly spaced vibrational energy levels. The energy of the different, evenly spaced levels can be calculated from:

$$E_{vib} = \left(\nu + \frac{1}{2}\right)\frac{h}{2\pi}\sqrt{\frac{k}{\mu}}$$

Equation 3

Since the selection rule for harmonic oscillator transitions is $\Delta\nu = \pm 1$ and energy levels are equally spaced, the energy difference between two consecutive levels will always be $E_{(\nu+1)}-E_\nu = f$, which is called the "fundamental frequency" of the band.

Vibration of actual chemical bonds cannot be explained using the harmonic oscillator model but rather using the model of an anharmonic oscillator by which energy levels are not evenly spaced. The anharmonic oscillator behaves like the harmonic oscillator but with the energy difference decreasing as ν increases (Morse's law):

$$E_{vib} = h\nu\left[1 - (2\nu + \Delta\nu + 1)\gamma\right]$$

Equation 4

Where, γ is the anharmonicity factor.

These changes from 0 to 2, 0 to 3 are known as overtones (first and second overtone correspondingly). Because of the decrease in energy change with the increaseν in the first overtone band will be an order of magnitude less intense than a fundamental band, the second less intense than the first and so forth.

Overtones appear between 780-2000 nm, depending on the overtone order and the bond nature and strength. The frequencies correspond to approximate multiples of the fundamental vibration.

Two or more vibrational modes can also interact to cause simultaneous energy changes, such as in the case of polyatomic molecules. These absorption bands are called combination bands. The frequencies will then be the sum of the multiples of each of the interacting frequency, for instance in the case of a triatomic molecule, there will be 3 first overtones, 3 second overtones etc.

In **Figure 6**, harmonic and anharmonic oscillator models are represented. The distance between vibrational levels are different in each model, i.e., in harmonic oscillator model the distance between levels are the same and in the other model the distance decrease. Besides that, it is possible to observe that for high internuclear distances between two atoms, anharmonic oscillator preview the dissociation but the other one don't.

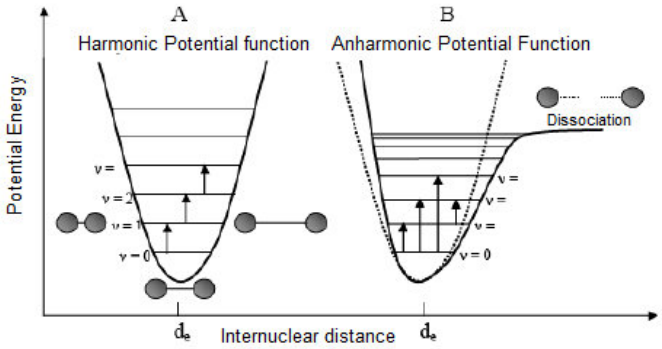


Figure 6- Harmonic Oscillator - Hooke's law (A) and Anharmonic oscillator- Morse's law (B) of a diatomic molecule[17]

3.2.3. Instrumentation [15, 16, 18]

A NIR spectrometer is generally composed of a light source, a monochromator, a sample holder or a sample presentation interface and a detector, allowing measurements.

The light source is usually a tungsten halogen lamp, since it is small and rugged. Detector types include silicon, lead sulfite (PbS) and indium gallium arsenide (InGaAs). The most expensive InGaAs detector combines the speed and size characteristics of the silicone detector with the wavelength range of the PbS detector.

Laboratory analyzers are intended for off-line or at-line measurements in quality control, research and other kinds of laboratories, i.e., high analyte sensitivity and reliability are required, while speed is of lower importance. Presently, grating and interferometer-based instruments are mainly in use for this purpose. The appropriate NIR measuring mode will be dictated by the optical properties of the samples. Transparent materials are usually measured in transmittance. Turbid liquids or semi-solids and solids may be measured in diffuse transmittance, diffuse reflectance or transflectance, depending on their absorption and scattering characteristics.

The company's spectrophotometer, microPHAZIR™, reads in diffuse reflectance.

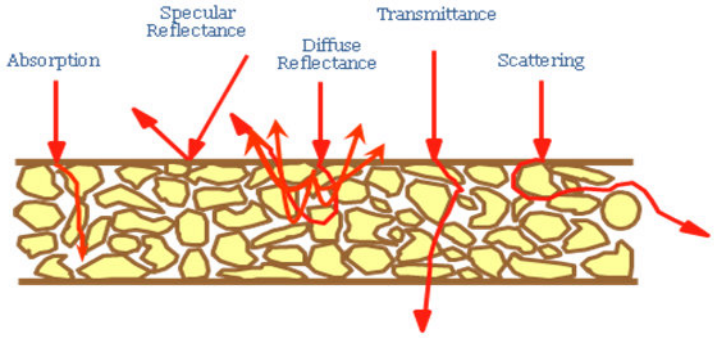


Figure 7 – NIR measuring modes *(retrieved from Thermo Fisher Scientific material support)*

The unit is based on the transmittance of the radiation absorption of the incident radiation and consequent reduction of intensity that reaches the detector, and the absorption determined from the difference between the detected intensity and incident intensity. Transmittance (T) is defined by **Equation 5**, where *A* is the absorbance, *I* the intensity after absorption transmitted intensity and I_0 the incident light.

$$T = \frac{I}{I_0} = 24^{-A} \qquad \textbf{Equation 5}$$

Diffuse reflectance occurs when radiation strikes a discontinuous matrix of the radiation is absorbed and another part is reflected in all directions, containing the spectrum information.

When radiation comes into contact with the sample, a part is absorbed by vibrational bands, and the fraction that is not absorbed can be reflected back. This last fraction is detected, allowing the calculation of the intensity of the absorbed radiation. Therefore, this method is used with samples in powder or fine particles which disperse well and absorb the incident radiation to a lesser extent and the reflectance (*R*) is reflected by the ratio between the intensity of reflected light (I_R) and the intensity of the incident light (I_0).

$$R = \frac{I_R}{I_0} \qquad \textbf{Equation 6}$$

The absorption of the radiation occurring in a sample obeys the Lambert-Beer law (**Equation 7**), which relates the absorption to the concentration of a particular constituent. In this equation, A is the absorbance, ε the molar absorptivity of each sample [$mol^{-1}L\ cm^{-1}$], *b* is the optical path of the radiation [cm] and *C* is the concentration [$mol\ L^{-1}$]. The absorbance of various constituents for the same wavelength is additive.

$$A = \varepsilon bC \qquad \textbf{Equation 7}$$

For diffuse reflectance, a theory by Kubelka and Munk, considers that the layer is composed of randomly distributed particles that evenly absorb and scatter the light, being much smaller than the size of the particle layer thickness. This is reflected by the Kubelka-Munk equation which yields a linear relationship between the reflection intensity and the concentration and assumes: infinite dilution in non absorbent matrix

coefficient and a constant spreading layer sample "infinitely thick", i.e. the increase thickness does not result in differences in reflection. This theory is expressed by **Equation 8**, where R is the reflectance values, *k* is the molar absorption coefficient, *C* is the concentration and *s* is the scattering coefficient of the sample.

$$F(R) = \frac{(1-R)^2}{R} = \frac{kC}{s} \qquad \textbf{Equation 8}$$

In any of the used methodologies (transmittance or reflectance), resulting spectrum from NIR spectroscopy are presented in absorbance, and depending on the data obtained are based on reflectance or transmittance is defined A= $\log\frac{1}{T}$ or A= $\log\frac{1}{R}$, respectively.

3.3 Chemometrics [15]

Chemometrics is an area of study that applies the mathematical and statistical chemical sciences, so as to delineate experimental procedures for maximum significant chemical information through the data methods and obtain knowledge of the chemical systems.

This technique allows the development of models to estimate properties for analysis of other properties. The quantitative analysis carried out in the laboratory are mostly time consuming and inexact, and are increasingly being replaced by instruments, including infrared spectroscopy techniques, which generate a large amount of signals. With the increasing sophistication of instrumental techniques, driven by the invasion of microprocessors and microcomputers in the chemical laboratory, new treatments of more complex data from a mathematical and statistical perspective, became necessary, in order to relate the obtained signals with the desired results. In the case of NIR spectroscopy, bands present in the spectrum arise from overlapping and combinations that generate a high number of signals, so it is necessary to use chemometrics multivariate analysis techniques for the determination of concentrations of the desired components.

The chemometrics have several applications including chemo-informatics, bioinformatics, image analysis, sensors and micro structures. As a major application area of NIR, the use of chemometrics algorithms involves using chemometric techniques and the pre-treatment of the NIR spectrum plotted in the development of qualitative and quantitative methods for the selection of the set of samples for calibration and set validation and identification of abnormal even samples (outliers).

3.3.1. Data pre-treatments

There are several spectrum pre-treatments that allow a significant improvement of the spectrum interpretation.

A concrete example of influence on the spectrum is the size of the particles in the sample, since the factor $\log\frac{1}{R}$increases with increasing particle size, due to the increase of the apparent optical path. As a consequence of the particle size effect, the baseline shifts, especially for wavelengths where high absorption occurs. For example, for two samples with the same composition but different particle size, there is more reflection of the smaller particles and the effects of specular reflection (reflection that occurs on the sample surface with the same angle of incidence) are reduced into smaller particles.

However this effect is not additive, but multiplicative, that is, the term is proportional to $\log\frac{1}{R}$ and can be attenuated in the multivariate calibration, using various methods of pre-treatment.

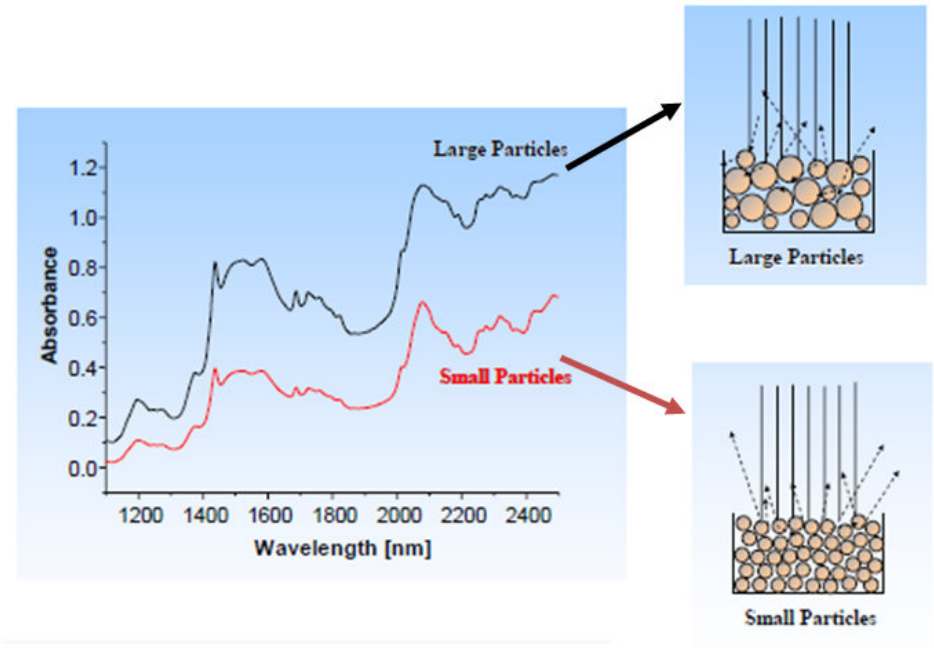


Figure 8 – Effect of particle size in spectrum *(retrieved from Thermo Fisher Scientific material support)*

For quantitative studies, pre-processing of spectrum reduces variations not directly related to the analyte concentration, such as random noise, baseline drift, light scattering, etc. Variable selection methods allow determining the spectrum region(s) where variations are specifically related to changes in the analyte concentration, so that uninformative variables can be excluded, thus enabling the construction of simpler and more robust models for routine analysis with NIR [19].

The most common spectrum pre-treatments are described below:

3.3.1.1. Derivatives

The spectrum derivation is one of the most widely used pre-treatment in NIR spectroscopy. It is used to minimize the overlapping bands and baseline variations problems. One of the most used methods is the one proposed by Savitzky-Golay (S. G.). The first derivative can correct baseline shifts and the second derivative corrects deviations which vary linearly with wavelength. A pre-processing step consisting of S.G. is very complex and relies on the convolution technique described by Savitzky and Golay in which a curve is fitted to a small section of the spectrum and then finds the slope of the tangent to this curve at the midpoint. Applying the Savitzky-Golay filter removes noise by applying a moving polynomial to the data.

3.3.1.2. Standard Normal Variate (SNV)

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

$$X_{i,m}^{SNV} = \frac{X_{i,m} - \bar{X}_i}{S_i}$$ **Equation 9**

Where $X_{i,m}^{SNV}$ *i, m* is the absorbance value from the row *i* (or spectrum *i*) and the column m (or variable m) once the pre-treatment is applied. $X_{i,m}$ is the original absorbance value from the *i* spectrum and S_i is the standard deviation from row *i* (spectrum$_i$).

3.3.1.3. Multiplicative Scatter Correction (MSC)

The Multiplicative Scatter Correction is a method used when multiplicative variations occur in the obtained NIR spectrum, correcting the effect of light scattering caused by the lack of homogeneity of the samples. It consists in calculating an average spectrum relating to all spectra in the set of calibration. Thereafter, each sample spectrum*i* (X_i) is transformed according to **Equation 24**, wherein X_i^{MSC} represents the transformed spectrum for the sample, i.e. coefficients *u* and *v* are chosen such that the difference between the transformed spectrum and the average spectrum is minimal.

$$X_i^{MSC} = \frac{x - b}{a}$$ **Equation 24**

For each spectrum (*x*) coefficients *a* and *b* are determined by linear regression so that the corrected spectrum becomes very similar to the mean spectrum of the calibration.

3.3.2. Multivariate analysis

Studies and investigations related to the scientific area usually generate a high amount of data. Given that they need to be treated to realize the completion of the assessment, the univariate study can give some ideas but does not respond with the desired interest. So for a good response, a study with multivariate analysis with the aid of chemometrics algorithms and concepts is necessary.

In multivariate analysis study for the qualification and quantification in NIR applications there are several algorithms, from which we highlight the principal component analysis, cluster analysis, regression by partial least squares, principal component regression, multiple linear regression between others.

However, this study essentially used two types of analysis: the principal component analysis as a qualitative method and partial least squares regression as a quantitative method.

3.3.2.1. Principal Components Analysis – PCA

The PCA is one of the most used chemometric techniques and consists in transforming the original variables into new variables called principal components (PCs). This powerful technique is used to obtain data reduction when there is correlation between them, providing the appropriate tools to identify the most important variables in the space of principal components.
The main components are the new variables generated by a mathematical transformation carried out in the original data matrix [15].

PCA is one of the multivariate methods of analysis and has been widely used with large multidimensional data sets. The use of PCA allows the number of variables in a multivariate data set to be reduced, while retaining as much of the variation present in the data set as possible[2, 17]. PCA is an unsupervised method of data compression and visualization widely used in NIR technology. As an unsupervised classification method, no information other than the NIR spectra are given to the algorithm, so the clustering occurs without orientation. This mathematical tool resolves the multivariate that, through linear combination, approximates the original spectrum. [15]

The coefficients (*loadings*) are chosen such that the new variables, unlike the primitives are not correlated among them. Since many variables are generated as new original, there seems to exist no data reduction. However not all the generated PCs are necessary to explain the variance of all data because of the PC characteristics. PC1 explains most of the variance in the data set, followed by PC2 which explains the second largest portion of the variance and so forth, i.e. while the original variables have the same statistical significance they have decreasing statistical significance, meaning that only a small number of PCs is needed to explain all the variance in the dataset. The other important feature is that each new main component is a linear combination, orthogonal to the previous one. The geometrical main components can be seen as projections from the original data in orthogonal axes (weights or *loadings*) covering the space variables, representing the main components, for example in a simple system of only two variables (and hence two main components), the main components can be seen as projections of the original data on orthogonal axes (weights or *loadings*) covering the space of the variables and representing a distance from this result (s*core*) [15].

The scores for each sample can be graphed constituting a map of the main components, and if the variance is retained on the first principal components, data may be represented on the map of principal components with two or three dimensions. These maps allow grouping samples with similar features and outliers detecting samples, i.e., samples which for some reason do not belong to the system [15].

PCA is a method that has the basic purpose of data reduction from linear combinations of the original variables and is therefore widely used in NIR spectroscopy, since the spectrum with over a thousand variables can be viewed in two dimensions. In addition to this, the model also allows only the relevant information to remain, excluding such variations associated with noise [15].

3.3.2.2. Partial Least Squares - PLS

The calibration process allows one to establish the relationship between the instrumental response and the property to be determined, using a set of representative samples. NIR spectroscopy provides large number response variables for each sample that may generally not be assigned to only one analyte. This has led to the development of methods to relate multiple variables with the property to be determined.

One method for quantitative analysis based on this mathematical transformation is Partial Least Squares (PLS) regression. This attempts to explain as much of the observed variation in the dependent variables as possible using the minimum of relevant factors contained in the spectrum data [20].

The PLS method is based on the reduction variable but unlike the PCA decomposition of the spectrum matrix, it takes place simultaneously with the decomposition of the matrix to determine the property.

In the development of calibration models of the parameters to be experimentally determined, the factors numbers define the model. The number of factors is a very important parameter because the greater number of factors, the greater the percentage of information included in the model and the lowest calibration errors, however the model's complexity increases. One must find a compromise between these parameters in order to obtain the simplest and most predictive model. One of the criteria for selection is the choice of number of factors that yields the lowest prediction error.

It is a technique of multivariate data analysis used to relate one or more response variables (Y matrix) with several independent variables (X matrix), correlating them to obtain a linear relationship, based on the use of factors (principal components). This is a model that determines quantitative correlation, therefore serving for constructing multivariate calibration curve.

3.3.2.3. Model Validation

Cross-validation is a validation technique based on the calibration data being used to evaluate the forecasting ability of PLS models for a given set of samples.

This cross-validation is performed by testing several sub-models, or, in this technique successively samples of the calibration set itself are eliminated. First, samples of the calibration data are eliminated. Thereafter, the calibration is carried out using the remaining samples before testing in the first sample compared with the actual y predicted by the model. The first sample is then placed in the calibration set and the second

sample is eliminated by repeating the procedure until all samples are submitted at same test. These parameters must be evaluated to test the quality of the model:

- **Root Mean Square Error of Calibration (RMSEC)**

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}i - yi)^2}{N-A-1}} \qquad \textbf{Equation 11}$$

Where N is the total number of spectrum collected and A stands for the number of factors used in the PLS model.

- **Root Mean Square Error of Cross Validation (RMSECV)**

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{Nc}(\hat{y}CVi - yi)^2}{Nc}} \qquad \textbf{Equation 12}$$

In **Equation 12**, $\hat{y}CVi$ is the cross validation estimation of sample I, yi the actual concentration of the samples and Nc is the number of calibration samples. RMSECV is a measure of the ability of the model to provide values of unknown samples.

- **Root Mean Square Error of Prediction (RMSEP)**

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{Np}(\hat{y}i - yi)^2}{Np}} \qquad \textbf{Equation 13}$$

In **Equation 13**, $\hat{y}i$ and yi are the PLS cross-validation prediction and measured reference values for i^{th}samples and the Np is the number of samples in the test set.

The difference between the expected and measured values can be expressed through the value of RMSEP. The number of the PLS model latent variables or factors (similar to PC) is established by the value of RMSECV obtained. A low value of RMSECV indicates a good model.

Other parameters such as the coefficient of determination (R^2) are also used to evaluate the calibration model.

In cross validation, all samples (*n*) are tested; one sample is left out, and then n-1 samples are used to perform the calibration and a prediction of the concentration of the sample which was not used for the calibration. This procedure continues until all samples have been deleted once. The minor error obtained by this validation is called RMSECV and is represented by **Equation 12**, where the number of calibration samples, *yi* the benchmark reference sample provided i.e. *yi* the predicted value for sample *i* given by the model are shown.

To terminate the PLS model, known concentration samples are needed, but these will belong to the validation set, not entering the calibration. This technique also gives an error, called RMSEP reflecting the difference between the predicted values of the sample *i* (*yi* predicted) and the reference values of the sample *i* (*yi* reference) of n samples of the validation set (**Equation 13**).

3.3.3. Quantitative Analysis

Quantitative methods are used to develop models to predict properties of unknown samples or the quantities present in the sample. For the construction of the model, these methods use the original spectrum or use the already pre-treated spectrum (e.g., reduction of variables).These methods can be classified as: linear or non-linear methods.

Before a NIR spectrometer can do any quantitative analysis, it has to be trained, i.e. calibrated using multivariate methods. The calibration process basically involves the following steps [18]:

1. Selection of a representative calibration sample set.
2. Spectrum acquisition and determination of reference values.
3. Development of a calibration model based on a chemometric multivariate procedure and determination of reference values f the analytical target property
4. Validation of the model
5. Prediction with calibration samples

In the **Figure 9** below, the methodology for developing a calibration model is summarized:

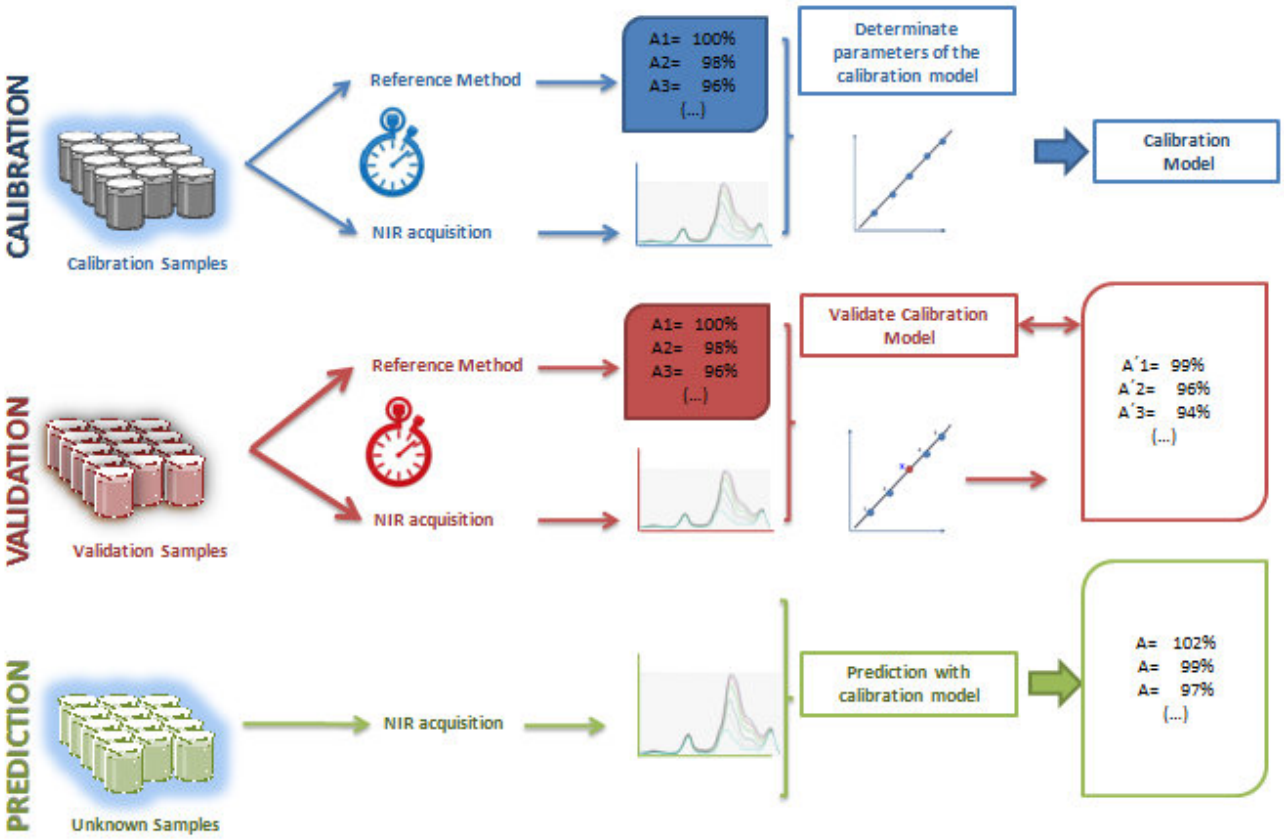


Figure 9 – Methodology for a calibration model developing

After constructing the model, it must be validated according to the guide lines in place.

