

UNIVERSITE DE SHERBROOKE

Faculté de génie

Département de génie chimique et de génie biotechnologique

ÉTUDE MULTIVARIEE DU COLLAGE DE  
POUDRES PHARMACEUTIQUES A ECHELLE  
INDUSTRIELLE

Thèse de doctorat

Spécialité : génie chimique

Mahamadou Dembélé

Sherbrooke (Québec) Canada

Mars 2023



# MEMBRES DU JURY

Ryan Gosselin

Directeur de thèse

Nicolas Abatzoglou

Codirecteur de thèse

Sophie Hudon

Évaluateur

Bouchra Nasri

Évaluateur

Federico Galli

Rapporteur



# RESUME

Le collage est l'un des problèmes les plus courants et les plus préjudiciables survenant lors de la fabrication des comprimés. C'est un phénomène qui consiste en l'adhésion de poudres pharmaceutiques sur la surface des outillages de compression. À cause du nombre important de paramètres impliqués dans son occurrence et leurs multiples interactions, l'étude et la compréhension du collage nécessitent une approche intégrative et simultanée où les paramètres interdépendants doivent être étudiés ensemble. La conduite de telles expérimentations simultanées à échelle laboratoire pourrait s'avérer très coûteuse. Cependant, une quantité importante de données est générée lors de la production industrielle des comprimés.

La présente thèse introduit, évalue et discute une nouvelle approche d'étude du collage basée sur l'analyse des données industrielles. Pour ce faire, deux études de cas furent conduites. La première s'est intéressée à une formulation à base d'ibuprofène où le collage était décrit de façon qualitative. Une centaine de variables prédictives sur 71 lots industriels de cette formulation furent collectées et analysées en utilisant des techniques d'analyse discriminante et nous avons pu en isoler six pour leur rôle dans le collage. En accord avec la littérature, ces six variables confortent le rôle de la lubrification, du contenu en eau des poudres et de la faible température de fusion de l'ibuprofène dans sa propension au collage. La seconde étude de cas a investigué trois fois plus de variables prédictives sur 28 lots industriels d'une formulation différente à base d'ibuprofène et de métocarbamol où le collage était décrit par une variable quantitative liée à sa sévérité. En utilisant la technique de régression sur les moindres carrés partiels, nous avons démontré que l'association de la force de compression et de la durée de granulation permettait d'expliquer 60 % des fluctuations du collage de cette formulation.

De façon complémentaire, les résultats de ces deux études de cas font la preuve de concept d'une nouvelle approche d'étude du collage directement à l'échelle industrielle. Comparée à des modèles expérimentaux, celle-ci s'avère moins coûteuse, mais surtout plus efficace en permettant l'étude simultanée d'un nombre important de paramètres et en prévenant les problématiques de reproductibilité et de représentativité rencontrées à échelle laboratoire. Cette thèse élabore également sur la méthodologie et des recommandations pratiques à suivre pour un déploiement efficace de cette nouvelle approche prometteuse dans l'étude du collage.

**Mots clés :** collage de poudres pharmaceutiques, fabrication de comprimés, ibuprofène, metocarbamol, analyse multivariée des données, analyse en composantes principales, analyse discriminante et régression sur les moindres carrés partiels.

# RÉSUMÉ EN VERSION ANGLAISE

Sticking is one of the most common and damaging issues that occur during tablet manufacturing. Sticking is the adhesion of powder onto tooling surfaces during compression. Because of the numerous factors involved in its occurrence, understanding tablet sticking requires the simultaneous investigation of these factors to clarify their possible interactions. However, conducting such a study experimentally can present a significant financial and technical burden. Nevertheless, substantial data is usually generated during industrial-scale tablet manufacturing, where sticking is inherently prevalent.

This thesis aims to introduce, evaluate, and discuss an industrial data analysis approach to study tablet sticking. This was achieved through two case studies: qualitative and quantitative respectively. The first case study investigates the qualitative sticking descriptor of an ibuprofen-based formulation. A total of 71 industrial lots of this formulation described by a hundred predictor variables were gathered and analyzed using discriminant analysis methods. We have identified six parameters, which in agreement with the literature data, highlight the role of lubrication, water content, and the low melting point of ibuprofen in its sticking tendency. The second case study explored three times more variables on 28 industrial lots of an ibuprofen and methocarbamol-based formulation where sticking severity was described by a quantitative variable. Using a partial least squares regression method, we showed that the association of compression force and granulation duration explains 60% of the sticking fluctuations of the studied formulation.

Together, the results of these two case studies demonstrate the proof of concept of a new industrial data analysis approach to tablet sticking. In comparison to experimental models, this is a financially more viable option but also a more efficient approach to the understanding of tablet sticking. It allows simultaneous investigation of an important number of parameters and prevents specific issues encountered at lab scales such as reproducibility and representativity. This thesis also elaborated on the methodology and best practices for an efficient application of this new promising approach to better study, understand, and resolve tablet sticking.

**Keywords:** tablet sticking, powder compaction, ibuprofen, methocarbamol, multivariate data analysis, principal component analysis, discriminant analysis, and partial least squares regression.

# REMERCIEMENTS

Cette thèse est le fruit d'un long travail qui a bénéficié de l'apport inestimable de nombreuses personnes. Je souhaite donc utiliser les premières lignes de ce document pour les témoigner de ma profonde gratitude. Mes premiers mots de remerciement vont à l'endroit de mes encadrants académiques et industriels, notamment Pr Ryan Gosselin, mon directeur de thèse, Pr Nicolas Abatzoglou, mon co-directeur de thèse, Pr François Gitzhofer, Sophie Hudon, Jean-Sébastien Simard et Jean-Maxime Guay. Merci pour tout l'accompagnement technique et moral que vous avez consenti dans la réalisation de ce travail.

Ryan Gosselin, Nicolas Abatzoglou et Sophie Hudon, eu égard à votre contribution exceptionnelle, je me dois de vous adresser des remerciements spéciaux. Vous êtes pour beaucoup dans mon développement intellectuel et personnel tout au long de cette thèse. Ryan, je me rappellerai toujours mes débuts dans ce projet où tu m'as petit à petit enseigné les bases et l'amour du codage en python, ainsi que les rudiments pour approfondir mes connaissances en analyse de données. Merci pour la constance de ton apport technique, réflexive et de direction à ce projet, mais surtout merci pour toute la sollicitude et la bienveillance que tu as bien voulu accorder à ma personne au travers de toutes les épreuves pendant la réalisation de ce projet. Nicolas, des capacités réflexives et humaines que j'ai acquises au cours de la réalisation de cette thèse, beaucoup sont aussi de ton fait. Des premiers cours que tu nous as donnés (EFD901) et nos différentes interactions, j'ai gagné une grande capacité d'autonomie, de réflexion, de recul et de gestion de projet de recherche qui ont été des compétences clés dans la réalisation de cette thèse. Aussi, ta sollicitude et ton assistance technique constantes ont été déterminantes pour ce projet. Merci pour tous ces apports qui ont non seulement édifié le projet de thèse, mais surtout impacté positivement mon intellect et ma personne. Sophie, malheureusement, il n'existe pas d'autres mots autres que merci pour témoigner une gratitude quel que soit son degré. Alors, merci. D'énormes pans de ce travail n'existeraient pas sans ton implication profonde à faciliter mon intégration chez Pfizer, m'aider à trouver les bons interlocuteurs pour avoir des données et bien sûr, tout l'appui technique depuis la collecte et l'analyse des données à la production des articles. Merci alors pour ton dévouement, ton temps, ton assistance technique et ta bienveillance envers ma personne et ce projet.

Mes sincères remerciements également à toute ma famille, notamment mon père Lassana Dembélé, qui s'est toujours battu contre vents et marées pour que je puisse m'engager et mener à terme ce projet de thèse, ma mère Djeneba N'Daou, pour toute l'affection et

l'accompagnement moral constants qu'elle m'a accordé tout le long de cette thèse, mais surtout un merci spécial à ma femme, Oumou Tamboura, pour toute sa sollicitude, sa bienveillance et son soutien moral constant durant toutes les épreuves de cette thèse. Je me rappelle encore les soirées interminables endurées ensemble lors de la préparation de la DPR, mais aussi le test du degré de vulgarisation de la carte mentale du collage. Tu as été la première personne à valider cette idée originale de la présente thèse. Merci pour tout.

Je ne saurais terminer ces remerciements sans une mention spéciale à tous mes amis de « GONBOUGOU », Sidi Mohamed Diarra, Fatoumata Sow, Boubacar Maïga, Yiriba Samaké et Seydou Haïdara, eux qui n'ont ménagé aucun effort pour mon accueil et mon intégration à Sherbrooke. Merci pour tout votre accompagnement moral, notamment toi, mon « letcha », Sidi Mohamed Diarra. Tu fus d'un soutien indéfectible au travers de toutes les épreuves pendant la réalisation de cette thèse. Mes remerciements également à mon très cher ami et co-porteur de thèse, Komlan Koumbogle, les discussions techniques interminables résonnent encore dans ma tête. Merci pour toute ton assistance technique et ta bienveillance à mon endroit. De même, merci à Frank Dega et Ketsia Yekpe, des amis en or rencontrés au cours de la réalisation de cette thèse et qui en sont pour beaucoup dans son avènement.

Enfin, je remercie la compagnie Pfizer Canada et l'ensemble du Groupe de Recherche sur les Technologies et Procédés de m'avoir accueilli et offert l'opportunité de parfaire mon éducation à leurs côtés. Merci pour tout l'accompagnement technique, matériel et moral durant toutes les années de la réalisation de cette thèse.



# TABLE DES MATIERES

## Table des matières

|  |    |
|--|----|
| Résumé.....  | 5  |
| Résumé en version anglaise.....                            | 6  |
| Remerciements .....  | 7  |
| Table des matières.....                                    | 9  |
| Liste des figures.....                                     | 13 |
| Liste des tableaux.....                                    | 14 |
| Liste des abréviations .....                               | 14 |
| Chapitre 1 : Introduction .....                            | 18 |
| 1. Mis en contexte et problématique : .....                | 18 |
| 2. Questions de recherche et objectifs :.....              | 20 |
| Chapitre 2 : État de l'art .....                           | 23 |
| 1. Les comprimés et leurs procédés de fabrication :.....   | 23 |
| 1.1 La compression directe : .....                         | 24 |
| 1.2 La granulation sèche : .....                           | 25 |
| 1.3 La granulation humide : .....                          | 26 |
| 2. La compression pharmaceutique : .....                   | 27 |
| 2.1 L'alimentation ou le remplissage :.....                | 27 |
| 2.2 L'arasage :.....                                       | 28 |
| 2.3 La compression proprement dite :.....                  | 28 |
| 2.4 L'éjection : .....                                     | 32 |
| 2.5 Les machines et outillages de compression : .....      | 33 |
| 2.6 Problèmes survenant au cours de la compression :.....  | 35 |
| 3. Le collage de poudres pharmaceutiques :.....            | 36 |
| 4. Les techniques d'évaluation du collage : .....          | 37 |
| 4.1 Les méthodes indirectes d'évaluation du collage: ..... | 37 |
| 4.2 Les méthodes directes d'évaluation du collage : .....  | 39 |

|   |            |
|---|------------|
| <b>5. Les causes du collage de poudres pharmaceutiques :</b> .....  | <b>40</b>  |
| 5.1 Le mécanisme conceptuel du collage : .....  | 40         |
| 5.2 Les paramètres relatifs à la formulation : .....  | 43         |
| 5.3 Les paramètres relatifs à la compression : .....  | 56         |
| 5.4 Les paramètres relatifs aux outillages :.....   | 62         |
| 5.5 Les paramètres relatifs à l’environnement :.....  | 68         |
| <b>6. Synthèse de l’état de l’art sur le collage et introduction du projet de thèse : la carte mentale du collage .....</b> | <b>71</b>  |
| <b>7. Les techniques d’analyse multivariée et leur intérêt dans l’étude du collage à échelle industrielle : .....</b>       | <b>75</b>  |
| <b>Chapitre 3 : étude multivariée qualitative du collage.....</b>   | <b>79</b>  |
| <b>Avant-propos .....</b>   | <b>79</b>  |
| <b>Résumé français .....</b>  | <b>80</b>  |
| <b>Résumé en anglais .....</b>  | <b>81</b>  |
| <b>1. Introduction : .....</b>  | <b>82</b>  |
| <b>2. Materials and methods:.....</b>   | <b>86</b>  |
| 2.1 Formulation :.....  | 86         |
| 2.2 Data collection and preparation: .....  | 86         |
| 2.3 Analysis methodology: .....   | 87         |
| <b>3. Results and discussions : .....</b>   | <b>89</b>  |
| 3.1 Sticking data : .....   | 89         |
| 3.2 Univariate data analysis : .....  | 89         |
| 3.3 Principal component analysis: .....   | 91         |
| 3.4 Linear discriminant analysis of dataset1:.....  | 92         |
| 3.5 Partial least squares-discriminant analysis of dataset2: .....  | 99         |
| 3.6 Discussion of the significant insights identified: .....  | 101        |
| 3.7 A new approach to studying tablet sticking and its related systematic methodology: .....                                | 108        |
| <b>4. Conclusion .....</b>  | <b>111</b> |
| <b>5. Acknowledgments:.....</b>   | <b>112</b> |

|  |            |
|--|------------|
| 6. Declaration of competing interest: .....                                    | 112        |
| <b>Chapitre 4 : étude multivariée quantitative du collage .....</b>            | <b>113</b> |
| <b>Avant-propos .....</b>  | <b>113</b> |
| <b>Résumé français .....</b>   | <b>114</b> |
| <b>Résumé en anglais .....</b>   | <b>115</b> |
| <b>1. Introduction : .....</b>   | <b>116</b> |
| <b>2. Materials and methods: .....</b>   | <b>119</b> |
| 2.1 Formulation : .....  | 119        |
| 2.2 Gathering, cleaning, and preparation of data: .....                        | 119        |
| 2.3 Analysis methodology: .....  | 120        |
| 2.4 Model validation: .....  | 120        |
| <b>3. Results and discussions: .....</b>                                       | <b>122</b> |
| 3.1 Sticking data: .....   | 122        |
| 3.2 Univariate data analysis: .....  | 123        |
| 3.3 Principal component analysis: .....  | 125        |
| 3.4 Partial least squares regression: .....                                    | 125        |
| 3.5 Discussion of the significant insights: .....                              | 133        |
| <b>4. Conclusion : .....</b>   | <b>138</b> |
| <b>5. Declaration of competing interest: .....</b>                             | <b>139</b> |
| <b>6. Acknowledgments: .....</b>   | <b>139</b> |
| <b>Chapitre 5 : Conclusions générales et perspectives .....</b>                | <b>140</b> |
| <b>Références .....</b>  | <b>143</b> |
| <b>Annexes .....</b>   | <b>154</b> |
| <b>Annexe 1 : la lubrification externe .....</b>                               | <b>154</b> |
| <b>1. Généralités sur les lubrifiants : .....</b>                              | <b>154</b> |
| <b>2. Description de l'approche de lubrification externe : .....</b>           | <b>155</b> |
| 2.1 La lubrification interne : entre nécessité et désavantages .....           | 155        |
| 2.2 La lubrification externe : Délivrer du lubrifiant là où nécessaire ? ..... | 157        |
| <b>3. Les avantages de la lubrification externe : .....</b>                    | <b>159</b> |

|   |            |
|---|------------|
| <b>4. Les inconvénients de la lubrification externe :</b>                         | <b>162</b> |
| <b>5. Conclusions :</b>   | <b>164</b> |
| <b>Annexe 2 : Les techniques d'analyse multivariée à réduction dimensionnelle</b> | <b>166</b> |
| <b>1. Analyse en composantes principales (ACP) :</b>                              | <b>166</b> |
| 1.1 Le critère de l'ACP :   | 167        |
| 1.2 La solution :   | 167        |
| 1.3 Le centrage des données :   | 168        |
| 1.4 La réduction des données :  | 168        |
| 1.5 La matrice de corrélation / la matrice de variance-covariance                 | 168        |
| 1.6 Les éléments propres  | 168        |
| 1.7 Le graphique des scores et le cercle de corrélation                           | 168        |
| <b>2. La régression multilinéaire :</b>   | <b>169</b> |
| 2.1 Estimation des paramètres du modèle :   | 169        |
| 2.2 Significativité du modèle :   | 170        |
| <b>Annexe 3 : la gestion des données manquantes</b>                               | <b>172</b> |
| <b>1. Introduction :</b>  | <b>172</b> |
| <b>2. Revue de littérature sur la gestion des données manquantes :</b>            | <b>173</b> |
| 2.1 Les mécanismes de données manquantes :  | 174        |
| 2.2 Les méthodes de traitement des données manquantes :                           | 177        |
| <b>3. Étude d'application des techniques de gestion de données manquantes :</b>   | <b>189</b> |
| 3.1 Matériels et méthodes :   | 189        |
| 3.2 Résultats et discussions :  | 191        |
| <b>4. Conclusion :</b>  | <b>194</b> |
| <b>Annexe 4 : Résumé comparatif des résultats du dataset1 avec LDA et PLS-DA</b>  | <b>195</b> |

# LISTE DES FIGURES

|  |     |
|--|-----|
| Figure 1 : les trois voies de fabrication des comprimés.....   | 24  |
| Figure 2 : Illustration des différentes étapes de l'opération unitaire de compression de poudre pharmaceutique en comprimés. ....                              | 27  |
| Figure 3 : Illustration des différents types de déformation/fragmentation. ....  | 29  |
| Figure 4: Modèle conceptuel du collage proposé par Paul et collab. (2017) [46]. ....   | 41  |
| Figure 5 : Carte mentale du collage. ....  | 72  |
| Figure 6 : Scatter plot of input variables in dataset1 colored according to the presence of sticking (y variable). ....  | 90  |
| Figure 7 : The score plot of the PCA model built on dataset1.....  | 91  |
| Figure 8 : Score (A), loading (B), and variable importance (C) plots of the first descriptive LDA model developed on dataset1.....                             | 93  |
| Figure 9 : Score (A), loading (B), and variable importance (C) plots of the optimized descriptive LDA model developed on dataset1. ....                        | 95  |
| Figure 10 : Histogram of the classification error rate for the predictive LDA models.....  | 97  |
| Figure 11 : Score (A), loading (B), and VIP (C) plots of the optimized PLS-DA model. ....  | 100 |
| Figure 12 : Histogram of the classification error rate for optimized PLS-DA (25 batches x 10 variables).....   | 101 |
| Figure 13 : Box plot of mean comparison between batches affected by sticking and those unaffected by sticking for V8. ....                                     | 102 |
| Figure 14 : Box plots of mean comparison between affected batches and unaffected batches for the environmental-related parameters.....                         | 104 |
| Figure 15 : Box plots of the mean comparison between affected and unaffected batches for the average disintegration time (V42). ....                           | 106 |
| Figure 16 : Box plots of the mean comparison between affected and unaffected batches for the total duration of stops during compression, in minutes (V44)..... | 107 |
| Figure 17 : Distribution of the number of sticking events (y-variable) reported across all gathered batches.....   | 122 |
| Figure 18 : Scatter plots of quantitative input variables. ....  | 123 |
| Figure 19 : Box plot displaying the mean comparison of the sticking events reported between different classes of qualitative input variables.....              | 124 |

|   |     |
|---|-----|
| Figure 20 : Score plot of the two first PCA components. ....  | 125 |
| Figure 21 : Performances of the first PLS model in calibration and validation. ....   | 126 |
| Figure 22 : Histogram of the evolution of RMSEC and RMSECV during the optimization procedure by filtering through VIP (threshold set to 1).....                   | 128 |
| Figure 23 : Overview of the optimized PLS model. ....   | 130 |
| Figure 24 : Predictive performances of the optimized PLS model. ....  | 132 |
| Figure 25 : Scatter plot of sticking behavior following the duration of colored layer granulation (V65) and the variation of bilayer compression force (V94)..... | 134 |
| Figure 26 : Représentation schématique d'un système de lubrification externe [99]......   | 158 |
| Figure 27 : carte de données manquantes des différents jeux de données utilisés. ....   | 191 |
| Figure 28 : Résultats d'évaluation des performances à la reconstruction des données originales. ....  | 193 |

## LISTE DES TABLEAUX

|  |     |
|--|-----|
| Table 1: caractéristiques des principaux types d'outillage de compression. ....                      | 35  |
| Table 2 : Repartition of the number of batches following the qualitative y variable modalities. .... | 89  |
| Table 3 : ANOVA results of the first PLS model.....  | 126 |
| Table 4 : ANOVA results of PLS models developed through the optimization process.....                | 128 |
| Table 5 : Résultats d'évaluation des performances à la reconstruction des données originales. ....   | 192 |

