SUIVI DU PROCÉDÉ DE COMPRESSION D’UN MELANGE DE POUDDRES PHARMACEUTIQUES: DÉVELOPPEMENT D’OUTILS ROBUSTES POUR UN SUIVI EN TEMPS RÉEL

Monitoring of multicomponent pharmaceutical powders in a compression process: development of a robust real time monitoring tools

Mémoire de doctorat
Spécialité: Génie Chimique

Pedro DURÃO

Jury: Ryan GOSSELIN (directeur)
Nicolas ABATZOGLOU (codirecteur)
Denis GROLEAU
Jean-Maxime GUAY
Jean-Michel LAVOIE
Rodolfo ROMAÑACH

Sherbrooke (Québec) Canada

Mai 2017
To my mother, to my father, to my sister
ABSTRACT

The way pharmaceutical industry develops and manufactures their products has been changing in recent years. The regulatory environment that they are obligated to comply has been pushing this change in order to endow this activity with state of art technology. The encouragement of the use of process analytical technology (PAT) to build the quality right from the design (QbD) is perhaps the most significant example of the new paradigm. The manufacturers are implementing this technology in new and existing products and benefiting from their advantages.

To implement PAT in a process, many steps must be taken (from the study of feasibility of the instruments until regulatory approval). This thesis describes the initial study (feasibility and model developments), prior to any submission of authorization, of the use of PAT tools (Near-Infrared (NIR), Red Green Blue (RGB) camera and Light Induced Fluorescence (LIF)) to monitor the compression process of a commercial multi-component blend.

After the potential of these tools was assessed, quantitative Partial Least Squares (PLS) models were able to be developed to monitor components with a concentration as low as 0.1 w/w % with a $R^2$ of 0.95. It was also proved that combining data from more than one tool was benefit for the accuracy of the model. The tools were also evaluated to their specificity by using a full factorial design where the models were built with simultaneous variations of concentration of some of the components. Even in this challenging case, the models built remained with an acceptable accuracy, considering the acceptance criteria used for dietary products such as multi-vitamins.

The work developed in this thesis contributed to the publication of 3 articles and 3 communications. Along with the proof of concept that it provided - which enlarged the opportunities for testing other probes - it also proved that is possible to monitor in-line the components in the feed frame. In this latter case, all the tools were accurate enough to monitor at least one component even if they are present in low concentration and part of multi-component blends.

Therefore, the industry can use this knowledge to monitor the compression process more adequately, increasing the range of tools used for the effect. Fundamental research can also be investigated as phenomena like segregation can be more accurately identified.

Keywords: PAT, LIF, NIR, RGB, Compression, PLS
RÉSUMÉ

La façon dont l'industrie pharmaceutique développe et manufacture ses produits a évolué au cours de ces dernières années. L'environnement réglementaire auquel elle est contrainte a provoqué ce changement dans le but de doter de technologies de pointe dans ses différentes activités. L’encouragement pour utiliser les technologies d’analyse de procédé (PAT) afin d’implémenter le concept de « Quality By Design » (QbD) est l’exemple le plus significatif de ce nouveau paradigme. Lentement, les industries implémentent ces technologies pour de nouveaux produits, mais également pour certains produits déjà existants, bénéficiant ainsi de leurs avantages.

Pour implémenter des PAT dans un procédé, plusieurs étapes doivent être franchies, de l’étude de faisabilité des instruments jusqu’à l’approbation réglementaire. Cette thèse décrit l’étude initiale (faisabilité et développement de modèles) avant toute demande d’autorisation d’utilisation d’outils PAT (proche infra-rouge (NIR), caméra RGB et fluorescence induite par laser (LIF)) pour suivre le procédé de compression d’un mélange commercial comprenant plusieurs ingrédients.

Après avoir établi le potentiel de ces différents outils, des modèles quantitatifs calculés par régression par moindres carrés partiels (PLS) ont été développés pour suivre les composants ayant une concentration aussi faible que 0,1 w/w%, avec un coefficient de détermination ($R^2$) de 0,95. Il a également été démontré que l’utilisation conjointe de données de plus d’un outil améliorait la précision du modèle. La spécificité de chacun des outils a également été évaluée à l'aide de plan d'expériences factoriels complets pour lesquels les modèles ont été construits en faisant varier simultanément la concentration de différents éléments. Même dans ces conditions, les modèles construits ont montré une précision acceptable, en considérant les critères d’acceptation utilisés pour les produits alimentaires comme les multivitamines.

Le travail présenté dans cette thèse a contribué à la publication de trois articles et de trois présentations orales. En plus de l’établissement de la preuve de concept, ce qui augmente les opportunités pour tester d’autres sondes, la possibilité de suivre en ligne la composition quantitative dans la ligne d’alimentation de la presse a également été établi. Dans ce dernier
cas, tous les outils sont suffisamment précis pour suivre au moins un des composants, même si celui-ci est présent en faible concentration et fait partie d’un mélange de plusieurs composants.

Conséquemment, l’industrie peut utiliser ses connaissances pour suivre le procédé de compression de façon plus adéquate en augmentant l’éventail des outils utilisés à cet effet. Une recherche fondamentale pourrait également investiguer divers phénomènes tels que la ségrégation, afin de mieux les comprendre.

**Mots-clés:** PAT, LIF, NIR, RGB, Compression, PLS
ACKNOWLEDGMENTS

It may sound a common place, but I am serious when I say that without any of the people listed below, I would never get this degree.

Firstly, I would like to thank my supervisor, Ryan Gosselin. It is an endless list, but thanks for all the incredible teaching, patience, support, belief in me and above all, friendship. I will always owe you, even if that day comes – the day when you finally realize that Portugal is an awesome country. I will surely miss the great mood, the uncomfortable feeling of my reflection when I enter your office (I hate that mirror!), the awesome computer desktop images and the “That’s all folks!”.

I would also like to thank my co-supervisor, Nicolas Abatzoglou, for his guidance throughout the project. I have learnt so much with his leadership, critical view and way of thinking.

I would like to thank Jean-Maxime Guay for all the availability and wise explanations that he always showed.

I shall not forget to thank the indispensable contributions of Henri Gauvin, Jacques Gagnier and Marc Couture for this project. All of you were remarkable and I will never forget your kindness and availability to help me.

I would like to thank all the graduated colleagues that helped me in this project: Oumaima (thanks for all the friendship and your patience to my “French”), Francis (how can I thank properly all the things you did for me?), Clémence (I really learn a lot from you, thanks for your dedication), Himmat (my partner, thanks for all the knowledge you passed me and proficuous debates), Emmanuel (I will never forget how you helped me on my arrival). I would also like to thank all the colleagues that took part of our weekly meetings and all the colleagues that later became friends that I met in this journey. It is impossible to list them all, but David, Mauricio, Tim, Jamie, Rajesh, Charlotte and Bárbara and Philippe, are certainly in this list.

I am very thankful to Jean-Sébastien Simard and Nicolas Abatzoglou for their managing of the Pfizer chair, for accepting me in the team and for giving me the financial and scientific support and conditions to develop this work. Your view of what a partnership between
academia and industrial must be, is a great example to the world and I am very proud of having been part of the team.

I also would like to show my gratitude to the members of the dissertation committee: Ryan Gosselin, Nicolas Abatzoglou, Denis Groleau, Jean-Maxime Guay, Jean-Michel Lavoie and Rodolfo Romañach.

I would like to thank Luísa, aka “Peste” for… basically everything. Her incredible long distance support, patience, determination, resilience and tolerance to me in Benfica issues, gave me the courage to overcome the obstacles one by one, even when Skype did not help. Awesome, right?

Finally, I would like to thank my parents, Maria Helena and José Joaquim, for their unconditional love, support, and understanding. I am really proud of you and I am sorry for making you use Skype as you never thought you would. Family appreciation is also extended to my sister Marta, who I really missed during my stay in Canada. Thanks for your support, understanding and the laughs you provided me in long distance calls.

“Merci à vous tous”.
CONTENTS

RÉSUMÉ .......................................................................................................................... iii
ACKNOWLEDGMENTS ................................................................................................. v
LIST OF TABLES ........................................................................................................... xi
LEXICON ....................................................................................................................... xiii
LIST OF ACRONYMS ..................................................................................................... xvii
CHAPTER 1 INTRODUCTION ................................................................. 1
  1.1 Context and problem to solve .............................................................................. 1
  1.1.1 Regulatory context ......................................................................................... 2
  1.1.2 Study of PAT in pharmaceutics ....................................................................... 3
  1.1.3 Advantages of monitoring the compression stage ............................................. 4
  1.1.4 Implementation of a PAT process ..................................................................... 5
  1.2 Definition of the research project ......................................................................... 6
  1.3 Project goals ......................................................................................................... 9
  1.4 Original contributions ......................................................................................... 9
  1.5 Document plan ..................................................................................................... 10
CHAPTER 2 STATE OF ART ................................................................. 13
  2.1 Background ........................................................................................................ 13
  2.1.1 Chemometrics ............................................................................................... 13
  2.1.1.1 Multivariate Data Analysis and Regression Methods ............................... 14
  2.1.2 Process Analytical Technology tools ............................................................... 23
  2.1.2.1 Imaging .................................................................................................... 23
  2.1.2.2 Near-Infrared spectroscopy ...................................................................... 27
  2.1.2.3 Light Induced Fluorescence .................................................................... 29
  2.2 Monitoring the compression stage – research topics ......................................... 32
  2.2.1 Influence of press parameters and powder properties ..................................... 33
  2.2.2 Imaging ......................................................................................................... 37
  2.2.3 Near-Infrared spectroscopy .......................................................................... 39
  2.2.4 Light Induced Fluorescence .......................................................................... 43
CHAPTER 3 Monitoring the concentration of flowing pharmaceutical powders in a tableting feed frame ................................................................. 47
  3.1 Introduction ........................................................................................................ 51
  3.1.1 Tableting ....................................................................................................... 51
  3.1.2 PAT tools ....................................................................................................... 51
  3.2 Materials and methods ....................................................................................... 52
  3.2.1 Set-up .......................................................................................................... 52
  3.2.2 Experimental design ..................................................................................... 54
  3.2.3 Software ....................................................................................................... 55
  3.2.4 Data treatment and presentation .................................................................... 56
  3.3 Results and discussion ....................................................................................... 57
  3.3.1 NIRS ......................................................................................................... 57
CONTENTS

CHAPTER 4 Using multiple Process Analytical Technology probes to monitor multivitamin blends in a tableting feed frame .................................................. 65
4.1 Introduction ............................................................................. 69
4.1.1 Application of PAT tools in the compression stage .................. 70
4.2 Materials, methods and techniques ......................................... 71
4.2.1 Experimental feed frame set-up ......................................... 71
4.2.2 Component selection and description .................................. 73
4.2.3 Experimental design .......................................................... 74
4.2.4 Data treatment ................................................................. 75
4.2.5 Dynamics ........................................................................ 77
4.2.6 Chemometrics .................................................................. 77
4.3 Results and discussion ........................................................... 80
4.3.1 Model building ................................................................. 80
4.3.2 Dynamics ........................................................................ 87
4.4 Conclusion ........................................................................... 90

CHAPTER 5 Specificity of Process Analytical Tools in the monitoring of multicomponent pharmaceutical powders ............................................... 91
5.1 Introduction ........................................................................... 95
5.1.1 Design considerations ....................................................... 96
5.1.2 Specificity of PAT tools ...................................................... 97
5.2 Materials and methods ........................................................... 98
5.2.1 Equipment ...................................................................... 98
5.2.2 Selection and description of components to analyse .......... 99
5.2.3 Design of experiments ...................................................... 100
5.2.4 Batch preparation and data acquisition .............................. 101
5.2.5 Data treatment ................................................................. 101
5.3 Results and discussion ........................................................... 101
5.3.1 Models building and selection .......................................... 101
5.3.2 Analysis of the residual error in interactions investigation ........ 105
5.3.3 Determination of the interactions ...................................... 108
5.4 Conclusion .......................................................................... 112

CHAPTER 6 CONCLUSION .................................................................. 114

ANNEXES .................................................................................... 118
REFERENCES ............................................................................. 125
LIST OF FIGURES

Figure 1.1: Chemical structure of the components to monitor .................................................. 7
Figure 2.1: PCA design ............................................................................................................ 16
Figure 2.2: Reorganization of an image – 3-Dimensional matrix to a 2-Dimensional matrix. 24
Figure 2.3: Calculation of score images .................................................................................. 25
Figure 2.4: Representation of the first pixel in the score plot .................................................. 26
Figure 2.5: Score plot and false composite image of softwood lumber .................................. 26
Figure 2.6: Overtones and combinations bands in NIR frequency .......................................... 28
Figure 2.7: Luminescence Jablonski diagram ........................................................................ 29
Figure 2.8: Five concentrations of a caffeine-lactose mixture a) NIR; b) LIF ......................... 30
Figure 2.9: Fluorescence of riboflavin in ethanol ................................................................... 31
Figure 2.10: VisioNIR sensor to analyze tablets .................................................................... 41
Figure 2.11: NIR probe installed in the feed frame ................................................................. 42
Figure 3.1: Feed frame set-up .................................................................................................. 53
Figure 3.2: Wave behavior of the feed frame .......................................................................... 54
Figure 3.3: Original acquisitions of the five mixtures (a) NIR; (b) RGB images; (c) LIF ...... 56
Figure 3.4: Unfolding of the 3-D RGB matrix into a 2-D matrix ........................................... 57
Figure 3.5: NIRS results ........................................................................................................ 58
Figure 3.6: LIF results ............................................................................................................. 59
Figure 3.7: RGB results .......................................................................................................... 61
Figure 3.8: Details of the first transition of RGB results ......................................................... 62
Figure 4.1: Feed frame set-up .................................................................................................. 72
Figure 4.2: Composition and organization of prepared batches ............................................ 74
Figure 4.3: Schematic representation of transition with a step change technique .................. 76
Figure 4.4: Mathematical development of mbPLS ................................................................. 79
Figure 4.5: Score plot of the PLS model of ferrous fumarate (NIR data) ............................... 81
Figure 4.6: PLS model (NIR): observations vs predicted concentration of ferrous fumarate .. 82
Figure 4.7: Predicted value of ferrous fumarate with a PLS model using data from (a) NIR and (b) all tools .......................................................................................... 83
Figure 4.8: Detailed transition of ferrous fumarate with a PLS model ................................. 88
Figure 4.9: k values and overall value of each studied vitamin .............................................. 89
Figure 5.1: % of concentration for each ingredient by level ................................................. 100
Figure 5.2: Score plot of the PLS model of ginseng developed from NIR data .................. 102
Figure 5.3: Comparison between adjusted $R^2_{\text{calibration}}$ values obtained for the same ingredients by different calibration designs ........................................................................................................ 105
Figure 5.4: Calibration (a) and validation (b) curve of riboflavin (LIF); calibration (c) and validation (d) curve of ginseng (LIF) ........................................................................................................................................ 109
Figure A.1: NIR probe dimensions .................................................................................................................. 118
Figure A.2: Feed frame radius .......................................................................................................................... 119
Figure A.3: Scheme of feed frame wheel with NIR probe and its acquisitions sizes .................. 120
Figure A.4: Acquisition area of the probe .......................................................................................................... 121
LIST OF TABLES

Table 2.1: Published literature concerning the compression stage .................................................34
Table 3.1: Experimental design........................................................................................................55
Table 3.2: First-order k values and 95% CI for every transition ..................................................63
Table 4.1: Equipment conditions .....................................................................................................72
Table 4.2: Reference values of tracked vitamins ..........................................................................73
Table 4.3: Adjusted $R^2$ values for predictive models of the vitamins .........................................84
Table 4.4: BIP values of data combined for the vitamins in study ..............................................86
Table 5.1: Parameters of acquisition and sample spot size ............................................................98
Table 5.2: Concentration ranges of ingredients under study .......................................................99
Table 5.3: Adjusted $R^2$ values of models built ............................................................................103
Table 5.4: Importance (%) and significance of each parameter in the selected models ............106
Table 5.5: Possibilities of interference from Table 4 .................................................................106
Table 5.6: t-test of the riboflavin model with LIF data (calibration) ........................................110
Table 5.7: t-test of the riboflavin model with LIF data (validation) ........................................111
Table 5.8: t-test of the ginseng model with LIF data (calibration) .............................................111
Table 5.9: t-test of the ginseng model with LIF data (validation) .............................................111
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Die</td>
<td>Cavity in which the powder coming from the feed frame enters to be compressed</td>
</tr>
<tr>
<td>Feed Frame</td>
<td>Part of the press machine that distributes the powder coming from the hopper to the dies</td>
</tr>
<tr>
<td>Hopper</td>
<td>Equipment that is filled with the powder that is supplied to the press</td>
</tr>
<tr>
<td>Paddle</td>
<td>Piece of the feed frame that pushes the powder from the hopper into the dies</td>
</tr>
</tbody>
</table>
# LIST OF SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_i )</td>
<td>Additive correction factor</td>
</tr>
<tr>
<td>( A )</td>
<td>Number of components / matrix / gain</td>
</tr>
<tr>
<td>( B )</td>
<td>Total number of blocks</td>
</tr>
<tr>
<td>( b_i )</td>
<td>Multiplicative correction factor</td>
</tr>
<tr>
<td>( b_p )</td>
<td>Vector of regression coefficients (( p = 1, 2, \ldots ))</td>
</tr>
<tr>
<td>( D )</td>
<td>Scaling metric</td>
</tr>
<tr>
<td>( E )</td>
<td>Matrix of noise</td>
</tr>
<tr>
<td>( J )</td>
<td>Width of an image</td>
</tr>
<tr>
<td>( k )</td>
<td>Rate constant ((s^{-1}))</td>
</tr>
<tr>
<td>( K )</td>
<td>Number of components / depth of an image</td>
</tr>
<tr>
<td>( M )</td>
<td>Observations (number of rows)</td>
</tr>
<tr>
<td>( m_i )</td>
<td>Mean of spectral measurements of sample ( i )</td>
</tr>
<tr>
<td>( N )</td>
<td>Number of variables / height of an image</td>
</tr>
<tr>
<td>( n )</td>
<td>Index of ( Y ) blocks</td>
</tr>
<tr>
<td>( P )</td>
<td>Matrix of loadings / matrix of normalized eigenvectors</td>
</tr>
<tr>
<td>( p_\alpha )</td>
<td>Loading of the ( \alpha )th component</td>
</tr>
<tr>
<td>( q )</td>
<td>Loading of the ( Y ) matrix</td>
</tr>
<tr>
<td>( Q )</td>
<td>Matrix of normalized eigenvectors</td>
</tr>
<tr>
<td>( \text{rpm} )</td>
<td>Rotations per minute</td>
</tr>
<tr>
<td>( s )</td>
<td>Seconds</td>
</tr>
<tr>
<td>( S_i )</td>
<td>Standard Deviation of ( k ) measurements (( k = 1, 2, \ldots ))</td>
</tr>
<tr>
<td>( t )</td>
<td>Time</td>
</tr>
<tr>
<td>( T )</td>
<td>Matrix of scores / score image</td>
</tr>
<tr>
<td>( t_\alpha )</td>
<td>Score in the ( \alpha )th component</td>
</tr>
<tr>
<td>( t_i )</td>
<td>PCA score</td>
</tr>
<tr>
<td>( u )</td>
<td>Score of the ( Y ) matrix / vector (eigenvector)</td>
</tr>
<tr>
<td>( \mu_{b,a} )</td>
<td>Weight vector of the ( a )th component</td>
</tr>
<tr>
<td>( w )</td>
<td>Loading weight</td>
</tr>
<tr>
<td>( W )</td>
<td>Super weight</td>
</tr>
<tr>
<td>( x )</td>
<td>Independent variable</td>
</tr>
<tr>
<td>( x_{ik} )</td>
<td>Spectral measurement of sample ( I ) at ( k )th wavelength</td>
</tr>
<tr>
<td>( X )</td>
<td>Matrix of data with ( M ) rows and ( N ) columns (( M, N = 1, 2, \ldots ))</td>
</tr>
<tr>
<td>( y )</td>
<td>Dependent variable</td>
</tr>
<tr>
<td>( \hat{y} )</td>
<td>Predicted relative concentration</td>
</tr>
<tr>
<td>( Y )</td>
<td>Predictive matrix with ( M ) rows and ( N ) columns (( N, M = 1, 2, \ldots ))</td>
</tr>
<tr>
<td>( \Lambda )</td>
<td>Eigenvalue</td>
</tr>
<tr>
<td>( \Delta )</td>
<td>Diagonal matrix of singular values</td>
</tr>
<tr>
<td>( \Lambda )</td>
<td>Diagonal matrix of eigenvalues</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>API</td>
<td>Active Pharmaceutical Ingredient</td>
</tr>
<tr>
<td>BIP</td>
<td>Block Importance in Prediction</td>
</tr>
<tr>
<td>CPCA</td>
<td>Consensus Principal Component Analysis</td>
</tr>
<tr>
<td>CPP</td>
<td>Critical Process Parameters</td>
</tr>
<tr>
<td>CQA</td>
<td>Critical Quality Attributes</td>
</tr>
<tr>
<td>DEM</td>
<td>Discrete Element Method</td>
</tr>
<tr>
<td>DoE</td>
<td>Design of Experiments</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>HPLC</td>
<td>High Performance Liquid Chromatography</td>
</tr>
<tr>
<td>LED</td>
<td>Light-emitting diode</td>
</tr>
<tr>
<td>LD</td>
<td>Laser diode</td>
</tr>
<tr>
<td>LIF</td>
<td>Light Induced Fluorescence</td>
</tr>
<tr>
<td>LV</td>
<td>Latent Variables</td>
</tr>
<tr>
<td>mbPLS</td>
<td>Multiblock Partial Least Squares</td>
</tr>
<tr>
<td>MIA</td>
<td>Multivariate Image Analysis</td>
</tr>
<tr>
<td>MLR</td>
<td>Multi Linear Regression</td>
</tr>
<tr>
<td>MSC</td>
<td>Multiplicative Signal Correction</td>
</tr>
<tr>
<td>MVDA</td>
<td>Multivariate Data Analysis</td>
</tr>
<tr>
<td>NIPALS</td>
<td>Nonlinear Iterative Partial Least Squares</td>
</tr>
<tr>
<td>NIR</td>
<td>Near-Infrared</td>
</tr>
<tr>
<td>PAT</td>
<td>Process Analytical Technology</td>
</tr>
<tr>
<td>PC</td>
<td>Principal Component</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>Principal Component Regression</td>
</tr>
<tr>
<td>PLS</td>
<td>Partial Least Squares</td>
</tr>
<tr>
<td>PSD</td>
<td>Particle Size Distribution</td>
</tr>
<tr>
<td>Q</td>
<td>Q – Residuals</td>
</tr>
<tr>
<td>QA</td>
<td>Quality Assurance</td>
</tr>
<tr>
<td>QbD</td>
<td>Quality by Design</td>
</tr>
<tr>
<td>R²</td>
<td>Goodness of fit</td>
</tr>
<tr>
<td>RGB</td>
<td>Red Green Blue</td>
</tr>
<tr>
<td>RMSEC</td>
<td>Root Mean Square Error Calibration</td>
</tr>
<tr>
<td>RMSECV</td>
<td>Root Mean Square Error Cross Validation</td>
</tr>
<tr>
<td>RMSEP</td>
<td>Root Mean Square Error Prediction</td>
</tr>
<tr>
<td>RTD</td>
<td>Residence Time Distribution</td>
</tr>
<tr>
<td>RTR</td>
<td>Real Time Release</td>
</tr>
<tr>
<td>SNV</td>
<td>Standard Normal Variate</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>T²</td>
<td>Hotelling’s distance</td>
</tr>
<tr>
<td>UDU</td>
<td>Uniformity Dosage Unit</td>
</tr>
<tr>
<td>UnV</td>
<td>Unit Variance</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra-violet</td>
</tr>
</tbody>
</table>
CHAPTER 1 INTRODUCTION

1.1 Context and problem to solve

Pharmaceutical industry is the subject of a strict legislation and regulation environment put in place to ensure that a quality product is delivered to the market. Regulatory agencies such as the Food and Drug Administration (FDA) in the United States, Health Canada and the European Medicine Agency, through the publication of checklists and guidelines, demand an extensive list of requirements right from the beginning of the development of a new pharmaceutical product up to post-commercialization monitoring. These demands require manufacturers to implement a system that facilitates the fulfillment of these requirements and, at the same time, to have the best possible tools to guarantee the production of a quality product while being able to keep economical competitiveness. Due to this regulatory environment and the difficult challenge that manufacturers face to fulfill it, a modification in the regulatory paradigm is not often promoted by the agencies, who have been historically conservative. In addition, a significant change in the processes, especially in products already approved, are not promoted by manufacturers, which must follow the recommendations of their regulators and therefore are reluctant in submitting modifications in their already approved processes and products. This means that only a breakthrough event (new findings, new technology) by either part could lead to a significant change in the regulatory concepts and / or modifications of processes already approved.

The implementation of Process Analytical Technologies (PAT) in pharmaceutics must be considered one of these events. Although their advantages had been recognized in many other industries before, once pharma regulatory agencies (especially the FDA) recommended their application, which converged with manufacturers growing will in adopting it, the paradigm changed significantly. At the same time the use of PAT triggered a new regulatory situation in which its application is critical to the fulfillment of requirements, scientists began studying and developing its use in many different processes, which has been leading to a boost in the PAT publications adapted to pharma. Therefore, the need to comply with the regulatory situation and the benefits for the manufacturer in applying PAT to their processes, concur for the pertinence of this work.
1.1.1 Regulatory context

After the publication by the FDA of the guideline [1] that encourages the use of PAT in the development and manufacturing of pharmaceutical products, the way pharmaceutical industry has been working on their processes has changed. Along with the publication of Q8(R2) [2], Q9 [3] and Q10 [4] guidelines and later with Q11 [5], these documents drove pharmaceutical industry towards a new scenario in which the focus was no longer in the final product but in the knowledge and understanding of the processes, following a risk-based approach. The concept lies in the assumption that the risk should be identified, mitigated and continuously evaluated. As mentioned in Q8(R2), the information obtained throughout the life cycle of a product should be included in the process guaranteed by the quality system implemented by the manufacturer. Continuous improvement is a policy that must be followed in this industry.

