Université de Lille 2
Année Universitaire 2017/2018

Faculté des Sciences Pharmaceutiques
et Biologiques de Lille

THESE
POUR LE DIPLOME D'ETAT
DE DOCTEUR EN PHARMACIE

Soutenue publiquement le 19 Octobre 2017
Par Mme Comyn Hélène

DEVELOPMENT OF AN ANALYTICAL METHOD FOR ACTIVE PHARMACEUTICAL INGREDIENT ASSAY BY NEAR INFRARED SPECTROSCOPY

Membres du jury :

Président : Jean-François Goossens, Professeur des Universités

Directeur, conseiller de thèse : Samd Guizani, Process Scientist, Ferring Pharmaceuticals

Assesseur(s) : Josselin Saunier, Project Manager, Ferring Pharmaceuticals

Faculté des Sciences Pharmaceutiques
et Biologiques de Lille

3, rue du Professeur Laguesse - B.P. 83 - 59006 LILLE CEDEX
☎ 03.20.96.40.40 - 🖷 : 03.20.96.43.64
http://pharmacie.univ-lille2.fr

Université Lille 2 – Droit et Santé

Président :	Professeur Xavier VANDENDRIESSCHE
Vice-présidents :	Professeur Alain DUROCHER Professeur Régis BORDET Professeur Eric BOULANGER Professeur Frédéric LOBEZ Professeur Murielle GARCIN Professeur Annabelle DERAM Professeur Muriel UBEDA SAILLARD Monsieur Ghislain CORNILLON Monsieur Pierre RAVAUX Monsieur Larbi AIT-HENNANI Madame Nathalie ETHUIN Madame Ilona LEMAITRE
Directeur Général des Services :	Monsieur Pierre-Marie ROBERT

Faculté des Sciences Pharmaceutiques et Biologiques

Doyen :	Professeur Damien CUNY
Vice-Doyen, 1er assesseur :	Professeur Bertrand DECAUDIN
Assesseur en charge de la pédagogie	Dr. Annie STANDAERT
Assesseur en charge de la recherche	Pr. Patricia MELNYK
Assesseur délégué à la scolarité	Dr. Christophe BOCHU
Assesseur délégué en charge des relations internationales	Pr. Philippe CHAVATTE
Assesseur délégué en charge de la vie étudiante	M. Thomas MORGENROTH
Chef des services administratifs :	Monsieur Cyrille PORTA

Liste des Professeurs des Universités - Praticiens Hospitaliers

Civ.	NOM	Prénom	Laboratoire
Mme	ALLORGE	Delphine	Toxicologie
M.	BROUSSEAU	Thierry	Biochimie
M.	DECAUDIN	Bertrand	Pharmacie Galénique
M.	DEPREUX	Patrick	ICPAL
M.	DINE	Thierry	Pharmacie clinique
Mme	DUPONT-PRADO	Annabelle	Hématologie
M.	GRESSIER	Bernard	Pharmacologie
M.	LUYCKX	Michel	Pharmacie clinique
M.	ODOU	Pascal	Pharmacie Galénique
Mme	RENNEVILLE	Aline	Hématologie
M.	STAELS	Bart	Biologie Cellulaire

Liste des Professeurs des Universités

Civ.	NOM	Prénom	Laboratoire
M.	ALIOUAT	El Moukhtar	Parasitologie
Mme	AZAROUAL	Nathalie	Physique
M.	BERTHELOT	Pascal	Onco et Neurochimie
M.	CAZIN	Jean-Louis	Pharmacologie – Pharmacie clinique
M.	CHAVATTE	Philippe	ICPAL
M.	COURTECUISSE	Régis	Sciences végétales et fongiques
M.	CUNY	Damien	Sciences végétales et fongiques
Mme	DELBAERE	Stéphanie	Physique
M.	DEPREZ	Benoît	Laboratoire de Médicaments et Molécules
Mme	DEPREZ	Rebecca	Laboratoire de Médicaments et Molécules
M.	DUPONT	Frédéric	Sciences végétales et fongiques
M.	DURIEZ	Patrick	Physiologie
M.	FOLIGNE	Benoît	Bactériologie
M.	GARÇON	Guillaume	Toxicologie
Mme	GAYOT	Anne	Pharmacotechnie Industrielle
M.	GOOSSENS	Jean François	Chimie Analytique
M.	HENNEBELLE	Thierry	Pharmacognosie
M.	LEMDANI	Mohamed	Biomathématiques
Mme	LESTAVEL	Sophie	Biologie Cellulaire
M.	LUC	Gerald	Physiologie
Mme	MELNYK	Patricia	Onco et Neurochimie
M.	MILLET	Régis	ICPAL
Mme	MUHR – TAILLEUX	Anne	Biochimie
Mme	PAUMELLE-LESTRELIN	Réjane	Biologie Cellulaire
Mme	PERROY	Anne Catherine	Législation
Mme	ROMOND	Marie Bénédicte	Bactériologie
Mme	SAHPAZ	Sevser	Pharmacognosie
M.	SERGHERAERT	Eric	Législation
Mme	SIEPMANN	Florence	Pharmacotechnie Industrielle
M.	SIEPMANN	Juergen	Pharmacotechnie Industrielle
M	TARTAR	André	Laboratoire de Médicaments et Molécules
M.	WILLAND	Nicolas	Laboratoire de Médicaments et Molécules

Liste des Maîtres de Conférences - Praticiens Hospitaliers

Civ.	NOM	Prénom	Laboratoire
Mme	BALDUYCK	Malika	Biochimie
Mme	GARAT	Anne	Toxicologie
Mme	GOFFARD	Anne	Bactériologie
M.	LANNOY	Damien	Pharmacie Galénique
Mme	ODOU	Marie Françoise	Bactériologie
M.	SIMON	Nicolas	Pharmacie Galénique

Liste des Maîtres de Conférences

Civ.	NOM	Prénom	Laboratoire
Mme	ALIOUAT	Cécile Marie	Parasitologie
M.	ANTHERIEU	Sébastien	Toxicologie
Mme	AUMERCIER	Pierrette	Biochimie

Mme	BANTUBUNGI	Kadiombo	Biologie cellulaire
Mme	BARTHELEMY	Christine	Pharmacie Galénique
Mme	BEHRA	Josette	Bactériologie
M	BELARBI	Karim	Pharmacologie
M.	BERTHET	Jérôme	Physique
M.	BERTIN	Benjamin	Immunologie
M.	BLANCHEMAIN	Nicolas	Pharmacotechnie industrielle
M.	BOCHU	Christophe	Physique
M.	BORDAGE	Simon	Pharmacognosie
M.	BOSC	Damien	Laboratoire de Médicaments et Molécules
M.	BRIAND	Olivier	Biochimie
Mme	CACHERA	Claude	Biochimie
M.	CARNOY	Christophe	Immunologie
Mme	CARON	Sandrine	Biologie cellulaire
Mme	CHABÉ	Magali	Parasitologie
Mme	CHARTON	Julie	Laboratoire de Médicaments et Molécules
M	CHEVALIER	Dany	Toxicologie
M.	COCHELARD	Dominique	Biomathématiques
Mme	DANEL	Cécile	Chimie Analytique
Mme	DEMANCHE	Christine	Parasitologie
Mme	DEMARQUILLY	Catherine	Biomathématiques
Mme	DUMONT	Julie	Biologie cellulaire
Mme	DUTOUT-AGOURIDAS	Laurence	Onco et Neurochimie
M.	EL BAKALI	Jamal	Onco et Neurochimie
M.	FARCE	Amaury	ICPAL
Mme	FLIPO	Marion	Laboratoire de Médicaments et Molécules
Mme	FOULON	Catherine	Chimie Analytique
M.	FURMAN	Christophe	ICPAL
M.	GELEZ	Philippe	Biomathématiques
Mme	GENAY	Stéphanie	Pharmacie Galénique
M.	GERVOIS	Philippe	Biochimie
Mme	GOOSSENS	Laurence	ICPAL
Mme	GRAVE	Béatrice	Toxicologie
Mme	GROSS	Barbara	Biochimie
M.	HAMONIER	Julien	Biomathématiques
Mme	HAMOUDI	Chérifa Mounira	Pharmacotechnie industrielle
Mme	HANNOTHIAUX	Marie-Hélène	Toxicologie
Mme	HELLEBOID	Audrey	Physiologie
M.	HERMANN	Emmanuel	Immunologie
M.	KAMBIA	Kpakpaga Nicolas	Pharmacologie
M.	KARROUT	Youness	Pharmacotechnie Industrielle
Mme	LALLOYER	Fanny	Biochimie
M.	LEBEGUE	Nicolas	Onco et Neurochimie
Mme	LECOEUR	Marie	Chimie Analytique
Mme	LEHMANN	Hélène	Législation
Mme	LELEU-CHAVAIN	Natascha	ICPAL
Mme	LIPKA	Emmanuelle	Chimie Analytique
Mme	MARTIN	Françoise	Physiologie
M.	MOREAU	Pierre Arthur	Sciences végétales et fongiques
Mme	MUSCHERT	Susanne	Pharmacotechnie industrielle
Mme	NIKASINOVIC	Lydia	Toxicologie
Mme	PINÇON	Claire	Biomathématiques

M.	PIVA	Frank	Biochimie
Mme	PLATEL	Anne	Toxicologie
M.	POURCET	Benoît	Biochimie
M.	RAVAUX	Pierre	Biomathématiques
Mme	RAVEZ	Séverine	Onco et Neurochimie
Mme	RIVIERE	Céline	Pharmacognosie
Mme	ROGER	Nadine	Immunologie
M.	ROUMY	Vincent	Pharmacognosie
Mme	SEBTI	Yasmine	Biochimie
Mme	SINGER	Elisabeth	Bactériologie
Mme	STANDAERT	Annie	Parasitologie
M.	TAGZIRT	Madjid	Hématologie
M.	VILLEMAGNE	Baptiste	Laboratoire de Médicaments et Molécules
M.	WELTI	Stéphane	Sciences végétales et fongiques
M.	YOUS	Saïd	Onco et Neurochimie
M.	ZITOUNI	Djamel	Biomathématiques

Professeurs Agrégés

Civ.	NOM	Prénom	Laboratoire
Mme	MAYES	Martine	Anglais
M.	MORGENROTH	Thomas	Législation

Professeurs Certifiés

Civ.	NOM	Prénom	Laboratoire
M.	HUGES	Dominique	Anglais
Mlle	FAUQUANT	Soline	Anglais
M.	OSTYN	Gaël	Anglais

Professeur Associé - mi-temps

Civ.	NOM	Prénom	Laboratoire
M.	DHANANI	Alban	Droit et Economie Pharmaceutique

Maîtres de Conférences ASSOCIES - mi-temps

Civ.	NOM	Prénom	Laboratoire
M.	BRICOTEAU	Didier	Biomathématiques
Mme	CUCCHI	Malgorzata	Biomathématiques
M.	FRIMAT	Bruno	Pharmacie Clinique
M.	GILLOT	François	Droit et Economie pharmaceutique
M.	MASCAUT	Daniel	Pharmacie Clinique
M.	ZANETTI	Sébastien	Biomathématiques
M.	BRICOTEAU	Didier	Biomathématiques

AHU

Civ.	NOM	Prénom	Laboratoire
Mme	DEKYNDT	Bérengère	Pharmacie Galénique
M.	PEREZ	Maxime	Pharmacie Galénique

Université de Lille 2 DROIT ET SANTÉ

Faculté des Sciences Pharmaceutiques et Biologiques de Lille

Faculté des Sciences Pharmaceutiques et Biologiques de Lille

3, rue du Professeur Laguesse - B.P. 83 - 59006 LILLE CEDEX
Tel. : 03.20.96.40.40 - Télécopie : 03.20.96.43.64
http://pharmacie.univ-lille2.fr

L'Université n'entend donner aucune approbation aux opinions émises dans les thèses ; celles-ci sont propres à leurs auteurs.

Abstract

Spectroscopic techniques are major tools in the chemistry field nowadays, with various applications. Among them, Near Infrared Spectroscopy (NIRS) is a booming tool for quantitative and qualitative analysis in the pharmaceutical industry.

The objective of the work that supported this thesis was to use NIRS in order to develop a fast analytical method for the assay of an API in a particular finish product. The aim was to have an alternative to the reference method currently used for batch release (HPLC).

This whole strategy includes the calibration model development, the comparison with the reference method, and the method validation according to internal and international guidelines on analytical methods.

This thesis presents the feasibility study performed in order to build the calibration model, and the method validation.

A satisfactory analytical method was developed for two different sample forms of the drug product, granules and powder forms (resulting from the crushing of the granules). Multi variate data analysis was used to build the model and assess its performance.

The model presents a precision (RMSECV) of 0.6% label claim, and was assessed at least equivalent to HPLC if not better in term of precision. The validation according to ICH guidelines was successfully performed, with all criteria (precision, linearity etc.) being satisfactory.

Résumé

Les méthodes spectroscopiques sont devenues des techniques incontournables dans le milieu de la Chimie, avec un immense champ de possibilités. Parmi celles-ci, la Spectroscopie Proche Infrarouge (NIRS) est une technique analytique en plein essor dans l'industrie pharmaceutique.

L'objectif du travail qui a supporté cette thèse était de développer une méthode d'analyse par spectroscopie proche infrarouge pour le dosage de principe actif dans un produit fini, à des fins de libération de lots. Cette nouvelle méthode, plus rapide, et plus économique, présente une alternative à la méthode de dosage actuelle par HPLC.

Ce projet inclut le développement du modèle de calibration, la comparaison du modèle obtenu avec la méthode de référence, et la validation de la méthode selon les réglementations internes et internationales relatives aux méthodes analytiques.

Cette thèse présente l'étude de faisabilité menée lors de la construction du modèle de calibration, ainsi que la stratégie de validation de la méthode et ses résultats.

Une méthode analytique satisfaisante a été développée, pour deux formes différentes du produit : sous forme granules ou poudres (résultant du broyage des granules). L'analyse de données multivariées a été utilisée pour la construction du modèle.

Le modèle a une précision (RMSECV) de 0.6% de la dose nominale, et a été démontré au moins équivalent à la méthode HPLC, si ce n'est meilleur en termes de précision.

La validation selon les recommandations ICH a été réalisée avec succès, tous les paramètres testés (linéarité, précision etc.) ayant rempli les critères.

Preliminary

Cette thèse a été écrite sur les bases du travail réalisé au cours de mon stage de fin d'études. Ce stage de six mois m'a permis de valider ma 6ème Année de Pharmacie ainsi que mon diplôme d'ingénieur en Génie des procédés.

Toutes les expériences décrites ont été réalisées au sein du Pilot Plant, unité pilote du Ferring International Centre SA (Saint Prex, Switzerland), ayant vocation à faire le lien entre la R&D et la Production chez Ferring.

Contacts:
Ferring Pharmaceuticals
Innovation Centre – Pilot Plant
M. Robert Rönnback: robert.roennback@ferring.com

Remerciements

Peu d'occasions sont données de remercier formellement tous les gens qui nous ont permis d'en arriver là où nous en sommes aujourd'hui, alors peut-être que la liste sera un peu longue…

Mes premiers remerciements vont bien sûr à mes parents. Vous nous avez transmis l'envie d'apprendre bien avant d'aller à l'école, et chaque jour à vos côtés était une nouvelle découverte, dans la simplicité. Merci pour tout le soutien moral et financier apporté au long de ces 7 années. Merci de m'avoir aiguillée sans jamais rien m'avoir imposé, merci de m'avoir supportée dans mes choix. Merci d'être toujours au bout du fil en cas de coup dur.

Je remercie mon jury d'avoir accepté de juger ma thèse. Merci Samd pour ta bienveillance depuis que je t'ai eu en maître de stage. Apprendre avec toi est passionnant. Merci M. Goossens d'avoir accepté d'être Président de ce jury malgré la distance. Merci Josselin d'être là aussi, je ne vais pas détailler ma gratitude, ça te ferait trop plaisir !

Merci à mes frères et sœur, Anne, Rémi et Simon. Grandir avec vous a été une école de la vie. Il est loin le temps où nous nous disputions le pot de Nutella, les relations se font parfois plus distantes, mais toujours constantes.

Merci au reste des coplègues du Pilot Plant : Bruno, Cédric, Anamaria… ne changez pas ! Ou en mieux (?!)

Merci aux PLA, Camille, Alexandra, Aurélie, François, Marie, Matthieu, Caro. Onze ans d'amitié, et pas une ride. Chaque moment passé avec vous est une cure de jouvence. Merci de m'avoir donnée une place dans ce monde au moment où je la cherchais le plus, et de m'avoir aimée pour ce que j'étais.

Merci à mes copines de pharma, merci Céline pour toutes ces heures à supporter les cours et les TP (on aura bien ri…), d'avoir toujours été là pour m'écouter. Et Clémence, que dire ? On n'a jamais trop fait dans les bla-bla, je ne commencerai donc pas aujourd'hui ! J'espère qu'on gardera cette tendance à se retrouver au même endroit au même moment, sans jamais l'avoir planifié.

Merci à mes copines de l'ENSIC, Cécile et Laurine. Cécile, pour avoir grandement contribué à diminuer mon nombre de rattrapages, et m'avoir sauvée de la dépression profonde d'une manière générale... ! Laurine, pour m'avoir prouvé qu'une jolie amitié peut naître de l'adversité, même quand c'était parti du mauvais pied. A toutes les deux, pour tous ces bons moments qui nous ont permis de tenir le coup au travers de deux années compliquées.

To Karishma and Katherine, my friends overseas. Thank you for making my year abroad such a wonderful experience. I always have a place to come back to whenever I need my England

cup of tea, and I'm enjoying every moment with you. Miss you so much girls, and miss my life there with you loads! To Fe, thank you for helping me writing the last words of this thesis, and for believing in me much more than I do myself!

Merci aux amis de longue date, Marion-jumelinette, Florence et Samuel, Merci Bénédicte pour toutes ces années partagées depuis le collège, pour m'avoir patiemment écoutée réciter la classification des angiospermes, le cycle de Krebbs et autres bizarreries (loin des Sciences Politiques !) pendant mes années pharmas, et particulièrement pour notre coloc inoubliable à Lille.

Enfin, merci à tous ceux que je n'ai pas encore cités mais qui comptent bien sûr beaucoup pour moi : ma grand-mère, Véro et Alain pour votre porte toujours ouverte, et les cousins au top. La team du LFB (vous me manquez !) que j'ai toujours tellement de plaisir à revoir – ce stage avec vous a été une bouffée d'air frais au milieu de ces sept longues années d'études.

J'en oublie probablement… pardon, pardon, et merci à tous !

Contents

Figures

Tables

Abbreviations and definitions

API	Active Pharmaceutical Ingredient
CG	Coated Granules
CV	Cross Validation : the entire data set of samples is split into individual samples or groups of samples, which are removed individually from the rest of the samples and tested as unknowns against a calibration model constructed using the rest of the samples. The characteristic statistic is the Standard Error of Cross Validation (SECV)
D1	First Derivative (spectral pretreatment)
DModX	Distance to model in the X-plane
DoE	Design of Experiment
FICSA	Ferring International Center S.A. : refers to the Swiss site
FTE	Full Time Employee
HPLC	High Performance Liquid Chromatography
IBD	Inflammatory Bowel Disease
ICH	International Conference on Harmonization
IR	Infra-Red
LOOCV	Leave One Out Cross Validation
MD	Mahalanobis Distance
MDI	Mahalanobis Distance Index
MDT	Mahalanobis Distance Threshold
MSC	Multiple Scattering Correction (spectral pretreatment)
MVDA	MultiVariate Data Analysis
NIR	Near InfraRed
NIRS	Near InfraRed Spectroscopy
PCA	Principal Component Analysis
PLS	Partial least square projection on latent structure: MVDA used to determine a relationship between spectra and a response.
QA	Quality Assurance
QC	Quality Control

RMSEE/RMSEC — Root Mean Square Error of Estimation/Calibration : measures the variability of the difference between the predicted and reference values for a set of calibration samples

$$SEC = \sqrt{\frac{\sum_{i=1}^{n}(y_{C,i} - Y_{C,i})^2}{n-p}}$$

Y_C = NIRS predicted value of calibration set
y_c = reference method value of calibration set
n = number of samples
p = number of coefficients, e.g. wavelength (MLR), principal components (PCR), factors (PLS)

RSMECV (=SECV) — Root Mean Square Error of Cross-Validation : measuring the difference between the NIRS procedure and reference method quantitative analyte values of the calibration set using a cross-validation method.

$$SECV = \sqrt{\frac{\sum_{i=1}^{n}(y_{CV,i} - Y_{CV,i})^2}{n}}$$

Y_{CV} = NIRS predicted value of calibration set
y_{cv} = reference method value of calibration set
n = number of samples

RMSEP[1] (=SEP) — Root Mean Square Error of Prediction : Measures the variability of the difference between the predicted and reference values for a set of independent validation samples, where:

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}(y_{V,i} - Y_{V,i})^2}{n}}$$

Y_v = NIRS predicted value
y_v = reference method value
n = number of samples

RTRT — Real Time Release Testing

SNV — Standard Normal Variate (spectral pretreatment)

SOP — Standard Operating Procedure

% w/w — Massic percentage

[1] *RMSEE/C/CV/P describe what is not explained by the model.