## LISTE DES ABREVIATIONS

| Abréviations | Définitions                        |
|--------------|------------------------------------|
| <b>Ace-H</b> | Acesulfame                         |
| <b>Ace-K</b> | Acésulfame potassium               |
| <b>ACP</b>   | Analyse en composantes principales |

| <b>Abréviations</b> | <b>Définitions</b>  |
|---------------------|---|
| <b>AFM</b>          | Microscopie à force atomique                                  |
| <b>ANOVA</b>        | Analyse de la variance  |
| <b>API</b>          | Ingrédient actif de la formulation                            |
| <b>BHB</b>          | N-butyl p-hydroxybenzoate                                     |
| <b>BPF</b>          | Bonnes pratiques de fabrication                               |
| <b>CAF</b>          | Caféine   |
| <b>CC</b>           | Croscarmellose  |
| <b>CEL</b>          | Celecoxib   |
| <b>CMCF</b>         | Méthode de la valeur la plus fréquemment rencontrée           |
| <b>CMCM</b>         | Technique de Monte-Carlo par chaîne de Markov                 |
| <b>CrN</b>          | Poinçon revêtu de nitrure de chrome                           |
| <b>DCPD</b>         | Phosphate dicalcique dihydraté                                |
| <b>DoF</b>          | Degré de liberté  |
| <b>EDS</b>          | Spectroscopie de rayons X à dispersion d'énergie              |
| <b>EM</b>           | Algorithme d'estimation et maximisation                       |
| <b>EU standard</b>  | Norme de fabrication européenne des outillages de compression |
| <b>FCS</b>          | Spécification conditionnelle pleine                           |
| <b>GLY</b>          | Glyburide   |
| <b>GMS</b>          | Glyceryl monostéarate   |
| <b>HCr</b>          | Chrome dur  |
| <b>HPLC</b>         | Chromatographie Liquide à Haute Performance                   |
| <b>HPMC</b>         | Hydroxy-propyl-méthyl-cellulose                               |
| <b>IBU</b>          | Ibuprofène  |
| <b>IBU-Na</b>       | Ibuprofène sodium   |
| <b>JN</b>           | Norme de fabrication japonaise des outillages de compression  |

| <b>Abréviations</b> | <b>Définitions</b>   |
|---------------------|--|
| <b>KNN</b>          | K proches voisins  |
| <b>LDA</b>          | Analyse discriminante linéaire   |
| <b>LOCF</b>         | Méthode de la dernière valeur connue   |
| <b>MA</b>           | Acide méfénamique  |
| <b>MAR</b>          | Données manquantes au hasard   |
| <b>MC</b>           | Méthylcellulose  |
| <b>MCAR</b>         | Données manquantes complètement au hasard  |
| <b>MCC</b>          | Microcristalline cellulose   |
| <b>MgSt</b>         | Stéarate de magnésium  |
| <b>MICE</b>         | Imputation multiple par équations chaînées   |
| <b>MLR</b>          | Régression multilinéaire   |
| <b>MNAR</b>         | Données non manquantes au hasard   |
| <b>MP</b>           | Méthyl parabène  |
| <b>MVDA</b>         | Analyse multivariée des données  |
| <b>NIPALS</b>       | Nonlinear Iterative Partial Least Squares  |
| <b>PCR</b>          | Régression sur les composantes principales   |
| <b>PLS</b>          | Régression sur les moindres carrés partiels  |
| <b>PLS-DA</b>       | Analyse discriminante par régression sur les moindres carrés partiels              |
| <b>POF</b>          | Force de décollement du comprimé de la surface du poinçon supérieur                |
| <b>RMSEC</b>        | Racine carrée de la somme des carrés des écarts résiduels de la calibration        |
| <b>RMSECV</b>       | Racine carrée de la somme des carrés des écarts résiduels de la validation croisée |
| <b>SEM</b>          | Microscopie électronique à balayage  |
| <b>SSF</b>          | Sodium stearyl fumarate  |
| <b>TFC</b>          | Tofacinitib  |



| Abréviations      | Définitions   |
|-------------------|---|
| <b>TiN</b>        | Poinçon revêtu de nitrure de titanium   |
| <b>TOF ou SCR</b> | Force de détachement du comprimé de la surface du poinçon inférieur                                 |
| <b>TSM</b>        | Tableting Specification Manual (norme de fabrication nord-américaine des outillages de compression) |
| <b>U</b>          | Poinçon non revêtu  |

# CHAPITRE 1 : INTRODUCTION

## 1. Mis en contexte et problématique :

Dans une quête perpétuelle d'amélioration de ses conditions de vie, l'homme n'a de cesse de réaliser des progrès. Certains sont des sauts et d'autres sont des bonds de géant. Parmi ces derniers, nous pouvons citer la science pharmaceutique qui est l'un de ces progrès ayant radicalement changé la qualité et l'espérance de vie des êtres humains. De nos jours, des médicaments sont disponibles pour une large majorité de maux et de nouvelles avancées sont accomplies chaque jour. De la pléthore de formes médicamenteuses existantes, les comprimés sont l'une des plus communes et des plus répandues [1,2]. Ils permettent de prévenir, guérir et soulager de nombreuses maladies plus ou moins graves. La popularité des comprimés est principalement due à leur grande stabilité physicochimique et leur large acceptation par les patients à cause de la simplicité de dosage et d'administration [1-5]. Les industriels apprécient également la forme « comprimé » à cause de sa commodité de fabrication [6-8].

La production de comprimés consiste en la compression d'un mélange de poudre particulaire ou granulaire constitué d'un ingrédient actif (API) et d'excipients. L'API confère ses vertus thérapeutiques au comprimé, tandis que les excipients lui donnent des propriétés physicochimiques particulières. Précédée ou non par une étape de granulation, l'opération de compression consiste en l'application d'une contrainte physique sur le mélange de poudre. Celle-ci est réalisée sur une machine appelée presse à comprimer. La contrainte physique est appliquée verticalement à l'aide d'outils appelés poinçons qui compriment la poudre dans une cavité appelée matrice. Au nombre de deux, les poinçons sont montés sur la presse de part et d'autre de la matrice et forment avec ceux-ci les outillages de compression.

Lors de la compression du mélange de poudre en comprimés, des particules de poudre ont tendance à adhérer à la surface des outillages. Connue sous le nom de collage, le phénomène consiste en l'accumulation d'une ou plusieurs couches de poudre de particules sur la surface des outillages de compression. Le collage aux matrices est peu reporté. Les études sont centrées sur l'adhésion aux surfaces des poinçons. Le phénomène de collage n'apparaît généralement que lors des campagnes de production à échelle industrielle où les possibilités d'investigation et de changement sont limitées [9,10]. À ce niveau, il a un effet dévastateur sur la productivité. Lorsque le collage apparaît à l'échelle industrielle, les productions sont interrompues pour nettoyer les équipements et les outillages. Les comprimés résultants sont considérés comme

défectueux et rejetés. Le nettoyage et le polissage récurrents des outillages induisent aussi leur usure prématurée [5,9]. À travers ces pertes de temps, de matière et problèmes mécaniques, le collage conduit inexorablement à l'augmentation des coûts et délais de production. Dans le contexte très concurrentiel de l'industrie pharmaceutique, cela pose un véritable frein à la compétitivité des entreprises. Les coûts de production faisant partie intégrante du prix de cession des médicaments, le collage revêt également un enjeu social.

Devant cet enjeu industriel et social, les industriels et les académiques mènent la riposte par l'intensification des recherches et des initiatives pour comprendre le mécanisme du collage et trouver des solutions pour y remédier. Plusieurs paramètres ont ainsi été rapportés dans la littérature comme ayant un effet sur le collage. Ces paramètres concernent tous les aspects de la compression : les propriétés des formulations (l'API et les excipients), les propriétés de surface des outillages, ainsi que les conditions opératoires et environnementales de la production. Cependant, les causes fondamentales permettant une description globale et complète du phénomène peinent à émerger [9,10]. Cela est principalement dû au fait que le collage résulte d'un équilibre dynamique et très fragile gouverné par un nombre particulièrement important de paramètres, mais surtout l'existence de multiples interactions entre ceux-ci. Les données de la littérature établissent l'implication d'environ une trentaine de paramètres dans l'occurrence du collage. Parmi ceux-ci, près de 50 % sont concernés par des effets d'interaction ou d'interdépendance. Par exemple, une importante force de compression permet généralement de réduire le collage, mais cet effet peut changer en fonction de la plasticité des particules d'API et de la rugosité de la surface des poinçons [11–16]. De même, la lubrification a généralement un effet bénéfique sur le collage, mais celui-ci peut varier en fonction de la forme du bout des poinçons et de leur type de revêtement [14,17–19]. À l'opposé, l'humidité relative peut favoriser le collage, mais son effet est modulé par la forme des cristaux de particule ainsi que l'hydrophobicité et la rugosité de la surface des poinçons [20,21]. L'effet de l'âge des poinçons dépend lui aussi du type de leur revêtement [14] et l'impact des propriétés d'un revêtement de surface change d'un API à l'autre [19,22]. En plus de tous ceux-ci, toute modulation d'une des nombreuses propriétés physicochimiques ou électrostatiques des API impacte grandement l'ensemble de l'équilibre des forces conduisant au collage [15,21–25].

L'étude, la compréhension et la résolution d'un tel phénomène multifactoriel nécessitent une approche intégrée et simultanée où les paramètres aux effets interdépendants doivent être étudiés ensemble [9,10,26]. Ceci est essentiel non seulement à une compréhension globale du phénomène, mais aussi à l'intercomparabilité de différentes études. En effet, les

travaux ne s'intéressant simultanément qu'à une poignée de paramètres peineront à décrire une compréhension globale complète et les multiples interactions rendront difficile le croisement d'études n'ayant pas été effectuées exactement dans les mêmes conditions.

Vu le nombre particulièrement important des paramètres et des interactions impliquées dans le collage, des expérimentations simultanées à échelle laboratoire pour le comprendre s'avèreraient très coûteuses financièrement. De plus, à cause des faibles vitesses et durée de compression, même une bonne étude dispendieuse du collage à échelle laboratoire demeurerait entachée par des problèmes de reproductibilité et de représentativité. En effet, les rapports de la littérature concordent et montrent que les longues séances de compression à haute cadence industrielle sont en partie des causes du collage [15,16,18,27]. La reproduction de telles conditions expérimentales à l'échelle laboratoire représenterait des coûts supplémentaires importants. Aussi, certains paramètres connus pour leur influence sur le collage sont techniquement difficiles à reproduire et à étudier à échelle laboratoire. Il s'agit entre autres de l'âge des poinçons, leur usure ou encore le temps de contact. Par conséquent, les modèles d'étude expérimentale à échelle laboratoire sont très limités à décrire de façon globale et complète la compréhension du collage. Cependant, une quantité phénoménale de données est générée lors de la fabrication des comprimés à l'échelle industrielle. Une potentielle exploitation effective de ces données permettrait d'ouvrir de nouvelles perspectives dans l'étude, la compréhension et la résolution du collage. La présente thèse s'intéresse à cette éventualité et se propose de répondre aux questions de recherche et objectifs décrits ci-dessous.

## **2. Questions de recherche et objectifs :**

Les questions de recherche au cœur de la présente thèse sont :

- a. Est-il possible d'exploiter les données industrielles pour étudier, comprendre et/ou résoudre le collage ?
- b. Comment une telle approche devrait-elle être conduite ?
- c. Quels types de résultats une telle approche pourrait-elle procurer ?
- d. Quels seraient les avantages et les limites de cette approche ?

L'objectif de la thèse est ainsi d'introduire, définir, évaluer et discuter la pertinence d'une nouvelle approche prometteuse dans l'étude, la compréhension et la résolution du collage. Celle-ci consiste à tirer parti de la quantité phénoménale de données générées lors de la production industrielle des comprimés pour développer de la connaissance autour du collage. Comparée à des modèles expérimentaux, cette nouvelle approche serait très peu coûteuse, mais

surtout elle pourrait s'avérer plus efficace. En effet, l'étude du collage avec une approche d'analyse des données industrielles permettrait d'étudier simultanément beaucoup plus de paramètres qu'une approche expérimentale. Les paramètres opératoires des opérations unitaires en amont de la compression pourraient également être investigués. Aussi, les faibles tolérances de variation des paramètres à l'échelle industrielle permettraient de mettre l'accent sur les variables auxquelles le collage est vraiment sensible. En plus de tous ceux-ci, l'étude du collage directement à échelle industrielle permettrait aussi de prévenir les problématiques de représentativité et de reproductibilité rencontrées à échelle laboratoire.

La nouvelle approche d'étude industrielle du collage proposée dans cette thèse est basée sur l'utilisation des techniques d'analyse multivariées (MVDA pour « Multivariate Data Analysis » en anglais). Ce sont un ensemble de méthodes permettant l'exploration, la modélisation et l'analyse de jeu de données décrites par plusieurs variables. Les techniques comme l'analyse en composantes principales ou la régression sur les moindres carrés partiels sont au cœur de la présente thèse. Elles permettent de réaliser une réduction dimensionnelle des données en résumant avec quelques variables latentes la majeure partie de l'information capturée par les nombreuses variables originelles d'un jeu de données. Ces méthodes sont parfaites pour l'analyse des données industrielles, car un nombre très important de variables est mesuré au cours de la production industrielle, mais peu de lots sont exécutés. Cette particularité génère des matrices de jeu de données rectangulaires qui manquent d'assez de degrés de liberté permettant l'utilisation des techniques classiques de régression ou de classification directement sur les variables originelles. De plus, les problèmes fréquents de multicollinéarité dans de tels jeux de données rectangulaires ne sont pas un challenge pour les méthodes d'analyse multivariée à réduction dimensionnelle.

En utilisant ces techniques d'analyse multivariée dans cette thèse, nous avons fait la preuve de concept de la nouvelle approche d'étude du collage à travers deux études de cas. La première s'est intéressée à une formulation à base d'ibuprofène où le collage était décrit par une variable qualitative. Pour cette formulation, 71 lots industriels décrits par une centaine de variables prédictives furent collectés et analysés en utilisant des techniques d'analyse multivariée de type discriminant. La seconde étude de cas a investigué trois fois plus de variables prédictives sur une formulation différente à base d'ibuprofène et de metocarbamol où le collage était décrit par une variable quantitative. Nous avons collecté et analysé un total de 28 lots industriels de cette seconde formulation en utilisant la technique d'analyse multivariée de régression sur les moindres carrés partiels. De façon complémentaire, les résultats de ces

deux études de cas démontrent clairement la pertinence de la nouvelle approche d'étude du collage à échelle industrielle. Elles illustrent également la méthodologie à utiliser selon le descripteur du collage, mais surtout elles discutent de façon détaillée les avantages, les limites et les perspectives d'une telle approche. Les deux études de cas ont chacune donné lieu à un article scientifique actuellement en processus de publication (l'un accepté et l'autre en révision des pairs). De plus, sur la base de l'expérience acquise dans la conduite des deux études de cas, nous avons élaboré sur des recommandations pour une meilleure efficacité dans le déploiement de la nouvelle approche d'étude du collage à échelle industrielle.

À la suite de cette introduction, le présent document de thèse aborde dans le deuxième chapitre, un état de l'art des connaissances autour du collage. Nous commençons par une description des généralités ayant trait à la production de comprimés avec un accent sur l'opération de compression. Nous nous intéressons ensuite au problème particulier du collage, en mettant l'accent principalement sur l'appréhension de la problématique de la présente thèse : l'aspect multifactoriel du phénomène et les difficultés des modèles d'étude expérimentaux à le décrire de façon complète. Nous proposons ainsi une description des techniques d'évaluation utilisées dans la littérature pour quantifier le collage avant d'élaborer sur ses causes. L'élaboration des causes commence par une description détaillée du mécanisme conceptuel permettant d'expliquer le phénomène, puis nous discutons de façon extensive l'implication de chacun des paramètres évoqués dans la littérature comme contribuant au collage. Dans cet exercice, nous avons prêté une attention particulière à analyser, comparer et confronter de nombreuses études traitant du même paramètre, afin de dégager un portrait complet de l'état des connaissances sur le collage au-delà de son caractère multifactoriel. Par la suite, nous proposons en guise de transition pour introduire les objectifs de la présente thèse, une synthèse de l'état de l'art des causes du collage, notamment à travers une proposition originale : la carte mentale du collage. Celle-ci résume de façon logique et visuelle l'équilibre complexe, dynamique et fragile qui conduit à l'occurrence du collage au cours de la compression. Le deuxième chapitre se termine par une introduction succincte aux techniques d'analyse multivariée et leur intérêt aux fins des objectifs de la présente thèse. Le troisième chapitre de cette thèse est ensuite consacré à la première étude de cas. Celle-ci est présentée sous le format de l'article scientifique qui en a découlé. De façon similaire, le quatrième chapitre élabore sur la deuxième étude de cas. Enfin, le dernier chapitre propose une conclusion générale faisant la synthèse de l'ensemble des résultats de cette thèse et les différentes perspectives de la nouvelle approche d'étude du collage à échelle industrielle.

## CHAPITRE 2 : ÉTAT DE L'ART

Ce chapitre présente l'état de l'art en relation avec le sujet de la thèse. Les généralités de la compression pharmaceutique sont abordées, puis l'accent est mis sur l'état des recherches autour de la compréhension du collage, notamment les méthodes employées couramment pour le mesurer, son mécanisme conceptuel et ses causes. Le chapitre 2 propose également une synthèse originale de l'analyse de l'état de l'art sous forme d'une carte mentale qui fait ressortir la trame rationnelle de la présente thèse de doctorat. Enfin, cette analyse approfondie est suivie d'une introduction aux techniques d'analyse multivariée et leur intérêt dans l'étude du collage de poudres pharmaceutiques.

### **1. Les comprimés et leurs procédés de fabrication :**

Les comprimés sont les formes de médicaments les plus communes et répandues par rapport à toutes les autres formes existantes [1,2]. Cet état de fait tient aux nombreux avantages de cette forme solide, notamment sa stabilité et sa simplicité de fabrication, manipulation et utilisation. En effet, les constituants d'une formulation pharmaceutique sont plus stables dans une préparation solide que liquide [3,4]. Ce paramètre est d'autant plus important que la stabilité physique, chimique et biologique de l'API et des excipients est cruciale pour l'efficacité des médicaments. La forme solide « comprimé » se prête également bien aux différentes opérations de conditionnement, stockage et distribution. De plus, les patients apprécient aussi la simplicité du dosage et de l'administration des comprimés [1,4,5]. Un autre avantage des plus importants est la simplicité technique et économique de la fabrication de cette forme solide [2,6–8].

Les comprimés sont fabriqués par compression sur une presse à comprimer par application d'une contrainte sur la poudre de formulation constituée de l'API et des excipients. L'API confère son effet thérapeutique à la formulation et les excipients, de nature diverse, procurent des propriétés physiques et chimiques particulières aux comprimés. Parmi les excipients généralement utilisés, nous pouvons citer les lubrifiants comme le stéarate de magnésium (MgSt), les liants comme la microcristalline cellulose (MCC) et la povidone, les désintégrants comme la croscarmellose sodique et les diluants comme le lactose ou le mannitol [1,2,5,8,28]. Ces excipients sont mélangés à l'API à des proportions bien déterminées, puis, la formulation résultante est éventuellement granulée et comprimée pour former des compacts solides et consistants, de poids et d'épaisseur identiques appelés « comprimés ». Cette production de

comprimés se fait généralement selon trois voies principales : la compression directe, la granulation sèche et la granulation humide [7,8]. La Figure 1 ci-dessous présente schématiquement les principales unités d'opérations constitutives des trois voies de fabrication des comprimés.

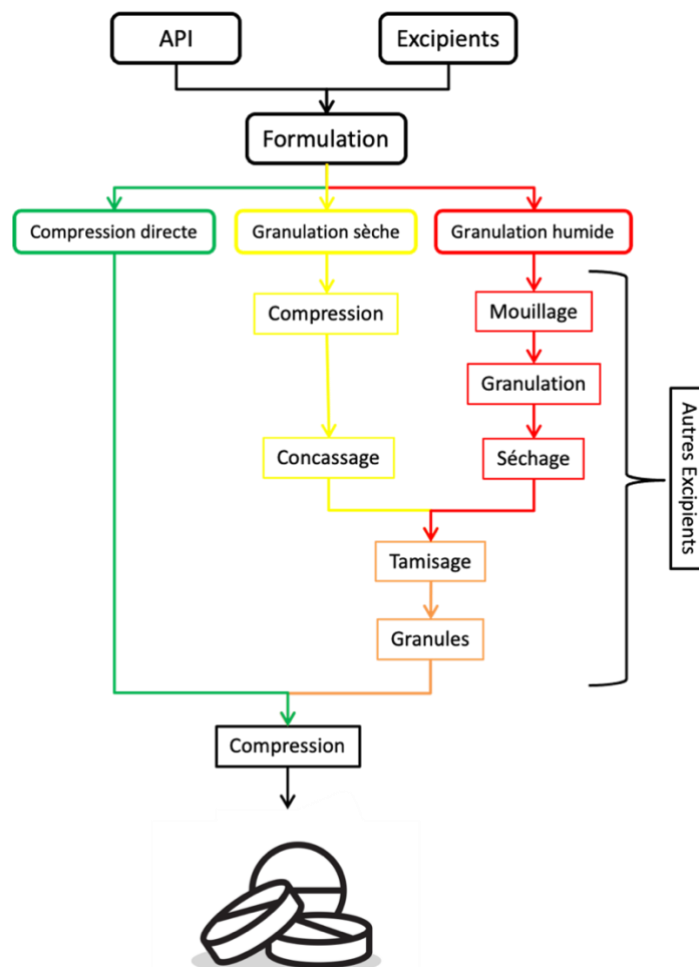


Figure 1 : les trois voies de fabrication des comprimés. Les trois premières boîtes colorées indiquent les trois voies de fabrication suivant par les opérations unitaires s'y rattachant.