The most frequently used multivariate regression methods in quantitative NIR analysis are principal component regression (PCR) and partial least-squares (PLS) regression. PCR uses the principal components provided by PCA to perform regression on the sample property to be predicted. PLS finds the directions of greatest variability by comparing both spectrum and target property information with the new axes, called PLS components or PLS factors [18].

In both cases, the optimum number of factors used to build the calibration model depends on the sample properties and the analytical target. Too many factors may lead to an "overfitted" model with a high regression coefficient and a low standard error of calibration (SEC), but a large standard error of prediction (SEP). Such model is not very robust and may fail when tested with an independent validation set [18].

3.3.4. Qualitative Analysis [16, 21, 22]

In qualitative analysis, sample properties that have to be related to spectrum variations have discrete values representing a product identity or a product quality, for example "good" or "bad" [22]. To solve the selectivity and interference problems of NIR spectrum, multivariate classification methods are used for grouping samples with similar characteristics. Multivariate classification methods, also known as pattern-recognition methods, are subdivided in "supervised" and "non-supervised" learning algorithms, depending on whether or not the class to which the samples belong is known.

3.3.4.1. Non-supervised classification

In non-supervised methods, also known as cluster analysis, the samples are classified without a prior knowledge, except the spectrum. These methods do not require any a priori knowledge about the group structure in the data, producing the grouping instead, i.e. clustering, itself. This type of analysis is often very useful at an early stage of an investigation to explore subpopulations in a data set, for instance different physical grades of a material. Cluster analysis can be performed with simple visual techniques, such as PCA or some hierarchical methods leading to so-called dendrograms.

3.3.4.2. Supervised Classification

"Supervised classification" methods, also known as discriminant analysis, are used to build classification rules for a number of pre-specified subgroups, i.e. the group structure of the training set is known. The classification rules are later used for allocating new or unknown samples to the most probable subgroup. Identity or good/bad quality are, thus, defined as belonging to a group with known properties. Algorithms of this type such as LDA (linear discriminant analysis), QDA (quadratic discriminant analysis), SIMCA (Soft Independent Modelling of Class Analogies) or KNN (K nearest neighbours) are typically used for constructing spectrum libraries [18].

Qualitative analysis by NIR spectroscopy usually relies on the use of spectrum libraries constructed by using one of the above methods. Such libraries are normally constructed using qualitative analytical tools included

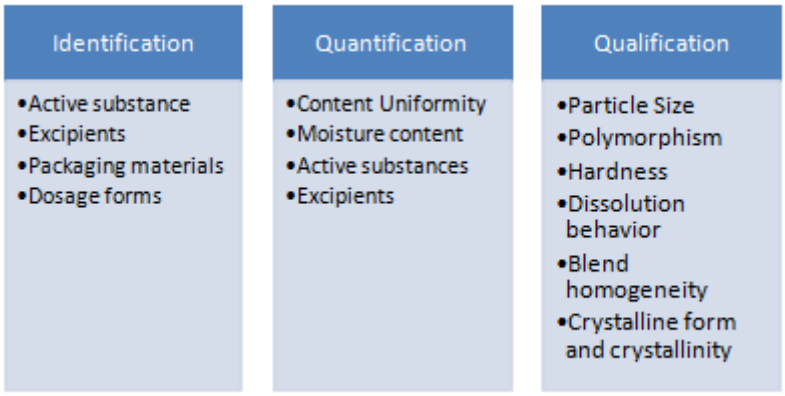
in the software bundled with commercially available equipment. Appropriate use of NIR libraries allows the chemical identification of products, simply by comparing the correlation coefficient between the spectrum for an unknown sample and those contained in the library, but also to determine whether they possess the desired physical properties.

3.4. NIR Applications [18]

NIR spectroscopy combined with multivariate data analysis opens many interesting perspectives in pharmaceutical analysis, both qualitatively and quantitatively. Fast and non-destructive NIR measurements without any sample pre-treatments may increase the analytical throughput tremendously [21, 23]. The use of fiber optic probes offers the opportunity for in-line and on-line process monitoring. The special feature of combined chemical and physical information allows for the assessment of a "spectrum signature" of raw materials, intermediates and final dosage forms, which in turn offers the possibility of a simultaneous determination of several sample characteristics.

NIR combined with multivariate data analysis has found increased use in the pharmaceutical industry. The most common application in this industry is the identification tests, but in the last decades it has proved its usefulness in other applications, like qualification, quantification tests and process monitoring and control.

Within the last years a growing number of research and review articles have reported on the great potential of NIR spectroscopy in pharmaceutical research, production, and quality control focusing on various "analytical targets", such as identity, content uniformity, moisture content, particle size, polymorphic and pseudopolymorphic forms, hardness, thermal and biopharmaceutical properties.



These applications can be applied in different applications:

- Off-line / at-line
- On-line / in-line

The Process Analytical Technology (PAT) initiative, issued by the Food and Drug Administration (FDA) in 2003, recommends to the pharmaceutical industry the development and employment of new technologies in the production and qualification of pharmaceutical products [24]. PAT clearly opens new perspectives to

the implementation of innovative technologies in quality and process control of raw materials, intermediates and final products. The PAT frameworks states that the monitoring efforts should be focused directly on the process, and not after each production step. PAT is aimed at promoting the analytical tools in pharmaceutical unit operations for monitoring and better understanding of critical process parameters [25,26,27].

3.4.1. Regulatory Aspects [16, 28]

NIR spectroscopy has a large number of advantages over other analytical techniques, and, thus, offers many interesting perspectives in pharmaceutical analysis. Actually, the major pharmacopoeias have generally adopted NIR techniques. The European and United States Pharmacopoeia both contain a general chapter on near-infrared spectrometry and spectrophotometry, respectively. These chapters address the suitability of NIR instrumentation for use in pharmaceutical analysis focusing mainly on operational qualification and performance verification comprising wavelength scale and repeatability, response repeatability, photometric linearity, and photometric noise. Only some limited guidance is provided in terms of developing and validating an application. The general legal requirements for instrumentation qualification procedures, namely design qualification (DQ), installation qualification (IQ), operational qualification (OQ), and performance qualification (PQ), are described in the Good Manufacturing Practices guideline (GMP).

Many pharmaceutical companies have successfully implemented NIR spectrometers in their quality control laboratories for routine use in raw material identification and qualification. This is based on the fact that major pharmacopoeias allow manufacturers to use analytical methods other than compendial ones for compliance testing, provided that they are validated according to parameters such as specificity, linearity, range, accuracy, precision, repeatability, reproducibility, detection limit, quantification limit, and robustness, as detailed in the U.S.P. Chapter 1225 on Validation of Compendial Methods and the general ICH Guidelines Q2A and Q2B on Validation of Analytical Procedures.

A number of recent sets of guidelines now advise the use of NIR spectroscopy as an universal method suitable for the identification of raw materials and other applications:

• European Pharmacopoeia, chapter 2.2.40.

• United States Pharmacopeia, chapter <1119>

• EMA, Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submission and variation

The EMA guideline is a new guideline specifically emitted for NIR methods (quantification, qualification and identifications methods): Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations. This guideline provides guidance on the use of NIR spectroscopy procedures including development, calibration and validation, when used with chemometric statistics, for qualitative and quantitative analysis.

This guideline aims to relate how NIRS should meet the requirements placed upon all pharmaceutical analytical procedures, taking into account the use of chemometric statistics and that NIRS procedures also

need to be robust with respect to the expected variability of the materials or products to be analyzed and the manufacturing processes used to prepare them.

This guideline introduces the concept of the NIRS procedure scope to facilitate continuous improvement and life cycle management. The NIRS procedure describes how the NIRS method and model are used for the intended purpose, within its defined scope. This guideline defines two different categories: model and method, considering that the method is the description of the key elements, mainly within the NIRS apparatus, which enable NIRS measurement of the analyte of interest and the model is the description of how the NIRS spectrum data measured using the NIRS method are related to the analyte of interest, generally employing chemometric software.

In addition, the EMA has published an addendum to this guideline with a goal to clarify the scope of the application of the guideline.

3.5. K-nearest neighbour algorithm (KNN)

Algorithm KNN is a linear and non-parametric supervised pattern recognition method. An unknown sample in prediction set is classified according to the majority of its K-nearest neighbours in training set. Parameter K has a great influence on the identification rate of KNN model, and the optimal value of K is determined in the KNN calibration process. Parameter K value is chosen according to such a way that the subsequent KNN classification yields optimal results are measured by a minimum prediction error The prediction errors for a given set of K values are estimated by cross validation, and the value of K that gives the highest prediction rate should be selected [29].

The distance between an unknown sample and all objects of the training set is calculated. The unknown sample is identified into a class with the target to which the distance is the nearest. For the KNN classifier, it is necessary to have an appropriate training set which is not too small, and a good discriminating distance could be obtained [30].

4. MATERIALS AND METHODS

All of the following described methods were developed and validated according to EMA's NIR guideline "Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submission and variation".

4.1. Materials and Equipment

In this work, a microPHAZIR™ Handheld NIR Analyzer spectrometer was used and data processing was performed with microPHAZIR™ software Polychromix Method Generation version 3.101 [31]. The PolychromixPhazir™ is a multi-purpose, hand-held spectrometer suitable for many chemometrics measurements [15].

The microPHAZIR™ was developed using Thermo Scientific's unique MEMS (Micro-Electro-Mechanical System) technology that enables the construction of spectrometers that have no moving parts, are small, and use little power. These characteristics are ideal for the fabrication of hand-held instruments. The spectrum range is between 1600-2400 nm and measurement interface is diffuse reflectance [31].

Figure 10 - microPHAZIR™ [31]

The principal measurements parameters of microPHAZIR™ are:

- Optical Resolution: 11 nm
- Spectrum range: 1600 – 2400 nm (6250 – 4170 cm^{-1})
- Measurement Interface: Diffuse reflectance
- Detector Setting: Single InGaAs

All samples involved in this work (Quantitative and Qualitative method development) existed in the Quality Control Laboratory, OM Pharma and were stored in an environment monitored room (temperature and humidity).

The samples set used to build the quantitative and qualitative model were previously analyzed by a reference method.

4.1.1. Samples – Quantitative Method

The construction of the calibration model was based on acquisition of spectrum of real and real contaminated samples. Determinations by NIR spectroscopy require the use of multivariate calibration models encompassing all potential sources of variability during routine analysis. Calibration models constructed using this strategy have advantages, because variability is necessary to build the model. This fact is very important to build a robust NIR calibration model; the calibration set has to contain the future expected variability that the model will meet in routine.

Calibration models were constructed, essentially, by using sets of samples obtained from different batches. So, real samples spectra were acquired over a period of six months. The acquisition of the NIR spectrum from each sample was performed while the sample was tested with reference method.

Variability different sources of variability such as number of batches, API concentration, different operators and temperature, are considered while developing the model. The spectrum acquisition was performed with an objective to cover all this variability.

The goal consisted in building a calibration model in a 80% (w/w) - 110% (w/w) in API range with 70 different batches (real samples and powder blend preparation) containing 80%, 85%, 90%, 95%, 100%, 105% and 110% (w/w) of API.

Commercial samples (finished product) invariably span a narrow API concentration range (around ± 5% of the label claim), so the blends of these samples are insufficient to construct calibration models. Since among all produced batches there is a high similarity between the samples, due to a well optimized manufacturing process, leading to very homogeneous and very close control parameters samples, there were no samples in the extremes. The variability offered by real samples was low comparatively to the concentration range. The real samples ranged between 97% - 102%, and this range is not sufficient to include the product specification. So, the preparation of powder mixtures was necessary.

Therefore, the API content range was expanded by using additional laboratory samples prepared by under-dosing (placebo addition) and overdosing (API addition) of the blend. The API addition in samples had the goal of increasing API concentration (105%–110%). Under-dosed samples were also prepared from production blend which, were supplied with known amounts of a placebo mixture containing the excipients in the same proportions as in the pharmaceutical preparation to obtain samples spanning the API concentration 85% - 95% API range.

Doping of blend samples with API was performed using five different batches of API; samples contaminated with Placebo involved the use of two different batches of each. A total of 70 samples were used to develop the method, of which 33 were real samples and 37 were placebo/API spiked samples. Doping of samples was conducted in order to maintain the real proportions of each component in the formulation.

To the spectrum acquisition, aliquots from production and spiked samples were placed in a plastic polyethylene bag and the reflectance spectra were recorded. For each sample, ten spectra were acquired in different positions inside the bag with the goal to increase the variability of the response of the sample. Between each spectrum acquisition the sample was stirred inside the bag.

4.1.1.1. Pharmaceutical Formulation and Manufacturing Process

The studied formulation consisted of a blend containing 500mg/555mg of API and 55mg/555mg of placebo (containing two excipients). This mixture is a production pre-phase with the goal of manufacturing capsules.

The production process, in summary, involves the addition of API and the other remaining ingredients in an appropriate mixer.

In routine, according to the product manufacturing instructions, when mixing phase is completed the In Process Control picks a representative sample of the batch with the purpose of being examined by the Quality Control and to evaluate the batch compliance with the specification. This is a long process and it is interrupted while waiting QC (Quality Control) analysis. The waiting time is about thirty minutes.

4.1.2. Samples – Identification Method

The samples used in Library development have been previously released in the Quality Control laboratory using the existing and approved methods and the results met the identification test specifications according monograph. Besides that, all certificates of analysis were verified against the results obtained for the different batches and the results were confirmed.

To select the batches of raw material to build the library, a survey of stored raw materials within the term of validity was carried out. A criterion of at least 3 batches for each type of raw material was determined to construct the library.

The majority of the samples were taken from the Raw Material archive at OM Pharma. To select each batch used in the library construction, a critical factor assessment of each raw material was performed, based on the size particle, water content and colour and/or appearance.

Due to the variability associated with each raw material, and since the physical characteristics differ inside the raw materials group, the amount of batches for each one varies in accordance with the respective variability. The reference samples must cover a vast range of variability so that the library is ready for these variations during the routine identifications.

Two sets of samples were used: one for the library construction and an independent on for validation purposes.

To the spectrum acquisition, aliquots from each raw material/batch were placed in a plastic polyethylene bag and the reflectance spectra were recorded. For each sample, five spectra were acquired in different positions inside the bag with the goal to increase the variability of the response of the sample. Between each spectrum acquisition the sample was stirred inside the bag.

Due to confidential character, during this work the raw materials will be identified by letters A to J. In the following table (**Table 1**) the total batches of each raw material used for construction of the library and its validation are summarized.

Table 1 - Batches of the raw materials in Library construction

API- raw material	Total Batches number
A	15
B	12
C	6
D	12
E	3
F	6
G	4
H	3
I	5
J	3

4.2. Methods

For the reference quantitative method to determinate the content in API in blend a titrant reagent was necessary, samples were weighed on an analytical balance and the sample dilution was performed with appropriate equipment. Titration was performed with an automatic burette; the time spent with this method is about 30 minutes.

For the identification method the raw materials were analysed in according to the respective monograph.

Prior to collecting data, the internal instrument verification test (Performance Qualification) has to be performed each day, in order to assure that the equipment is working properly and they are in agreement with the equipment PQ.

The spectrum acquisition involved in this work was made as shown in **Figure 11**.

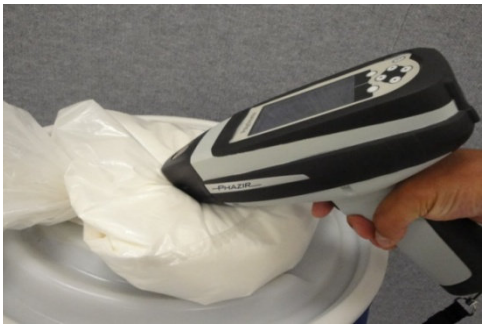


Figure 11 – Spectrum acquisition mode

4.2.1. Quantification Method

Different spectrum pre-processing methods were used in order to enhance the information searched for the study, to decrease the influence of the side information contained in the spectrum and finally to construct the calibration model. The pre-processing methods were applied in combination with the whole spectrum or different spectra regions.

Several models were developed to evaluate the most appropriate model. The selection of the model with the best predictive capacity was based on the investigation of number of principal factors and the calculation RMSECV, RMSEP and R^2. To evaluate the results, RMSECV and determination of R^2 were considered. For a good model the RMSE values should be low and R^2 should be as nearest of one as possible and small differences between RMSEP and RMESCV must be observed.

The results obtained for the quantitative method are shown in **Table 5**, **Chapter 5.**

4.2.2. Identification Method – API Library

To ensure the identity of a raw material, some of NIR´s advantages are of utmost interest: speed, non-destructive identification, direct measurement without the need of sampling. The chemical identity of a raw material is confirmed by comparing spectrum of a spectrum library, nonetheless, this library must be built and validated allowing the confirmation or non-confirmation of the raw material identity.

The NIR identification library was developed for ten solid APIs. The experimental part of this work was divided into three sections: sample selection, library development and library validation (internal and external). For library construction the following steps were considered, as summarized in **Figure 12**:

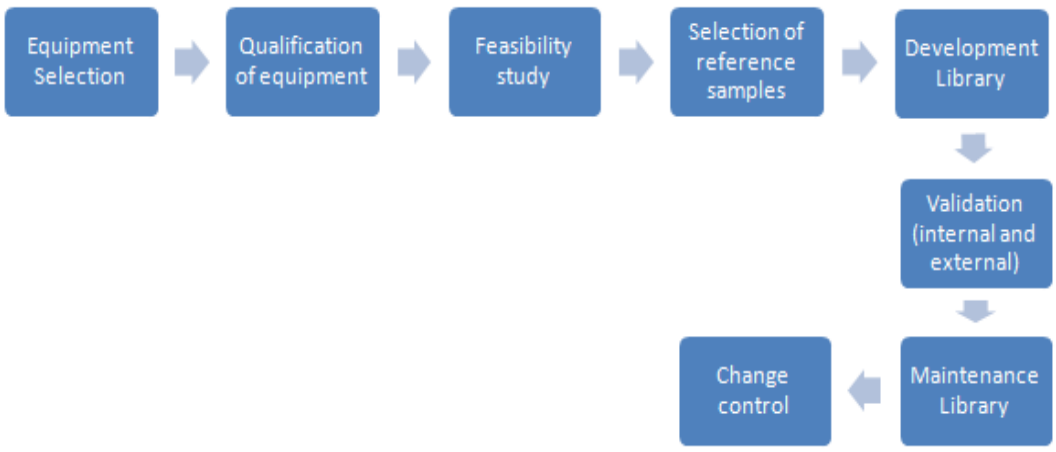


Figure 12 - Library construction steps

4.3. Library development

The library construction involves mathematical treatments to reduce the spectrum complexity. In this work's particular case, the used algorithm was the one available with the software: K-Nearest Neighbour (KNN) algorithm. To build the library, several parameters were taken into account, and these are listed in the table below:

Table 2 – Library development

Stages	Library Development
Purpose	• Library scope definition
Selection of samples/spectrum for calibration set	• Spectrum data should be acquired for the calibration sets; • Sample variability may be built into the library (moisture, particle size, residual solvents, degradation products, other physical/chemical properties, retained samples, temperature (especially for liquids), operator, presentation, etc.)
Display data	• All spectrum should be visually examined to check if there are anomalies or the presence of outliers; • Potential outliers must be investigated and only excluded for valid analytical reasons, and any exclusion must be documented.
Calibration set selection	• The number of samples required for each material group will depend on the discriminate algorithm used and the complexity of the application
Data pre-processing	• If necessary, data can be mathematically treated to reduce spectrum complexity
Selection of wavelength range	• Specifically range can be selected
Library construction	• Library structure is software dependent and users requirement; • Library can also be split into sub-libraries to ensure the required level of specificity and the transformations must be the same within each sub-library but may be different from each other;
Algorithm selection	• The choice of the algorithm is dependent on the user or equipment • It is recommended to use the simplest available algorithm
Threshold determination	• Initially, internal validation should be performed using the software default values, or those recommended by the manufacturers. • Library thresholds can be modified following internal validation of the library, • Once the threshold values have been set, the internal validation should be repeated to prove acceptable discrimination between different groups while maintaining acceptance of a material to its group. This may be an iterative process. • A threshold below 0,95 is not acceptable for the algorithm use.

A sufficient amount of samples for each raw material was introduced into clear polyethylene bags to perform the acquisition of the spectrum. For each raw material batch, five spectra were collected by three different operators in different days.

A reference library was constructed with NIR spectrum of 10 different API. All spectrum obtained were examined to check for absence of anomalies or the presence of outliers and no problems were found. All materials were incorporated into one library.

4.4. Library Validation (Internal and External)

Besides the library performance assessment, the parameters definition is needed for its validation, such as specificity and robustness parameters.

The goal of an analytical procedure validation is to ensure that it is suitable for its intended purpose. For library validation it is necessary to perform an internal and external validation. In the following table all the important parameters that must be accounted for both validations are listed.

Table 3 – Library Validation

<table>
<tr><td>Internal Validation</td><td colspan="2">• Evaluation of Library performance using the same the samples used to create the library;
• Typical steps:
• Verification that spectrum used to create the library are correctly identified;
• Confirmation that the distributions for materials in the library do not overlap;</td></tr>
<tr><td colspan="3">After successful internal validation, performance databases should be verified using samples that didn’t generate the database.</td></tr>
<tr><td rowspan="2">External Validation</td><td>Specificity</td><td>• Potential challenge should be presented to the database (reference library). The challenge should be rejected;
• Samples of materials represented by the library, but not used to create it, must give positive identifications when analyzed;
• NIR should be sufficiently specific to discriminate between batches that comply with the tests parameters and batches that do not, in the same way as for the reference method.</td></tr>
<tr><td>Robustness</td><td>• This tests the effect of minor changes to normal operating conditions on the analysis. Typical changes:
• Effect of environmental conditions (e.g. temperature, humidity) on the analysis;
• Effect of sample temperature on the analysis
• Changes in pre-processing and calibration algorithm parameters (e.g. derivative gap/segment, distance threshold etc.)</td></tr>
</table>

The internal specificity refers to the ability of the spectrum library to properly identify the samples, from which the library was derived, as a match for that compound. This internal specificity is dependent on the samples chosen to generate the spectrum library.
External specificity refers to the proper identification of samples, which were not used to generate the spectrum library. External validation is performed with independent batch that were not used in the construction of the library. The spectra of unknown samples were compared to that library.

The external validation was performed as follows:

- **Specificity**: Ability to assess unequivocally the analyte:
    - **Positive control** measurement precision: the library was challenged 5 times with one unknown sample of the raw material to be identified;
    - **Negative control** measurement precision: the library was challenged with two unknown samples;
- **Robustness**: To assess possible critical factors in the acquisition of the spectrum and, consequently, in the response of the model, the following was evaluated:
    - **Different Operators –** The spectrum acquisition was performed by three different operators;
    - **Changes in Pre-processing –** In the library constructions more than one pre-processing was tested with the goal to evaluate which of them gave the best response;
    - **Different Temperatures (samples) –** The spectrum acquisition was performed with different test samples temperatures;
    - **Different Temperatures (environment) –** The spectrum acquisition was performed with different room temperature;
    - **Different in bag thickness –** The evaluation of the thickness of the bag for the acquisition of the spectrum for later identification was tested;

All batches used for validation have been released and complied with the existing primary methods.