Introduction

Since the last decades, the demand for product quality improvement has continuously increased in the chemical, petrochemical, food, pharmaceutical industries notably. With the rise of new technologies, alternatives to time-consuming, conservative and not environmentally friendly analytical methods - such as HPLC, GC, NMR – are being developed. Spectroscopic techniques are one of these new advantageous technologies, and among them, NIR has proved itself an extremely powerful tool for industrial quality control and process monitoring.

To this respect, a Near Infrared spectrophotometer was acquired by Ferring's Pilot Plant, in 2013, to be used for the analytical and process control of one of Ferring's bestselling product. This product is available as a solid form.

Historically, HPLC and Potentiometry have been the reference methods for API assay in this product. However, these methods are time-consuming, and involve the use of a large quantity of solvents. The rationale to find an alternative method is therefore to reduce the analytical lead time, reduce costs and environmental fingerprint.

The objective of this work was to develop a fast at-line NIRS method for the assay of API in a finished drug product, for which the drug is formulated as granules.

This thesis presents the basics of Near Infrared Spectroscopy and chemometrics, and its application in developing a validated analytical method to determine API content in a pharmaceutical product

Part 1: Use of Near Infrared Spectroscopy to build analytical methods

1. Near InfraRed Spectroscopy: Principles and Applications

1.1 Electromagnetic spectrum

Spectroscopy consists in studying the interaction of light with matter. The light is a radiation of electromagnetic type, characterized by a wavelength λ ranging from 100 nm to 1 cm.

The electromagnetic spectra is divided into several types of radiations: the most energetic ones being gamma and X-rays, followed by the ultraviolet, the visible and the infrared radiations. Finally, the lowest energetic radiations are the micro and radio waves. These different domains are summarized on Figure 1 below [1].

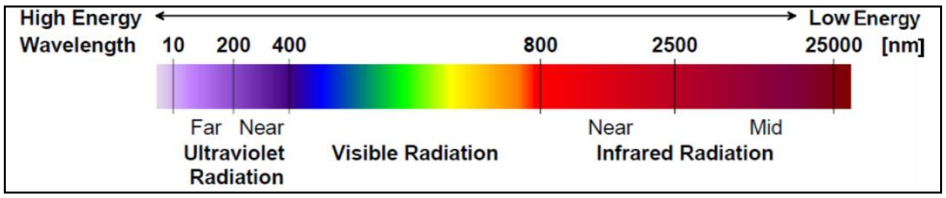


Figure 1 : Light electromagnetic spectrum

The Near Infrared domain ranges from 800 to 2500 nm. Near Infrared and Mid infrared (2500 to 25000 nm) correspond to the vibrational spectroscopy.

Absorbance spectra are represented versus wavelength λ (nm) or wavenumber k (cm^{-1}, with k = 1/λ). Thus, the NIR domain ranges from 800 and 2500 nm, corresponding to 12500-4500 cm^{-1}.

1.2 Interaction of radiation and matter

IR spectroscopy is based on the absorption of light by the substance to be measured. This absorption excites molecular vibrations and rotations, which have frequencies that are the same as those within the infrared range of the electromagnetic spectrum.

A molecule is composed of atoms structured together with covalent bonds, and NIR spectra of a chemical compound can be observed because of the vibrations of these atomic bonds.

Each bond vibrates with its own energy ($E=hv_1$), which depends on the atoms involved in this bond (harmonic oscillations model). These vibrations are the fundamental vibrations v_1. There are other vibrations, called overtones. The frequency of overtone vibrations is a multiple of the fundamental vibration frequency.

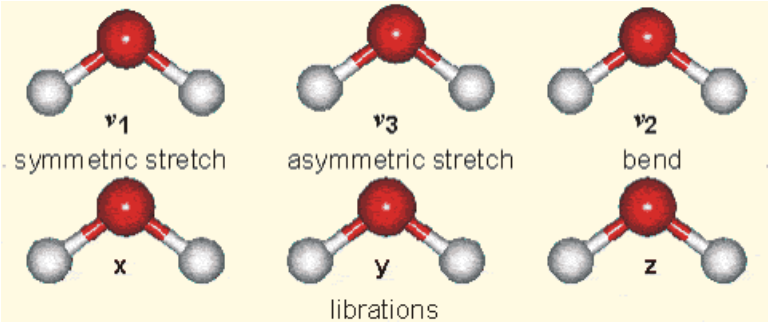


Figure 2: Molecular vibrations [2]

Each bond can absorb energy. Near Infrared spectroscopy measures different types of inter atomic vibrations. These vibrations can be attributed to specific groups (e.g. –O-H, C-H, C=C ...). The final Infrared spectra will contain absorption due to overtones and combination between groups.

1.3 Measurement modes

There are two measurement modes in classical NIR spectrometers: Transmittance and reflectance.

When in transmittance (illustrated on Figure 3), the incident light goes through the sample with an intensity I_0. Then the intensity I_T of the out coming light is collected on the other side and transformed into absorbance unit using Beer Lambert law: $A = \log(\frac{I_0}{I_T})$.

Transmittance is used for liquid samples, with a concentration for which the Beer Lambert law can be valid.

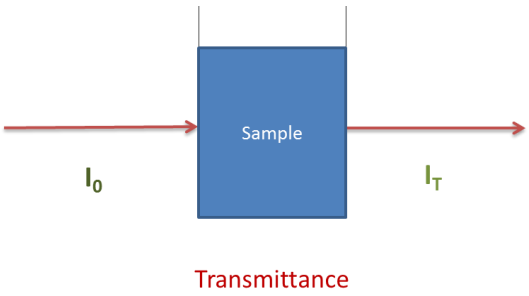


Figure 3: Transmittance NIR analysis

For solid forms or turbid liquids, reflectance mode is used. The intensity collected by the probe results from the light that was reflected and diffused by the sample. The absorbance is computed from the same equation than transmittance.

The intensity of the reflected light depends on:

- The intensity of the incident light
- Absorbance phenomena from the atomic bonds i.e. the chemical properties of the sample
- Scattering phenomena within the sample i.e. the physical properties of the sample (Figure 4)

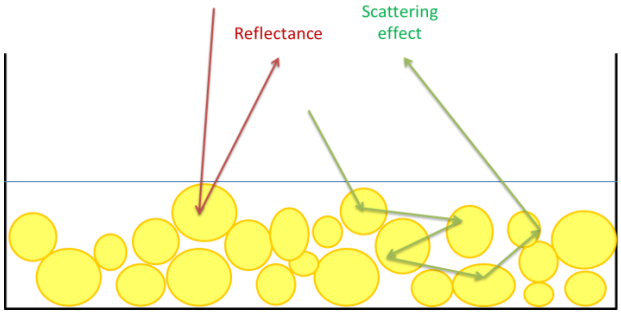


Figure 4 : Incident light trajectory

1.4 Advantages of NIRS

NIRS has become an analytical technique of great interest for the pharmaceutical industry because of the following advantages [3]:

- Short analysis time (< 5min)
- Generally non-destructive
- No (or reduced) sample preparation needed, making possible on/in-line analysis
- Applicable to liquid and solid forms
- Both quantitative and qualitative (identification) with possibly only one calibration model
- Includes physical information about the sample
- Various measurement modes available (reflection, transmission, transflection…)
- Environmentally friendly

1.5 Applications in the Pharmaceutical Industry

NIRS has wide potential uses in Quality Control, such as identification and assay of materials (raw materials, intermediates and finished products), as well as verification of their physicochemical properties. Conventional manufacturing is based on batch processes. Quality attributes of the drug are measured at/off-line, in the Quality Control laboratory, on samples selected at the end of the process. These samples are meant to be representative of the whole batch, and results within the specifications allow the release of the finished product and its access to the market. All these are “at-line” applications, where NIR is useful but disconnected from the process.

Today, with the development of new technologies, opportunities arise to improve pharmaceutical development, manufacturing, and quality assurance through innovation in product and process development, process analysis, and process control [4].

That is why pharmaceutical companies (but not exclusively) are now starting to see farther than “quality by testing”, and adopt “quality by design” approaches. For this latter, quality of the finished product is ensured through in-depth process understanding and control and monitoring. For these purposes, Process Analytical Technology (PAT) tools are essential. Indeed, to understand a process, key sources of process variations have to be identified and closely controlled. Otherwise, the probability to deliver a poor quality product increases. PAT tools are able to monitor those critical steps, and are particularly useful when considering a continuous manufacturing process. They may be used as part of a Real Time Release Testing (RTRT) strategy.

Among them is NIR spectroscopy. One great advantage is the possibility to separate the spectrometer from the point of sampling, thanks to small moveable and wireless probes [5], which makes it very handy to use at any point of a process. The very short analysis time it requires makes NIR a very good option for online analysis as well, along with being non-destructive and non-invasive.

Overall, either online moveable probes or classical at-line set-ups have various applications, and some examples will now be presented.

1.5.1 Chemical and pharmaceutical characterization

First and now probably most common application of NIR in the pharmaceutical industry is the chemical characterization of analysed product, to identify and determine content of an ingredient of interest. Most molecules vibrate when submitted to NIR frequencies.

NIR spectra can be used to identify materials. Spectra of an unknown compound can easily be compared with a library of spectra, and techniques to discriminate between similar products are very sensitive, for example distinguishing between different polymorphs of a same molecule and excipient analogues [6]. Another example will be seen later in this thesis (cf. Part 2 - 4. Comparison between Uncrushed and Crushed samples). Hence, not only NIR can identify a chemical ingredient, but can also give information on its quality. It goes without saying that poor quality of raw materials could have a very negative impact on process performance (bad flowability, reduced solubility, etc.).

In order to quantify an ingredient in a sample, the methodology is similar for any method development. First, a feasibility study is performed to assess the response of the compound when analysed with NIR, and check there is no issue with the physical form of the sample. Then, a calibration takes place, for which different concentrations of the compound of interest

are measured. Once a correlation is found between spectra and concentration, validation samples are prepared to ensure the right predictability of the calibration model [7].

1.5.2 Monitoring and controlling of a crystallisation process

API quality is critical for drug manufacturing process. The crystallisation of APIs process must ensure to deliver the right polymorph, chemical purity and particle size distribution before being used onward. These kinds of processes can be very sensitive to small variations of process parameters, which is why a tight control is necessary. NIR has been used by Janssen to measure the water content of a compound with a precision of +/- 0.02%, corresponding to detecting a variation of 200g water in 1000kg of homogeneous crystallisation mixture [8].

Method development included potential variable parameters in order to be as robust as possible: range of water concentration, temperature, assay, different batches of material.

In a practical way, NIR was used online to measure water content in the crystalliser, and then control the flow of water feeding this crystalliser.

1.5.3 Monitoring of a continuous granulation process

Wet granulation is very common in pharmaceutical manufacturing processes, in order to improve compactibility and flowability of powder mixtures prior to tabletting. Some parameters can have a critical influence on the granulate properties. For example, temperature and moisture content of mixes can have an impact on the solid state of particles and lead to different types of polymorphs which may not be desirable [9]. Indeed, if a change in chemical and/or physical characteristics of a product occurs, their processability (compression etc.), solubility, stability and – last but not least – bioavailability may be affected, leading to a completely different product. First, NIR coupled with Multivariate Data Analysis can be used to better understand the process and determine which parameters have the biggest influence on granulate properties (mixer speed, feed rate of the process, temperature, liquid concentration, equipment design etc.). Second, these information can be used to monitor and control processes, for example adjusting temperature or liquid input during the granulation process.

1.5.4 Monitoring of a tableting process

Tablets account for two thirds of the total pharmaceutical solid forms production, hence tabletting processes are very common. NIR has been used to monitor different parameters related to process and product quality during a compression operation: content uniformity, compression force, tensile strength, moisture and mean particle size [10]. For these applications, NIR can be used because physical properties of the product are reflected in the spectra. For example, the scattering effect is impacted by the density of the analyte. As was illustrated on Figure 4, the light trajectory will be impacted by the configuration of the particles in the tablet. This made possible the development of models able to predict the compression

force that was applied on a tablet, or measure the tensile strength (both being correlated) of the finished product. Same principle is applied for mean particle size determination.

As far as chemical properties are concerned, such as API and moisture content, multivariate data analysis coupled with NIR becomes particularly interesting to take into account the above mentioned physical properties of blends or tablets. It then becomes easier to understand how each variable may influence the others, and develop a reliable measurement model.

2. Spectral data: Acquisition & Analysis

2.1 Spectral measures

For this project, the spectral measures were performed on a Fourier transform spectrometer MATRIX F Duplex from BRUKER OPTICS. The spectra are obtained by reflectance with an at-line probe Q412/A-Ex, BRUKER, connected to the spectrometer by an optic fibre.

The acquisition of absorbance spectra is achieved with the OPUS 7.0 software, provided with the equipment by BRUKER. The following measurement settings have been used for spectral acquisition:

- Spectral range: 12489.2 – 3995.9 cm^{-1}.
- Resolution: 16 cm^{-1}

Each sample is measured six times and an average spectrum is calculated using OPUS. This average spectrum is used as reference spectrum for each sample.

The parameters presented above are derived from previous feasibility studies performed in Ferring lab, for which a Design of Experiments (DoE) was performed to assess the impact of each measurement parameter on the collected spectra.

2.2 Data preprocessing

Data preprocessing is an essential step of the calibration. Indeed, the raw spectra are usually noisy and combine both physical and chemical information from the sample. For instance, random variations - such as offsets or differences in linear baselines - appear between the spectra. These variations may be due, for example, to differences in the sample preparation, sample thickness, particles size, noise due to the measure, etc. The objective of data preprocessing is to increase as much as possible the correlation between the spectral data and the parameter of interest to be predicted, here the content in API.

Data preprocessing includes the choice of spectral filters, also known as pretreatments, and spectral range selection.

Pretreatments are mathematical transformations applied to the spectra in order to eliminate or minimize variability unrelated to the property of interest. Thus, they are used to attenuate the noise, the influence of physical properties or emphasize some variations for example. The following pretreatments were tested during the development [11] :

- SNV (Standard Normal Variate), which normalizes a spectrum by first calculating the average intensity value and subsequent subtraction of this value from the spectrum. Then the sum of the squared intensities is calculated and the spectrum is divided by the square root of this sum. This method is used to account for different samples thickness, or to minimize the influence of particles size for example

- D1 (first derivative), is the slope at each point of the original spectrum. It peaks where the original spectrum has maximum slope and it crosses zero where the original has peaks. This method emphasizes steep edges of a peak over relatively flat bands. Hence it is used to emphasize pronounced, but small features over a broad background.
- D2, has the same purpose than D1, but with a more drastic effect. It emphasizes more every small variation of the spectra.
- MSC (Multiplicative Scatter Correction), performs a linear transformation of each spectrum so that it best matches the average spectrum of the whole set [7].

Combinations of pretreatments are also possible (e.g. D1+SNV) and frequent for quantitative models. Indeed, they correct the scattering effects of the samples, that is to say decrease the noise and remove the offsets due to physical properties of the samples [12].

Spectral range selection involves selecting a subset of spectral regions from which the PLS model provides the smallest error in prediction. Adequate selection enhances the robustness and performance of the model.

2.3 Multivariate data analysis

Given the consequent amount and the complexity of the data to be analysed, statistical tools such as MultiVariate Data Analysis (MVDA) have to be used. Indeed, the dataset obtained from the spectral measures is as follows:

N Samples	X Variables				Y Responses (Content)
	λ1	λ2	λ...	λk	
Spectra 1	x1	x2		xk	y1 %
Spectra 2	Absorbance values				y2 %
Spectra ...					...
Spectra n	x1	x2		xk	yn %

Table 1 : Spectral dataset

To the n samples analysed (e.g. Calibration sets, with known Content y1 to yn) correspond as many absorbance values as wavenumbers. These wavenumbers and the corresponding absorbance values are called Variables and referred to as X.

Principal Component Analysis (PCA) and Projection on Latent Structures (PLS) are the two methods used to analyse the dataset.

As all analytical methods, X can be defined as:

X = information + structured noise (instrumental, temperature, humidity, etc.) + random noise

PCA allows to get a simple overview of the information present in the dataset, displaying the systematic variation within the X matrix. It explores relationships both among the observations and within the variables.

This analysis results in a summary showing how the observations are related, as well as possible deviating observations or cluster of observations. To exemplify, PCA could differentiate the samples according to their different batch numbers, or excipient types, etc.

PLS is similar to PCA, but is used to perform a regression modelling between different blocks of variables, in this case the wavenumbers X and the response Y (content in API). The aim of PLS is to give a model able to predict Y from X for future observations.

MVDA also has the advantage of providing the user with graphical results, making it really visual and easy to manipulate, interpret and present.

To summarize, using PCA and PLS on a large and complex dataset, one is able to:

- Obtain an overview of data
- Classify and discriminate the observations and the variables

} PCA → Qualitative analysis

- Perform a regression modelling between variables : PLS → Quantitative analysis

From a mathematical point of view, MVDA propose projection-based methods. A K-space is constructed by attributing to the K X-variables one orthogonal axis in a co-ordinate system. Hence, each row of the dataset can now be positioned as a point in this K-dimensional space, resulting in swarms of points. Figure 5 below is an example of the K-dimensional space construction, where K=3 to make the dimensional space easier to visualize.

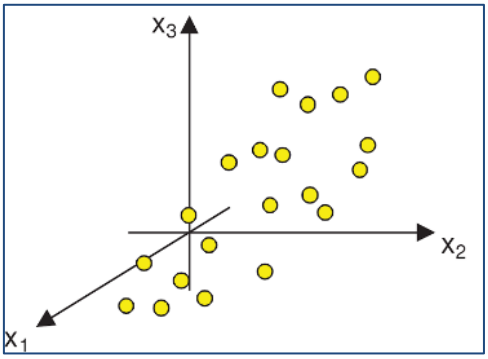


Figure 5 : Construction of a K-dimensional space (K=3) [13]

The concept is to simplify this representation into a lower-dimensional space - usually from 2 to 5 dimensions - in order to obtain an overview of the spread of the points. Thanks to the graphical overview, one will be able to detect clusters of observations, trends or outliers, as similar samples will be close to each other.

Statistically, the method finds lines, planes and hyperplanes in the K-dimensional space that approximate the data as best as possible in the least square sense. Each new axis is called

principal component, and the new co-ordinates in the principal components space are defined as *scores*. The centre point of this new space is the average observation in the dataset.

Principal Components Analysis:

The construction of the components is sequential and orthogonal. The first component (PC1) is the line which best fits the shape of the points swarm, and accounts for the maximum variance in the data (Figure 6). Each observation is projected onto this line to get its new co-ordinate in the PC-dimensional plane: a score.

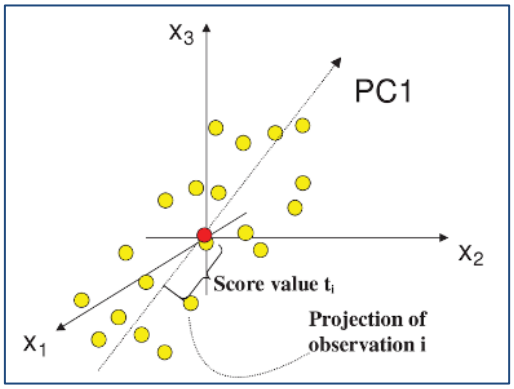


Figure 6 : Construction of PC1 [13]

If one component is not enough to explain the systematic variability in the dataset, a second principal component is computed. It is another line, orthogonal to PC1, reflecting the second largest source of variation in the data (Figure 7).

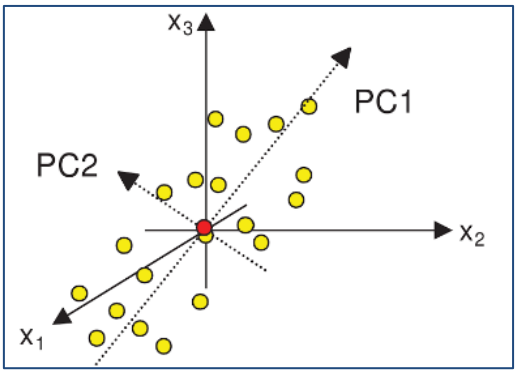


Figure 7 : Construction of PC2 [13]

PC1 and PC2 result in a plane within the K-dimensional space. By projecting the observations on this plane, a score plot is obtained (Figure 8), allowing to visualize the structure of the dataset, if there is any defined pattern for example.

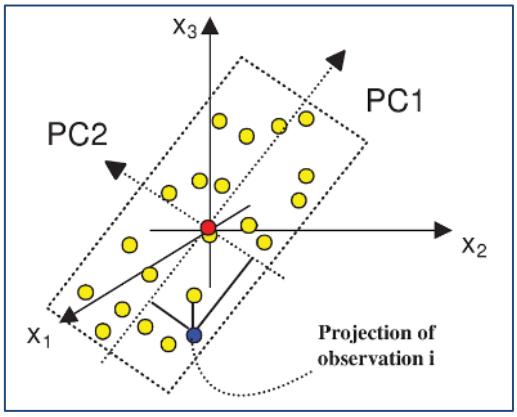


Figure 8 : Score plot construction [13]

Projection on latent squares:

As in PCA, each observation can be plotted. However, the major difference in PLS is that each row in the data table corresponds to two graphical points, one in the X-space and one in the Y-space, resulting this time in two swarms of points, as illustrated on Figure 9.