In this new paradigm, PAT plays a critical role as it allows gathering information in different steps that are critical to understand the underlying phenomena that are occurring. In this strategy, not only it helps in the identification of the risks but in their mitigation, monitoring, control and revaluation. In fact, it is intended to implement the concept of quality right from the early stages of development (Quality by Design (Qbd) [1]), in which a design space and a control strategy in the manufacturing stage are perfectly justifiable and a product is made within the defined specifications [6] limited by the control space (usually controlled by PAT tools). The advantages lie in the increase of safety and efficacy of the drugs and also in the efficiency of their manufacture chain as quality is improved and economical savings are obtained. This new paradigm was later reinforced through the publication of the guideline on process validation, in which the data coming from this technology could be used to validate the process [7]. The traditional approach of 3 batches validation is now complemented by what is called a continuous validation system.

In spite of the attempt to adjust all the regulation to this new paradigm, the extensive set of documents ruling pharmaceutical activity that has not been updated, has raised issues about compliance when PAT is implemented, especially as monitoring tools. In fact, some of the requirements to control a batch are adjusted to the traditional techniques and thus, are inadequate to the PAT paradigm, which has been a drawback in its implementation. One of the typical examples occurred in tablet analysis. For many years it was only possible to comply
with the Uniformity of Dosage Units (UDU) test if 30 tablets were tested and no sample was outside the ± 25% of label claim. This statement did not take into account any statistical normal distribution and due to the larger number of samples obtained by PAT, the possibility of having a sample out of specification was very high. Therefore, until the implementation of a new regulation in March 2013 [8] that took into account the sampling size (chapter 2.9.47 of European Pharmacopeia – Demonstration of uniformity of Dosage Units using Large Sample Sizes), the use of PAT was somehow restricted in this case. Since a discussed strategy and the ultimate aim of this work is to confirm tablets conformity by analysing the powder that leads to their constitution, this chapter must be taken into account in a future implementation of the process. In the case some of the methods and recommendations stated in the regulatory guidelines are difficult to fulfill using PAT methods, the critical point is to guarantee quality of the product. Therefore, it is the manufacturer’s function to develop and prove that their methods are accurate and equivalent to the ones regulation considered to be the reference.

1.1.2 Study of PAT in pharmaceutics

Although the use of PAT in the development of new products allows taking full advantage of this technology, the application of PAT as a control tool in the manufacture of products which are already present in the market represents a valuable resource for this technology as well. For instance, the concepts of continuous manufacturing rely in the correct implementation of this technology throughout the manufacturing steps [9].

The implementation of techniques to monitor the critical quality attributes (CQA) of the product either by monitoring directly or through the control of the critical processes parameters (CPP) is never easy and comprises several steps. Due to the specificity of processes and products, it is hard to use methods beyond their intended use even if they were previously applied, since PAT usually depends on the development of a specific model for each case. However, previous knowledge might provide some hints of which CQAs / CPPs must be monitored, the tools that can be used and / or the restraints that must be considered.

Since tablets are the most common pharmaceutical dosage form, the development of techniques that can be used to monitor their manufacturing steps (granulation, drying, blending or coating) has been the subject of several studies that are reported in the literature. On the other hand, not many studies have focused in the compression stage. In the literature,
monitoring of the compression stage has been divided in a) analysis of the intact tablet either in or off-line and b) monitoring the circulating powder in the feed frame of a press. The latter will be the focus of this work as many advantages can be listed.

1.1.3 Advantages of monitoring the compression stage

Despite the fact that the powder must be controlled in the preceding stages, especially its concentration and homogeneity, a fast and reliable measurement of its constitution at this stage might be important for several reasons. Thus, a simple identification of the components that constitute the powder entering the press machine might detect some errors occurring in previous stages, such as an error in formulation manufacturing (absence or wrong selection of a component, for instance). Although these episodes are rare, they can still occur, especially when a direct compression technique is applied, since it usually comprises just a simple mixing before compression. Another advantage lies in the possibility to quantify one or more components and guarantee the homogeneity of the powder that is delivered to the dies overcoming at the same time the identified problem of the representativeness of the sample obtained from thief sampling after blending [10]. In fact, identification of non-homogeneity whether segregation is present or not, is of the utmost importance at this stage. Moreover, the occurrence of the latter is potentiated during the transport to the hopper of the press and its discharge to the feed frame [11] (both are usually made by gravity). In addition, also contributing to this phenomenon is the mixing mechanism occurring inside the feed frame of the press [12], especially in multicomponent blends that due to their higher number of components, have a higher probability to have differences between them (especially particle size and shape). Thus, the use of monitoring tools might detect the occurrence of material that do not meet product specifications, allowing the person in charge to control the process, keeping the tablets under specification. A feedback control mechanism can be implemented based in the data obtained by these tools - for instance, through an installation of a controller that can stop immediately the press machine or in the case of continuous manufacturing, it can “communicate” with other equipment by adjusting their process parameters (speed rotation in a blender, for instance). The quantification of the component might also be used as a complement (in a first stage) to the uniformity dosage unit (UDU) test, usually performed at the end of the batch. As batch historical data is obtained, the suppression of this test by
traditional methods such as Ultra-Violet (UV) or High Performance Liquid Chromatography (HPLC), which are costly and do not provide immediate results, might be considered. This will allow the manufacturer to release the product immediately after manufacturing - **Real Time Release – RTR** (assuming the criteria for acceptance is met after analysis of data obtained through PAT tools) [13]. The quantification of the components to ensure the homogeneity of the blend is even more important when a certain Active Pharmaceutical Ingredient (API) with a narrow therapeutic margin is present (usually in low concentrations). The adequate monitoring allows important ensuring in the clinical safety for these types of formulations. Another advantage lies in the improvement in terms of quality that the implementation of PAT tools represents at this stage due to sampling available for the entire batch. In fact, regulation obligates the manufacturer to test up to 30 tablets (usually taken at different times during the batch) to evaluate their UDU [14] if traditional methods are used, independently of the number of tablets produced by batch. In this technique, since the powder present in the feed frame is subject to continuous analysis from the tools in place, it is possible to assess the uniformity required for the tablets. Moreover, in terms of the total mass of powder tested (in comparison with the sum of the mass of the 30 tablets required by regulation) and also in terms of the limit of 25 previously mentioned it is possible to test a larger quantity in comparison with traditional methods, allowing a stricter control of the quality of the batch [8] (consult Annex A.1 for the calculation in this case).

Due to the benefits listed above, a need to develop a technique to monitor effectively the powder in the compression stage is perfectly identified. Although the specificity for a product is recognized, the evaluation of the proposed set-up and the potential of the selected monitoring tools to be used in PAT field can be extrapolated and might represent a breakthrough in the monitoring of this manufacturing step.

1.1.4 Implementation of a PAT process

The implementation of a PAT analytical technique intended to monitor and control a process goes through different stages. The first step is the recognition of the parameter, or attribute, to be monitored. Based on its characteristics and on the features of the analytical technique, a tool is proposed and a feasibility test must be performed (a preliminary test whose goal is to confirm the potential of the tool to monitor the parameter or attribute). The
following step includes the qualification of the analytical technique (installation, operation and performance). During this stage, analytical standards are used and the stability of the analytical technique must be assessed. For NIR, the wavelength and photometric scale must be validated [15] and parameters such as wavelength uncertainty, photometric linearity and spectrophotometric noise must be determined. Once the equipment is qualified, the method must be validated, that is, it must “demonstrate that the analytical technique is suitable for the intended purpose” [15]. Parameters such as specificity, linearity, range, accuracy, precision and robustness must be evaluated. These parameters are usually assessed using analytical reference data and statistics. The model for the parameter in study must be developed with an adequate design, to guarantee its robustness and after validation, it can be submitted to the regulator. Due to its paradigm, the PAT monitoring process is considered a continuous validation process.

In this thesis, only the first part of this process (feasibility test) is developed. The models developed in this thesis are part of this stage and are not intend to be used for any type of submission.

1.2 Definition of the research project

As mentioned in the previous section, the development of a methodology (proposed set-up, tools, data acquisition and analysis strategy) that can accurately monitor in real time the concentration of powder in the feed frame of a tablet press, is of the utmost importance, not only because of the regulatory demand but also due to the advantages listed. The objective of this thesis is: to use PAT tools as a process control strategy. Three different analytical techniques are proposed: Near-Infrared (NIR) spectroscopy, Light Induced Fluorescence (LIF) and Red Green and Blue (RGB) Imaging (camera), installed in the feed frame of a press machine and constituted by two intermeshing wheels. In a configuration like this, one of the wheels is responsible to immediately direct the powder that falls under gravity from the hopper into the feed frame to the dies, while the other recirculates the exceeding powder and ensures the filling requirements of the dies. Wahl et al. [16] stated that due to the larger residence time, the powder that reaches the second wheel has been more stressed and therefore is more likely to suffer phenomena like de-mixing and segregation. For these reasons, the authors considered that spectra of the powder obtained over the first wheel are more
representative of the constitution of the tablet and therefore, the probes must be placed in this position. However, since segregation is more likely to occur over the second wheel, placing the probes at this location might ensure product uniformity throughout the batch. In this work, the design of the press machine, specifically the position of the feed frame, made it impossible to install the probes over the first wheel and, therefore, the PAT tools were installed over the second wheel. A trigger was installed in the set-up, so the area of analysis remained the same to: a) mitigate the effect of potential variation due to the change of properties of the powder that can occur due to higher shear stresses (particle size and shape) and b) to remove any biased data that could happen due to paddle detection (especially in spectroscopic tools – although its influence is usually removed by pre-processing methods).

Throughout this thesis, these analytical techniques were selected to monitor the concentration of 5 selected ingredients in a multicomponent / vitamin blend immediately before being compressed into tablets, which therefore gives a reasonable assumption of the constitution of the tablet. Considering previous studies that highlight potential segregation phenomena that might occur in this stage (see Section 2.2) and the potential replacement or reduction of final tablet testing (for example, UDU testing), quantification of the ingredients is of the utmost importance. Since it is not feasible to monitor all the 31 components of the blend, the components selected were the vitamins ascorbic acid, beta-carotene and riboflavin and the supplements ferrous fumarate and ginseng extract (Figure 1.1).

![Chemical structure of the components to monitor](image)

*Figure 1.1: Chemical structure of the components to monitor: a) ascorbic acid; b) beta-carotene; c) riboflavin; d) ferrous fumarate; e) ginsenoside (major constituent of ginseng)*
They were selected based on their properties and the characteristics of the tools. Analysis of the chemical structure shows several O-H (ascorbic acid, riboflavin, ginseng), N-H (riboflavin) and C=O (ascorbic acid, riboflavin, ferrous fumarate) bonds, which are usually seen in NIR. It also shows some constituents with several aromatic rings (riboflavin, ginseng) that are usually seen by LIF and some of them present a distinguishable color (yellow for riboflavin, reddish brown for beta-carotene and ferrous fumarate and light brown for ginseng). The decision to monitor the concentration of the components and not any other parameter was due to the fact that at this stage, the concentration (and therefore, homogeneity) of the blend is usually an identified CQA. These will allow guaranteeing tablet uniformity and ensuring with more safety that UDU for these components is met throughout batch manufacture.

Although this commercial blend is not regulatory considered to be a pharmaceutical product, its manufacture obeys to pharmaceutical requirements. For the purpose of this study, it is more adequate than a traditional pharmaceutical formulation due to the number of ingredients it has on its constitution. In fact, the complexity of the blend chosen to be monitored, allows extrapolating these findings to simpler cases in which less components are present (if these tools are successfully used to monitor these components in a complex blend, in cases where the number of components is lower – which is the case of the majority of pharmaceutical dosage forms - it is expected to perform even better). Moreover, since it is a complex blend and it is impossible to track all the components, marker constituents might be defined to extrapolate the monitoring to other components. In addition, the possibility to use more than one tool simultaneously for the same process was evaluated. This factor is of the utmost importance because many situations might require more than one monitoring tool: in cases where the formulation has more than one component required to be monitored (i.e., a formulation with more than one API) that are not responsive to the same tool; in the case different CQA are required to be monitored (i.e., the concentration of a component and humidity level of the mixture). For the purpose of this work it is also intended to evaluate if the simultaneous monitoring of a component by more than one tool, enhances its prediction and the advantages that come from their simultaneous use. At the same time, this work also assesses the potential of these PAT tools to work in-line, in this manufacturing stage or in any other, if the required conditions for these tools to work, are met (the component must be NIR responsive, fluorescent and show a distinctive color, respectively).
1.3 Project goals

This thesis has one general objective that can be decomposed in several secondary objectives (defined in bullets):

*Develop a methodology that is able to accurately monitor in-line the concentration of the powder inside the press machine:*

- Investigate the feasibility and potential of a set-up to be used in the in-line monitoring of the compression process;

- Investigate the potential of RGB camera and LIF tool (NIR had already been studied) to work as a PAT tool in the in-line monitoring, specifically in a feed frame of a press;

- Develop models and test their robustness for selected components that can accurately predict their concentration, even for low dose components;

- Propose strategies that can improve the accuracy of those models, mainly through chemometric techniques (ex: combining data from several tools);

- Draw conclusions about the extrapolation to other processes, taken into account the conditions of this project, mainly the multicomponent /vitamin constitution of the powder used;

- Evaluate the capacity of the tools to study the kinetics of the powder;

1.4 Original contributions

This project produced 3 different papers which included some original contributions:

*Demonstration and suggestion of the practical applicability of LIF and RGB camera in the in-line monitoring of the compression stage:* Although NIR was also used in this project, its applicability as a PAT tool in many other processes and at this stage of the manufacturing process is well known (most of the PAT monitoring in the literature is done using NIR). Nevertheless, the in-line NIR probe used in this project had never been used to monitor a complex blend. On the other hand, the in-line applicability of LIF is a novelty in this project as well as the use of a RGB camera used in-line at this manufacturing stage. Taking advantage of the different features of the components present in the selected blend (the distinguishable color
of the component, necessary to be detected by the RGB camera or their fluorescent properties which is critical for LIF use) it was possible to show the utility of these tools in the PAT context.

**Possibility to quantify in-line low dose components in a feed frame of a tablet press even for multicomponent blends:** Low dosage component quantification is a difficult task even for off-line techniques, which usually involves sample preparation to isolate the components intended to be quantified. The difficulty of that measurement increases when that component is present in a multicomponent blend due to signal interference. The possibility that such a concentration can be detected in-line is a breakthrough in process monitoring and opens the possibility that RTR can be applied in blends that have these types of components, for instance, contraceptive medicines. In this project, among the tools used, LIF demonstrated to be a suitable tool in monitoring low dosage component.

**Demonstration of the benefits of using simultaneous tools in the monitoring of one component:** The PAT studies reported in the literature are usually presented by testing a single tool to monitor a certain parameter (or CQA). In this project, the usefulness of combining two or more tools was demonstrated, since the accuracy of the models developed for quantification of the components was improved when data from both tools were combined. This fact might be especially relevant for complex blends and for low dosage components if the tools are not sufficiently responsive to the component they intend to monitor.

### 1.5 Document plan

This thesis is divided in two different parts: State of art and project development.

In the first part (chapter 2), some critical concepts used throughout this thesis are introduced, followed by a state of art review relatively to the PAT tools used in the work. Taking into account the nature of this project, the work that has been done on the compression stage is highlighted.

The second part, which comprises the project development is constituted of three different chapters (3, 4 and 5). In chapter 3, a qualitative evaluation of the potential of the tools to detect changes in concentration is performed. Thus, it is intended to confirm the suitability of these tools and to validate the set-up by ensuring that the position where these tools were installed (feed frame) was adequate to detect those variations. Chapter 4 is intended to
evaluate the behavior of these tools in a real scenario, since the data was obtained in a running commercial (deactivated) press, using for the effect a marketed multicomponent blend. The quantification models of some of the components, the benefits of combining data obtained from more than one tool and the dynamics of the powder inside the feed frame were also assessed. Finally, in chapter 5, it is intended to evaluate the selectivity of each tool, considering concomitant interference between components. Thus, quantification models of these components were developed to evaluate their robustness when present in a complex blend.
CHAPTER 2  STATE OF ART

This project encompasses many interrelated concepts. One of them is the link between PAT applications and Multivariate Data Analysis (MVDA) due to the amount of data that is generated through this technology and the benefits of interpreting them through mathematical strategies. The first part of this chapter provides the concepts that are necessary to understand this work. Therefore, the basics of chemometric techniques (section 2.1.1) that will be used in this thesis are presented, followed by the concepts of the PAT tools used (section 2.1.2). The mathematical transformation required in imaging analysis is highlighted. Finally, and taking into consideration the overall goal of the project (that is, in-line monitoring of the compression stage by installing in the press feed frame three different tools – NIR, RGB camera and LIF), the PAT application of these analytical techniques is reviewed, with special emphasis on their use at the compression stage (when reported).

2.1 Background

2.1.1 Chemometrics

The amount of information obtained from various analytical techniques, whether they are “off” (“sample analyzed away from the process”), “at” (“sample analyzed in closed proximity to the process”), “on” (“sample is diverted from the process and may return after being analyzed”) or “in” (“sample is not removed from the process”)—line in a process [17], brought a “problem” to the pharmaceutical industry: its analysis. The development of tools allowing the establishment of some relations between the studied variables and / or their relation with the final product, became important to accurately determine the analyzed parameters, speed up the process and take full advantage of the PAT benefits. So, chemometrics became part of the pharmaceutical industry world, as its concept not only relate to the use of statistical methods for reduction and extraction of useful information present in multivariate data that can be used to monitor and predict specifications, but also to the capability to design and select optimal measurement procedures and experiments [18]. Thus, chemometrics are useful to detect relationships between process attributes and final product specifications, enhancing process understanding which leads to an efficient control of pharmaceutical processes [19].
Several chemometric techniques have been developed in recent years [20]. These techniques are much more important as they extract quality information that can be used for the manufacturer to assure final product quality. The advantage of monitoring a pharmaceutical process is not only to comply with the quality assurance (QA) requirements, but also to allow the manufacturer to continuously verify that the process follows the predicted course, comparing it to a previously validated model and to modify the process in due time if something unexpected occurs. The following topics cover the chemometric techniques that will be used throughout this work.

2.1.1.1 Multivariate Data Analysis and Regression Methods

The huge amount of data that analytical techniques provide makes it necessary to implement some mathematical strategies in order to analyze them. The development of a model that must be validated before being industrially applicable and that will allow that a certain parameter or specification can be predicted, is one of the goals of data treatment.

When different variables are present, one of the most used mathematical regression approaches is the Multi Linear Regression (MLR), where the independent variables predict the parameter(s) in study (dependent variable(s)), according to Equation 2.1:

\[
y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_i x_i \quad \text{Equation 2.1}
\]

where \( y \) = dependent variable; \( b_i \) = regression coefficient; \( x_i \) = independent variable

However, this technique is not appropriate when dealing with the kind of information usually obtained by PAT tools, since in these cases, many variables either are correlated or influence each other, leading to an unreliable prediction of the \( y \). Therefore, the use of latent variable models, which consist of data reduction through the establishment of relations between variables, is usually the resource used, making MVDA one of the most important strategies in this area. The combination of the variables present in original data infers mathematically the formation of a lesser number of variables called “Latent Variables” (LV) which is precisely the foundation of data reduction [21]. This consists in reducing the original
data variables \( (x_i) \) by establishing relationships between them. The LV formed are orthogonal between each other, that is, uncorrelated.

Data are typically presented in an \( \mathbf{X} \) matrix where the rows \( (M) \) represent the observations (samples), and the columns \( (N) \) represent the variables in the study (in a NIR spectrum, the wavelength). The parameter to be predicted is represented in another matrix \( (\mathbf{Y}) \) which can be constituted of one or more variables, depending on the number of parameters to be predicted and the correlation between them. Usually, parameters that do not show a good correlation are predicted individually to increase the robustness of the model for each one of these parameters.

When working with images, the previous concepts are applied in a special way and are referred to as Multivariate Image analysis (MIA). In this concept, data are acquired in image format and useful information is extracted by using MVDA concepts.

This chapter presents the main concepts of three of the most applied techniques of MVDA which will be used throughout this project.

**Principal Component Analysis and Principal Component Regression**

Principal Component Analysis (PCA) is one of the first evaluations when dealing with MVDA. It is formed by building a vector in the \( \mathbf{X} \ (M \times N) \) matrix that explains most of the variance, that is, a vector that crosses all dimensions of data (variables) in a plane and lies in a position where the overall distance of the points to that vector is the smallest possible - first Principal Component (PC). Since there are points below and above the plane, this distance is computed as a sum of squares. The second PC is the vector that explains the following major source of variance of the data (although smaller than the first one), and so on, being all PC components orthogonal to each other (Figure 2.1\( \ a \) and \( b \)). Therefore, the last PC represents little information, which is regularly considered as unimportant (noise) to the set. Extracting this part of the information leads to a reduction of the matrix rank, where the number of columns is the same as the number of PC chosen to characterize data, facilitating data interpretation. The maximum number of PCs that can be represented is given by the number of objects -1 or by the number of variables (depending on which one is smaller).
The coefficients for the new variable vectors are the loadings \((p_i)\) \((see\ Equation\ 2.2)\). When gathered, all the loadings constitute a matrix \(P\ (N\times K)\), \((K = \text{number of components})\), which results from the transformation of the original data set relating to the variables. When projected in a loading plot, the relation of the variables with the PCs might be observed.

Also in this mathematical development, the distance of every object (rows) to the vectors (PCs) (columns) that better explains the variance is calculated. This distance is called scores \((t)\) and when gathered, they constitute the matrix \(T\ (M\times K)\). Their representation is plotted in a score plot \((Figure\ 2.1\ c)\). In this graph, the relations between the objects and how they behave against the PCs can easily be seen.

Therefore, the \(X\) data are represented by \(Equation\ 2.2\ [23]\):

\[
X = T * P' + E \quad Equatin 2.2
\]

where \(T = \text{matrix of scores} (M\times K)\), \(P = \text{matrix of loadings} (N\times K)\), \(E = \text{matrix of noise} (M\times N)\), calculated after subtracting the original matrix by the matrix originated after chosen PCs were considered.

The product of \(T\) and \(P'\) is equal to the sum of the considered PCs, that is

\[
X = t_1 * p_1' + t_2 * p_2' + t_\alpha * p_\alpha' + ... + E \quad Equation 2.3
\]
where \( t_\alpha \) = scores of \( PC_\alpha \) and \( p_\alpha \) = loading of \( PC_\alpha \).

- Data mining

PCA is intimately linked with data reduction, which is possible due to the properties of the matrixes, mainly the concepts of eigenvalues and eigenvectors. Although there are other mathematical techniques for data reduction, the most common are singular value decomposition (SVD) and nonlinear iterative partial least squares (NIPALS). The first, which is an adaptation for rectangular matrixes of the eigendecomposition for square matrixes, decomposes a matrix into the following equation:

\[
A = P\Delta Q^T \quad \text{Equation 2.4}
\]

Where \( A \) = rectangular matrix; \( P \) = normalized eigenvectors of \( AA^T \); \( \Delta \) = diagonal matrix of the singular values (\( \Delta = \Lambda^{0.5} \), \( \Lambda \) = diagonal matrix of the eigenvalues of matrix \( AA^T \) and \( A^TA \)); \( Q \) = normalized eigenvectors of \( A^TA \).

The scores (T) and loadings (P) of equation 2.2 are related to “\( P\Delta \)” and “\( Q^T \)” of equation 2.3, respectively. This highlights the importance of calculating the eigenvalues / eigenvectors of matrixes.

As NIPALS is concerned, it is more often used for large amounts of data and when only a few PCs are needed, since no covariance matrix is calculated. The description of the algorithm is present in section A.2.

Therefore, PCA technique is useful for reducing the \( X \) data, removing the noise from it and allowing the observation of the relations between objects, overcoming the limitation of MLR when there is the need for a certain \( y \) parameter or specification to be predicted. In this case, this regression method is named Principal Component Regression (PCR) and it is a two-step procedure, being the first the PCA transformation of the \( X \) matrix followed by the application of the MLR mode [23]. However, since the PCA application is independent from the \( y \), this is, the model of fitting \( X \) could be different from the prediction ability of \( y \), it is necessary to increase the number of PC one by one and test the model in order to have a good
predictability, which obviously limits the goal of truncating data. Therefore, a new regression method is more adequate and it is presented in the following section.

- **Partial Least Squares**

  Partial Least Squares (PLS) method relies on the concept of PCA, but since it is a regression method, a Y matrix comprised of one or more parameters in study (the ones to be predicted) is added to the equation. This method differs from PCR because in the X matrix transformation, there is a maximization of covariance between the X and Y matrixes. In the mathematical development of this algorithm, other matrixes come into play, such as the loading weight (w) as well as the loadings (q) and scores (u) of the Y matrix. For a detailed explanation of the mathematical development of this equation, reference must be consulted [24].

  The results of PLS produce the percentage of variance explained for the Y matrix, as well as the information used to achieve it in X data. A high percentage of the Y explained with a low percentage of X data used, represents a major goal of modeling using chemometrics and allows a good prediction of the parameter in study. However, these results could only be obtained if the relation between the variables of X data is strong enough.

  In some cases, the data used to predict the parameter Y can be obtained from more than one source (two different equipment, for example). In these cases, it is common to merge these sets of data in only one larger X matrix. However, it can be beneficial to use a different chemometric technique, which is described in the next section.

- **Multiblock**

  With the variety of tools available and the possibility to include historic measurements in the study of a process (usually manufacturers keep some records that might be useful in the development of a model), there are many sources of data (different equipment used - NIR, LIF, traditional techniques, etc.) which generates different type of information that must not be interpreted as being part of each other, although it was obtained from a sample of the same composition. Thus, if the origin of the information is known, it is useful to apply a technique that segments data in blocks. Based on this, Consensus PCA (CPCA) and multiblock PLS (mbPLS) have been reported as good algorithms to work with, at least as a complementary
tool for traditional PCA and PLS. For a more detailed explanation of this algorithm, section 4.2.6, Figure 4.4 and the reference must be consulted [20]. MbPLS allows the inclusion of all types of information in a single matrix, although the mathematical development is able to separate it in blocks or segments. Although the overall prediction of the model is not improved when compared to PLS [25], the interpretation of the contribution to each block to the prediction of the $Y$ is of the utmost importance, especially in the development stage and / or when more than one process tool is being tested, as it is the case in this project.

However, even if better and more adequate methodology is applied, there is no certainty that a model could be successfully developed. In fact, model building involves lots of steps, which are detailed in the following section.

- **Modeling**

Building a model based in MVDA requires lot of data and an adequate extraction of its useful features. Several steps are usually recommended to be followed:

- **Exploratory data analysis**

  It is the first step in model building and consists of the first overview of data, identifying the relation between the objects and variables. It is important for outlier detection, which is defined as a sample that differs strongly from the others. In model building, a negative effect will occur and inaccurate predictive results might occur [19]. Outlier identification in spectra is based in Leverage (position in relation to other samples) and Q-residuals (referred below). Regarding the $y$ parameter, the $y$ residual student test is usually performed. Outlier identification could also be performed in the following step of modeling which is the pre-processing stage (confirms the suspicious samples and / or identifies newer ones). The complementary identification in these steps is critical for outlier detection and removal.