4.5. Library Maintenance

If new suppliers of some API are approved by OM Pharma, a sample of these products should be assessed by the built library. If library shows to be able to identify the product, a re-validation is required to increase the variation of information on the construction of the library, even if it proves able to identify. If this does not prove suitable for the identification of this raw material, library enrichment is required and necessary.

5. RESULTS AND DISCUSSIONS

5.1. QUANTIFICATION MODEL

The aim of this study was to develop a NIR chemometric method suitable for the direct quantification of API in powder blends.

5.1.1. NIR spectrum Acquisition - Interference Evaluation

Evaluation of the interference in the spectrum acquisition is a very important and critical factor. A good knowledge of the possible interferences in the sample spectrum acquisition assures a good result in the development of the calibration model.

Preliminary experiments were carried out to examine the effect of possible interferences. The ICH Q9 Guideline describes the principles and tools of quality risk management to guide development efforts. A risk assessment was conducted by involving process development internal experts, analytical development, API pilot plant and manufacturing.

An Ishikawa Diagram (**Figure 13**) was created to map the potential variables that can have an impact on the desired NIR method quality attributes. In total, six categories were identified: equipment, measurements, environment, material, operator and bag.

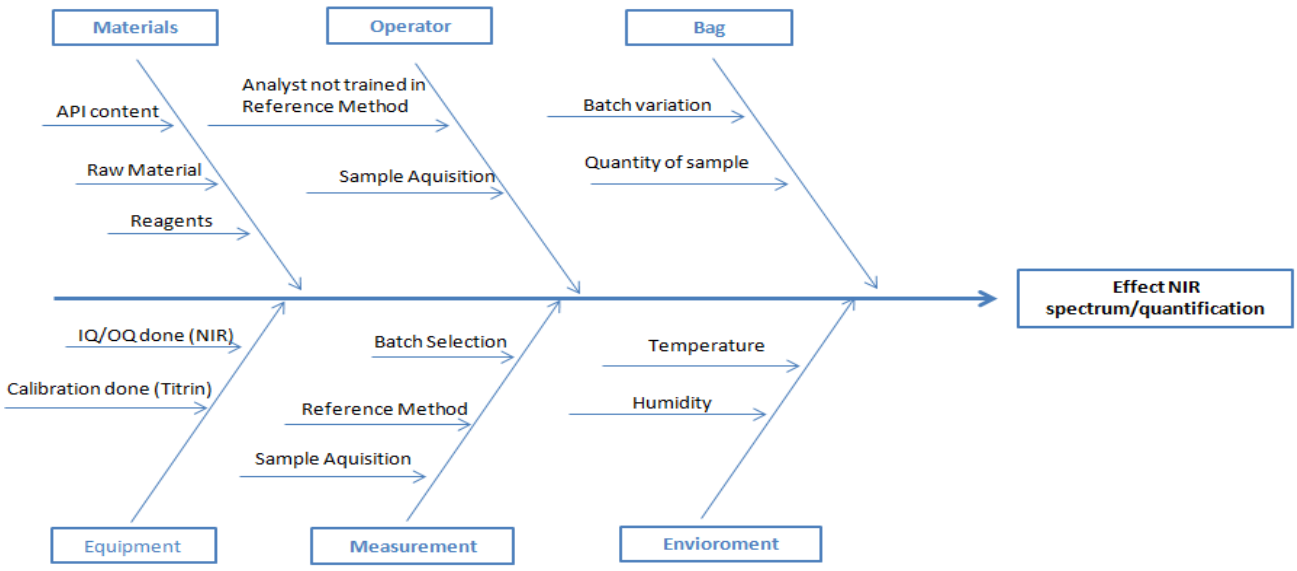


Figure 13 - Ishikawa diagram in critical factors evaluation

For this work, each of the possible interferences have been carefully evaluated and, if possible, tested.

In the following sections the impact in the spectrum data of each mentioned factor (**Figure 13)** is evaluated. It follows, in each case, a comparison of the raw data (untreated spectrum) and pre-treated spectrum (spectrum with the pre-treatment method used for quantification). The possible interferences detected by

the Ishikawa diagram (**Figure 13**) were evaluated by a PCA. This analysis was carried out to assess clusters of samples/results, making it possible to evaluate whether or not there is an interference of each parameter in the analysis.

5.1.1.1. Materials

This is an important variable in the quantification model development, being the one holding all the primary information for its construction, namely the API content of each tested sample, the raw materials and reagents used for the API determination by the reference method.
The raw materials involved in the formulation (API and excipients) were obtained regarding the quality requirements in the European Pharmacopeia or other supporting document.

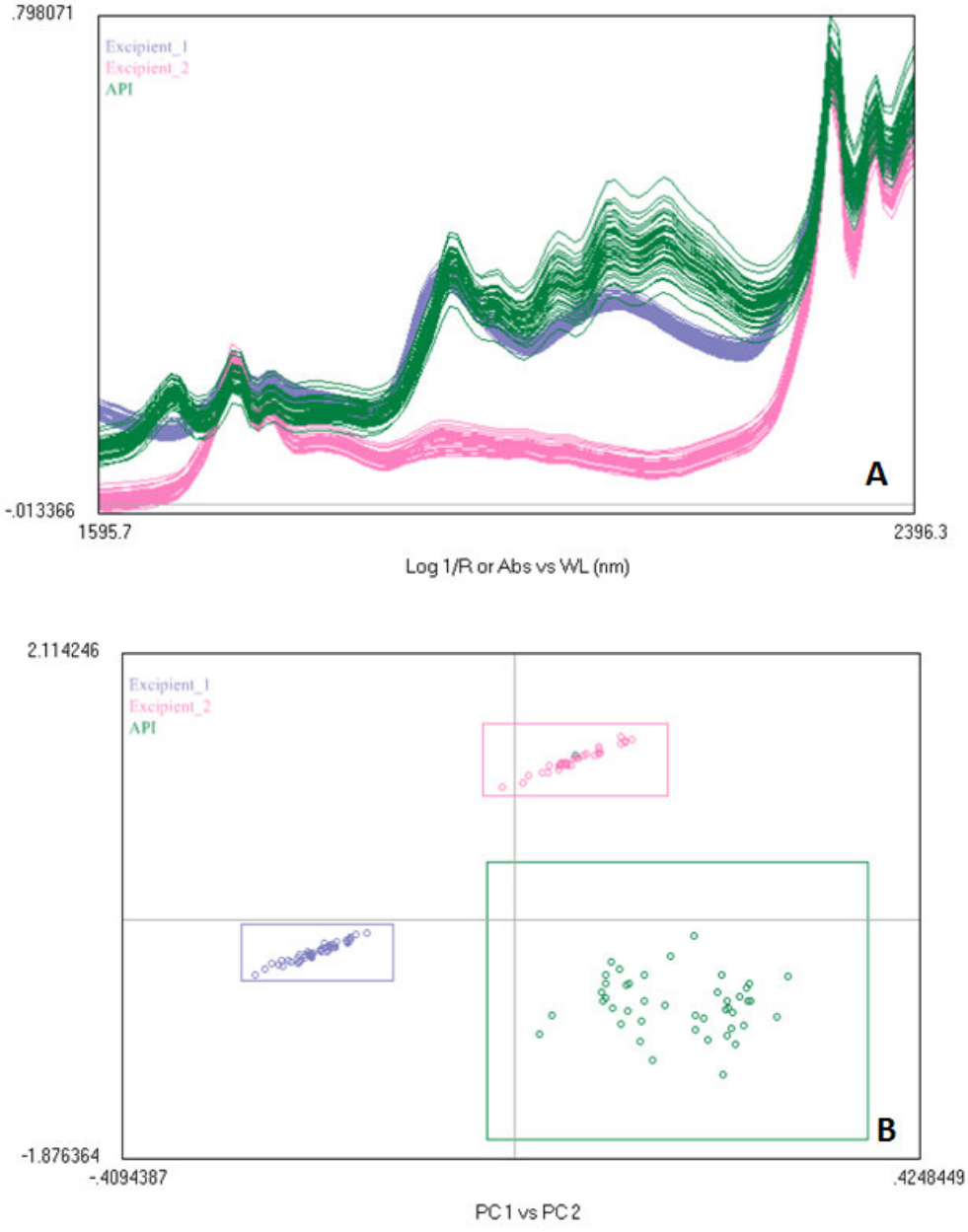


Figure 14 – **A:** Formulation raw materials untreated spectrum; **B:** PCA score plot of formulation raw material, no-pre-treated spectrum (PC1 x PC2)

To evaluate the influence of the formulation raw materials, 10 random batches were selected. Looking at the untreated spectrum in **Figure 14 A**, a difference between them is clearly visible. These raw materials have significantly different chemical structures, explaining the different spectrum. This difference is also visible in the principal components analysis, as seen on **Figure 14 B**, where three different groups were obtained.

The obtained spectra were submitted to pre-treatments in order to reduce its dispersion, as well as to obtain individual clusters for each component of the formulation. Nonetheless, each cluster variance is lower when comparing to the raw, untreated data. These results are shown in **Figure 15**:

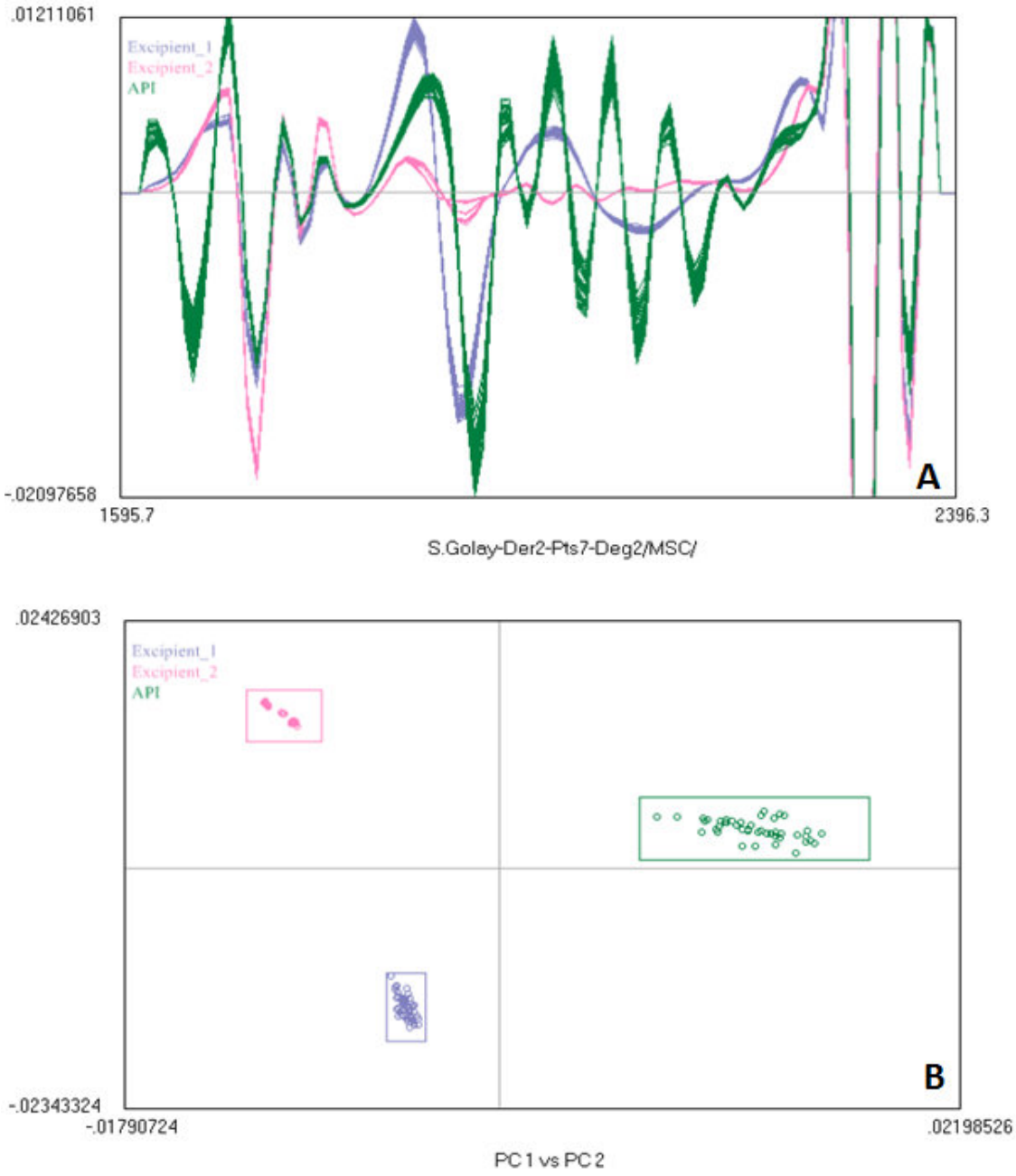


Figure 15 – **A:** S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) followed by MSC pre-processed NIR reflectance spectrum of raw material. In this plot, X and Y axes represent the wavelength and absorbance, respectively; **B:** PCA score plot of components analysis of formulation raw material, pre-treated spectrum (S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC) (PC1 x PC2)

Looking at **Figure 15**, one can identify the formation of groups of each raw material present in the formulation. This point out to the fact that each raw material is characterized by different components and that none overlaps the main raw material, the API.

5.1.1.2. Operator

The samples quantification by the reference method was performed only by a trained analyst. This train includes equipment, software and method train.
Three batches spectrum were obtained by two analysts, in order to evaluate the possible variations due to the spectrum acquisition.

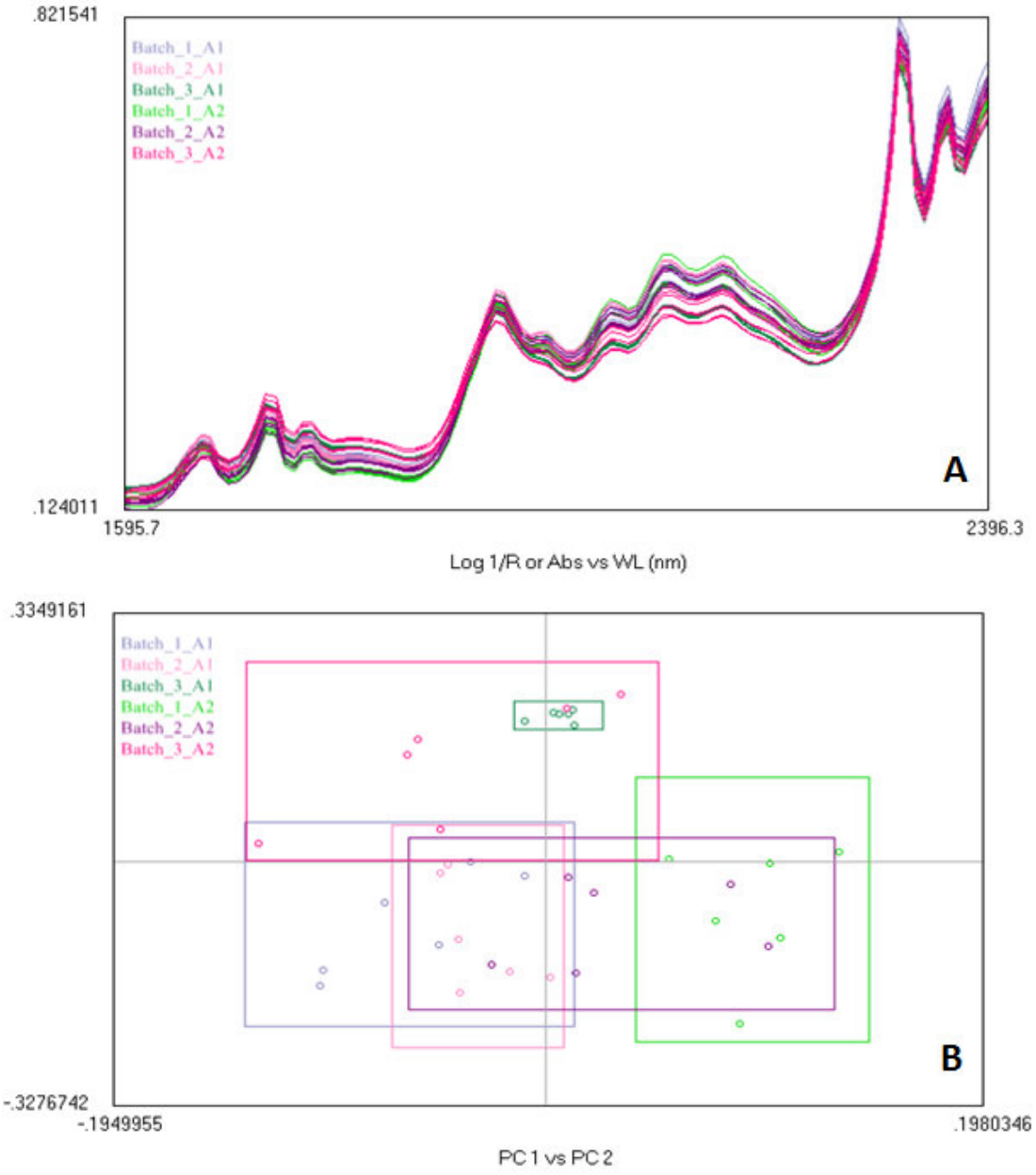


Figure 16 – **A:** Untreated spectrum of the acquisition of three batches by two operators; **B:** PCA score plot analysis of measurements performed by two operators at the same batch of blend, no pre-treated spectrum (S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC) (PC1 x PC2)

Even though the raw data allow a proper separation of the individual analysis, this can be improved with a spectrum treatment, reducing the clusters variance. These results are shown in **Figure 17**:

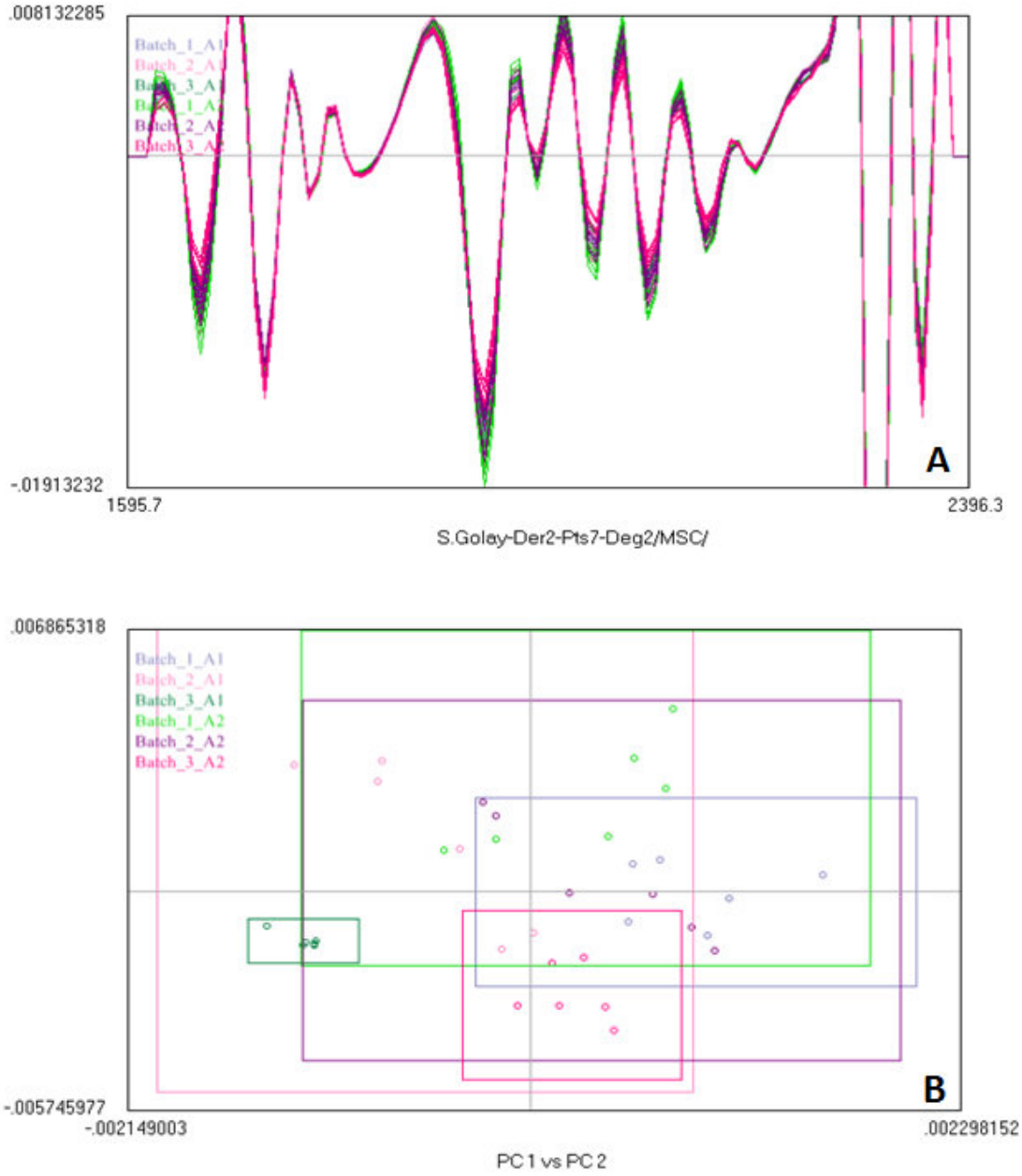


Figure 17 – **A:** S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) followed by MSC pre-processed NIR reflectance spectrum of measurements performed by two operators at the same batch of blend. In this plot, X and Y axes represent the wavelength and absorbance, respectively; **B:** PCA score plot analysis of measurements performed by two operators at the same batch of blend, pre-treated spectrum (S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC)

Looking at the PCA data, there is no evidence of this factor bringing any variability to the obtained results, showing a reduced variance associated with each cluster, thus not being a critical factor for the process.

5.1.1.3. Polyethylene Bag

The bag interference was evaluated through the spectrum acquisition of three available bag batches and a principal component study.