Comme illustré dans la Figure 1, la compression directe est la voie la plus rapide de fabrication des comprimés. Les voies de fabrication impliquant la granulation sont plus longues. Elles impliquent des étapes de compression supplémentaire, de concassage et de tamisage pour la granulation sèche et des étapes de mouillage-granulation, de séchage et de tamisage pour la granulation humide.

### 1.1 La compression directe :

La compression directe est la méthode la plus rapide et la plus efficace de fabrication des comprimés, car elle nécessite très peu d'étapes et d'équipements [2,3,6–8]. L'API est mélangé



aux excipients et la formulation résultante est directement compressée sur une presse pour former les comprimés. Comparé aux autres voies de fabrication, c'est le procédé le plus économique, car le temps, l'espace et l'énergie nécessaires sont minimaux [3,28,29]. Les risques de contamination sont également limités et la méthode peut être employée pour les API sensibles à la température et à l'humidité, car il n'y a pas d'étape de mouillage ou de séchage [6,29].

En dépit de ses indéniables atouts, le recours à la compression directe est restreint par de nombreuses contraintes qu'elle pose sur le choix des constituants de la formulation, notamment de bonnes propriétés de compressibilité, de compactibilité, une bonne fluidité et une faible ségrégation [3,5,7,8]. Le choix des excipients, mais surtout des API est largement limité par ces contraintes, car la plupart n'ont pas de bonnes propriétés de compressibilité. Leur tendance à la ségrégation peut aussi compromettre l'uniformité de la formulation et l'homogénéité des comprimés. L'ensemble de ces facteurs combinés limite le recours systématique à la compression directe. En outre, des techniques dites de granulation sont utilisées pour améliorer les propriétés des poudres avant de les comprimer. Comme son nom l'indique, la granulation consiste en la formation de granules de taille supérieure. Elles permettent d'améliorer les propriétés de compressibilité et d'écoulement, mais aussi de limiter la ségrégation des particules de poudre en homogénéisant la distribution granulométrique.

## **1.2 La granulation sèche :**

La granulation sèche est une méthode alternative à la compression directe qui n'implique pas l'utilisation d'un liquide de mouillage jouant le rôle de liant dans le lit de particules. C'est une voie plus longue et plus coûteuse que la compression directe, mais nécessite moins d'étapes que la granulation humide. Elle est généralement employée lorsque la formulation a de mauvaises propriétés physiques pour une compression directe et que celle-ci est aussi sensible à l'ajout de liant sous forme liquide ou à la température pour passer en granulation humide. Les liants utilisés dans la granulation sèche sont très souvent des poudres d'amidon, de gélatines, de sucres ou de dérivés cellulosiques [7,8]. Pour former des granulés en l'absence d'humidité, la granulation sèche implique une compression primaire. Celle-ci permet de compacter et de réduire la distribution de tailles du mélange particulaire de façon à produire un mélange granulaire de taille uniforme qui s'écoule librement [1,30]. Le produit résultant de la compression primaire est ensuite concassé, tamisé et éventuellement complété par un mélange final avant d'être finalement compressé à nouveau pour former des comprimés.

Selon la technique employée pour la compression primaire, nous pouvons distinguer deux méthodes de granulation sèche : la granulation par plaques où une presse à comprimer forme de grandes plaques et la granulation sous rouleaux compacteurs où le lit de poudre passe à travers des rouleaux de compression et forme des feuillets de matière (rubans). Cette dernière est la méthode largement employée dans l'industrie, car la méthode d'alimentation dans la granulation par plaques conduit fréquemment à des problèmes de densification [1,3,30]. Lors de la granulation sous rouleaux compacteurs, l'alimentation est assurée par une trémie avec une vis sans fin, conférant une meilleure uniformité.

### **1.3 La granulation humide :**

La granulation humide est la méthode de granulation la plus utilisée dans l'industrie. Elle implique l'utilisation d'un liquide de mouillage qui permet une agglomération préliminaire des particules de poudre sans application de pression [3,5,7,30]. Le liquide de mouillage est généralement un mélange de solvant très souvent volatil comme l'eau, l'éthanol ou l'isopropanol et de divers liants [3,7,30,31]. Ces derniers favorisent l'établissement de liaisons interparticulaires et améliorent donc la cohésion interne des granulés.

La granulation humide consiste à un mélange des constituants de la formulation, qui sont ensuite mouillés et granulés, puis séchés, tamisés si nécessaire pour homogénéiser la taille des particules et ils sont finalement comprimés. Ces nombreuses étapes constituent la limite principale de la granulation par voie humide, car elles allongent les délais de production et augmentent les risques de perte de rendement. Aussi, les étapes de mouillage et de séchage limitent l'utilisation de cette voie de fabrication pour les API thermolabiles et sensibles à l'humidité.

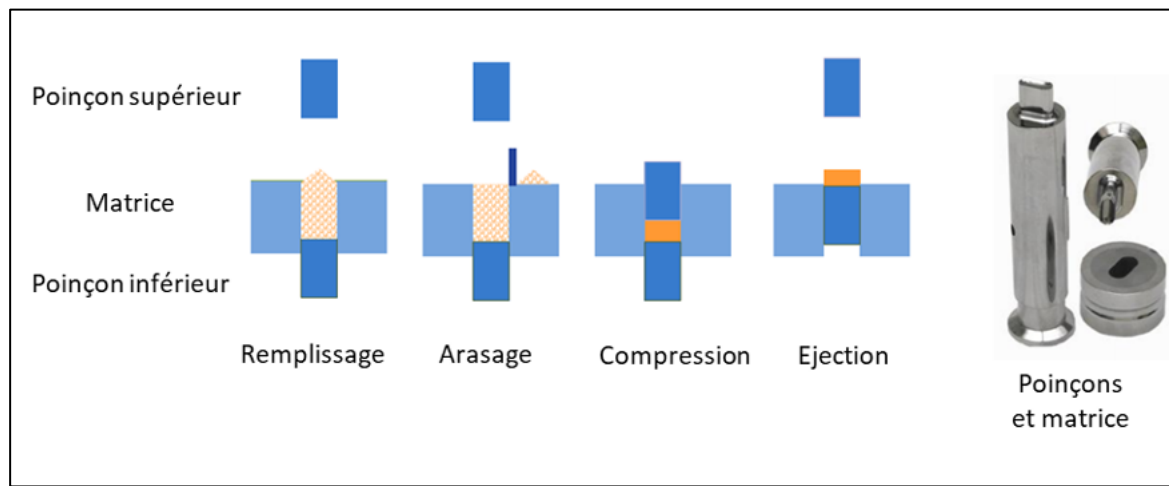
Cependant, la granulation humide demeure la méthode la plus employée dans l'industrie, car elle présente de nombreux autres avantages, notamment la limitation de l'émission de poussières par réduction des particules fines, la limitation de la ségrégation et l'amélioration de la cohésion et de la coulabilité [7]. Nombreuses variantes de granulation humide existent et plusieurs sont aussi en développement. Une bonne revue détaillée de ces variantes et les récents progrès a été publiée par Shanmugam et collab. (2017) [30].

En somme, suivant les propriétés physiques et chimiques des API et des excipients, les comprimés peuvent être fabriqués en compression directe ou après des étapes de granulation sèche ou humide. Toutefois, l'opération unitaire finale de toutes ces voies est commune et

consiste en la compression de poudre de mélanges particulières ou granulaires sous forme de doses uniques de comprimés.

## **2. La compression pharmaceutique :**

La compression pharmaceutique est l'étape ultime de la fabrication de comprimés. Elle consiste en l'application d'une pression axiale par les poinçons sur une quantité définie de poudre encastrée dans la matrice. La Figure 2 ci-dessous présente les principales étapes de la compression.



*Figure 2 : Illustration des différentes étapes de l'opération unitaire de compression de poudre pharmaceutique en comprimés.  
Images de Petr Zámotny [32].*

L'opération unitaire de la fabrication de comprimé peut être divisée en quatre étapes distinctes : le remplissage de la cavité de compression, l'arasage de l'excédent de poudre, la compression proprement dite par l'application de la force de compression et l'éjection du comprimé néoformé.

### **2.1 L'alimentation ou le remplissage :**

Elle consiste au remplissage de la cavité de compression par la poudre de formulation. Lors de cette étape, le poinçon du bas se positionne dans la matrice pour former la cavité ou la chambre de compression. La position du poinçon et la géométrie de la matrice déterminent ainsi le volume de la cavité. Le sabot d'alimentation de la presse assure ensuite le remplissage des chambres de compression avec la poudre de formulation qui s'écoule à partir de la trémie de la presse [4,5]. Les propriétés d'écoulement des poudres sont très importantes à cette étape, car elles influent sur la quantité et l'homogénéité du contenu des comprimés [31]. Selon l'équipement utilisé, la poudre s'écoule soit par gravité ou plus généralement par un mécanisme

automatisé couplé à la gravité. Nous assistons alors à un réarrangement primaire des particules s'écoulant dans la cavité pour former un empilage [5,7,31].

## **2.2 L'arasage :**

Pour assurer un remplissage total, la cavité de compression est surremplie de poudre à rebord et l'étape d'arasage consiste à l'évacuation de cet excès de poudre. Cela est réalisé grâce au retour du sabot d'alimentation. L'arasage permet ainsi de calquer le volume exact de poudre à comprimer préalablement déterminé par la hauteur du poinçon inférieur à l'aide d'une came, appelée came de dosage [7,8,31].

## **2.3 La compression proprement dite :**

C'est l'opération de mise sous forme comprimés proprement dite où la poudre contenue dans la cavité de compression est soumise à une pression pour former un compact. Le processus peut être scindé en 3 sous-étapes distinctes : le tassement ou la densification, la consolidation et la décharge.

### **2.3.1 Le tassement ou la densification :**

Au début de l'étape de compression, la contrainte est appliquée graduellement sur le lit de poudre encastré dans la cavité de compression. Il se produit alors un réarrangement secondaire par densification [5,10]. Au fur et à mesure que la contrainte exercée augmente, les particules se réarrangent et se rapprochent par des glissements et des rotations [4,8]. La porosité diminue graduellement, car l'air contenu dans le lit de poudre est chassé, forçant les particules à se rapprocher. Les points de contact des particules augmentent et nous assistons à la formation d'un lit de poudre plus dense.

### **2.3.2 La consolidation :**

Toujours sous contrainte, la densification atteint un point critique où les particules sont immobilisées et aucun réarrangement n'est plus possible. À ce moment, la formation d'un compact est induite par une série de processus simultanés de déformations et de fragmentation qui aboutiront à la formation de liens entre les particules mises en contact [4,7,8]. Ce sont ces liens interparticulaires qui maintiennent la structure du compact. En fonction de leurs propriétés mécaniques, la vitesse et l'intensité de la contrainte appliquée, les particules en mélange peuvent se déformer selon trois mécanismes : déformation élastique, déformation plastique et fragmentation (voir Figure 3).

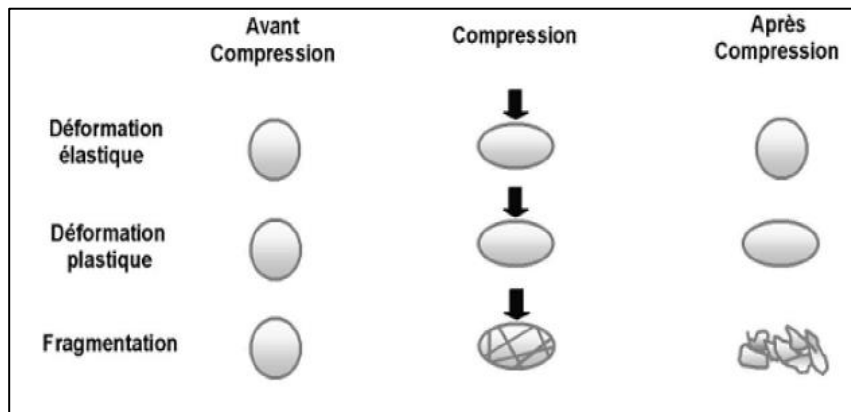


Figure 3 : Illustration des différents types de déformation/fragmentation.  
Images de Amal TITA-GOLDSTEIN [7].

- **La déformation élastique :**

La déformation élastique est la déformation réversible que subit une particule soumise à une contrainte comprise dans ses limites d'élasticité [5,7]. Cette déformation est simultanée à l'application de la contrainte et disparaît à l'arrêt de celle-ci.

- **La déformation plastique :**

C'est la déformation irréversible que subissent les particules ductiles soumises à d'importantes pressions [5,7,8]. Ce sont des déformations qui se produisent sans changement de volume. Contrairement aux déformations élastiques, elles induisent des changements permanents sur les formes des particules (voir Figure 3).

- **La fragmentation :**

La fragmentation est la fracture des particules en fragments de tailles plus petites qui peuvent subir une toute nouvelle série de réarrangements par glissements et rotations les uns par rapport aux autres, suivi d'une densification et de processus de déformations élastiques recouvrables ou plastiques permanentes [5,7,8].

Dans la littérature, plusieurs équations sont utilisées pour décrire les comportements de déformation des particules, mais l'équation de Heckel est la plus utilisée dans la fabrication des comprimés. Cette relation tire parti de l'évolution de la porosité en fonction de la pression de compression pour caractériser la densification des poudres. L'équation de Heckel est basée sur l'hypothèse que la densification suit une cinétique de premier ordre [5,7,8].

$$\ln\left(\frac{1}{\varepsilon}\right) = K \cdot P + A$$

Équation 1 : Equation de Heckel

Où  $\epsilon$  est la porosité et P la pression de compression en Pascals (Pa). A est une constante déterminée à travers l'équation caractérisant le réarrangement initial des particules. K est la pente de la courbe. Sa raideur indique une grande proportion de déformation plastique.

Du point de vue de la formation des comprimés, les particules constitutives du lit de poudre vont se déformer suivant les trois mécanismes décrits ci-dessus. Pour des contraintes de faibles intensités n'excédant pas leur domaine d'élasticité, elles se déforment de façon réversible. En fonction de l'intensité et de la durée d'application de la force, les particules ductiles commencent à se déformer de façon irréversible. Les particules fragiles quant à eux se fragmentent et créent des densifications locales importantes. Les nouvelles particules formées, plus petites, subissent également des déformations réversibles et irréversibles et se fragmentent jusqu'à une taille critique [5,7,8]. À la suite de cette série de déformations et de fragmentation, les surfaces de contact entre les particules augmentent et des liens se forment entre elles. Parmi ces liens, nous pouvons citer le verrouillage mécanique des particules, les forces intermoléculaires (les forces de Van der Waals, les forces électrostatiques et les ponts hydrogènes), les ponts solides et les ponts liquides.

- **Les interactions moléculaires :**

Ce sont des forces attractives qui se développent entre les molécules au sein du milieu granulaire sur de courtes distances grâce à l'augmentation des surfaces de contact [5,32]. Il s'agit entre autres des forces électrostatiques, les forces de Van der Waals et les ponts hydrogènes.

**Les forces électrostatiques :** Elles se développent entre deux particules de charge  $q_1$  et  $q_2$  distantes de  $r$  avec une intensité donnée par la loi de coulomb [7,31].

$$F_{\text{électrostatique}} = \frac{|q_1 \cdot q_2|}{4 \cdot \pi \cdot \epsilon_0 \cdot \epsilon_r \cdot D^2}$$

*Équation 2 : La loi de coulomb*

Où  $\epsilon_0 = 8,854 \cdot 10^{-12} \text{ C}^2\text{J}^{-1}\text{m}^{-1}$ ,  $\epsilon_r = 80$  et  $D$  sont respectivement la permittivité du vide, la permittivité relative de l'eau (rapport entre permittivité absolue de l'eau et une référence, permittivité du vide) et la distance entre les deux charges. Ces forces sont attractives entre des particules de charges opposées et répulsives lorsque les charges sont identiques. Elles sont effectives sur de très courtes distances comme illustrée par la loi de coulomb.

Les **forces de Van der Waals** : ce sont des forces attractives qui agissent à des distances interatomiques jusqu'à 10 nm [8,31]. Elles dépendent des particules en contact par la constante de Hamaker. L'interaction de Van der Waals est la résultante de trois types d'interactions : les forces dipolaires (terme de Keesom), les forces de polarisation (terme Debye) et les forces de dispersion (terme de London). Ces dernières sont les plus importantes qui contribuent à l'interaction totale résultante et elles se développent, quelle que soit la nature des surfaces en interaction [7,8,31]. Dans l'hypothèse de particules sphériques, la force peut être déterminée par la relation suivante :

$$F_{van\ der\ waals} = \frac{A \cdot R}{12 \cdot D^2}$$

*Équation 3: La force de Van der Waals*

Où A, R et D sont la constante de Hamaker, le rayon des particules et la distance de séparation des particules respectivement.

**Les ponts hydrogènes** : ce sont des forces intermoléculaires s'exerçant également sur de courtes distances (0,176 nm) grâce à des atomes d'hydrogène qui s'associent à des atomes électronégatifs comme l'oxygène, l'azote, le fer ou le chlore. Sensibles à la température, ces liaisons ont une intensité intermédiaire entre les forces de Van der Waals et des liaisons covalentes [7,31].

- **Les ponts solides :**

Ce sont des interactions de nature diverse qui ont lieu aux points de contact des particules par fusion partielle, cristallisation, solidification d'éléments en fusion, intervention de liant, ponts de neige, ponts de glace résultants du gel de ponts capillaires, cimentation ou des réactions chimiques comme la formation de sels [5,7,8,10,31,32]. Ce sont des liaisons fortes structurellement et les paramètres influant sur leur intensité sont tout aussi variés que leur nature.

- **Les ponts liquides : forces capillaires**

Lorsque la poudre est humide, des ponts liquides peuvent s'établir dans la structure par des forces capillaires. Le liquide contenu dans la poudre forme un ménisque entre les différentes particules [7,31]. La présence de ces ponts liquides dans un milieu granulaire génère de la cohésion.

### - **Verrouillage mécanique :**

Ce lien s'établit entre des particules de formes irrégulières souvent acquises par déformation plastique et fragmentation. En fonction de leur géométrie, leur surface et leur position, les unes par rapport aux autres, les particules peuvent s'encaster ensemble et maintenir une cohésion sans liens physiques ou chimiques proprement dites [7,8,31].

En fonction des propriétés physicochimiques des particules et les paramètres de compression, notamment l'intensité de la force de compression et sa durée d'application, plusieurs des mécanismes décrits ci-dessus entrent dans la formation de liens de cohésion dans des proportions variées. Le compact est ainsi formé et sa cohésion interne tient en l'intensité de ces liens. Cette formation du compact met fin à la consolidation.

### **2.3.3 La décompression :**

Lors de l'application de la contrainte pour former le compact, les particules emmagasinent de l'énergie et une partie de celle-ci est restituée par retour élastique à l'arrêt de l'application de la contrainte [8,31]. Ce retour est le recouvrement dû à la déformation élastique des particules contenues dans le compact. Il est plus ou moins important en fonction des conditions de compression (force et vitesse de compression), mais aussi des propriétés mécaniques de la poudre en compression. La décompression peut ainsi conduire soit à la rupture des surfaces interparticulaires, donc le clivage du comprimé, soit à une augmentation modérée de l'épaisseur du comprimé [5,31]. Le taux de retour élastique (ER) peut être déterminé à l'aide de la relation suivante :

$$ER = \left( \frac{h - h_p}{h_p} \right) * 100$$

*Équation 4: Taux de retour élastique*

Où  $h_p$  et  $h$  sont l'épaisseur du comprimé sous compression et après éjection respectivement.

### **2.4 L'éjection :**

C'est la dernière étape de l'opération unitaire de compression. De façon concomitante à l'arrêt de la contrainte appliquée, le poinçon du bas remonte avec le comprimé formé jusqu'à la surface de la table de la matrice. Le comprimé est ensuite dégagé par une barre latérale de détachement. Ce processus se fait en même temps que le début de la formation d'un nouveau



comprimé impliquant la descente du poinçon inférieur pour reformer la chambre de compression et l'arrivée du sabot au-dessus de celle-ci pour le remplissage.