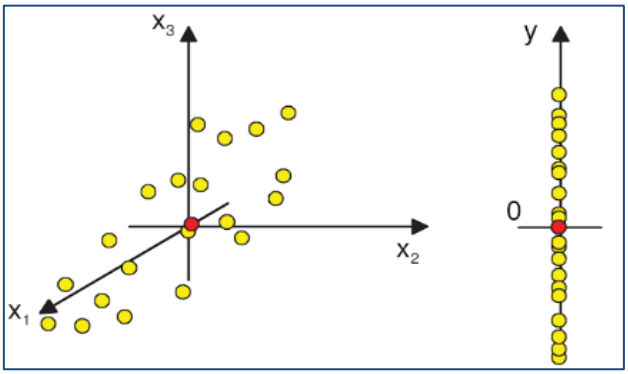


Figure 9 : PLS - X and Y-spaces

The aim of PLS is to describe the relationship between the positions of the observations in both spaces, the predictor one (X) and the response one (Y). To do this, principal components are used again.

The first component is a line in the X-space which best fits the swarm of points and provides a good correlation with the y-vector.

As in PCA, by projecting the observation i on PC1, its new co-ordinate is obtained and called score t_{i1}. All the scores t_{i1} form the first X-score vector $\mathbf{t_1}$. This is illustrated on the left-hand part of Figure 10.

The vector $\mathbf{t_1}$ is used to estimate y, referred to as $\widehat{y_{(1)}}$. As seen on the right-hand part of Figure 14, one component is not enough in this case to predict the y responses accurately. The difference between real and predicted responses is called residuals.

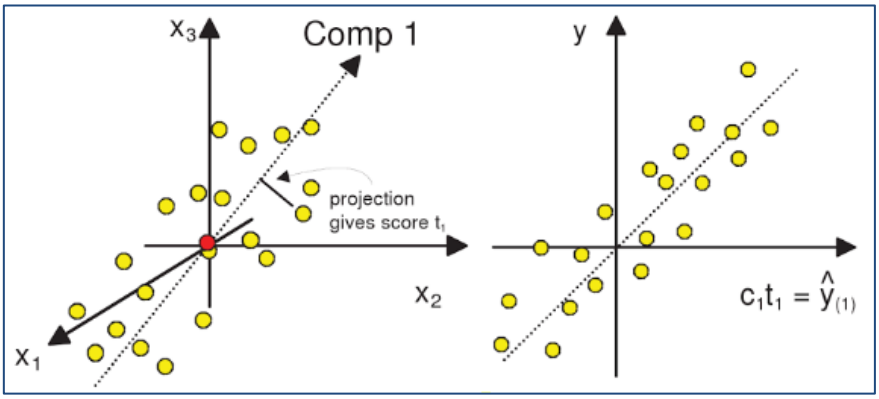


Figure 10 : PLS - First component construction

The residuals can also be represented as a vector $\mathbf{f_1}$, equal to $y - \widehat{y_{(1)}}$. As illustrated on Figure 11 below, this residual vector obtained after computing the first component $\mathbf{t_1}$ is much shorter than the vector representing the measured data. This means that PC1 accounts for a large part of the variation in **y**.

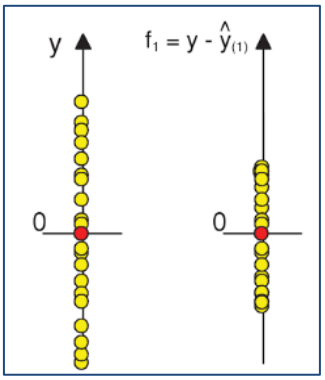


Figure 11 : Measured vector y vs. Residual vector f_1

Similarly to PCA, a second component may be calculated, in order to estimate as well as possible the residuals $\mathbf{f_1}$ from the remaining variation in the X-space (Figure 12). This second component still is a line orthogonal to PC1 and passing through the average observation of the dataset. The second score vector is called $\mathbf{t_2}$.

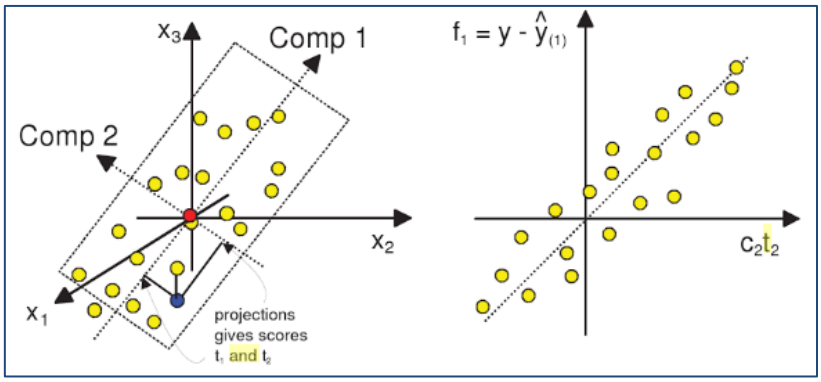


Figure 12 : Second component construction

The model performance is evaluated thanks to the right-hand part of Figure 12. Once more, the tighter the scatter around the diagonal, the stronger the correlation between **X** and **y** in the second PLS dimension.

Now, the two components can be combined to assess the predictive ability of the PLS model. The combination leads to a better estimation of **y** than the two components separately (Figure 13). Indeed, it combines the information of PC1 and PC2 in the X-space.

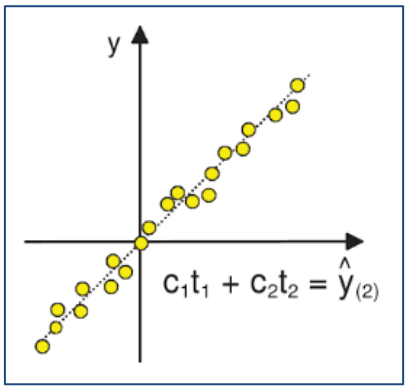


Figure 13 : Estimation of y with 2 Principal Components

To summarize this part, in order to create a good predictive model, accurate and robust, the following parameters need to be optimized:

- Pretreatments used, in order to reduce the effect of noise and remove physical information that would not be relevant for the purposes of calibration
- Absorbance bands selected, so that the most relevant bands remain in the calibration, typically those which are the most specific of the analyte
- Number of components used, enough components must be included in the model in order to predict the variable of interest, but including too many would lead to overfitting and lack of model robustness.

In the scope of this development, the MVDA study was performed as follows. According to the optimization tool “Quant 2” in OPUS software, a list of the best pretreatments and spectral range is given for the selected data, along with recommendations for the number of components to include.

However, the OPUS proposed models cannot be used as such, because the data are usually over fitted with a too high number of components. In OPUS, precision prevails over robustness. Therefore, to obtain a robust model it is important not to use the recommendations regarding the number of components. Hence to adjust the model, the OPUS recommendations are tested and adjusted with the software SIMCA, which allows more flexibility in the data manipulation.

Therefore MVDA was realized with OPUS software for a general approach, then SIMCA 13 was used for in-depth analysis and optimization of the model.

Part 2: Development of an at-line analytical method for API quantification in granules

1. Objective

The objective of this work was to develop and validate a fast analytical method to determine the API content in one of Ferring main solid product. This product is formulated as granules with a slow release coating. It is available on the market at 3 different doses: 1, 2 and 4g. The granules are packaged into Sachets, filled to achieve the right dose.

The analytical method developed is based on Near Infra-Red Spectroscopy (NIRS). Currently, the method used for release purposes is HPLC. Hence the objective is to provide a faster and at least as reliable method for the at-line assay of API in the finish product, as an alternative to HPLC.

2. Model construction

2.1 Overview of Model Development

All NIRS method developments follow the same scheme, represented on Figure 14 below.

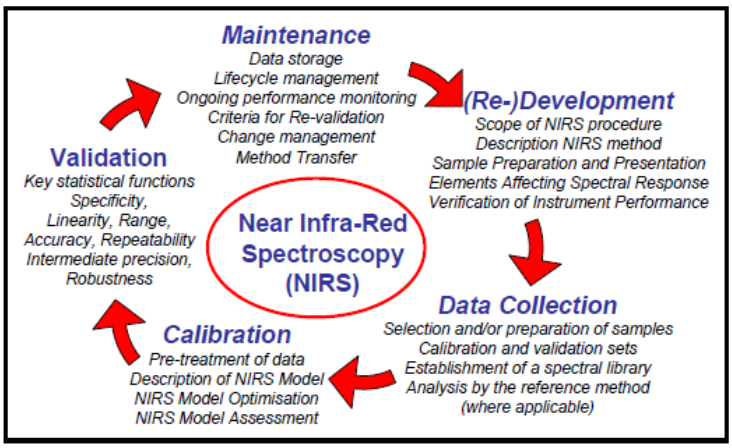


Figure 14 : NIRS method development steps [7]

The objective of the project was to develop a model that will allow the measurement of API content for Sachets commercial form. To do so, a calibration step has to be performed. The range of the calibration was set up in this particular case at 80 to 105% label claim (see 2.3.1 for further details).

To analyse the data from the calibration, MVDA was used. Thus, a PLS model was created using 15 calibration samples (see 2.3 for details on PLS). Once a satisfying calibration model

is obtained, an external validation has to be performed with 12 validation samples, in order to assess the model prediction power.

This external validation delivers two key parameters:

- RMSEP value, which represents the difference between the real values and the predicted values. RMSEP corresponds to a standard deviation and provides an estimation of the precision of the model.

- The regression equation between the reference and the predicted values. This equation $Y_{Pred} = a*Y_{Ref} + b$, provides an indication of the accuracy of the model. An ideal model has no systematic bias (b=0) nor relative bias (a=1).

In a second step, the 27 samples were pooled together to build the final model. The final model precision is assessed using the “Leave-p-out cross validation” method, providing a value of RMSECV. RMSECV is also considered as a standard deviation and provides an estimated method precision.

This final model was used to analyse Sachets of 1g and 2g dose. Given that the coated granules come from the same manufacturing process, regardless of the quantity to be filled during packaging (1g, 2g or 4g), there is no difference related to the sachets dosage.

These Sachets were also analysed using the reference HPLC method. Results from both methods were compared.

Once a satisfactory model was defined, analytical validation took place. The model was thus tested for linearity, specificity, accuracy, precision and robustness.

2.2 Preliminary studies

2.2.1 Comparison of samples physical properties

2.2.1.1 Objective

The objective of this study was to assess the feasibility of developing the model based on powder mixes, in order to reconstitute finish products with the different contents necessary for the calibration model.

An underlying objective was to assess the feasibility of performing the future content analysis on the coated granules directly i.e. without sample preparation. The point was to skip the step consisting in crushing the granules, which is currently performed for the HPLC analysis and the NIRS methods for suppositories and tablets content dosage.

2.2.1.2 Method

Spectra of non-crushed coated granules were taken. Then the granules were crushed and a new measurement was performed. Furthermore, spectra of powders blends corresponding to 100% label claim were taken.

All these spectra were analysed with MVDA methods to determine if the three classes could be differentiated.

2.2.1.3 Results and discussion

Spectra Analysis

The spectra obtained are presented on Figure 15 below:

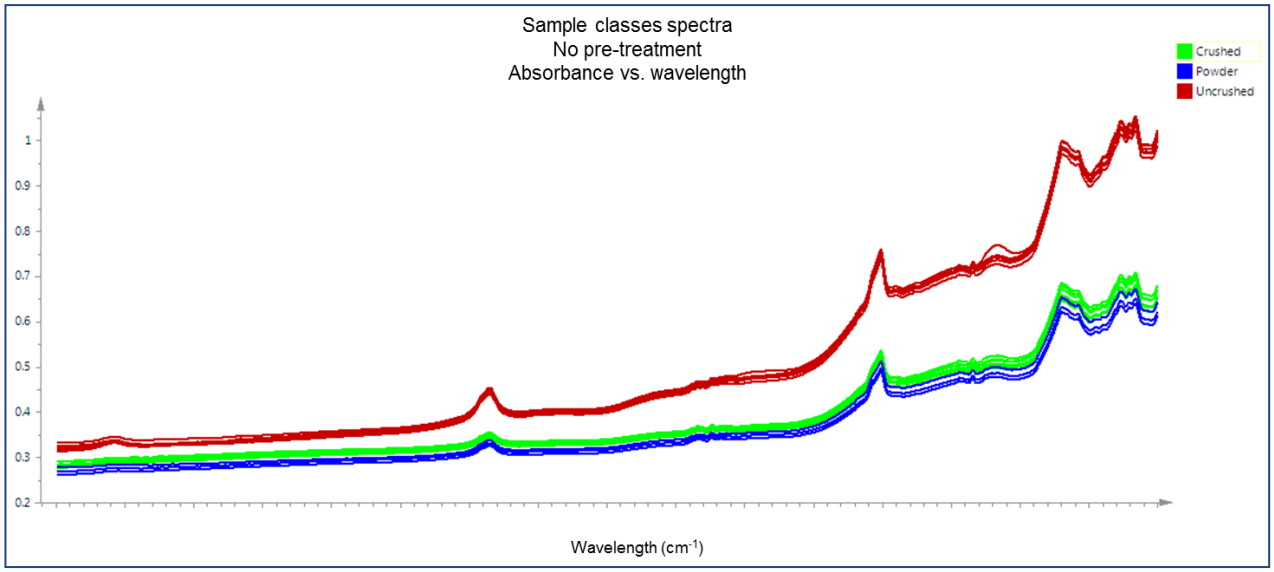


Figure 15 : Sample classes spectra

NIRS Absorbance measurements are influenced by the physical properties of the sample (ex: the particle size distribution in the sample, the smoothness of the powder surface …). These physical properties induce different light scattering effects leading to different spectra. On the figure above (Figure 15), it can be seen that the spectra recorded for uncrushed material are very different from the spectra for powder or crushed materials, as far as absorbance is concerned. This result was expected since the particle size of uncrushed material is much larger than the powder mixes or the crushed material. It can also be seen that powder mix and crushed material yield to differentiable spectra, but in this case the difference is much smaller given that the particles sizes are closer.

To reduce the impact of the physical properties effect and to focus the analysis only on the chemical information (i.e. the API content in this case), the spectra can be mathematically pre-treated using filters. Therefore, the different slopes and offset observed on Figure 15 can be corrected (Figure 16).

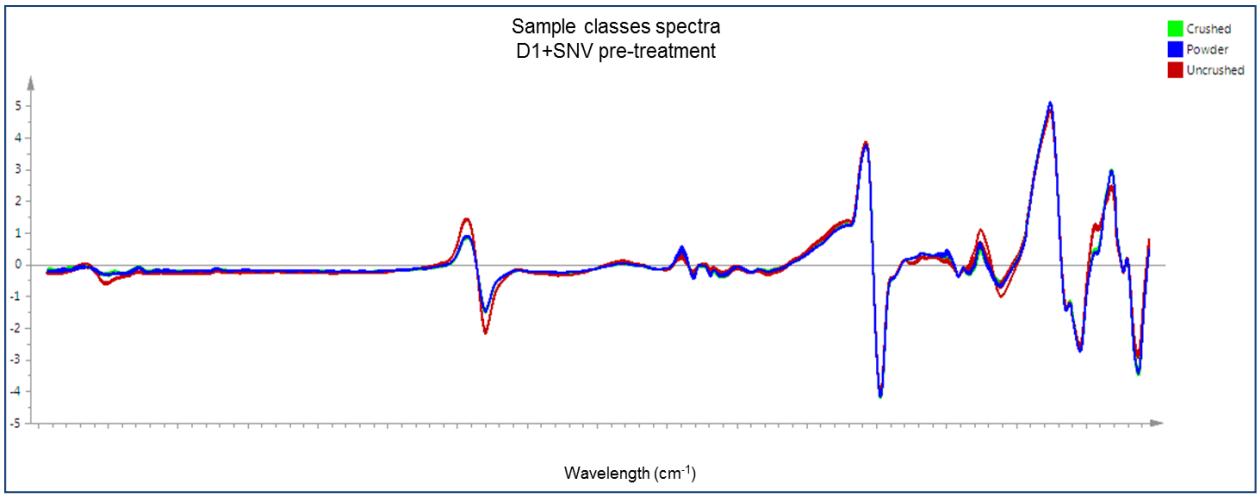


Figure 16 : Sample classes spectra - Pre-treated spectra (D1 + SNV)

On this Figure 16, where the original spectra presented on Figure 15 have been processed by applying normalization (SNV: Standard Normalization Variate) and first derivative (17 smoothing points), it can be concluded that the differences between the spectra have been reduced. Only the uncrushed material can still be visually differentiated from the powder and crushed, these two being overlapping. This indicates that the API content might be determined directly on the uncrushed granules. But to confirm this, in depth analysis using MVDA tools has to be performed.

MVDA Analysis

To analyse the information contained in all the spectra, a Principal Component Analysis was performed using SIMCA software.

The PCA is performed with setting in the software the three different classes, to see how the variations into the spectral dataset may explain the differences between these classes.

The best results are obtained with the combination SNV + D1 pre-treatments. On the first score plot below (Figure 17), it appears that the first component, accounting for 99.7% of the dataset variability, is able to explain the physical difference between the uncrushed granules on one hand, and the crushed granules and reconstituted 100% label claim blends on the other hand. The second component is differentiating the crushed granules and the reconstituted 100% label claim sets, but accounts for only 0.3% of the dataset variability.

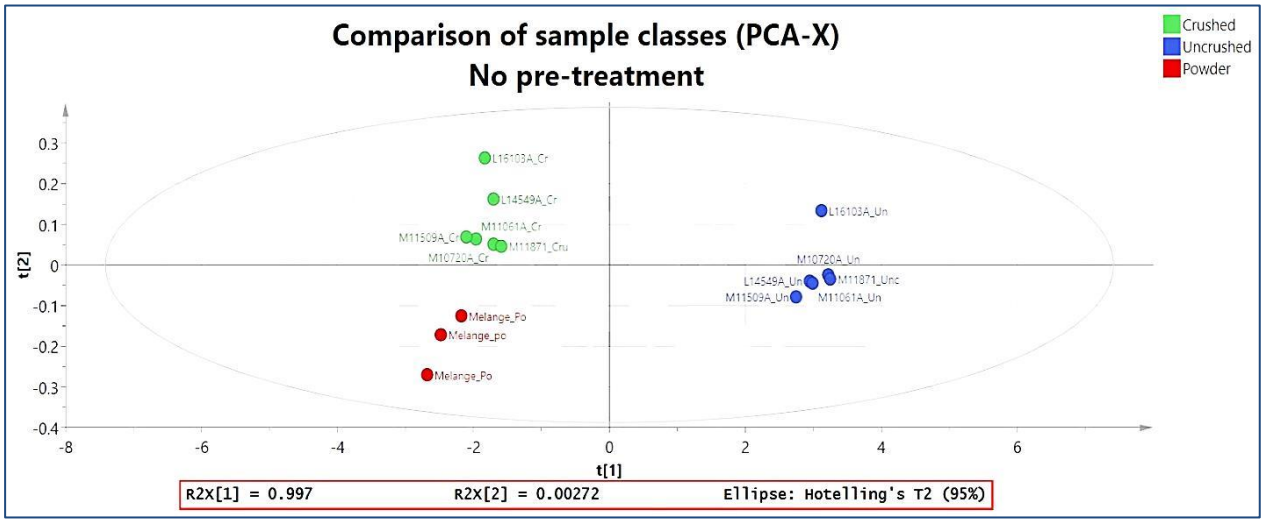


Figure 17 : Comparison of sample classes - PCA without pre-treatment

When a SNV+D1 pre-treatment is applied, the different classes are even harder to differentiate, and the crushed granules are not distinguished from the reconstituted 100% label claim sets anymore (Figure 18).

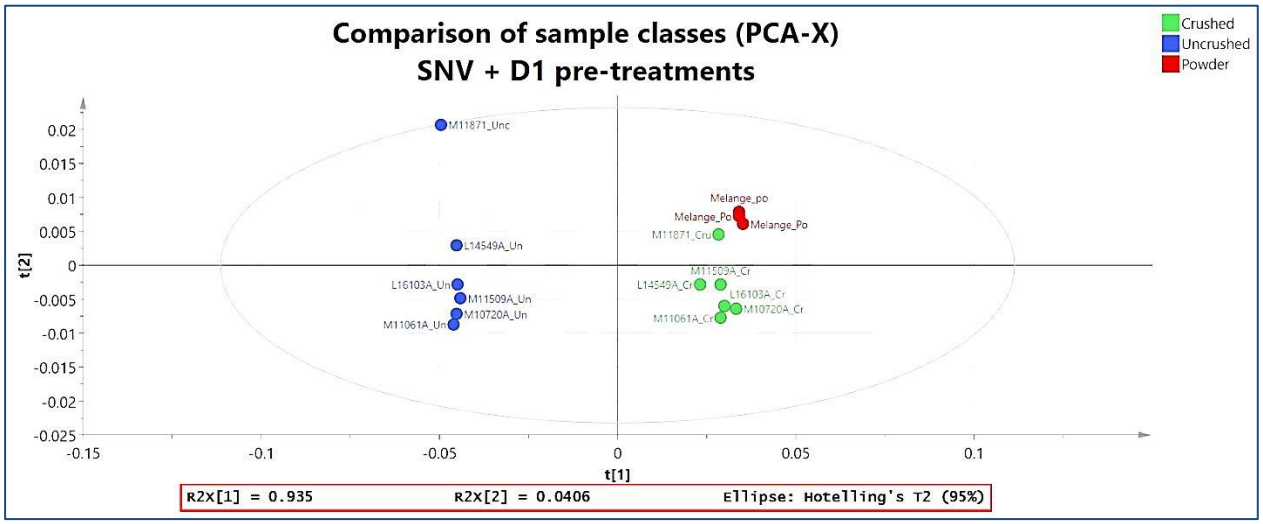


Figure 18 : Comparison of sample classes – Samples pretreated with SNV+D1

The first conclusion from these results is that it is feasible to develop the calibration model from powder reconstituted mixes, given the similarity between the crushed granules and reconstituted 100% label claim sets spectra.