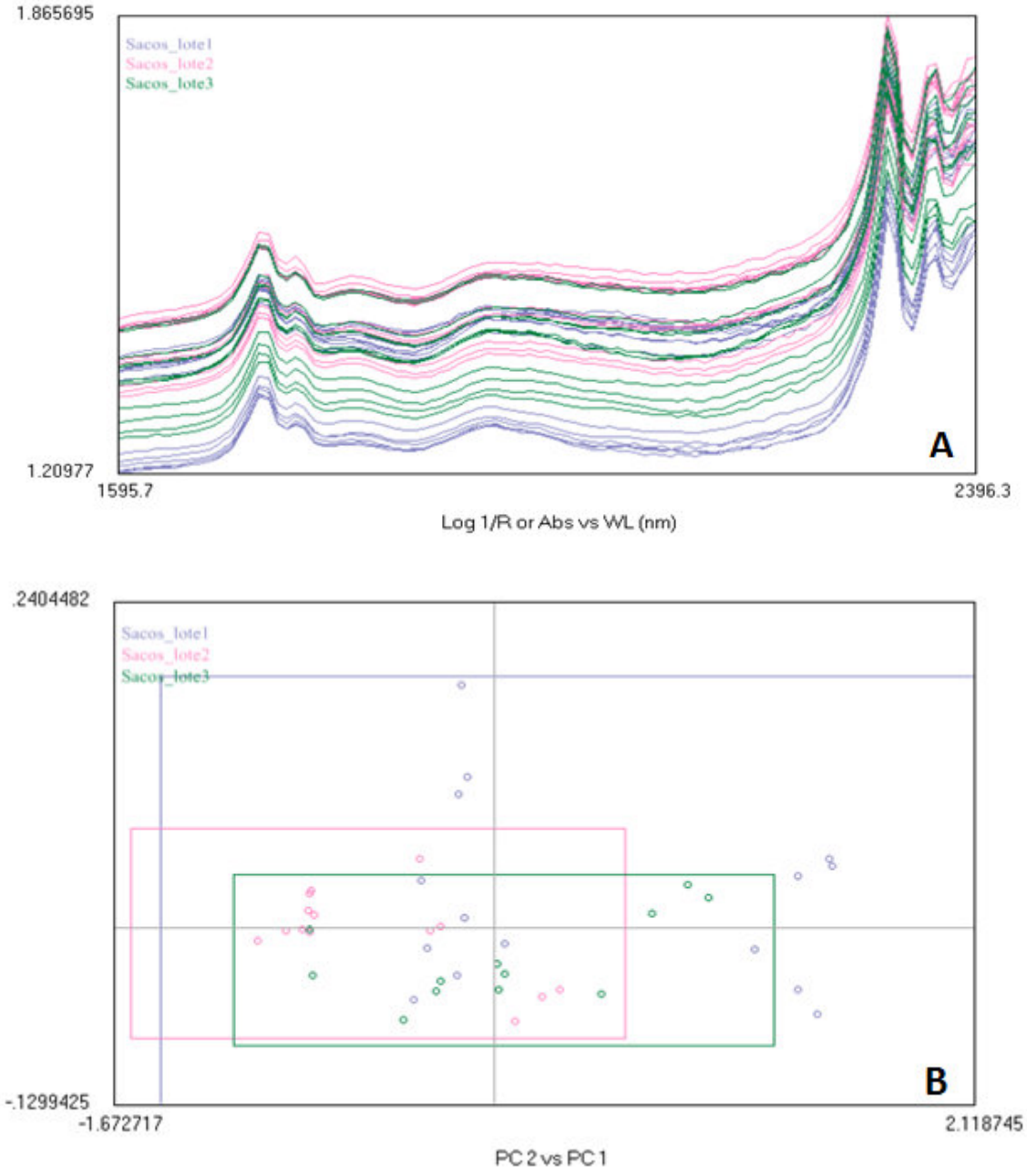


Figure 18 – **A:** Evaluation of inter-batch bags interference, no pre-treated spectrum; **B:** PCA score plot analysis of measurements performed inter-batch bags, no pre-treated spectrum

The spectrum dispersion is significant when changing the bags batches, as well, as in the same batch. The PCA analysis with the raw data shows this dispersion. In order to improve this, the data was submitted to a pre-treatment so the variability could be reduced, as seen on **Figure 19**.

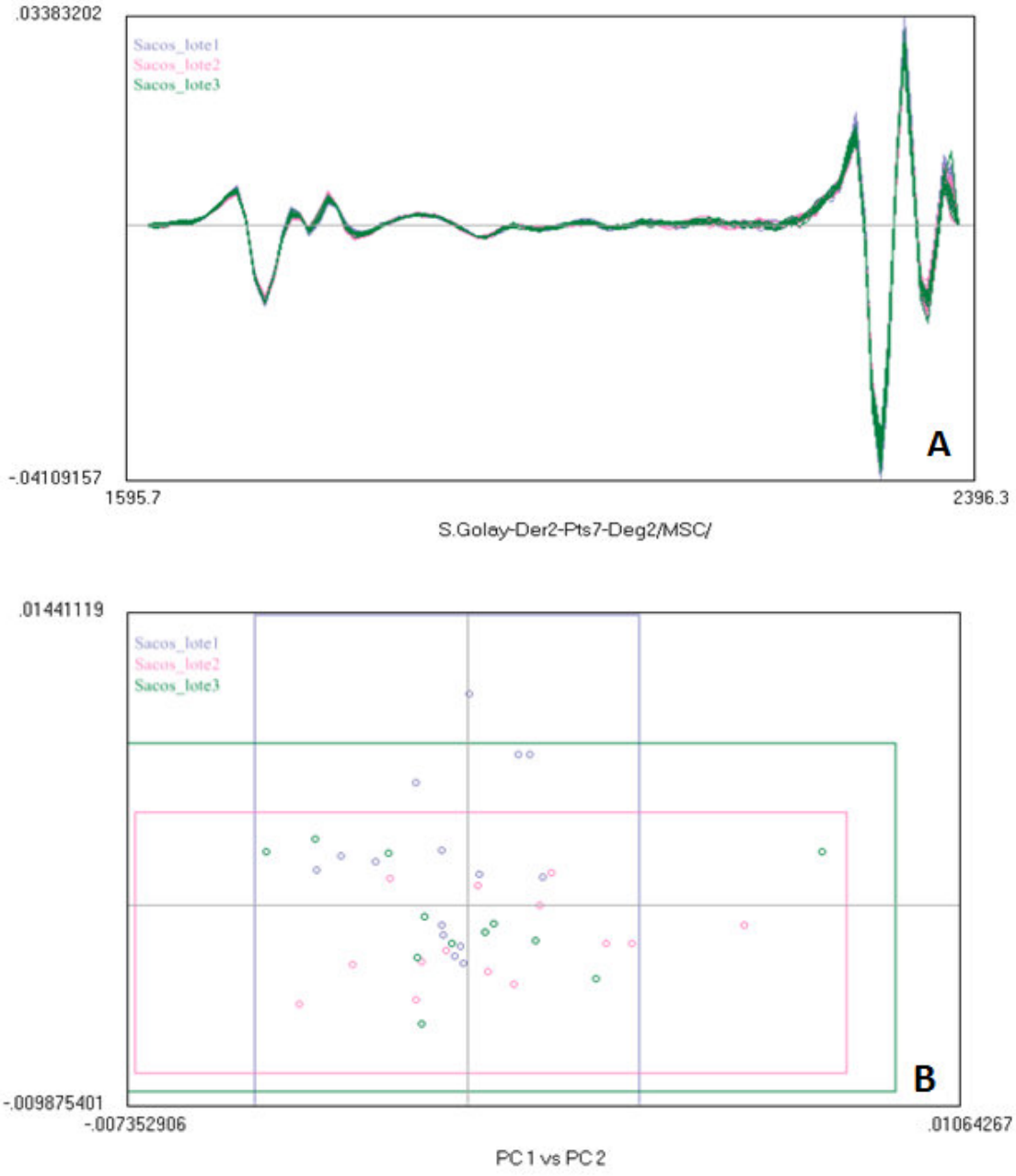


Figure 19 – **A:** S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) followed by MSC NIR reflectance spectrum of measurements in evaluation of inter-batch bags. In this plot, X and Y axes represent the wavelength and absorbance, respectively; **B:** PCA score plot analysis of measurements performed inter-batch bags, pre-treated spectrum (S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC)

The variability was reduced, as seen on **Figure 19 A**, as well as the PCA analysis, in which the variance is lower when comparing to the raw data.

The inter-bag batch variance is not significant, since looking at the PCA data, even when using different types of bags, it is not possible to differentiate them in the obtained data.

To further evaluate this, the same sample spectrum was acquired in the three different types of bags.

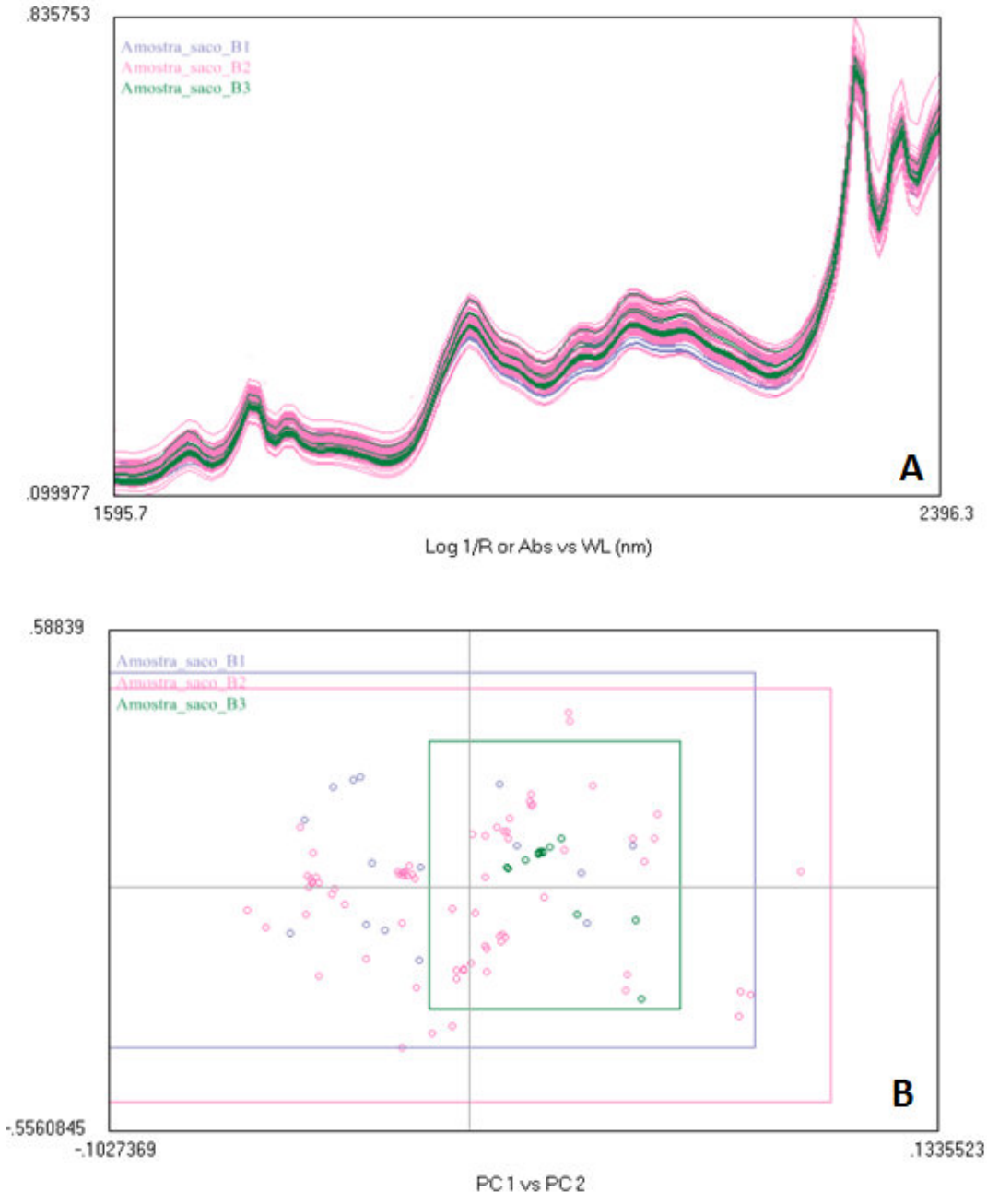


Figure 20 – **A:** Evaluation of inter-batch bags interference in sample spectrum acquisition, no pre-treated spectrum; **B:** PCA score plot analysis of measurements performed inter-batch bags, no pre-treated spectrum

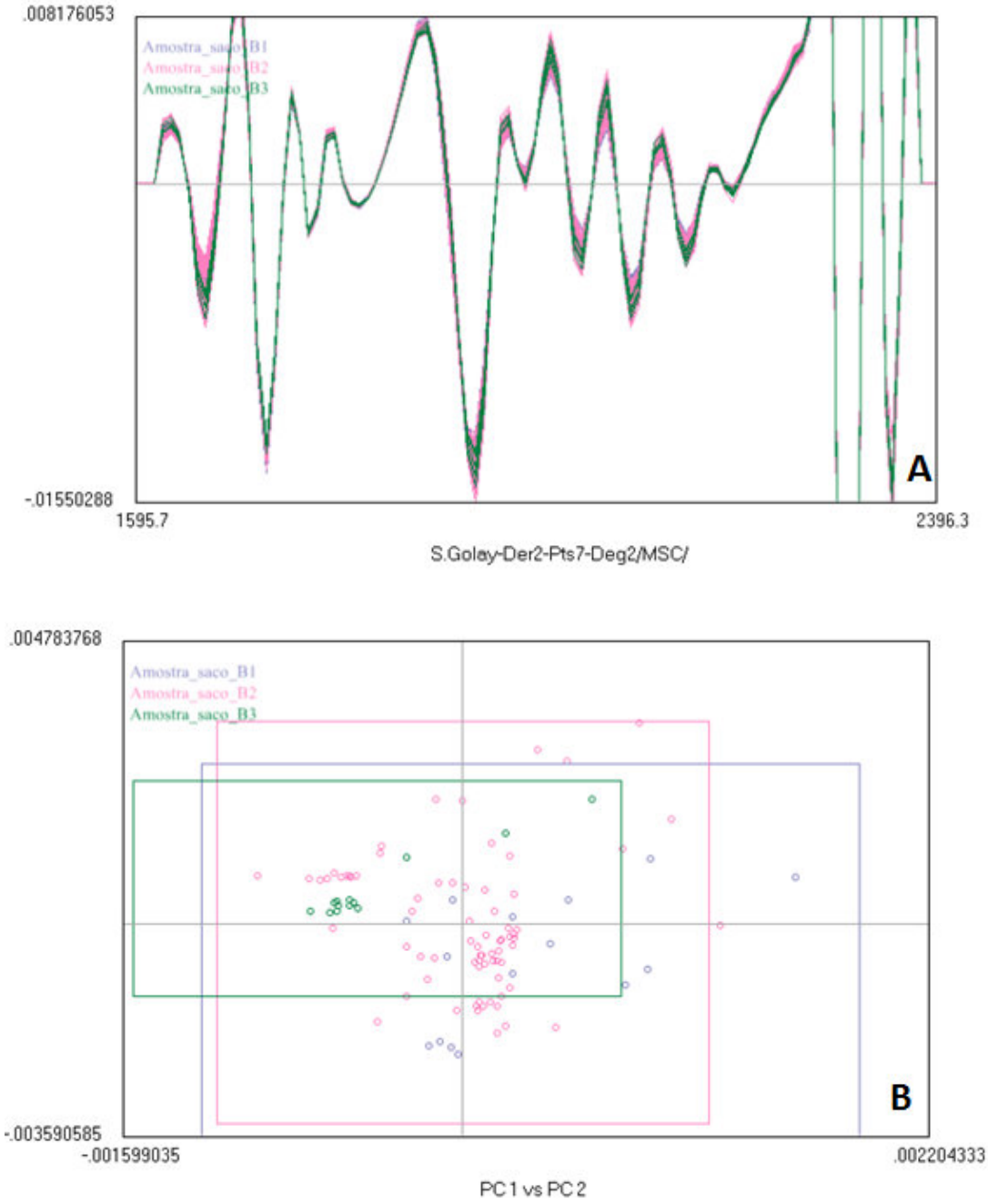


Figure 21 – **A:** Evaluation of inter-batch bags interference in sample spectrum acquisition, pre-treated spectrum (S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC)); **B:** PCA score plot analysis of measurements performed inter-batch bags, pre-treated spectrum (S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC))

Despite that with the PCA analysis of the raw data allowed a separation in clusters for each of the components, seen in **Figure 20 B**, by applying pre-treatments (**Figure 21 A, B**) a PCA analysis with lower variance was obtained.

In short, according to the PCA analysis, it can be said that the bags are not a critical factor in this process.

Nonetheless, since all the samples spectra are acquired through the bags, it was necessary to find its absorption zone, so the respective spectra were analyzed to perform this assessment:

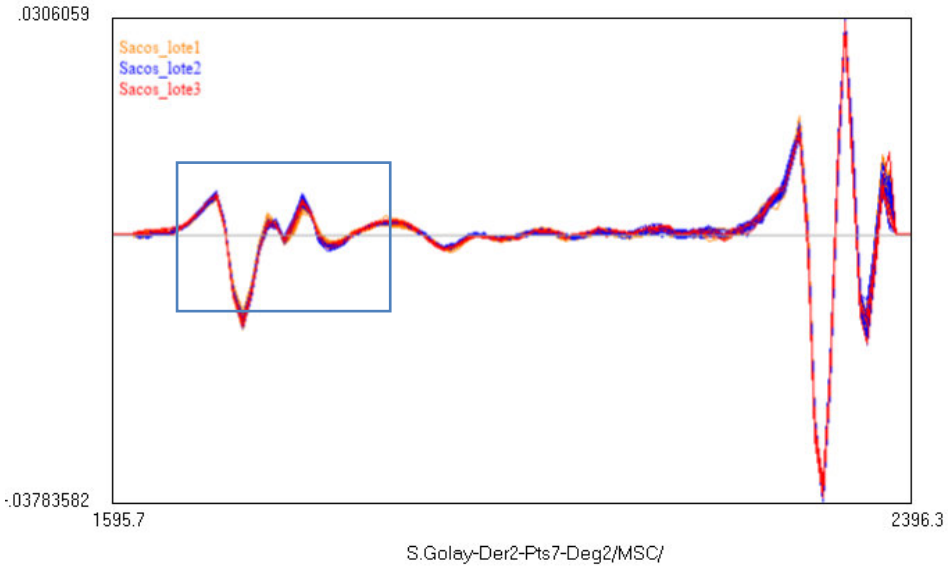


Figure 22 – Polyethylene bag used in the sample spectrum acquisition, pre-treated spectrum (S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC)

Looking at the spectrum, the evident polyethylene bags absorption zone is in the 1650 nm – 1920 nm range.

With the goal to evaluate the effect of sample amount on the acquisition of the spectrum several numbers of samples were tested. To evaluate their interference, a principal component analysis of the data was performed.

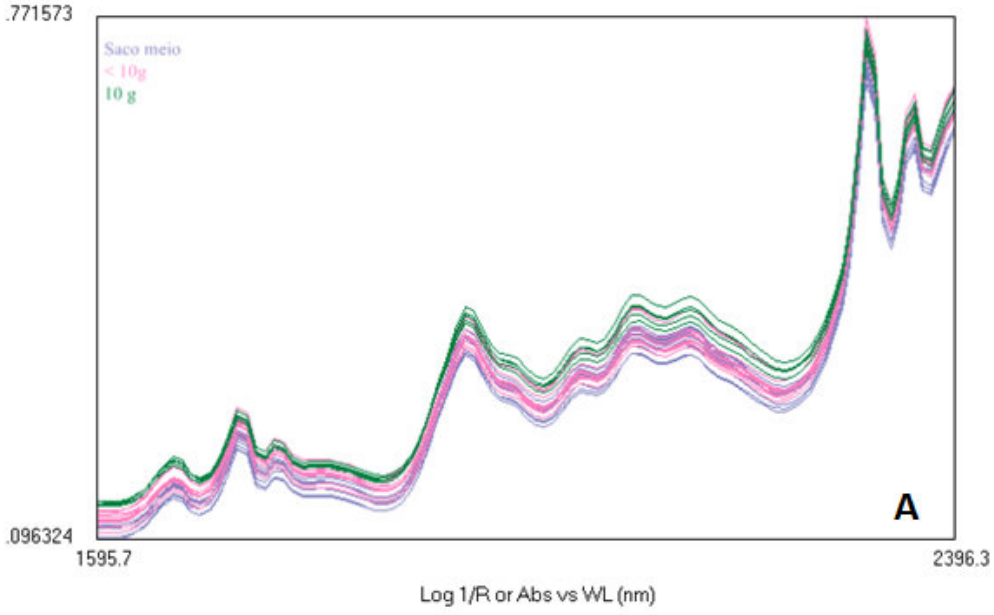


Figure 23 – **A:** Evaluation of sample amount in sample spectrum acquisition, no pre-treated spectrum;

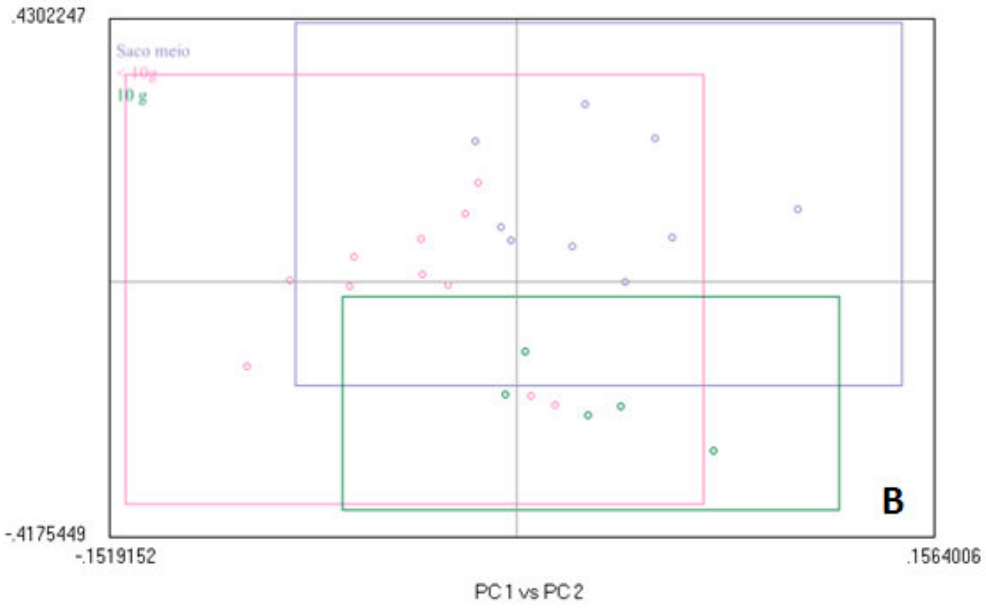


Figure 23 – **B:** PCA score plot analysis in evaluation of sample amount in sample spectrum acquisition, no pre-treated spectrum

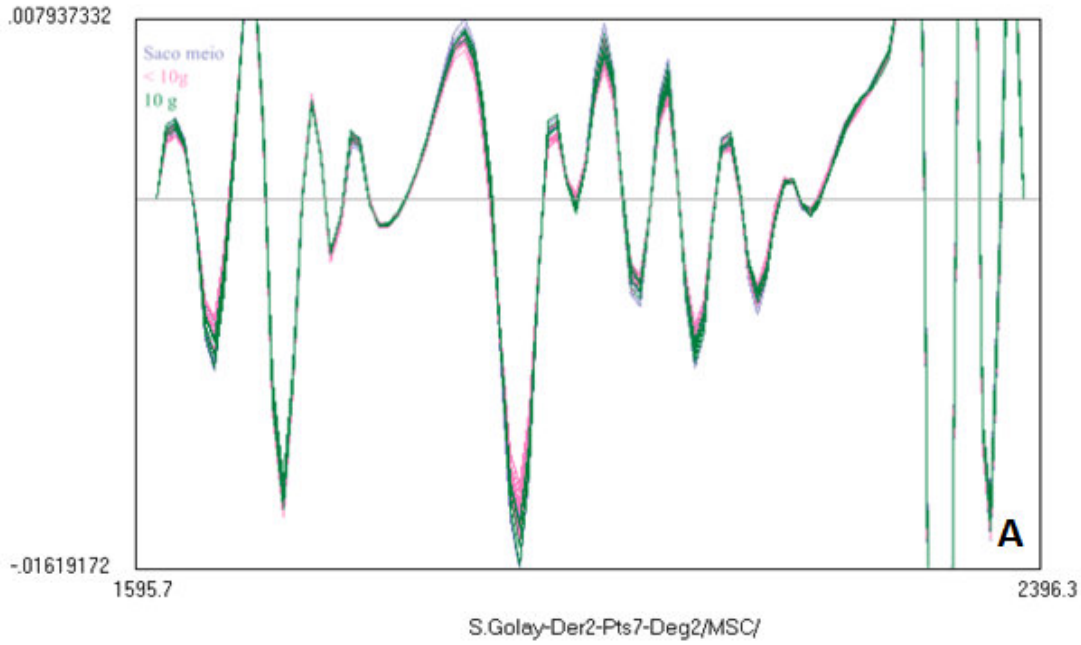


Figure 24 – **A:** Evaluation of sample amount in sample spectrum acquisition, pre-treated spectrum (S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC)

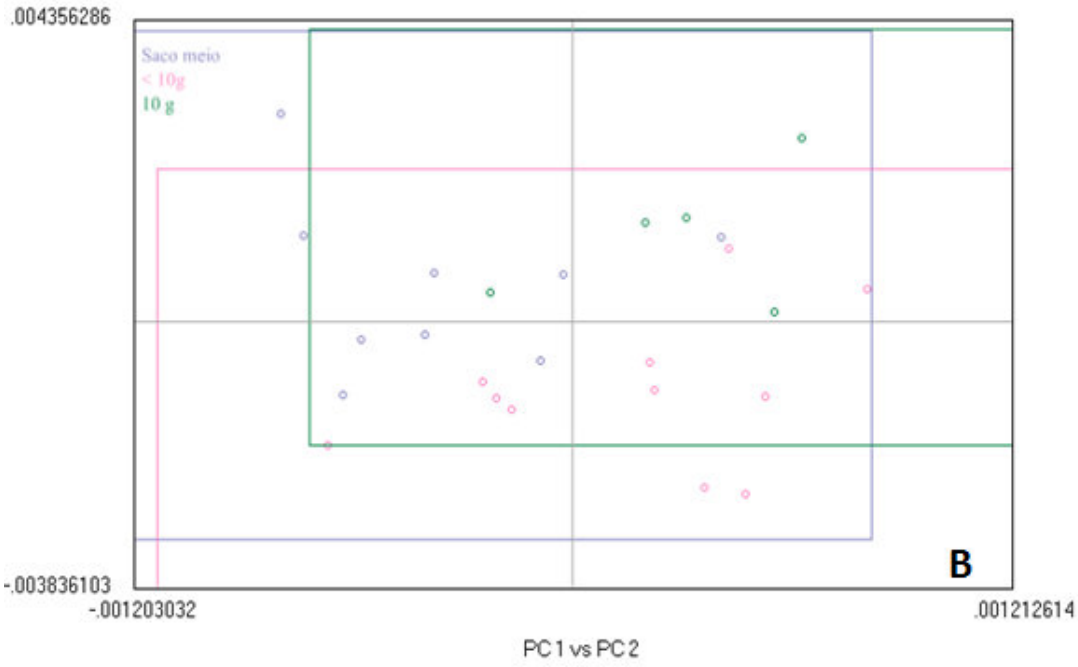


Figure 24 –B: PCA score plot analysis in evaluation of sample amount in sample spectrum acquisition, pre-treated spectrum (S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC)

Applying pre-treatments to the raw data (**Figure 24**) was crucial to obtain a lower variance, when comparing to the raw spectrum data (**Figure 23**).