- **Data pre-processing**

  Pre-processing is a crucial step in model building, performed to highlight the characteristics of the data facilitating its analysis, and also to correct some differences between samples attributed to factors other than their composition (i.e. equipment characteristic or powder difference - particle size, etc.). As an example, in spectroscopy, phenomena like
baseline offset and multiplicative effects are common. Preliminary observation of data allows the selection of an area of interest (in a spectrum, for example), which is also part of the pre-processing.

There are many types of pre-processing techniques that might be applied (individually or in combination). Their selection depends almost exclusively on the type of data being studied. One of the most used is autoscaling, which is calculated based in the following equation:

\[
x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{S_k}
\]

\textit{Equation 2.5}

where \(x_{ik}\) = variable of index \(i\) (row) and \(k\) (column); \(\bar{x}_k\) = column mean; \(S_k = \sqrt{\frac{\sum_{i=1}^{N} x_{ik} - \bar{x}_k}{N-1}}\)

Autoscaling is useful to remove the influence of absolute values and variation difference along the variable, thus, enhancing the variation response between them. This treatment is mandatory when working with variables whose values have different scales or orders of magnitude.

Spectra are perhaps one of data types where pre-treatment is more important. Apart from selection of appropriate bandwidth (which some authors consider also part of data pre-treatment), some corrections might need to be done. Phenomena such as multiplicative variations between spectra may be solved using techniques like standard normal deviation (SNV) – \textit{Equation 2.6} - or multiplicative signal correction (MSC) – \textit{Equation 2.7} [26][27].

\[
x_{ik}^* = \frac{(x_{ik} - m_i)}{S_i}
\]

\textit{Equation 2.6}

\[
x_{ik}^* = \frac{(x_{ik} - a_i)}{b_i}
\]

\textit{Equation 2.7}

where \(x_{ik}\) = spectral measurement at the \(i\)th wavelength for the \(i\)th sample; \(m_i\) = mean of the \(k\) spectral measurements for sample \(I\), \(S_i\) = standard deviation of the same \(k\) measurements; \(a_i\) = additive correction factor calculated from regression method; \(b_i\) = multiplicative correction factor calculated from regression method.
While in the case of SNV, mean and standard deviation of variables intensity (wavelengths) of the spectra are used, $a_i$ and $b_i$ estimation in MSC is a result of “a linear fit to a reference spectrum” [26]. Since SNV does not involve fitting, it has the disadvantage of being sensitive to noisy entries [28]. In MSC, problems may arise if the reference spectrum (usually the average of a set) is not representative of new data [29].

Another likely phenomenon is the baseline offset and in this case, derivatives are usually the preferred technique not only because they completely remove baseline variation between spectra, but also enhance the variability along them due to differences in their slope (in the case of 1st derivative) and subsequent changes, allowing a better resolution for overlapped features [26]. This converts a “subtle” signal into a detectable one, allowing that property to be accurately identified and quantified. The use of the second derivative enhances the effect of the first derivative.

The use of spectroscopy in this project requires the use of these techniques (especially in the case of NIR). Selection of the overtone areas based on the sample properties as well as the removal of noisy regions in the spectra should be done before preprocessing. However, further spectra selection could be performed after pre-processing is done. After all these steps are performed, the latent variable model can then be computed and therefore the latest stage of modeling could be achieved.

✓ Model characterization

The type of modeling (PCA, PLS, etc.) and consequent choice of the number of components that the model integrates (selection made after analysis of the results), are the last steps in model building. An industrially applicable model must be tested (validation) after being developed (calibration). Different strategies are used to validate the model, as the validation samples might come either from the batch used to develop the model, or from a different set of samples. According to it, many statistical indicators should be analyzed, depending on the type of data set used for both stages:

- **Root Mean Square Error of Calibration (RMSEC)** – Measure of the average difference between predicted and measured response values at the calibration stage [23];

\[
RMSEC = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N-A-1}}
\]

Equation 2.8
where $\hat{y}_i = y$ obtained after testing the calibration equation on the calibration data; $A =$ number of PLS factors

- **Root Mean Square Error Prediction (RMSEP)** – Measure of the average difference between predicted and measured response values at the validation stage [23];

$$RMSEP = \sqrt{\frac{1}{np} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2} \quad \text{Equation 2.9}$$

where $N =$ number of samples in test

- **Root Mean Square Error Cross Validation (RMSECV)** – Measure of the average difference between predicted and measured response values of samples from the calibration set that were placed aside;

$$RMSECV = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_{CV,i} - y_i)^2} \quad \text{Equation 2.10}$$

where $\hat{y}_{CV,i} =$ estimate for $y_i$ based in the calibration equation without the sample $i$.

- **Hoteling’s distance ($T^2$)** – Non-euclidean distance of each observation in relation to the average. If an observation is beyond the 95% limit, it indicates that observation is different from the others, but it does not necessarily mean that it is not part of the model.

$$T^2 = (x_i - \bar{x})S^{-1}(x_i - \bar{x})^T \quad \text{Equation 2.11}$$

where $S^{-1} =$ inverse of the variance-covariance matrix of $X$; $x_i =$ observation; $\bar{x} =$ sample mean

- **Q – Residuals (Q)** - Off plane distance of samples relative to the plane formed. Most of the time, samples whose distance is beyond the 95% limit, may represent an outlier of data, that is, a sample that is not part of the model and should be ignored for model building / validation. However, its interpretation must be carefully made as it might represent an unpredictable behavior of the system [27].
\[ SPE(Q) = \sum_{j=1}^{i}(x_{ij} - \hat{x}_{ij})^2 \]  \hspace{1cm} \text{Equation 2.12}

where \(x_{ij}\) is an element of \(X\) and \(\hat{x}_{ij}\) is an element of \(\hat{X}\) (model predicted)

- \(R^2\) – Fit of the equation obtained after data analysis.

\[ R^2 = \frac{SS(\text{explained})}{SS(\text{total})} \]  \hspace{1cm} \text{Equation 2.13}

If a quantitative prediction is the goal, calibration \(Y\) values to be added might be obtained by off-line traditional techniques, being the validation set used to test if the new tools are able to follow the developed model. In addition to the validation performed in the model, there are some cases where \(Y\) values obtained by off-line traditional techniques are used to confirm and compare the efficacy of the model, or even as a validation set.

2.1.2 Process Analytical Technology tools

2.1.2.1 Imaging

Imaging as a PAT tool is a broad concept. In fact, different kinds of techniques are said to apply imaging but the basis amongst them are slightly different. In common, they all have the capacity to provide information about spatial distribution which is useful when characterizing blend homogeneity, for example [30]. However, its use in pharmaceutical industry is mainly deterministic, that is, it is applied to control certain predictable factors (tablet presence in the blistering process of a pharmaceutical, for example). However, in this project, its stochastic applicability will be tested. Due to the randomness of the processes studied (different spatial colorations appear in every image), it is necessary to detect and to interpret the features of the image to infer other characteristics (for instance, to quantify a component). This feature is more technically and accurately demanding and efforts in this area have not always been successful in other fields [31].

Imaging is usually associated with spectroscopic information (UV-visible, NIR or Raman). However, in this project the definition of imaging is used in the “regular” context, that is, a RGB camera is used to take pictures in 8 bit format and their analysis is performed
with chemometric techniques. The final image of a RGB camera in a 8 bit format is the result of assembling the matrices acquired in Red, Blue and Green band, whose values are defined in a Grey Scale from 0 (Black) to 255 (White). Therefore, the treatment of data is made in that matrix whose size is defined by the number of pixels of width (N), the number of pixels of height (J) and three matrices in depth (Red, Blue and Green) (K) (N×J×K). As they represent the same area, these three images are highly correlated, which can be seen after PCA treatment. To do that, this 3-Dimensional matrix requires a transformation to a 2-Dimensional matrix to be processed and analyzed by the data treatment techniques referred above. This transformation is possible due to a reorganization of data: when working with images, a vertical unfolding is the preferred technique in which the N and J are multiplied, having a final matrix size of ((N×J)×K (Figure 2.2 a) and b)). Once this reorganization is done, the image context is lost (although it is reversible).

Figure 2.2: Reorganization of an image – 3-Dimensional matrix to a 2-Dimensional matrix

One of the major problems in applying imaging as a PAT tool is the difficulty to process such an amount of data. In order to reduce computer memory demand and to take advantage of the equivalents given by the eigenvalues / eigenvectors properties, reduction of a 3-Dimensional matrix to a 2-Dimensional 3×3 matrix is necessary to successfully apply imaging in real time (Figure 2.2 c)). This allows for an original image comprised of a 3-Dimensional matrix to be reduced up to a 3×3 matrix (in case of RGB imaging). In case of an image set, the final 3×3 matrix is the result of the sum of all images. Therefore, once the 3×3 matrix is obtained, the loadings vectors of the image (or of an image set) are easily obtained through SVD (Figure 2.3 a)). Then, the correspondent score vectors are calculated using the unfolded
matrix \((I \times J) \times K\), which are then used to calculate the score vectors \((Figure 2.3\ b))\). All the three score vectors \(t_a\) (in RBG, \(a = 1, 2, 3\)) are then refolded to obtain a score image \(T_a\) of size \(N \times J\) \((Figure 2.3\ c))\).

As in MVDA, the graphs are useful for the interpretation of data in MIA. However, in MIA the score plot is represented by a color-coded 2-Dimensional density histogram since there is an overlap of observations. All pixels are “transferred” to the score plot according to their score image values \((Figure 2.4))\). Pixels with the same features are represented in the same region of the score plot, leading to an overlap of points. Since the score plot is color coded, point overlap is seen through a higher intensity in that region (the brightest intensity possible will be represented in white color). This representation of the score plot facilitates its

\[ \begin{array}{c}
0,
\end{array} \]

\[ \begin{array}{c}
J
\end{array} \]

\[ \begin{array}{c}
A
\end{array} \]

\[ \begin{array}{c}
K
\end{array} \]

\[ \begin{array}{c}
N
\end{array} \]

\[ \begin{array}{c}
I
\end{array} \]

\[ \begin{array}{c}
L
\end{array} \]

\[ \begin{array}{c}
\text{Figure 2.3: Calculation of score images (Adapted from Geladi and Grahn [32])} \]

interpretation and analysis and is critical for model development, being the selection of the area of interest usually made by trial and error [32]. The intensity of the color and the scores are especially important when building a model used to trace batch manufacturing. \(Figure 2.5\) was published in a work by Bharati and MacGregor [33] and demonstrates the importance of region selection for the features in study (in this case, the defects of softwood lumber).

The extraction of the information when analyzing images is the hardest step in modeling, especially when the features in study are subtle and not frequent [34]. Since the variance explained by the first PC is much higher than the variance explained by the second PC, especially in MIA, horizontal location of the feature is usually much more important than the vertical one. However, it is the location in the score plot of the attribute intended to be monitored that determines the importance of each component for a given study.
The analysis of score plots is important and facilitates interpretation in model building. Monitoring the score plot might allow an evaluation of the status of an ongoing process. However, since it is difficult to constantly monitor changes in this plot, many other informations might be extracted from a series of images such as brightness and brightness uniformity, average brightness of non-luminous area, average color of the whole image, average color of the luminous region and number of colors [35].

![Figure 2.4: Representation of the first pixel in the score plot (Adapted from Geladi and Grahn [32][36])](image)

![Figure 2.5: Score plot and false composite image of softwood lumber [33]](image)

Due to the color features that imaging techniques provide, it might be used to quantify and qualitatively identify components in a blend. However, its capability depends highly on
several parameters, mainly the characteristics of the camera chosen (resolution, capacity to acquire high frames per second, adequate optical parts and a light source [37]) and the capacity of the computer to process that amount of data. Texture analysis is among the features that RGB imaging can also perform usually with the help of a Wavelet algorithm [38] but it is complementary with MIA since the information is lost during unfolding [39]. Moreover the existence of different heights in the sample turns the image blurry, which makes this technique only possible to achieve when a flat area of analysis is obtained and a small height between the camera and the sample is obtained, which is expected in a dynamic pharmaceutical manufacturing. Other parameters such as the sensor noise, shutter time and exposure time influence the quality of the image obtained [37]. The major limitation of RGB imaging is the fact that a contrast must exist between the particle intended to be quantified and its host, although some techniques exist (hue, saturation, intensity - HSI) to enhance these differences [40].

2.1.2.2 Near-Infrared Spectroscopy

NIR spectroscopy is widely used in industry and has been the most used tool in the PAT initiative. Its importance is quite evident since it lead to the release of a specific guideline by European Medicines Agency (Europe Regulator) and FDA relative to the subject [41][17]. Its specificity, speed of response, wavelength stability, larger sample analysis and the fact of being non-destructive turns it into an attractive tool. This technology takes advantage of the fact that the combination of fundamental vibrations and overtones, which every chemical bond exhibit, are seen at NIR frequencies [42][43]. It ranges from 4000 cm⁻¹ to 12500 cm⁻¹ (800 - 2500 nm) which is the region where those features can be measured, which involves no sample preparation (contrary to what happens if the mid IR region is investigated, for instance) [26].

Observed spectral features are based in anharmonicity, Fermi resonance and different dipole moments. The first one is responsible for the specific bands in this region, as the energy of the molecule is not symmetrical due to its intramolecular forces. The second allows a split to other frequencies of two or more absorptions bands originally in the same frequency, in a polyatomic molecule [42]. No detailed explanation will be provided about this as it is beyond the scope of this report. However, reference on the subject is readily available [42].
Functional groups possessing one or more hydrogen atoms are generally seen in this region, especially the ones interacting with carbon, nitrogen or oxygen, although other low frequency bands such as C=O and C=C (especially when an alkene group is attached) might also be present [44]. Figure 2.6 represents the overtones and combinations, as well as the wavelengths where they can be observed [45]:

NIR is used in the analysis of many compounds - liquids, solids or gaseous and to characterize physical state as well [43]. Changes in the intensity of peaks in NIR spectra represent a different concentration of the compound or a state transition, which allows quantitative determination along with the qualitative characterization, which is useful to detect phenomena as segregation, for instance. The fact that H₂O overtones are present in the NIR range makes it a perfect tool to detect moisture in a composition due to his capacity to quantify water content. However, in aqueous solutions, its capacity to quantify other components is compromised due to high water absorption [43]. NIR complies with the characteristics that all PAT tools must have (fast, non-destructive and accurate), making it one of the preferred equipment used in this context. It also allows for a large number of molecules to be quantified at the same time and its transferability to industrial application is not difficult [46]. Its major limitation is its high detection limit, especially when compared to other tools (see LIF section below).

Once the composition of the multivitamin mixture is known, characteristic bands could be selected and a close monitoring and spectral selection could be performed. Thus, phenomena like segregation could be detected.

Figure 2.6: Overtones and combinations bands in NIR frequency [45]
2.1.2.3 Light Induced Fluorescence

Luminescence is a type of spectroscopy characterized by the emission of light after an excited state has been acquired. If the source of excitation proceeds from photons, it is called photoluminescence which can be further divided in fluorescence (LIF) or phosphorescence.

Explanation of luminescence phenomena is based in Jablonski’s work and it is represented in Figure 2.7. Exposition of a fluorophore material to a light source in the region of UV-visible light leads an electron from a fundamental state ($S_0$) to an excited state ($S_2$ or $S_1$) – transition 1. After a vibrational relaxation (transition 2), the electron either emits fluorescence (transition 4) or goes to a lower excited state (transition 3). At this stage, after a new vibrational relaxation, the electron can emit fluorescence (transition 4), decay without emitting radiation (transition 5) or less likely go to a singlet and triple excited states (transition 6). If this is the case, the electron goes to phosphorescence (transition 7), or goes to the external conversion route. This phenomenon is named photobleaching and it is still not yet fully understood, although it is known that the presence of anti-oxidants in the mixture reduces this effect [47].

![Figure 2.7: Luminescence Jablonski diagram](image)

LIF is characterized by its high sensitivity, low detection limit (concentration as low as 0.01% might be observed) and dynamic range as well as a fast response, high spatial resolution and sensing capabilities [26][48]. The use of a wavelength for emission and another
one for absorption represents an advantage for fluorescence, since the specificity of the method is highly increased. Different sources indicate that the sensitivity of fluorescence is approximately 100–1000 times greater than the absorption of spectrophotometric methods [49]. In Figure 2.8, it is observed how LIF demonstrates clear advantages of sensitivity over NIR in a five concentration (0.05% - 3% (w/w)) caffeine-lactose mixture.

The main disadvantage of this technique lies in the fact that it can only be applied when compounds are fluorophores (substances that emits fluorescence after excitation by a source of light with specific wavelength and intensity). However, since in pharmaceutical industry the active substance is usually the compound that needs to be quantified and it is likely to have in its constitution one or more aromatic rings (which is a property that leads to fluorescence) and taking advantage of the fact that the majority of excipients do not have fluorescence properties (which contributes to high method specificity), this technique has a high potential of use [50][51]. Another limitation lies in the fact that in order to be analyzed, the formulation must be a “dried blend, semisolid or liquid” in which a constant change in the monitoring window is required [52].

![Fluorescence spectra](image)

*Figure 2.8: Five concentrations of a caffeine-lactose mixture a) NIR Spectra; b) LIF spectra [48]*

The fluorescence intensity is also affected by pH fluctuations and temperature. Self-quenching behavior at high concentrations is observed, leading to a redundant result in which despite the increase in fluorophore concentration, emission intensity decreases because absorption speed is higher than emission. Saturation of the detector might occur when the fluorophore concentration increases, limiting the linear response range [52].
When many fluorophores are present in a sample, more sensitive equipment is required to overcome the problem of overlapping.

The LIF equipment used in this project (presented below) has some features that increase the possibility to obtain better prediction. It allows the selection of multiple excitation wavelengths and acquires the full emission spectrum. However, a single excitation wavelength was chosen (280 nm) since it was the wavelength that provided a higher emission for the API and a lower emission for the matrix [53] (reason why other constituents such as microcrystalline cellulose did not represent a problem in this study). Contrary to many other probes that have a photomultiplier detector, it produces a full emission spectrum. This avoids the univariate analysis characteristic of photomultiplier detectors and therefore allows to monitor more than one component at a time even if their spectral emissions overlap. This feature is especially important, since the powders in this work are constituted of several active ingredients, some of them with fluorophore properties.

Another important aspect is the choice of the emission source. Light-emitting diode (LED) arrays and Laser diode (LD) arrays are the best ones, since they do not heat (do not change sample characteristics), do not give any background noise, emission signal precision is better and provide a good dynamic control [26].

A fluorescence spectrum is different from a NIR spectrum, which is evident when comparing Figure 2.8 and Figure 2.9. With a fluorescence spectrum, a smoother line is obtained when comparing to NIR. The presence of many components increases the noise in NIR spectra, making model development much harder in this case. Therefore, in LIF spectra, chemometric treatment is not as demanding as for NIR spectra, leading to more accurate results.

![Figure 2.9: Fluorescence of riboflavin in ethanol](image)

*Figure 2.9: Fluorescence of riboflavin in ethanol [54]*
The ability to detect low concentration components is the most important and distinguishable feature in relation to the others technologies present. Since this work will be developed in a multivitamin mixture whose composition have substances that emit fluorescence and are present in a low concentration, LIF is the perfect tool to monitor these components.

However, the benefits of installing PAT tools are not restricted on monitoring a certain component. The information they provide increase process understanding and allow model calibration design to be more adequate. Therefore, the influence of some parameters (process or formulation related) can be more successfully investigated and included in model building. The interaction between these two factors and the extension in which they modify the final product specifications must be understood to define an accurate control space in which the manufacturer can work without compromising final product quality. In this sense, PAT tools play a critical role in this development as they allow detection of underlying phenomena. This information is then linked with the powder study performed under controlled conditions (rheological tests). Therefore, the concepts that guide powder rheology have to be known in order to establish this connection. These concepts are discussed in the following section.

2.2 Monitoring the compression stage – research topics

Compression is the last step in tablet manufacturing, which consists in a mechanical transformation of powder. As referred, powder is transferred from the hopper into the feed frame, whose constitution encompasses one or more rotating wheels whose paddles that have the function of feeding the powder continuously into the dies, where they are finally compressed. In spite of its importance, an extensive research is not reported, although it can still be divided in two major groups (Table 2.1). There are some studies focused on the influence of different parameters of the press (either operating parameters – feed frame speed, die disk speed – or design parameters – paddles number, die dimension, number of wheels in the feed frame) in the powder behavior (modification of flow properties, segregation phenomenon, wave behavior, etc.), in powder properties and also in the acquisition by the tools (distance to probe, etc.). These types of tests are performed by running simulations (mainly through Discrete Element Method (DEM) technique) or through the acquisition of
experimental data. On the other hand, some studies have focused on in-line monitoring of the process itself, either by assessing (usually, quantifying a component) the circulating powder with a probe installed in the feed frame or by testing the intact tablets recently formed. In this latter strategy, the techniques are usually performed off-line, although two works refer its in-line applicability [55][56]. Table 2.1 represents all the studies published so far regarding the in-line study of the tablet press, when applicable. In the case of tablet analysis, the extensive off-line analysis that can be performed is not listed, since it is beyond the scope of this project. However, the application of the tools used in this study for that purpose, is highlighted.

In the next sections, the outcomes of the studies made regarding the influential parameters of the press and the role of powder properties (Section 2.2.1) are presented, followed by the work that has been done using the probes chosen for this work especially when used in-line for component tracking (Sections 2.2.2, 2.2.3 and 2.2.4)

2.2.1 Influence of press parameters and powder properties

The feed frame is a piece that is very different between presses, regarding its geometry, number and direction of rotation of wheels, shape and number of paddles, etc. Since they constitute a mixing mechanism, they are able to change the properties of the powder and influence die fill weight variability [12][57]. Therefore, they have been the subject of some studies that use simulations to predict powder behavior [58]. However, due to different feed frame designs, the diversity of operative parameters (different feed frame and dies speed) and the differences in powder properties, the conclusions obtained from these studies must be contextualized. In fact, all these factors require that in the implementation of a monitoring tool, the conditions used when the model was developed must be respected. Therefore, these studies demonstrate that a careful approach must be taken when feed frame monitoring is used. However, in spite of the great variability of this stage, these studies help to understand which conditions can potentiate a good monitoring strategy.

Many studies demonstrate that higher paddle speed originates a lower weight variability in the dies [57][58] since the availability of the powder is higher. However, it also implicates a higher shear stress to the powder, which leads to particle breakage [59] and thus, can induce some undesirable phenomena such as segregation, especially in formulations with a higher tendency to segregate [12][60] (components with important differences of particle size and
<table>
<thead>
<tr>
<th>Scope</th>
<th>Tool(s)</th>
<th>Conclusion(s)</th>
<th>Reference number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>Influence of press parameters and powder properties</strong>&lt;br&gt;a) <strong>Simulation methodology</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study of segregation inside feed frame</td>
<td>DEM</td>
<td>Occurrence of size segregation on the feed frame and in die filling, mainly affected by paddle wheel speed</td>
<td>[12]</td>
</tr>
<tr>
<td>Study of paddle speed, number and height</td>
<td>DEM</td>
<td>Faster paddle speed, higher number of paddles and lower paddle height lead to lower die fill weight variation;</td>
<td>[58]</td>
</tr>
<tr>
<td>Evaluate the influence of paddle shape, rotation direction and speed; Direction of powder flow, Residence Time Distribution (RTD) and die filling variability; Effect of powder cohesion</td>
<td>DEM</td>
<td>Flow and RTD varies with paddle shape, rotation direction and speed; More weight stable tablets are obtained for higher paddle speeds; Cohesive particles present larger weight variability</td>
<td>[61]</td>
</tr>
<tr>
<td>Study of RTD of particles and relating operating and feed frame conditions</td>
<td>DEM</td>
<td>Particle size segregation occurs when small particles have faster exit speeds; RTD of particles is affected by paddle and dies speed;</td>
<td>[62]</td>
</tr>
<tr>
<td>b) <strong>Experimental data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluate influence of paddle speed in weight variability</td>
<td>Sieve, Pycnometer, PFT&lt;sup&gt;(1)&lt;/sup&gt;</td>
<td>Paddle speed influence tablet weight variability in low flowing powders</td>
<td>[57]</td>
</tr>
<tr>
<td>Study the effect of design and initial powder characteristics in flow characteristics</td>
<td>(Off-line): GDR&lt;sup&gt;(2)&lt;/sup&gt;, LDA&lt;sup&gt;(3)&lt;/sup&gt;</td>
<td>Feed frame geometry, operating conditions, die system conditions and initial particle size distribution (PSD), changes the final PSD. Paddle speed increment in small feed frames does not affect PSD.</td>
<td>[59]</td>
</tr>
</tbody>
</table>
1. **Influence of press parameters and powder properties**

   **b) Experimental data (continuation)**

<table>
<thead>
<tr>
<th>Scope</th>
<th>Tool(s)</th>
<th>Conclusion(s)</th>
<th>Reference number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation of the effect of powder properties, die size, rotary disc speed, paddle speed on flow pattern, applied shear and die filling uniformity</td>
<td>Digital imaging / NIR</td>
<td>Powder properties, speed and geometry of feed frame and dies affect flow pattern inside the feed frame; Shear and RTD are function of blend properties and operating conditions; Cohesion and die filling variability decreases with paddle speed</td>
<td>[60]</td>
</tr>
<tr>
<td>Study of the effect of blend composition, and feed frame and die parameters on powder and tablet properties</td>
<td>Quantachrome Instruments Autotap, GDR(^{(2)})</td>
<td>For lubricated blends, feed frame parameters had a large impact on powder hydrophobicity and powder properties; Shear promotes flow but has an effect in tablet hardness and dissolution</td>
<td>[63]</td>
</tr>
<tr>
<td>Optimise the probe distance and study the influence of powder movement</td>
<td>NIR</td>
<td>Optimization of distance; Possibility to minimize agitation effect in the signal;</td>
<td>[64]</td>
</tr>
<tr>
<td>Influence of paddle wheel speed, probe location and mass throughput rates</td>
<td>NIR</td>
<td>Paddle wheel rotational speeds and probes location must be optimized for different geometries.</td>
<td>[65]</td>
</tr>
<tr>
<td>Understand the die filling process; study different parameters that can affect model prediction</td>
<td>NIR</td>
<td>NIR can be used to determine mass changes inside the feed frame; paddle wheel speed has significant impact in dynamics; Distance between probe and powder has effect on model prediction;</td>
<td>[66]</td>
</tr>
</tbody>
</table>

2. **Monitoring**

   **a) Powder**

<table>
<thead>
<tr>
<th>Scope</th>
<th>Tool(s)</th>
<th>Conclusion(s)</th>
<th>Reference number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluate content uniformity of tablets by monitoring of the powder</td>
<td>NIR</td>
<td>Successful monitoring of powder with an API of 30% w/w and 2 excipients</td>
<td>[16]</td>
</tr>
<tr>
<td>Track an API</td>
<td>NIR</td>
<td>4% potency component was successfully identified</td>
<td>[64]</td>
</tr>
<tr>
<td>Scope</td>
<td>Tool(s)</td>
<td>Conclusion(s)</td>
<td>Reference number</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>---------------</td>
<td>------------------</td>
</tr>
<tr>
<td>2. Monitoring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Powder (continuation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantification of a mixture with two APIs</td>
<td>NIR</td>
<td>Good correlation between off-line tablet analysis and powder concentration</td>
<td>[65]</td>
</tr>
<tr>
<td>Monitor the concentration of API (5-15% w/w)</td>
<td>NIR</td>
<td>In-line models presented more accuracy and less RMSEC than off-line models due to an influence of paddle speed.</td>
<td>[66]</td>
</tr>
<tr>
<td>Evaluate the potential to monitor an API in a feed frame</td>
<td>NIR</td>
<td>Univariate analysis led to a good correlation coefficient</td>
<td>[67]</td>
</tr>
<tr>
<td>b) Tablets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantification of API on tablet surface</td>
<td>LIF</td>
<td>Quantitative correlation between tablet surface and off-line analysis</td>
<td>[51]</td>
</tr>
<tr>
<td>In-line monitoring of a continuous tablet process</td>
<td>NIR</td>
<td>Authors were able to develop calibration models with low error for in-line tablet monitoring</td>
<td>[55]</td>
</tr>
<tr>
<td>Applicability of in-line NIR sensor in to quantify API and excipients</td>
<td>NIR</td>
<td>Simultaneous quantification of API and excipients was achieved</td>
<td>[56]</td>
</tr>
<tr>
<td>Hardness prediction, API determination, Uniformity of dosage</td>
<td>NIR</td>
<td>PLS models developed off-line were able to predict hardness, API (15-30 mg) and evaluate the UDU.</td>
<td>[68]</td>
</tr>
<tr>
<td>Content uniformity, hardness, disintegration time, friability;</td>
<td>NIR</td>
<td>PLS models developed off-line allowed the prediction of the parameters in study (API of 200-300mg)</td>
<td>[69]</td>
</tr>
<tr>
<td>Compare quantitative off-line LIF measurements of tablet surface with a reference method (UV)</td>
<td>LIF</td>
<td>Two different APIs with a range of 1.64-4.75% and up to 20% were successfully quantified through analysis of tablet surface</td>
<td>[70]</td>
</tr>
</tbody>
</table>

*(PFT: Pharma Flow tester; GDR: gravitational displacement rheometer; LDA: Laser Difraction Analyzer)*
density in its constitution). Similar effects are expected when higher residence time inside the feed frame is potentiated (i.e., decreasing die speed [63]). Paddle speed can also affect the dynamics of the feed frame [66], as the wave behavior usually seen in this equipment due to the paddle, varies. This can also lead to variations of the signal obtained in the installed probe [65], since at higher paddle speeds, dilatation of the powder occurs. For non-contact probes, this information is important to evaluate the best S/N ratio and to know other characteristics of the compression stage that might be occurring (for instance, a different baseline shift might be used to determine the mass inside the feed frame [66]). Since in this work the NIR probe works as a non-contact probe, this information represents an indication that a preliminary study regarding S/N ratio needed to be performed.