Lors de l'éjection, au fur et à mesure que le comprimé émerge de la cavité de compression, il se détend également dans la largeur par recouvrement de déformation élastique jusqu'à un état d'équilibre stable [5,33]. Lorsque la cohésion du compact est faible ou que la détente élastique est très importante, le comprimé nouvellement formé peut se fissurer par rupture des liaisons interparticulaires.

## **2.5 Les machines et outillages de compression :**

### **2.5.1 Les machines de compression : presses**

Une machine de compression est un appareil électromécanique qui permet de fabriquer des comprimés à partir d'une poudre de formulation [34]. Depuis près de 2 siècles où la première structure de presse fut brevetée par William Brockedon, leur principe de fonctionnement n'a pas changé. C'est toujours une contrainte appliquée à un lit de poudre dans une matrice qui conduit à la formation de comprimés. Toutefois, de nos jours, les presses se sont largement modernisées et sont désormais plus performantes, instrumentées et contrôlées par des ordinateurs pour une meilleure productivité, un meilleur contrôle et suivi du procédé [5,7,31]. On distingue principalement trois types de machines de compression : les presses excentriques qui sont dotées d'une seule station de compression (presses à station unique), les presses rotatives dotées de plusieurs stations de compression (presses multistation) et la dernière révolution, les presses de simulation qui permettent de mimer les profils de compression de plusieurs types d'appareil.

#### **- Les presses à station unique :**

Ce sont des machines de compression ayant une seule station de compression statique qui est composée d'un poinçon supérieur, d'une matrice et d'un poinçon inférieur. La force de compression est généralement appliquée par le poinçon supérieur uniquement et le poinçon inférieur demeure statique lors de la compression. À cause de cette application de contrainte unilatérale, les comprimés produits présentent généralement une densification inhomogène [5,8]. La limite de ces presses réside aussi dans leur capacité de production limitée. Celle-ci est déterminée par la vitesse de l'unique station de compression. Par exemple, une presse à station unique Manesty F3 ne peut produire qu'environ 3500 à 5000 comprimés par heure.

#### **- Les presses rotatives :**

Ce sont des machines de compression ayant plusieurs stations de compression dynamiques qui tournent pour comprimer la poudre. Les matrices sont enchâssées dans la table centrale rotative appelée « tourette ». Les nombreux sets de poinçons (haut et bas) sont fixés sur la périphérie autour de cette table et lorsque celle-ci tourne, les poinçons sont entraînés dans des mouvements verticaux à l'aide de cames et de rouleaux de précompression et de compression. Les presses rotatives se caractérisent par l'application de contraintes biaxiales, assurant une meilleure homogénéité de la densification des comprimés. Ces presses ont été développées spécialement pour augmenter les capacités de production. Le nombre de comprimés dépend du nombre de stations et de la vitesse de la tourette. Lorsque celle-ci fait seulement une soixantaine de tours par minute avec vingt stations, la presse produit plus de 70 000 comprimés à l'heure.

- **Les simulateurs de compression :**

Ce sont des machines possédant une seule station de compression, mais elles sont spécialement instrumentées et outillées pour simuler le profil de compression (types et profils de poinçons et vitesse de compression) de nombreux autres types de presse [5]. De petites quantités de poudre sont utilisées sur ces presses pour réaliser des études de développement de formulation ou pour tester l'effet de plusieurs paramètres de procédés, notamment les paramètres de mise à l'échelle.

### **2.5.2 Les outillages de compression : poinçons et matrices**

L'outillage utilisé sur les presses pour produire des comprimés comprend les poinçons et la matrice. Ces outils sont généralement faits en acier. Les fabricants incorporent à leur constitution différents autres éléments d'alliage en faible proportion pour diverses raisons, notamment pour améliorer leur dureté, leur résistance à la corrosion et leur stabilité à haute et basse température [4]. On peut citer parmi ceux-ci : le carbone, qui est un élément permettant d'augmenter la dureté et la résistance mécanique de l'acier ou le chrome qui améliore en plus de la dureté, la résistance à l'usure, l'abrasion et la corrosion ou encore le tungstène qui est utilisé pour maintenir la dureté à des températures élevées [4].

De ces alliages, les fabricants d'outillage désignent et prototypent différents types de set de poinçons et de matrices. La Table 1 ci-dessous synthétise les caractéristiques des principaux types d'outillage disponibles.

Table 1: caractéristiques des principaux types d'outillage de compression.

| Type d'outillage | Diamètre du poinçon | Diamètre de la matrice | Longueur du poinçon |
|------------------|---------------------|------------------------|---------------------|
| B                | 19 mm               | 30,15 mm               | 133,6 mm            |
| D                | 25,4 mm             | 38,1 mm                | 133,6 mm            |
| BB               | 19 mm               | 24 mm                  | 133,6 mm            |
| DB               | 25,4 mm             | 30,15 mm               | 133,6 mm            |

## 2.6 Problèmes survenant au cours de la compression :

Lors de la compression des poudres pharmaceutiques en comprimés, différents problèmes peuvent survenir et donner lieu à la production de comprimés non conformes qui ne pourront constituer que des pertes. Les plus communs sont l'hétérogénéité du poids des comprimés, une faible cohésion interne, le laminage et le collage.

L'un des attributs qualité des comprimés est l'uniformité de leur contenu. Cela permet de garantir l'administration de doses conformes. Comme décrit plus haut, c'est le volume de la cavité de compression déterminée par le volume de la matrice et la position du poinçon inférieur qui définit la quantité unitaire de poudre à comprimer. À densité fixe, le poids est alors un indicateur direct de l'uniformité du contenu des comprimés. De grosses variations de poids impliquent alors un problème d'uniformité du contenu. Plusieurs facteurs peuvent contribuer à ce problème, notamment les propriétés physiques et d'écoulement des poudres, mais aussi les conditions d'opération de la presse [4,5].

Les comprimés peuvent être sujets également à une insuffisance de dureté, c'est-à-dire une faible cohésion interne. Ce paramètre est important pour la tenue mécanique du comprimé afin que ce dernier puisse résister aux différents stress subis par la suite, notamment lors du conditionnement, l'expédition, la distribution et même l'utilisation quotidienne. Plusieurs paramètres affectent la dureté des comprimés, notamment la force de compression, la lubrification, le temps de contact, et les propriétés physiques du compact formé comme son épaisseur [1,4,5].

Après la formation du compact, des phénomènes de clivage ou de délaminage peuvent aussi se produire. Les comprimés résultants se trouvent briser dans leur épaisseur. Le clivage

ou le délaminage sont généralement les conséquences d'un important retour élastique [4]. Ils dépendent donc principalement des propriétés de la poudre ainsi que les pressions de compression.

L'ensemble de ces problèmes et beaucoup d'autres représentent des challenges scientifiques et techniques pour l'industrie pharmaceutique, car ils limitent leur productivité et leur compétitivité. En plus de ceux présentés dans cette section, l'un des problèmes les plus fréquemment rencontrés également au cours de la compression est le collage de poudres pharmaceutiques. C'est-à-dire l'adhésion de la poudre à comprimer sur les surfaces des outillages. Ce phénomène est le sujet d'étude principal de cette thèse de doctorat et fait l'objet de discussions plus détaillées dans les sections et chapitres suivants.

### **3. Le collage de poudres pharmaceutiques :**

Le collage est le phénomène décrit par l'adhésion de poudres sur la surface des outillages de compression, c'est-à-dire la surface des poinçons (haut et bas) et les parois de la matrice. Pour tenter de le décrire précisément, plusieurs définitions existent dans la littérature, car il en existe plusieurs variantes. Pour résumer, on distingue trois variantes principales de collage : l'adhésion de la poudre dans les inscriptions gravées sur la surface des poinçons (déclinée en anglais par « picking »), la couverture de la face des poinçons par un film de poudre (appelée « filming » en anglais) et la couverture de la surface des poinçons ou des matrices de plusieurs couches de poudre (dénommée « sticking » en anglais) [5,8,9]. Dans ce projet, nous référerons toujours au terme unique de collage pour désigner l'ensemble de toutes ces variantes. Ainsi, le terme de collage décrit tout phénomène d'adhésion de poudre sur les surfaces des poinçons et/ou des matrices, l'adhésion des poudres à la surface des poinçons étant la plus reportée et la plus étudiée.

Le collage survient très peu durant les phases de développement de la formulation lorsque peu de comprimés sont produits. Il n'apparaît généralement que dans les phases industrielles de production où les possibilités de modifications et d'investigations sont très limitées [9,14,29,35,36]. À ces échelles, l'occurrence du collage porte de sérieux préjudices à la production : (i) il rallonge les coûts et délais de production, (ii) il impacte la qualité des comprimés produits et (iii) il induit une usure prématurément sur les presses et les outillages.

En effet, lorsque le collage survient, la production est arrêtée pour nettoyer, polir et/ou changer les poinçons avant de la redémarrer. Les comprimés en ressortent avec un aspect dégradé, des surfaces rugueuses, des gravures incomplètes et des variations de poids qui

conduisent tous à leur rejet et leur destruction [9,29]. L'occurrence du collage peut également conduire à l'usure prématurée des poinçons à travers les nombreuses opérations de nettoyage et de polissage, mais aussi, il peut conduire à la démolition de la barre de dégagement des comprimés de la surface des poinçons à cause des plus grandes forces d'éjection requises [5,7,8]. Eu égard de ces nombreuses conséquences sur la production et la compétitivité industrielle, plusieurs initiatives ont été lancées pour tenter d'élucider le problème du collage. De grands progrès ont été accomplis, mais à ce jour, aucune cause racine permettant une compréhension globale n'a encore été identifiée [9,10,26,29]. Cela est dû à l'aspect multifactoriel très prononcé du phénomène, mais aussi l'indisponibilité pendant longtemps de techniques d'évaluation pertinentes pour le mesurer.

#### **4. Les techniques d'évaluation du collage :**

Pour étudier un phénomène, il faut tout d'abord pouvoir le quantifier de façon précise. L'un des plus grands freins à l'émergence de connaissances sur le collage fut, pendant longtemps, l'indisponibilité de techniques de mesures fiables et précises. Cela est principalement dû au confinement des outillages sur une presse en opération, mais aussi le fait que le collage n'apparaît généralement qu'à l'échelle industrielle où les arrêts pour réaliser des mesures sont prohibés. Dans cette section, nous présentons de façon succincte les différentes techniques d'évaluation du collage. Cela permet d'avoir une perspective sur la difficulté d'une synthèse bibliographique, mais surtout de voir comment cette variété de techniques et leur caractère indirect pour beaucoup, limitaient les capacités de compréhension globale du collage.

##### **4.1 Les méthodes indirectes d'évaluation du collage:**

Pendant très longtemps, l'étude du collage n'impliquait pour la plupart que des méthodes d'évaluation indirecte. Elles se fondent principalement sur l'exploitation des aspects théoriques de l'adhésion ou de l'opération de compression. Parmi ces techniques, on peut citer les plus communément employés comme : (1) la microscopie à force atomique (AFM) qui permet de mesurer la force d'adhésion en différentes surfaces, (2) la force de décollement du comprimé de la surface du poinçon supérieur (POF pour « pull-off force » en anglais), (3) la force de détachement du comprimé de la surface du poinçon inférieur (TOF pour « take-force » ou aussi référée comme SCR pour « scraper pressure » en anglais), (4) la force d'éjection et (5) les mesures de la rugosité de la surface des poinçons ou des comprimés.

Ces méthodes indirectes d'évaluation du collage accusent de plusieurs lacunes qui pourraient mitiger les conclusions découlant de leur emploi. Leur nature indirecte est la

principale lacune. Par exemple pour l'AFM, une forte adhésion n'implique pas nécessairement du collage et ses conditions d'opération sont très différentes des conditions réelles d'une compression de poudre [5,7]. De plus, de récentes publications démontrent même l'inefficacité de certaines méthodes indirectes à fidèlement transcrire le phénomène du collage. Par exemple, la TOF ou SCR fut employé pendant des années par de nombreuses équipes de recherche pour étudier le collage, mais les récentes données de la littérature étaye ses limites, voir son inefficacité [37].

L'évaluation du collage avec la TOF consiste à mesurer la force nécessaire à la barre de dégagement des presses pour dégager le comprimé nouvellement formé de la surface du poinçon inférieur. Le postulat de cette technique est que la TOF représenterait la force d'adhésion entre la surface du comprimé et la surface du poinçon. Saniocki et collab. (2013) ont évalué la pertinence de la TOF pour la mesure du collage de formulations à base d'ibuprofène (IBU) [37]. Les données obtenues avec la TOF furent comparées dans les mêmes conditions à celles fournies par une méthode directe où la quantité de poudre collée sur la surface des poinçons est dosée par Chromatographie Liquide à Haute Performance (HPLC). Les résultats obtenus montrent que le collage mesuré par dosage HPLC augmente proportionnellement avec la proportion d'IBU tandis que les données de TOF restent au même niveau, quelle que soit la proportion d'API [37]. De plus, en utilisant seulement les matrices d'excipient, la TOF mesurée est tantôt plus élevée ou plus faible que celle mesurée pour les formulations correspondantes susceptibles de coller davantage en raison de la présence d'API.

Saniocki et collab. (2013) ont en outre introduit une approche différenciée pour l'interprétation des données de TOF [37]. Cette approche est basée sur l'existence ou non du collage. D'une part, lorsque la force d'adhésion entre la surface du poinçon inférieur et la surface du comprimé est plus élevée que la force de cohésion dans le comprimé, c'est-à-dire qu'en présence de collage, la force nécessaire pour déplacer le comprimé de la face du poinçon, la TOF, représente plus une mesure directe de la force de cohésion entre les couches de particules de comprimé (celle collée sur la surface des poinçons et les particules de surface du comprimé). D'autre part, lorsque la force d'adhésion du comprimé à la surface du poinçon inférieur est plus faible que la cohésion dans le comprimé, c'est-à-dire qu'en l'absence de collage, la TOF mesure alors la force d'adhésion entre les surfaces du poinçon et du comprimé. Cette interprétation différenciée permet de comprendre les liens entre le collage et la TOF, mais elle montre surtout son impertinence pour une évaluation systématique du collage lors d'expérimentations.

## **4.2 Les méthodes directes d'évaluation du collage :**

Les méthodes directes sont basées sur l'évaluation visuelle ou la quantification de la masse de poudre adhérente sur la surface des poinçons. On distingue : (1) l'évaluation visuelle de la surface des poinçons, (2) la pesée de la poudre collée sur la surface de poinçons spéciaux à bout détachable, (3) la quantification de la poudre collée sur la surface des poinçons par HPLC et (4) les techniques d'imagerie. Comparativement aux méthodes indirectes, toutes ces méthodes directes permettent une évaluation plus précise et fiable du collage.

De nos jours, la pesée et la quantification par HPLC de la poudre collée sur la surface des poinçons sont les méthodes les plus employées. La quantification du collage par pesée du poinçon a été utilisée pour la première fois par Sendall et Satniforth en 1986 [38]. Depuis, Mullarney et collab. (2012) l'ont optimisé en introduisant des poinçons à bout détachable où c'est le bout des poinçons qui est désormais pesé au lieu du poinçon en entier [39]. Ceci a permis d'améliorer considérablement la sensibilité de la méthode de quantification par pesée. L'HPLC a aussi été utilisée dans de nombreuses études pour quantifier précisément la poudre collée sur la surface des poinçons en utilisant des spectromètres UV-Visible ou des spectromètres de masse [11,12,35,40]. La principale limite de l'utilisation de ces méthodes pour étudier le collage réside dans la nécessité d'arrêt de la compression pour procéder aux mesures. De plus, elles sont destructives et nécessitent beaucoup de temps pour les manipulations de démontage et de remontage des outillages.

Des méthodes d'évaluation par imagerie commencent aussi à être introduites pour visualiser et quantifier le collage [29,35,41–43]. Parmi ceux-ci, l'une des plus employées à date est la microscopie électronique à balayage (SEM), qui couplée à l'analyse d'image 3D (SEM/3D) permet de quantifier de façon qualitative le collage. Lorsque la SEM est couplée à la spectroscopie de rayons X à dispersion d'énergie (SEM/EDS), il permet d'identifier les constituants chimiques de la poudre collée sur la surface des poinçons [35]. Al-Karawi et collab. (2017) ont aussi introduit une méthode d'étude du collage par imagerie consistant à réaliser un moulage haute résolution de la surface des poinçons suivie de mesures 3D [42]. Un total de 58 paramètres décrivant la surface des poinçons sont ainsi mesurés et traités par ACP. En utilisant cette approche, les auteurs ont étudié la tendance au collage de différentes textures de surface de poinçon et les résultats conduisent aux mêmes conclusions qu'une quantification par HPLC de la poudre collée sur la surface des poinçons [42]. Toutefois, la méthode est rapportée moins discriminante et à cause de la sensibilité des matériaux d'impression, elle est inutilisable pour des poudres contenant des composants acides. Dans l'ensemble, ces

techniques d'évaluation par imagerie sont aussi intéressantes, mais elles sont hors lignes, chronophages, et souvent qualitatives.

Plus récemment, Thomas et collab. (2020) ont proposé une nouvelle technique d'évaluation du collage qui ne nécessite pas l'arrêt de la compression [43]. Cette mesure en ligne est basée sur la réflexion d'un faisceau laser par la surface du poinçon durant la courte période entre chaque opération unitaire de compression. La méthode consiste en l'utilisation d'une source laser et d'une photodiode fixée sur le sabot d'alimentation de la presse. La source laser et la photodiode sont toutes placées à angle égal ( $41,5^\circ$ ) relativement à la surface du poinçon de telle sorte qu'après chaque opération de compression, lorsque le sabot passe pour remplir la matrice pour la compression suivante, la source laser illumine la surface du poinçon du haut et la lumière réfléchiée par celle-ci est captée par la photodiode [43]. Les résultats reportés par Thomas et collab. (2020) pour une première application de cette technique montrent que l'intensité de la réflexion corrèle bien avec la couverture de la surface des poinçons. Plus celle-ci est couverte, plus l'intensité de la réflexion diminue. Cette technique est alors très prometteuse dans l'étude du collage, car elle est simple, en ligne, rapide et non destructive. Elle pourrait toutefois présenter des limites pour des phénomènes de collage intenses, où la diminution de l'intensité de la réflexion saturerait après une couverture complète de la surface du poinçon.

Eu égard de la variabilité impressionnante des techniques d'évaluation employées pour étudier le collage, la revue de littérature sur les causes du collage présente quelques spécificités. Nous avons notamment tenu à identifier pour chaque référence, les techniques employées, mais surtout une attention particulière a été portée au croisement des données reportées afin de ne soutenir et analyser que des résultats concordants provenant de plusieurs auteurs avec différentes méthodes d'évaluation du collage.

## **5. Les causes du collage de poudres pharmaceutiques :**

### **5.1 Le mécanisme conceptuel du collage :**

De façon conceptuelle, le collage fut décrit pendant longtemps comme le résultat du rapport d'intensité entre deux groupes de force : la somme des forces d'interaction contribuant à la cohésion des comprimés (interactions particules-particules au sein de la poudre) et la somme des forces d'interaction contribuant à l'adhésion des particules sur la surface des outillages de compression. Suivant ce modèle conceptuel (modèle 1), les particules se détachent de la surface du lit de poudre et adhèrent à la surface des outillages lorsque l'intensité des forces d'adhésion



des particules à la surface des outillages dépasse l'intensité de l'adhésion de ces mêmes particules aux autres particules constitutives de la poudre (cohésion).

Ce premier modèle conceptuel est incapable d'expliquer deux observations claires corroborées par toute la littérature sur le collage : l'accumulation de plusieurs couches de poudre sur le poinçon et surtout l'enrichissement de celles-ci en API. En effet, suivant le modèle conceptuel 1, le collage cesserait dès que la surface des poinçons sera entièrement recouverte d'une première couche de poudre, puisqu'au cours des compressions subséquentes, la surface des poinçons ne sera plus exposée à la poudre. Cela est contraire aux données de la littérature qui rapportent une accumulation de plusieurs couches de poudre à la surface des poinçons pour de nombreuses formulations [9,14–16,29,36,44–46]. De plus, le modèle conceptuel 1 présume que la poudre collée sur la surface des poinçons sera de même composition que la formulation, ce qui est également contraire à la quasi-totalité des données de la littérature. Face à ces insuffisances, Paul et collab. (2017) ont introduit un nouveau modèle conceptuel (modèle 2) plus complexe et plus complet basé sur le rapport d'intensité de trois forces au lieu de deux forces (voir Figure 4) [46].