Three sample amount were tested, acquiring 5 scans per each amount sample: enough amount to fill 50% of the bag volume ("Saco meio"), less than 10g amounts ("<10g"), and 10g amounts ("10g"). The acquisition of samples with less than 10g was not viable since this amount is not enough to cover the posterior side of the bag where the acquisition is performed.

With the PCA analysis (**Figure 24**), one can conclude that sample amount is not a critical factor, as long as it is enough to ensure that light dispersion does not occur while acquiring the spectrum. The minimum sample amount is 10g.

5.1.1.4. Equipment

All equipment involved in this work complies with the IQ, OQ and PQ tests.

5.1.1.5. Measurement

The reference method for its API content is a validated method according to the respective guidelines. Every determination was verified in order to assess the existence of any calculation or analyst error. The acceptance criteria (maximum inter sample $RSD_{\%}$ (relative standard deviation) of 5%) was always achieved.

5.1.1.6. Environment

Spectrum acquisition was performed in a laboratory with a 20ºC – 23ºC temperature variation. Furthermore, the development was performed over a 6 month period, which allowed different environmental conditions (humidity) to be tested and incorporated into the method.

5.1.2. Spectrum Investigation

To develop the quantitative model, the formulation's raw-materials spectrum had to be known. 10 random batches of each were selected and 10 scans were acquired per batch.

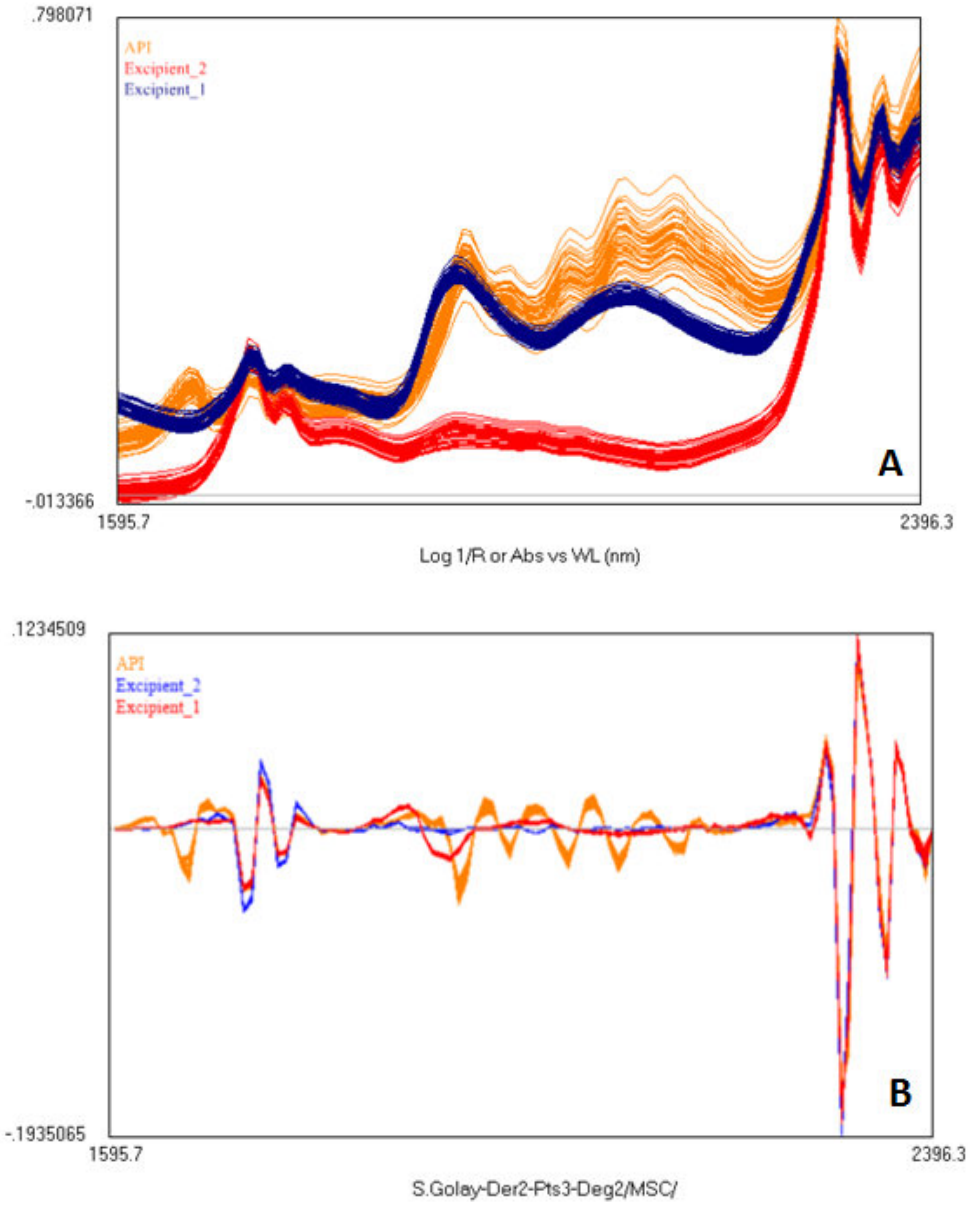


Figure 25 – **A:** Spectrum data of the formulation components, no pre-treated spectrum; **B:** S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) followed by MSC pre-processed NIR reflectance spectrum of measurements the formulation components. In this plot, X and Y axes represent the wavelength and absorbance, respectively.

In order to build the multivariate calibration models, different pre-processing methods were used in combination with different spectrum regions. The NIR spectrum for three samples with 80% (w/w), 100% (w/w) and 110% (w/w) in API are shown in **Figure 26**.

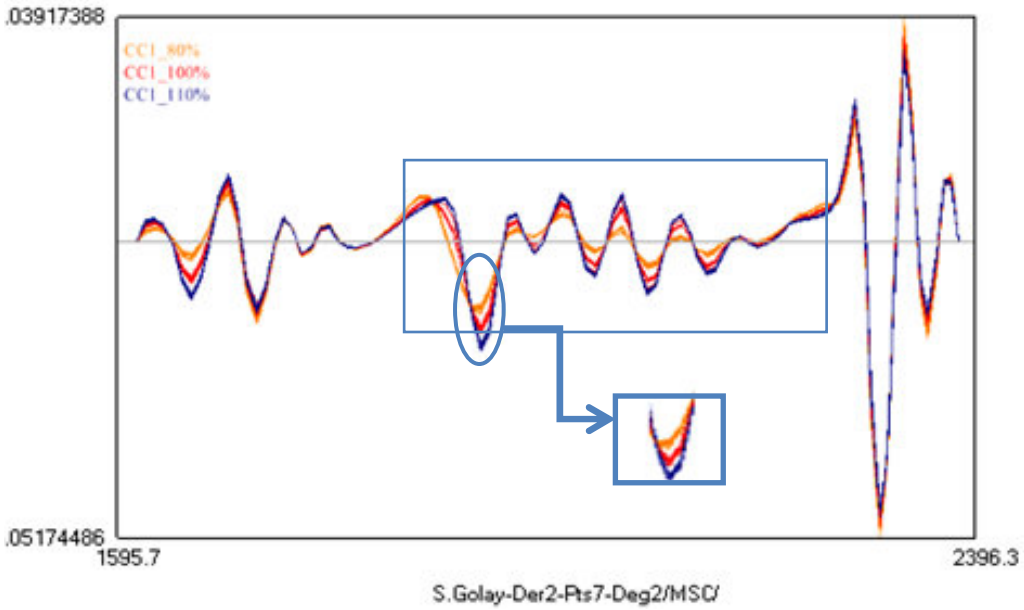


Figure 26 – NIR spectrum of formulation with three different content of the API, pre-treated spectrum (S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC). In this plot, X and Y axes represent the wavelength and absorbance, respectively.

The model was built with real and contaminated samples, followed by the application of the pre-processing of PCA analysis. This analysis helps to assess whether the sample group is divided into two distinct groups: contaminated samples and real samples, and helps identifying outliers that were not previously detected.

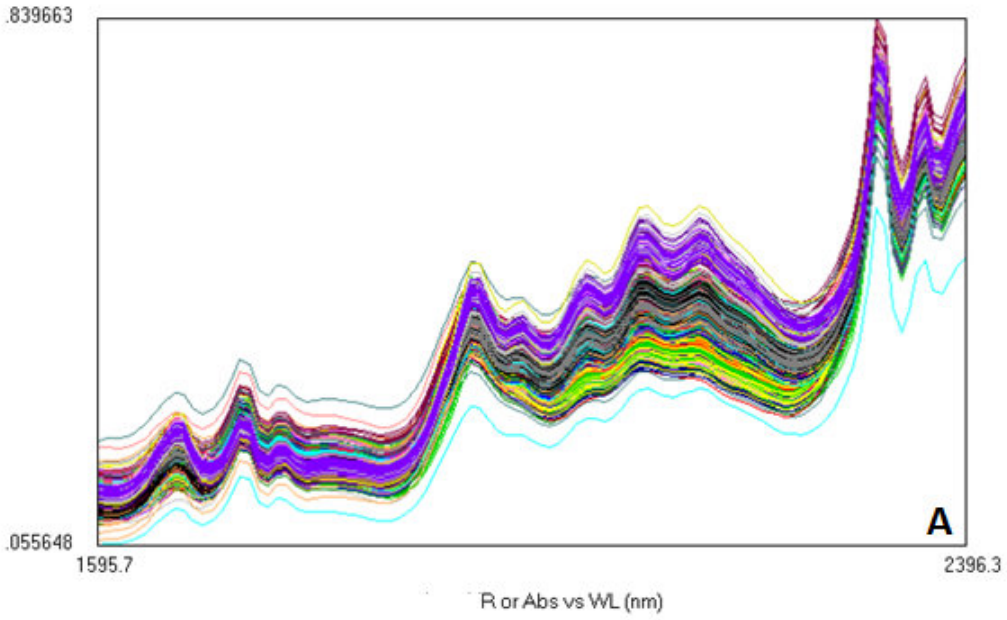


Figure 27 – **A:** No pre-processed NIR reflectance spectrum of measurements performed in all samples involved in quantification model. In this plot, X and Y axes represent the wavelength and absorbance, respectively;

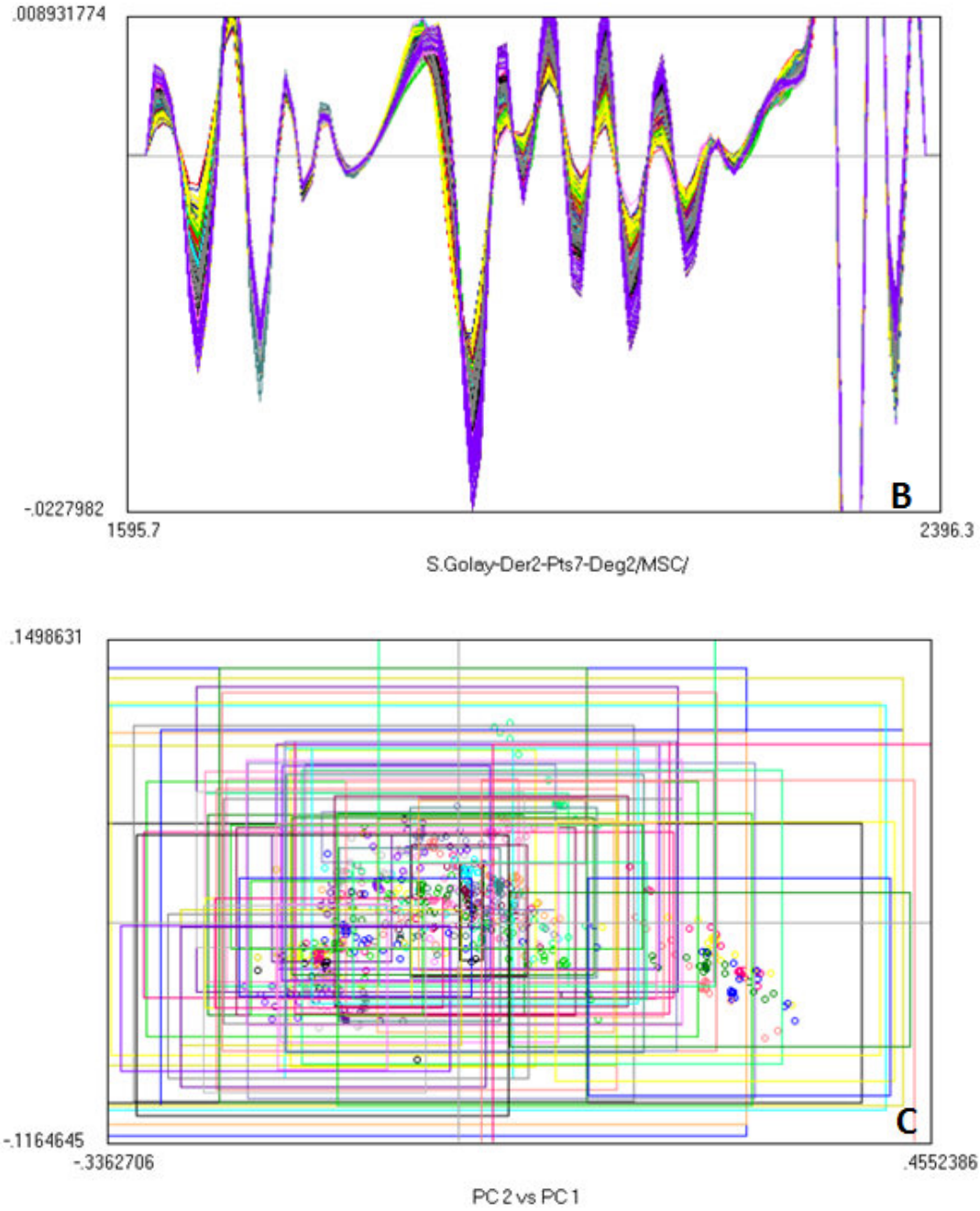


Figure 27 – **B:** S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) followed by MSC pre-processed NIR reflectance spectrum of measurements performed in all samples involved in quantification model. In this plot, X and Y axes represent the wavelength and absorbance, respectively; **C:** PCA score plot (PC2 x PC1) analysis of real and contaminated samples, S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC pre-treated spectrum

Looking at **Figure 27**, no visible differences allowing the separation of both sample groups are shown. The PCA analysis presented in **Figure 27C** does not reveal the grouping of the samples according to the type of sample; low variance.

As seen in **Figure 28**, the intense spectrum peaks of API are mainly seen in the region of 1830 - 2300 nm, presenting significant differences between absorption bands. Before the application of any pre-processing methods, the 1840 - 1940 nm spectrum range was selected for further analysis.

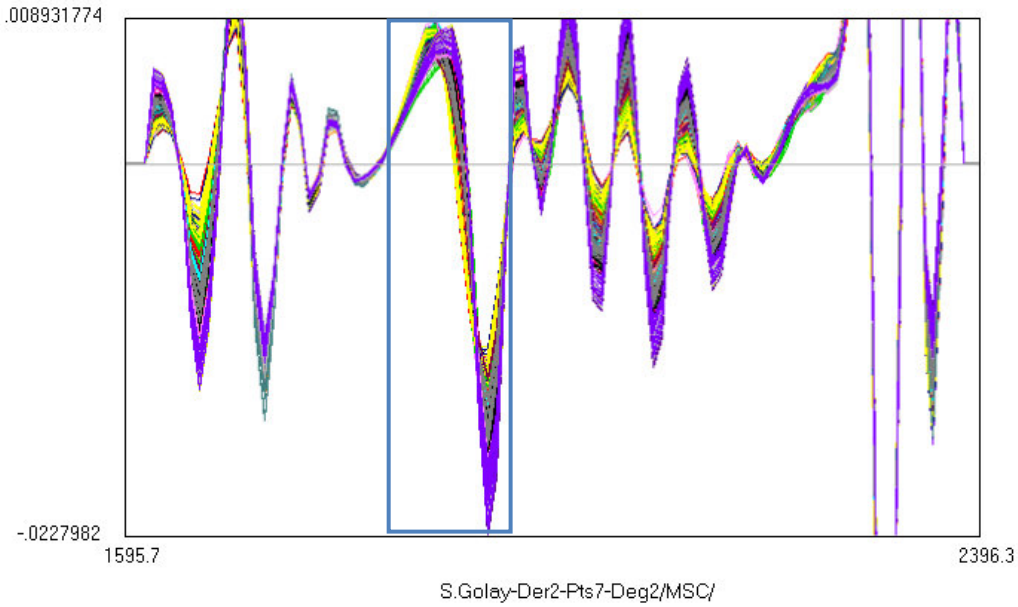


Figure 28 – Selection of the specific wavelength, S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) and MSC pre-treated spectrum. In this plot, X and Y axes represent the wavelength and absorbance, respectively.

As the goal of the final method is to monitor drug concentration, a specific drug wavelength was chosen instead of the whole spectrum. The band in the 1840 - 1940 nm range was selected because at this location a strong correlation between peak height and concentration exists.

5.1.3. Spectrum data pre-treatment

Several factors that can lead to changes in the NIR spectrum, can be eliminated or mitigated by the application of pre-treatments, which aim to remove systematic variations, correct the baseline, minimize noise (smooth), improve the definition of superposed peaks in the same region, indicating the parameters of interest thereby increasing the selectivity and removing irrelevant information and reducing the effects caused by differences in particle size. Despite these corrections, the choice of pre-treatment must ensure that it does not eliminate chemical and physical relevant information to the system under study.

The first step in model development was to investigate the most adequate pre-processing method. The most suitable set of pre-treatments applied to the model samples was based upon the choice of the number of factors and the calculation of RMSECV and RMSEC. A PLS regression was performed with the calibration set and cross-validation was carried out for model validation.

To evaluate the viability of the model including the value of R^2 and their errors is necessary to inform the model of *y*-values. The *y*-values correspond to the theoretical content determined by the reference method. Each of the samples used to build the model is matched to the *y*-value. The reference values for the samples were obtained by direct weighing of the representative sample of each mixture; an average API value for each batch was obtained by analyzing with reference method. The model predictive ability is an important factor to evaluate the model viability and is given by RMSEP. Other parameters are important in model evaluation. These evaluations are shown in **Table 5: Development of NIR model quantification / Parameters of NIR model quantification**.

5.1.4. Determination of the optimal Factors number

For PLS analysis, the spectra pre-treatment methods and the number of factors are critical parameters and both may be related. The number of factors to be used in the PLS model is very important because too few components will generate a poor model (with low predictive capability). Using too many, on the other hand, generates an overffited model, one which is also modelling noise in the sample calibration set, thus generating a low RMSEC (%, w/w) but performing poorly in the validation set.

The number of factors value is automatically chosen by the software (minimizing the cross-validation error) and should be critically examined by whoever is developing the model.

The software chooses, by default, 10 factors from which it will choose the ones considered sufficient to hold all the information needed to predict the quality parameters. In this model, the first 4 factors were enough to fulfil this requirement, as shown in **Figure 29**.

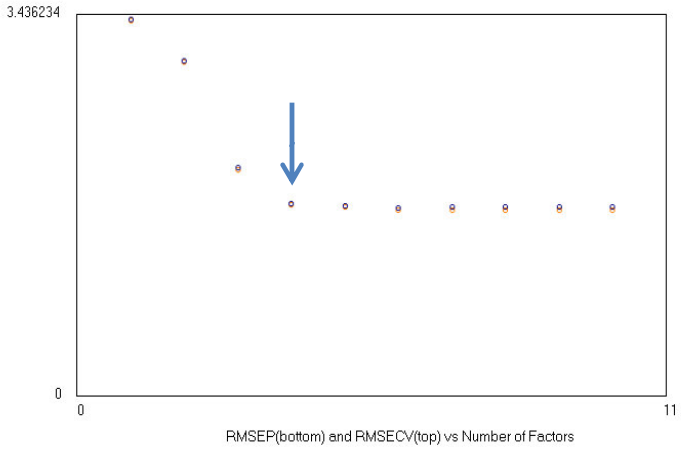


Figure 29 - Number of factors to be used in the PLS model, after performing the cross validation, the number of factors was assigned as 4;

Table 4 - Parameters of NIR model quantification (*software comparison*)

Parameters	PMG	SIMCA 13.0/MODDE 9.0
R^2_C:	0,957	0,960
R^2_{CV}:	0,957	0,962
RMSEC (%, w/w):	1,72	1,73
RMSECV (%, w/w):	1,73	1,75
R^2_V:	0,961	0,967
RMSEP (%, w/w):	1,67	1,69

In summary, **Figure 29** shows that factors 1-4 explain the majority of the observed spectrum and concentration variation. This plot demonstrates a decrease in the amount of error (RMSECV (%, w/w)) with successive used factors. This plot diagnostic is used to determine the optimal number of factors for the model. A minimum number of factors were used in the model to avoid over-fitting the data which would result in the model performing poorly with samples not used in the model development.

The number of PLS factors shows the evaluation of the RMSECV (%, w/w) value, the best number of PLS factors corresponding to RMSECV (%, w/w) minimum value. The loadings plots for components 1 – 4 obtained in the model are presented in **Figure 30.**

In order to obtain the percentage contribution of each of the 4 factors, software SIMCA 13.0/MODDE 9.0. was used. Calculation of the model associated errors was also performed with the same software to evaluate if both software's calculations provided significantly equal results (**Table 4**).