In terms of powder properties, Mendez et al. [63], demonstrated that lubricated powders (especially the cohesive) are more susceptible to shear in terms of their flow properties as they show a lower dilatation and flow index and a higher hydrophobicity. This might lead to “overlubrication” of the powder which decreases its cohesion that have an impact on tablet properties (dissolution and hardness). This information enhances the importance of monitoring this stage of the manufacturing process (most of the formulations have lubricants in their constitution) and the specificity of the model developed for each blend.

### 2.2.2 Imaging

Imaging alone is not a common PAT tool for in-line monitoring of pharmaceutical processes, although its usefulness and potential are quite evident. The concept is broad, although imaging is usually a feature of an equipment that could also perform another type of analysis, usually spectroscopic (NIR, Raman), allowing the possibility to characterize the sample both spatially and qualitatively. As this project is concerned, the information obtained by imaging will be complemented with the data obtained by other tools, but the equipment itself is dedicated (a RGB camera will take pictures). Therefore, taking into account the goal of this tool in this work, this section is focused on the studies published in which imaging has been use as a dedicated equipment with stochastic goals.

Monitoring of coating process is the pharmaceutical manufacturing step where imaging has been most tested and some results are promising, although no reports were found illustrating its use as an in-line tool, which highlights the novelty of this project. A review
from Knop and Kleinebudde [71] referred the possibility to use imaging while working with pellets, evaluating their difference before and after coating, if a narrow size distribution, a uniform coating and a large number of measured particles are obtained. One of the most important experiments was performed by Kennedy and Nierbergall [72]. They used digital imaging in batch samples and through standard deviation and mean of their optical densities calculation, they were able to monitor the uniformity and quantify the thickness of the samples, despite the few numbers obtained (around 200). Also at-line, Kucheryavski et al. [73] demonstrated a strong relation between image features and pellets growth, comparing the pixel area occupied by pellets during the process, after having removed some samples throughout the coating stage and inserted them in a 5 cm box for picture acquisition. The use of Angle Measure Technique (a chemometric method) showed better robustness than wavelet transformation. Mozina et al. [74] also dynamically demonstrated the advantages of this tool, using a monochromatic camera and developing and validating a model that was capable to detect agglomeration and coating thickness, after separation of the pellets according to their shape and size (so the measurement of the coating thickness was not affected by agglomeration phenomena) and calibration of pixel size with bearing balls. Applying image analysis to optical microscopy to monitor the coating of the pellets based in the mean difference of projected area of coated and uncoated pellets, was a study performed by Larsen et al. [75] whose conclusions proved the robustness and the advantage of this method due to a lower need of sampling (only 1000 pellets). Use of imaging to assess the cosmetic attribute of coating was reported by García-Muñoz and Gierer [76] using off-line images to determine the coating uniformity end-point. They also inserted an on-line camera inside the pan coater (being this case one of the rare examples where on-line imaging was applied in a pharmaceutical process) and by using PCA adaptive modelling approach were able to identify the time where color of coating stopped changing.

Imaging has also been studied in the blending step, although fewer studies are reported and none of them were performed with an RGB camera. Despite this, Whightman et al. [77] were able to quantify the components present in the mixture, after adding a solidifying agent to the mixture and analyzing it after the mixture was sliced. They used parameters as average coloring to quantify the components. Also Realpe and Velázquez [78] proved imaging can be used for quantification of mixtures and exhibit the same precision as NIR, having used mean
gray values of several mixtures to calibrate the model, which is an indication of imaging applicability in this project. Another work by Daumann et al. [79] referred imaging as a suitable method to monitor efficient mixing as they put a color tracer in a mixture, isolated it through image reprocessing and transformed the analysis in a colored binary code to monitor the efficiency of blending throughout time through the analysis of empirical variance of the trace concentration (color). The use of imaging to evaluate certain phenomena such as their tendency to the blends to segregate has also been reported, taking advantage of its feature to determine the size of the particles [80].

Although none of the works described above has used this tool in-line, there have studies in other industries that have successfully applied it in this condition for the qualitative and quantitative characterization of a process. One of the most important was performed by Yu and MacGregor [35] in which the performance of a boiler system was evaluated in-line by a RGB camera through the acquisition of pictures of flames generated by combustion. Based on the color difference of the flames, the authors were able to predict, among other things, the concentration of components present in the combustion reaction or originated by it, which is what it is intended in this work. Also in the food industry, the implementation of an in-line RGB camera in the production of snack foods was used to predict its coating concentration and distribution [81][82]. The use of an RGB camera was used in another study that was able to monitor in-line the flotation process, after a selected region comprising the features of interest was chosen in the score plot and ulterior wavelet texture analysis was applied do distinguish between two features of interest [83]. The change in the score plot was used to follow the manufacturing process. The chemical industry has also successfully applied imaging as on-line tool, especially in crystallization processes to monitor the shape and size of crystals [84].

### 2.2.3 Near-Infrared Spectroscopy

Near-Infrared is one of the most used equipment in PAT applications as it perfectly complies with the characteristics demanded for these types of tools. Along with the water content referred above, there are many others parameters that can be measured with NIR whether in, on, at or off-line. A review article by Luypaert et al. [85] discussed all applications and studies that had been done with this technology until then, highlighting its usefulness in
raw material identification, particle size measurement, tablet hardness measurement and also in components quantification, reason why it is commonly used to monitor drug potency. Also for that reason, it is a suitable tool for segregation detection [86]. Moreover, properties like powder particle size, pH, moisture content, flow related properties, flowability and angle of repose have also been successfully modeled and could be predicted by this tool [87]. NIR has also demonstrated its ability to detect process changes and interferences as well as to study the kinetic of a mixture [88][89][90].

Due to the scope of this thesis, the in-line application represents the major point of interest. In fact, unlike imaging whose applications have been restricted to the manufacturing stages of blending and coating, studies using in-line NIR spectroscopy have been explored in a much larger extension. Along with these two manufacturing steps [91][92], there are many studies referring to its utility in granulation [93], drying [94], compression [65] and freeze-drying [95]. In case of (wet) granulation, drying and freeze-drying, NIR spectroscopy utility is mainly justified by its capability to monitor water content. In the coating stage, as the core is coated, higher reflectance is observed, which leads to a change in spectra and enables the monitoring of coating end point as well as determination of coating thickness [96]. In blending and compression stages, NIR spectroscopy is a powerful method to determine homogeneity (which identifies blending end point) and for drug quantification (which identifies potential segregation phenomena), which is the feature of interest in this project. Its capacity to determine particle size is also important to understand the behavior of powder and recently, its use as a tool to assess in real time the powder density has been reported [97], parameter that is critical to maintain uniformity of the blend. As a first choice PAT, its applicability has also been applied in several continuous manufacturing stages [98].

Taken into account the focus of this project in the monitoring of the compression stage, it is pertinent to highlight the most significant reports that were published regarding this manufacturing stage and using NIR. As mentioned above, two different strategies to monitor the compression stage can be distinguished: a) Monitoring of the tablet and b) Monitoring the powder in the feed frame. In the first case, many off-line studies report the success in the quantification of the components and other tablet parameters by acquiring data from intact tablets [68][69]. On the other hand, an in-line strategy to use intact tablets to monitor the compression stage was suggested by Karande et al. [56]. Despite having installed the probe in
the press machine, their analysis was performed in the product final form (tablet) and not in the powder. They were able to develop a model to quantify both the API and the excipients and proved it could detect powder phenomena such as segregation, since in one of the batches a difference in concentration between the components was observed, phenomenon which the authors attributed to the hopper geometry and to the difference in particle size and density of the components. Another in-line analysis of tablet was done by Järvinen et al. [55], as part of a continuous tabletting system. In this work, acetaminophen tablets with a concentration of 20-30% (w/w) were successfully predicted based in a VisioNIR system placed at the exit of the dies (Figure 2.10). Despite of these studies and as mentioned in a recent review made by Laske et al. [99], analysis of the tablet by spectroscopic methods present in general a lower S/N ratio when compared with powder analysis due to the higher specular reflection observed in tablets along with a more surface analysis (depth of penetration in tablets is also lower). Therefore, monitoring the circulating powder in the feed frame seems to present a better strategy to control the compression stage at least by the use of spectroscopic methods, as suggested in this work. However, there are not many reports about the use of NIR in the feed frame. A work by Liu and Blackwood [64] was one of the first that placed an in-line NIR probe above a window

Figure 2.10: VisioNIR sensor to analyze tablets [55]
in the feed frame of a rotary press machine. They were able to detect a variation in concentration of 70% in a 4% (w/w) drug load. Ward et al. [65], used a similar setup (Figure 2.11) and placed a powder constituted by placebo and then a layer of API (in the first test with a concentration of 3.5% (w/w) and in the second with a concentration of 30% (w/w)) in the hopper and were able to detect concentration changes in the feed frame, having also established a good relation between potency detection and off-line results obtained by traditional techniques. More recently, a work by Wahl et al. [16], demonstrated the ability of NIR to monitor in-line a composition of 30% API with 2 excipients whose model was developed off-line. In this study, the authors were also able to detect segregation by the end of their run, which were attributed to the discharge from the hopper. Another study by Šašić et al. [67], pointed the potential of a NIR mounted in a feed frame to monitor the API, after a univariate profile, selected at the typical band of absorption of the API, was developed and correlated with off-line analysis. Therefore, all these studies indicate the utility of NIR to quantify components in the feed frame in a different range of concentrations.

Many of the studies concluded about some technical parameters that are critical to implement this technique. Liu and Blackwood [64] studied the influence of probe position in

Figure 2.11: NIR probe installed in the feed frame [65]
spectra acquisition and concluded that the dynamic measurement of spectra could be performed if the probe was in an accurate position and adequate data pre-processing was done. The same conclusion was observed by Ward et al. [65] whose work also investigated the influence of mass flow rate inside the feed frame and the paddle speed (varied between 10 and 40 rpm). They observed a close relation with off-line measurements when mass flow rate was changed, but in the case of paddle speed, there was a slight bias in the NIR measurements. Another important technical study was performed in 2007 by Li et al. [100]. The authors studied the influence of beam size of the in-line NIR sensor in the blending stage and suggested that this factor is important when building a model to quantify components. In other study, Mateo-Ortiz et al. studied the influence of many other parameters that could be obtained through NIR [66], as well as factors that must be considered when this tool is applied in the monitoring of the feed frame. They concluded that the baseline of the NIR spectrum could be used to assess the mass inside the feed frame. The baseline was also shown to be affected by the paddle speed due to the changes in powder density it originated, which put the powder closer to the probe. They also demonstrated that the NIR prediction is strongly affected by paddle speed which confirmed the specificity of each of the models developed for a certain condition. However, in this project, installation of a trigger in the set-up was done to minimize the noise by selecting a specific area of acquisition.

2.2.4 Light Induced Fluorescence

As imaging, LIF applications in pharmaceutical industry are rare. Its application is mainly used in monitoring powder blending (whether detecting the end point or homogeneity) through drug quantification and in the quantification of tablets.

In the case of blending, all studies reported in literature showed a good capability to detect homogeneity and signal the end point, despite some limitations of the method due to the API concentration. In one of the initial studies performed by this tool using an on-line strategy, Lai et al. [101] proved a linear relation could be obtained from a concentration as low as 0.1% to 1% API (w/w) of Triamterene, demonstrating the potential of this tool to monitor high potency drugs. However, a nonlinear behavior was observed for higher API concentration, which was believed to be due to the saturation of the photomultiplier tube. In a later study, the same author [52] was able to monitor the concentration of Triamterene as low as 0.02% w/w in the
In this work, the potential of this tool to detect segregation was also shown, since segregation was statistically demonstrated, after analyzing the obtained LIF signal in theoretically homogenous blends. This feature is of the upmost importance when monitoring the feed frame and reveals the importance of the installation of this tool in this piece of equipment.

A study by Karumanchi [49] demonstrated the applicability of LIF to monitor qualitatively the blending stage and is well correlated with conventional HPLC techniques. More recently, Guay et al. [53] installed a LIF probe in a lab-scale V-blender and were able to develop an accurate model to follow the API (concentration between 1.44 and 4.19% w/t) in a study that proved the applicability of LIF as a PAT tool. This work validated the option of selecting LIF in this work, as it was able to track in-line low dosage components.

In the case of tablets, two main studies must be highlighted. In the first case, Lai et al. [51] were able to establish a linear correlation for concentrations of Triamterene tablets between 1.64% and 4.75% w/w just by analyzing their surface and monitored in a high surface rate – 3000 tablets/min. In another study, Domike et al. [70] using the same concept, was able to establish a correlation for tablets containing caffeine in a concentration up to 20% w/w.

Since it is a relatively new technique, these studies also present a lot of technical information that must be taken into consideration in the implementation of this tool. For instance, the dynamic capability of this method has been proved and it was shown that both blending homogeneity and drug content can be adequately predicted using this kind of setup, even if a single emission wavelength is used (in this case, mixtures components did not interfere between them) [51][52]. Also the penetration depth of LIF was investigated. Lai et al. [51] found that in tablets prepared by 3D Printing procedure, 30% of the signal of LIF was detected in a 1 mm depth, leading to the conclusion that if this tool can penetrate this much in a loose powder, it is expected that the penetration will be much lower when working with denser powders. Since in this project a free flowing powder was used, this depth was considered for calculating the mass analyzed by each observation – Annexes A1. Light penetration inside the sample will depend on the characteristics of the powder being studied, mainly the particle size and its density, as well as the selected wavelength. Shah et al. [102] reported a study where the density of the powder was correlated with the response from the LIF. They observe that a linear response from the LIF was obtained when the density
increased up to 4% w/w, but from this value on, a negative trend was observed, confirming the influence of the density of the powder on the results obtained by this tool. In his study, Guay et al. [53], concluded that changes in moisture, pressure (hydrostatic forces, for instance), particle size and especially the presence of colored components, affect the LIF signal and must be controlled in routine, which is a relevant information for this work since many other colored components are present in the studied blend.

Some improvements are also reported in the equipment. Recently, Dickens et al. [103] published a work where a LED-array fluorescence sensor was developed to investigate the improvement of the sensitivity of the equipment. They used a dynamic emission that led to a broader detection capability which allowed them to detect concentrations of the component in study (Tryptophan) as low as 0.001% (w/w).

All the studies referred above indicate the potential of these tools to be used in the proposed conditions. However, the development of monitoring tools is not straightforward and goes through different stages. In the next three chapters, the papers that arose from this work are presented. They chronologically describe the logical and typical sequence of studies made when the goal is to implement a PAT tool in a process. Therefore, in chapter 3 a qualitative analysis of the potential of the tools is assessed and a technical evaluation of the set-up is evaluated. After its successful completion, a quantitative evaluation was performed in the press (chapter 4) in which the benefits of simultaneous use of tools is determined, as well as an evaluation of blend kinetics. Finally, in chapter 5, the evaluation of the robustness of the three PAT tools is explored.
NOTE

The following chapter is based on a paper entitled “Monitoring the concentration of flowing pharmaceutical powders in a tablet feed frame” and was published in “Pharmaceutical Development and Technology”.

It has been included here as part of Pedro Filipe Marchão Palmeiro Durão’s work. Due to a misunderstanding with the journal, the first author is indicated as Ryan Gosselin, while it should have been Pedro Filipe Marchão Palmeiro Durão. Demands to the journal to make corrections upon reception of proof copies have not been successful.

Permission to include this chapter in this thesis has been requested to the Faculty and has received special approval by the vice dean of research.
CHAPTER 3  Monitoring the concentration of flowing pharmaceutical powders in a tableting feed frame

Title in french: Suivi de la composition de poudres pharmaceutiques lors de leur écoulement dans la trémie d'alimentation d'une presse à comprimés

Authors and affiliations:

Ryan Gosselin, M.Eng., Ph.D, Associate Professor, Université de Sherbrooke, Département de génie chimique et génie biotechnologique, 2500 boul. de l'Université, Sherbrooke, Québec, Canada, J1K 2R1.

Pedro Durão, M.Pharm., Ph.D. Candidate, Université de Sherbrooke, Département de génie chimique et génie biotechnologique, 2500 boul. de l'Université, Sherbrooke, Québec, Canada, J1K 2R1.

Nicolas Abatzoglou, M.Eng., Ph.D., Full Professor, Université de Sherbrooke, Département de génie chimique et génie biotechnologique, 2500 boul. de l'Université, Sherbrooke, Québec, Canada, J1K 2R1.

Jean-Maxime Guay, M.Eng., Pfizer Global Supply, Process Analytical Sciences Group, Montreal, Quebec, Canada

Date of acceptance: 22 September 2015

Status: Published

Summary

Content: The first step in the implementation of a PAT tool is to evaluate its potential to monitor the selected CQA. In this paper, it is intended to qualitatively evaluate the potential of the selected tools (NIR, RGB camera and LIF) to detect differences in concentration of certain components in the proposed set-up. The components to monitor were selected according to their characteristics and of the tools in place. Five different blends (with six components) were used and the concentration of three components varied between them. The trial was performed by simulating the conditions usually found on a press.

Results: Through the use of MVDA, the variation of the first PC - calculated after PCA analysis - was plotted throughout the experiment for each of the probes. Assuming the variation of the score values had a relation with the concentration of the components, it was demonstrated that all tools were able to detect a difference in the concentration at least in one point of the experiment. In addition, it was possible to correlate the variation of the score values with the concentration of the components, allowing to determine which components could be monitored by each of the tools.

Contributions to the thesis: This paper allowed to confirm the potential of RGB imaging and LIF spectroscopy to be used in-line in the monitoring of components (the applicability of NIR had already been evaluated in other studies) in the proposed set-up. It also confirmed the potential of LIF to be used in the quantitative evaluation of low dosage components.
Abstract

The use of process analytical technology (PAT) tools is increasing steadily in the pharmaceutical industry. Such tools are now located throughout the process. When producing tablets, the tableting step itself may be the ideal moment to assess final product composition. Being the last unit operation in tablet production where the elements are still free flowing, it is relatively straightforward to ascertain the composition of the blend in real time. However, a single probe cannot be expected to monitor the composition of every component of a multicomponent blend. In this study, 3 PAT tools (light-induced fluorescence spectroscopy, near infrared spectroscopy and color (RGB) imaging) simultaneously checked the composition of powder blends flowing through the feeding unit (feed frame) of a tablet press. The results demonstrate the potential of these tools in monitoring changes in concentration of a multicomponent mixture in real time, providing users with means to both scrutinize the process and better understand phenomena occurring inside the feed frame.

**Keywords:** Tablet Press; PAT; Imaging; Near-infrared; Fluorescence
Résumé français:

L’utilisation des technologies d’analyse de procédés (Process Analytical Technology, PAT) est en augmentation dans l’industrie pharmaceutique. Ces technologies sont maintenant implantées directement dans les procédés de fabrication. Lors de la production de comprimés, l’étape de compression est idéale pour évaluer la composition finale du comprimé. Cette étape est la dernière durant laquelle les différents éléments composant le mélange sont libres, il est donc relativement aisé de vérifier la composition du mélange en temps réel. Cependant, il n’est pas réaliste d’espérer qu’une unique sonde puisse déterminer la concentration de tous les composés d’un mélange complexe. Lors de cette étude, trois technologies différentes (spectroscopie de fluorescence induite par la lumière, proche infra-rouge et imagerie couleur (RGB)) ont été utilisées simultanément pour vérifier la composition du mélange s’écoulant de l’unité de remplissage (trémie d’alimentation) de la presse à comprimer. Les résultats de cette étude démontrent le potentiel de ces technologies pour suivre en temps réel les changements dans la composition du mélange, cela permettant aux utilisateurs de suivre précisément le procédé ainsi que d’avoir une meilleure compréhension des phénomènes se produisant dans le bol d’alimentation.

Mots-clés: Presse à comprimé; PAT; Imagerie; Proche Infra-rouge; Fluorescence
3.1 Introduction

Since the 2004 publication of “Guidance for Industry PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance” [1], the use of PAT tools by the pharmaceutical industry has been the subject of considerable scientific work. Although PAT tools (e.g. NIR spectroscopy) are now commonly employed for off-line measurements, their true potential may lie in their application to in-line or on-line measurements [69]. The information provided by these tools makes it possible to monitor the process in real time, leading to better process understanding, improved quality of the final product and the reduced need for post-production destructive sample testing.

3.1.1 Tableting

Tableting is the last stage in the manufacturing chain in which the product is still in free-flowing powder form. As such, powder monitoring at this step makes is possible to ensure that the product meets quality specifications. Late-stage monitoring can quantitatively determine the composition of the mixture being compressed, assess mixture homogeneity and detect undesirable phenomena, such as agglomeration and segregation. The latter could occur anywhere inside the process, from the intermediate bulk container, to the hopper or even inside the feeding unit (i.e. feed frame) itself. Gauging powder quality inside the feed frame may reduce the need for off-line tablet analysis which is destructive, time-consuming and costly. Therefore, PAT tools in this step of the manufacturing process are fundamental to enable real-time release by reducing the necessity of further quantitative API testing.

The present work evaluates the potential of combining three tools, NIRS, LIF and color RGB imaging, to monitor variations in concentration of a dry particulate system, inside the feed frame of an industrial tableting press.

3.1.2 PAT tools

Different technologies have been applied as PAT tools: for example, spectroscopy such as NIRS and Raman are among the probes most commonly deployed for this purpose [104] because of their numerous advantages (e.g. high acquisition rates, non-destructiveness, non-invasiveness). Color imaging has emerged as one of the most promising tools for PAT
purposes [76], although its usefulness has been reported in few studies from the pharmaceutical industry. LIFS, a technology which takes advantage of the fluorescent properties of components, is gaining acceptance in the field, especially when low-dose components are present [53]. All these tools generate large amounts of data, which usually require multivariate analysis [21].

Several studies, reported in the literature involving process monitoring via PAT tools, have focused on solid particles, since they represent the bulk of pharmaceutical production [105][106]. Different stages in the manufacturing process, such as blending and granulation, have been investigated in PAT applications with the help of these tools [104]. However, despite the importance of the tableting step in final product quality, not much work has been done on the utility of PAT tools to monitor powder flows entering the tableting stage. Reports in the literature demonstrate the possibility of monitoring the composition of powders present in the feed frame of a press machine and detecting segregation phenomena with NIRS technology. Ward et al. [9], used a step change technique to demonstrate the suitability of two NIR probes to detect differences in the concentration of the mixtures (API contents ranging from 3.5% to 30%) and linked their potency to off-line measurements of the tablets. They also studied the influence that certain parameters (throughput rate and rotation speed of the feed frame speed) had in the signal output of the tool. Also, work by Wahl et al. [16] monitored the composition of the API (30%) and of two excipients (65% w/w in total) using an in-line NIR probe, after the development of an off-line PLS model. Along with the detection of drug concentration, they also proved the utility of this technique to detect segregation in the process, which was later confirmed by the off-line tablet analysis.