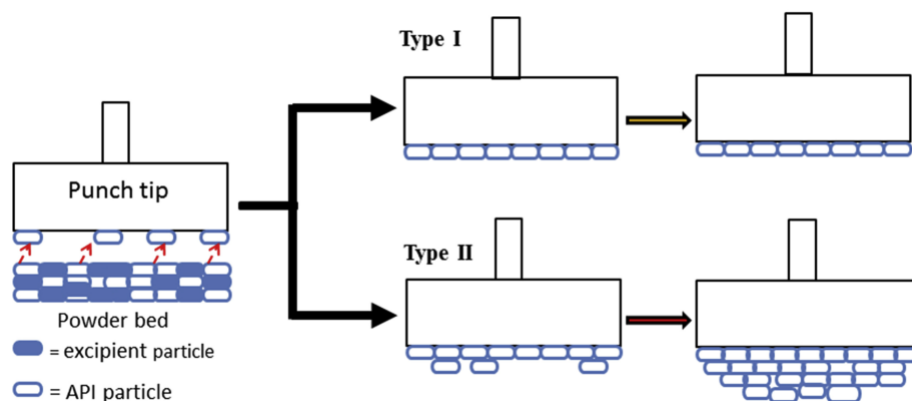


Figure 4: Modèle conceptuel du collage proposé par Paul et collab. (2017) [46].

Suivant ce deuxième modèle conceptuel, il existe deux types de collage : un collage modéré de type I où une seule couche de particules adhère à la surface du poinçon et un collage sévère de type II, plus problématique, où plusieurs couches de particules s'accumulent à la surface des poinçons [46]. Ces deux types de collage surviennent selon le rapport d'intensité de trois forces : les forces d'adhésion entre les particules d'API et la surface des poinçons ( $F_1$ , API-Poinçon), les forces de cohésion des particules d'API dans la matrice d'excipients ( $F_2$ , API-Excipients) et les forces de cohésion entre particules d'API ( $F_3$ , API-API). Le collage modéré de type I survient lorsque  $F_1 > (F_2 \text{ et } F_3)$  où l'adhésion des particules d'API au poinçon est prédominante [46]. Lorsque la surface des poinçons est complètement recouverte d'une

première couche de particules d'API, l'aggravation subséquente du collage est gouvernée par le rapport d'intensité entre F2 et F3. La prédominance de la cohésion API-Excipients (F2) permet de confiner le collage au type I tandis qu'une plus grande cohésion API-API (F3) conduit à l'occurrence du collage sévère de type II [46]. Dans ce dernier cas, l'affinité des particules API-API favorise leur accumulation sur la surface du poinçon. Ce modèle conceptuel proposé par Paul et collab. (2017) est le plus abouti pour expliquer le phénomène du collage dans son ensemble. Elle cadre parfaitement avec la quasi-totalité des données de la littérature, mais surtout elle est assez flexible pour prendre en compte le caractère multivarié prononcé du phénomène. Les mêmes types d'interaction régissent les trois forces gouvernant l'occurrence du collage (F1, F2 et F3). Il s'agit principalement de forces de Van der Waals, de forces électrostatiques, de liaisons covalentes, de ponts hydrogènes, de forces capillaires, de ponts solides ou encore de verrouillages mécaniques (voir description à la section 2.3). En fonction des propriétés des particules et des conditions de la compression, l'ensemble de ces interactions varie constamment en nombre et en intensité au cours de la compression. Le collage est donc la résultante d'un équilibre dynamique entre des forces d'adhésion et de cohésion.

Cet équilibre dynamique conduisant au collage est également très fragile, car elle est sensible à un nombre très important de paramètres, mais surtout leurs multiples interactions. Les données de la littérature établissent clairement le lien entre le collage et plusieurs des propriétés des particules d'API, des formulations, de la surface des outillages, ainsi que des paramètres opératoires et environnementaux des procédés de production des comprimés. D'une part, nous pouvons distinguer les paramètres dont l'augmentation est reportée avoir un effet bénéfique sur le collage comme la taille des particules, la force de cohésion intrinsèque API-Excipient, la lubrification, le temps de contact, la force de compression, la forme de la tête des poinçons et l'hydrophobicité de leur surface. D'autre part, l'augmentation de nombreux autres paramètres est reportée avoir un effet délétère sur le collage. Il s'agit entre autres de la proportion d'API, la cohésion intrinsèque API-API, les propriétés physicochimiques et électrostatiques des API, l'humidité relative, les propriétés adhésives, l'hétérogénéité de la surface des poinçons ainsi que la vitesse et la durée des séances de compression. De nombreuses interactions entre tous ces paramètres furent également reportées. Dans les sections suivantes, nous décrivons de façon détaillée l'implication de chacune de ces composantes dans l'occurrence du collage. Dans cet exercice, une attention particulière fut portée à bien faire ressortir les caractères spécifiques du collage qui limitent son étude, sa compréhension et sa résolution de façon expérimentale à échelle laboratoire.

## **5.2 Les paramètres relatifs à la formulation :**

### **5.2.1 L'ingrédient actif de la formulation :**

Les résultats de la littérature établissent que la majeure partie de la fraction de poudre collée sur la surface des poinçons est constituée d'API et que le collage est proportionnel à sa proportion dans la formulation [21,27,35,37,44,46,47].

Neilly et collab. (2009) ont étudié le collage à l'aide de la SEM et la EDS [47]. En comparant les images de compacts formés à ceux de la fraction de poudre collée sur la face des poinçons, ils ont montré que cette dernière était majoritairement constituée de l'API. McDermott et collab. (2010) ont mené des études semblables et sont arrivés à la même conclusion avec deux formulations différentes, notamment l'ABT089 et l'ABT279 [35].

Saniocki et collab. (2013) ont investigué l'effet de la proportion d'API sur le collage de formulations à base d'IBU à différentes proportions avec différentes matrices d'excipient [37]. Pour la quantification du collage, la poudre collée sur la surface des poinçons après des cycles de 25 compressions fut dissoute dans du méthanol et dosée par HPLC. Les résultats obtenus démontrent que, quelles que soient la force de compression et la matrice d'excipient, la quantité de poudre collée sur la surface des poinçons augmente proportionnellement avec la proportion d'API dans la formulation [37].

En 2016, des chercheurs de l'université du Minnesota et des centres de recherche du groupe pharmaceutique Pfizer ont étudié le collage de 24 formulations différentes en API [46]. Dans cette étude, le collage fut quantifié différemment par pesée du bout détachable des poinçons supérieurs après chaque 0, 25, 50, 75 et 100 cycles de compression pour chacune des 24 formulations différentes. Les résultats montrent que la propension au collage est API spécifique et pour les formulations les plus collantes contenant seulement 10 % d'API, la masse collée sur les poinçons est constituée d'environ 93 à 96 % d'API [46].

Plus récemment en 2017, des conclusions similaires ont aussi été rapportées par Badal et collab. [21]. Les auteurs ont étudié différemment le phénomène d'adhésion entre les particules de poudre et la surface des poinçons en utilisant une sonde colloïdale en acier montée sur un microscope à force atomique. Cette technique permet de mesurer les forces d'interaction entre différentes surfaces. Les auteurs ont ainsi évalué l'adhésion de particules d'IBU, de MCC et de lactose à une surface métallique en acier. Les résultats montrent qu'il existe une faible force d'adhésion de type Van der Waals entre les particules d'excipients (MCC et lactose) et

la surface métallique d'acier tandis que les particules d'API (IBU) montrent une plus grande force d'adhésion à la surface métallique d'acier grâce à des interactions capillaires [21]. L'apparition et l'amplitude de ces forces capillaires sont dues à une condensation rapide de l'humidité résiduelle ou ambiante permettant la formation de ponts aqueux entre les surfaces [21].

L'ensemble des études ci-dessus impliquant différents API et techniques d'évaluation du collage (imagerie, quantification par HPLC, par pesée ou mesure de forces d'adhésion par AFM) confortent l'hypothèse que ce sont les particules d'API qui adhèrent le plus sur la surface des poinçons et que le collage est proportionnel à la proportion de celui-ci dans la formulation. Cette conclusion ne peut pourtant pas être mise à profit pour énoncer des solutions au collage, car la proportion d'API dans la formulation n'est pas modulable. Elle est essentielle à l'effet thérapeutique du médicament. Mais bien au-delà, cette spécificité du collage aux particules d'API rajoute même plus de complexité dans l'étude et la compréhension du collage. En effet, de nombreux travaux démontrent clairement que le mécanisme même du collage est API dépendant [10,15,46]. Il existerait ainsi autant de mécanismes conduisant au collage que d'API, de telle sorte que des études impliquant différents API sont difficilement comparables et extrapolables. Cet état de fait résulte de l'existence d'effets d'interaction entre l'API et de nombreux autres paramètres impliqués dans le collage. La littérature permet notamment de montrer que l'effet de la rugosité ou du revêtement de la surface des poinçons [15,19,22] ainsi que celui de la force et de la durée de compression [13,15,16,42] change substantiellement d'un API à l'autre. Ces nombreuses interactions et le positionnement même des API au centre du phénomène du collage sont spécifiques à leurs propriétés physiques, chimiques ou électrostatiques.

### **5.2.2 Les propriétés physicochimiques des particules:**

Les principales propriétés physicochimiques des particules étudiées pour leur rôle dans le collage de poudres pharmaceutiques sont l'énergie de surface, le comportement mécanique des particules, la cohésion API-API, la surface spécifique de contact et la taille des particules.

#### **- L'énergie de surface des particules :**

Hooper et collab. (2017) ont étudié l'impact de la forme des cristaux d'IBU sur leur propension au collage [48]. Quatre formes cristallines d'IBU ont été générées par recristallisation dans différents solvants (hexane, toluène, acétonitrile et éthanol). La propension au collage fut évaluée sur une presse Gamlen dotée d'une jauge qui permet de mesurer la force de

détachement de la base mobile de la matrice de compression. Les résultats obtenus dans cette étude ont montré que la propension au collage de l'IBU changeait avec sa forme cristalline, où la forme la plus régulière obtenue par recristallisation dans l'éthanol montrait la plus grande propension au collage, mais aussi l'énergie de surface la plus élevée [48]. L'étude de la surface des particules a permis de déterminer que le changement de forme cristalline s'accompagne d'un changement dans l'énergie de surface spécifique dû à une présence plus ou moins importante de faces cristallines contenant des groupements fonctionnels polaires, spécifiquement d'acide carboxylique [48]. L'ordre de grandeur des énergies de surface des quatre formes cristallines correspond à l'ordre de grandeur de leur propension au collage, démontrant ainsi un lien de proportionnalité entre l'énergie de surface spécifique modifiée par la forme cristalline et la propension au collage.

Des conclusions similaires sur l'effet de l'énergie de surface ont aussi été rapportées dans plusieurs autres travaux, notamment ceux de Al-Karawi et collab. (2018) sur la propension au collage de l'IBU et l'ibuprofène sodium (IBU-Na) [13], ceux de Paul et collab. (2019) sur l'acide libre d'acésulfame et son sel de potassium [25], mais aussi les travaux de Wang et collab. (2020) sur le célécoxib (CEL) et son cocrystal de L-proline [49]. Ces deux dernières études ont aussi investigué le rôle des propriétés mécaniques et cohésives des particules d'API dans le collage.

- **La plasticité et la cohésion API-API :**

Paul et collab. (2019) ont investigué de façon approfondie les comportements de collage de la forme libre d'acésulfame (Ace-H) et sa forme de sel de potassium (Ace-K) [25]. Le collage fut évalué par pesée du bout détachable des poinçons et les auteurs ont démontré que l'exposition de groupements fonctionnels polaires et électronégatifs à la surface de l'Ace-H contribuait à un début plus rapide de son collage aux poinçons grâce notamment à des interactions électrostatiques. De plus, ils ont montré que la plus grande plasticité et la forte force de cohésion intrinsèque de l'Ace-H par rapport à sa forme de sel de potassium induisent également pour celui-ci un collage sévère de type II suivant le modèle conceptuel du collage décrit plus haut [25]. La grande plasticité de l'Ace-H permet une plus grande capacité de déformation des particules conduisant à l'augmentation de la surface de contact entre les particules d'API déjà collées sur le poinçon et les particules d'API de comprimés subséquents. La plus grande cohésion intrinsèque permet aussi aux particules d'API de comprimés subséquents d'avoir plus d'affinité pour la couche d'API déjà collée sur les poinçons que la matrice d'excipients.

Les travaux de Wang et collab. (2020) ont aussi livré des conclusions similaires sur le rôle de l'énergie de surface et des propriétés mécaniques des particules [49]. Les auteurs avaient étudié la capacité d'un cocrystal du célécoxib identifié par approche computationnelle à réduire sa propension au collage grâce à des propriétés mécaniques différentes. Les résultats rapportés ont montré que le produit de cocrystallisation permet de réduire substantiellement la propension au collage du CEL. La quantité de poudre du cocrystal collée sur la surface des poinçons après 50 cycles de compression ne représentait environ que 10 % de la quantité de poudre de CEL collée sur la surface des poinçons [49]. Des analytiques approfondis ont permis aux auteurs d'attribuer la performance du cocrystal à sa faible plasticité et l'exposition minimale de groupes fonctionnels électronégatifs à sa surface.

- **La taille des particules :**

De façon générale, il fut admis pendant longtemps que les particules de petite taille favorisaient le collage de poudres pharmaceutiques, notamment à cause de leurs propriétés électrostatiques et l'importance des surfaces d'échange spécifiques. Bien que cette affirmation demeure vraie, les récents travaux de Paul et collab. (2020) y ont apporté plus de nuances et permettent une meilleure compréhension de la complexité de l'effet de la taille des particules sur le collage [50]. Les auteurs ont investigué le rôle de la taille des particules dans le comportement de collage des sept API différents qui sont : l'IBU, le CEL, la caféine (CAF), l'acide méfénamique (MA), le glyburide (GLY), le méthyl parabène (MP), et le phosphate dicalcique dihydraté (DCPD) [50]. Différentes tailles de particules de ces sept API furent générées par recristallisation ou par broyage et le collage fut évalué par pesée du bout détachable du poinçon supérieur après plusieurs cycles de compression sous 100 et 200 MPa. Les résultats de cette étude montrent que la réduction de la taille des particules conduit à une augmentation du collage avec certains API (CEL, GLY et MA) tandis qu'elle conduit à sa diminution avec les autres API (CAF, DCPD, IBU et MP) [50]. Des analyses détaillées au niveau moléculaire ont permis de montrer que la réduction de la taille des particules s'accompagne de plusieurs autres changements qui permettent d'expliquer l'effet contradictoire observé. Il s'agit notamment de l'énergie de surface, la surface de contact spécifique, la plasticité et la force de cohésion API-API des particules [50]. La réduction de la taille des particules contribue à l'initiation du collage selon son impact sur l'énergie de surface spécifique résultant de l'exposition plus ou moins importante de nouvelles faces de cristal plus ou moins polaires susceptibles à des interactions électrostatiques. L'augmentation de l'exposition de groupements fonctionnels électronégatifs a ainsi favorisé une initiation plus rapide du collage du CEL tandis que l'inverse a ralenti

l'initiation du collage de L'IBU [50]. L'exacerbation du collage (type II) à la suite de la réduction de la taille des particules est quant à elle, dictée par son impact sur la force de cohésion API-API, résultante des changements de comportements mécaniques et de forme cristalline. Une forte cohésion API-API résultant de la réduction de taille des particules conduit ainsi à une aggravation du collage tandis que dans le cas inverse, la réduction de la taille des particules renforce davantage la cohésion API-excipient et contribue à la diminution du collage comme reporté pour l'IBU dans l'étude [50].

En définitive, l'état actuel de la littérature permet d'avoir une très bonne compréhension du rôle des propriétés physicochimiques des API dans leur propension au collage. Celle-ci cadre bien avec le modèle conceptuel proposé par Paul et collab. [46]. Nous pouvons ainsi affirmer sur la base de la littérature que l'initiation du collage est dictée en partie par l'exposition de groupements fonctionnels électronégatifs sur les faces des cristaux de particules d'API, conditionnant leur énergie de surface. Le collage multicouche de type II est quant à lui dépendant de la cohésion API-API et le comportement mécanique des particules, notamment une forte plasticité qui augmente les surfaces de contact API-punch, mais aussi API-API. L'effet de la taille des particules n'est que la résultante des bouleversements qu'elle induit sur ces autres propriétés physicochimiques.

Cette compréhension de l'effet des propriétés physicochimiques des API ajoute aussi un cran de complexité dans l'étude et la compréhension du collage de façon globale. En effet, les études présentées dans cette section montrent toutes que l'équilibre des propriétés physicochimiques est dynamique et très sensible. D'une forme cristalline à l'autre, d'une taille de particules à l'autre, ou d'une condition opératoire à l'autre, l'effet des paramètres discutés ci-dessus peut changer rapidement et substantiellement. Ainsi de la section précédente où le collage était décrit comme API dépendant, cette section introduit la notion de collage qui serait plus propriétés ou grade d'API dépendant. Ce ne sont plus seulement des études réalisées avec des API différents qui sont difficilement comparables, mais même des études utilisant des API identiques peuvent être difficilement comparables lorsque des différences fondamentales physiques ou chimiques existent ou sont induites par le traitement des poudres avant la compression. De plus, les différentes propriétés physicochimiques discutées ici interviennent dans le collage à travers des mécanismes purement d'interaction avec d'autres paramètres impliqués aussi dans le collage, notamment les caractéristiques de la surface des poinçons (revêtement, rugosité et géométrie), mais aussi les paramètres de la compression comme la force ou la vitesse de compression.

Sur le plan solution, le rôle des propriétés physicochimiques des API pourrait être mis à profit via l'ingénierie des cristaux de particules pour prévenir le collage. De nombreux travaux sont actuellement conduits dans cette direction avec des résultats encourageants, notamment tous ceux présentés dans cette section [22,25,48–50]. On pourrait aussi ajouter les travaux de Chen et collab. (2020) qui, similairement à ceux de Wang et collab. (2020), ont démontré que la cristallisation sphérique du célécoxib assistée par un polymère comme l'hydroxy-propyl-méthyl-cellulose (HPMC) permettait de réduire sa propension au collage grâce à une fine couche de HPMC qui vient prévenir physiquement les interactions avec la surface des poinçons [45]. Néanmoins, il est important de se rappeler que les propriétés physicochimiques des particules d'API sont essentiellement aussi les éléments constitutifs de leur effet thérapeutique. Ainsi, la fenêtre d'action des solutions d'ingénierie des cristaux est limitée de façon pratique, car elle doit concilier avec l'impératif de l'effet thérapeutique.

### **5.2.3 Les propriétés électrostatiques des particules :**

Lors de la production de comprimés, les particules de poudre entrent en contact les uns avec les autres ainsi qu'avec les différentes surfaces des équipements de production. Ces contacts engendrent une charge électrique à leur surface par un processus connu sous le nom de triboélectrification [51]. Phénomène très répandu, la triboélectrification est le transfert d'électrons qui a lieu entre les surfaces de deux matériaux de nature différente lorsqu'ils sont mis en contact. L'amplitude du transfert est fonction de la nature des deux matériaux, leur surface de contact et l'apport d'énergie mécanique lors de leur frottement [5,7]. La durée de la charge sur la surface dépend du temps de relaxation qui est fonction de la permittivité et de la résistivité des matériaux [52–54]. Cet effet peut s'étirer des minutes et des heures pour les poudres pharmaceutiques à cause de leur petite taille, leur faible densité et leur nature isolante [54]. Cette charge électrostatique des poudres peut les rendre susceptibles à plus de ségrégation, d'hétérogénéité et d'adhésion [52–54]. Eu égard de l'importance de l'exposition des surfaces des particules de poudre lors de la fabrication des comprimés, notamment lors des étapes de mélange, de granulation, de transport, et de compression, plusieurs auteurs se sont alors intéressés au rôle des propriétés électrostatiques résultantes sur le collage.

Enes et collab. (2012) ont évalué et comparé l'intensité de la charge triboélectrique de près d'une vingtaine d'API et de plusieurs excipients. Les particules sont chargées par le principe du « Shaking » où la poudre est introduite dans un contenant en acier inoxydable et secouée horizontalement, puis versée doucement dans une coupe de faraday muni d'un électromètre pour mesurer la différence de potentiel [54]. Le ratio charge sur masse est utilisé



pour quantifier la charge des particules et la proportion massique des particules restées adhérer au contenant d'acier inoxydable est utilisée pour estimer l'adhésion. Les résultats obtenus montrent que la triboélectrification des API conduit à une charge d'amplitude plus importante et variée (entre -323 nC/g à 170 nC/g) que celle des excipients (entre -11,3 nC/g et 3,82 nC/g) [54]. Cette différence de comportement électrostatique pour les excipients est attribuée à leur faible surface d'échange (grande taille de particule) et leur hydrophilicité comparativement aux API qui sont généralement de plus petite taille. Les résultats rapportés pour l'adhésion suivent la même tendance que le comportement de charge triboélectrique où les particules d'API adhèrent plus à l'acier inoxydable du shaker (entre 10 % et 90 %) que les excipients (entre 0,8 % et 8 %) [54].