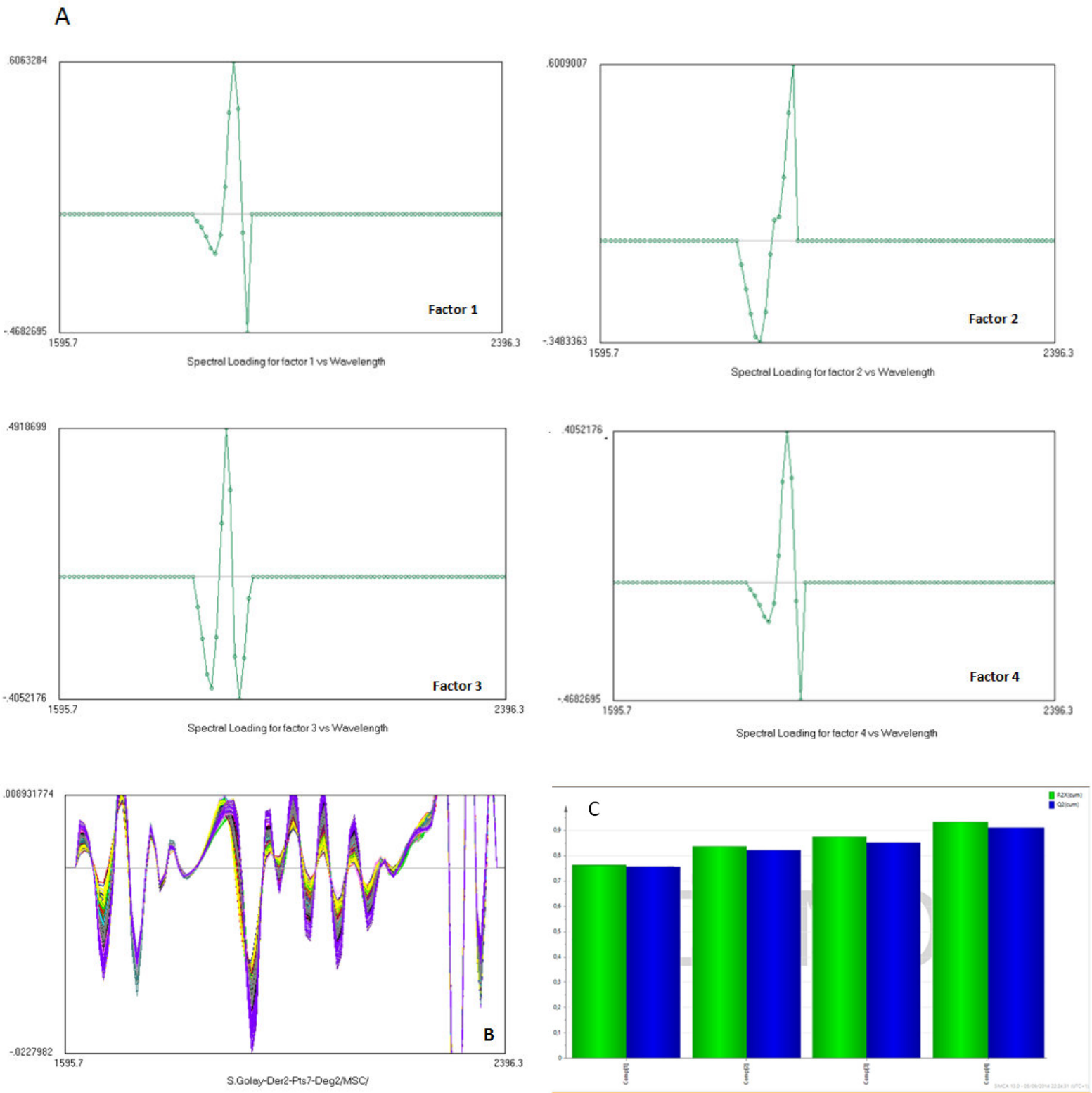


Figure 30 –A: Loadings plots for factors 1 – 4. In this plot, X and Y axes represent Wavelength and Spectrum Loadings, respectively; **B:** S.G. (derivative: 2; Smooth Pts: 7, Degree: 2) followed by MSC pre-processed NIR reflectance spectrum of measurements performed in all samples involved in quantification model. In this plot, X and Y axes represent the wavelength and absorbance, respectively; **C:** Contribution of same components *(Data obtained with software SIMCA 13.0/MODDE 9.0)*

In the loadings plots (**Figure 30 A**), the first component seems to have significant peaks between 1886 and 1921 nm. Signals in this range of the NIR spectrum might be water and in this case this is a clear possibility since the API is monohydrate. The second, third and fourth components show one peak in the area between 1895 – 1940 nm, which is also in the typical aromatic (ArOH), again, a specific characteristic of the API molecule. Looking at the loading plots analysis it is visible that the largest intensity peaks associated with each factor corresponds to the Chemical characteristics of the API that are not found on any of the other formulation component.

In the **Figure 30 C** shows the contribution of the factors and shows that the four factors were necessary to describe the model, which capture 92.9% of the data variance in the spectrum.

The Hotelling's T^2 Range (**Figure 31**) plot displays the distance from the origin in the model plane (score space) for each observation.

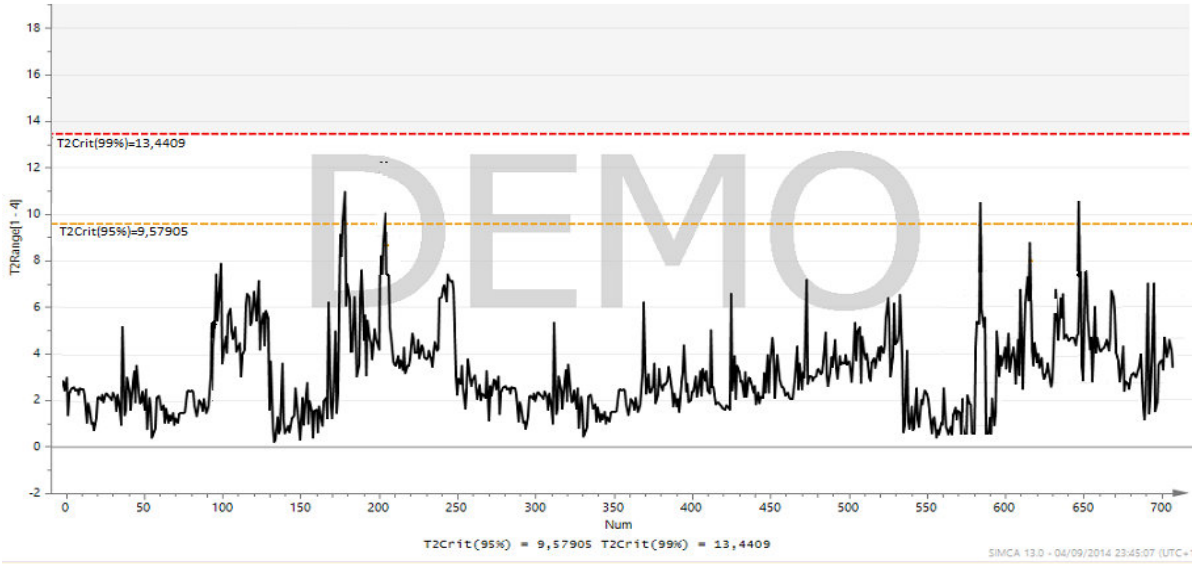


Figure 31 – Hotelling T^2 statistical process control chart. The horizontal lines are the 95% (orange line) and 99% (red line) control limits of the statistics. *(Data obtained with software SIMCA 13.0/MODDE 9.0)*

Data used in the calibration model are mainly found in the confidence interval of 95%. In **Figure 33**, one can verify that 4 of the 700 individual data are outside this limit, representing only 0.6% of the whole data.

5.1.5. Model development and validation

Cross validation, as well as validation according to ICH Guidelines was carried out for the calibration model. In NIR spectroscopy, spectrum pre-processing and spectrum regions used for building the calibration models are of crucial importance. Spectrum pre-processing is usually performed to remove unwanted scattering features incorporated in NIR spectrum which are often due to the differences in size of the constituent particles and other interfering factors that do not provide any information about the chemical concentration of the analyte of interest. The quantification model was built with a total of 70 batches; this is equivalent to approximately 700 spectra.

The software used for data processing and construction of the model allows the application of different types of pre-treatments: baseline offset, baseline liner, normalize max, normalize range, normalize peak, norm unit vector, normalize area, S.G. (derivative, smooth pts, degree), SNV and MSC. Among these methods, only two (S.G. (2, 7, 2) and MSC) were selected.

The effects of light scattering caused by differences in the size and shape of the particles can be corrected by applying the dataset multiplicative signal correction (MSC). The result of application of this pre-treatment in the data is represented in **Figure 28**. Other pre-treatments were tested but the best results were obtained with this selection.

Finally, the calibration set contained 50 samples (500 spectrum), and the remaining 20 samples (200 spectrum) constituted the prediction set. The determinations by reference method (*y*-values) in the calibration set almost covered the entire range in the prediction set. It included both production batches and the specially produced batches to give the largest variation in API content.

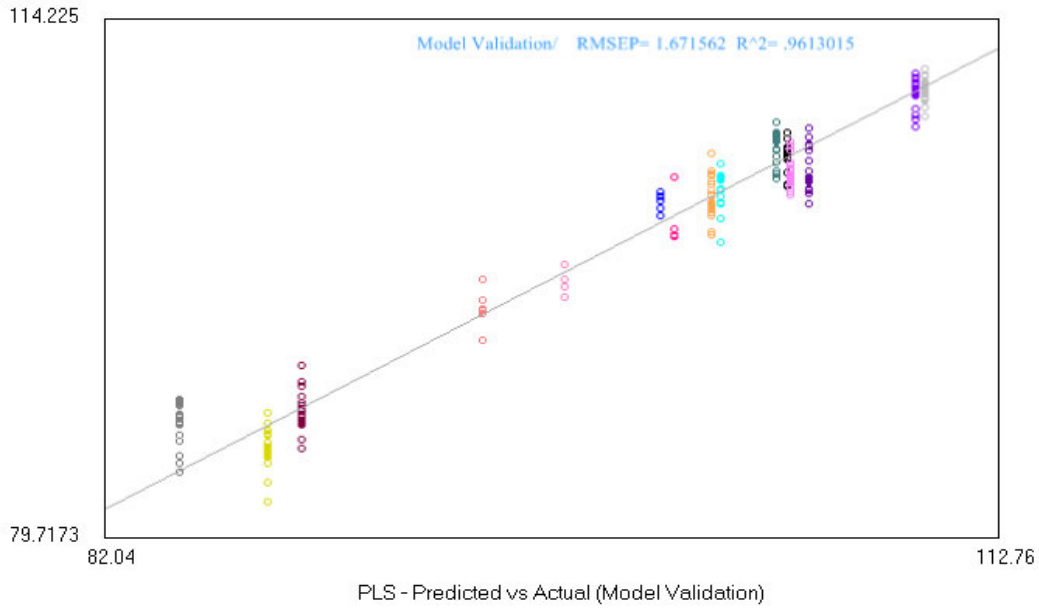


Figure 32 – Validation result (RMSEP (%, w/w) = 1.67; R^2= 0.961). In this plot, X and Y axes represent the Predicted Values and Actual Values, respectively.

The validation set (**Figure 32**) was used to evaluate the prediction capability of the PLS model constructed from the calibration set. Twenty batches that were not used in the calibration set, ranging between 85% and 110% of the stated amount, were used in this set.

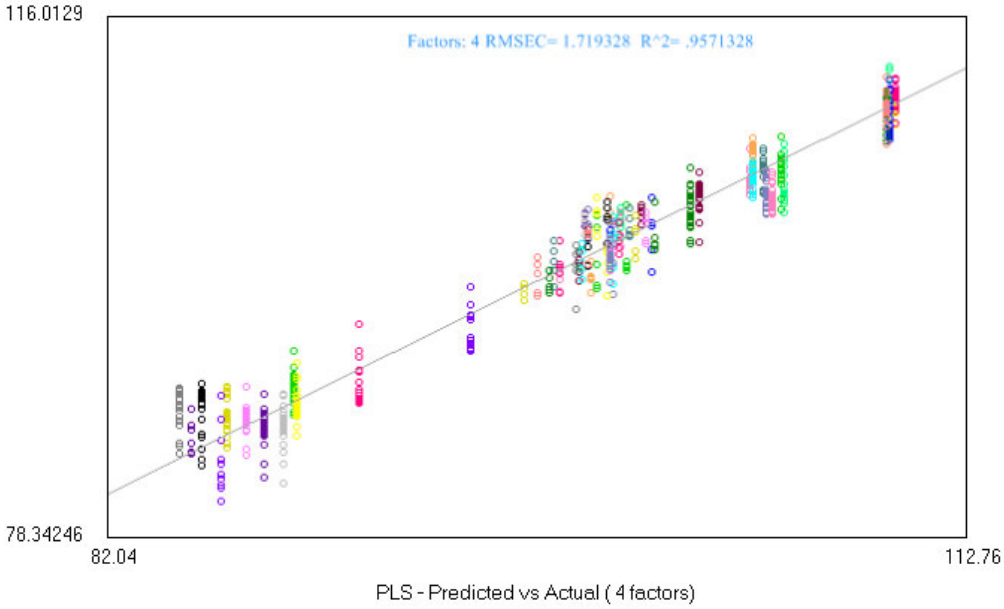


Figure 33 – Quantification Model (RMSEC (%, w/w) = 1.72; R^2= 0.957). In this plot, X and Y axes represent the Predicted Values and Actual Values, respectively.

Figure 33 represents the quantification model containing the 50 samples, equivalent to approximately 500 spectra.

5.1.6. Model Selection

The selection of the final obtained model was based on the evaluation of several models. The obtained results for the best three models are summarized in the table below. The chosen model, and the one discussed in this work is model 1.

Every model described in **Table 5** was obtained according to the description in the Materials and Methods chapter. Models 1 to 3 correspond to the ones developed to the quantification, and the best on was chosen by a comparison of the models described in **Table 5**:

Table 5 - Development of NIR model quantification / Parameters of NIR model quantification		
Model	**Results**	**Selected Parameters**
MODEL 1	Spectrum range selected (nm):	1840 – 1940
	Spectrum pre-treatment:	S.G. (Derivative: 2; Smooth Pts: 7; Degree: 2) + MSC
	Number of PLS factors:	4
	Development (n=50 batches)	
	R^2_C:	0.,957
	R^2_{CV}:	0.957
	RMSEC (%, w/w):	1.72
	RMSECV (%, w/w):	1.73
	Validation (n=20 batches)	
	R^2_V:	0.961
	RMSEP (%, w/w):	1.67
MODEL 2	Spectrum range selected (nm):	1830– 1940
	Spectrum pre-treatment:	S.G. (Derivative: 2; Smooth Pts: 9; Degree: 2) + SNV
	Number of PLS factors:	5
	Development (n=50 batches)	
	R^2_C:	0.951
	R^2_{CV}:	0.950
	RMSEC (%, w/w):	1.82
	RMSECV (%, w/w):	1.83
	Validation (n=20 batches)	
	R^2_V:	0.954
	RMSEP (%, w/w):	1.84
MODEL 3	Spectrum range selected (nm):	1840 – 2280
	Spectrum pre-treatment:	S.G. (Derivative: 2; Smooth Pts: 7; Degree: 2) + MSC
	Number of PLS factors:	8
	Development (n=50 batches)	
	R^2_C:	0.952
	R^2_{CV}:	0.953
	RMSEC (%, w/w):	1.78
	RMSECV (%, w/w):	1.79
	Validation (n=20 batches)	
	R^2_V:	0.962
	RMSEP (%, w/w):	1.69

R^2_C: R^2 of calibration; R^2_{CV}: R^2 of cross-validation; R^2_V: R^2 of validation

Model 1 was selected due to its biggest predictive potential, best factor number, its RMSECV (%, w/w), RMSEP (%, w/w) and R^2. The RMSECV (%, w/w) was used as the main model accuracy indicator. These parameters need to be evaluated to test the quality of the model.

This project had the goal to substitute a validated reference method (titration method) by a fast response NIR quantitative model. The model development was made incorporating the titration method results which has a maximum $RSD_{\%}$ value of 5.0% between samples.

According to the literature review, equivalent pharmaceutical industry applied models showed significantly low error values (< 1), nonetheless, the majority of these works was based on HPLC or ultra violet-visible methods, which have lower technique associated errors when comparing to a titration technique. The obtained R^2 value is higher than 0.95. Considering the above facts related to the intrinsically reference method errors, one can conclude that the error obtained for this model was acceptable.

Even though, this error can be improved in the future with the enrichment of the PLS model by adding new samples.

5.2. QUANTITATIVE MODEL VALIDATION

The NIR quantification model was validated according to the validation guidelines of the International Conference on Harmonization (ICH) and the European Medicinal Agency products (EMA), assessing the parameters described in the **Table 6**:

Table 6 - Validation Requirements for NIR Method Quantification

ICH Q2A Validation Parameter	Test
Specificity	+
Linearity	+
Accuracy	+
Precision	
Repeatability	+
Intermediate Precision	+
Robustness	+
Detection Limit	-
Quantification Limit	-

- Not normally evaluated; + normally evaluated

5.2.1. Specificity

Specificity of the model to quantify the API was demonstrated by accuracy and robustness performed during the validation of the method. These parameters are indicative that the selected range to the model has a high influence on the predictive values and including maximum spectrum difference between high and low content for the API analyte.

5.2.2. Linearity

The linearity over the 80% (w/w) - 110% (w/w) API content range with a total of seven concentrations (according to ICH Guideline Q2 (R1) recommending a minimum of five concentrations) was confirmed based on the R^2 value of regression lines obtained from the NIR model predicted values plots and reference values at the calibration and test set validation stages. This test was supported with the R^2 and analysis of residuals.

The **Table 7** shows the obtained results for the tested samples for linearity by two methods.

Table 7 – Linearity (experimental results)

Concentration Level (%)	% RM	$RSD_{\%\ (RM)}$	%NIR	$RSD_{\%\ (NIR)}$	Bias (%)
80	80.1	0.3	80.2	1.3	-0.145
85	85.7	1.3	85.9	0.8	-0.207
90	91.6	2.5	91.4	0.3	0.215
95	95.3	1.7	95.1	0.5	0.206
100	100.8	0.3	100.8	0.3	-0.026
105	105.1	1.8	105.0	0.5	0.169
110	110.9	1.0	111.0	0.4	-0.137

RM: Reference method

As seen in the linearity profile and in **Table 7**, the bias has values between −0.026 and 0.215 in the range 80% (w/w) - 110% (w/w).

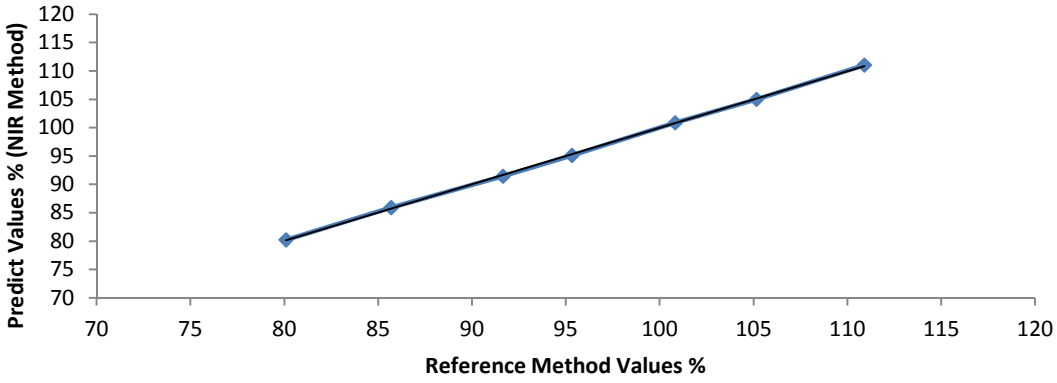
The bias can be expressed as:

$$Bias = \frac{\sum_{i=1}^{n}(y_i - Y_i)}{n}$$ **Equation 13**

Where:

Y: NIRS predicted value
y: Reference method value
n: number of samples

With the obtained results (**Table 7**), a calibration line graph (**Graph 3**) was constructed with the data from the two methods and their statistical analysis (**Table 8**):



Graph 3 – Linearity (Reference method *vs* NIR method)

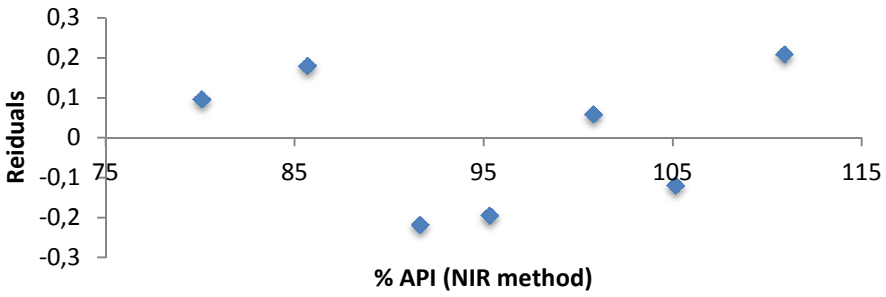
Table 8 –Linearity Results – Statistical Test			
	Coefficients	**Inferior Limit 95.0%**	**Superior Limit 95.0%**
Intercept (b)	0.3668	-1.4394	2.1730
Var X1 (a)	0.9960	0.9772	1.0148

To evaluate both methods similarity with the linearity test is important to ensure that:

- Intercept interval includes ‘’zero’’
- Slope confidence interval includes ‘’one’’

Looking at data summarized in **Table 8**, we can conclude that both methods are equivalent, and there are no differences between each, since the linearity test meets the requirements: *b*= [-1.4394; 2.1730] and *a*= [0.9772; 1.0148].

The analysis of the residuals in the linearity test is an important factor. By residual analysis plot (**Graph 4**) is possible to detect problems in curve fitting, e.g., deviations from linearity, presence of atypical samples and dependence among the errors.



Graph 4 – Residuals Analysis of the Linearity test

The **Graph 4** indicates that the errors scatter is randomly around the zero line indicating that the model is suitable.

5.2.3. Accuracy

Accuracy was determined according to the ICH specifications, which required three samples, one at each of the three concentration levels according to the model range (80%, 100% and 110% (w/w)), also being important to evaluate the accuracy in the range of specifications of the product in real applications (95% (w/w), 100% (w/w) and 105% (w/w)). The samples on each range were prepared with three different batch of API. In summary, the accuracy was evaluated in five different concentrations.

Each sample was scanned and five individual spectra were obtained per sample and the API concentration was predicted using the calibration model.

Table 9 – Accuracy Results

Concentration Level (%)	Accuracy_1			Accuracy _2			Accuracy _3		
	% RM	%NIR	%Recovery	% RM	%NIR	%Recovery	% RM	%NIR	%Recovery
80	80.1	81.8	102.2	79.9	80.0	100.2	80.6	82.6	102.4
95	95.3	95.0	99.7	95.3	96.7	101.5	96.1	96.2	100.1
100	100.8	100.5	99.7	99.4	100.4	101.0	100.2	99.0	98.8
105	105.1	105.7	100.6	104.6	106.3	101.6	105.0	105.6	100.5
110	110.9	111.0	100.1	108.8	110.6	101.6	110.9	111.8	100.8

Accuracy was evaluated in terms of percentage recovery compared to the reference method and the results obtained by the quantitative NIR validated method were compared with the values of the reference method. To compare the two methods, a percentage recovery between 98% and 102% goal was determined. The results for this test are summarized in **Table 9**.