3.2 Materials and methods

3.2.1 Set-up

The tableting feed frame is part of a Manesty Novapress 37-station rotary tablet press (Figure 3.1). It consists of two intermeshing wheels (each with 10 paddles). Powder enters the feed frame in the inlet chute directly onto the counter clock-wise rotating wheel (right), then goes on to the clock-wise rotating wheel (left), where it is scanned by probes before being
conveyed to dies via an opening in the bottom of the feed frame. This configuration makes immediate monitoring possible before it is transferred to the dies and compressed.

![Feed frame set-up](image)

*Figure 3.1: Feed frame set-up. (A) Powder Inlet, (B) RGB imaging system (including camera, lens and ring light), (C) NIRS probe, and (D) LIFS probe*

The original Plexiglas® window, located above one of the wheels immediately before the powder reaches the tablet punches, was modified to accommodate NIRS (JDSU MicroNIR™ 1700 Spectrometer, 900-1700 nm, 6.2 nm resolution), RGB camera (Basler Pilot 5 MP piA2400-12gc), and LIFS diode array detector (Prozess Technologie, Model 801, with 302-1149 nm, 3.4 nm resolution; St. Louis, MO). LIFS and NIRS probes are located 1 mm above the paddles where an opening in the Plexiglas® window allows RGB pictures to be taken directly without any interference. These three probes are triggered at each rotation by a sensor on the paddle shaft. To avoid interference between the probes (e.g. stray light from the RGB camera may influence LIFS measurements), both the lighting and acquisition steps are activated asynchronously by the trigger. As such, all three probes acquire data at different times over each paddle rotation.

Paddle rotation (kept at constant 6 rpm) creates crests and troughs in the circulating powder (*Figure 3.2*). The 10-blade wheel commonly employed in Manesty presses creates wide, flat troughs between the paddles and crests rising a few millimeters above the paddles.
Based on this arrangement, the best acquisition area is defined for each probe. For both spectroscopic tools (NIRS and LIFS), the parameter taken into consideration is the signal-to-noise (S/N) ratio, while the focus and area of acquisition are used for the RGB camera.

Preliminary results indicate that non-contact probes (i.e. NIRS and RGB) work better when acquiring troughs. As the trough surface is relatively flat, greater measurement repeatability can be achieved. In contrast, the contact probe (i.e. LIFS) yielded better results when acquiring crests.

![Image](image.png)

*Figure 3.2: Wave behavior of the feed frame (A) Trough; (B) Crest*

### 3.2.2 Experimental design

The mixtures were composed of six components and three of them were monitored – referred to as vitamins I, II and III – in a calcium carbonate, vitamin E and magnesium stearate matrix. *Table 3.1* represents the experimental design and tools for detecting each transition. Concentrations in the mixtures are representative of those present in other common multivitamin products. The three components studied were selected for their ability to be detected by NIRS, RGB or LIFS probes. As such, each or the transitions presented in *Table 3.1* could be monitored by at least one of the PAT tools.
Thus, vitamin I is known to be NIRS-responsive, the dark brown color of vitamin II can be tracked by RGB, and vitamin III’s fluorescence makes it a suitable candidate for monitoring by LIFS. It is noteworthy that vitamin III’s yellow tint may also be useful in RGB analysis as it is distinguishable from the dark brown color of vitamin II.

The mixtures were processed in a 6.5 quart V-Blender (filled to 50%) before being manually added to the press hopper. When the hopper ran out of powder, the feed frame was stopped and the next mixture was added for the following run. The feed frame was not cleaned between runs in order to monitor composition transitions. As such, changes in composition of the inlet stream can be likened to successive step changes in composition. As mentioned, each probe measures the blend once every rotation of the feed frame paddle, which translates into approximately 50 acquisitions for each mixture.

### 3.2.3 Software

The trigger and lighting systems were controlled by Labview (National Instruments, Austin, TX). LIFS data were collected by NovaPac (Prozess Technologie, St. Louis, MO). Data analysis was undertaken with in-house Matlab (MathWorks) algorithms.
3.2.4 Data treatment and presentation

The five mixtures (Table 3.1) were analyzed via LIF, NIR and RGB. An overview of the original data is presented in Figure 3.3.

![Figure 3.3: Original acquisitions of the five mixtures (a) NIR spectra, (b) RGB images and (c) LIF spectra](image)

For NIRS, the data were pretreated with standard normal variate before being centered and scaled to unit variance (UnV). For RGB imaging, the original 5 MP images were cropped to the region of interest. The final images (900 x 400 pixels) were first re-shaped into a 2-D matrix (Figure 3.4), centered and analyzed via MIA [34].
Spectral data obtained by LIFS were smoothed with a Savitzky-Golay filter (second-order polynomial) and UnV. Principal component analysis (PCA) was used to interpret the results throughout this work.

3.3 Results and discussion

3.3.1 NIRS

As the presence of a multicomponent matrix (6 components were tested) makes it difficult to quantify single components because of mutual interference, the full spectral range (914-1700 nm) is used in the two-component PCA model (97.2% of variance explained $t_1$: 82.2% and $t_2$: 15.0%). Figure 3.3 a) illustrates a representative NIR spectrum sample of each one of the mixtures. The main differences in the spectrum profile occurred between 1400 and 1700 nm. Figure 3.3 b) reports the NIRS signal variation of $t_1$ throughout the experiment (for clarity, only the first component is shown). Vertical lines in the graphs represent moments when different mixtures are added to the hopper.

The results suggest that the NIR signal appears to be unchanged by the first transition (from mixture 1 to mixture 2), but varies sharply after the addition of mixture 3 (2 to 3). This smooth variation in $t_1$ values seems to be unaffected by the addition of mixture 4 (3 to 4). The results concur with the initial hypothesis (Table 3.1) that the NIRS probe would only be capable of detecting variations in vitamin I content. When mixture 5 (the center point) is
added, \( t_1 \) returns to an intermediate value (\( t_1 = -5 \)) as would be expected, despite the fact that the \( t_1 \) signal is not stable when the last mixture was added.

As mentioned, the first component explains most of the variance and exhibits clear temporal transitions. Explaining only 15% of the variance, the temporal transitions in the second component are far less distinct. *Figure 3.5 b*) illustrates the \( t_1-t_2 \) score plot in which the process follows a clock-wise trajectory. While not obvious in the first component, the use of the second component confirms however that the NIR probe can distinguish between mixtures 3 and 5, processed at 700 and 1300 seconds, respectively.

*Figure 3.5: NIRS results: a) scores of the first component (\( t_1 \) vs time, b) \( t_1-t_2 \) score plot. Both are color-coded to illustrate time*
3.3.2 LIFS

As is the case in quantitative NIRS, it is often useful in LIFS to select a component-specific spectral interval to achieve accurate results and avoid interference by other components. Although vitamin III has a characteristic fluorescent band (around 582 nm), most spectra were characterized (300-800 nm) to both determine if LIFS could be undertaken for on-line analysis and its robustness to variations of nonfluorescent components. The excludes part of the spectra (800-1135 nm) devoid of features and only representing noise in the NIRS portion of the spectrum, as is common with this technology (Figure 3.6 a)).

![Graph showing LIF results](image)

**Figure 3.6:** LIF results: a) scores of the first component ($t_1$) versus time, b) $t_1$-$t_2$ score plot. Both are color coded to illustrate time.
Two-component PCA represents 85.7% of data variance ($t_1$: 65.3%, $t_2$: 20.4%). It is well-established that the presence of color variations may impact the performance of this tool [53]. Here, the dark brown color of vitamin II could influence the signal. Nevertheless, Figure 3.6a), which represents variation of $t_1$, illustrates the response of the tool as it follows vitamin III concentration throughout the experiment (mixtures 1 to 2, 3 to 4 and 4 to 5). Figure 3.6b) represents the PCA of the two first components, in which the process clearly follows a counter clock-wise trajectory in the score plot. These results confirm the capability of this probe to distinguish low-dose blends, in much the same manner as was discussed for NIRS.

3.3.3 RGB

Two-component PCA of the RGB images represents 92.3% of data variance ($t_1$: 70.1%, $t_2$: 22.2%). Figure 3.7a) illustrates variations of both PCA $t_1$ throughout the experiments. As expected when comparing the raw images (Figure 3.3b)) with the $t_1$ variation, a clear transition is seen when mixture 2 is added, which corresponds to the presence of vitamins II and III (Figure 3.7a)). Unlike the NIRS signal, steady state is reached with the RGB signal around 500 s in mixture 2.

Transition between mixtures 3 and 4, which differ by only 2 w/w % of vitamin III but cause significant changes in $t_1$ values, clearly illustrate the sensitivity of the method. With steady state time points occurring during the processing of mixture 1 (1 to 282 s), the noise on $t_1$ appears to be approximately 750 (dimensionless). Thus, transition 3 to 4 had a S/N ratio of 8.3, placing it between typical values for the limit of detection of 3 and the limit of quantification of 10. The final transition, due to the addition of mixture 5, proves once more the sensitivity of this tool for color components, as this change corresponds to decreased vitamin II and III concentrations.

These results reveal that the tool is able to detect variations in vitamin III concentrations similar to those present in multivitamin compounds. As for vitamin II, the results do not make it possible to determine whether signal variation was due to the combination of either components or whether it was simply due to vitamin III.

Figure 3.7b) illustrates the $t_1$-$t_2$ score plot. Unlike the observations made for NIRS and LIFS, the RGB observations appear to remain roughly in the same location over time (following an arc-shaped pattern). This appears to indicate that color imaging is far less
capable of disguising between the mixtures. Rather, observations made at 400, 900 and 1300 s are all located in approximately the same area of the $t_1$-$t_2$ score plot. Further components do not make it possible to distinguish between these time points. Visual inspection of the acquired images confirms that these 3 time points present very similar color levels.

Figure 3.7: RGB results a) scores of the first component ($t_1$) vs time, (b) $t_1$-$t_2$ score plot. Both are color-coded to illustrate time

3.3.4 Dynamics of transitions

As results clearly indicate that the mixtures can be distinguished and that clear transitions are visible between the steady states, kinetic models were computed on the PCA $t_1$ scores. It is
important to note that the following discussion is based on the behavior of $t_1$, thus inferring a linear relation between probe output and the concentration of the analyte.

Step changes between batches are used to calculate transition dynamics between each batch. A first-order kinetics model is used for this purpose:

$$t_i = A \exp(-kt) \quad \textit{Equation 4.1}$$

where $t_i$ = PCA score (dimensionless), $A$ = gain (dimensionless), $k$ = rate constant ($s^{-1}$) and $t$ = time (s).

To illustrate these dynamics, Figure 3.8 presents the first transition observed by RGB (addition of mixture 2). A first-order model of $t_1$ scores is calculated by minimizing the sum-of-squares. In this case, $k = 0.175 \pm 0.019$ s$^{-1}$ (95% confidence interval (CI): 95% CI). Such a small CI ($\approx$10% of the parameter value) confirms that first-order models appear to match the experimental results. Table 3.2 presents all the calculated $k$ values with their 95% CI.

![Figure 3.8](image)

\textit{Figure 3.8: Details of the first transition of RGB results – vitamin III profile}

\textit{Table 3.2} reports first-order parameters (k) calculated for each probe at each transition. The results show that all model parameters were statistically significant and characterized by relatively small CI. Nevertheless, they reveal some variability in numerical values of the parameters, both within the transitions of a single probe and between probes.
Table 3.2: First-order k values and 95% CI for every transition

<table>
<thead>
<tr>
<th>Mixture</th>
<th>NIRS</th>
<th>RGB</th>
<th>LIFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixture 1</td>
<td>-</td>
<td>0.175 ± 0.019</td>
<td>0.069 ± 0.005</td>
</tr>
<tr>
<td>Mixture 2</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mixture 3</td>
<td>0.024 ± 0.001</td>
<td></td>
<td>0.052 ± 0.006</td>
</tr>
<tr>
<td>Mixture 4</td>
<td></td>
<td>0.070 ± 0.020</td>
<td>0.107 ± 0.014</td>
</tr>
</tbody>
</table>

The fact that k values can differ for a single probe may be related to varying powder compositions throughout the study. In this sense, if varying powder properties influence the throughput rate of the feed frame, they should impact the dynamics perceived by all three probes. However, these tools acquired different portions of powder flow within the feed frame (troughs: NIRS and RGB; crests: LIFS). While these areas are linked, there is no reason why they would exhibit exactly the same dynamics for all components of the mixture. Since NIRS only tracked a single component which was not detectable by the other probes, it was not possible to directly compare the k values. However, RGB and LIFS shared three transitions during the experiments. As indicated in Table 3.2, the dynamics measured via RGB and LIFS appear to be correlated. The k values calculated in the first transition (mixture 1 to mixture 2) decrease in the transition from mixture 3 to mixture 4 and increase once again in the transition from mixture 4 to mixture 5. This correlation between the dynamics, measured via RGB and LIFS clearly supports the hypothesis that variations in calculated k values may be linked to changes in overall rheology of the blend and may improve process understanding.

3.4 Conclusion

This work was intended to determine the potential of NIRS, LIFS and RGB in monitoring the composition of flowing powder in the feed frame of a pharmaceutical tableting press. Using a six-component mixture, changes in the concentration of three ingredients, all present in pharmaceutically relevant concentrations, were detected by PAT tools.

The results indicate that these probes are capable of monitoring the composition of flowing powders in real time. While three probes were used in this study, the goal here was
not necessarily to propose a set-up in which all of them would always be employed simultaneously. Rather, the goal was to have a versatile system in which the right probe, or probes, would be operated as needed. Here, all three probes successfully monitored the intended products (vitamins I, II and III). This is important as it limits the correlation between probes and leads to better process understanding.

Our work reports the first application of both RGB imaging and LIFS to monitor powder flow in real time inside a feed frame. With such data, it was established that the kinetics of each transition step was well-represented by a first-order model. Although we do not yet have a comprehensive understanding of the underlying phenomena occurring inside the feed frame, it is logical to attribute our observations to different rheological behaviors of the powder mixtures. These probes, alone or combined, have proven to be suitable in studying feed frame dynamics and, consequently, detecting undesirable phenomena, such as segregation.

**Acknowledgements**

The authors thank Pfizer and the Natural Sciences and Engineering Research Council of Canada (NSERC) for their funding. The technical contribution of Jacques Gagné, Henri Gauvin and Marc Couture (Université de Sherbrooke) is gratefully acknowledged. Special thanks for Ovid da Silva for editing this manuscript.
CHAPTER 4 Using multiple Process Analytical Technology probes to monitor multivitamin blends in a tableting feed frame

*Title in french:* Utilisation de plusieurs sondes PAT pour le suivi de mélanges multivitaminés dans le bol de remplissage d'une presse à comprimé

*Authors and affiliations:*

**Pedro Durão,** M.Pharm., Ph.D. Candidate, Université de Sherbrooke, Département de génie chimique et génie biotechnologique, 2500 boul. de l’Université, Sherbrooke, Québec, Canada, J1K 2R1.

**Clémence Fauteux-Lefebvre,** M.Eng., Ph.D, Pos-Doc, Université de Sherbrooke, Département de génie chimique et génie biotechnologique, 2500 boul. de l’Université, Sherbrooke, Québec, Canada, J1K 2R1.

**Jean-Maxime Guay,** M.Eng., Pfizer Global Supply, Process Analytical Sciences Group, Montreal, Quebec, Canada

**Nicolas Abatzoglou,** M.Eng., Ph.D., Full Professor, Université de Sherbrooke, Département de génie chimique et génie biotechnologique, 2500 boul. de l’Université, Sherbrooke, Québec, Canada, J1K 2R1.

**Ryan Gosselin,** M.Eng., Ph.D, Associate Professor, Université de Sherbrooke, Département de génie chimique et génie biotechnologique, 2500 boul. de l’Université, Sherbrooke, Québec, Canada, J1K 2R1.

*Date of acceptance:* 8 November 2016

*Status:* Published

*Reference:* Talanta 164 (2017) 7–15
Summary

Content: After the assessment described in the previous chapter regarding the potential of these tools to be used in-line in the monitoring of the components in the feed frame, the next step was to evaluate the accuracy of quantitative models. Therefore, in this chapter, PLS models are developed for five components present in a multivitamin blend. This blend was selected to be studied due to its complex formulation, which would test the accuracy of these tools for the targeted components.

The benefits of combining data from several tools to monitor a component were also evaluated as well as the modification of rheological properties of the blend was assessed through evaluation of its kinetics, obtained after changing the concentration of some components.

Results: Four of the five vitamins in this study could be modeled with accuracy by at least one of the tools, including the monitoring of a low dosage component (by LIF). Combining data from several tools proved to be a good solution to increase accuracy prediction of a component. The contribution of each tool for that prediction (important to select which tool to use in case it is not possible to use all the recommended tools) can also be obtained through chemometrics. It was also concluded that blend kinetics was not affected by the difference of concentration of a component between blends.

Contributions to the thesis: The tools in study (NIR, RGB camera and LIF) demonstrated to be capable to be used as quantitative instruments in the in-line monitoring of a component in a feed frame, even if a complex blend is in place. Moreover, in case of LIF, a low dosage component was able to be tracked with accuracy. It was also demonstrated that the strategy of combining data from different tools promotes better accuracy of the models developed and thus, monitoring of a CQA by more than one tool is encouraged.
Abstract

As PAT implementation grows in the pharmaceutical industry, more studies are being performed to evaluate its suitability in new applications and processes within the manufacturing chain. As the last step in tablet production, the compression stage represents a critical phase that ensures product quality. In-line control put in place at this stage has the potential to detect powder blends that are out of specification limits and, thus, help to improve product quality.

The objectives of the present project are to quantify the composition of a commercial 31-component multivitamin powder blend in real time on an industrial feed frame, using 3 different PAT tools: LIF, NIR and RGB color imaging.

To do so, the concentrations of 5 components (beta-carotene, riboflavin, ferrous fumarate, ginseng and ascorbic acid) were alternately changed and monitored with one or many probes. Transition periods between batches served to quantify different powder flow dynamics with sequential composition step changes.

The results showed that 4 out of 5 components, each present in commercially-relevant concentrations, could be monitored by one or more tools. Flow dynamics were measured and found to vary significantly in different powder blends.

Keywords: Pharmaceutical; Multivitamin; Tablet Press; Monitoring; PAT; Kinetics
Résumé français:

L’implantation de PAT dans l’industrie pharmaceutique est en pleine croissance. De ce fait de plus en plus d’études sont réalisées afin d’évaluer leur pertinence dans de nouvelles applications et dans de nouveaux procédés sur les chaînes de fabrication. Etant la dernière étape dans la fabrication de comprimés, la compression représente une étape critique dans la qualité du produit. Les contrôles en ligne mis en place lors de cette étape permettent de détecter des mélanges dont la composition est hors des limites de spécification et ainsi d’améliorer la qualité du produit.

L’objectif de cette étude est de quantifier la composition d’un mélange multi-vitaminique commercial contenant 31 éléments, en temps réel, dans une trémie d’alimentation de type industriel en utilisant trois technologies analytiques de procédé différentes : LIF, NIR and imagerie couleur (RGB).

Dans ce but, les concentrations de cinq éléments (béta-carotène, riboflavine, fumarate ferreux, ginseng et acide ascorbique) ont été alternativement modifiés et contrôlés à l’aide d’une ou de plusieurs sondes. Les périodes de transitions entre chacun des lots ont permis de quantifier les différentes dynamiques d’écoulement des poudres en changeant de manière séquentielle les concentrations entre les éléments.

Les résultats montrent que 4 des 5 composés, tous présents à des concentrations représentatives d’un produit commercial peuvent être mesurées par une ou plusieurs technologies. Les dynamiques d’écoulement ont été mesurées et il a été mis en évidence que ces derniers variés de façons significatives en fonction de la composition du mélange.

Mots-clés: Pharmaceutique Multi-vitamine; Presse à comprimé; Suivi en ligne; PAT, Cinétique
4.1 Introduction

Pharmaceutical science and engineering are strongly regulated fields because of their potential impacts on human health. Numerous requirements are in place to ensure that only high-quality products are manufactured. Historically, before being released on the market, final product testing happens through validated and frequently-destructive techniques (e.g., dosage uniformity, dissolution). If these analyses fail to meet prevailing norms, the entire production batch is typically rejected (only exceptionally can it be reprocessed), which causes significant delays and financial losses [107]. However, the paradigm has been changing in recent years.

After the publication of Q8 [2], Q9 [3] and Q10 [4] guidelines, along with 2004 FDA PAT guidance [1], the focus has remained on monitoring, understanding and controlling processes, with intermediate and final product testing as complementary methods, to ensure product quality (especially when a PAT approach is implemented). This is done by controlling critical quality attributes of the process, preferably in real time, which could avoid subsequent out-of-specification features. Problems of sample representativeness and destruction are also reduced through PAT. These advantages have led to the development of PAT adapted for different steps of the manufacturing chain, such as blending, granulation, drying, compression and coating [104]. Although their usage throughout the manufacturing process is relevant, certain critical steps (e.g., drying and blending end-points as well as tableting) may yield more expedient knowledge of final product quality.

Selection of the most suitable PAT tools is, therefore, of the utmost importance. It depends on processing line characteristics (e.g., line speed, geometric configuration of targeted equipment) as well as on sample and component properties to be tracked (e.g., chemical composition, physical features, state, concentration). It is also clear that monitoring multicomponent mixtures can be challenging because of component-component interference [108]. In light of these constraints, NIR and Raman spectroscopy are often chosen (many APIs are responsive to such techniques), but the results of visible imaging [76], fluorescence [49], focused beam reflectance [109] and acoustic emissions [110] have also been reported. Except for visible imaging which is mainly used in packaging (although some have demonstrated its utility as a PAT tool in coating [74], blending [78] and granulation [111] processes), most of these techniques have been tested in drug product manufacturing.
4.1.1 Application of PAT tools in the compression stage

A literature search has shown that, in comparison to other manufacturing stages (e.g. blending, drying), the compression phase has not been explored extensively despite its many advantages. While 70% of pharmaceutical dosage forms are tablets [112], tableting has seldom been considered for PAT applications. Being the final step in the manufacturing chain, where the product can be accessed in powder form, quantitative control at this stage might make it possible to avoid further and unnecessary testing of the final product, potentially leading to real-time product release [113] (dissolution testing is often mandatory). Monitoring of powder entering the dies could provide information on phenomena occurring in powder blends during discharge through the hopper or inside the feed frame (mainly segregation and agglomeration) [86]. In segregated batches, it is expected that significant differences exist in the concentrations of ingredients within powders that can be detected by the right PAT tools [114]. They also make it possible to study powder flow dynamics [52] inside the feed frame, which can be of particular interest in the development stage, especially in relation to laboratory powder rheology testing [115]. In fact, different powders are expected to have different flow kinetics because of their properties (such as particle size and shape, density), the environment (external humidity) and interaction between them and the feed frame (attrition). Since free flowing powders tend to move freely and allow faster mingling, powder kinetics are expected to change in environments that potentiate their mixing (paddle-driven, forced movement).

Most work on the compression stage has been centered on quantifying tablet composition by NIR [68] and Raman spectroscopy [116], rather than via in-line monitoring of flowing powder passing through the press. However, some studies have demonstrated that powder monitoring is possible at this stage of the process. Ward et al. [65] assessed circulating powders present in the feed frame and reported the possibility of quantifying APIs by NIR. They related the signals obtained with component concentrations. Wahl et al. [16] tested NIR in the feed frame and detected segregation in mixtures being processed.

In CHAPTER 3, the possibility of monitoring powder components in the feed frame with NIR, LIF spectroscopy and RGB color imaging is suggested. In that work, the relative concentrations of a 6-component powder blend were varied. All 3 probes were able to monitor the composition of certain blended components. RGB imaging tracked color components, NIR
traced some relatively organic high-concentration components, and LIF followed low-dose fluorescent components. Our study disclosed the potential of LIF as a PAT in laboratory settings. Moreover, LIF discerned a broad range emission spectrum, which contributed to overcoming the lack of specificity of some equipment in tracking these components. It is also possible to integrate a selective excitation band and thus increase component selectivity. Despite upgrades of these technologies, the general deployment of LIF must take into account the interference that some components might present (especially the colored ones) in the resultant signal as well as other sources of variability, such as particle size distribution and moisture [53]. In this sense, multivariate techniques and spectral range selection can quantify and increase the specificities of tracked components. All these features make LIF a powerful tool for monitoring low-dose components (as low as 0.1% w/w) [101].

In the present work, 3 different PAT tools (NIR, LIF and RGB) were installed in the feed frame of an industrial tableting press to monitor the composition of 5 key components (beta-carotene, riboflavin, ferrous fumarate, ginseng and ascorbic acid) of a 31-component powder blend indicative of typical multivitamin blends. First, this study seeks to ascertain if these PAT tools are capable of detecting composition step changes in complex multicomponent blends. Second, it attempts to determine if they can quantify powder flow dynamics inside the feed frame. Our investigation of dynamic behaviour is intended to evaluate if small concentration differences between batches significantly alter powder throughput rates, a step forward in better process understanding.

4.2 Materials, methods and techniques

4.2.1 Experimental feed frame set-up

A 37-station Manesty Novapress rotary tablet press was employed in this study. Using 19 of the stations to limit unnecessary powder consumption, its speed was set to produce 530 tablets/min. Tablet weight was adjusted to 1.76 g in every run. The feed frame of the press consisted of 2 intermeshing wheels with 10 paddles rotating at 20 rpm. The PAT tools were located on top of the feed frame, right before the powder exited to the dies, to best represent tablet composition.

The PATs involved were: NIR (Viavi MicroNIR™ Pro 1700 ES, 900-1,700 nm, 6.2 nm resolution), RGB camera (5 MP Basler Pilot piA2400-12gc) with light system, and LIF
(Prozess Technologie, iPAS 801, diode array detector, 302-1,149 nm, 3.4 nm resolution). Figure 4.1 is a schematic of the entire set-up.

![Figure 4.1: Feed frame set-up: (A) Powder inlet, (B) RGB imaging system (including camera, lens and ring light), (C) NIR probe, and (D) LIF probe](image)

A sensor located in the feed frame and activated every second rotation (every 6 s) ensured data acquisition and storage. Taking crests and troughs in the powder surface (caused by the paddles) into account, specific acquisition areas were defined earlier for each tool to maximize signal-to-noise (S/N) ratios. In terms of tool positioning, study areas and acquisition sequence, the reader is referred to CHAPTER 3.

PAT acquisition parameters were defined in preliminary tests. Their conditions are enumerated in Table 4.1.