Second, there is an offset difference between the uncrushed and crushed granules absorbance. The difference remains on some wavelength bands despite the pre-treatments applied. This will have to be taken into account for the further development, to determine if it is possible to reduce the samples physical differences without affecting the content measurement.

2.2.2 Raw Materials study

2.2.2.1 Objective

The objective of this study was to assess the impact of the excipients on the API content analysis of the drug product.

2.2.2.2 Method

The spectra of each ingredient was individually taken. Different batches were included for each one. The powder is put directly into the plate and the measurement is performed.

2.2.2.3 Results and discussion

The graph below (Figure 19) represents the spectra obtained for each ingredient (API, Excipient 1, Excipient 2). This way, it becomes clear on which bands the excipients are expected to impact on the absorbance of the mixed samples.

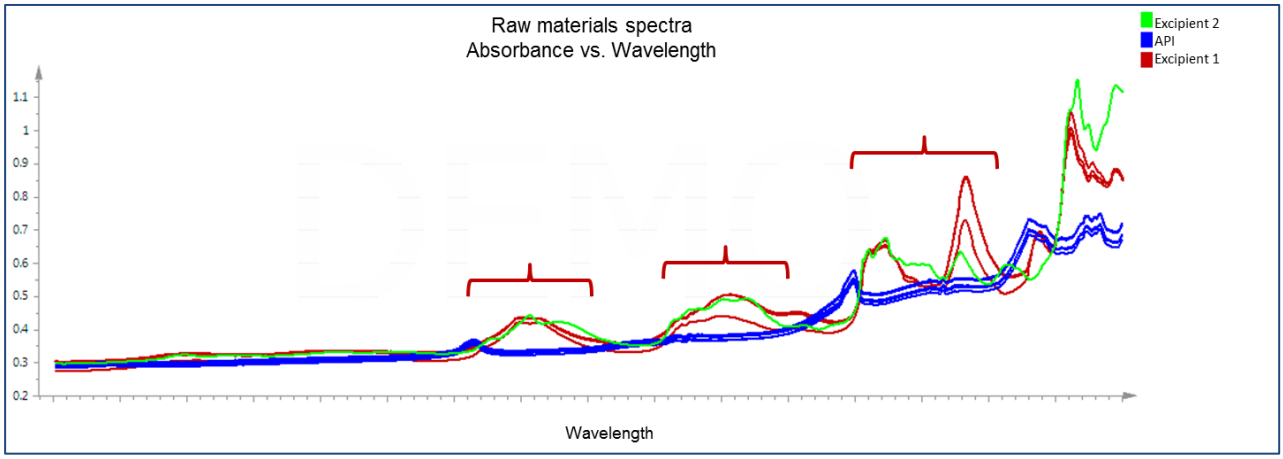


Figure 19 : Spectra of raw materials

The Excipient 1 peaks amplitude is slightly different. It appears that the wavelengths impacted correspond to the water absorbance bands (7400-6200 and 5800-5500 cm^{-1}). Hence the Excipient 1 is known to have a hygroscopic behaviour. This will be taken into account during the model development, to ensure the robustness to humidity variation in the samples.

2.3 Calibration

2.3.1 Calibration and validation ranges

The calibration range was set from 80 to 105% label claim, knowing that the maximum percentage that may be reached in the Sachets is theoretically 106%. This would mean the

Sachet is entirely filled with pure API, which is unrealistic. In terms of mass proportion, the granules are actually dosed at 95% API.

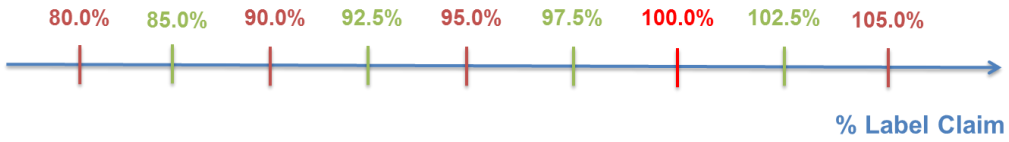
The 100% label claim for the 1g dose, from which the calibration and validation samples composition were calculated, is presented in Table 2 below.

Ingredient	Composition Sachet 95% 1g	% w/w
API (mg)	**1000.00**	**94.42%**
Excipients (mg)	**59.10**	5.58%
Total sachet mass (mg)	**1059.10**	100.00%

Table 2 : Composition of Drug Product Sachet 1g

The target amount of one individual sample is 12g equivalent API – which corresponds to 12 sticks 1 g, 6 sticks 2g, 3 sticks 4g. This leads to a total sample mass of 12.7092g.

The following points were chosen for calibration (in red) and validation sets (in green):



Each ingredient was weighed individually, then the powders were homogeneously mixed using a laboratory blade crusher.

The table in Annexe 1 summarizes the real API content of each sample.

2.3.2 Spectra analysis

The raw spectra of the calibration and validation sets are presented in

Figure 20 (at-line measurement of the powder mix). The spectra are coloured according to the corresponding theoretical API content (in percentage of label claim) of the samples (from 80% in blue to 105% in red).

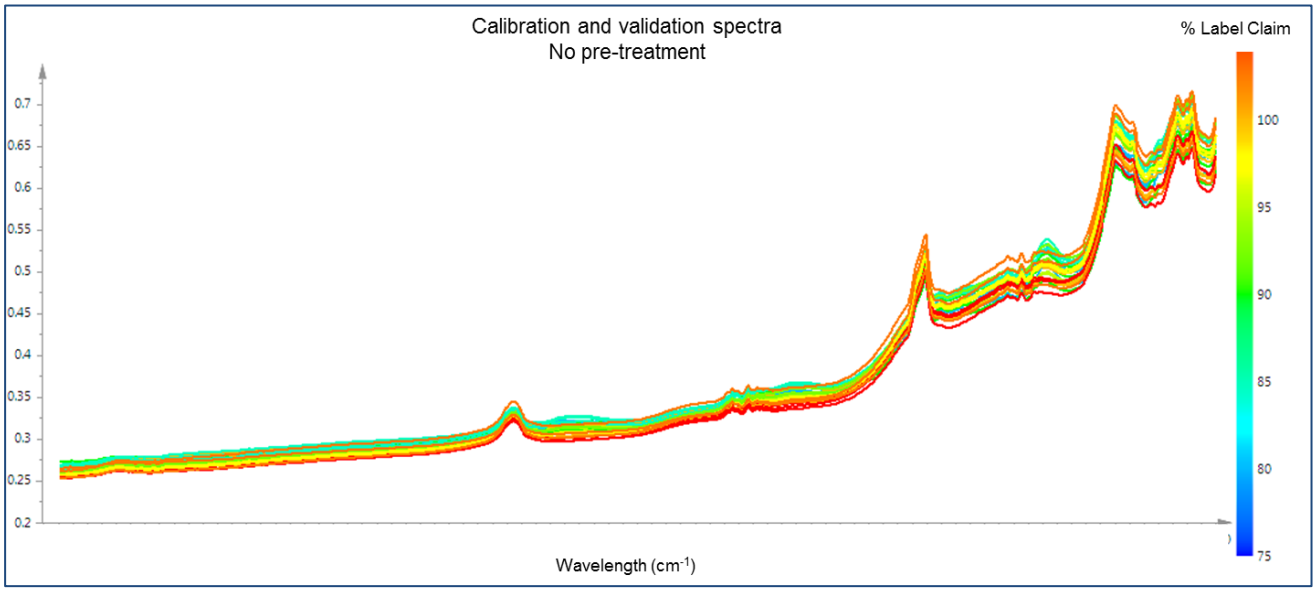


Figure 20 : Full spectra (no-pre-treatment) of the calibration and validation sets

As seen on this figure, there is no pattern following the content in API, the signal is noisy, which is expected as far as raw data are concerned. That is why a preprocessing step of the data is mandatory, to be able to focus the information on the chemical properties of the sample.

During this development phase on SIMCA, the following pre-treatments were also assessed: no pre-treatment, SNV, MSC, D1+SNV, D1+MSC and D2. These pre-treatments have given worse results compared to D1 + SNV + MSC, so they will not be presented in this report.

Figure 21 shows the spectra pre-treated by SNV on the whole wavelength range. This shortening was performed to specifically exclude the noisy extremes of the spectra. The specific absorbance bands corresponding to API, Excipient 1 and Excipient 2 can be easily noticed, as the order of colour shades follows the API content. Hence these specific absorbance bands are particularly interesting to build the model.

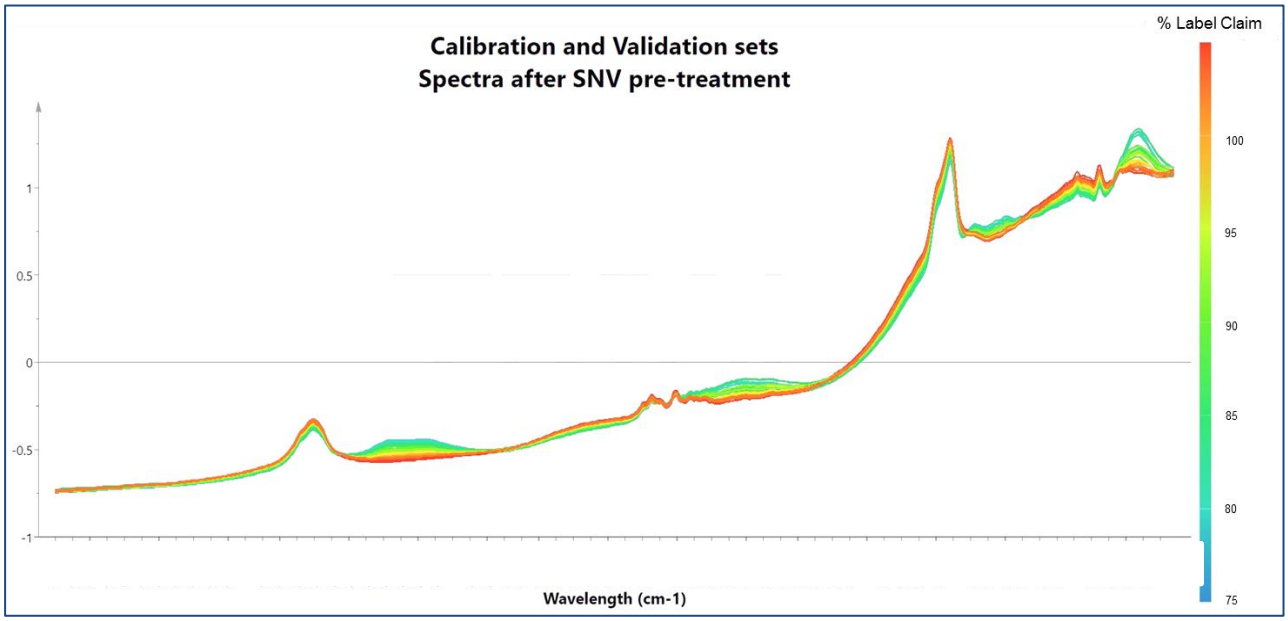


Figure 21 : Calibration and validation spectra sets – Cut and pre-treated by SNV

Figure 22 represents the same spectra as Figure 21 but first pre-treated with a first derivative pretreatment.

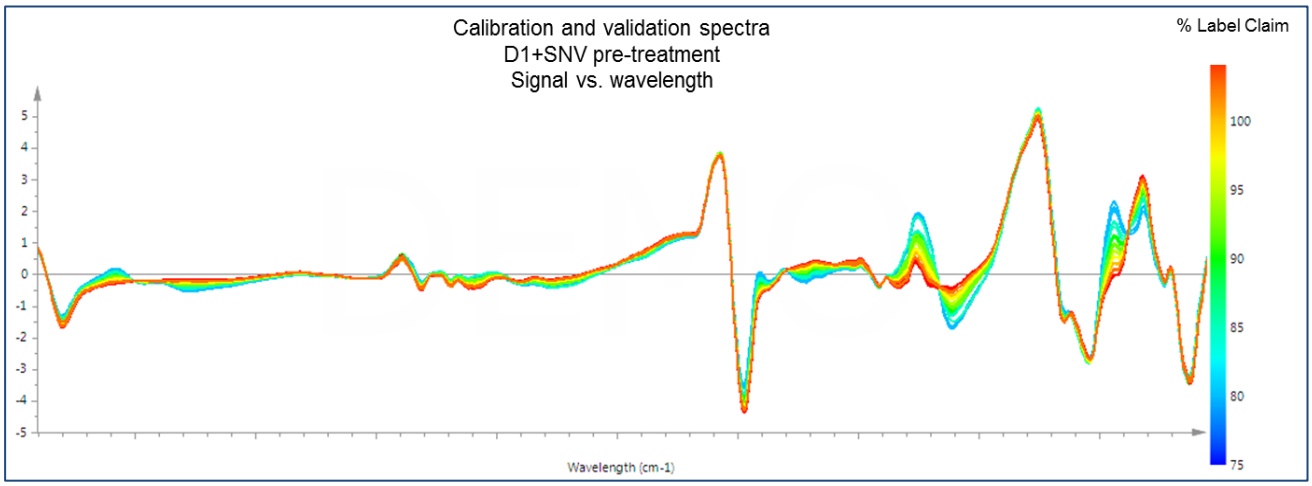


Figure 22 : Calibration and validation spectra sets
Cut and pre-treated by D1+SNV

To conclude this part, it is possible, thanks to the mathematical pretreatments available, to organize the spectra according to their API content: The pretreated data are then used with the MVDA analysis to build the model.

2.3.3 Principal Components Analysis

For explorative purposes, a PCA was performed to detect any trend in the dataset due to the variations of the samples.

The score plot on Figure 23 shows the PCA of the 27 spectra filtered by D1 + SNV on the whole wavelength range in order to observe the chemical information.

The dots are coloured according to their content in percentage of label claim (upper graph of Figure 23), and API type.

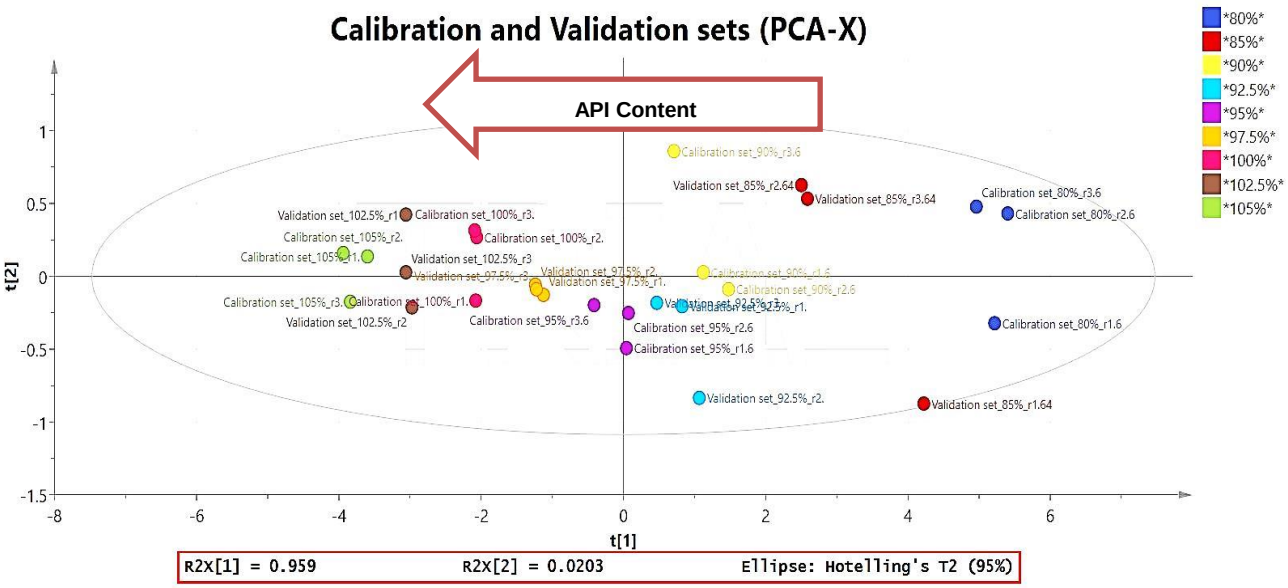


Figure 23 : PCA score plots of all spectra filtered by D1+SNV

On the upper graph of Figure 23, it is observed that the API content follows the PC1 axis, meaning that the highest difference between all the spectra is due to the variation of content (R2X[1] = 95.3%).

The PC2, explaining 2.61% of the variation in the variables dataset, does not seem to be correlated to any of the variation that was intentionally included in the model construction, as the samples are randomly spread along the axis.

One outlier is observed outside of the ellipse corresponding to 95% of the confidence interval (Validation set – 85% - 1st repetition), but is of no significant impact on the model because its leverage on the power is not significantly high (see 2.4.3).

To conclude this part, the PCA confirmed that it should be feasible to build a model to predict the API content based on the NIR spectra.

2.3.4 Final preprocessing parameters

In order to determine the relationship between spectra and the API content, the Partial Least Square projection on latent structure (PLS) is now used.

The research of the best model parameters in terms of pretreatments, spectral range selection, etc. follows an iterative process. At first, a satisfactory model was found combining D1+SNV pretreatments on a large spectral band. However, when measuring the content of intact granules, a systematic underestimation of about 4% was observed, meaning that this model still included a significant amount of physical information.

Moreover, the removing of the spectral bands related to water absorbance was resulting in a relative loss of precision of the model. Hence, other optimizations had to be performed in order to meet all the project criteria (analysis on Uncrushed granules, robustness, etc) and led to more appropriate parameters, which are going to be presented in this part.

The OPUS optimization tool showed that the combination of D1 + SNV on the whole spectra, in combination with MSC (Multiple Scattering Correction) on specific calibration bands provided particularly high precision (i.e low RMSEP values) in terms of API content prediction, with only one component. Consequently, this combination of pre-treatments were finally assessed in SIMCA software.

Thus, the final preprocessing parameters are as follows:

The calibration and validation sets were pretreated on the whole wavelength range by first derivative (D1) followed by standard normal variate correction (SNV). Following these pretreatments, the spectra were pretreated with MSC on the specific calibration bands selected for the model development. These bands are selected so that to correspond to API and excipients major absorption bands. Furthermore, in this range, the main absorbance bands of the water are partially excluded to give a robust model against water content variation.

The PLS was then performed with the following settings:

- Spectral calibration range: as defined
- Variables scaling: centering (standard setting for NIRS model)
- Number of Cross Validation-groups (cf. 2.3.5 for details): 7 (default setting)

The final spectra used for the model is represented on Figure 24.

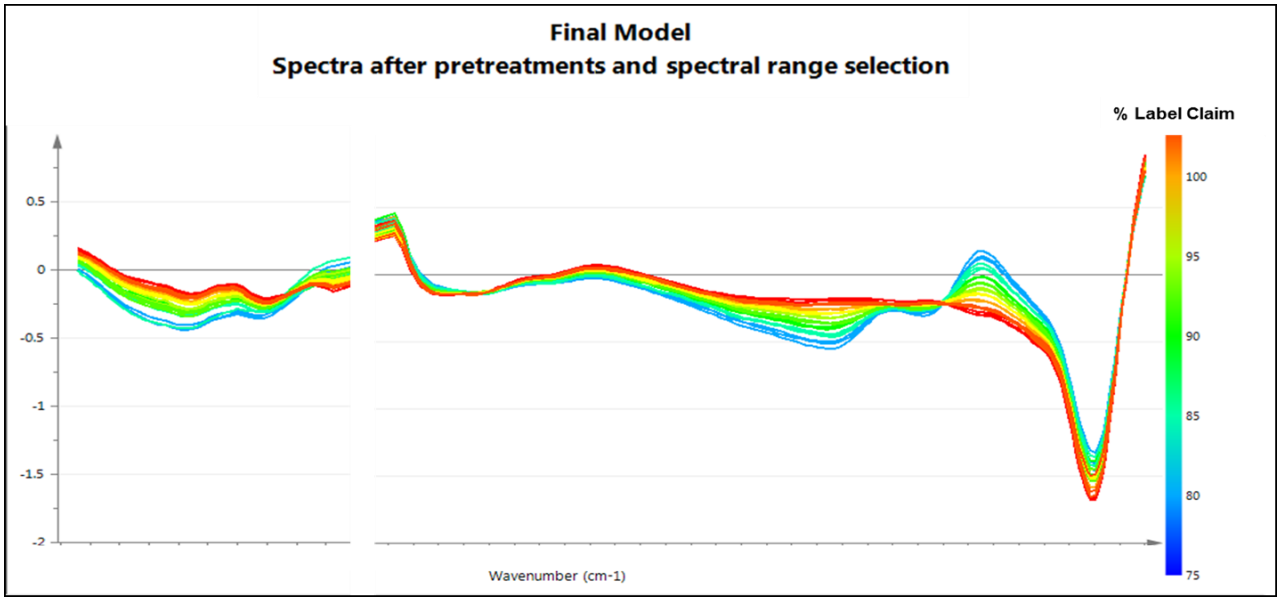


Figure 24 : Final Model - Spectra after pre-treatments and bands selection

As can be seen on this figure, where the spectra are coloured according to %label claim, the information is mainly focused on the API content, as the colour shades follow the API content along the whole selected bands.