Ghori et collab. (2014) ont aussi rapporté des conclusions similaires sur les comportements de charge des API et des excipients, ainsi que le lien avec l'adhésion [52]. Les travaux ont porté sur l'étude de l'influence de 6 différents grades d'un mélange de HPMC et de méthylcellulose (MC) sur les propriétés triboélectriques et d'adhésion de la flurbiprofène. De façon similaire à l'étude de Enes et collab. (2012), les particules sont chargées par le principe du « Shaking », le ratio charge sur masse permet de quantifier la charge et la proportion massique des particules restées coller au contenant d'acier inoxydable permet d'estimer l'adhésion. Ces deux paramètres, charge et adhésion furent ainsi étudiés pour plusieurs échantillons (flurbiprofène pure, chacune des fractions particulières des 6 différents grades de mélange HPMC/MC et plusieurs proportions de mélange entre le flurbiprofène et les polymères de cellulose).

Pour chacun des grades du polymère de cellulose, il a été rapporté une corrélation positive entre la capacité de charge triboélectrique et l'adhésion avec des coefficients de corrélation de 81 % à 98 % [52]. Quant aux propriétés des particules, la charge et l'adhésion sont rapportées baisser lorsque la taille des particules augmente. Cette corrélation est attribuée à l'augmentation de la surface d'échange spécifique pour les particules de petites tailles, augmentant les contacts et les échangeant d'électrons entre particules et surface. Une corrélation négative est par ailleurs rapportée entre la charge/adhésion et l'hétérogénéité du polymère HPMC/MC, la taille des molécules et leur viscosité. En effet, lorsque le ratio HPMC/MC du polymère augmente (de 0 à 38 %), la charge et l'adhésion diminuent (de 57 nC/g à 10 nC/g et de 33 % à 11 % respectivement) [52]. Pour des proportions semblables (38 % à 40 %), la charge et l'adhésion continuent de diminuer (de 12 nC/g à 4 nC/g et de 11 % à 5 %

respectivement) lorsque la taille des molécules et la viscosité augmentent (de 88 000 g/mol à 125 000 g/mol et de 4351 cps à 79 229 cps respectivement) [52].

Quant au flurbiprofène pur, les auteurs rapportent une charge négative très élevée en absolu (-243 nC/g) et une plus grande adhésion (74 %) par rapport à tous les grades du polymère de cellulose [52]. Ces observations sont attribuées au haut degré de cristallinité et de résistivité du flurbiprofène. Pour les mélanges de flurbiprofène avec les différents grades du polymère de cellulose, la charge et l'adhésion chutent au fur et à mesure que la proportion de polymère augmente de 0,5 à 15 % avec un plateau autour de 5 à 10 % pour la charge [52]. Aucun plateau n'est observé par ailleurs pour l'adhésion. De plus, pour les polymères de cellulose ayant les plus faibles charges, même de faibles proportions de mélange à 0,5 % (w/w) avec le flurbiprofène permettent de réduire de moitié la charge et l'adhésion [52]. La charge et l'adhésion sont même neutralisées lorsque la proportion de ce mélange atteint 15 % (w/w). Les auteurs ont attribué ces effets à une réorganisation due à la différence de charge où des particules de charges différentes s'attirent fortement, créant des liaisons covalentes entre l'API et les excipients [52]. Ainsi, plus la différence de potentiel de charge triboélectrique est grande entre l'API et les excipients, plus la charge totale résultante est atténuée, conduisant également à une plus faible adhésion à cause d'une plus forte cohésion de l'API dans la matrice d'excipients.

Plus récemment, Samiei et collab. (2017) ont aussi étudié l'influence des propriétés électrostatiques de neuf formulations différentes en API sur leur propension au collage [55]. Les propriétés électrostatiques furent évaluées à l'aide d'un CDTA (Charge Decay Time Analyser). Cet appareil mesure la différence de potentiel générée par une décharge à effet couronne sur la surface de l'échantillon ainsi que le temps de dissipation de cette charge. Les auteurs ont ainsi utilisé le potentiel de surface générée (V), les temps pour que ce potentiel chute à 36,8 % (Time 1/e) et à 10 % (Time 10 %) pour caractériser les propriétés électrostatiques de neuf API pures, d'une matrice d'excipient composé de MCC et de MgSt et leurs différents mélanges [55]. Le collage fut évalué par la pesée du bout détachable du poinçon supérieure après une série de 100 compressions sur une presse à station unique, Manesty F3. Les résultats obtenus concordent avec les travaux précédents et démontrent une propension plus élevée à la charge électrostatique des API pures par rapport aux excipients et aux formulations. Les API testés ont un potentiel de surface initial qui varie entre -1092 V et -1080 V, ainsi qu'un Time 10 % qui n'est pas atteint au bout de 9 heures tandis que les mélanges ont un potentiel de surface plus faible dans les mêmes conditions (entre -708 V à 1050 V) et un

Time 10 % de 6,9 à 35,6 secondes [55]. Ces données montrent bien un meilleur comportement électrostatique (plus faible charge et vitesse de dissipation rapide) induit par l'ajout d'excipients aux formulations. Pour la corrélation des propriétés électrostatiques avec la propension au collage, Samiei et collab. (2017) ont rapporté une faible corrélation entre le potentiel de charge initial des API et des mélanges avec le collage de l'ordre de 17 % et 58 % respectivement. Mais une bonne corrélation positive ( $R^2 \cong 89\%$ ) est rapportée avec le temps de dissipation de la charge [55]. Ainsi, la propension au collage d'une formulation augmente avec sa tendance à maintenir sa charge électrostatique de surface sur la durée.

L'ensemble de ces études montrent que les API ont un mauvais comportement électrostatique par charge triboélectrique (forte charge et lente dissipation) par rapport aux excipients. Ce qui fut attribué à la relative grande taille des particules d'excipients et de leur hydrophilicité. Le mélange API-Excipient permet d'améliorer le comportement électrostatique des API. Ces propensions à la charge triboélectrique contribuent à l'adhésion des particules sur les surfaces métalliques à travers des interactions électrostatiques, notamment le temps de dissipation de la charge des formulations qui corrèle bien avec des données de collage mesurées par pesée du bout détachable des poinçons. Le rôle des propriétés électrostatiques des API dans le collage s'aligne bien avec la compréhension du rôle des propriétés physiques et chimiques décrite plus haut, notamment le rôle de l'énergie de surface spécifique dans l'initiation du collage. On comprend aussi aisément qu'elle contribue de la même façon à la complexité du phénomène de collage. L'amélioration du comportement électrostatique (réduction de la capacité de charge, du temps de dissipation et de la cohésion API-matrice d'excipient) pourrait être mise à profit pour atténuer la propension au collage des poudres par une meilleure sélection de la matrice d'excipients.

#### **5.2.4 La matrice d'excipients :**

Suivant le modèle conceptuel du collage (section 5.1), l'aggravation du collage par l'accumulation de plusieurs couches de poudre à la surface des poinçons (collage de type II) est gouvernée par le rapport de force entre la cohésion API-API et la cohésion API-Excipient. En accord avec ce modèle, plusieurs travaux de la littérature rapportent qu'une importante cohésion API-Excipient permet de limiter le collage [22,35,45,52,55–57].

Paul et collab. (2018) ont évalué le rôle de quatre différentes matrices d'excipients sur le collage de l'IBU et du CEL [57]. Les matrices d'excipients évaluées sont le MCC de grade avicel PH102 et avicel PH105(n), l'hypermellose et un mélange (3:1) d'amidon et

d'avicel PH102. Les formulations sont constituées chacune de 20 % (w/w) d'API, 0,25 % (w/w) de MgSt complétés avec les différentes matrices d'excipients. Le collage a été évalué dans cette étude par la pesée du bout détachable des poinçons. Pour l'ensemble des matrices d'excipients, les résultats montrent que la propension au collage du CEL est plus importante que celle de l'IBU. Les auteurs ont relié cette observation à une faible cohésion API-API de l'IBU par rapport au CEL [57]. Toutefois, pour les 2 API, la propension au collage à travers les différentes matrices d'excipients suit le même ordre: avicel PH105(n) < avicel PH102 < hypermellose < mélange amidon et Avicel PH102. Les forces de cohésion API-Excipient sont rapportées suivre l'ordre opposé [57]. Ainsi, pour un API donné, une matrice d'excipients favorisant la cohésion API-excipient permet de réduire le collage.

Des conclusions similaires ont également été rapportées par Capece et collab. (2019) [56]. En investiguant le rôle de la surface spécifique des particules dans le collage, les auteurs ont démontré qu'une surface spécifique d'API plus faible relativement à la surface spécifique de l'excipient conduisait à une plus grande cohésion API-Excipient et permettait la réduction du collage. Dans la même lignée, les travaux de Saddik et collab. (2021) confortent aussi le rôle important de la cohésion API-Excipient dans l'occurrence du collage [22]. Cette dernière étude démontre clairement que les excipients peuvent contribuer à réduire le collage à travers divers mécanismes, notamment la réduction de la surface de contact entre les particules d'API et la surface des poinçons, mais aussi leur impact sur le comportement électrostatique des API comme décrit plus haut (section 5.2.3). L'excipient réduit la force d'adhésion des particules d'API aux poinçons en réduisant la capacité et le temps de dissipation de la charge électrostatique des particules d'API. En plus, le différentiel de charge entre API et excipient conduit généralement à l'établissement de liaisons covalentes au sein du lit de poudre, favorisant une forte cohésion API-Excipient, ce qui permet de réduire également le collage.

Les données de la littérature établissent ainsi clairement que le choix d'une matrice d'excipients convenable permet de réduire la propension au collage d'une formulation donnée et cette propriété peut être utile dans les étapes de développement d'une formulation afin de prévenir le collage. Plusieurs auteurs ont même proposé des index de collage basé sur cette propriété afin de faciliter le criblage des matrices d'excipients en développement [17,22,56,58]. Capece et collab. (2019) ont par exemple proposé le différentiel de force de cohésion API-excipient et API-API rapporté à la fraction de surface spécifique de l'API relative à la surface spécifique de l'excipient [56]. Cet index prend ainsi en compte autant le rapport des forces de cohésion, mais aussi la disponibilité de surface de contact entre les particules d'API et les

poinçons. Les résultats proposés par Capece et collab. (2019) confirment la pertinence de l'index proposé [56].

Bien qu'il soit clair qu'une plus grande cohésion API-Excipient soit bénéfique au collage, les interactions fondamentales qui le gouvernent restent assez fragiles et moins comprises. Par exemple, McDermott et collab. (2010) ont investigué l'effet de différents grades de MCC sur la propension au collage d'une formulation à base d'ABT-089 [35]. Les deux grades de MCC utilisés sont le Vivapur 102 avec des particules à surface rugueuse et le Avicel PH102 avec des particules à surface lisse. À l'opposé d'une conclusion rationnelle où les surfaces plus rugueuses du Vivapur 102 permettraient une cohésion API-Excipient plus importante et donc moins de collage, les résultats obtenus par McDermott et collab. (2010) ont montré plus de collage pour la formulation avec du Vivapur 102 relativement à l'Avicel PH102 [35]. Loin d'être affirmatifs dans l'explication de ces résultats, les auteurs postulent que la rugosité des particules de Vivapur 102 réduirait la disponibilité des API à la lubrification utilisée dans l'étude. On comprend alors que la force de la cohésion API-Excipient peut être influencée par d'autres paramètres qui pourraient être reliés ou non à la nature stricte des excipients.

### **5.2.5 La lubrification :**

La lubrification d'une formulation permet d'améliorer ses propriétés d'écoulement et de réduire la friction aux parois des matrices durant la compression et l'éjection. Elle peut remplir également une fonction antiadhérente en prévenant le collage. Nous présentons ci-dessous l'état de l'art de l'effet de ces substances sur le collage, notamment le MgSt qui est le lubrifiant le plus couramment employé dans les formulations pharmaceutiques.

De façon générale, il est admis que la lubrification permet de réduire le collage [16–18,29,35,59,60]. Dans leurs travaux sur le collage avec le SEM et l'EDS, McDermott et collab. (2010) ont investigué l'influence du MgSt sur le collage de l'ABT089 et les résultats montrent que l'augmentation du contenu en MgSt conduit à une baisse significative de la quantité d'API collée sur la surface des poinçons [35]. Les travaux de Sendall et collab. (1986) avaient aussi abouti à des conclusions similaires avec le macrogol [38]. De même, Swaminathan et collab. (2017) ont également rapporté un effet bénéfique de la lubrification avec le MgSt sur le collage de formulations à base de lactose et de mannitol [16]. Les auteurs ont utilisé un poinçon instrumenté pour mesurer le profil de la force d'adhésion lors de la phase de décompression et les résultats obtenus confirment qu'une concentration croissante de MgSt permet de réduire le

collage. En utilisant d'autres méthodes de mesure du collage, des résultats similaires avec le lactose ont aussi été rapportés par Toyoshima et collab. (1988) et Osamura et collab. (2017) [18,51]. Cet effet bénéfique du MgSt sur le collage est consensuellement expliqué par un enveloppement des particules d'API par le lubrifiant, ce qui les rend moins disponibles à l'adhésion à la surface des outillages de compression. Ando et collab. (2020) ont investigué et confirmé au niveau mécanistique l'effet de la lubrification avec le MgSt [23]. En se servant de la chromatographie gazeuse inverse, les auteurs ont évalué le changement dans les propriétés de la surface des particules d'un API confidentiel lors du mélange avec le MgSt. Les résultats obtenus montrent que ce sont des composants d'énergie de surface qui diminuent au cours du mélange avec le MgSt [23]. Au fur et à mesure que ces composants d'énergie de surface baissent, le collage de la formulation étudiée baisse également [23].

Plus récemment, Patel et collab. (2021) ont aussi étudié l'effet de la lubrification avec le MgSt et le sodium stearyl fumarate (SSF) sur le collage d'une formulation de ketoprofen à la surface de cinq différents types de poinçons (en acier S7 ou M340 ou revêtu de chrome dur HCr, de nitrure de zirconium ZrN ou de nitrure de chrome CrN) [17]. Les résultats obtenus montrent que l'augmentation de la concentration des deux types de lubrifiant permet de réduire le collage à tous les cinq types de poinçons. Mais, d'un lubrifiant à l'autre ou d'une concentration à l'autre, l'ordre de performance des cinq types de poinçon par rapport au collage n'était pas toujours le même. Par exemple, sans ajout de lubrifiant, l'ordre de performance au collage des différents types de poinçon était CrN>HCr>M340>S7>ZrN, tandis qu'avec 1 % de MgSt ou de SSF, cet ordre était quasiment inversé, soit S7>M340>CrN>HCr>ZrN et CrN>M340>S7>HCr>ZrN respectivement [17]. Il existe donc une forte interaction entre la lubrification et le type de poinçon utilisé ou son revêtement.

Cet effet d'interaction avait également été rapporté par Al-Karawi et collab. (2017). Les auteurs ont investigué l'effet de quatre différents types de lubrifiant (le MgSt, le glyceryl monostéarate-GMS-, le SSF et le lubritose) sur le collage de l'IBU à quatre différents types de revêtement de poinçons [14]. Un effet d'interaction significatif entre le type de revêtement et le type de lubrifiant fut ainsi rapporté et la plus grande sensibilité au collage a été identifiée pour la lubrification avec le lubritose, suivi par le GMS, puis le MgSt et enfin le SSF, qui montrait les meilleures propriétés antiadhérentes [14]. L'emploi de lubrifiants doit alors se faire de concert avec les autres paramètres de la compression pour s'assurer de son efficacité.

En plus de l'existence d'effets d'interaction, plusieurs études mitigent l'effet bénéfique même de la lubrification sur le collage de formulations à base d'ibuprofène. Roberts et collab.

(2004) ont étudié l'effet du type et de la concentration de lubrifiant sur le collage de formulations à base d'IBU [12]. Divers niveaux de lubrifiants (MgSt et acide stéarique) variant de 0,25 à 2 % furent testés et le collage a été quantifié après chaque cycle de compression en immergeant la face du poinçon dans de l'éthanol suivi d'une analyse par spectrophotométrie à 264 nm. Les résultats obtenus par Roberts et collab. (2004) montrent que la propension au collage de l'IBU est proportionnelle à la concentration de MgSt, tandis qu'elle n'est pas affectée par la concentration d'acide stéarique [12]. L'effet délétère du MgSt sur le collage de l'IBU a été attribué à l'abaissement du point de fusion de celui-ci et l'affaiblissement de la cohésion interne des compacts, tous induit par le lubrifiant. Les travaux de Mullarney et collab. (2012) soutiennent aussi ces conclusions d'un effet délétère du MgSt sur le collage de formulations à base d'IBU [39]. En utilisant une méthode différente de quantification du collage, les auteurs avaient montré que le collage augmentait avec la concentration de MgSt pour une formulation à base d'IBU, tandis qu'elle le réduisait avec une formulation à base de mannitol [39]. Par opposition à ces deux travaux, une étude conduite récemment par Nakamura et collab. (2016) dans des conditions similaires aux travaux de Roberts et collab. (2004) conclut elle, à des résultats complètement contradictoires où le collage d'une formulation à base d'IBU diminuait lorsque la quantité de MgSt augmente [58].

Ainsi, pour les formulations à base d'IBU, l'impact de la lubrification apparaît mitigé, tantôt délétère, tantôt bénéfique. Cet état de fait est typique du phénomène de collage où tellement de paramètres interagissent qu'en définitive, même des études quasi similaires peuvent donner des résultats contradictoires à cause probablement de différences opératoires ou matérielles souvent considérées mineures. Du reste, les données de la littérature permettent tout de même de conclure à un effet bénéfique globale de la lubrification sur le collage. La lubrification pourrait donc être une solution efficace pour réduire le collage. Cependant, moins de recours y sont faits, car son incorporation dans la formulation a des effets délétères sur les attributs qualité des comprimés, notamment une réduction de leur cohésion interne, le rallongement de leur temps de dissolution et de désintégration. Ces effets secondaires sont directement reliés au mécanisme d'action même du lubrifiant et de sa quantité dans la formulation. Pour tirer profit de l'effet bénéfique globale de la lubrification sur le collage sans ses inconvénients sur les attributs qualité des comprimés, la lubrification externe a fait son apparition dans les années 1990. Elle consiste en la pulvérisation continue de poudres de lubrifiant convoyée par de l'air comprimé sur les surfaces des poinçons et matrices au cours de la compression. Le lubrifiant est ainsi acheminé aux interfaces entre la poudre et les outillages

pour réduire l'adhésion et les frictions, sans former des comprimés mécaniquement faibles avec de long temps de dissolution et de désintégration. Une description détaillée de la technique de lubrification externe incluant les inconvénients et les challenges de son implémentation est proposée à l'annexe 1 du présent document.

### **5.3 Les paramètres relatifs à la compression :**

#### **5.3.1 Les procédés de production des poudres:**

De la section 5.2 précédente, il fut conclu que le collage est principalement dû aux particules d'API, notamment leurs propriétés physiques, chimiques et électrostatiques. Les procédés de production des poudres par nature sont les opérations qui ont le plus d'impact sur ces propriétés lors de la fabrication des comprimés. Se faisant, il apparaît alors clairement que les voies de traitement des poudres avant compression jouent un grand rôle dans leur propension au collage.

Bejugam et collab. (2013) ont comparé les occurrences de collage de diverses formulations produites par compression directe et par granulation sèche sous rouleau compacteur [61]. Les auteurs ont évalué le collage à l'aide d'un indicateur semi-quantitatif variant de 1 à 10 suivant l'intensité du collage. Cet indicateur est bâti par intégration des données sur la quantité de poudre collée sur la surface des poinçons, la rugosité des surfaces des comprimés, la difficulté dans l'éjection et la fréquence de nettoyage des poinçons [61]. Les résultats obtenus montrent que la propension au collage est très élevée pour tous les API formulés avec la compression directe comparativement à la granulation sèche. La granulation sèche améliorerait les propriétés de compressibilité des API conférant une meilleure cohésion, ce qui réduit leur forte tendance au collage comparée à la compression directe. Chatteraj et collab. (2018) postulent aussi que la compression directe est plus susceptible au collage du fait que l'API ne soit pas assez lié dans la matrice d'excipient [9]. La granulation humide, quant à elle, serait meilleure à cause de l'utilisation des liants qui enveloppent les API et réduisent leur exposition comparativement à la granulation sèche qui ne permet pas un enveloppement complet [9].

Ces résultats et hypothèses présument que les performances au collage diminuent de la granulation humide à la compression directe en passant par la granulation sèche. Cependant, il n'a pas été établi à date qu'une voie de fabrication serait supérieure à une autre pour tout type de formulation. Néanmoins, on peut conclure à un effet probable des voies de fabrications sur le collage à travers leur influence sur les propriétés physiques, chimiques et électrostatiques des particules, notamment la forme et la taille des particules, leur énergie de surface, leur charge



triboélectrique ou encore leur humidité résiduelle. Des paramètres comme les temps de mélange, de granulation ou encore de séchage modulent probablement une ou plusieurs de ces propriétés des API et des formulations. Les méthodes de production des poudres rajoutent ainsi un niveau de difficulté supplémentaire dans l'étude et la compréhension globale du collage, car des poudres non formulées dans les mêmes conditions pourraient conduire à des particules de poudres aux propriétés différentes. Toutes choses qui conditionnent grandement la propension au collage des formulations.