<table>
<thead>
<tr>
<th>Equipment</th>
<th>Delay time (ms)</th>
<th>Integration time (μs)</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIR</td>
<td>470</td>
<td>1 000</td>
<td>50</td>
</tr>
<tr>
<td>LIF</td>
<td>780</td>
<td>80</td>
<td>1</td>
</tr>
<tr>
<td>RGB</td>
<td>1 850</td>
<td>300</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4.1: Equipment conditions (number of samples: observations averaged for each acquisition)
Both NIR and RGB (including the light system) were controlled by Labview software, while LIF data were acquired by NovaPac software from Prozess Technologie.

### 4.2.2 Component selection and description

A commercial 31-component mixture, composed of vitamins and minerals, served as reference blend throughout the study. This mixture was selected not only because it contained suitable components to be monitored by NIR, LIF and RGB, but also because its complex composition represented an important test of tool adequacy in industrial contexts. From this mixture, 5 vitamins (beta-carotene, riboflavin, ferrous fumarate, ginseng and ascorbic acid) were selected for quantification throughout the study by alternately overaging their concentrations. Pyridoxine, another vitamin, was the 6th component: its concentration was varied at 2 levels, although it was not quantified in this work. Its inclusion in the study attempted to assess detection by the analytical tools as an indication of their eventual quantification capacity. *Table 4.2* presents the usual concentration range of selected vitamins in this type of formulation.

*Table 4.2: Reference values of tracked vitamins*

<table>
<thead>
<tr>
<th>Vitamin</th>
<th>Beta-carotene</th>
<th>Riboflavin</th>
<th>Ferrous fumarate</th>
<th>Ginseng</th>
<th>Ascorbic acid</th>
<th>Pyridoxine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration % (w/t)</td>
<td>&lt;1.0</td>
<td>&lt;2.0</td>
<td>2.0-8.0</td>
<td>2.0-10.0</td>
<td>5.0-15.0</td>
<td>&lt;1.5</td>
</tr>
</tbody>
</table>

To select these components, the NIR response, color and fluorescence properties, along with PAT tool characteristics (sensitivity, detection limit), were taken into consideration to ensure that at least one of the tools could monitor each component. Along with these criteria, vitamins with a different concentration range (from low-concentration beta-carotene, riboflavin and pyridoxine to high-concentration ascorbic acid) were selected to test the detection limits of the PAT tools in a complex matrix and to ensure that some of them would be detectable.

Riboflavin, ferrous fumarate and ascorbic acid were expected to present a NIR signal, but riboflavin concentration could prevent its detection. The fluorescent properties of riboflavin,
ginseng and pyridoxine made them suitable for LIF, and the distinctive color features of beta-carotene, riboflavin, ferrous fumarate and ginseng were thought to be appropriate for RGB analysis, although the beta-carotene and riboflavin concentrations could hinder their detection.

4.2.3 Experimental design

The trial design attempted to create conditions for the development of quantitative models to track each of the 5 components, facilitating the study of powder flow dynamics at the same time. The goal was to prepare individual batches in which the concentration of a single component differed from the reference batch. *Figure 4.2* presents the study’s design relative to label claims. It must be noted that all concentrations of the remaining 25 components were kept constant in all batches.

In this sense, the 3 first batches (batches 1 to 3) were prepared to serve as reference for the following batches. Thereafter, 2 overage levels were chosen for beta-carotene (batches 4 and 5), riboflavin (batches 7 and 8), ferrous fumarate (batches 10 and 11) and ginseng (batches 13 and 14). Three batches with different ascorbic acid concentrations were intercalated between them (batches 6, 9 and 15) along with a batch prepared with a lower concentration of pyridoxine (batch 12).

![Figure 4.2: Composition and organization of prepared batches (numbers represent the overage % of each vitamin in terms of label claims)](image-url)
About 7.5 kg were prepared for each batch, taking into account a validated pharmaceutical manufacturing process with equipment and time adapted from proprietary, industrial, scale-down methodology. The experiments called for the consecutive addition of each batch to the hopper of the tableting press when the quantity of powder remaining from the previous batch ran down. Timing of the change was triggered by an automated low powder-level sensor installed on the press.

4.2.4 Data treatment

- Preliminary selection

The 15 batches analyzed represent a total of 1,209 observations: 1 acquisition at each 6 s (around 80 observations per batch). The 3 probes were positioned in rapid succession to ensure that they had: 1) acquired equivalent samples, and 2) gathered the same number of acquisitions. Some observations were excluded from the analysis when powder levels were too low. As batch change-over took place after the machine detected low powder levels in the hopper, some acquisition occurred during out-of-specification conditions. These periods could easily be identified in the data as they drastically altered the spectral characteristics of the NIR and LIF probes and led to unfocused RGB images. After these exclusions, 1,008 observations were recorded in the study. It is worth mentioning that faulty acquisition by one of the tools excluded the time point as a whole so that correspondence between observations and tools remained unaltered.

In NIR and LIF, this step was straightforward and gave matrices of $1209 \times 255$ (LIF) and $1209 \times 126$ (NIR). In RGB, 3D image data (consisting of red, green and blue signals for each pixel) were analyzed by MIA as reported by Macgregor et al. [34], and converted to matrices of $1209 \times 1024$.

- Model points selection

Since composition was altered by sequential step changes, a transitory period between 2 steady-states was expected (Figure 4.3).

Acquisitions were therefore divided into 3 distinct datasets: transition points (red symbols) served to evaluate powder flow dynamics but not to characterize model fit as true
powder concentrations were not known during these times. Steady state points were divided into 2 datasets: calibration points (green symbols) and test set (yellow symbols).

![Figure 4.3: Schematic representation of transition with a step change technique ("□" in green color represents calibration points, "○" in yellow color represents test set, and "◊" in red color represents transition points)](image)

Transition points were defined for each vitamin with data from the tool that presented the sharpest transition identified by Principal Component Analysis plotting of the first principal component (PC) over time. They were selected by visualization of that curve, considering the period from the beginning of the curve until the signal stabilized. These observations (totaling 329 points), marked as transition points, were not part of either calibration or testing of the models built. The 679 remaining observations were then split into calibration (343) and test (336) points.

After selecting and separating data into these groups, PLS models were developed for each vitamin (except pyridoxine) with each PAT tool. Considering the 5 vitamins and 3 tools, this represents 15 PLS models. In addition, data from all tools were concatenated into a single large matrix (679×1,405) and used to produce 5 more PLS models, 1 for each of the vitamins. After analysis of the R² values obtained, an extra model comprising concatenation data from RGB and LIF (679×1,279) was built for riboflavin. Another one, comprising NIR and RGB data (679×1,150), was constructed for ferrous fumarate. Finally, data combination was also undertaken between NIR and LIF (679×381), so that a model could be generated for ginseng as well. This step attempted to evaluate if the inclusion of inadequate tools to monitor the
vitamin was responsible for the addition of noise to the total model. Before building the models, NIR data were pretreated with SNV followed by auto-scaling, RGB data were pretreated with MIA and auto-scaled, and LIF data were treated with a first-order Savitzky-Golay filter followed by auto-scaling. All data were treated and models built with PLS Toolbox and in-house Matlab software.

4.2.5 Dynamics

To study powder flow dynamics, data from all tools were combined and PLS models were built for each of the vitamins (except pyridoxine). The tools were combined because not only did they provide the best prediction (discussed in CHAPTER 3) in comparison to their use alone but also because preliminary modelling showed that all 3 probes identified similar powder dynamics for any given transition. The concentration step changes throughout this trial created transition periods. The study consisted of selecting time periods when the concentration of a given vitamin was changing by plotting predicted concentration through those transition periods. A kinetic equation was fitted to calculate its rate constant ($k$). Owing to the nature of our study design, 3 different transitions were observed for beta-carotene, riboflavin, ferrous fumarate and ginseng, with 4 transitions for ascorbic acid.

As in a previous work [104], a first-order kinetics model was found to adequately fit concentration transitions observed in trial design. The formula is defined in Equation 4.2:

$$\hat{y} = A exp(-kt) \quad \text{Equation 4.2}$$

where $\hat{y}$ is predicted relative concentration (dimensionless), $A$ is gain (dimensionless), $k$ is the rate constant (s$^{-1}$) and $t$ is time (s).

4.2.6 Chemometrics

- **PLS**

Chemometrics are intimately connected with PAT due to the amount of correlated data these tools generate. PLS analysis is one of the most documented methods because of its success in establishing relationships between observation $X$ ($N \times K$) and one or more attributes
in study $Y$ $(N \times M)$, such as the concentration of an ingredient. In its mathematical development, it seeks to maximize covariance between $X$ and $Y$ through the development of score $T$ $(N \times A)$, where $A$ is the number of components), and $W$ is loading weight $(K \times A)$. In this study, the PLS technique was applied to generate models for the prediction of different vitamins.

When working with multiple datasets, as is the case here with NIR, LIF and RGB, it is often relevant to merge the data into a single large $X$ matrix after adequate scaling. While this may improve final prediction, the individual contribution of each original data matrix is lost. It may therefore be interesting to maintain the individual identity of the original data blocks in the model in the form of a mbPLS model.

- **Multiblock PLS**

  Data concatenation is useful when several data sources are in place. Since several tools were used in this study (NIR, LIF and RGB), it is possible to include the technique and therefore merge all information in a single matrix. However, although simple PLS analysis is often conducted to obtain the model, it is helpful in certain cases to differentiate some blocks (usually defined parts of data coming from different tools) of merged data to collect information. In this case, mbPLS analysis may be considered [20]. In such mathematical development, prediction does not improve when compared to regular PLS analysis, but it is possible to interpret the amount of variance coming from each block and contributing to the overall value of prediction. Its mathematical development is based on PLS being the greatest difference with the merging of scores calculated from each block into a larger matrix $T$ (super-block), so that:

$$ T = [t_1 \ t_2 \ ... \ t_I] \quad Equation\ 4.3 $$

$T$ is then used to compute super weights ($W_T^T$) which describe the relative importance of each block [117]. Figure 4.4 shows the schema of mbPLS mathematical development.

When only 1 variable is modelled (as is the case in this work), steps 6 and 7 of Figure 4.4 are redundant since $u$ equals $Y$. 
Block importance in prediction (BIP)

In PLS analysis, the explanatory variables that play an important role in model prediction (thus allowing inference of which variables are more important to the prediction of Y) can be highlighted through calculation of the variable importance of projection. This index is the result of contribution of the variable in every component, taking into account the percentage of variability explained by that component. In mbPLS, the same computing method is adapted to calculate BIP which is represented in Equation 4.4 [119]:

\[
BIP^\text{n}_b(A) = \sqrt{B \cdot \sum_{a=1}^{A} \frac{\mu_{b,a} ||P_{t,a}y_n||_2^2}{\sum_{l} \mu_{b,a} ||P_{t,l}y_n||_2^2}}
\]

Equation 4.4

where:
A = Total number of components (1, 2, … , a)
B = Total number of blocks (1, 2, … , b)
n = Index of Y blocks (n = 1, 2, … , N)
\(\mu_{b,a}\) = Weight vector of the \(a^{th}\) component
D = Scaling metric
The equation shows that $\sum_k BIP_k^N(A)^2 = B$, which determines that a block with a number higher than 1, represents an important contribution to the model.

4.3 Results and discussion

4.3.1 Model building

- **Analysis and interpretation**

The analysis of each PLS model began with the qualitative evaluation of score plots. This step aimed to identify batches containing overaged vitamins in the plot and to determine if any other batches stood out, indicating potential interactions between certain components that may have an impact on the developed model. Therefore, the score plot presented a preliminary indication of the potential of the vitamin to be monitored, but no definitive conclusions could be drawn without qualitative analysis of the calibration and test samples.

To illustrate data analysis, the model obtained with NIR for ferrous fumarate was used throughout this section. The score plot of the first 2 PCs is presented in *Figure 4.5* and illustrates typical representation of this experiment.

In the plot, different concentrations of ferrous fumarate (batches 10 and 11: 35% and 70% of overage) are grouped by color and style. Clear separation is seen between them and the rest of the batches, especially along $t_1$ (37.75% of variance explained). However, although the model is built in relation to ferrous fumarate, 4 other groups of observations (surrounded by circles in the mentioned figure) are identified. They correspond to batches 6, 9, 14 and 15. While these batches all have the same nominal concentration of ferrous fumarate, their matrices are different enough to be detected as different entities. However, only batch 14 varied along $t_1$ (like overaged batches 10 and 11), which is an indication that the vitamin overaged in this batch might have a negative impact on the prediction of ferrous fumarate. Therefore, although $t_2$ explains 40.71% of existent variance, the effect of batches 6, 9 and 15 in predicting ferrous fumarate in the model remains unclear.
As mentioned, score plots were analyzed for 23 PLS models. The objective was to evaluate which other vitamins could potentially have an impact on modeling of the vitamin being predicted. While interpretation of the first 2 PCs of the score plot provides an indication, we must keep in mind that the final models used to monitor a vitamin are usually developed with more PCs, thus mitigating the influence that this variability might have in model prediction. Therefore, the isolated batches in Figure 4.5 might not be significantly relevant in the final model, as shown in Figure 4.6, where a predicted concentration is plotted with 2 PCs (Figure 4.6 (a)) and 6 PCs (Figure 4.6 (b)). Both graphs illustrate that when changes in composition are present, the predicted concentration samples are located around the theoretical value. However, in Figure 4.6 (a), along with underestimation of the concentration of the studied vitamin around 170%, it is observed that batches 9 and 14 are far from the theoretical value (these batches were already isolated in the corresponding score plot presented in Figure 4.5). When more components are used (as seen in Figure 4.6 (b)), the prediction of the vitamin is more accurate and batches 9 and 14 are around their theoretical value.

This allows us to conclude that despite the fact that some vitamins might have the potential to interfere with one another, thus, contributing to the model, the inclusion of more PCs might overcome the problem. It is also important to highlight the fact that, as seen in both graphs, a change in concentration leads to a transitory period until the signal stabilizes. When
Figure 4.6: PLS model (NIR): observations vs predicted concentration (in terms of label claims) of ferrous fumarate; (a) 2 PCs, (b) 6 PCs (dark line: theoretical Y value; “□” in green color represents calibration points, “◊” in yellow color represents test set, and “◊” in red color represents transition points)

the latter stage occurs, it means that concentration of the component in the feed frame is constant. These transitory periods were used to ascertain the dynamics of the mixture, which is discussed later.
- Effects of data fusion

The effects of data concatenation (i.e., combining NIR, LIF and RGB data) on model error can be seen in Figure 4.7 through comparison of predicted concentration in the transitory period of ferrous fumarate of a PLS model (4 PCs), using data from the calibration set of NIR only and when data from all tools are combined.

![Figure 4.7: Predicted value of ferrous fumarate with a PLS model using data from (a) NIR and (b) all tools. The dark line represents the theoretical Y value](image)

From this figure, we can see that when data from all tools are considered (Figure 4.7 (b)), predictions closely follow the expected values (dark bars), while in the case of NIR, the data lie mainly over (135%) and under (170%) the expected value, respectively causing over- and
underestimations of predicted concentration. The decreased bias in Figure 4.7 (b), observed throughout the analyses, clearly underlines the advantages of multiple PAT probes in monitoring complex blends.

- Model results and BIP values

The predictive power of PLS models is represented by their adjusted $R^2$ values. Table 3 reports the adjusted $R^2$ for calibration and the test points for each vitamin in study according to the number of components, using individual data (from each tool individually) as well as through their combination (i.e., data concatenation), with all 3 probes, or, when appropriate, the 2 most useful tools for a given product.

**Table 4.3: Adjusted $R^2$ values for predictive models of the vitamins (Cal.: calibration; PC: principal components. Background color represents the best tools for each case)**

<table>
<thead>
<tr>
<th></th>
<th>NIR</th>
<th>RGB</th>
<th>LIF</th>
<th>3 tools combined</th>
<th>2 best tools combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta-carotene</td>
<td>0.25</td>
<td>0.20</td>
<td>0.49</td>
<td>0.91</td>
<td>0.84</td>
</tr>
<tr>
<td>Riboflavin</td>
<td>0.34</td>
<td>0.22</td>
<td>0.78</td>
<td>0.95</td>
<td>0.94</td>
</tr>
<tr>
<td>Ferrous fumarate</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>Ginseng</td>
<td>0.94</td>
<td>0.93</td>
<td>0.63</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>Ascorbic acid</td>
<td>0.89</td>
<td>0.86</td>
<td>0.54</td>
<td>0.91</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Beta-carotene, present in low concentrations (<1.0 w/t %), appears reddish-brown to the human eye. According to the results, LIF was the only tool capable of predicting it, although the adjusted $R^2$ value ($R^2_{cal} = 0.91$) was correlative low compared to the other vitamins. However, since beta-carotene is not fluorescent, this value is likely the result of a phenomenon called auto-absorption (in the visible region) due to its color [53]. RGB and NIR were not expected to monitor beta-carotene considering its low concentration. For these reasons, data concatenation did not significantly improve the predictive capabilities of the models ($R^2_{cal} =$
0.93) and, since neither NIR nor RGB gave good results, none of the tools was able to predict the vitamin at this concentration.

Riboflavin (yellow) was also present in low concentration (<2.0 w/t %). For the same reason and due to its fluorescence properties, LIF was the only tool that delivered good prediction. Nevertheless, its prominent color allowed RGB to detect some variation, but its low concentration prevented better prediction with this tool alone. When all data were combined, with more PCs (4 vs 5), the accuracy of the calibration model increased ($R^2_{\text{cal}} = 0.96$) in opposition of what happened with the test set. When only the 2 best tools were combined (RGB and LIF), and for the same number of PCs, the model slightly increased prediction ($R^2_{\text{cal}} = 0.97$) in comparison to the values obtained when all tools were combined. Although the difference was not very large, it was an indication that the presence of NIR data added noise to the model, decreasing its predictive power.

As far as ferrous fumarate (reddish-brown and 2.0-8.0 w/t %) is concerned, it could be monitored by all 3 tools. In fact, its chemical composition and concentration above the NIR detection limit allowed it to be monitored by this tool ($R^2_{\text{cal}} = 0.97$). Also, its color was critical for RGB tracking ($R^2_{\text{cal}} = 0.96$), and the reported influence that color has on the LIF signal justified its result ($R^2_{\text{cal}} = 0.94$). Since all tools were able to monitor it (although LIF might be considered a misleading model), analyzing all data simultaneously led to an increase in model accuracy at the same time as the number of PCs was decreased ($R^2_{\text{cal}} = 0.97$, 2 PCs). A similar conclusion was drawn when the 2 most suitable tools (NIR and RGB) were combined ($R^2_{\text{cal}} = 0.96$, 3 PC). These results show that although the difference was not significant, removal of a tool which has good prediction power caused a decrease in the $R^2_{\text{cal}}$ result of the combined model (NIR and RGB).

Ginseng (light brown color) was present in a concentration range of 5.0-15.0 w/t %. Its chemical composition and fluorescence properties made it a suitable candidate to be monitored by both NIR and LIF. In this case and despite its color, the RGB tool could not monitor it, probably because of the staining effect rather than a defining effect it had on the mixture. The combined data from all tools confirmed the already good results obtained with these tools alone ($R^2_{\text{cal}} = 0.97$, 5 PCs). When the 2 best tools were combined (NIR and LIF), and for the same number of components, $R^2_{\text{cal}}$ remained constant, but $R^2_{\text{val}}$ improved slightly. As with riboflavin, there was a slight difference of $R^2$ values between combined data with 2 or 3 tools,
although it was also true that combination not including the tool with the worst prediction (RGB) presented better results.

As for ascorbic acid (5.0-15.0 w/w %), it is white and known to be NIR-responsive. Therefore, NIR but not RGB was able to monitor it \( (R^2_{\text{cal}} = 0.89) \). As in situations where the component in study was not fluorescent, the \( R^2 \) value obtained with LIF was justified by the color variation of the mixture (adding ascorbic acid will dilute the concentration of the other components), leading to a misleading model with this tool.

In the case of pyridoxine, its low concentration and fluorescent properties made LIF the only suitable tool to track it. Thus, the initial goal set was to detect a change in the LIF signal after varying its concentration. However, as mentioned earlier, ferrous fumarate also has a significant LIF signal which biases the pyridoxine signal (see concomitant variation of these components in batch 12). Therefore, a conclusion about LIF’s potential in monitoring pyridoxine would still be speculative and it is left for future study.

These results show that the combined use of at least 2 PAT tools can improve component predictions, often without increasing mathematical complexity of the models (number of PCs). However, it is not always possible, or desirable, to deploy multiple PAT tools to monitor a process. In this sense, the tools that are more important must be evaluated by mbPLS models. Table 4.4 reports significance of the BIP values (BIP values of significant tools >1) of each tool based on the PLS models presented in Table 4.3, developed for each of the vitamins, when data from all tools are combined [119].

<table>
<thead>
<tr>
<th>Vitamins</th>
<th>NIR</th>
<th>RGB</th>
<th>LIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta-carotene</td>
<td>0.96 (0.79-1.13)</td>
<td>1.00 (0.82-1.16)</td>
<td>1.05 (0.90-1.22)</td>
</tr>
<tr>
<td>Riboflavin</td>
<td>0.67 (0.48-0.90)</td>
<td>1.05 (0.84-1.27)</td>
<td>1.28 (1.13-1.44)</td>
</tr>
<tr>
<td>Ferrous fumarate</td>
<td>1.39 (1.29-1.51)</td>
<td>0.87 (0.75-1.00)</td>
<td>0.74 (0.66-0.82)</td>
</tr>
<tr>
<td>Ginseng</td>
<td>1.17 (1.00-1.34)</td>
<td>0.48 (0.35-0.63)</td>
<td>1.36 (1.13-1.53)</td>
</tr>
<tr>
<td>Ascorbic acid</td>
<td>1.21 (0.92-1.45)</td>
<td>0.61 (0.45-0.80)</td>
<td>1.18 (1.00-1.41)</td>
</tr>
</tbody>
</table>

Through analysis of the results in Table 4.4, the significance of each tool in this combination matrix for the prediction of each vitamin could be highlighted. As for beta-
carotene, the fact that the model developed was not as good as others made the information non-significant relative to its contribution. Although LIF demonstrated better prediction (as mentioned, its prediction was due to other causes), it was observed that the contribution of the other tools was also significant, despite their low predictive power when used alone. This information raises the possibility that the vitamin might be monitored by these tools in different conditions (e.g., a less complex type of matrix, a more concentrated component). Riboflavin’s color and fluorescence properties made it detectable by these tools with greater significance than NIR, which probably might be related to its low concentration in the mixture. The inverse occurred in the case of ferrous fumarate, as NIR demonstrated that it was a more adequate monitoring tool. Finally, ginseng and ascorbic acid had a larger possibility of being detected with NIR and LIF. Overall, as expected, it might be concluded that tool selection must depend on the vitamin to be monitored, its specificity for the technology used, with vitamin concentration being an important factor in the decision. LIF monitoring is preferable when concentration is low and fluorescence properties are apparent. RGB monitoring was demonstrated to be less efficient, although it was able to show a significant contribution with a low concentration component, such as riboflavin. Therefore, it might be a tool to consider when the component in study has color and no fluorescent properties. As for NIR, a significant contribution of the models was demonstrated, except in cases when the vitamin concentration in study was low (beta-carotene and riboflavin).

4.3.2 Dynamics

- Feed frame as ideal vessel

The rheological properties of each component present in powder (particle size, shape and density) can influence its relative flow and behavior as a whole. Therefore, different concentrations of the same component in a multicomponent powder, like the one used in this work, might present different flow dynamics. Figure 4.8 illustrates the transition observed (from batch 10 to 11), using the predicted concentration obtained with our example (NIR and ferrous fumarate), with a first-order equation (red curve) fitted by minimization of the sum of squared errors (SSE). The k of the real vessel calculated through Equation 4.2 had a value of 0.067 s⁻¹ (0.056-0.076, with 95% CI). If continuous stirred-tank reactor behavior of the feed frame is considered, and taking into account powder density, the total and constant volume of
the feed frame occupied by the powder and the throughput rate, an ideal vessel would have a k of 0.021 s\(^{-1}\).

![Graph showing transition of ferrous fumarate with PLS model (NIR, 5 PCs)](image)

*Figure 4.8: Detailed transition of ferrous fumarate with a PLS model (NIR, 5 PCs)*

From comparison of k values for an ideal vessel and feed frame, it can be concluded that ideal behavior is far from being obtained. One possible explanation is linkage of the powder’s transport mechanism from the feed frame to the dies. In fact, when it enters the feed frame, a certain amount of powder is immediately directed towards the dies because of forced movement of the first wheel’s paddles. This powder never reaches the second wheel (which is, at the same time, directing powder towards the dies), making transition faster. Since the PAT tools presented here are located over the second wheel, it is expected that k is lower than that obtained theoretically for the first wheel.

- **k calculation**

Although the feed frame does not behave as an ideal vessel, the transitions studied fitted well in a first-order kinetics model, as seen in *Figure 4.8*. All transitions studied (17 in total, since pyridoxine was not included) were subject to individual testing of its dynamics to calculate their k value computed from a first-order equation. However, due to the low S/N ratio, which can be inferred from *Figure 4.8*, there was a large associated error which prevented any conclusion regarding differences or trends within batches. Therefore, it was assumed that the same dynamic would be observed for the same vitamin overage. This led to
the computation of a single $k$ value by simultaneously minimizing the SSE of each run of a given vitamin. Figure 4.9 presents the calculated $k$ for all the vitamins in study along with the associated error. From its analysis, we can conclude that no significant differences in dynamics were apparent between components, with the exception of beta-carotene. However, as highlighted earlier, the model for this vitamin was not as good, biasing the calculation of its dynamic values (which could be verified by the large associated error). Although some differences were expected to occur because of different component rheologies, mixture complexity and the small differences between them (in some cases, a few grams in a 7.5 kg batch) were most likely responsible for similar mixing behavior of all batches.

![Graph showing k values and overall value of each studied vitamin](image)

*Figure 4.9: k values and overall value of each studied vitamin*

It can be concluded that the dynamic behaviors of all powders passing through the feed frame were similar, which allowed calculation of the $k$ associated with the mixture, using all 14 transitions (beta-carotene was not included) observed during the trial and allowing minimization of the associated error in its calculation. The result obtained from the optimization of all transitions and the strict confidence interval associated with it allows good definition of the general $k$ in the trial $[0.093; 0.105]$ (“Total” in Figure 4.9).
4.4 Conclusion

The objective of this work was to monitor the composition of certain key components in a multivitamin powder blend in the feed frame of a tableting press in real time. Three PAT tools monitored the circulating powder blend: LIF, NIR and RGB color imaging.