2.3.5 PLS on Calibration sets

A first calibration model is now constructed with the 15 calibration sets (80, 90, 95, 100, and 105%) using PLS, then its predictive ability is assessed thanks to the validation samples.

Figure 25 illustrates the R^2Y and Q^2 values of the model (cumulated), with one, two and three components (Comp[x]).

R2Y accounts for the fraction of all the Y's explained by the corresponding component in the least square sense, and can be interpreted as the goodness of fit of the model. R2Y can be interpreted in the same way as the coefficient of determination R^2 for a typical linear regression.

Q2 is the fraction of the total variation of the Y's that can be predicted by a component according to cross-validation, and can be interpreted as the goodness of prediction of the model. Q2 can be interpreted in the same way as a predicted R^2 for a typical linear regression.

Increasing the number of components from one to two does not significantly improve the model's predictive ability (Q^2 for components 1 and 2 are almost equal). A decrease in Q2 means that the analysis has started fitting noise in the model, hence the corresponding component is not improving the predictive ability in terms of chemical information. This case is not desirable.

Moreover, the fewer component the model is based on, the more robust it is. Thus, a one-component model was selected for further development.

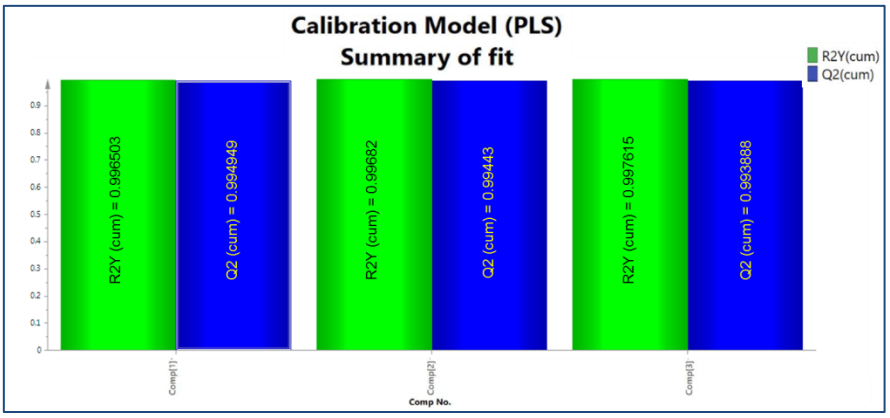


Figure 25 : Summary of fit – Calibration Model

Figure 26 plots the predicted values obtained with NIRS model against the observed values determined by weighing. The NIR model used to predict the content has a precision of 0.61 % of label claim (RMSECV). The RMSECV statistical test consists in performing a calibration leaving a set number of samples out, and predicting these samples with the new submodel

developed. The algorithm performs this step by step until all defined set of samples have been left out and predicted. In this case, at each iteration, 7 samples are taken out of the calibration model for cross-validation. The distance between the real and the predicted value is quantified and results in the Root Mean Square Error of Cross Validation (see formula in the Abbreviations and Definitions), which is the first indicator of the predictive ability of the model [7].

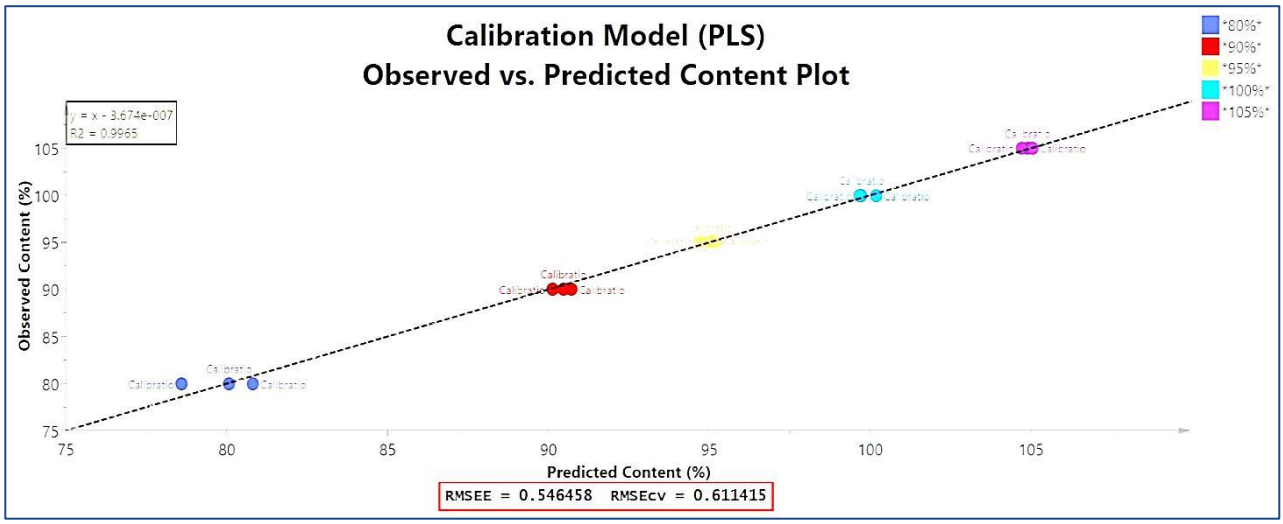


Figure 26 : Observed vs. Predicted content – calibration set (15 samples)

The sets at 80% are slightly spread over the calibration line. This may be due to the fact that they contain the highest amount of excipients, which some of them are known to be hygroscopic. Hence the content prediction may be affected in the region of 80% content. However, from a practical point of view, the drug product samples to be assayed with this method are centred on 100% and show a relatively low variability around this target (the specifications limits are set at +/-5%). Therefore, the accuracy of content determination in the region of interest (95-105%) should not be negatively affected.

To summarize, the calibration model obtained with the previous PLS parameters gave the following results:

- Number of samples: 15
- Number of principal components chosen: 1
- $R^2Y = 0.997$
- $Q^2Y = 0.995$
- RMSECV = 0.61 %label claim

2.3.6 Validation sets prediction

The external validation consists in testing the calibration model using a new set of known content samples. These 12 samples are prepared in the same way as the calibration samples and their spectra are recorded. Then, based on these new spectra, the calibration model is

used to estimate the content of the samples. The estimated API content in the validation sample should fall close to the true content.

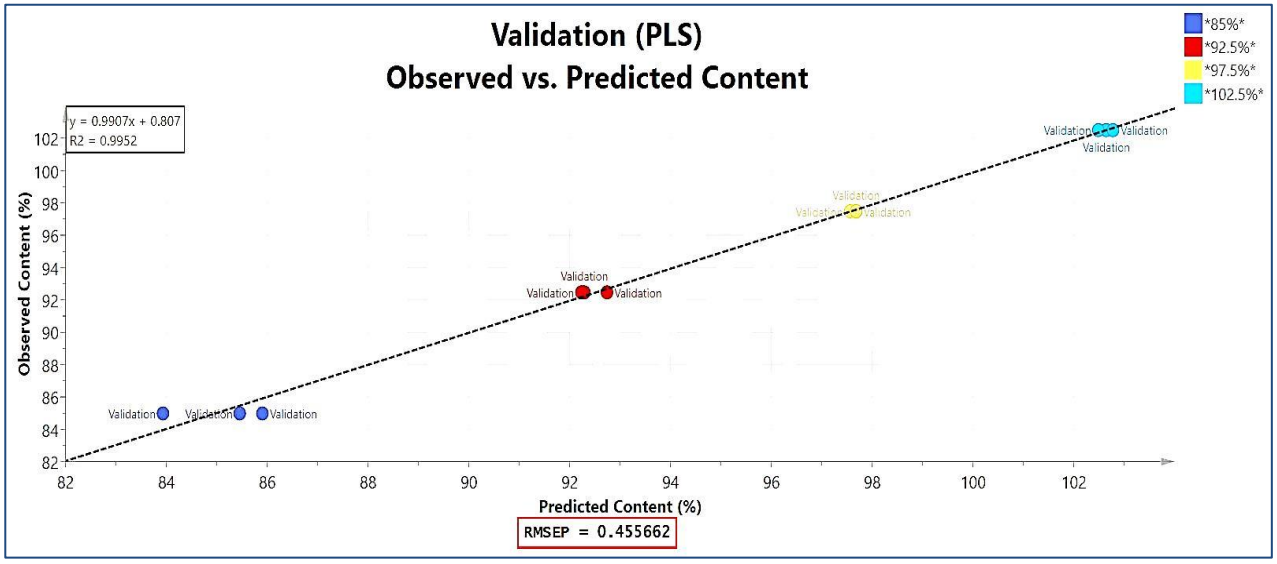


Figure 27 : Observed vs. Predicted – Validation set (12 samples)

The external validation (data set not included in the calibration model) produces a RMSEP value. This is a more reliable estimate of the method precision compared to the RMSECV obtained by cross-validation.

The final error of prediction is RMSEP = 0.46% which is low and in the same order of magnitude than the cross-validation RMSECV.

As observed on Figure 26, there is a slight spreading of sets 85% over the regression line, for the same reasons discussed for the calibration.

To summarize:

- Number of samples used to build the model: 15 (calibration set)
- Number of samples to predict: 12 (validation set)
- Number of principal components: 1
- $R^2Y = 0.991$
- RMSEP = 0.46

Based on these results, it is concluded that the PLS model used to predict the API content from the NIRS spectra (applying the spectral range selection and mathematical pre-treatments described in section 2.3.4) is reliable in term of precision and accuracy.

2.4 Final PLS model

After the external validation, the 27 samples (calibration and validation sets) are pooled together in order to build a more powerful and robust PLS model, and assess the overall RMSECV. Figure 28 shows the overall RMSECV of the model which is ≈ 0.58%, very satisfactory.

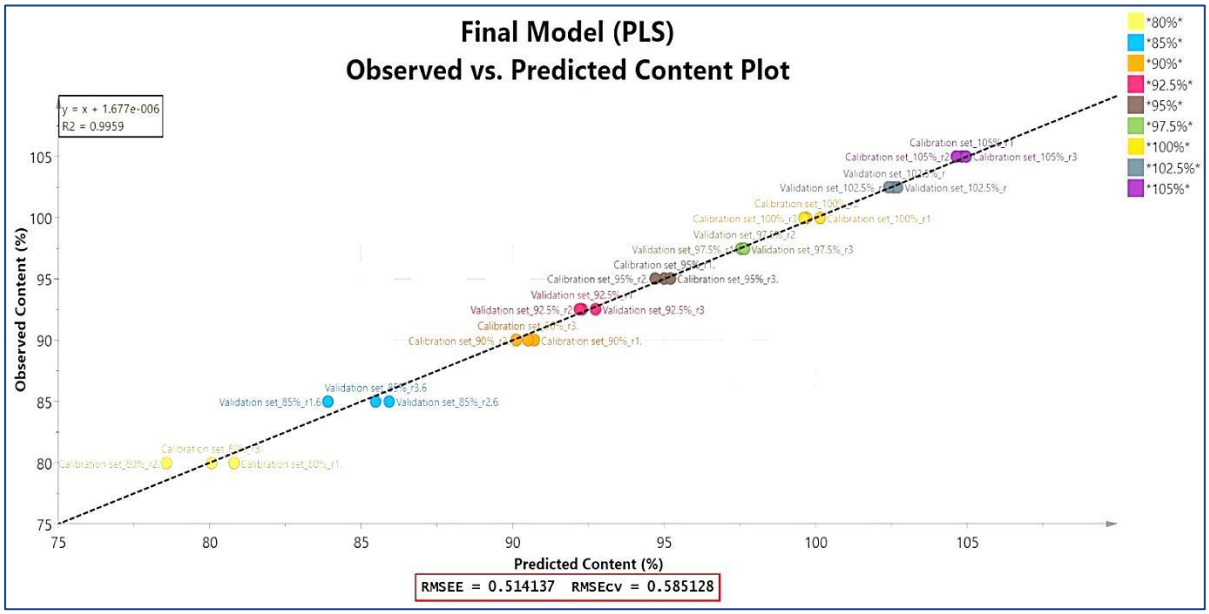


Figure 28 : Observed vs. Predicted content – Calibration and Validation sets (27 samples)

Several statistical tools are available to assess the relevance of the model, some of them will be presented in the next points.

2.4.1 Coefficient Plot

On the loading column plot (Figure 29), the impact of each selected absorbance wavelength is observed.

During the model construction, the absorbance bands were selected in order to have a statistical significance to calculate the API content. This means that the error bars at 95% confidence level (in red) do not cross the X-axis.

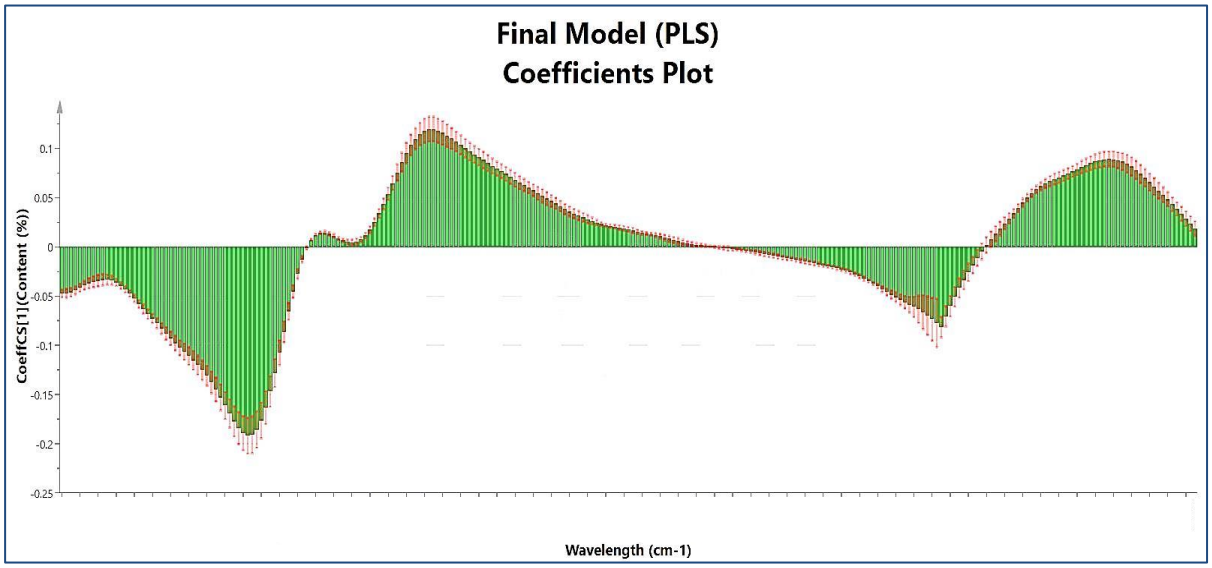


Figure 29: Loading column plot

2.4.2 DModX for PLS

The DModX graph (Figure 31) reports the residual distance to the model of each spectra in the X-dimensional space. In other words, after processing the spectrum with the PLS method, some X-data remain unexplained because it could not be correlated to the Y responses. The residual distance between the observation and the model (see Figure 30) can be computed and interpreted to determine if the spectrum is an outlier.

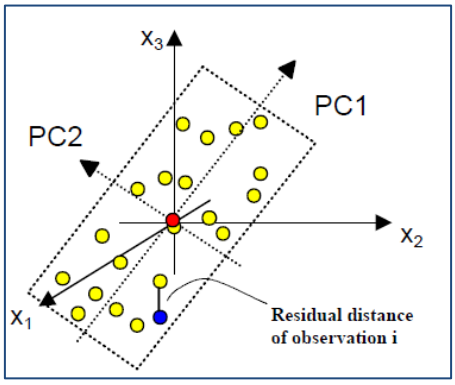


Figure 30 : Representation of residual distance

An observation above the threshold of twice the Dcrit (0.05) could be considered as unusual. Therefore, in this figure one outlier is observed, corresponding to validation set_85%_r1). It is kept in the model because its leverage in the model is weak (see Figure 32). Moreover, it is important to note that there is no noticeable sequence of samples with a persistent high DModX. Such a sequence could lead to a not desirable shift of the model plane.

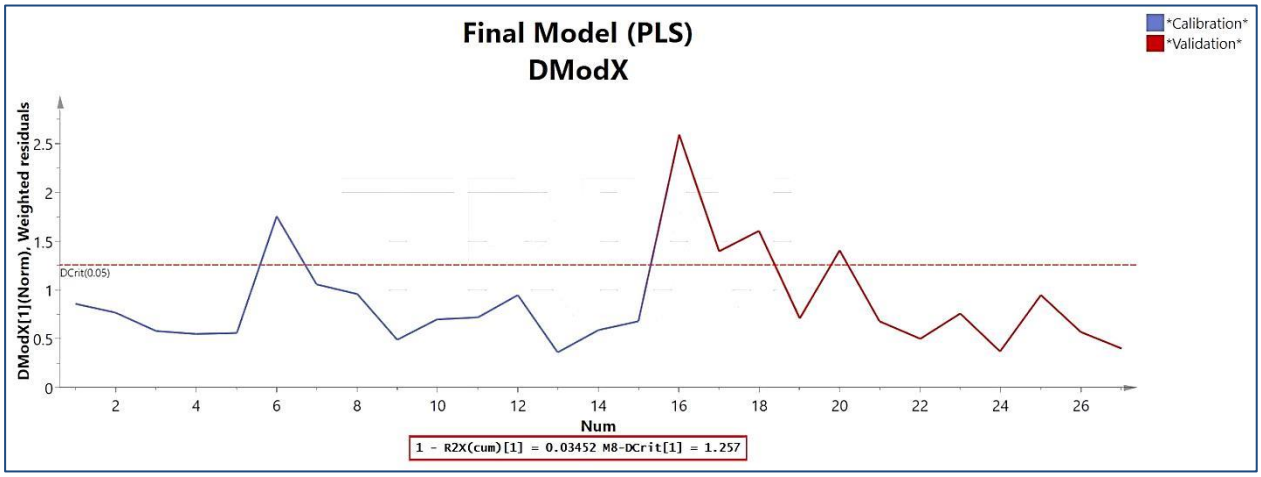


Figure 31 : DModX - Final Model

2.4.3 Hotelling's T^2 for PLS

The Hotelling's T^2 or leverage graph (Figure 32) is an estimation of the sample point's influence on the PLS model. An observation with a high leverage value exerts a large influence on the PLS model and could be considered as an outlier.

It was observed that the boundaries of the model (80 % and 105% of label claim) have a slightly higher importance compared to the other points. Indeed, it is usual to have a higher leverage on the extreme sides of a quantitative model.

No outlier was observed, meaning that all the samples are useful and have comparable weight on the model to predict the API content.

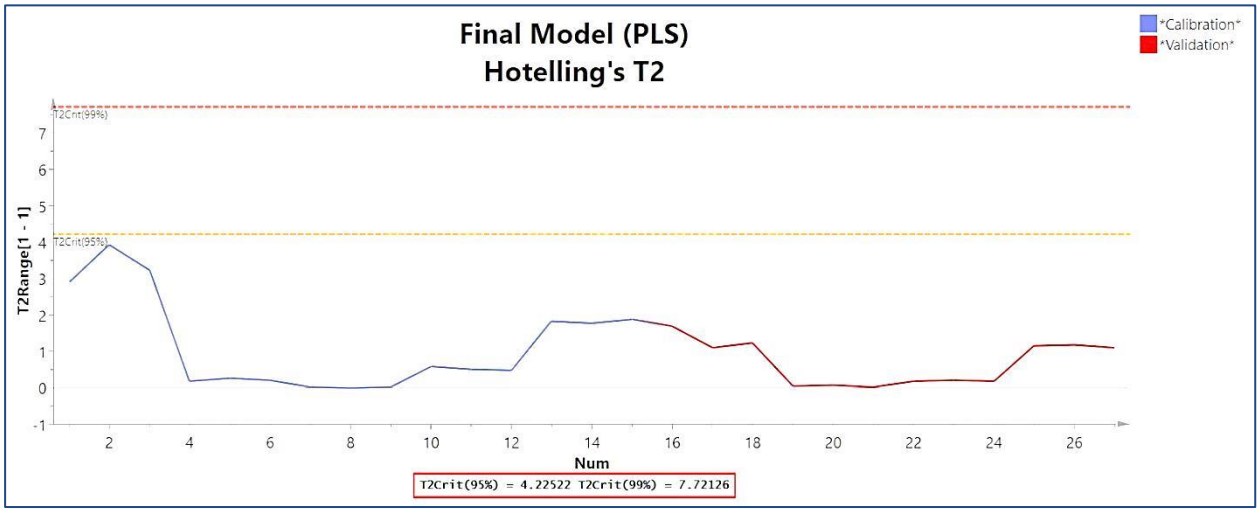


Figure 32 : Final Model - Hotelling's T2

2.5 Conclusion on model development

The model obtained with the calibration set has an error of prediction (RMSEP) < 0.6 %.