### **5.3.2 La durée de la compression :**

Les données de la littérature établissent unanimement que le collage est proportionnel à la durée de la compression. McDermott et collab. (2010) ont rapporté sur l'influence de plusieurs paramètres, dont la durée de la compression, sur la propension au collage de deux API (ABT089 et ABT279) [35]. Les résultats obtenus montrent que l'intensité du collage augmente avec la durée de la compression (le nombre de cycles consécutifs en compression), quelles que soient les conditions de compression. Les rapports de proportionnalité sont rapportés linéaires avec des coefficients de corrélation très élevés même pour les plus longs cycles de compression (7000 cycles pour ABT089 avec un  $R^2$  de 0.9978 et 5000 cycles pour ABT279 avec un  $R^2$  de 0.9955) [35]. Ces fortes corrélations s'expliquent par l'accumulation continue de poudre à la surface des poinçons, mais aussi de potentiels pics locaux de température à la surface des poinçons facilitant l'adhésion [29,35].

Nombreuses autres études rapportent des conclusions similaires sur la proportionnalité entre la durée de la compression et le collage [9,14–16,29,36,45,46]. En fait, cette relation est même connue comme l'un des challenges majeurs limitant la capacité et la portée des études de développement à petite échelle sur le collage. En effet, ce n'est qu'une fois que les formulations sont rendues à l'échelle industrielle que le collage est très souvent observé après de longs cycles de compression [5,9,10,26]. L'opportunité de détecter et prévenir le collage est donc largement amenuisée au moment du développement de la formulation où une plus grande marge de manœuvre dans le choix des excipients est disponible. De plus, la reproduction du collage à échelle laboratoire devient alors très difficile ou très coûteuse en temps et en argent à cause de la quantité importante de poudre nécessaire à de longues séances de compression. Lorsqu'il arrive à être reproduit, la question de la représentativité se posera également. Les simulateurs de compression décrits plus haut sont des presses spécialement désignées pour aider les études de développement au travers de ces nombreux challenges de reproductibilité et de représentativité du collage entre les échelles laboratoire et industrielle.

### **5.3.3 Force de compression:**

Les données de la littérature fournissent des conclusions contradictoires quant à l'impact de la force de compression sur le collage.

D'une part, plusieurs auteurs rapportent que l'augmentation de la force de compression a un effet bénéfique pour le collage [13,15,27,35,36,51,62,63]. Par exemple, Kakimi et collab. (2010) ont rapporté sur le rôle de la force de compression dans le collage de formulations à base de n-butyl p-hydroxybenzoate (BHB) [62]. Le collage fut évalué à travers la force d'adhésion (contrainte de cisaillement) entre les particules BHB de la surface des comprimés et le poinçon inférieur et les résultats obtenus confirment que la force d'adhésion diminue lorsque la force de compression augmente. Kakimi et collab. (2010) attribuent ce lien à une forte cohésion dans le comprimé, qui serait plus grande que les forces d'adhésion des particules à la surface des poinçons [37,62]. Des conclusions similaires ont également été rapportées par Saniocki et collab. (2013) avec une formulation à base d'IBU [37]. Les compressions sont réalisées sur une presse rotative avec trois niveaux de force de compression (7, 12 et 19 kN). Après la production de 25 comprimés, la poudre collée sur la surface des poinçons inférieurs est dissoute dans du méthanol et dosée par HPLC. Les résultats obtenus par Saniocki et collab. (2013) montrent également que, quelle que soit la proportion d'API, le collage diminue avec la force de compression [37].

D'autre part, à l'opposé des travaux illustrés ci-dessus, certains rapports de la littérature concluent plutôt à un effet délétère de la force de compression sur le collage [11,16,39,48,64–66]. Par exemple, Paul et collab. (2017) ont étudié l'influence de la force de compression (50 à 200 MPa) sur la propension au collage de cinq formulations différentes en API (CEL, IBU, MA, GLY et Tofacitinib - TFC) [64]. La quantité de poudre collée sur le poinçon supérieur à chaque 25 compressions d'une série de 100 compressions a été utilisée pour évaluer le collage. Les résultats obtenus montrent que l'intensité du collage augmente avec la force de compression et l'effet est plus marqué pour des composés ductiles [64]. Les auteurs attribuent cela à une grande couverture de la surface des poinçons par les particules d'API, mais aussi une plus grande force de cohésion entre particules d'API. En effet, les données sur la plasticité des API étudiés montrent une corrélation entre la tendance au collage et les propriétés mécaniques des formulations conduisant à une plus grande couverture de la surface des poinçons [64]. Ainsi, il existe donc une interaction entre la force de compression et les propriétés mécaniques des particules. Les particules ductiles, susceptibles à plus de déformations plastiques, conduisent à une plus grande couverture de la surface des poinçons

contrairement à des particules fragiles tendant à se fragmenter ou celles qui se déforment élastiquement [36,64]. Pourtant, l'existence de cette seule interaction ne permet pas d'expliquer la grande diversité des résultats de la littérature. Par exemple, pendant que Paul et collab. (2017) rapportent un effet délétère de la force de compression sur le collage des formulations à base d'IBU [64], Sanioccki et collab. (2013) rapportent l'exact contraire dans des conditions opératoires quasi similaires [37].

Ces conclusions contradictoires de l'effet de la force de compression sur le collage tiennent essentiellement du caractère multifactoriel prononcé du phénomène. Des travaux récents prenant en compte plusieurs autres paramètres de façon simultanée ont permis d'apporter plus de compréhension sur le rôle de la force de compression dans le collage [10–14]. En plus de l'interaction avec les propriétés mécaniques des particules décrite ci-dessus, ces travaux démontrent clairement l'existence d'autres effets d'interactions significatives entre la force de compression et les propriétés de la surface des poinçons, notamment la rugosité et le type de revêtement [10–14].

En 2003, Robert et collab. (2003) ont étudié simultanément l'effet de la rugosité de la surface des poinçons, le type de revêtement, la force de compression et la proportion sur le collage de formulations à base d'IBU [11]. Les compressions sont effectuées avec trois poinçons différents : un nouveau poinçon avec une faible rugosité, un ancien poinçon avec une forte rugosité et un poinçon revêtu de chrome par électro-placage avec une faible rugosité. La fraction d'ibuprofène collée sur le poinçon du haut fut quantifiée par spectroscopie à 264 nm. Les résultats dépouillés par analyse de la variance (ANOVA) ont montré que la proportion d'API, la force de compression, la rugosité de la surface des poinçons, ainsi que l'interaction de tous ces trois facteurs avaient une influence significative sur le collage [11]. Le nouveau poinçon à faible rugosité présente un profil de collage proportionnel à la force de compression, tandis que l'ancien à forte rugosité se montre insensible aux variations de la force de compression avec une plus faible propension au collage. Les auteurs ont attribué cette meilleure performance au collage de l'ancien poinçon à forte rugosité par l'existence d'imperfections à sa surface qui agiraient en réduisant les forces d'attraction entre la surface du poinçon et du comprimé [11]. Les essais effectués avec le poinçon revêtu de chrome montrent que l'application du revêtement augmente le collage à faible force de compression et le réduit à forte force de compression. Les auteurs suggèrent que la tendance inverse de l'effet du revêtement serait due à une augmentation de l'adhésion par l'effet des charges statiques entre

les particules (qui se seraient chargées durant le mélange et/ou le remplissage du poinçon) et la surface des poinçons revêtus, charges qui s'annihileraient à forte force de compression [11].

Plus récemment, Al-Karawi et collab. (2017) ont aussi étudié simultanément l'effet de trois facteurs (la force de compression, le type de lubrifiant et le revêtement des poinçons) sur le collage de formulation à base d'IBU [14]. Deux niveaux de force de compression (5 kN et 19 kN), quatre différents types de revêtement (non revêtu -U-, le chrome dur-HCr-, le nitrure de titane -TiN- et le nitrure de chrome -CrN-) et quatre types de lubrifiant (MgSt, GMS, SSF et le lubritose) furent testés suivant un plan d'expérience D-optimal. Les auteurs ont ainsi rapporté que tous les facteurs étudiés avaient un effet significatif sur le collage [14]. De plus, ils rapportent un effet d'interaction significatif entre la force de compression et le type de revêtement : à faible force de compression, la quantité de poudre collée sur la surface des quatre différents types de revêtement augmentent selon l'ordre suivant : CrN < TiN < U < HCr et à forte force de compression, les différences en matière de collage s'estompent [14]. Les auteurs attribuent cette tendance de diminution du collage avec l'augmentation de la force de compression à une plus grande cohésion au sein du comprimé.

Au vu de ces nombreux effets d'interaction, les conclusions contradictoires de la littérature quant au rôle de la force de compression apparaissent compréhensibles. Ainsi, en fonction des propriétés mécaniques des API utilisés dans les différentes études, mais aussi de l'état de la surface des poinçons employés, la force de compression peut avoir un effet bénéfique ou délétère sur le collage. Bien plus d'effets d'interactions pourraient également intervenir. La force de compression apparaît donc comme l'un des paramètres clés dans l'occurrence du collage. La compréhension de son rôle ou l'identification de ses niveaux optimaux pour prévenir le collage nécessite un jeu d'équilibre fin avec tellement d'autres paramètres que son investigation isolée n'apporterait probablement que beaucoup plus de questions que de réponses.

#### **5.3.4 La vitesse de compression :**

Les données de la littérature concordent et indiquent que l'augmentation de la vitesse de compression a un effet délétère sur le collage [9,13,27,62]. Khedekar (2017) a investigué l'effet de plusieurs paramètres de compression, dont la vitesse sur le collage de formulations à base d'IBU [27]. Le collage fut évalué par la pesée des résidus de poudre collée sur la surface des poinçons après des cycles de 10 compressions et les résultats obtenus montrent que le collage augmente avec la vitesse de compression. Avec différentes autres méthodes d'évaluation du

collage, des conclusions similaires ont aussi été rapportées par Kakimi et collab. (2010) et Al-Karawi et collab. (2018) [13,62]. Cet effet délétère de la vitesse de compression sur le collage est consensuellement expliqué par une transmission inadéquate de la pression à haute vitesse de compression, conduisant à une augmentation de la proportion des déformations élastiques par rapport aux déformations plastiques dans le milieu granulaire [13,27,62]. Il en résulte alors un affaiblissement important de la cohésion interne du lit de poudre et donc plus de collage.

Suivant le mécanisme de la compression sur une presse rotative, la transmission inadéquate de la force de compression n'est pas directement due à la vitesse de compression. En effet, l'augmentation de la vitesse de compression impacte plutôt un autre paramètre plus critique de la compression : le temps de contact ou « dwell time » en anglais. Ce paramètre correspond à la durée pendant laquelle le lit de poudre subit la contrainte maximale de la force de compression. Ainsi, plus ce temps de contact est important, plus la durée des contacts interparticulaires au sein du lit de poudre est importante, conduisant à une densification et une cohésion plus importante et homogène. L'augmentation de la vitesse de compression ne fait en réalité que diminuer ce temps de contact et par conséquent elle conduit à plus de collage.

Tout comme pour la durée de la compression, le rôle de la vitesse de compression ainsi décrit, rajoute également un niveau de complexité supplémentaire dans l'étude et la compression du collage de façon globale. En effet, les grandes cadences de compression à échelle industrielle sont aussi admises de façon générale comme l'une des principales causes du collage. Ces cadences de compression sont difficiles ou très coûteuses à reproduire à l'échelle laboratoire. Au vu de l'effet qui en découle, le déficit de reproductibilité pose tout de suite la question de la représentativité des modèles d'étude du collage à échelle laboratoire. Les faibles vitesses de compression expérimentées au laboratoire conduisent à des temps de contact importants et donc plus de cohésion et moins de collage. Ceci explique aussi la difficulté d'observation du collage lors des études à échelle laboratoire, notamment au cours du développement des formulations.

Afin de surpasser ces lacunes de reproductibilité et de représentativité des modèles d'étude expérimentaux du collage à échelle laboratoire, Ito et collab. (2018) ont récemment développé et introduit de nouveaux designs de poinçons de laboratoire permettant de mimer les temps de contact et de consolidation généralement rencontrés à l'échelle industrielle [60]. Les modifications de design consistent essentiellement en la réduction du rayon de la partie plate de la tête des poinçons. Cela permet de réduire la surface de contact entre la tête des poinçons et les rouleaux de compression, ce qui contribue à réduire également la durée

d'application de la contrainte maximale lors de la compression. Les résultats rapportés par Ito et collab. (2018) confirment l'efficacité de ces designs de poinçons à réduire le temps de contact à des niveaux même plus faibles que ceux rencontrés à échelle industrielle. Dans le futur, plus de recours devraient être faits à ces designs de poinçons pour améliorer la reproductibilité et la représentativité des modèles d'étude expérimentaux du collage à échelle laboratoire. D'autant plus que les récentes données de la littérature montrent que le temps de contact est un excellent paramètre pour assurer la représentativité lors de la montée ou de la descente en échelle [14].

#### **5.4 Les paramètres relatifs aux outillages :**

Le collage se développant sur la surface des outillages, de très nombreux paramètres se référant à ces outillages sont rapportés dans la littérature. Cependant, le collage de poudre sur les matrices est peu abordé, certainement à cause de la faible prévalence de cette forme de collage, mais aussi l'accès difficile à la matrice. Les paramètres rapportés dans la littérature concernent quasiment tous les aspects des poinçons : la forme, le profil, l'âge, les inscriptions et les propriétés des surfaces, comme la rugosité, l'hydrophobicité ou encore la finition.

##### **5.4.1 La géométrie des poinçons :**

La forme du bout, mais aussi de la tête des poinçons peuvent grandement influencer le collage des poudres pharmaceutiques.

Comme expliqué dans la section 5.3.4 précédente, la forme de la tête des poinçons peut moduler le temps de contact lors de la compression. Se faisant, elle peut grandement impacter le collage. Khededar (2017) a investigué l'effet de la forme des têtes des poinçons sur le collage de formulations à base d'IBU [27]. Les auteurs ont évalué quatre poinçons (P1, P2, P3 et P4) avec différents diamètres de tête plate qui sont respectivement  $\cong 0$  mm, 7,9 mm, 15,9 mm et 19,7 mm. À vitesse de compression égale, le diamètre de tête plate est proportionnel au temps de contact, car la surface de contact entre les rouleaux de compression et la tête du poinçon sera importante. Par exemple, à une vitesse de compression de 20 RPM, les temps de contact des quatre poinçons P1, P2, P3 et P4 sont respectivement d'environ  $\cong 0$  msec, 50 msec, 101 msec et 125 ms [27]. En accord avec l'effet du temps de contact décrit dans la section 5.3.4 précédente, les résultats obtenus par Khededar (2017) montrent que le collage diminue avec l'augmentation du diamètre de tête plate des poinçons [27].

La forme du bout des poinçons peut aussi moduler la propension au collage des poudres pharmaceutiques. En effet, Osamura et collab. (2017) ont rapporté que les performances anti-

collantes du MgSt peuvent varier en fonction de la forme du bout des poinçons [18]. Les auteurs avaient évalué la performance de quatre concentrations croissantes de MgSt (0 %, 0,5 %, 1 % et 3 %) à réduire le collage de formulations à base de lactose sur la surface de quatre types de poinçons avec des bouts de forme différente. Le poinçon de type 1 possède un bout plat avec une ligne sécante. Le poinçon de type 2 a une forme convexe tandis que le poinçon de type 3 est de forme concave. Le poinçon de type 4 est quant à lui de forme convexe et possède une ligne sécante et des gravures. Les résultats rapportés par Osamura et collab. (2017) montrent que la propension au collage est dépendante de la forme du bout des poinçons. Du poinçon de type 1 au type 4, la quantité de MgSt nécessaire pour prévenir le collage était de 0,5 %, 0,5 %, 3 % et 1 % [18]. La propension au collage est donc plus importante pour le poinçon de forme concave suivi du poinçon de forme convexe avec une ligne sécante et des notes de gravures. Les poinçons de forme plate ou convexe ont de faibles propensions au collage similaires. Des conclusions similaires sont également rapportées par Reed et collab. (2015) [19].

Les données de la littérature indiquent ainsi que la géométrie des poinçons peut bouleverser les équilibres de forces tendant à réduire ou à exacerber le collage. Ce paramètre supplémentaire ainsi que ces potentielles interactions doivent alors être considérés dans les études pour tenter de comprendre le collage de façon globale.

#### **5.4.2 La rugosité de la surface des poinçons:**

Plusieurs travaux dans la littérature rapportent que la rugosité de la surface des poinçons a un effet délétère sur le collage [10,13,19,20,29,40,42]. Bunker et collab. (2011) ont mesuré par AFM la force d'adhésion entre des particules de lactose et deux surfaces (polie et non polie) de poinçon en nitrure de chrome et les résultats rapportés montrent que la force d'adhésion augmente avec la rugosité de la surface des poinçons [20]. La force d'adhésion moyenne à la surface non polie est de 19,2 nN, tandis que celle à la surface polie est de 14,2 nN [20]. Les auteurs suggèrent que l'abondance des aspérités augmente la surface de contact spécifique, conduisant à plus de collage. Wang et collab. (2015) ont rapporté des conclusions similaires sur l'effet délétère de la rugosité sur l'adhésion de particules de MA [40]. Le collage fut évalué à l'aide d'un analyseur de texture et une HPLC. Le principe est simple : une bille à roulement en acier inoxydable de rugosité connue est placée dans une matrice évacuée et la poudre est ensuite comprimée. Après consolidation, la bille est extraite par extrusion et la poudre collée sur sa surface est dissoute dans du méthanol et quantifiée par HPLC. Les résultats rapportés par Wang et collab. (2015) montrent que lorsque la rugosité moyenne de la surface du métal passe de 289 nm à 407 nm, la quantité de poudre collée est presque triplée [40]. Plus

récemment, Al-Karawi et collab. (2017) ont également confirmé qu'une surface plus rugueuse conduisait à plus de collage pour une formulation à base d'IBU [42]. Les auteurs ont développé une méthode d'étude de la surface des poinçons qui consiste à réaliser un moulage haute résolution suivie de mesures 3D. Un total de 58 paramètres de surface sont ainsi mesurés et traités par ACP. En utilisant cette nouvelle approche, Al-Karawi et collab. (2017) ont étudié la propension au collage de différentes textures de surface de poinçons et les résultats obtenus conduisent aux mêmes conclusions qu'une quantification par HPLC de la poudre collée sur la surface des poinçons, à savoir que le collage augmente avec la rugosité de la surface des poinçons [42].

À l'opposé de ces précédents travaux, de nombreux autres auteurs rapportent plutôt un effet bénéfique de la rugosité sur le collage [11,67–69]. Schuman et Searle (1992) ont étudié la propension au collage d'une substance confidentielle à différents revêtements de poinçons [68]. Par évaluation visuelle, les auteurs rapportent moins de collage pour une surface de poinçon rugueuse comparativement à une surface plus lisse. Ces observations furent attribuées aux imperfections de la surface rugueuse qui agiraient comme des locus brisant les forces d'adhésion entre le comprimé et le poinçon [68]. Des conclusions similaires quant à l'effet bénéfique de la rugosité des poinçons sur le collage furent rapportées également par Roberts et collab. (2003), Uemura et collab. (2007) et Uchimoto et collab. (2013) [11,67,69].

Les récents travaux de Takeushi et collab. (2020) apportent tout de même un peu plus de lumière sur les conclusions contradictoires de l'effet de la rugosité sur le collage [15]. Les auteurs ont investigué la capacité de prévention du collage par l'utilisation de poinçons avec différents niveaux de rugosité. Le collage a été évalué par dosage spectrophotométrique et deux modèles d'API furent utilisés, notamment l'acétaminophène et l'IBU. Les résultats obtenus montrent que l'augmentation de la rugosité des poinçons permet de diminuer le collage de la formulation à base d'acétaminophène, tandis qu'elle n'a aucun impact ou aggrave même dans certaines conditions le collage de l'IBU [15]. Les auteurs expliquent cette observation par l'existence d'une interaction entre la rugosité des poinçons et les propriétés mécaniques des particules. La rugosité permet de réduire le collage en agissant comme des locus qui limitent la surface de contact avec les particules d'API. Mais, en présence de particules ductiles comme l'IBU où les déformations plastiques prédominent, la grande capacité de déformations empêche la rugosité de réduire substantiellement les surfaces de contact, d'où une inefficacité à réduire le collage de cet API aux poinçons.