The results of the calculation of the recovery percentage are calculated by **Equation 14**:

$$\%\ Recovery: \frac{API_{result\ by\ NIR\ method}}{API_{result\ by\ Reference\ Method}} \times 100 \qquad \textbf{Equation 14}$$

Accuracy was also estimated based on bias values obtained to each individual's concentration. The bias that is the mean difference between reference and NIR method data was considered.

The results are summarized in **Table 10**:

Table 10 –NIR Results for Accuracy (Bias)

Concentration Level (%)	Bias
80	- 0.1162
95	- 0.0274
100	0.0106
105	- 0.0647
110	- 0.0618

As seen in the accuracy profile and in **Table 10**, the bias has values between −0.0618 and 0.0106 for every concentration levels.

To the accuracy result in the reference method validation, within the same assay, a recovery percentage of 100.6% and a 0.34% RSD was obtained. Regarding the NIR developed method the obtained values were of 100.7% recovery percentage and 0.99 %RSD.

5.2.4. Precision

Precision was evaluated at two levels: Repeatability (intra-assay precision) and Intermediate Precision (repeatability over different days by a different operator). The intermediate precision as well as the repeatability, calculated through the $RSD_{\%}$, are not larger the 3%. This limit was defined considering the variability of the reference method.

5.2.4.1. Repeatability

Repeatability is an indicator of precision under the same conditions over a short interval of time; this test was assessed according to the ICH guidelines. For this test, six readings of a single sample at 100% target concentration of three different real batches were determined with the calibration model.

The $RSD_{\%}$ value of the six samples was determined to evaluate the method precision, defining a maximum $RSD_{\%}$ between samples of at most 3%. The results for this test are summarized in **Table 11**.

Table 11 – Repeatability Results

	Repeatability _1	Repeatability _2	Repeatability _3
Scan	**%NIR**	**%NIR**	**%NIR**
1	99.71	97.4	99.86
2	101.95	98.5	99.72
3	98.68	97.9	99.55
4	98.49	99.6	100.00
5	99.04	98.8	99.87
6	99.66	97.6	99.52
Average $_6$ (%):	99.59	98.31	99.76
SD_6:	1.26	0.83	0.19
RSD_6 (%):	1.27	0.84	0.19

The $RSD_{\%}$ values obtained in the analysis of the three batches complied with the criteria regarding the overall $RSD_{\%}$, indicating that the model was adequate in terms of repeatability of predictions by the applied model.

5.2.4.2. Intermediate Precision

Intermediate precision, like repeatability, is an indicator of precision, but this test evaluated different random events possible influence during the analysis. This test was performed by analyzing three different real batches on two different days and by two different analysts, each carrying out the analysis on a different day.

The results for the precision test (Repeatability and Intermediate Precision) carried out by two analysts on different days were compared with the results obtained for the reference method. The results are summarized in **Table 12**:

Table 12 - Comparison of results (NIR *vs* Reference Method)				
	Repeatability _1	**Repeatability _2**	**Repeatability _3**	
% NIR:	99.6	98.3	99.8	**Analyst 1**
% RM:	100.2	97.8	99.5	
Average (%):	**99.89**	**98.06**	**99.63**	
RSD (%):	**0.4**	**0.4**	**0.2**	
% NIR:	99.3	99.3	100.8	**Analyst 2**
% RM:	100,2	97.8	99.5	
Average (%):	**99.77**	**98.56**	**100.15**	
RSD (%):	**0.6**	**1.1**	**0.9**	

The $RSD_{\%}$ value of the twelve samples (six samples by each analyst) was determined and a value lower than 6% (highest acceptable value) was obtained, indicating that the model was adequate to predicted API concentration when submitted to different analysis conditions. The results for this test are summarized in **Table 12**.

The obtained results with the NIR method were compared with the reference values in terms of error (**Table 13**). For this calculation the NIR results were considered the average of the individual results obtained in Repeatability test by two analysts.

Table 13 - Error analysis – Comparison Methods			
Sample	**Results (%)**		
	RM	**NIR Method**	**Error**
Repeatability _1	100.2	99.5	0.74
Repeatability _2	97.8	98.8	-1.02
Repeatability _3	99.5	100.3	-0.78

RM: Reference Method

The error associated with different methods was determined by the following formula:

$$Error(\%\, w/w) = Result\ (\%\, w/w, API)_{\,ReferenceMethod} - Result\ (\%\, w\ /w, API)_{NIRmethod} \qquad \textbf{Equation 15}$$

The difference in results between the two methods is not significant, fulfilling defined $RSD_{\%}$ for repeatability. Both parameters had satisfactory values for all studied samples.

5.2.5. Robustness

Robustness is a measure of the method's capacity to remain unaffected when submitted a small but deliberate variations.

The following parameters were studied:

5.2.5.1. Different operators;
5.2.5.2. Different temperatures;

5.2.5.1. Different Operators:

This parameter was evaluated in Precision (5.2.4).

5.2.5.2. Different Temperatures

To evaluate the temperature influence two tests were carried out:

- Variation of the sample temperature (40ºC during 24 hours)
- Variation of room temperature (32ºC – spectrum acquisition)

These tests were performed using three different real batches. The results are summarized in **Table 14**:

Table 14 - Robustness: Temperature Effect

	Sample temperature effect (40ºC during 24 hours)						Environment temperature (32ºC)					
	Batch 1		Batch 2		Batch 3		Batch 1		Batch 2		Batch 3	
	% NIR	% RM	% NIR	% RM	% NIR	% RM	% NIR	% RM	% NIR	% RM	% NIR	% RM
Average (%):	98.2	100.2	99.5	97.9	98.0	99.5	99.7	100.2	99.4	97.9	98.5	99.5
(yi - xi) (%):	2.0		1.6		1.5		0.5		1.5		1.0	

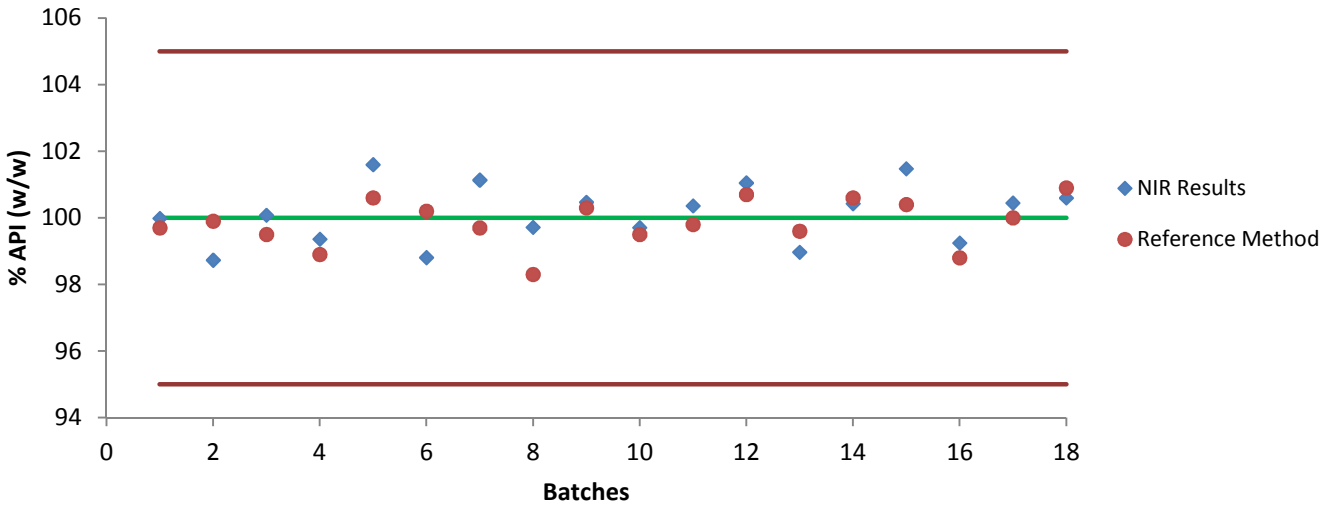
The results were compared with reference method; the error (yi - xi) was calculated by **Equation 17**:

$$Error(\%\, w/w) = \left|Result\ (\%\, w/w, API)_{\ ReferenceMethod} - Result\ (\%\, w\ /w, API)_{NIRmethod}\right| \qquad \textbf{Equation 17}$$

5.2.6. Real time applications

The NIR model development showed good results from the initial calibration and its validation confirmed the accuracy of the developed NIR calibration model to predict the concentration of the API in blends. However, the next goal was to apply the developed method for real time application in Quality Control Laboratory. Calculation of $RSD_{\%}$ is an easy way to compare NIR data from various blends with blend content data obtained with traditional method. An additional statistical test (Student's *t*-distribution) was performed to evaluate the results.

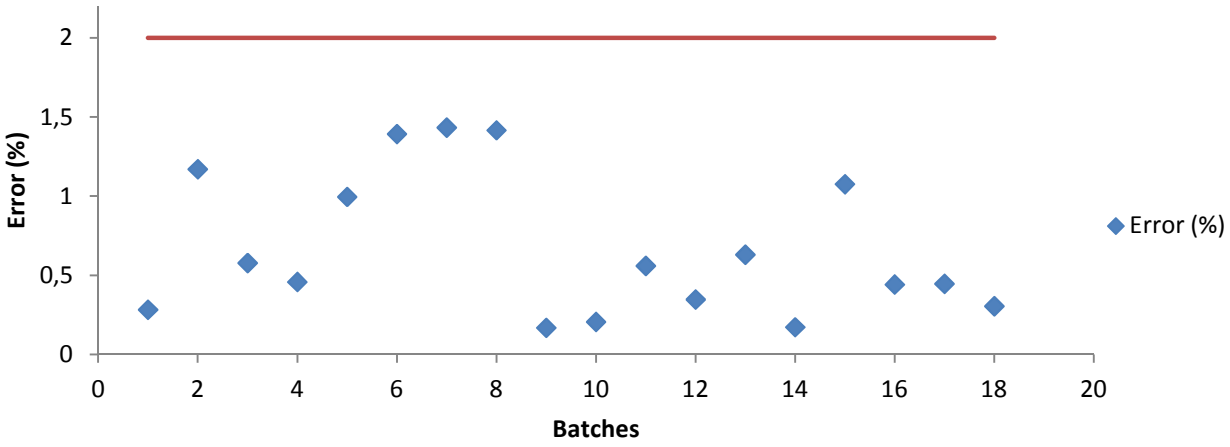
To compare the quantification model performance 18 batches were tested with both methods. The results were translated into a Control Chart (**Graph 5**).



Graph 5 – Chart Control 1 – Evaluation of results by NIR method and Reference method

Both methods originated results complying the specification criteria (95% - 105%). By a rough evaluation, a low difference between both methods for the same batch analysis is visible, and in general, all the results are close to 100% API (w/w).

To assess the NIR method performance, a maximum of ±2% between NIR and reference method difference was established. This was evaluated in the form of a control chart, calculating the error through the following **Equation 17**, visible on **Graph 6**.



Graph 6 – Chart Control 2 of error between both methods

The results are inside the established error limits, as expected, since the RMSEP (%, w/w) and other errors obtained during the method development were around 1.7%.

To compare both analytical methods, *t*-Distribution statistical test was applied to the obtained average results of each method.

Results evaluation is performed based on the determination of $t_{calculated}$ comparatively to t_{table}; $t_{calculated}$ is calculated through **Equation 18**:

$$t = \frac{\bar{d}\sqrt{n}}{S_d} \qquad \textbf{Equation 18}$$

Where:

$\bar{d}$: errors mean ($y_i - x_i$)
n: number of samples
S_d: errors standard deviation

Methods similarity evaluation is done with the obtained results for $t_{calculated}$ according to:

If $|t_{calculated}| > t_{table}$, systematic errors, not equivalent methods

If $|t_{calculated}| \leq t_{table}$, unsystematically errors, equivalent methods

In **Table 15**, the obtained results with 18 batches analysis are summarized:

Table 15 – Results by both methods

Batch	Method		$d_i = (y_i - x_i)$
	Reference Method (x_i)	NIR (y_i)	
1	99.7	100.0	0.28
2	99.9	98.7	-1.17
3	99.5	100.1	0.58
4	98.9	99.4	0.46
5	100.6	101.6	1.00
6	100.2	98.8	-1.39
7	99.7	101.1	1.43
8	98.3	99.7	1.42
9	100.3	100.5	0.17
10	99.5	99.7	0.21
11	99.8	100.4	0.56
12	100.7	101.0	0.35
13	99.6	99.0	-0.63
14	100.6	100.4	-0.17
15	100.4	101.5	1.08
16	98.8	99.2	0.44
17	100.0	100.4	0.45
18	100.9	100.6	-0.30
		Average:	0.263
		sd:	0.78
		n:	18
		$t_{calculated}$:	**1.43**
		t_{table}:	**2.11**

NOTE: $t_{calculated}$ was performed with a 95% confidence interval

Looking at **Table 15**, we can conclude that $|t_{calculated}| \leq t_{table}$, meaning that no significant difference exist between each method.

5.3. Method impact in Quality Control – OM Pharma

Associated with product's analysis comes time spent in Quality Control. Considering that the time spent with the reference method was 30 minutes and with the NIR method was 5 minutes (spectra acquisition and results evaluation), NIR method represents about 14% of the time per year needed to analyze all batches when compared with the reference method.

The impact of cost and associated to the new methodology requires an economic analysis more detailed that is not the scope of this work. However it should be noted that the gain in terms of time is directly related to costs reduced (consumables) associated with the reference method.
With this assessment, it is clear that the NIR quantification method implementation at Quality Control is profitable in terms of costs and time.

5.4. API - LIBRARY DEVELOPMENT

Different pre-treatments were applied in order to obtain the best separation between the different groups of raw materials. For library development, some pre-treatments were investigated, however, during this work only three of them were discussed because they were the ones showing the best overall result (**Table 16**).

Table 16 - Results obtained for different pre-treatment

Parameters	Model 1	Model 2	Model 3
	Step 1: S.G. (derivative: 1; smooth pts: 3; degree: 2) **Step 2:** Normalize Range	**Step 1:** S.G. (derivative: 2; smooth pts: 3; degree: 2) **Step 2:** Baseline Linear	**Step 1:** S.G. (derivative: 1; smooth pts: 3; degree: 2) **Step 2:** Baseline Linear
Spectrum Range (nm):	1670 -2300	1670 -2300	1670 -2250
Algorithm:	Spectrum-Match- KNN	Spectrum-Match- KNN	Spectrum-Match- KNN
Minimum Correlation Threshold assigned:	0.95	0.95	0.95
Maximum Valid IDs at threshold:	0.97	0.88	0.95
Total mismatches:	0	0	0
% Sample correctly identified:	100	100	100

NOTE: All presented methods have been developed for $k = 1$

From the table above, it is possible to conclude that the best model results comparatively to total mismatches, % of correctly identified samples and maximum valid IDs (threshold) were obtained using the Model 1 (**Figure 34**).

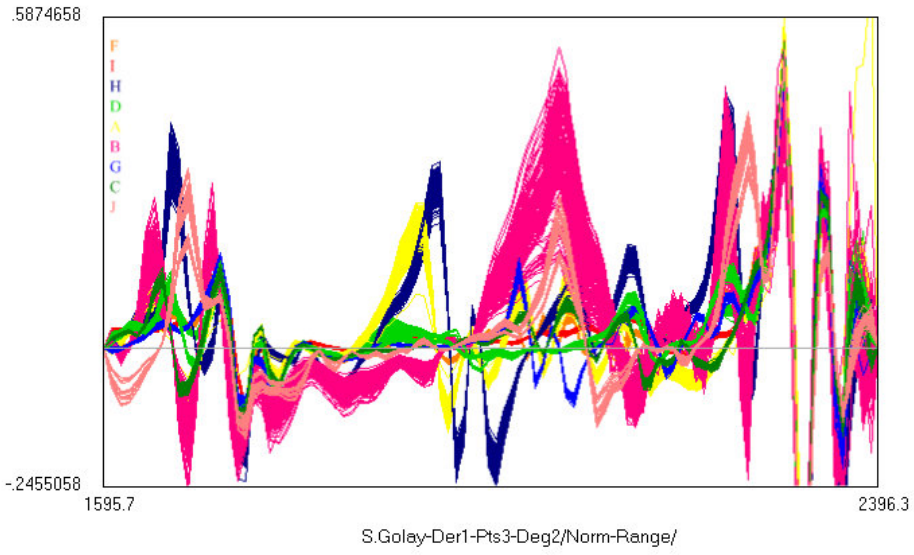


Figure 34 – Spectrum of pre-treatment 1 application (Model 1)

From the pre-treatments shown in **Table 16**, only Model 1 has an acceptable threshold for the method development, even though every pre-treatment identified all the raw material groups correctly, as seen in **Table 16** (Total mismatches).

In the figures below, the obtained results for each of the models are shown:

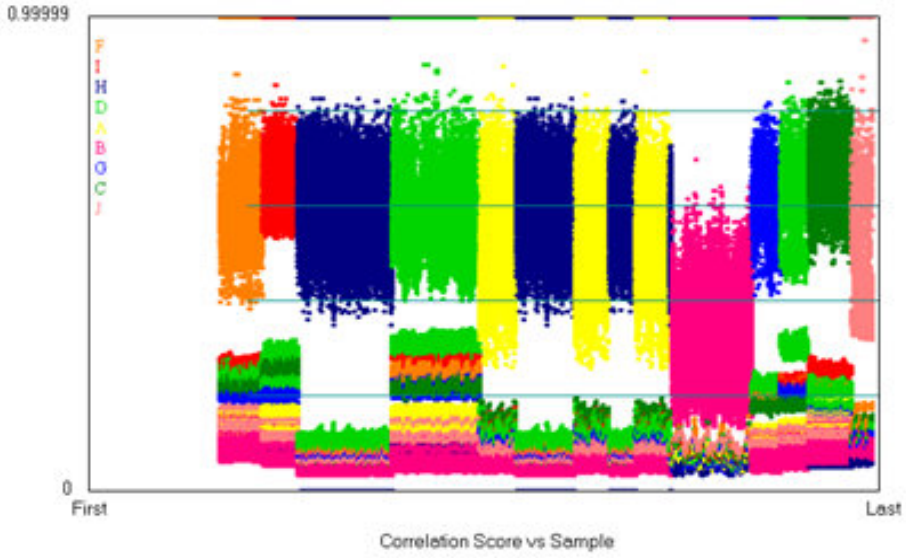


Figure 35– Pre-treatment 1 result (Model 1)

Looking at **Figure 35** results, the raw materials are clearly separated from each other. These graphs show (from top to bottom) the similarity order of the analysed materials. The coloured bars on the top are each of the raw materials and in the way down its closest material. The second colour always matches the first one for every material, attesting the method correct identification capability and selectivity.

In the second pre-treatment, the Golay polynom conditions were changed, as well as a spectrum base line adjustment (**Figure 36**).

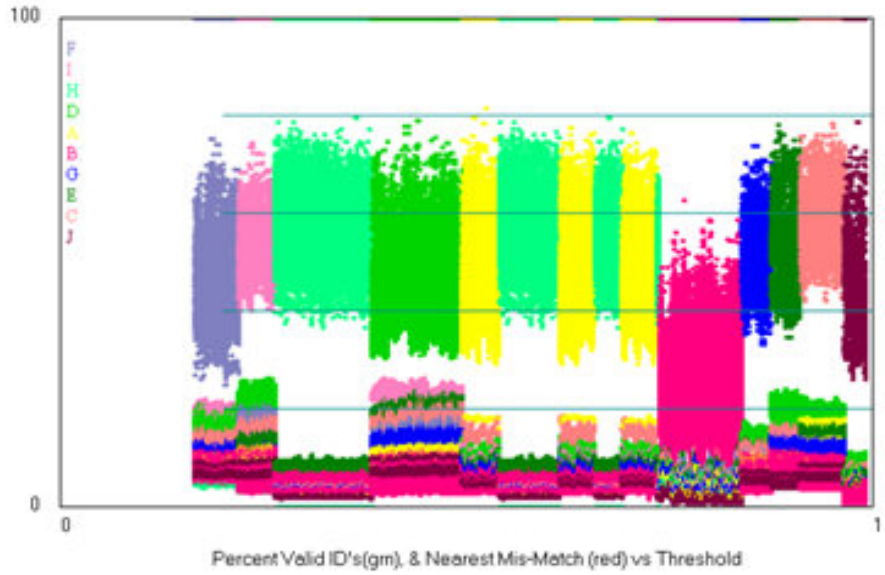


Figure 36 – Pre-treatment 2 result (Model 2)

The second and third methods arose with the need to better separate the raw-material B, as seen in the corresponding graph, where this material shows the lowest difference, comparing with the remaining raw materials.

In the Model 3 (**Figure 37**), the spectrum zone was reduced in order to eliminate as maximum noise as possible.

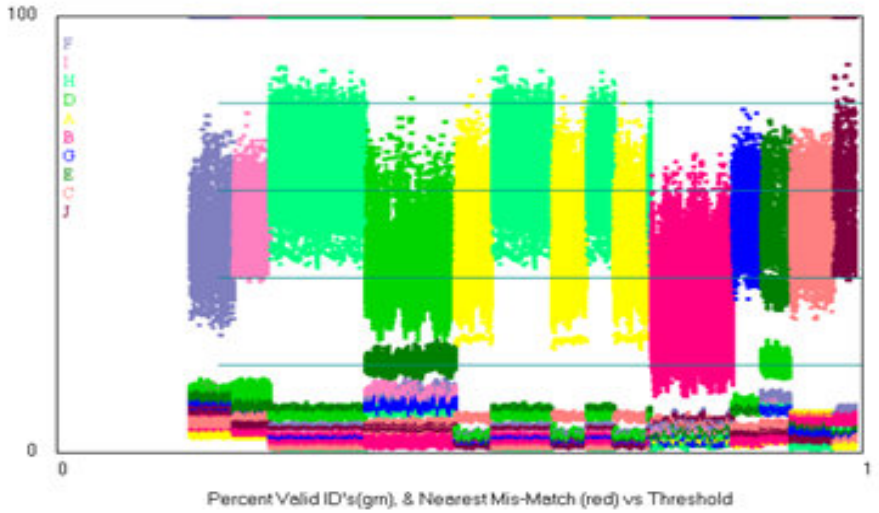


Figure 37 – Pre-treatment 3 result (Model 3)

In this improvement attempt, other identifications were deteriorated, as well as the threshold.
This raw-material has a significant variability due to its size particle and also its colour, influencing the discrepancy in its spectrum, thus, hampering the construction of a model that can totally separate this raw material from every other.

In this specific case, transmittance differences were visible when acquiring samples from the same batch, a fact that can be explained with three factors: the different applied forces when acquiring spectrum, the powder compression originating different free spaces between the particles and thus the light path in the matrix, and the size particle causing the same effect as powder compression.

Raw material B size particle was one of the critical factors used in the selection of its batches, since the selection criteria very wide. Nonetheless, the identification model, being validated according to the Guidelines, showed positive results for the identification of every raw material. Even though there is a need to further improve the model, it shows good development results such as the threshold.

5.5. LIBRARY VALIDATION

The performance characteristics for the validation of the NIR identification method according to the guidelines are described in the **Table 17**:

Table 17 - Validation Requirements for NIR Method Identification

ICH Q2A Validation Parameter	Test
Specificity	+
Linearity	-
Accuracy	-
Precision	
Repeatability	-
Intermediate Precision	-
Robustness	+
Detection Limit	-
Quantification Limit	-

- Not normally evaluated; + normally evaluated

5.5.1. Library validation

Library validation included an internal validation and an external validation step.

5.5.2. Internal Validation

The internal specificity refers to the ability of the spectrum library to properly identify the samples, from which the library was derived. An acceptable internal specificity would demonstrate that none of the spectrum used to build the library would be considered a statistical outlier.