To assess the robustness of the model, the prediction results of calibration, cross-validation and external validation are compared. The value of RMSEE (calibration set), RMSECV (cross-validation) and RMSEP (external validation set) are all three less than 0.6% of API content and the R^2 value is higher than 0.99.

According to these results, the model developed is stable and ready for ICH validation.

Number of principal components	Spectral filters	RMSEE	RMSECV	RMSEP
1	D1+SNV MSC on bands selected for modelling	0.51	0.58	0.46

Table 3 : Summary of models settings and precision results

2.6 Final model on OPUS

In order to be actually used for content determination, the final model had to be implemented on OPUS, on which a user program can be created, to perform the analysis. The analysis program was built based on the 27 samples (calibration + validation set).

The samples have been pretreated according to Table 3.

The PLS was performed on OPUS with the following settings:

- Number of samples: 27
- Number of principal components: 1
- Spectral range: as defined
- Mean centering (standard setting for NIRS model)
- Number of samples excluded for cross validation: 7

The final model obtained with OPUS gave the following results:

- R^2Y = 99.7%
- RMSECV = 0.43
- Bias = -0.0137 (has to be as close as possible to zero)
- The Mahalanobis distance threshold = 0.31

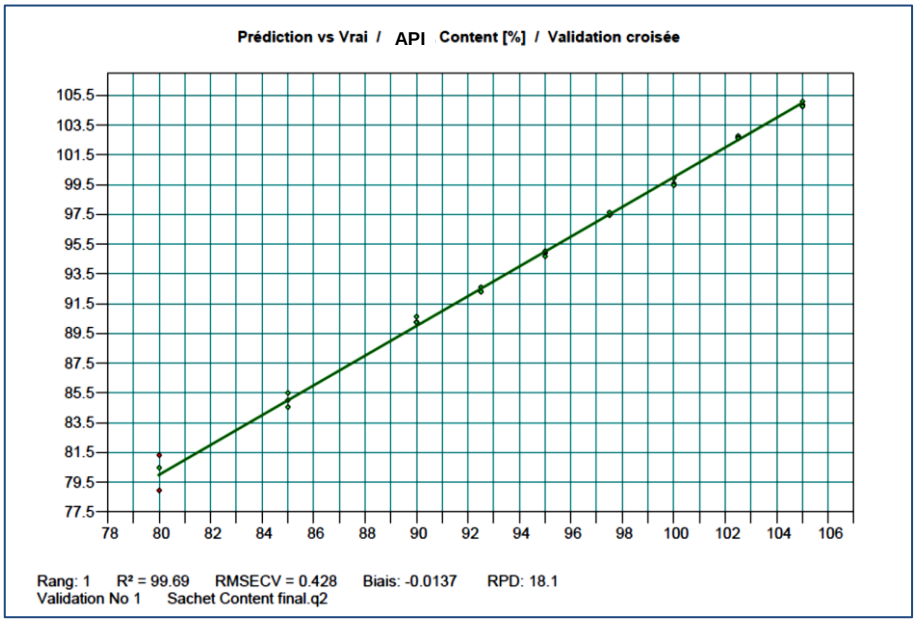


Figure 33 : Final Model - OPUS implemented model

These parameters are similar to what was found during the SIMCA development, hence this final model was stored as the reference model for the routine analysis. The resulting program allows to display the content result directly when measuring a sample's spectra.

The following analysis (2.7 ; 3 ; 4 ; 5) were performed using this model.

2.7 Specificity of the model

The specificity of the model was evaluated using 2 OPUS statistics tools: the Mahalanobis Distance (MD) and the Residuals. New commercial samples were tested using the method developed in section 2.6.

2.7.1.1 Mahalanobis Distance

To ensure the detection of outliers by NIRS, a Mahalanobis distance (MD) test is included in the OPUS analysis software. This test ensures that the spectrum of the analysed sample does not diverge from what is expected according to the calibration data [14]. It is similar to the Hotelling's T^2 presented in 2.4.3.

The spectral structures of the complete calibration data set are compared to the structure of the analyte spectrum. More precisely, MD is defined as the difference between the measured spectrum of the analyte and the mean value of all spectra in the calibration data set. MD integrates both the variation between samples at a given wavelength and inter-wavelength variations.

If the spectrum contains structures which do not fit the calibration range, an increase of the Mahalanobis distance is observed.

OPUS determines a Mahalanobis Distance Threshold (MDT) by computing the mean value and the standard deviation from the calibration set spectra (in this case, the spectra of 27 samples). Assuming normal distribution, a one-sided limit covering a probability of 99,999% is defined. For the 27 samples in the final model, the MDT was calculated to be 0.31.

To detect outliers, the MDI (Mahalanobis Distance Index) is displayed in analysis reports. The MDI is the ratio of the MD (calculated on a given spectrum and using the model developed) to the MDT:

$$\mathrm{MDI}=\frac{\mathrm{MD}_{sample}}{\mathrm{MDT}}$$

Theoretically, a MDI superior to 1 indicates that the sample spectra is an outlier (its distance from the model is higher than the threshold). Conversely, a MDI < 1 indicates a spectra that can be reliably analysed using the model developed in this study. This value can be increased in the method in order to be less restrictive and to account for some variability that could appear in the future (e.g. changes in raw materials).

2.7.1.2 Spectral Residuals

The model cannot explain the total variance within the spectral dataset. The part of spectral information that is not explained is called residuals, and corresponds to the difference between the raw data and the data processed with the PLS model. Higher residuals indicate a possibility of outliers. The difference is calculated between the measured spectrum x_i and the spectrum expected from modelisation i.e. from scores [15]. The residual value is interpreted by calculating a F-value and a FProb for each sample tested. If the FProb value is larger than 0.99, then the sample is reported as an outlier (99% confidence threshold). Spectral residuals are similar to the DModX calculated with SIMCA (cf. 2.4.2).

2.7.1.3 Specificity of the model against suppositories and tablets

To assess the specificity of the method, spectra of suppositories and tablets based on the same API at different %label claims were analysed with the Sachets method. API content, Mahalanobis distance and spectral residuals are reported in Annexe 2

Annexe 2The majority of the spectra have a Mahalanobis distance lower than the threshold. This is due to the fact that the Sachet model was built with only one component explaining the variation of API content. The variation due to the excipients is very low as the formulation of granules is 95% of API w/w, hence it is not taken into account in the model. The only thing the model can detect when analysing tablets and suppositories is the variation of API.

The only spectra seen as outliers in terms of MD are Tablets 120%, for which the spectra are distant enough from the sachets calibration model to be distinguished.

However, when the spectral residuals are analysed, the distance between the Sachets model and the other formulations is logically detected. Therefore, the model is specific to sachets and can identify the suppository and tablet spectra as outliers thanks to the residuals.

2.7.1.4 Specificity of the model against raw materials

The spectra of raw materials previously measured (cf. 2.2.2) were analysed with the quantification OPUS method to see if they were detected as outliers.

Results are reported in Table 4.

Sample		Predicted API Content (%)	MD	Limit	Residuals	F Value	F Prob	Outlier
Raw Material	**Batch Number**							
API	B25213	106.08	0.088	0.31	0.00814	2.7	0.888	-
	B25343	106.06	0.088	0.31	0.00739	2.22	0.852	-
	B30195	105.86	0.085	0.31	0.00748	2.28	0.857	-
	B30384	106.17	0.09	0.31	0.00852	2.96	0.903	-
	B32058	105.67	0.082	0.31	0.00528	1.13	0.704	-
Excipient 1	31228	392.58	55	0.31	0.472	9090	1	MD/Res
	79881	395.21	56	0.31	0.482	9460	1	MD/Res
	118185	388.92	54	0.31	0.441	7900	1	MD/Res
Excipient 2	55024	263.54	18	0.31	0.679	18800	1	MD/Res
	65605	266.78	19	0.31	0.686	19100	1	MD/Res

Table 4 : Results of specificity analysis against raw materials

As expected, the excipients are detected as outliers in terms of MD and spectral residuals. Nevertheless, the API is not. This was expected given that pure API is very similar to the 105% label claim sets included in the calibration range. Its predicted content as well is consistent (around 106%).

3. Comparison between NIRS method and HPLC method on uncrushed and crushed granules

In order to validate the NIRS method, the equivalence with the HPLC analysis has to be demonstrated.

A set of 10 samples composed of commercial Sachets 1g & 2g was used for this test. The sticks were picked up in the same way as the reference HPLC sampling method, so that the NIR results could be compared with the HPLC ones.

Six spectra were measured for each sample before and after crushing the granules. The average content obtained from the 6 spectra is used as reference value for the sample.

The API content was predicted using the final PLS model. The specification range is 95 – 105% of API label claim.

The results for each commercial batch with the three different analysis are reported in Table 5 below.

Batch Number	% Label claim		
	HPLC	NIR (Uncrushed)	NIR (Crushed)
L14849A	100.71	100.47	100.39
L14850A	101.27	100.25	100.21
L16102A	101.13	100.55	100.41
L16034B	100.71	100.43	100.48
L16038D	100.40	100.45	99.97
L17061B	101.37	100.22	100.13
M12195A	101.86	99.811	99.85
M12197A	101.41	100.15	99.83
M12200A	102.30	100.09	99.84
M12034A	101.43	100.05	99.72
Mean	**101.26**	**100.25**	**100.08**
Distance HPLC vs. NIR	**-**	**1.00%**	**1.17%**

Table 5 : HPLC and NIR methods comparison

The relative distance between the mean results of the two methods is calculated using the following equation:

$$RD\ \% = \frac{2 \times |Mean_{HPLC} - Mean_{NIR}|}{Mean_{HPLC} + Mean_{NIR}} \times 100$$

The acceptance criteria is that the RD% should not be greater than 2%, which is the case here with a RD% equal to 1.17% and 1.00% between HPLC and NIR (crushed and uncrushed respectively). Based on these results, it can be concluded that the three API assay methods provide comparable measurements, which makes the NIRS analysis suitable for routine analysis.

4. Comparison between Uncrushed and Crushed samples

4.1 Method

One of the objective of the assay method development was to assess the feasibility of performing a measure on the granules, without the crushing step. Given that the calibration model was built on powder samples, it is therefore necessary to check that the method is valid on Uncrushed granules as well as it is on Crushed ones.

To do so, the previous content results (Table 5) of commercial granules were used. The equivalence between Uncrushed and Crushed assays is tested with a paired T-test on the means of each class.

The specificity of the model is tested as well with MD and Residuals to check that the Uncrushed granules are not detected as outliers compared to the crushed form.

4.2 Results and discussion

The t-test is performed with Minitab software. The results are presented on Figure 34 below.

Paired T-Test and CI: Uncrushed, Crushed

Paired T for Uncrushed - Crushed

	N	Mean	StDev	SE Mean
Uncrushed	10	100.198	0.233	0.074
Crushed	10	100.130	0.296	0.093
Difference	10	0.0684	0.2357	0.0745

95% CI for mean difference: (-0.1003, 0.2370)
T-Test of mean difference = 0 (vs ≠ 0): T-Value = 0.92 P-Value = 0.383

Figure 34 : Paired T-test analysis of population means (Uncrushed vs. Crushed)

Considering these results, there is no statistical difference (p-value > 0.05) between the two classes mean, which signifies that the method is suitable for both sample forms.

4.3 Specificity tests on commercial granules

The specificity test results (MD and Residuals) are presented in Annexe 3.

Because the model has been built using crushed samples, some uncrushed granules are seen as outliers (high residuals) with the current parameters. It is concluded that the model still contains some physical information despite the pre-treatments applied. However, the content prediction is not statistically affected.

5. Model Robustness on Uncrushed coated granules

The previous methods were developed for crushed forms, and their robustness was already tested and found satisfying according to ICH guidelines. Before going further to the method validation, and given that it is the first time a method is used on granules form, it was decided to test the robustness of the method on one critical parameter which is the granules size.

Commercial granules are separated according to their sizes using sieves of different mesh size: 850 and 1000µm, then analysed with the Sachet method. The corresponding bulk content is measured.

The results obtained are reported in Table 6.

Batch Number	Sample	Content Mean (%)
L14408A	Bulk	**100.55**
	850	**99.35**
	1000	**99.61**
L14409A	Bulk	**100.23%**
	850	**98.88**
	1000	**99.08**
L16014A	Bulk	**100.52**
	850	**99.34**
	1000	**99.27**
L16104A	Bulk	**100.55**
	850	**99.31**
	1000	**99.30**

Table 6 : Robustness study on granules size results

The mean content value of the three classes are compared with an ANOVA test (Figure 35).

One-way ANOVA: Bulk, 850, 1000

```
Null hypothesis         All means are equal
Alternative hypothesis  At least one mean is different
Significance level      α = 0.05

Analysis of Variance

Source  DF    Adj SS    Adj MS  F-Value  P-Value
Factor   2  0.000382  0.000191    46.42    0.000
Error    9  0.000037  0.000004
Total   11  0.000419

Means

Factor  N     Mean    StDev         95% CI
Bulk    4  1.00463  0.00153  (1.00233, 1.00692)
850     4  0.99222  0.00229  (0.98993, 0.99452)
1000    4  0.99313  0.00219  (0.99083, 0.99542)

Pooled StDev = 0.00202952
```

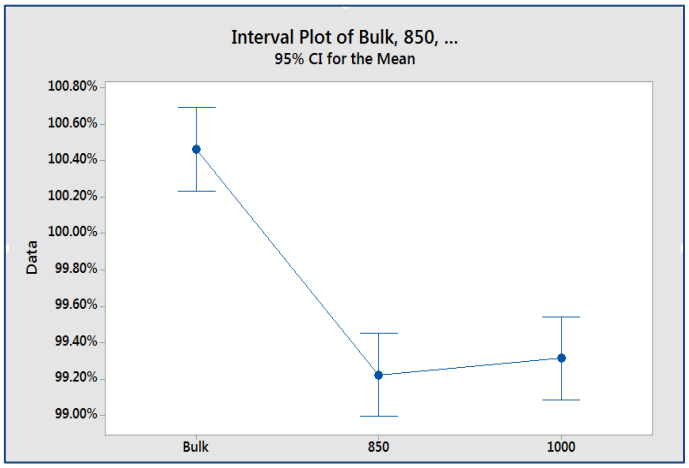


Figure 35 : ANOVA test results

According to these results, the observed differences in predicted content (approximately 1%) is statistically significant (p-value < 0.05). More precisely, the bulk predicted content is statistically different from the sieved granules. This difference may be due to the difference of light scattering within the samples. The granules size being different, the light does not follow the same trajectory within the samples, thus the collected absorbance is different.

The crushing step allows to reduce the physical differences, mainly in particle size distribution, between the tested samples. Hence, for release testing purpose, it is more robust and reliable to include a crushing step before the NIR spectra are collected.

However, the Uncrushed method still showed very satisfactory results in term of content prediction. Consequently this method without crushing could have an interesting application in on-line process content monitoring, as no sample preparation step would be required.

6. Conclusion on the model

The purpose of these tests was to develop a fast analytical assay method for at-line measurement of API content in Sachets 1g, 2g and 4g using NIR spectroscopy. The second purpose was to compare the developed method to the reference one (HPLC).

- Model precision:

The precision of the calibration model determined by external validation is 0.46 % of API label claim (RMSEP).

The final model composed of the calibration and validation set gives a precision (obtained by cross validation) of 0.58 % of API label claim (RMSECV).

This precision of approximately 0.6 in % of API label claim is fully satisfactory compared to the specification range (95-105% of label claim).

- Model accuracy:

The part-to-part comparison on commercial samples tested with the QC-HPLC method and the NIRS method shows that these two methods are not statistically different.

- Model specificity:

As far as the crushed granules are concerned, the final model is restrictive enough to identify the suppository and tablet spectra as outliers, thanks to MD and spectral residuals. Therefore, the sachet model is considered as specific.

Given the MD obtained when assessing specificity (2.7 and Table 4), the MDI can be reliably set at 1.5 in the OPUS method.

- Model robustness :

Preliminary robustness study has shown that the content determination is slightly sensitive to the granule size if the method is applied on uncrushed granules. Therefore, the method to be validated includes a crushing of the granules prior to the NIR spectra measurement. A complete robustness study will be included as part of the method validation.

- Comparison with the reference method (HPLC) :

The relative distance between both methods was calculated and found satisfactory (< 2%). Moreover, the HPLC method has a RSD% ranging from 0.4 to 1.0 [16], which is slightly less precise than the NIRS method (RMSECV < 0.6%). Thus the NIR method can be used as an alternative to HPLC to determine the API content in the finish product.

This feasibility study demonstrated that it is possible to use near infrared spectroscopy for API content in Sachets 95%.

Given the application of the analysis to batch release, it was decided that the validation would be performed on crushed granules. However, the NIR assay analysis on Uncrushed granules would be a valid approach for process monitoring applications.

Part 3: Analytical method validation

The validation of NIR method is a mandatory step at the end of development, as it is for any analytical method. The aim of validation is to guarantee that each of the future results obtained in routine analysis will be close enough to the real content. The validation process is documented by regulatory authorities.

Method Validation was performed according to the ICH Harmonized tripartite guideline on Validation of analytical procedures [17] and European Medicines Agency guidelines on the use of Near Infrared Spectroscopy (NIRS) by the pharmaceutical industry and the data requirements for new submissions and variations [18].

The validation strategy set up covers the requirements for model and method validation:

- Statistical Spectral quality test and Outlier Handling
- Risk assessment of variables
- Specificity
- Linearity
- Range
- Precision
- Accuracy
- Method robustness

1. External validation set

In order to validate the method for API assay, an external validation set is used to test the predictive ability of the model. The same formulations as for the calibration set (80%, 90%, 95%, 100%, and 105% of %API label claim) are prepared and analysed with the NIR method.

The external validation set was prepared according to the composition stated in Table 2.

2. Specificity

2.1 Acceptance criteria

Due to the intrinsic properties of NIR PLS models, the prediction of API content is based on the API and excipients absorbance bands. Therefore, the specificity of the model has to be demonstrated during validation, especially proving that this method is able to discriminate the sachets from other forms of products based on the same API and raw materials present on the production site.

Specificity is evaluated using the Mahalanobis distance (MD) to the model, and the spectral residuals.

Specificity is evaluated using the following samples:

- 10 x Tablet 1g
- 10 x 1g Suppository
- 13g PLACEBO Sachet 95%
- Raw materials: Excipients (approximately 13.00 g)

Acceptance criteria

- Mahalanobis distance index for Sachets 95% is below the defined MDI of 1.5
- Mahalanobis distance for all other samples is above the defined threshold of 0.31. These samples are detected as outliers by OPUS software, meaning their Mahalanobis distance index is higher than 1.5
- Sachets are not seen as outliers by the OPUS method i.e. their FProb is lower or equal to 0.99
- All other samples are detected as outliers i.e. FProb > 0.99

2.2 Results

In the table below, the MDI is the ratio of the Mahalanobis Distance to the Mahalanobis Distance threshold of 0.31. Samples with an MDI superior to 1.5 are systematically rejected.

Spectral residuals outliers are directly indicated as such on the OPUS report, after calculation of the spectrum distance to the calibration spectra. The software computes a value of FProb related to the quantification of the spectral residuals. If FProb > 0.99, the sample is reported as an outlier.

Sample	Batch number	Mahalanobis Distance Index	Spectral Residuals FProb	Sample rejected
Sachet 95%	L14409A	0.1	0.88	**No**
Tablet	L16671	0.2	1	**Yes**
Suppositories	L13030	0.01	1	**Yes**
Placebo Sachet 95%	M12502#1	> 133	1	**Yes**
Excipient 2	139400	> 42	1	**Yes**
Excipient 1	118185	> 127	1	**Yes**

Table 7: Results for Specificity

The Mahalanobis distance for the Sachets 95% is inferior to the threshold, and they are not detected as outliers in terms of spectral residuals either. Concerning the other samples, they

all have a MDI superior to 1.5, or/and are detected as outliers by the spectral residuals analysis. Therefore, this specificity test demonstrates that the analytical method is specific to the Sachets 95%.

3. Linearity

3.1 Acceptance criteria

The linearity is evaluated based on the linear regression (by the method of least squares) performed for the external validation test (80, 90, 95, 100, 105% label claim). The linear regression is performed on the external validation set, in which API content is determined by weighing to avoid introducing the uncertainty of the HPLC reference analytical method

Parameter to be reported	Specification for Validation
Correlation Coefficient	Reported for information
Y-Intercept	Confidence interval at 95% of the intercept of the regression includes 0
Slope of The Regression Line	Confidence interval at 95% of the slope includes 1
Residual Sum of Squares (RSS)	Reported for information
Residual analysis	Normal distribution of residuals in the full range of analysis.
Coefficient of Determination (R^2)	R^2 > 0.99 [7]

Table 8: Criteria for Linearity

Key parameters for linear regressions (slope, y-intercept and coefficient of determination R^2) are reported in Table 9.