Of the 31 components present in the blend, 5 (beta-carotene, riboflavin, ferrous fumarate, ginseng and ascorbic acid) were chosen and monitored in situ with multiple composition step changes. Four of the 5 vitamins, all present in industrially-relevant concentrations, could be monitored by at least one of the tools, according to their physical characteristics (e.g., color) and concentrations. The combination of more than one tool (i.e., via data concatenation) made it possible to limit validation bias, enhancing model quality. With tool combination, all models developed were validated with an adjusted $R^2$ value higher than 0.83 and up to 0.97 with just 2 PCs. The results demonstrate the potential of these tools to be implemented in the feed frame for monitoring concentrations.

The flow dynamics of each transition were evaluated and calculated with first-order model kinetics. No significant differences in circulation rates were detected between components, probably owing to high complexity of the blend, the relatively small changes in overall blend composition, and the small rheological differences caused by these variances.

Acknowledgements

The authors thank Pfizer and the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding this project. The technical contributions of Jacques Gagné, Henri Gauvin and Marc Couture (Université de Sherbrooke) are gratefully acknowledged. Special thanks to Ovid da Silva for editing this manuscript.
CHAPTER 5 Specificity of Process Analytical Tools in the monitoring of multicomponent pharmaceutical powders

Titre en français: Spécificité des outils analytiques de procédé dans la surveillance des poudres multicomposants pharmaceutiques

Authors and affiliations:

Pedro Durão, M.Pharm., Ph.D. Candidate, Université de Sherbrooke, Département de génie chimique et génie biotechnologique, 2500 boul. de l’Université, Sherbrooke, Québec, Canada, J1K 2R1.

Clémence Fauteux-Lefebvre, M.Eng., Ph.D, Postdoc, Université de Sherbrooke, Département de génie chimique et génie biotechnologique, 2500 boul. de l’Université, Sherbrooke, Québec, Canada, J1K 2R1.

Jean-Maxime Guay, M.Eng., Pfizer Global Supply, Process Analytical Sciences Group, Montreal, Quebec, Canada

Nicolas Abatzoglou, M.Eng., Ph.D., Full Professor, Université de Sherbrooke, Département de génie chimique et génie biotechnologique, 2500 boul. de l’Université, Sherbrooke, Québec, Canada, J1K 2R1.

Ryan Gosselin, M.Eng., Ph.D, Associate Professor, Université de Sherbrooke, Département de génie chimique et génie biotechnologique, 2500 boul. de l’Université, Sherbrooke, Québec, Canada, J1K 2R1.

Date of acceptance: Submitted

Status: Submitted

Reference:-
Summary

Contents: In the previous chapter, the potential of the tools to quantify certain components was evaluated, even if they are part of a multicomponent blend and / or present in low concentration. However, to implement this technology as monitoring / control tools, the models developed must take into consideration all the variations that can occur in the process. Therefore, it is essential that a good design is built to obtain a robust model and therefore accurately predict the target component. This chapter is aimed at evaluating the specificity of these tools to the component intended to be monitored. Using an off-line testing, two full factorial designs of four vitamins were applied to evaluate the accuracy of the models built and assess any potential interference that could exist in the signal obtained from these tools.

Results: As expected, since a significant difference in design was established in comparison with the one studied in the previous chapter (simple step change in chapter 5 vs a full factorial design in the present chapter), the models developed presented a lower R^2. However, if the complexity of the blend is considered, the results demonstrated a good selectivity for the targeted components. Nevertheless, some interference (especially in the LIF tool) was identified and although it did not prove to be statistically significant, special attention to this factor must be taken in future implementations.

Contributions to the thesis: This chapter intends to propose a general method to evaluate PAT robustness. To do that, the 3 tools were used to evaluate the selected components in the multicomponent blend used, in order to track them with an acceptable degree of accuracy, taking into account the regulatory environment. For other blends (those whose stricter regulation), it is important to take into account certain interferences that might occur and that could deplete the required selectivity of the tools. The interference occurring when LIF is applied might play an important role in the prediction of the target component.
Abstract

The application of Process Analytical Technologies in pharmaceutics manufacturing has been the subject of many studies. Active pharmaceutical ingredient monitoring in real time throughout the manufacturing process is commonly the target of many such implementations. The tools in place must be sensitive to, and selective of, the parameter(s) to be monitored, i.e., in the case of component quantification, they must be sensitive to the component in question and robust against all others. In the present study, 4 different ingredients (riboflavin, ferrous fumarate, ginseng and ascorbic acid) in a multicomponent blend were monitored by 3 different tools (near infrared spectroscopy, laser-induced fluorescence, and red-green-blue camera) using a full factorial design. The goal was to develop efficient and robust concentration-reading/prediction models able to assess and monitor component interference. Despite relatively high complexity of the blend studied, the 3 tools demonstrated reasonable specificity for the tracked ingredients, taking into account larger acceptance criteria typical of dietary products. In certain cases, some interference might lead to biased predictions, highlighting the importance of good calibration. The tools tested and the methodology proposed have divulged their potential in monitoring these components, despite the complexity of the 31-component blend.

Keywords: Process analytical technology; near-infrared; light induced fluorescence; color imaging; specificity; interference
Résumé français:
Les recommandations des institutions de réglementation sur la mise en œuvre de la “Quality by Design” et les avantages qui en découlent dans le développement pharmaceutique, en particulier dans la phase de production, conduisent l'industrie pharmaceutique à appliquer des PAT dans leurs procédés. La surveillance de l’API en temps réel tout au long du procédé de fabrication est habituellement la cible de plusieurs de ces implémentations. Les outils mis en place doivent être sensibles et robustes aux paramètres à surveiller, c'est-à-dire, dans le cas de quantification d'une composante, ces outils doivent être sensibles à la composante et sélectifs à tous les autres, ce qui est influencé par la complexité de la matrice.

Dans cette étude, quatre différents API (riboflavine, fumarate ferreux, ginseng et acide ascorbique) présents dans un mélange multicomposant ont été monitorisés par trois outils différents (spectroscopie proche infrarouge, fluorescence par laser et caméra rouge, vert et bleu) utilisant un plan factoriel complet. L'objectif est de développer un modèle d'acquisition / prédiction de concentration efficace et robuste, capable d'évaluer et de prendre en compte l'interférence des composants surveillés. Malgré la complexité relativement élevée du mélange étudié, les outils ont démontré une spécificité modérée pour les ingrédients étudiés, en tenant compte des plus grands critères d'acceptation spécifiques aux produits diététiques. Néanmoins, dans certains cas, une certaine interférence pourrait conduire à des données biaisées, mettant en évidence l'importance d'un bon plan de calibration. Les outils testés et la méthodologie proposée ont démontré leur potentiel dans la surveillance de ces composants, malgré la complexité du mélange.

Mots-clés: PAT, NIR, LIF, RGB Imagerie, Spécificité, Interférence
5.1 Introduction

Pharmaceutical manufacturing companies are increasingly aware of the benefits of replacing conventional product testing (usually based on laboratory work – UV spectroscopy, HPLC) by state-of-the-art PAT. The latter provides non-destructive analysis, statistically representative process sampling and real-time information [120]. Regulatory recommendations have been directing pharmaceutical product development, specifically manufacturing steps towards process monitoring and understanding its limits rather than focusing on the product itself [5], which is at the heart of PAT. This approach, which defines QbD as the methodology to be followed in pharmaceutical development, assigns an essential role to PAT to obtain as much information as possible throughout the development process and the optimization of existing processes. Therefore, PAT:

(a) in the case of a new product, can be part of the first development stages (API definition [121] and characterization [122] – identification, impurity detection, crystallinity, humidity level) – or even applied through the final development stages, and
(b) in the case of already-available commercial formulations, can serve to redesign, monitor and control existing manufacturing processes.

Successful PAT implementation largely depends on the identification of CQA – “physical, chemical, biological or microbiological property or characteristic that should be within an appropriate limit range, or distribution to ensure the desired product quality” [2]. In fact, through risk analysis [3], CQA are identified and tested (design space), defined (control space), monitored and controlled by PAT tools. Once CQA or CPP have been identified, the first step in PAT implementation is tool selection for monitoring via a feasibility study.

In PAT implementation during the manufacturing stage, many CQA might be adapted with prior knowledge of similar processes, such as water quantity and particle size in wet granulation [93] or API quantification in blending [91]. However, tool selection to monitor CQA is strongly product-dependent. For instance, in component quantification, the choice of monitoring tool depends on its response to that tool. In these processes, NIR is largely accepted as a tool to monitor water content [93] while several techniques (NIR [98], Raman [123], LIF [53]) have demonstrated suitable characteristics and high potential in API quantification.
5.1.1 Design considerations

Although published data indicate which tools might be used to monitor certain CQA, process and product specificity make preliminary feasibility studies imperative in PAT implementation. Since PAT tools to monitor CQAs are usually identified through the evaluation of model accuracy, it is critical to take robustness into consideration, so that tool potential can be correctly assessed. On this subject, and given the common use of multivariate techniques, such as PLS in model development, Xiang et al. [124] pointed out the importance of calibration in model building. In their study, which analyzed API in tablets with NIR as monitoring tool, the importance of a randomized design was highlighted, not only as a way to reduce the potential of chance correlations (correlations not related to the target parameter, between process data (X) and predicted parameter (y)), but also to increase model specificity. It proved that randomized designs presented better specificity with varying parameter values, such as excipient and hardness, than correlated designs. The problem of correlations in calibration design was demonstrated in other fields by Rhiel et al. [125], as models developed in ovary cell cultures, based on samples with high correlations between metabolites, only worked if that same correlation was maintained.

In fact, creating ideal conditions to achieve the correct design could be difficult and would be even harder when a complex mixture is assessed. In case of mixtures with several components, their study requires calibration runs to take all variables into account, increasing the number of necessary experiments to obtain reliable results, especially when components within the blend do not have clearly distinguishable features and/or lead to a distinguishable response by a given monitoring tool. Design of experiment (DoE) techniques are important to overcome this problem and reduce the number of runs that a full factorial design would demand [126]. Scheibelhofer et al. [127] underscored the dependence of model performance on chosen DoE. In some cases, despite lacking in robustness, simpler designs are intended to investigate general effects [128], especially when a PAT tool is used to determine the entire blend as a sum of its APIs and excipient [129], relative to others that monitor a specific component [130].
5.1.2 Specificity of PAT tools

The quantitative prediction of components (usually an API) is the most common CQA described in the literature. Often, a single PAT tool monitors several components in blends by multivariate techniques [129][131][132]. However, the goal is usually to monitor the blend as a whole (as a result of all interactions established between all components of the blend and the monitoring tool), as in the determination of mixture endpoints, and not to develop a method to individually quantify each component. To do so, regulations require analyte assessment for technique specificity [41][17]. All too often, this requirement is not considered extensively in published reports. In model development of API quantification, agents potentially interfering with the tool are usually not included in model design and are commonly replaced by binary or tertiary mixtures [133], so that the possibility of interference with other components is decreased. This indicates potential, but does not challenge the tool with more difficult, i.e., realistic, conditions.

Tool selection must take into consideration: (a) the CQA intended to be monitored (b) the environment (humidity, matrix complexity), (c) their own features (in which their specificity and accuracy to the CQA play a critical role), and (d) evaluation after adequate experimental design. In the present work, we propose to test a multivitamin blend with different components, employing 3 different PAT tools: NIR, RGB imaging and LIF. Despite its common use as a PAT tool, NIR is less specific than other spectroscopy tools, such as Raman or Fourier Transform Infrared owing to the fact that the peaks are generally the result of contributions from chemical groups in more than 1 tracked component of the analyzed mixture [134]. However, the fact is that it is both non-destructive and contains characteristic bands, making it possible to identify and quantify the component. Thus, NIR is the most widely-used tool in every step of manufacturing processes. LIF specificity might be enhanced if the equipment is capable of varying the range of excitation and emission specific to each molecule [53]. In the RGB imaging, specificity depends on spatial resolution of the camera and is highly affected by color similarities between components.

Work described in CHAPTER 4 demonstrated the potential of these tools to quantify components in complex mixtures. A multicomponent blend enhanced the inherent difficulty in tracking a single component because of possible signal overlapping. Four components were successfully modeled (riboflavin, ferrous fumarate, ginseng, ascorbic acid) based on data
obtained in-line in a feed frame tablet press by alternately changing the concentration of each component in the study. However, the DoE in this study did not consider the potential impact that simultaneous variations in the concentration of ingredients might have on the accuracy of the tool and, therefore, on the correlation that might exist between components.

In light of the latter, the present study seeks to fill the gap by specifically assessing interactions that might occur between components of a single blend and evaluating model adequacy obtained through an extensive (full factorial) design. By testing the same complex mixtures used in the work developed in CHAPTER 4, as well as the same PAT tools, specificity of the latter is evaluated by identifying which ingredients can interfere with predictions of others, and determine how they are manifested in the results. The expected outcome is a general method to ascertain if confounding is present in a multicomponent blend.

5.2 Materials and methods

5.2.1 Equipment

Powders were profiled by NIR spectrometer (Viavi MicroNIR™ 1700, 900-1,700 nm, 6.2-nm resolution), RGB camera (5 MP Basler Pilot piA2400-12ge), and LIF spectrometer (Prozess Technologie, Model 801, 302–1149 nm, 3.4nm resolution; St. Louis, MO). Table 5.1 defines the parameters of acquisition.

<table>
<thead>
<tr>
<th>Equipment</th>
<th>Integration time (µs)</th>
<th>Number of averaged samples</th>
<th>Spot size (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIR</td>
<td>500</td>
<td>50</td>
<td>15 × 4</td>
</tr>
<tr>
<td>LIF</td>
<td>80</td>
<td>1</td>
<td>4×4 (round)</td>
</tr>
<tr>
<td>RGB</td>
<td>300</td>
<td>1</td>
<td>25 × 22</td>
</tr>
</tbody>
</table>

NIR acquisition was undertaken with Viavi Labview-based software, RGB acquisition with “Measurement and Automation Explorer” (National Instruments) with images of 2,454×2,056 pixels, and LIF acquisition with NovaPac software from Prozess Technologie.

Acquisition was conducted off-line. NIR and LIF data were obtained by direct contact between the probes and sample (both with average penetration depth of about 1 mm,
depending on powder density), while RGB imaging data were acquired with previously-determined, constant distance for clear and focused images.

5.2.2 Selection and description of components to analyse

A commercial 31-component multivitamin mixture was selected as the reference product. Four ingredients (riboflavin, ferrous fumarate, ginseng and ascorbic acid) from the mixture were selected for tracking and to study the application of PAT tools. Table 5.2 presents concentration ranges in these formulations.

Table 5.2: Concentration ranges of ingredients under study: typical values of different multivitamin blends

<table>
<thead>
<tr>
<th>Component</th>
<th>Riboflavin</th>
<th>Ferrous fumarate</th>
<th>Ginseng</th>
<th>Ascorbic acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration % (w/t)</td>
<td>0.0-2.0</td>
<td>2.0-8.0</td>
<td>2.0-10.0</td>
<td>5.0-15.0</td>
</tr>
</tbody>
</table>

The ingredients were selected because they are common in dietary products and, at the same time, they have properties that are representative of many components in pharmaceutical formulations, especially in terms of particle size distribution, shape, and density differences between components that could enhance the segregation phenomenon. In addition, the components to be monitored were also selected by taking into account their physical and chemical properties detected by at least one of the tools in place. Therefore, NIR responses, color features (RGB imaging) and fluorescence properties (LIF), along with a different range of concentrations (to evaluate the capacity of these tools), were criteria for their selection. The same blend and ingredients were analyzed using these apparatuses in CHAPTER 4 with an in-line configuration (i.e., powder flowing through the feed frame of a tabletting press). Riboflavin (yellow and fluorescence properties) was modeled by LIF despite its low concentration; ferrous fumarate (reddish brown color, NIR-responsive) by NIR and RGB; ginseng (NIR-responsive and fluorescent) by NIR and RGB, and ascorbic acid (NIR-responsive) by NIR. These models were based on variation in concentration of a single component, one at a time, without considering interactions between ingredients, the aim of this work.
5.2.3 Design of experiments

Previous research attempted to develop models for each ingredient in a design where component concentration varied individually. The present work intends: a) to evaluate goodness of fit ($R^2$) of the models built while simultaneously varying the concentrations of other components (therefore, giving robustness to the models), and b) to identify interactions that could occur between components, i.e., characterize ingredient(s) whose presence might influence accurate ingredient prediction intended to be monitored as well as the type of error encompassed (under- or over-estimation) by possible interference.

Two separate 2-level, full factorial designs, each applied to all 4 ingredients, were tested. They were aimed at: a) building robust models, b) identifying and detecting interactions that might occur in a broad range of concentrations, and c) potentiating the probability of interactions between components. In the first full factorial ($2^4=16$ batches) composed of levels 0 (ingredients in the reference concentration) and 1, 3 center points (level 0.5) were included (batches 17-19). The second full factorial was constituted of levels 1 and 2, and no center points (batches 16 and 20-34). Figure 5.1 reports overage concentrations relative to label claims for each level of ingredient under study.

![Figure 5.1: % of concentration for each ingredient by level (levels “0”, “0.5”, “1” and “2”, represented by blue, cyan, yellow and red, respectively)]
5.2.4 Batch preparation and data acquisition

All batches were prepared from the same reference mixture taken from the manufacturing line (therefore prepared by a validated method). The reference mixture was then divided into 34 batches (400 g each). The 4 ingredients were weighed and added to each batch to obtain the desired concentration, and the mixture was processed in a 4-L in-house-built blender using proper scaled-down methodology to define blending time (100 seconds). A representative sample of each batch was analyzed off-line by the 3 tools in place. 16 acquisitions per tool were achieved for every batch.

5.2.5 Data treatment

The 16 acquisitions obtained from each of the 34 batches (544 acquisitions per tool, in total) were divided equally into calibration and test sets. The NIR data were pretreated with Standard Normal Variate and scaled to UnV. RGB data were analyzed by MIA [34], then scaled to UnV. LIF data were treated with first-order Savitzky-Golay filter, then scaled to UnV. Therefore, a total of 12 PLS models were developed (4 components and 3 tools). Individual models, which gave better $R^2$ values, were then characterized by residual error related to each parameter to identify which ingredients presented significant interference in the model developed (by ANOVA). After their selection, t-testing evaluated which ingredients possibly interfered in the monitoring of each ingredient by each tool.

All data treatment, model building and statistical analysis were conducted with PLS Toolbox and in-house software using Matlab.

5.3 Results and discussion

5.3.1 Models building and selection

As mentioned in Section 5.2.3, each ingredient studied had 4 different concentrations (levels) distributed in all batches. Concentration differences were known to be large enough to be detected by the tools. Therefore, when a PLS model is developed, and assuming that no interaction is present, it should be able to distinguish 4 distinct sample types, each clearly clustered in PLS score plots. On the other hand, if strong interaction could bias the data, the 4 clusters would be blurred. Although score plot representation is purely qualitative, it
represents a preliminary conclusion of possible interference and must be observed for all 12 PLS models developed.

Considering, as an example, the score plot obtained for ginseng when NIR data were tested (Figure 5.2). In CHAPTER 4, this tool was shown to be able to monitor ginseng variations, keeping the ratios of all other components unchanged. In the present case, however, where such constraints were removed, it was observed that, as concentrations increased, the samples tended to shift to the right. This fact is much clearer at the highest level (yellow ◊) where samples are mainly located in the right top quadrant of the graph. However, not all samples at this level are located in that region as some appear in the center of the plot, along with others of lower concentration. It indicates that some interference may be occurring between ginseng and another compound when NIR is modeled. It might, in turn, lead to biased estimation of ginseng concentration.

Analysis of score plots for all other models generally gave similar results (not presented). However, due to the qualitative nature of this analysis, few conclusions can be drawn from these results. For this reason, adjusted R$^2$ was computed for each model as key indicator of model quality. Table 5.3 indicates the calibration and test values of adjusted R$^2$ for the models built as well as the number of PC.

![Figure 5.2: Score plot of the PLS model of ginseng developed from NIR data (8 PC): sample concentration increased in the sequence of circles (red o), squares (light blue □), stars (dark blue ●), and diamonds (yellow ◊)](image-url)

Table 5.3 indicates the calibration and test values of adjusted R$^2$ for the models built as well as the number of PC.
Table 5.3: Adjusted $R^2$ values of models built (Cal.: calibration; PC: principal components)

<table>
<thead>
<tr>
<th></th>
<th>NIR</th>
<th></th>
<th></th>
<th>RGB</th>
<th></th>
<th></th>
<th>LIF</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cal.</td>
<td>Test</td>
<td>PC</td>
<td>Cal.</td>
<td>Test</td>
<td>PC</td>
<td>Cal.</td>
<td>Test</td>
<td>PC</td>
</tr>
<tr>
<td>Riboflavin</td>
<td>0.62</td>
<td>0.57</td>
<td>8</td>
<td>0.95</td>
<td>0.78</td>
<td>8</td>
<td>0.72</td>
<td>0.55</td>
<td>8</td>
</tr>
<tr>
<td>Ferrous fumarate</td>
<td>0.85</td>
<td>0.83</td>
<td>8</td>
<td>0.95</td>
<td>0.70</td>
<td>8</td>
<td>0.73</td>
<td>0.64</td>
<td>8</td>
</tr>
<tr>
<td>Ginseng</td>
<td>0.85</td>
<td>0.81</td>
<td>8</td>
<td>0.93</td>
<td>0.70</td>
<td>8</td>
<td>0.90</td>
<td>0.85</td>
<td>7</td>
</tr>
<tr>
<td>Ascorbic acid</td>
<td>0.78</td>
<td>0.71</td>
<td>8</td>
<td>0.85</td>
<td>0.39</td>
<td>8</td>
<td>0.70</td>
<td>0.60</td>
<td>7</td>
</tr>
</tbody>
</table>

From Table 5.3, it can be inferred which tools are better suited to monitor each ingredient. Riboflavin, which is yellow, present in low concentration (<2.0%) and fluorescent, was a good model when monitored by RGB ($R^2_{calibration} = 0.95$) and had lower values when monitored by LIF ($R^2_{calibration} = 0.72$). NIR was not able to accurately predict this ingredient, probably because of its low concentration in the mixture. As for ferrous fumarate, its brownish-red color and high concentration (2.0-10.0%) made it suitable for tracking by NIR ($R^2_{calibration} = 0.85$) and RGB ($R^2_{calibration} = 0.95$). As with LIF, although this ingredient does not have fluorescence properties, it can be predicted by its color outcome due to the visible spectroscopy effect, which is not recommended. The light brown color, intermediate concentration (2.0-10.0%), fluorescence properties and NIR-responsiveness of ginseng made it trackable by all tools. NIR was most accurate ($R^2_{calibration} = 0.78$) in the case of ascorbic acid, a white powder present in high concentrations (5.0-15.0%) among the ingredients studied, although RGB calibration gave a higher value ($R^2_{calibration} = 0.85$). However, RGB led to lower values ($R^2_{test} = 0.39$), preventing it from being considered as accurate (the white color of ascorbic acid makes it impossible for RGB to monitor since it is diluted in the blend). As with LIF, the absence of fluorescence in this component culminated in lower results obtained.

**Practical utility of the tools**

Analysis of the results in Table 5.3 reveals no surprises regarding the tools which are able to monitor each of the ingredients as they are in accordance with the conclusions obtained from CHAPTER 4. Therefore, assuming that $R^2 \geq 0.70$ represents acceptable fit for the model, taking complexity of the blend and the quantification tolerance of these components into account, riboflavin can be monitored by RGB and LIF, ferrous fumarate by NIR and RGB, ginseng by all tools, and ascorbic acid by NIR. Modeling of riboflavin by NIR (low
of ferrous fumarate by LIF (known interference of the tool with color components that give biased results) and of ascorbic acid by RGB (low R^2_{test}) and LIF (low R^2_{calibration}) proved to be ineffective.

It can be observed that R^2 values are generally lower (Figure 5.3) in comparison to the results obtained in CHAPTER 4. This is attributed to design, since relative concentration of the ingredients remains constant, but varies in full factorial design. Although such is the case with NIR and LIF, RGB leads to better prediction in complex designs. These results must be interpreted very carefully since RGB is more sensitive to changes in acquisition conditions (process dynamics, vibrations), i.e., prediction is significantly different when the model is developed in-line or off-line. This fact may have improved RGB models for ginseng monitoring beyond the results obtained in CHAPTER 4. Finally, the higher R^2_{calibration} values obtained for riboflavin, when monitored by NIR in the present study compared to a simpler design, has no significance since the tool did not give adequate values in both cases.

Interpretation of these R^2 values must consider the constraints and the goal for which these tools are installed. A general overview indicates that all tools are capable of monitoring at least 1 ingredient (lowest value of 0.72 in riboflavin monitoring by LIF). Although this value appears low, it is important to consider that the model is developed for low-concentration ingredients (<2%) present in a 31-component multivitamin blend after complex calibration. A previous study with a simpler design led to a R^2_{calibration} of 0.95 (CHAPTER 4). Moreover, these types of formulations, with a large number of components, are characteristic of dietary products (like multivitamins) and are rarely found in pharmaceutical formulations whose number of components is significantly lower. Along with the API, which is usually the component to be monitored, a typical tablet has few other excipients in its constitution, such as diluents, binders, desintegrants and lubricants, rarely higher than 10 in sum. Acceptance criteria are broader in dietary products than in pharmaceuticals (constituted by an API), mainly due to the degradation these kinds of products undergo throughout their shelf life [135]. Therefore, the margin of error of the models developed is likely within accepted criteria of these ingredients (criteria that mainly depend on the manufacturer). Taking all the mentioned constraints into account, the acceptability of these values also indicates the potential of these tools to more accurately monitor other components with the same characteristics in less complex blends.
5.3.2 Analysis of the residual error in interactions investigation

In the development of every model, associated errors can derive from the method itself (e.g., under-/over-parameterization) or from the information present in data and not relevant to model building. In the latter case, it is possible to identify and quantify the relative relevance of parameters (the components alone and their interactions) to this error in the model. Their evaluation is critical to establish the importance that they might represent in the tool used for monitoring and the interactions that might occur, which could produce unreliable measurements. This analysis was conducted for all models selected as adequate in the previous section. The relative importance of each parameter (contribution to overall error, expressed in %) was ascertained by ANOVA (95% confidence interval: 95% CI) to determine which were significantly affecting model prediction:

$$\% \text{ parameter}_i = \frac{SS_i}{SS_{error}} \times 100\% \quad \text{Equation 5.1}$$

where $SS_i$ is the sum of squares relative to the parameter under study; $SS_{error}$ is the sum of squares relative to total error in the group (parameters).
To facilitate interpretation of the results, and since the significance of this analysis lay in study of the interference that the parameters alone might cause (A, B, C and D), interactions between them (AB, AC, AD, BC, BD, CD, ABC, ABD, ACD, BCD, ABCD) were not considered. The importance of each parameter on error in the models, described in Section 5.3.1 as well, is enumerated in Table 5.4 where parameters that showed statistical significance (95% CI) are highlighted.