As acceptance criteria (best spectrum match), the correlation coefficient (CC) values have to be higher than 0.95 (the correlation threshold defined in the method development).In addition, this value has to be the highest in order to have the correct identification of a raw material, when compared with the other library groups.

The following **Table 18** shows the best match identification group (*Top Match - TM*) for each batch of each raw material and the second and third (*2nd and 3rd Match, 2ND and 3RD respectively*) best match identification group as well as its correlation coefficient. With the obtained results in the internal validation it was possible to draw the map shown in **Figure 38**, which allowed subsequent internal auxiliary validation testing.

	A	B	C	D	E	F	G	H	I	J
A	TM		2ND	3RD					3RD	
B		TM	3RD			2ND				2ND
C			TM	3RD		3RD			2ND	
D				TM	2ND				3RD	
E				2ND	TM	3RD			3RD	
F				3RD		TM			2ND	
G				2ND	3RD		TM			
H				3RD	2ND			TM		
I				2ND		3RD			TM	
J		3RD	3RD			2ND				TM

Figure 38 – Map of mismatches obtained in Internal Validation

In **Table 18**, the obtained results for the Internal validation performed to 10 raw materials was summarized:

Table 18 - Internal Validation results

Raw Material	Identified as Top Match		Identified as 2nd Match		Identified as 3rd Match	
	Identification	CC	Identification	CC	Identification	CC
A	A	0.9999	C	0.7936	I D	0.6277 0.7664
B	B	0.9999	F J	0.5779 0.4243	C	0.4641
C	C	0.9999	I	0.8997	D F	0.8670 0.8672
D	D	0.9999	E	0.9531	I	0.9379
E	E	0.9999	D	0.9551	F	0.8839
F	F	0.9999	I	0.9152	D	0.9027
G	G	0.9999	D	0.8567	E	0.8401
H	H	0.9999	E	0.5472	D	0.4709
I	I	0.9999	D	0.9293	F	0.9168
J	J	0.9999	F	0.6331	B C	0.5495 0.6027

The results demonstrated that all samples used to build the library were correctly identified.

5.5.3. External Validation

External specificity refers to the proper identification of samples, which were not used to generate the spectrum library. The spectra of unknown samples were compared to the library.

During external validation, the following parameters were assessed:

5.5.3.1. Specificity - positive control;

5.5.3.2. Specificity - negative control;

5.5.3.3. Robustness.

5.5.3.1. Specificity - Positive Control

The positive control acceptance criteria require that all positive control spectra compared to the library meet the acceptance criteria (CC > 0.95 and CC obtained has to be the highest value).

The results of the positive control measurement are shown in **Table 19**, the CC value corresponding to the mean result for each measurement.

Table 19 - External Validation - Positive Control					
Raw Material External Batch	**External Batch Identified as Top Match**		**Raw Material External Batch**	**External Batch Identified as Top Match**	
	Identification	**CC**		**Identification**	**CC**
A	A	0.9974 0.9977 0.9977 0.9985 0.9981	**F**	F	0.9994 0.9980 0.9992 0.9992 0.9991
B	B	0.9911 0.9901 0.9951 0.9896 0.9770	**G**	G	0.9998 0.9996 0.9996 0.9994 0.9957
C	C	0.9997 0.9998 0.9991 0.9993 0.9994	**H**	H	0.9993 0.9998 0.9994 0.9997 0.9993
D	D	0.9882 0.9978 0.9985 0.9984 0.9983	**I**	I	0.9997 0.9997 0.9998 0.9996 0.9993
E	E	0.9996 0.9994 0.9996 0.9982 0.9978	**J**	J	0.9995 0.9995 0.9998 0.9997 0.9994

The results demonstrated that the library correctly identified all external samples.

5.5.3.2. Specificity - Negative Control

To perform the specificity - negative control, it was necessary to know which groups of raw materials could be *confused*. This information was obtained from the internal validation results, **Table 20** – Internal Validation Results. The referred table defines the *Top Match* (the correct identification) and the two other groups that have the next highest CC values (*2nd* and *3rdMatch*). These two groups were used to perform the negative control.

The results of the negative control measurement are shown in **Table 20**.

Table 20 - External Validation - Negative Control

Selected material in library (Model ID)	Raw material (Negative control)	Identified as (*Top Match*)	Top CC
A	C	C	0.9983
	D	D	0.9977
	I	I	0.9987
B	F	F	0.9980
	J	J	0.9985
	C	C	0.9982
C	I	I	0.9979
	D	D	0.9984
	F	F	0.9981
D	E	E	0.9997
	I	I	0.9980
E	D	D	0.9979
	F	F	0.9980
	I	I	0.9979
F	I	I	0.9993
	D	D	0.9970
G	D	D	0.9967
	E	E	0.9996
H	E	E	0.9994
	D	D	0.9980
I	D	D	0.9979
	F	F	0.9991
J	F	F	0.9973
	B	B	0.9633
	C	C	0.9973

The results demonstrated that the library correctly identified all negative control samples.

5.5.3.3. Robustness

Robustness is a measure of the method's capacity to remain unaffected when submitted a small but deliberate variations.

The following parameters were studied:

5.5.3.3.1 Different operators;

5.5.3.3.2. Changes in pre-processing;

5.5.3.3.3. Different temperatures;

5.5.3.3.4. Different bag thickness

The first and second parameters were covered as part of the development of the library.

5.5.3.1. Different Operators

During the library development, three operators collected five spectra from each batch of each raw material, so, these parameters were evaluated during the library development. No significant differences were found between the spectra obtained with the three operators.

Despite not having shown significant differences, to evaluate this parameter another operator performed the analysis of each raw material. The obtained results were summarized in **Table 21**:

Table 21 - Robustness: Operator

Selected material in library (Model ID)	Identified as(*Top Match*)	Top CC
A	**A**	0.9979
B	**B**	0.9911
C	**C**	0.9988
D	**D**	0.9972
E	**E**	0.9996
F	**F**	0.9989
G	**G**	0.9998
H	**H**	0.9981
I	**I**	0.9994
J	**J**	0.9986

5.5.3.2. Changes in pre-processing

Changes in pre-processing are an important parameter to have into account during library and method development. This is essential to prevent the mismatch between groups of raw materials. In this work, the best groups separation was obtained using the following pre-processes: S.G. (derivative: 1, smooth points: 3; degree: 2), Normalize Range over the spectrum range 1670 to 2300 nm. **Figure 39** shows the spectrum after applying the described pre-treatment.

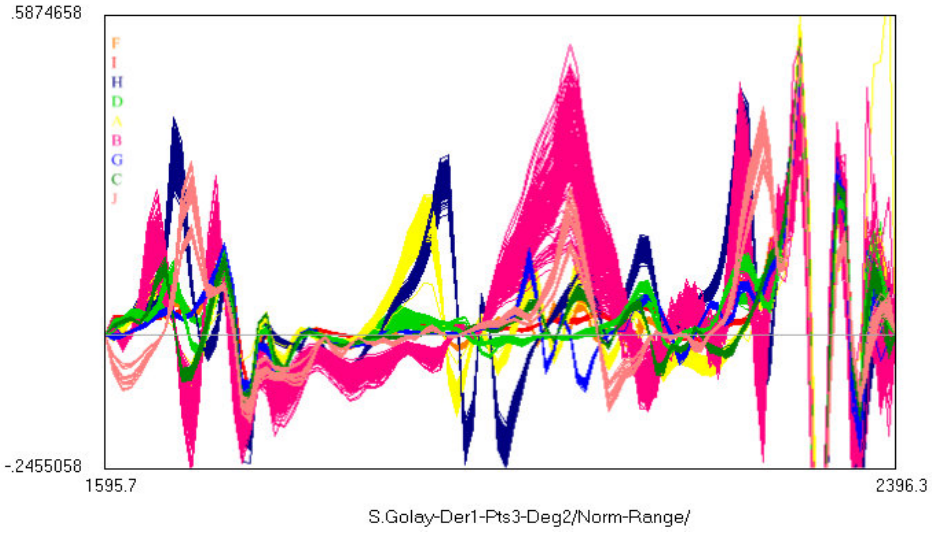


Figure 39 – Raw material spectrum after the pre-treatment application

5.5.3.3. Different temperatures:

The temperature is a critical factor in the acquisition of NIR spectrum. Thus, to evaluate the influence of temperature, two tests were carried out:

- Variation of the sample temperature (cold / hot)
- Variation of room temperature

To assess the robustness of the developed method, samples were kept for 60 minutes under 40 ° C and -8 ° C conditions.

As well as the sample temperature, the environment temperature can also be a critical factor for the acquisition of spectrum. Taking as a reference temperature of 20ºC - 25 ºC for the conditions of spectrum acquisition, to evaluate the variation of environmental temperature the acquisition of the spectrum was carried out in a room with an ambient temperature of 32 ° C. The results are listed in **Table 22**.

Table 22 – Robustness: Temperature Effect

Selected material in library (Model ID)	Identified as (*Top Match*)	Sample Temperature Conditions		Environment Temperature
		Top CC (about -8ºC)	Top CC (about 40ºC)	Top CC (32ºC)
A	**A**	0.9984	0.9948	0.9963
B	**B**	0.9758	0.9913	0.9874
C	**C**	0.9994	0.9984	0.9988
D	**D**	0.9988	0.9967	0.9987
E	**E**	0.9995	0.9981	0.9992
F	**F**	0.9994	0.9990	0.9985
G	**G**	0.9996	0.9990	0.9994
H	**H**	0.9999	0.9989	0.9988
I	**I**	0.9996	0.9990	0.9979
J	**J**	0.9995	0.9993	0.9990

5.5.3.4. Different bag thickness:

Raw material reception is performed in the warehouse under humidity and temperature controlled conditions, and its storage is a critical factor since it must guarantee its conformity, most commonly being done in polyethylene bags. Analysis through this polymeric material would be the fastest possible way, not requiring any sampling, and without the need to breach the packaging to assess the material identification.

However, not every raw material spectrum was acquired through the original packaging due to some being coloured or because some of the batches were reference samples. To minimize this variation, every sample was placed inside a transparent polyethylene bag.
To assess the different thicknesses of these bags, that can influence the identification, different thicknesses were simulated.

The obtained results were summarized in **Table 23**:

Table 23 - Robustness: Effect of the bag thickness

Raw Material	Test: 3 x NTB		Test: 5 x NTB		Test: 7 x NTB	
	Top Match	Top CC	Top Match	Top CC	Top Match	Top CC
A	NO-ID	NO-ID				
B	NO-ID	NO-ID				
C	C	0.9770	NO-ID	NO-ID		
D	D	0.9901	D	0.9557		
E	D	0.9842	D	0.9774		
F	F	0.9812	F	0.9554	NO-ID	NO-ID
G	G	0.9728	NO-ID			
H	NO-ID	NO-ID				
I	I	0.9872	I	0.9569		
J	J	0.9656	NO-ID			

NTB: normal thickness of the bag used to build library; **NO-ID:** when the library does not identify the raw material

If the first test did not comply, the next test was not performed. The test was carried out until obtaining an identification result.

Regarding the tested parameters for robustness it is possible to conclude that the method developed is indeed robust.
For the later test (thickness evaluation) non consistent results were obtained, however, the methodology currently adopted for acquiring spectrum is the same as the methodology by which the library was developed and validated. Samples of raw material are introduced into a polyethylene bag, and the spectrum acquisition is performed.

In summary, it can be concluded that the method is robust for normal operating conditions and that changing temperature (-8ºC to 40ºC) and bag thickness will not lead to false identifications. Bag thickness can influence the identification, but with the normal operating conditions described in **4.1.2. Samples - Identification Method** no false identifications were obtained.

5.6. Method impact in Quality Control – OM Pharma

Focusing only on the applicability of this identification method in the API involved in this work the analysis is associated a time spent in Quality Control. Considering that the time spent in the analysis per batch (each batch is composed by 5 containers and no consider a sampling time) with the reference method identification was 45 minutes and with the NIR method was 15 minutes. NIR method represents about 25% of the time per year needed to analyze all API batches when compared with the reference method.

The impact of cost and associated to the new methodology requires an economic analysis more detailed that is not the scope of this work. However it should be noted that the gain in terms of time is directly related to costs reduced (consumables) associated with the reference method.
NIR identification method implementation at Quality Control is profitable in terms of costs and time.

6. CONCLUSION

First, a NIR method was developed for a direct assay of API in a blend, a critical step on a particular pharmaceutical form production. The method was validated in terms of specificity, linearity, precision, accuracy and robustness and the results demonstrated that the developed method is appropriate for the quantification in a range 80% to 110% (w/w, API).

The proposed method for API assay was in good agreement with the reference Titration method in use at OM.

Second, an identification method was developed for the direct API identification in the warehouse. The identification method was validated (internal and externally) and the results demonstrated that the developed method is appropriate for the identification of all API solid OM Pharma.

The proposed methods can be used as a tool for rapid characterization of API in this process involved both in the phase of reception at the OM Pharma Warehouse as well as the control of the finished product associated with their use.

The developed procedures are adequate for the determination of the API content as well as identifying APIs without any sample preparation and may be useful for the future use and development of Process Analytical Technology in OM Pharma.

NIR spectroscopy, when in association with chemometric techniques, allow the construction of calibration models, able to quantify the API content in a several components mixture.

NIR technique presents several advantages, namely operator errors being almost eliminated, sample preparation is minimal and time reduction. This does not happen in other reference methods. Also, it is able to simultaneously determine several parameters with high speed and precision, with low costs, as well as not producing any environmentally harmful residues.

By using NIR methods it is possible to carry out extremely accurate quantitative analysis in the production control (control of the final production phase) and quality control of raw materials in the pharmaceutical manufacturing. This implementation helps the laboratory and manufacturing workers to improve productivity by reducing analysis time and the analyzer can also aid regulatory compliance. By using the NIR methods for API quantification and raw material analysis, OM Pharma has saved a significant amount of time in the laboratory, as well as improving productivity, and reducing operating costs.

Library maintenance is an important factor because if new suppliers of some API are approved by OM Pharma, a sample of these products should be assessed by the built library. When the new raw material (from new supplier) is tests with the library a re-validation can be required.

The current work opens several possibilities at OM for using the proposed methods and approaches to other products and quality assurance and control in routine activities at the company.

7. FUTURE WORK

Perfecting the calibration model is an important factor for the reduction of the associated error, even though it demonstrated itself precise for the API determinations and that it met all the required criteria. This improvement must be a continuous methodology, by incorporating new batches as they are received, with the goal to improve the method robustness and thus lower the error.

This work approached the API quantification in the mixing phase but in the future, the application of this NIR quantification method to finished product can also be considered.

The built library is now at an implementation phase and an excipients identification library is already being built.

The API library may, nonetheless, be simplified when it comes to spectrum acquisition. When initially built, OM Pharma did not have enough batches in their original packaging. To overcome this fact the acquisition was made with retention samples. At this moment, the library is being improved and a query on which raw-materials can be acquired directly in the original packaging was performed, in order to eliminate the sampling and so its acquisition can be made *in loco* (Warehouse)

For this work only a API library was built, nonetheless, OM Pharma is developing and building a library for every raw-materials alongside the evaluation of the spectrum acquisition.

While developing this project, an API identification followed by its quantitative assay when incorporated in the formulation was performed. Both these steps were discussed and the results are compiled in this work. Nonetheless, it is of great interest, and a future goal, to perform the API qualitative NIR characterization, such as water content, particle size and potency.

This work has boosted the use of this technology; in the future this project will be continued and enhanced. Due to NIR spectroscopy capabilities and advantages, in the future, the Quality Control pretends to develop new projects in order to implement this technique to other applications.

8. REFERENCES

[1] Aenugu, H. P. R.; Kumar, D. S.; *et al*; *Near infrared Spectroscopy - An Overview,* ChemTech, Vol. 3, No. 2 (2011), 825 – 836

[2] Luypaert, J.; Massart, D. L.; Heyden, Y. V.; *Near-infrared spectroscopy application in pharmaceutical analysis,* Talanta, 72 (2007), 865 – 883

[3] Xie, L; Ye, Xingqian, Liu, D.; Ying, Y.; *Quantification of glucose, fructose and sucrose in bayberry juice by NIR and PLS,* Food Chemistry, 114 (2009), 1135 –1140

[4] Li, W.; Xing, L.; Cai, Y.; Qu, Haibin; *Classification and quantification analysis of Radix scutellariae from different origin with near infrared diffuse reflection spectroscopy,* Vibrational Spectroscopy, 55 (2011), 58 –64

[5] Porfine, A.; Rus, L.; Vonica, A. L.;*High-throughput NIR-chemometric methods for determination of drug content and pharmaceutical properties of indapamida powder blends for tableting,* Journal of Pharmaceutical and Biomedical Analysis, 70 (2012), 301 – 309

[6] Moes, J. J.; Ruijken, M. M.; *et al*; *Application of process analytical technology in tablets process development using NIR spectroscopy: Blend uniformity, content uniformity and coating thickness measurements,* International Journal of Pharmaceutics , 357 (2008), 108 – 118

[7] Vanarase, A. U.; Alcalà, M;*et al*; *Real-time monitoring of drug concentration in a continuous powder mixing process using NIR spectroscopy,* Chemical Engineering Science, 65 (2010), 5728 – 5733

[8] Jamrógiewicz, M.; Cal, K.; Gruszeca, M.; *et al*; *Determination of API content in a pilot-scale blending by near-infrared spectroscopy as a first step method to process line implementation,* Acta Poloniae Pharmaceutica – Drug Research, Vol. 70, No. 3 (2013), 419 – 429

[9] Jarvinen, K.; Hoehe, W.; *et al*.; *In-line monitoring of the drug content of powder mixture and tablets by near-infrared spectroscopy during the continuous direct compression tableting process,* European Journal of Pharmaceutical Sciences, 48, (2013), 680 – 688

[10] Tamuta, I.; Rus, L.; Iovanov, R.; Rus, L. L.; *High-throughput NIR-chemometric method for determination of drug content and pharmaceutical properties of indapamide tablets,* Journal of Pharmaceutical and Biomedical Analysis, 84 (2013), 285 – 292

[11] Ziémons, E.; Mantanus, J.; Lebrun, P.; Rozet, E.; Evrard, B; Hubert, Ph; *Acetaminophen determination in low-dose pharmaceutical syrup by NIR spectroscopy,* Journal of Pharmaceutical and Biomedical Analysis, 53 (2010), 510 – 516

[12] Liew, C. V.; Karande, A. D.; Heng, P. W. S.; *In-line quantification of drug and excipients in cohesive powder blends by near infrared spectroscopy,* International Journal of Pharmaceutics Chemical, 386 (2010), 138 – 148

[13] Xie, Y.; Tao, W.; Morrison, H; *et al*.; *Quantitative determination of solid-state forms of pharmaceutical development compound in drug substance and tablets,* International Journal of Pharmaceutics, 362 (2008), 29 –36

[14] Blanco, M; Peguero, A.; Influence of physical factors on the accuracy of calibration models for NIR spectroscopy, Journal of Pharmaceutical and Biomedical Analysis, 52 (2010), 59 –65

[15] Almeida, F. M. N.; *Espectroscopia de Infravermelho Próximo com Transformada de Fourier (FT-NIR) na caracterização de farinhas para alimentação Pueril,* Dissertação para obtenção do grau de mestre em Engenharia Biológica, (2009)

[16] Volta e Sousa, L. M. G..; *Using NIR spectroscopy for Raw Materials Identification,* Dissertação para obtenção do grau de mestre em Engenharia Farmacêutica, (2010)

[17] Ghadimi. A.; Sadrzadeh, M.; Mohammadi, T.; *Prediction of ternary gas permeation through synthesized PDM membranes by using Principal Component Analysis (PCA) and fuzzy logic (FL),* Journal of Membrane Science, 360 (2010), 509 – 521

[18] Reich, G.; *Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical application,* Advanced Drug Delivery Reviews, 57 (2005), 1109 – 1143

[19] Baptista, P.; Felizardo, P.;Menezes, J. C..; *Multivariate near infrared spectroscopy models for predicting the methyl esters content in biodisel,* Analytica Chimica Acta, 607 (2008), 153 – 159

[20] Eustaquio, A.; Graham, P.; *et al*; *Quantification of paracetamol in intact tablets using near-infrared transmittance spectroscopy,* Analyst, 123 (1998), 2303 – 2306

[21] Schaefer, C,;Clicq. D.; *et al*; *A process analytical technology (PAT) approach to control a new API manufacturing process: Development, validation and implementation,* Talanta, 120, (2014), 114 –125

[22] Amorim, S. R.; Klier, A. H.; Angelis, L. H.; *Controle de Qualidade na indústria farmacêutica: identificação de substâncias por espectroscopia no Infravermelho,* Ver. Bras. Farm., 94 (3), (2013), 234 – 242

[23] Alvarenga, L.; Ferreira, D.; Menezes, J. C.; Lochmann, D.; *Tablet identification using near-infrared spectroscopy (NIRS) for pharmaceutical quality control,* Journal of Pharmaceutical and Biomedical Analysis, 48 (2008), 62 – 69

[24] Sarraguça, M. C.; Lopes, J. A.; *Quality control of pharmaceuticals with NIR: from lab to process line,* Vibrational Spectroscopy, 49, (2009), 204 – 210

[25] Karande, A. D.; Heng, P. W. S.; Liew, C. V.;*In-line quantification of micronized drug and excipients in tablets by near infrared (NIR) spectroscopy: Real time monitoring of tableting process,* Journal of Pharmaceutics, 396 (2010),63 – 74

[26] Wahl, P. R.; Fruhmann, G.; *et al*; *PAT for tableting: Inline monitoring of API and excipients via NIR,* European Journal of Pharmaceutics and Biopharmaceutics, 87, (2014), 271 – 278

[27] Tomuta, I.; Iovanov, R.; Bodoki, E.; Leucuta, S. E.; *Identification of meloxicam and excipients on intact tablets by near infrared spectroscopy and chemometry,* Farmacia, Vol. 58, No. 5 (2010)

[28] European Medicines Agency, *Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and data requirements for new submission and variation*;

[29] Chen, Q.; Cai, J., Wan, X.; Zhao, J.; *Application of linear/non-linear classification algorithms in discrimination of pork storage time using Fourier transform near infrared (FT-NIR) spectroscopy*; Food Science and Technology 44 (2011) 2053 – 2058

[30] Jiang, H.; Liu, G.; Xiao, X. et al; *Monitoring of solid-state fermentation of wheat straw in a pilot scale using FT-NIR spectroscopy and support vector data description*; Microchemical Journal 102 (2012) 68–74

[31] MicroPHAZIR™ User Manual, Thermo Scientific Handheld Near-Infrared Analyzer

[32] Eustaquio, A.; Graham, P.; Jee, R. D.; *et al*; *Quantification of paracetamol in intact tablets using near-infrared transmittance spectroscopy,* The Analyst, 123, (1998), 2303 – 2306

[33] Blanco, M.; Bautista, M.; Alcalà, M; *API determination by NIR spectroscopy across pharmaceutical production process – Research Article,* AAPS PharmaSciTech, vol. 9, No. 4 (2008), 1130 – 1135

[34] Schonbichler, S. A.; Bittner, L. K. H. *et al*; *Simultaneous quantification of verbenalin and verbascoside in Verbena officialis by ATR-IR and NIR spectroscopy,* Journal of Pharmaceutical and Biomedical Analysis, 84 (2013), 97 –102

[35] Pierna, J. A. F.; Vermeulen, P.*; et al*; *Calibration Transfer from Disoersive Instruments to Handheld spectrometers,* Applied spectroscopy, 64 (2010), 644 – 648

[36] G. Downes; *Technical Report 212 – Developing and assessing a wood meal calibration for the Polychromix Phazir™,* Cooperative Research Center for Forestry, June 2011

[37] Pharmaceutical Analysis Science Group, *Guidelines for the development and validation of near infrared (NIR) spectroscopic methods*;