3.2 Results

% of API label claim		80	90	95	100	105
Weighed % of API label claim	1	78.769	89.730	94.901	99.998	105.000
	2	79.941	90.150	95.220	99.988	104.998
	3	80.363	90.213	95.519	99.996	104.998
Measured % of API label claim	1	79.999	89.998	94.9973	100.113	105.140
	2	80.000	89.997	95.000	100.210	105.070
	3	79.997	89.994	94.9954	100.357	105.355
Y-Intercept		**-1.924**				
95% CI for Y-Intercept		(- 4.26, 0.41)				
Regression Line Slope		**1.0212**				
95% CI for Regression Line Slope		(0.9965, 1.0460)				
R^2		**99.84%**				

Table 9: Results for Linearity

These results were obtained by performing a regression between the real and predicted content, as plotted on Figure 36 below.

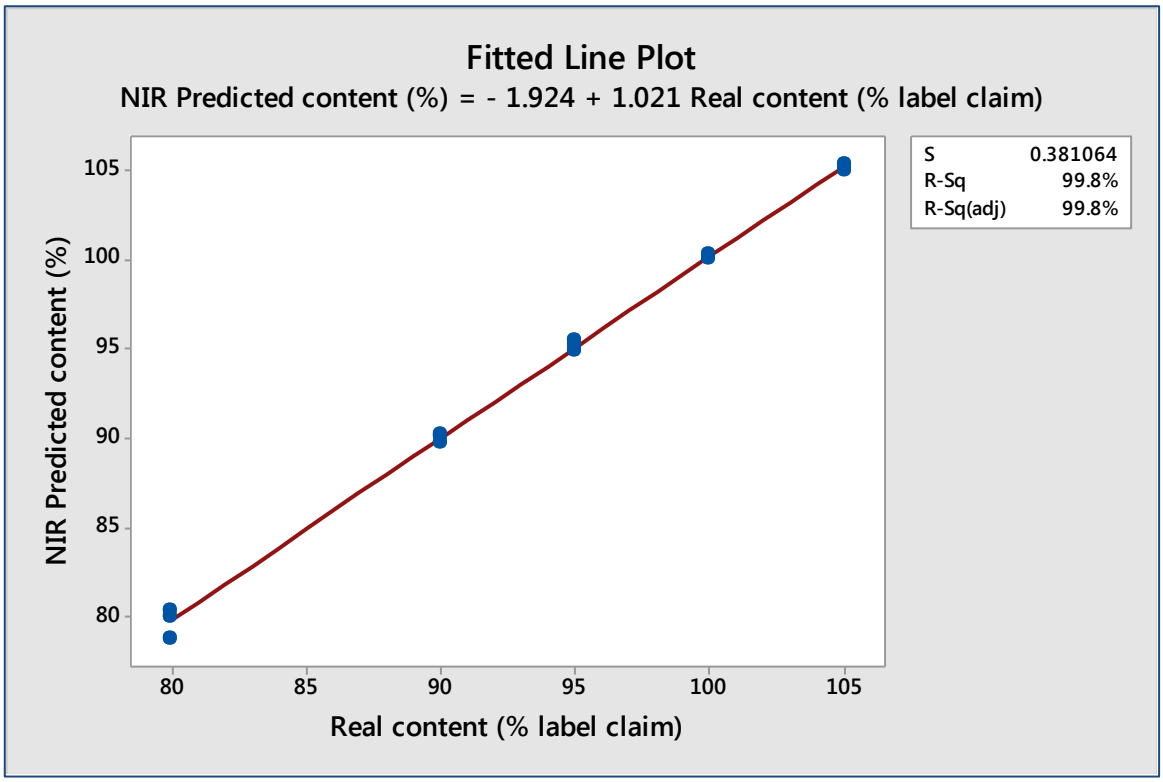


Figure 36 : Theoretical vs. Predicted for external validation set (15 samples)

As can be seen in Table 9 and Figure 36 above, R^2 is superior to the acceptance criteria of 0.99. Moreover, the CI at 95% of Y-intercept includes 0 and the CI at 95% of the slope includes 1. Hence the method is considered linear on the range 80-105% label claim.

4. Accuracy

4.1 Acceptance criteria

The accuracy is evaluated by computing the difference between the predicted value and the theoretical true value for the external validation test set (as described in 1). For each sample, the recovery percentage is calculated:

%Recovery = Predicted content / theoretical content * 100

Acceptance criteria

- Individual %Recovery: 98-102% (According to "Accuracy" in internal SOP)
- Bias not statistically different from zero

4.2 Results

% of API label claim		80	90	95	100	105
%Recovery	1	98.463	99.702	99.898	100.114	100.133
	2	99.926	100.170	100.231	100.223	100.068
	3	100.457	100.243	100.551	100.361	100.340
Mean for Recovery (%)		99.615	100.039	100.227	100.233	100.181
%Recovery Stdev		1.033	0.294	0.326	0.124	0.142
%RSD for Recovery		1.037%	0.294%	0.326%	0.123%	0.142%

Table 10: API Recovery

The graph below (Figure 37) plots the individual %Recovery values (grey dots) as well as the mean %Recovery (blue dots) for each %Label claim set. The corresponding confidence intervals (CI) at 95% are also represented. The red dotted lines represent the limits for individual %Recovery (98-102%) as defined in the protocol.

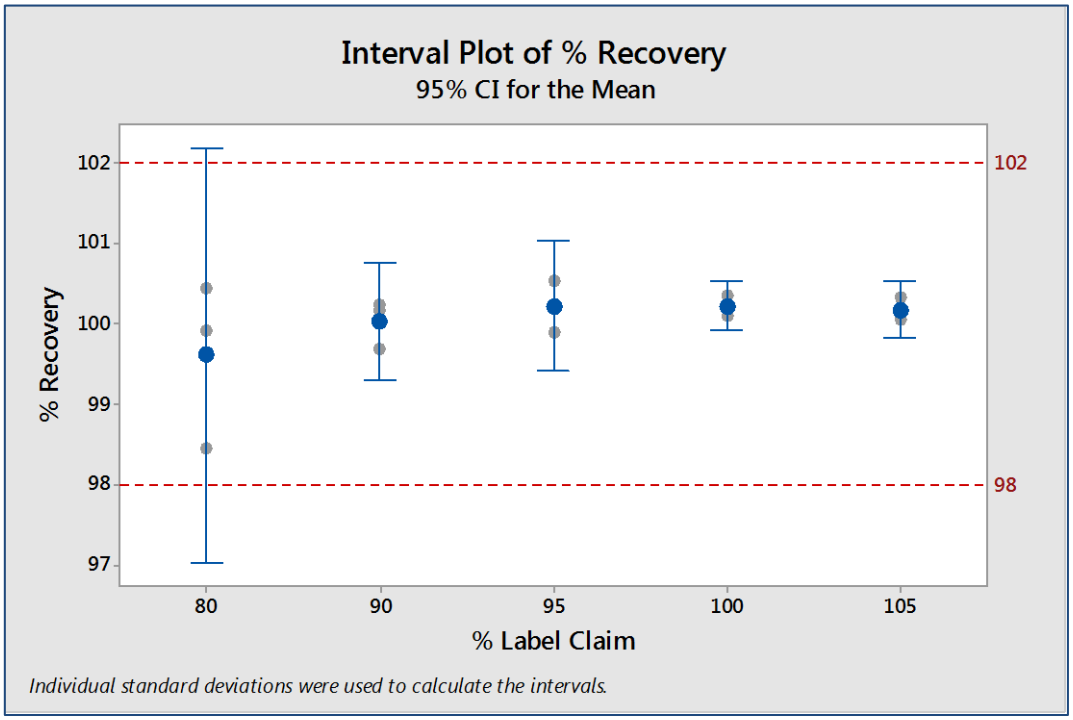


Figure 37: 95% CI for Recovery Mean (Confidence intervals based on pooled repetition error)

Mean %Recovery	**100.059**
95% CI for the %Recovery Mean	**99.786 - 100.332**

Table 11: Recovery Percentage

Conclusion:

Acceptance criteria	**Individual %Recovery 98-102%**	**Complies**
	Bias not statistically different from 0	**Complies**

All individual %Recovery were found within the target values (98-102%). The 80% sets are more spread in terms of %Recovery, although still in the specification range. This phenomenon had already been noticed during the feasibility study and is attributed to the variability of the excipients (hygroscopicity, etc.). However, the sets within the specification range for content determination (95-105% Label Claim) are much more tightened around the average %Recovery, which is fully satisfactory for the application of this analytical method.

Percentage Recovery meets the acceptance criterion. The method is accurate.

5. Precision

As indicated in the ICH guidelines [17], precision of an analytical procedure is the degree of scatter between series of measurements on a homogeneous sample under prescribed conditions. Precision is evaluated through repeatability and intermediate precision.

5.1 Repeatability

Repeatability is observed when the same operator measures the same sample repeatedly with the same method and equipment. Repeatability is assessed using the results of the external validation set for the following percentages of API label claim: 80, 100 and 105%.

Acceptance criteria

- RSD ≤ 2.0 %

% of API label claim	80	100	105
Predicted content Mean	79.691	100.227	105.188
Mean Recovery	99.615	100.233	100.181
CI 95% for Mean Recovery	97.050 – 102.181	99.925 – 100.540	99.828 – 100.533
%RSD on recovery	1.037	0.294	0.326
Acceptance Criteria	RSD < 2%	RSD < 2%	RSD < 2%
Criteria fulfilled	**Yes**	**Yes**	**Yes**

Table 12 : Method Repeatability and acceptance criteria

5.2 Intermediate precision

Intermediate precision shows the precision of the analytical method carried out in the same laboratory but by different technicians, using different equipment (if possible), at different days.

Intermediate precision is carried as follows:

- Two different operators analyse samples independently
- Each operator performs the analysis on a different day
- Six samples at 100% of API label claim are measured by each operator

Acceptance criteria

- $\frac{(Mean_{Operator\ 1} - Mean_{Operator\ 2})}{Mean_{Total}} \leq 2\%$
- RSD (n=12) ≤ 2%

	Real content (% Label Claim)	NIR Predicted content (% Label Claim)	%Recovery	%Recovery Mean	%Recovery RSD (%)
Operator A	99.999	99.584	99.585	99.695	0.142
	99.999	99.870	99.871		
	99.999	99.554	99.554		
	100.001	99.597	99.596		
	99.996	99.697	99.701		
	99.997	99.859	99.862		
Operator B	99.996	99.889	99.893	99.947	0.114
	99.997	100.033	100.036		
	99.995	99.808	99.814		
	100.001	100.043	100.042		
	99.998	100.063	100.065		
	99.999	99.828	99.830		

Table 13: Results for Intermediate Precision

Acceptance criteria	**%RSD (n=12) for API < 2%**	**0.128**	**Complies**
	$\frac{(Mean_{Operator\ 1} - Mean_{Operator\ 2})}{Mean_{Total}} < 2\%$	**0.251%**	**Complies**

Table 14: Acceptance criteria for Intermediate precision

All the criteria complies with the requirements, thus the precision of the method is validated.

6. Range

Range is established by confirming that the analytical procedure provides an acceptable degree of linearity, accuracy and precision for the samples of the external validation set (covering the range of the analytical procedure).

Acceptance criteria

Linearity, Accuracy and Precision are demonstrated on the range of 80 – 105 % of the tested concentration.

Conclusion:

The range of this analytical method was demonstrated through the tests for linearity, accuracy and precision. As can be seen in the previous paragraphs, the method was demonstrated to be linear, accurate and precise. Thus the range for API assay is validated between 80 and 105% of API label claim.

7. Method robustness and Risk Assessment of variables

A DoE was executed to test the method's robustness. The aim of this study was to document the influence of expected analytical method parameters variation on the quality of the results (change in accuracy was chosen as a response in this DoE).

Potential causes of variation (based on FMEA score)	Action already done	Recommended action
Sample chemical and physical properties		
Variation in raw material properties	➢ Considered in method development: different batches and suppliers of raw materials (API, Excipient 1, Excipient 2) have been used to develop and optimize the model	No recommended actions
Sample water content	➢ This has been considered in model development: the wavelength ranges used exclude mainly the absorbance bands of the water. ➢ Sample preparation and measurement are performed in a monitored room.	No recommended actions
Sample particle size	➢ All samples used for method development and used during this validation are crushed into fine powder.	Test in DoE the impact of incomplete crushing leading to bigger particle size
Sample preparation method		
Sample size	➢ Mix corresponding to 12 Sachets 1g and 6 Sachets 2g have been used for model development	Test in DoE the impact of a different sample size
Crushing step	➢ Crushing method has been optimized during previous developments to get an easy and fast method to prepare sample.	Test in DoE the impact of variation in this method (number of crush and time of each crush, crusher position, time separating crush and NIRS analysis)
Sample surface layering and smoothing	➢ Sample surface is quickly layered and smoothed with a spatula between each measurement.	No action recommended
Sample homogeneity	➢ Based on previous experiments, each result is based on six measurements, with a manual shuffling step between each measurement.	Test in DoE the impact of measurement without manual shuffling.
Sample dish	➢ All measures for method development were made using a dish of 7cm diameter	Test in DoE the impact of using a dish of 9.5 cm diameter to be similar to the other NIR analytical methods
Sample positioning	➢ Sample is positioned under the probe onto a rotating sample holder. This increases the surface scanned.	Test in DoE the impact of the rotating sample holder speed and position.
Measurement method		
Software parameters	➢ All measurement settings are fixed in OPUS LAB software. Operator has restricted access only.	No recommended actions.
Equipment drift	➢ Lifecycle management	No recommended actions.

Table 15: Overview of potential causes of variation in the model

DoE Variables	Low level	High level	Standard Settings
Sample size	80% of the sample size (equivalent to 8g API)	120% of the sample size (equivalent to 16g API)	100% of the sample size (equivalent to 12g API)
Number of crushes	2	6	4
Crusher angle (Yes/No)	No	Yes	Yes
Manual shuffling step between each reading	No	Yes	Yes
Analytical dish diameter (cm)	7	9.5	9.5
Rotating sample holder speed	No rotation	Speed 1	Speed 1
Rotating sample holder position	Off centred	Centred	Centred
Time separating crush and analysis	None	>2h	None

Table 16: DoE parameters for robustness testing

The sachet batches were produced at FICSA. For these batches, the mean API content is determined by averaging the value obtained of four samples measured using standard settings.

The impact of each parameter on model accuracy (output of DoE) is assessed by computing the ratio of predicted content over the mean content of the batch.

Acceptance criteria

No acceptance criteria, the DoE results have to be analysed, discussed and corrective action taken if necessary.

DoE Results

The results of the DoE are represented in Figure 38 (Half-Normal Probability Plot).

The half-normal probability plot is a graphical tool that helps assess which factors affect the response the most. This plot thus enables the selection of effects that “stand out” to include them in a model. Large effects (absolute values) appear in the upper-right section of the plot. The lower-left portion of the plot contains effects caused by noise rather than a true effect of the factors under investigation.

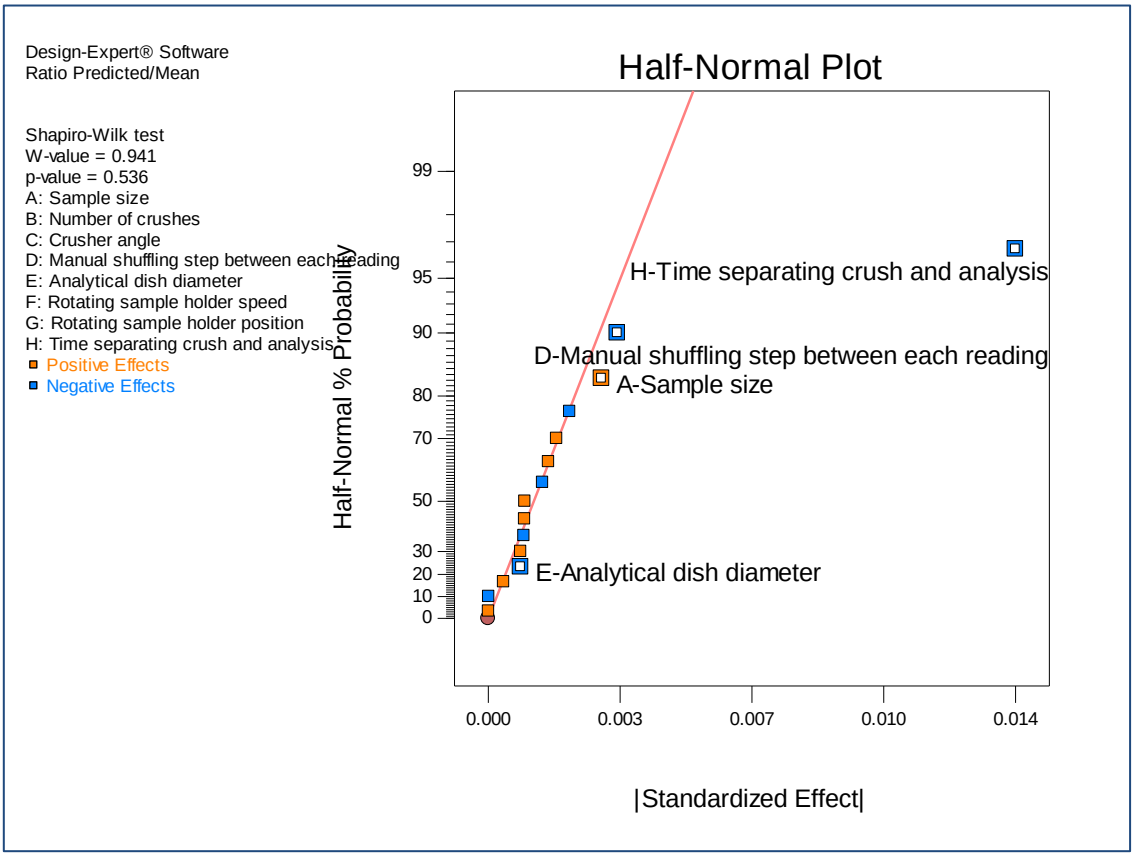


Figure 38: Half-Normal Plot

As can be seen on Figure 38, three effects could be included in the model: Sample size, Manual shuffling step between each reading and Time separating crush and analysis. The statistical significance of the selected effects in the model is assessed in an ANOVA (Table 17).

ANOVA for selected factorial model					
Analysis of variance table [Partial sum of squares - Type III]					
Source	**Sum of Squares**	**Mean Square**	**F Value**	**p-value Prob > F**	
Model	8.137E-004	2.034E-004	36.55	< 0.0001	significant
A-Sample size	*3.387E-005*	*3.387E-005*	*6.08*	*0.0313*	
D-Manual shuffling step between each reading	*4.398E-005*	*4.398E-005*	*7.90*	*0.0169*	
E-Analytical dish diameter	*2.769E-006*	*2.769E-006*	*0.50*	*0.4952*	
H-Time separating crush and analysis	*7.331E-004*	*7.331E-004*	*131.71*	*< 0.0001*	
Residual	6.123E-005	5.566E-006			
Cor Total	8.749E-004				

Table 17: ANOVA table for selected factorial model

Values of "Prob > F" less than 0.05 indicate that model terms are significant. In this case, the three selected effects (A, D and H) have a statistical significant impact on accuracy. However, when studying the individual influence of the three factors on the response, it appears that A and D's impact on the content prediction is in the same magnitude than the noise of the method. Hence, from a practical stand point, no significant impact is asserted for these two factors. As far as the H factor is concerned (Time between crush and analysis), the impact is larger (Figure 39) and calls for limiting the time between crushing and NIR measurements to maximum 45 min, according to Figure 39. This time limit was thus included in the SOP.

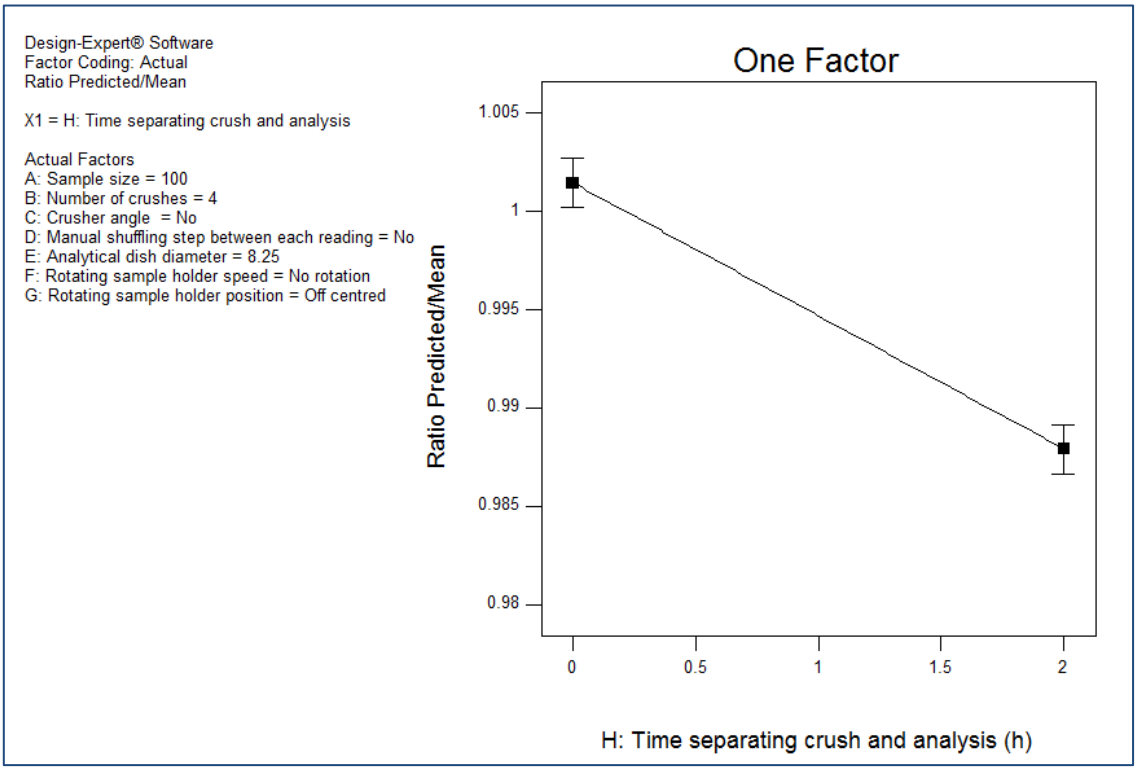


Figure 39: Influence of Factor H on response

Regarding the Factor E impact (Analytical dish diameter), the p-value is above 0.05, which means the impact on the measure is not significant. This is confirmed when looking at its influence on the response (Figure 40), which is of about 0.1%. Hence it can be neglected.