Table 5.5 includes the outcomes to facilitate analysis.

Table 5.4: Importance (%) and significance of each parameter in the selected models (Cont.: Contribution)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Riboflavin</th>
<th>Ascorbic acid</th>
<th>Ginseng</th>
<th>Ferrous fumarate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cont.</strong></td>
<td>p-value</td>
<td>Cont. p-value</td>
<td>Cont. p-value</td>
<td>Cont. p-value</td>
</tr>
<tr>
<td><strong>NIR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ascorbic acid</td>
<td>0.4</td>
<td>0.873</td>
<td>16.8</td>
<td>0.005</td>
</tr>
<tr>
<td>Ginseng</td>
<td>1.0</td>
<td>0.741</td>
<td>2.2</td>
<td>0.518</td>
</tr>
<tr>
<td>Ferrous fumarate</td>
<td>3.6</td>
<td>0.318</td>
<td>1.2</td>
<td>0.673</td>
</tr>
<tr>
<td><strong>LIF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Riboflavin</td>
<td><strong>26.4</strong></td>
<td><strong>0.000</strong></td>
<td><strong>8.2</strong></td>
<td><strong>0.040</strong></td>
</tr>
<tr>
<td>Ginseng</td>
<td>1.5</td>
<td>0.616</td>
<td>0.5</td>
<td>0.844</td>
</tr>
<tr>
<td><strong>RGB</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Riboflavin</td>
<td>7.1</td>
<td>0.131</td>
<td>1.7</td>
<td>0.604</td>
</tr>
<tr>
<td>Ginseng</td>
<td>0.5</td>
<td>0.865</td>
<td>1.9</td>
<td>0.583</td>
</tr>
<tr>
<td>Ferrous fumarate</td>
<td>0.9</td>
<td>0.739</td>
<td>5.9</td>
<td>0.152</td>
</tr>
</tbody>
</table>

Table 5.5: Possibilities of interference from Table 4 (bold underlined values are significant at α = 0.05)

<table>
<thead>
<tr>
<th>Tools</th>
<th>Ingredients</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIR</td>
<td>Ascorbic acid</td>
<td>Most significant parameters</td>
</tr>
<tr>
<td></td>
<td>Ginseng</td>
<td>Ascorbic acid</td>
</tr>
<tr>
<td></td>
<td>Ferrous fumarate</td>
<td>Ferrous fumarate</td>
</tr>
<tr>
<td>LIF</td>
<td>Riboflavin</td>
<td>Riboflavin</td>
</tr>
<tr>
<td></td>
<td>Ginseng</td>
<td>Ginseng</td>
</tr>
<tr>
<td>RGB</td>
<td>Riboflavin</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Ginseng</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Ferrous fumarate</td>
<td>Ferrous fumarate</td>
</tr>
</tbody>
</table>
As might be expected, the results first reveal that the most significant parameter affecting the prediction error of a component is the component concentration itself. In general terms, this can be explained by the fact that we are working with multiple concentrations, and extreme concentrations are slightly biased (e.g., high concentrations tend to be underestimated).

The results show that secondary components sometimes also significantly affect predictions (riboflavin and ginseng monitoring by LIF). However, in other cases, they do not, even if they are responsive to the monitoring tool. Taking the model developed for ascorbic acid with NIR as an example, it is evident that, although other components (ginseng and ferrous fumarate) contribute to the error (4.6 and 3.6, respectively), probably because they are both NIR-responsive, the multivariate approach was able to build models without significant interference.

Nevertheless, in ginseng monitoring, none of the models developed (for any of the tools) showed significant parameters, although with LIF, ginseng was nearly significant (p-value = 0.057), unlike the other tools (p-values of 0.278 and 0.445 with NIR and RGB, respectively). These observations disclose that LIF is much more selective for ginseng than the other tools, despite goodness of the model obtained with NIR ($R^2_{calibration} = 0.81$). It is possible that the model developed with this tool might be the result of correlation of several parameters (NIR is responsive to all ingredients under study).

The absence of parameter significance was also witnessed with component modeling by RGB, except in the case of ferrous fumarate. Such findings lead to the conclusion that, in RGB modeling, the error observed is largely method-related (many color components could correlate with this ingredient), which could eventually limit utility of the tool.

It is also apparent from this work that parameters are only significant when their contribution is at least around 8% of error. Although this percentage cannot be generalized and mostly depends on study type (especially the DoE approach) and concentrations involved, it can be assumed that lower error percentages will not be significant. An obvious case is the model developed for riboflavin with LIF. In that case, 2 significant parameters can be observed: riboflavin with a contribution of 26.4%, and ascorbic acid with a contribution of 8.2%. It clearly indicates that the presence of ascorbic acid might interfere with riboflavin prediction by this tool. Although there is no clear-cut explanation of this correlation, since
ascorbic acid does not have fluorescent properties and, thus, does not have the potential to directly affect the LIF signal, many other mechanisms can occur in such complex blends (e.g., increasing ascorbic acid concentrations lead to dilution of other fluorescent components). In contrast, certain components, expected to contribute significantly to modeling, do not seem to cause any interference, e.g., the presence of ferrous fumarate in ginseng monitoring by LIF, because of its color properties. However, in this specific case, ferrous fumarate contributes to error of 6.1%, which corresponds to the highest contribution of the non-interfering components of all models. Although not significant, the result indicates that some interference might actually occur, especially since the monitored ingredient is not significant as well.

It is concluded that the latter 2 interactions – one mathematically significant, and the other making a high contribution – were most significant in the models developed and defined for testing. All other components proved to be non-important in the models developed.

5.3.3 Determination of the interactions

Analysis of residual errors makes it possible to determine which ingredients have the potential to interact with each other, but does not explain the concentration range in which interference plays a role, or the effect of that interference. These 2 factors are discussed in this section.

The previous section noted that riboflavin modeling with LIF might be affected by the presence of ascorbic acid. Also, the possibility was considered that ferrous fumarate might have an impact on ginseng prediction when the same tool was tested. The effect of these interactions is depicted in Figure 5.4 (a) to (d), through calibration and validation curves of these models, using a color code system to identify concentrations of the interfering parameters (ascorbic acid and ferrous fumarate, respectively) present in each concentration of the studied ingredients (riboflavin and ginseng). Trial design in this work did not provide information on component interaction at lower concentration (level 0) with another component present in higher concentration (level 2) and vice versa. However, all the studied ingredients had 2 levels (0 and 2) at which interference was determined. At level 0 of the studied component, the interfering parameters were at levels 0 and 1, whereas at level 2 the interfering parameters were at levels 1 and 2. Therefore, study of interference was possible by analyzing the results of modeling curves. Taking as an example the data presented in Figure 5.4 (a), it is
apparent that, for riboflavin level 0 (concentration = 128%), there was no significant
difference in terms of distance to the sample curves of both ascorbic acid concentrations.

Figure 5.4: Calibration (a) and validation (b) curve of riboflavin (LIF) - color code system
representing the concentration of ascorbic acid; calibration (c) and validation (d) curve of ginseng
(LIF) - color code system representing the concentration of ferrous fumarate

Nevertheless, it can be seen from the validation curve (Figure 5.4 (b)) that samples with
high ascorbic acid concentrations (level 1, light blue samples) above the calibration line
outnumber samples at low concentrations (level 0, dark blue samples). Besides, on the other
extremity of the curve (riboflavin level 2, concentration = 220%), both in calibration and
validation, it is observed that higher concentration samples are located above lower
concentration samples. In validation curves, both samples are even located below the curve,
which might be an indication that their presence leads to riboflavin underestimation. As far as LIF monitoring of ginseng is concerned, it can be seen that, in both the calibration and validation curves, separation between samples with different concentrations of the interfering component is more obvious at lower ginseng level (level 0, 135%), than when the concentration of this component is higher (level 2, 300%). At this concentration, ferrous fumarate might have an effect on ginseng overestimation if ferrous fumarate is present. At higher concentrations (level 2), this influence does not seem so strong.

_Are they true interfering agents?_

To evaluate if the difference in position is significant between each group of samples with the same concentration of ingredient under study but varying concentrations of the interfering agent, t-testing was performed with unknown variance and 95% CI. Statistical differences between the averages of each pair indicate that the concentration of each interfering parameter plays a role in prediction of the ingredient to be monitored (riboflavin and ginseng, respectively). These tests were performed for both calibration and validation curves. As seen in Table 5.6 to Table 5.9, the only pair of statistically different samples (|t-value| > t-critical) is the interaction of ascorbic acid at the higher riboflavin concentration (level 2) in the calibration curve, and the interaction of ferrous fumarate with ginseng at the lower concentration of this ingredient (level 0). This proves that ascorbic acid concentration does, in fact, play a role in riboflavin prediction, although, as studied in the previous section, this parameter was not significant to the model.

<table>
<thead>
<tr>
<th>Table 5.6: t-test of the riboflavin model with LIF data (calibration)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Component predicted</strong></td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Riboflavin (concentration)</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>
In the case of ascorbic acid interaction in riboflavin monitoring, its influence is most likely due to the dilution effect that higher quantities of ascorbic acid have on total fluorescence emitted by fluorescent components of the blend. Although this interaction seems
to occur only with higher quantities of riboflavin and ascorbic acid, it remains inconclusive since the calibration curve shows a significant effect of possible overestimation, whereas validation, despite the fact that it is not significant, is located below the calibration curve. This interference of ascorbic acid was not expected at the beginning of the trial and reinforces the need for robust model design in the late stages of implementation.

As far as ginseng monitoring is concerned, the presence of ferrous fumarate seems to overestimate it at a low concentration. This effect vanishes when ginseng concentration increases, regardless of simultaneous increment of the interfering ingredient. It was expected because of the spectroscopic influence this component has on fluorescence. Being significant only in low doses, it is of utmost importance for monitoring ingredients with higher concentrations but might represent a setback when the goal is to monitor low-dose components in the blend.

Therefore, it can be said that these components (ascorbic acid and ferrous fumarate) are likely to behave as interfering agents in prediction of the studied ingredients (riboflavin and ginseng, respectively) with LIF monitoring.

5.4 Conclusion

This work centers on the use of PAT tools to monitor multicomponent pharmaceutical powder blends. Specifically, it seeks to develop a method for studying model robustness (i.e., the predicted concentrations of a given component) when levels of the other components are free to vary. To illustrate the method, 4 ingredients (riboflavin, ferrous fumarate, ginseng and ascorbic acid), typically present in multivitamin blends, were monitored by 3 probes (NIR, LIF and RGB imaging).

At least 1 tool was capable of accurately predicting the concentrations of all ingredients, showing that, even in complex mixtures usually found in multivitamins, such tools may represent an option to monitor some components ($R^2_{min} = 0.72$). These models were developed with simultaneous variation of other components and, for that reason, presented higher prediction error than previous work ($R^2_{min} = 0.95$), which enhanced the importance of good calibration design. In general, the models developed were robust against interference by the other ingredients being studied, demonstrating good specificity, which is a good indicator of their utility as a PAT tool. Nevertheless, the prediction of some components proved to be
affected by the concentrations of other components. This was the case with riboflavin prediction by LIF, which appeared to be impacted by ascorbic acid, probably due to high concentration of the vitamin in the mixture (dilution factor). The presence of color components could also change prediction when monitored by LIF. Therefore, when implementing these monitoring tools, it is important to investigate possible interactions that could occur, especially when more than one color component is present (especially important in the case of RGB and LIF) and/or if the mixture has more than 1 fluorescent component in its constitution. As in this work, a concentration range must be studied, since varying levels of interfering components differentially affect prediction of the ingredient to be monitored.

The present work demonstrates the importance of robust design, and even in complex blends, there is potential to use these tools as monitoring instruments, especially when a larger acceptance interval criterion is tolerated, as is the case with commercial multivitamin formulations. This outcome may potentially apply to these tools as a complement or even as a replacement of traditional quantification techniques in such formulations and to show that their specificity is reasonable, even in complex mixtures.

Acknowledgments

The authors thank Pfizer and the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding. The technical contribution of Jacques Gagné, Henri Gauvin and Marc Couture (Université de Sherbrooke) is gratefully acknowledged. Special thanks to Ovid da Silva for editing this manuscript.
CHAPTER 6 CONCLUSION

Due to the changes in the regulatory environment and the benefits provided by the implementation of PAT, the use of this technology is becoming more common in the development of new products and monitoring of known processes. In fact, its use as a monitoring technique to control the manufacturing processes of existing products is being reported in literature more often in recent years. Therefore, the most common steps of manufacturing processes (blending, granulation, drying, coating) have been the subject of most these studies. On the other hand, although tablets are the most common pharmaceutical dosage form, monitoring of the compression stage has not been the subject of many studies, especially with non-spectroscopical tools.

This work demonstrates the possibility to monitor in-line the compression stage using three different tools (NIR, RGB camera and LIF) under the defined conditions. After an initial study with a simple blend to assess the potential of these tools and to evaluate the proposed set-up, these tools were used to monitor the constitution of a multicomponent powder flowing in the feed frame. Results demonstrated their ability to monitor the selected components in the proposed blend under the conditions defined, including a low dosage component in which LIF demonstrated a significant accuracy. It was also investigated the advantages of combining data from the tools in place. Results show that data combination has a beneficial effect in the component prediction as it increases the model accuracy. This finding might lead manufacturers to apply more than one tool to monitor their processes in cases where the accuracy is not guaranteed by only one tool or in cases where more than one CQA / CPP (quantification of another component, for example) is intended to be monitored.

A further study performed off-line, reinforced the applicability of these tools to be used as monitoring tools, as robust models were developed from two full factorial designs based on variation of concentration of four components. Although the accuracy of these models might be adequate for these type of products (multivitamins, classified as dietary products) their prediction values would not be adequate for pharmaceutical products, especially those with a narrow therapeutic margin whose reliability is very much needed. On the other hand, the formulation tested represented an extreme case due to their large number of components, which is hardly found in pharmaceutics. Therefore, the potential and applicability of these
tools to monitor a process, especially the compression stage of pharmaceutics is clearly demonstrated in this project. However, it must be stated that the specificity of the blends (powder properties) and the operating parameters (paddle and die speed) affect the model developed and must be tested prior to any implementation.

The knowledge obtained in this project opens the possibility to use these tools to monitor the tablet press as long as the features of the tools are appropriate to the components to track. The effective monitoring of components in the tablet press also provides resources to study and detect undesired phenomena such as segregation, which is especially important in formulations whose properties enhance this possibility. In this sense, further studies should direct their focus in tracking components present in formulations known to be segregative in order to study this effect in the feed frame and to optimize the operating conditions (number and shape of paddles, feed frame speed) that can reduce this effect.

In terms of component quantification to complement and eventually to replace the traditional UDU test, it is critical to develop a robust model. Further studies should focus on developing and testing robust models (with an adequate DoE) while changing the operating conditions (feed frame speed, for instance). The variability of the operating conditions must be evaluated in the validation of the model, to guarantee that a reliable measurement is obtained and to prove to the regulatory agencies that whole sets of variability have been included in the development. Screening of possible interactions should be studied in order to guarantee the specificity. Moreover, the possibility of combining the tools must be assessed and other tools that have been successfully used in other applications (Raman spectroscopy, chemical imaging) can also provide a significant output of real time press monitoring.
Conclusion en français

En raison de l'évolution de l'environnement réglementaire et des avantages livrés par la mise en œuvre de la PAT, l'utilisation de cette technologie devient de plus en plus courante dans le développement de nouveaux produits et dans le suivi de procédés existants. En fait, son utilisation comme technique pour contrôler les procédés de fabrication des produits existants a davantage été rapportée dans la littérature au cours des dernières années. Par conséquent, les étapes plus courantes des procédés de fabrication (mélange, granulation, séchage, revêtement) ont fait l'objet de la majorité de ces études. D'autre part, bien que les comprimés soient de la forme galénique la plus courante, la surveillance de l'étape de compression n'a pas fait l'objet de nombreuses études, notamment avec des outils non spectroscopiques.

Ce travail montre la possibilité de surveiller en ligne l'étape de compression en utilisant trois outils différents (NIR, caméra RGB et LIF) dans des conditions définies. Après une étude initiale avec un mélange simple pour évaluer le potentiel de ces outils et évaluer le montage proposé, ces mêmes outils ont été utilisés pour surveiller la constitution d'une poudre multicomposante s'écoulant dans le cadre d'alimentation. Les résultats ont démontré la capacité des outils à surveiller les composants choisis dans le mélange proposé, dans les conditions définies, y compris un composant à faible dosage dans lequel la LIF a démontré une précision significative. On a également étudié les avantages de combiner les données des outils en place. Les résultats montrent que la combinaison de données a un effet bénéfique dans la prédiction des composants, car elle augmente la précision du modèle. Cette constatation pourrait amener les fabricants à appliquer plus d'un outil pour surveiller leurs processus dans les cas où l'exactitude n'est pas garantie par un seul outil ou dans le cas où plus d'un CQA / CPP (quantification d'un autre composant, par exemple) doit être surveillé.

Une autre étude réalisée hors ligne a renforcé l'applicabilité de ces outils pour des fins de contrôle, car des modèles robustes ont été développés à partir de deux modèles factoriels complets basés sur la variation de concentration de quatre composantes. Bien que la précision de ces modèles puisse être adéquate pour ces types de produits (multivitamines, classés comme produits diététiques), leurs prédictions ne seraient pas adéquates pour les produits pharmaceutiques, en particulier ceux dont la marge thérapeutique est faible et dont la fiabilité est très nécessaire. D'autre part, la formulation testée représentait un cas extrême en raison de
leur grand nombre de composants, ce qui est peu probable dans une formulation pharmaceutique typique. Par conséquent, le potentiel et l'applicabilité de ces outils pour surveiller un processus, en particulier l'étape de compression des produits pharmaceutiques, est clairement démontré dans ce projet. Cependant, il faut préciser que la spécificité des mélanges (propriétés de la poudre) et les paramètres de fonctionnement (vitesse de la pale et de la matrice) affectent le modèle développé et doivent être testés avant toute mise en œuvre.

Les connaissances acquises dans ce projet ouvrent la possibilité d'utiliser ces outils pour contrôler la presse à comprimés aussi longtemps que les caractéristiques des outils sont appropriées aux composants à suivre. La surveillance efficace des composants dans la presse à comprimés fournit également des ressources pour détecter des phénomènes indésirables tels que la ségrégation, ce qui est particulièrement important pour les formulations dont les propriétés améliorent cette possibilité. Dans ce sens, d'autres études devraient être orientées vers le suivi des composants présents dans les formulations connues pour être ségrégatives, afin d'étudier cet effet dans le cadre d'alimentation et d'optimiser les conditions de fonctionnement (nombre et forme des palettes, vitesse du cadre d'alimentation) qui peuvent réduire cet effet.

En termes de quantification de composants pour compléter et éventuellement remplacer le test UDU traditionnel, il est essentiel de développer un modèle robuste. D'autres études devraient se concentrer sur le développement et l'essai de modèles robustes (avec un DoE adéquat) tout en changeant les conditions de fonctionnement (vitesse d'avance, par exemple). La variabilité des conditions de fonctionnement doit être évaluée lors de la validation du modèle afin de garantir une mesure fiable et de prouver aux organismes de régulation que des ensembles complets de variabilité ont été inclus dans le développement. Le criblage des interactions possibles doit être étudié afin de garantir la spécificité. En outre, la possibilité de combiner les outils doit être évaluée et d'autres outils qui ont été utilisés avec succès dans d'autres applications (spectroscopie Raman, imagerie chimique) peuvent également fournir une prédiction significative de la formulation d’un comprimé à la sortie de la presse.
ANNEXES

A.1 CALCULATION OF THE MASS ANALYZED

- NIR

The volume of the material analyzed depends on the integration time set for the probe (and the number of samples averaged), distance of the probe to sample, feed frame speed and penetration depth (which depends on cohesion properties of the powder).

- Integration time
  The integration time set for this probe was 50 ms (1 000 µs × 50 samples).

- Distance of the probe to sample
  Probe size is 1.5 × 0.4 cm (red arrows, Figure A.1)

![NIR probe dimensions](image)

*Figure A.1: NIR probe dimensions*

Previous tests demonstrated that at a distance of 11 mm to the sample (as the setup in chapter 3), the probe detects a range up to 2.5 cm, which means that at this distance, the probes has a spot size of 2.5 cm × 1.4 cm (0.5 cm more for each side).
> *Feed frame speed*

The speed of the feed frame was set to **20 rpm** in the trial described in chapter 4.

Therefore, taking into account the radius (r) of the feed frame (*Figure A.2*):

![Feed frame radius](image)

*Figure A.2 : Feed frame radius*

\[
r = 11.75 \text{ cm}
\]

\[
P = 2 \times \pi \times r = 73.8 \text{ cm}
\]

Since the feed frame is constituted by 10 paddles, the outer circular sector (l) between two paddles is 7.38 cm.

However, the probe is not located in the outer edge and the powder moves differently according to its radial position in the feed frame. Therefore, it is necessary to calculate the sectional distance at each of the probes’ edge. Taken into account the position of the probe (between radius 10.75 and 8.25 cm):

\[
P_{\text{outer edge}} = 2 \times \pi \times 10.75 = 67.5 \text{ cm} \rightarrow 6.75 \text{ cm (sectional distance between paddles)}
\]

\[
P_{\text{inner edge}} = 2 \times \pi \times 8.25 = 51.8 \text{ cm} \rightarrow 5.18 \text{ cm (sectional distance between paddles)}
\]
With a speed of 20 rpm, it takes 0.3 seconds = 300 ms for a paddle to course the space between paddles. With an integration time of 50 ms for NIR, the distance covered in the circular section of each edge is:

\[
D_{\text{outer edge}} = \frac{50 \text{ ms}}{300 \text{ ms}} \times 6.75 \text{ cm} = 1.13 \text{ cm}
\]

\[
D_{\text{inner edge}} = \frac{50 \text{ ms}}{300 \text{ ms}} \times 5.18 \text{ cm} = 0.86 \text{ cm}
\]

Figure A.3: Scheme of feed frame wheel with NIR probe and its acquisitions sizes (in cm)
Therefore, the total area of analysis ($T_{\text{AREA}}$) of the probe, represented in Figure A.4 is:

$$T_{\text{AREA}} = A + B$$

Since: $A = 2.5 \text{ cm} \times 1.4 \text{ cm} = 3.50 \text{ cm}^2$

$$\text{Cross sectional area} = \frac{r \times l}{2}$$

$$B = D - C = \frac{10.75 \text{ cm} \times 1.13 \text{ cm}}{2} - \frac{8.25 \text{ cm} \times 0.86 \text{ cm}}{2} = 2.52 \text{ cm}^2$$

$$T_{\text{Area}} = 3.50 + 2.52 = 6.02 \text{ cm}^2$$


- **Depth**
  The depth of the equipment for free flowing powders (as it is the case), is assumed to be around \(1 \text{ mm} = 0.1 \text{ cm}\).

- **Mass calculation**
  An observation taken by NIR in this experiment was able to quantify a volume of:
  \[
  T_{\text{volume}} = 6.02 \text{ cm}^2 \times 0.1 \text{ cm} = 0.602 \text{ cm}^3
  \]
  Taking into account the bulk density of the powder for this experiment (0.81 g/cm\(^3\)), the mass analyzed for each observation was equal to:
  \[
  m_{\text{analyzed}} = 0.810 \times 0.602 = 0.488 \text{ g}
  \]

- **LIF**
  In the case of LIF, since it is a contact probe, the acquisition size is the same as the probe size (0.4 cm diameter). Therefore, the volume analyzed by LIF only depends on the probe size, integration time, feed frame speed and penetration depth.

- **Integration time**
  The integration time set for this probe was 80 µs (1 sample).

- **Feed frame speed**
  As mentioned, the feed frame speed was set to 20 rpm.

  Taken into account the position of the probe (between radius 10.5 and 10.1 cm):
  \[
  P_{\text{outer edge}} = 2 \times \pi \times 10.5 = 67.5 \text{ cm} \rightarrow 6.60 \text{ cm (sectional distance between paddles)}
  \]
  \[
  P_{\text{inner edge}} = 2 \times \pi \times 10.1 = 51.8 \text{ cm} \rightarrow 6.35 \text{ cm (sectional distance between paddles)}
  \]
  With an integration time of 80 µs = 0.08 ms for NIR, the distance covered in the circular section of each edge is:
  \[
  D_{\text{outer edge}} = \frac{0.08 \text{ ms}}{300 \text{ ms}} \times 6.60 \text{ cm} = 0.002 \text{ cm}
  \]
\[ D_{\text{inner\ edge}} = \frac{0.08 \, ms}{300 \, ms} \times 6.35 \, cm = 0.002 \, cm \]

Therefore, it can be considered that the acquired area for the LIF corresponds to the size of its probe:

\[ A_{\text{LIF}} = \pi \times 0.2^2 = 0.13 \, \text{cm}^2 \]

➢ Depth

The depth of the equipment for free flowing powders (as it is the case), is assumed to be around 1 mm = 0.1 cm.

➢ Mass calculation

An observation taken by LIF in this experiment was able to quantify a volume of:

\[ V_{\text{volume}} = 0.130 \, cm^2 \times 0.010 \, cm = 0.0013 \, cm^3 \]

Taking into account the density of the powder for this experiment (0.81 g/cm³), the mass analyzed for each observation was equal to:

\[ m_{\text{analyzed}} = 0.810 \times 0.0013 = 0.0011 \, g \]
A.2 DATA MINING - NIPALS

Algorithm description:

1. Choose a start $t_i$.
2. Calculate $p$ ($p = X^T.t / (t^T.t)$);
3. Normalize $p$ ($p \rightarrow \|p\| = 1$)
4. Calculate $t$ ($t = X.p / (p^T.p)$);
5. Calculate the residual matrix ($E = X - t.p^T$)

*Figure A.5: Algorithm of NIPALS*
REFERENCES


REFERENCES


[36] G. Szatvanyi, C. Duchesne, G. Bartolacci, Multivariate Image Analysis of Flames for Product


[41] EMA, Guideline on the use of Near Infrared Spectroscopy (NIRS) by the pharmaceutical industry and the data requirements for new submissions and variations, 2014.


REFERENCES


References