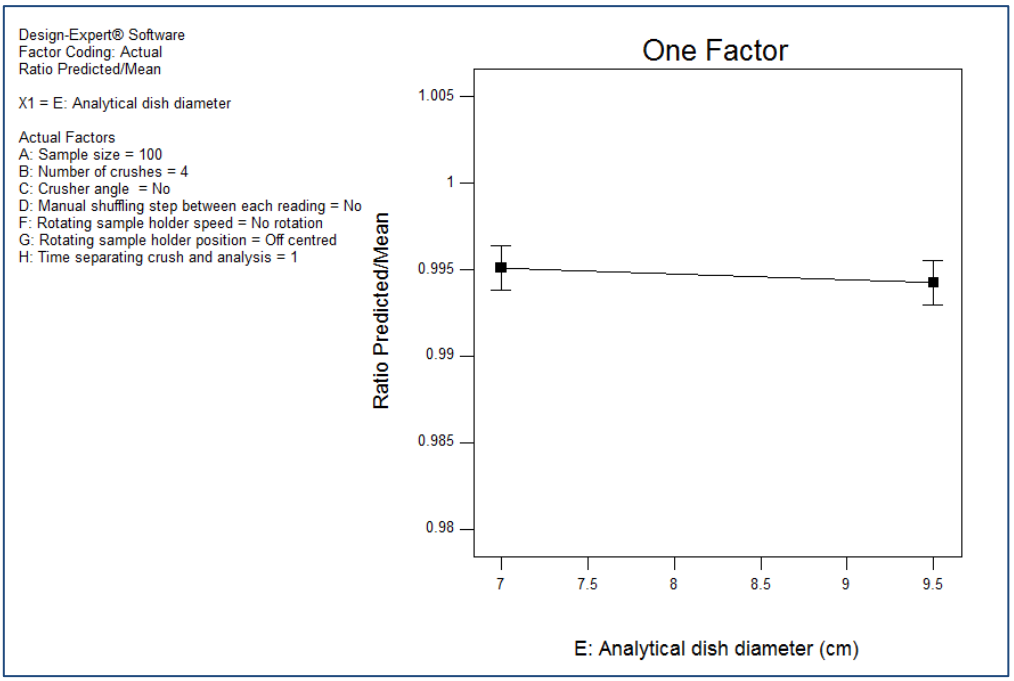


Figure 40 : Influence of Factor E on response

Conclusion on DoE results:

DoE results demonstrated that variation in only one parameter may affect measurement accuracy, this parameter being the time between the sample crush and the analysis with NIR. To ensure a correct analysis, it was specified in the analytical method procedure to perform the NIR analysis within 45 minutes after the crushing step, according to the results presented on Figure 39.

Therefore, the method can be considered robust as long as the agreed period between crush and analysis is respected.

Moreover, as there is no significant influence of the dish diameter on the response, the 9.5cm diameter dish was validated for further analytical uses, so that the analysis procedure is similar for the three forms (suppository, tablet, and sachet).

8. Conclusion on Validation

Analytical method validation on Sachets 95% was successfully performed. The parameters of the method are reported in Table 18 below.

Parameter	Procedure scope
Instrument	Bruker : FT-NIR Spectrometer Matrix-F Duplex (n° 961.00)
Software	OPUS version 7.0
Mode	Reflectance
Number of spectra/sample	6
Sample presentation	At-line mix of crushed sachets
Concentration range	Concentration range 80-105 % of API label claim
Spectral pre-processing	D1 combined with Standard normal variate (SNV) on full spectrum, then MSC on specific quantification bands

Table 18: Procedure scope

All generated data for specificity, accuracy, precision (repeatability and intermediate precision), linearity, range and robustness experiments gave satisfactory results.

These results meet the specifications laid out by the ICH guidelines on validation procedures [17] and EMEA guidelines on the use of NIRS [18].

Conclusion

The main objective of this work was achieved, as API assay method by NIRS in Sachet 95% was successfully developed.

This development is an iterative process where the model has to be adjusted in order to find the best parameters for the given purposes. The developer has to modify the pretreatments, the spectral bands to include or not in the calibration, the number of principal components to be used, etc. In the end, a balance between precision and robustness has to be found.

Now that the method is validated, this method for API assay in the finished product is being transferred to the Quality Control Laboratory to be used in routine. Prior to be used for batch release, the methods also have to be filed and approved by the regulatory authorities as alternative analytical methods to HPLC.

The implementation of NIRS methods will allow important savings for the company in terms of time and resources, hence it will help to reduce costs. Moreover, the methods are as reliable as HPLC and even seem to improve the precision of the result.

Furthermore, NIRS could have many other innovative applications for the company in the future. For example, it could be used for in-process control applications, allowing to monitor some steps of the process such as the coating. If the concept is further developed, NIRS would be really interesting as part of a real time release strategy as well.

Finally, NIRS has the potential to be used on different product types such as liquid or semi-liquids forms, hence could be of a great interest for other Ferring products.

NIRS is taking more and more importance in the pharmaceutical industry. This is evidenced by the regulatory authorities starting to publish specific guidelines about how to develop and validate NIRS methods (EMA, 2014 [19] ; FDA 2015 [20]). Hence it will be the responsibility of pharmaceutical companies such as Ferring to keep a watchful eye on the current trends in order to be in the race with the last innovative technologies in pharmaceutics.

Bibliography

[1] Indatech, “Optical spectroscopy techniques for process analysis,” 2016.

[2] M. Chaplin, “Sea within the Great Barrier reefWater Absorption Spectrum,” 01 July 2017. [Online]. Available: http://www1.lsbu.ac.uk/water/water_vibrational_spectrum.html. [Accessed 18 July 2017].

[3] Tomuta, Ian et al., “Development and validation of NIR-chemometric methods for chemical and pharmaceutical characterization of meloxicam tablets,” *Drug Dev Ind Pharma,* pp. 1520-5762, 2014.

[4] “Guidance for Industry: PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance,” U.S. Department of Health and Human Services - Food and Drug Administration, 2004.

[5] “MicroNIR Handheld Products,” VIAVI Solutions Inc., [Online]. Available: https://www.viavisolutions.com/en-us/osp/products/micronir-spectrometers/micronir-handheld-products. [Accessed 05 September 2017].

[6] S. Halsey, “Ensuring raw material quality,” *Ensuring raw material quality,* pp. 12-15, 2009.

[7] S. Rosa, P. Barata and J. Martins, “Development and validation of a method for active drug identification and content determination of ranitidine in pharmaceutical products using near-infrared reflectance spectroscopy: A parametric release approach,” *Talanta,* vol. 75, p. 725–733, 2008.

[8] K. De Smet, T. Van Den Kerkhof and A. Prudic, “The benefits of NIRS in monitoring and controlling a crystallisation process,” *European Pharmaceutical Review,* pp. 40-44, 2016.

[9] M. Fonteyne , J. Vercruysse, D. Córdoba Díaz, D. Gildemyn, C. Vervaet, J. P. Remon and T. De Beer, “Real-time assessment of critical quality attributes of a continuous granulation process,” *Pharmaceutical Development and Technology,* p. 18(1): 85–97, 2013.

[10] “Monitoring of the manufacturing process for ambroxol hydrochloride tablet using NIR-chemometric methods: compression effect on content uniformity model and relevant process parameters testing,” *Drug Dev Ind Pharm,* p. 41(11): 1877–1887, 2015.

[11] Bruker, OPUS - User Manual - Quant, 2011.

[12] Huang, Jun et al., “Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy,” *American Pharmaceutical Review,* January 2010.

[13] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikström and S. Wold, Multi- and Megavariate Data Analysis, Umetrics Academy, 2006.

[14] J.-P. Conzen, Multivariate Calibration : A practical guidefor developing methods in thequantitave analytical chemistry, Bruker Optik GmbH, 2014.

[15] *NIR Multivariate Calibration, Bruker, 3rd Edition, 2014.*

[16] *Q-3.2.P.5.3 Validation of Analytical Procedure PD-ANA-385:.*

[17] *ICH Harmonised Tripartite Guideline: Validation of Analytical Procedures: Text and Methodology Q2(R1).*

[18] *EMEA Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations, EMEA/CHMP/CVMP/QWP/17760/2009 Rev2, 27 January 2014.*

[19] E. M. Agency, *Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations,* 2014.

[20] Food and Drug Administration, U.S. Department of Health and Human Services, Development and Submission of Near Infrared Analytical Procedures - Guidance for Industry, Pharmaceutical Quality/CMC, March 2015.

[21] “Gastroenterology,” Ferring Corporate Website, [Online]. Available: http://ferring.com/en/products/gastroenterology/#expanded-0. [Accessed 18 August 2016].

[22] Ferring Pharmaceuticals, “Gastroenterology,” 2016. [Online]. Available: http://ferring.ca/en/products/gastroenterology/#expanded-0. [Accessed 04 June 2017].

Annexes

Annexe 1 : Real content of calibration and validation sets

NIR average spectra name	API (mg)	Excipients 1 and 2 (QS 12709 mg)	Real content (% of label claim)
Calibration set_80%_r1	9600.2	3109.2	80.0%
Calibration set_90%_r1	10800.4	1909.9	90.0%
Calibration set_95%_r1	11400.7	1309.5	95.0%
Calibration set_100%_r1	12000.0	709.3	100.0%
Calibration set_105%_r1	12600.0	109.2	105.0%
Calibration set_80%_r2	9600.0	3109.2	80.0%
Calibration set_90%_r2	10800.0	1909.3	90.0%
Calibration set_95%_r2	11400.3	1309.3	95.0%
Calibration set_100%_r2	12000.0	709.2	100.0%
Calibration set_105%_r2	12600.0	109.2	105.0%
Calibration set_80%_r3	9600.2	3109.6	80.0%
Calibration set_90%_r3	10800.1	1909.6	90.0%
Calibration set_95%_r3	11400.1	1309.3	95.0%
Calibration set_100%_r3	12000.3	709.6	100.0%
Calibration set_105%_r3	12600.1	109.2	105.0%
Validation set_85%_r1	10200.1	2509.2	85.0%
Validation set_92.5%_r1	11100.0	1609.2	92.5%
Validation set_97.5%_r1	11700.2	1009.4	97.5%
Validation set_102.5%_r1	12300.4	409.2	102.5%
Validation set_85%_r2	10200.0	2509.3	85.0%
Validation set_92.5%_r2	11100.0	1609.2	92.5%
Validation set_97.5%_r2	11700.0	1009.2	97.5%
Validation set_102.5%_r2	12300.0	409.2	102.5%
Validation set_85%_r3	10200.0	2509.2	85.0%
Validation set_92.5%_r3	11100.0	1609.2	92.5%
Validation set_97.5%_r3	11700.0	1009.2	97.5%
Validation set_102.5%_r3	12300.2	409.2	102.5%

Annexe 2 : Results of specificity analysis against tablets and suppositories with Sachets method

	Sample	Predicted Content (%)	MD	Mah.Limit	Residuals	F Value	FProb	Outliers
Tablets spectra	80% - repetition 1	97.15	0.0055	0.31	0.1120	510	1	Res
	80% - repetition 2	96.91	0.0047	0.31	0.1120	515	1	Res
	80% - repetition 3	96.72	0.0041	0.31	0.1040	441	1	Res
	90% - repetition 1	99.70	0.019	0.31	0.1010	413	1	Res
	90% - repetition 2	99.41	0.017	0.31	0.1070	464	1	Res
	90% - repetition 3	99.07	0.015	0.31	0.0995	403	1	Res
	100% - repetition 1	102.40	0.042	0.31	0.1140	527	1	Res
	100% - repetition 2	102.09	0.039	0.31	0.1120	507	1	Res
	100% - repetition 3	101.98	0.038	0.31	0.1000	407	1	Res
	110% - repetition 1	106.92	0.1	0.31	0.1490	901	1	Res
	110% - repetition 2	106.51	0.095	0.31	0.1430	831	1	Res
	110% - repetition 3	105.14	0.075	0.31	0.1240	628	1	Res
	120% - repetition 1	201.46	7.2	0.31	1.3600	75000	1	MD/Res
	120% - repetition 2	146.00	1.7	0.31	0.6480	17100	1	MD/Res
	120% - repetition 3	133.40	0.96	0.31	0.4790	9350	1	MD/Res
	85% - repetition 1	98.38	0.011	0.31	0.1050	450	1	Res
	85% - repetition 2	97.71	0.0078	0.31	0.1000	410	1	Res
	85% - repetition 3	98.59	0.012	0.31	0.1080	475	1	Res
	85% - repetition 4	98.54	0.012	0.31	0.1060	460	1	Res
	95% - repetition 1	100.65	0.026	0.31	0.1050	449	1	Res
	95% - repetition 2	100.01	0.021	0.31	0.0938	358	1	Res
	95% - repetition 3	101.21	0.031	0.31	0.1090	488	1	Res
	95% - repetition 4	101.04	0.029	0.31	0.1050	451	1	Res
	105% - repetition 1	104.21	0.063	0.31	0.1170	560	1	Res
	105% - repetition 2	102.61	0.044	0.31	0.1020	420	1	Res
	105% - repetition 3	104.52	0.067	0.31	0.1290	678	1	Res
	105% - repetition 4	103.94	0.059	0.31	0.1190	575	1	Res
	115% - repetition 1	112.23	0.2	0.31	0.2080	1760	1	Res
	115% - repetition 2	107.69	0.11	0.31	0.1490	906	1	Res
	115% - repetition 3	114.94	0.27	0.31	0.2430	2410	1	Res
	115% - repetition 4	111.95	0.2	0.31	0.2050	1700	1	Res

	Sample	Predicted Content (%)	MD	Mah.Limit	Residuals	F Value	FProb	Outliers
Suppository spectra	80% - repetition 1	87.04	0.032	0.31	0.3130	4000	1	Res
	80% - repetition 2	87.29	0.029	0.31	0.3100	3910	1	Res
	80% - repetition 3	86.96	0.032	0.31	0.3140	4030	1	Res
	90% - repetition 1	91.24	0.0053	0.31	0.2410	2370	1	Res
	90% - repetition 2	90.97	0.0064	0.31	0.2410	2370	1	Res
	90% - repetition 3	90.87	0.0068	0.31	0.2450	2440	1	Res
	100% - repetition 1	92.91	0.00099	0.31	0.1860	1410	1	Res
	100% - repetition 2	93.99	1.9E-005	0.31	0.1830	1360	1	Res
	100% - repetition 3	93.69	0.00014	0.31	0.1850	1390	1	Res
	110% - repetition 1	95.83	0.0017	0.31	0.1400	801	1	Res
	110% - repetition 2	95.80	0.0017	0.31	0.1390	782	1	Res
	110% - repetition 3	95.77	0.0016	0.31	0.1380	778	1	Res
	120% - repetition 1	97.41	0.0065	0.31	0.0997	405	1	Res
	120% - repetition 2	96.80	0.0043	0.31	0.0995	403	1	Res
	120% - repetition 3	97.41	0.0065	0.31	0.0993	401	1	Res
	85% - repetition 1	89.53	0.013	0.31	0.2730	3040	1	Res

	Sample	Predicted Content (%)	MD	Mah.Limit	Residuals	F Value	FProb	Outliers
	85% - repetition 2	89.47	0.014	0.31	0.2730	3040	1	Res
	85% - repetition 3	89.20	0.015	0.31	0.2670	2890	1	Res
	95% - repetition 1	92.66	0.0014	0.31	0.2160	1890	1	Res
	95% - repetition 2	92.63	0.0015	0.31	0.2090	1790	1	Res
	95% - repetition 3	92.92	0.00096	0.31	0.2050	1720	1	Res
	105% - repetition 1	95.04	0.00048	0.31	0.1630	1090	1	Res
	105% - repetition 2	95.25	0.00073	0.31	0.1590	1030	1	Res
	105% - repetition 3	95.21	0.00068	0.31	0.1580	1010	1	Res
	115% - repetition 1	96.77	0.0042	0.31	0.1180	570	1	Res
	115% - repetition 2	96.89	0.0046	0.31	0.1190	574	1	Res
	115% - repetition 3	96.81	0.0043	0.31	0.1170	555	1	Res

Annexe 3 : Results of specificity analysis on commercial granules

Sample		Predicted API Content (%)	MD	Limit	Residuals	F Value	F Prob	Outlier
L16038D	Crushed	99.97	0.021	0.310	0.0021	0.180	0.325	-
	Uncrushed	100.45	0.025	0.310	0.0127	6.560	0.984	-
L16102A	Crushed	100.41	0.024	0.310	0.0053	1.160	0.710	-
	Uncrushed	100.55	0.025	0.310	0.0129	6.800	0.985	-
L17061B	Crushed	100.13	0.022	0.310	0.0055	1.240	0.725	-
	Uncrushed	100.22	0.023	0.310	0.0161	10.600	0.997	Res
M12034A	Crushed	99.72	0.019	0.310	0.0102	4.270	0.951	-
	Uncrushed	100.05	0.022	0.310	0.0132	7.100	0.987	-
M12195A	Crushed	99.85	0.020	0.310	0.0024	0.237	0.370	-
	Uncrushed	99.81	0.020	0.310	0.0146	8.710	0.994	Res
M12197A	Crushed	99.83	0.020	0.310	0.0083	2.800	0.894	-
	Uncrushed	100.15	0.022	0.310	0.0135	7.400	0.989	-
M12200A	Crushed	99.84	0.020	0.310	0.0044	0.785	0.617	-
	Uncrushed	100.09	0.022	0.310	0.0137	7.680	0.990	-
L14849A	Crushed	100.39	0.024	0.310	0.0038	0.574	0.545	-
	Uncrushed	100.47	0.025	0.310	0.0153	9.490	0.995	Res
L14850A	Crushed	100.21	0.023	0.310	0.0027	0.286	0.403	-
	Uncrushed	100.25	0.023	0.310	0.0141	8.090	0.992	Res
L16034B	Crushed	100.48	0.025	0.310	0.0079	2.550	0.878	-
	Uncrushed	100.43	0.024	0.310	0.0131	7.030	0.987	-

Université de Lille 2
FACULTE DES SCIENCES PHARMACEUTIQUES ET BIOLOGIQUES DE LILLE
DIPLOME D'ETAT DE DOCTEUR EN PHARMACIE
Année Universitaire 2017/2018

Nom : Comyn
Prénom : Hélène

Titre de la thèse : DEVELOPMENT OF AN ANALYTICAL METHOD FOR ACTIVE PHARMACEUTICAL INGREDIENT ASSAY BY NEAR INFRARED SPECTROSCOPY

Mots-clés : Analyse proche-infrarouge, méthode analytique, développement de modèle, analyse de données multivariées, validation analytique

---

Résumé :

Les méthodes spectroscopiques sont devenues des techniques incontournables dans le milieu de la Chimie, avec un immense champ de possibilités. Parmi celles-ci, la Spectroscopie Proche Infrarouge (NIRS) est une technique analytique en plein essor dans l'industrie pharmaceutique.

L'objectif du travail qui a supporté cette thèse était de développer une méthode d'analyse par spectroscopie proche infrarouge pour le dosage de principe actif dans un produit fini, à des fins de libération de lots. Cette nouvelle méthode, plus rapide, et plus économique, présente une alternative à la méthode de dosage actuelle par HPLC.

Ce projet inclut le développement du modèle de calibration, la comparaison du modèle obtenu avec la méthode de référence, et la validation de la méthode selon les réglementations internes et internationales relatives aux méthodes analytiques.

Cette thèse présente l'étude de faisabilité menée lors de la construction du modèle de calibration, ainsi que la stratégie de validation de la méthode et ses résultats.

Une méthode analytique satisfaisante a été développée, pour deux formes différentes du produit : sous forme granules ou poudres (résultant du broyage des granules). L'analyse de données multivariées a été utilisée pour la construction du modèle.

Le modèle a une précision (RMSECV) de 0.6% de la dose nominale, et a été démontré au moins équivalent à la méthode HPLC, si ce n'est meilleur en termes de précision.

La validation selon les recommandations ICH a été réalisée avec succès, tous les paramètres testés (linéarité, précision etc.) ayant rempli les critères.

Membres du jury :

Président : Jean-François Goossens

Assesseur(s) : Josselin Saunier

Membre(s) extérieur(s) : Samd Guizani, Process Scientist, Ferring Pharmaceuticals (Suisse)