Les résultats de Takeushi et collab. (2020) permettent de confirmer également une interaction entre la rugosité des poinçons et la force de compression pour la formulation à base d'IBU. À faible force de compression, l'augmentation de la rugosité permet de réduire le collage, tandis qu'elle n'a pas d'effet ou l'aggrave à forte force de compression [15]. Le comportement mécanique des poudres étant dépendant de la contrainte subie, l'existence de cette interaction n'est pas surprenante. À faible force de compression, bien que les particules d'API soient ductiles, la contrainte appliquée n'excède pas leur domaine d'élasticité. Se faisant, la rugosité de la surface des poinçons permet effectivement de réduire la surface de contact avec les particules, et donc le collage. Mais à forte force de compression, les particules se déforment de façon plastique et épousent les aspérités de la surface, d'où l'incapacité de la rugosité à réduire les surfaces de contact et donc le collage. Par ailleurs, si la force de compression est suffisamment importante, du verrouillage mécanique pourrait même intervenir et contribuer à une aggravation du collage.

En définitive, la rugosité de la surface des poinçons a une grande influence sur le collage des poudres pharmaceutiques, mais cet effet est fortement dépendant des propriétés mécaniques des poudres et de la force de compression. De ce fait, tout comme la force de compression, l'étude de façon isolée de l'effet de la rugosité des poinçons sur le collage n'apporterait probablement que plus de questions que de réponses. Une approche intégrative et simultanée est donc nécessaire pour une meilleure compréhension du collage.

### **5.4.3 Le revêtement de la surface des poinçons :**

Suivant le modèle conceptuel du collage présenté à la section 5.1, l'initiation du collage résulte de l'intensité de l'adhésion entre les particules d'API et la surface des poinçons ( $F_1$ ). La nature du revêtement de cette surface est donc un paramètre très important dans l'occurrence du collage pour une formulation donnée. L'application de revêtement permet de modifier substantiellement les propriétés physiques, chimiques et électrostatiques des surfaces. De nombreux revêtements de surface ont donc été évalués par différentes équipes de recherche.

Schumann et Searle (1992) ont montré que la rugosité de la surface de poinçons revêtus de nitrure de chrome par bombardement ionique leur conférait une meilleure performance au collage comparativement à un poinçon non revêtu ou revêtu de chrome par électro-placage [68]. En utilisant des formulations à base d'acide ascorbique, Uemura et collab. (2007) ont également montré que les poinçons à surface durcis prévenaient mieux le collage que des poinçons revêtus de chrome par électro-placage [69]. Uchimoto et collab. (2013) ont rapporté

des résultats similaires quant à l'effet de l'irradiation par faisceau d'électrons et du durcissement des métaux [67]. Ces derniers ont procédé au durcissement de la surface d'un poinçon par décharge de carbure de tungstène. La surface d'un autre poinçon fut traitée par irradiation de faisceau d'électrons d'une couche de sulfure de tungstène, de fluorure de sulfure et de carbone. Ces traitements produisent des surfaces rugueuses avec une gamme d'aspérités de 3 $\mu$ m et 2 $\mu$ m respectivement [67]. Un troisième poinçon plus lisse obtenu par électro-placage de chrome fut également évalué. Après 200 cycles de compression, la quantité de poudre collée sur la surface des poinçons traités par durcissement et par irradiation de faisceau d'électrons était de 60  $\mu$ g tandis que celle collée sur la surface du poinçon traité par électro-placage était de 176  $\mu$ g [67]. Cette performance relative des poinçons traités par bombardement ionique, durcissement ou irradiation de faisceau d'électrons fut consensuellement attribuée à la présence d'aspérités rugueuses sur leur surface qui réduiraient la surface de contact et les forces d'adhésion entre les poinçons et les poudres [67–69].

Plus récemment, Saddik et collab. (2021) ont montré que l'initiation du collage est à la fois API et surface de poinçon spécifique [22]. Leur étude a consisté en l'évaluation du collage de 5 API différents (l'acétaminophène, le salicylamide, l'IBU, le flurbiprofène et le kétoprofène) à la surface de 5 revêtements de poinçon différents (en acier S7 ou M340 ou revêtu de chrome dur-HCR-, de nitrure de zirconium-ZrN- ou de nitrure de chrome -CrN-). Le temps avant l'apparition du collage fut employé comme indicateur et les résultats obtenus montrent qu'aucun revêtement n'est capable d'éliminer le collage pour tous les API évalués [22]. D'une formulation à l'autre, la performance des revêtements est substantiellement différente. Par exemple, il faut près de 15 minutes pour voir du collage du salicylamide sur une surface revêtue de HCr, tandis qu'il y apparaît 10 fois plus vite en moyenne ( $\cong$ 1,5 minute) pour les autres API [22]. Cette interaction entre la surface du poinçon et le type d'API n'est pas surprenante, car suivant le modèle conceptuel du collage (section 5.1), ce sont les interactions intermoléculaires entre API et poinçon qui gouvernent l'initiation du collage. La variété de performance des revêtements en fonction de l'API avait aussi été rapportée par Reed et collab. (2015) [19]. Cette étude avait brossé un portrait global de la performance des revêtements de poinçon à réduire le collage, car 24 revêtements différents furent évalués avec 5 formulations différentes et les profils de collage obtenus pour chacun des poinçons montrent une très grande variabilité de performance à réduire le collage des 5 formulations étudiées. De plus, pour une des formulations spécifiquement, aucun des 24 revêtements de surface ne permettait de réduire substantiellement le collage [19].

Outre l'interaction avec l'API, d'autres effets d'interaction avec le revêtement de surface des poinçons ont été reportés, notamment ceux avec le lubrifiant, la force de compression ou encore l'âge des poinçons [11–14,17]. Al-Karawi et collab. (2017) avaient évalué la performance au collage de quatre types de revêtements de poinçons après une période d'utilisation courante de trois ans [14]. Les résultats ont montré que pour des poinçons revêtus de nitrure de chrome ou de chrome dur, la propension au collage ne varie pas au bout des trois ans d'utilisation tandis qu'elle augmente substantiellement pour des poinçons non revêtus ou revêtus de nitrure de titanium [14]. Les auteurs attribuent la sensibilité au collage de ce dernier à sa faible résistance à l'usure par corrosion.

Quant à l'effet d'interaction entre le revêtement de surface et la force de compression, les données de la littérature concordent et montrent que l'effet des revêtements de surface n'est observable qu'à faible force de compression. Sous une forte force de compression, les différences en termes de propension au collage s'estompent [10–14,29]. Cette interaction est expliquée par la plus grande cohésion au sein du comprimé résultant de l'application d'une forte contrainte. Al-Karawi et collab. (2018) ont investigué le collage de formulations à base d'IBU et d'IBU-Na sur différents types de revêtement de poinçons [13]. Pour chacune des formulations, des plans d'expérience ont été préparés pour tester simultanément l'effet de la force de compression, de la vitesse de compression et de 5 différents types de revêtements de poinçons (U, CrN, TiN, HCr, Carbone de type diamant pour DLC). Après production de 27 comprimés, le collage fut quantifié par HPLC et les résultats montrent que les tous les trois paramètres étudiés, mais aussi plusieurs de leurs interactions ont un effet significatif sur le collage de l'IBU et de l'IBU-Na [13]. De façon globale, Les revêtements en CrN et TiN se révèlent les plus performants tandis que les revêtements U, HCr et DLC sont peu efficaces dans la réduction du collage de l'IBU et de l'IBU-Na. Les différents revêtements de cette étude ont été soumis à des analyses d'AFM, d'angle de contact et de texture. La bonne performance des revêtements de CrN et TiN a ainsi été reliée à de faibles interactions AFM et d'angles de contact, mais aussi une bonne homogénéité et luisité de la surface des poinçons. A contrario, de fortes interactions AFM et angles de contact ainsi qu'une surface hétérogène et rugueuse sont reliés aux faibles performances anti-collage des revêtements en HCr, DLC et U [13].

L'ensemble des données de la littérature montrent ainsi que le revêtement de surface des poinçons exerce une grande influence sur la propension au collage des poudres pharmaceutiques. C'est l'une des pistes privilégiées dans la littérature pour tenter de réduire l'adhésion. Malheureusement, la multiplicité et la variété de ses interactions empêchent un

recours systématique. De plus, cela complexifie l'étude du collage, où les conclusions d'investigations utilisant différents revêtements de poinçons sont difficilement comparables.

### **5.5 Les paramètres relatifs à l'environnement :**

Plusieurs rapports de la littérature ont discuté de l'implication de l'humidité relative et de la température sur la propension au collage des poudres pharmaceutiques. Par exemple, il est systématiquement reporté que la faible température de fusion de l'IBU jouerait un rôle important dans sa propension au collage [11–13,20,29,70,71]. Chatteraj et collab. (2018) rapportent que des augmentations localisées de température dues à des frictions interparticulaires pourraient monter à plus de 25 °C au-dessus de la température ambiante [9], créant un différentiel de température pouvant contribuer à l'adhésion des particules grâce à des ponts capillaires résultant de la condensation. Ces importantes augmentations de température pourraient aussi favoriser le collage à travers des phénomènes de fusion rapide à la surface des poinçons. En dépit de ces nombreuses hypothèses, il n'existe pour l'heure aucune donnée scientifique démontrant une implication directe de la température dans le collage, surtout pour ce qui est de la température ambiante. L'impact de cette dernière est probablement relié au mécanisme par lequel l'humidité relative module le collage des poudres. En effet, les données de la littérature montrent que l'humidité relative favorise le collage à travers la formation de ponts capillaires à la surface des poinçons [20,21,24,39,72–74].

Danjo et collab. (1997) ont rapporté que l'augmentation du contenu en eau de poudres à base de lactose influait leur propension au collage [74]. En utilisant plusieurs temps de séchage après granulation, les auteurs avaient réussi à simuler cinq niveaux différents de contenu en eau de la poudre (1,9 %, 2,1 %, 2,5 %, 3 % et 6,2 %). L'impact de ces derniers sur l'adhésion de la formulation aux poinçons a été évalué par la mesure de la force de détachement des comprimés du poinçon inférieur et la rugosité de la surface des comprimés produits. Les résultats rapportés par Danjo et collab. (1997) montrent que la propension au collage augmente avec le contenu en eau des poudres de lactose jusqu'à 3 %, puis il diminue pour des niveaux supérieurs d'eau. Les auteurs postulent que le contenu en eau des poudres renforce la force d'adhésion entre les particules et le poinçon grâce à des interactions capillaires, mais à des niveaux importants, le contenu en eau agirait comme un lubrifiant et limite le collage [74]. Le contenu en eau des poudres peut être modulé par plusieurs paramètres lors de la fabrication des comprimés, notamment la quantité des agents de mouillage lors de la granulation, le temps, la température et l'efficacité du séchage, mais aussi l'absorption d'humidité relative durant de

longs procédés de fabrication avec plusieurs heures de stockage d'intermédiaires des différentes unités d'opération.

Bunker et collab. (2011) ont étudié sous différents niveaux d'humidité relative, les forces d'adhésion entre une particule de lactose et différents revêtements de surface des poinçons (CrN, DLC et HCr) [20]. La même particule de lactose fut utilisée pour tester par AFM l'effet d'une humidité relative à variation progressive ascendante et descendante (10 %-30 %-60 %-30 %-10 %) sur l'adhésion aux différentes surfaces de poinçons. Les résultats obtenus montrent de façon globale que l'augmentation du niveau d'humidité relative est proportionnelle à la force de l'adhésion des particules à toutes les surfaces de poinçons testées [20]. Cette relation proportionnelle entre l'humidité relative et l'adhésion est attribuée à la formation de ponts capillaires de façon concomitante avec l'augmentation de l'humidité relative. En effet, plusieurs travaux ont rapporté qu'au-delà du niveau d'humidité relative critique, l'absorption de l'eau dans l'atmosphère environnante permet la formation de ponts liquides dans les interstices entre les particules et la surface des poinçons [20,75,76].

Le phénomène est donc aussi dépendant de la température et des propriétés des matériaux en présence, particules et poinçons. Pour ces derniers, Bunker et collab. (2011) rapportent un effet d'interaction [20]. L'effet de l'humidité relative change en fonction du revêtement de la surface des poinçons. Le collage aux revêtements en HCr et en DLC est plus sensible à l'humidité relative que la surface revêtue de CrN. Pour cette dernière, une augmentation de 10 % à 30 % d'humidité relative n'a aucun impact sur la force d'adhésion, mais une forte augmentation est observée (24 nN à 90 nN) lorsque l'humidité relative passe de 30 % à 60 % [20]. En face, on note des forces d'adhésion largement plus élevées dès les plus faibles niveaux d'humidité relative pour les deux autres revêtements de surface (HCr et DLC). Par exemple, l'adhésion au HCr augmente de 85 nN à 115 nN et à 135 nN lors de la progression ascendante de l'humidité relative [20]. Les mêmes sensibilités sont aussi observées dans la course descendante de l'humidité relative où la force d'adhésion mesurée baisse très rapidement pour la surface revêtue de CrN, tandis qu'elle baisse très lentement pour la surface revêtue de HCr et de DLC. Pour cette dernière, l'adhésion ne baisse pas dans le délai imparti de l'expérimentation (30 minutes par niveau d'humidité relative). Cet effet d'interaction entre l'humidité relative et le revêtement des poinçons est attribué aux propriétés hygroscopiques des surfaces [20]. En mesurant l'angle de contact avec l'eau des revêtements utilisés dans leur étude, Bunker et collab. (2011) ont montré que la surface revêtue de CrN était la plus hydrophobe comparativement aux surfaces revêtues de HCr et de DLC, qui sont hydrophiles.

Les ponts liquides peuvent donc se former même à de faibles niveaux d'humidité relative dépendamment de l'hygroscopie de la surface. Ces résultats cadrent bien avec les travaux sur le revêtement des poinçons décrits dans la section 5.4.3 précédente. L'inefficacité des revêtements de HCr à prévenir le collage comme rapporté par plusieurs auteurs pourrait ainsi s'expliquer par leur profil hydrophile qui favorise l'adhésion.

Plus récemment, Badal et collab. (2017) ont rapporté des conclusions similaires [21]. Ces derniers avaient en outre démontré le rôle important de la topographie des interfaces particules de poudre et surface de poinçons dans la formation des ponts aqueux conduisant au collage. Ce rôle est essentiellement gouverné par le rayon de kelvin, qui est le rayon de pore le plus grand dans lequel la condensation capillaire peut se produire. Badal et collab. (2017) ont démontré que des interfaces de surfaces rugueuses à aspérités locales plus grandes que le rayon de kelvin limitaient la formation de ponts capillaires [21]. À l'inverse, pour des particules d'API à surface moins rugueuse comme l'IBU, les aspérités locales formées à l'interface avec les poinçons sont plus faibles que le rayon de kelvin, favorisant la formation de ponts capillaires et donc plus de collage en fonction de l'humidité relative, mais aussi de la température [21].

En résumé, les paramètres environnementaux de la compression jouent aussi un grand rôle dans le collage à travers des mécanismes de fusion, d'adsorption d'eau ou de condensation rapides à l'interface entre les particules et les poinçons. Les interactions entre l'effet de l'humidité relative et les propriétés de surface des particules et des poinçons contribuent aussi au caractère multivarié du phénomène de collage des poudres pharmaceutiques. Ainsi, les études pour tenter de comprendre le collage doivent aussi nécessairement prendre en compte les paramètres environnementaux.

## **6. Synthèse de l'état de l'art sur le collage et introduction du projet de thèse : la carte mentale du collage**

Pour bien illustrer la problématique de la présente thèse, nous avons résumé dans la Figure 5 ci-dessous l'ensemble des connaissances disponibles actuellement (section 5) pour expliquer le collage : la carte mentale du collage. Cette proposition originale permet d'appréhender visuellement la complexité du phénomène de collage.

La carte mentale comprend quatre groupes de paramètres (formulation, poinçon, méthode et environnement), ainsi que deux super-paramètres dont tout changement impacte fortement l'ensemble des équilibres conduisant au collage : le type d'API et le type de revêtement de la surface des poinçons. Les relations déclinées en vert réduisent le collage, lorsque la cause source augmente et les relations déclinées en rouge font augmenter le collage, lorsque la cause source augmente. Une flèche noire pleine indique les relations à cause qualitative ou quand l'impact n'est pas complètement linéaire. Une flèche noire en pointillé indique les différents effets d'interaction. Un système de numérotation composé de la première lettre du groupe d'appartenance des paramètres et d'un chiffre séquentiel permet de les identifier au travers de la carte mentale.

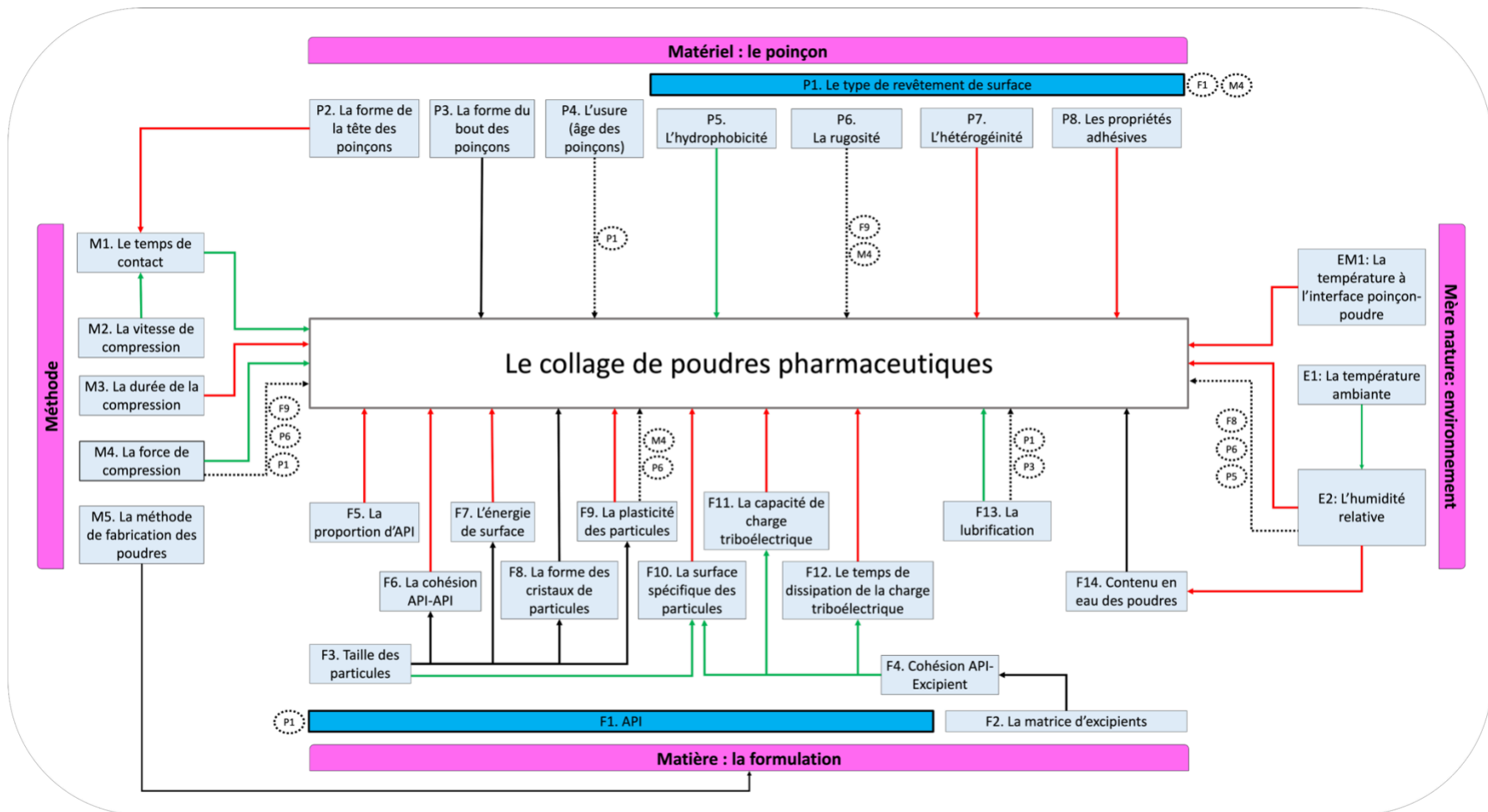


Figure 5 : Carte mentale du collage.



Suivant la carte mentale (Figure 5), le collage est la résultante d'un équilibre de forces déterminé par les propriétés des particules et des formulations, les propriétés de surface des outillages, ainsi que les paramètres opératoires et environnementaux des procédés.

Tout d'abord, du point de vue de la formulation, ce sont les particules d'API qui jouent le plus grand rôle. Ainsi, l'augmentation de leur proportion dans la formulation contribue à plus de collage. Cette susceptibilité des particules d'API au collage est principalement due à leurs propriétés électrostatiques (capacité de charge et temps de dissipation de celle-ci) et leurs propriétés physicochimiques (énergie de surface, forme cristalline, plasticité et surface d'échange spécifique). Une augmentation de toute composante de ces propriétés favorise l'occurrence du collage. L'effet de la taille des particules, quant à elle, dépend de son impact sur ces propriétés physicochimiques. À cause de la susceptibilité du collage à autant de propriétés des API, celui-ci est désigné comme super paramètre. En effet, d'un API à l'autre ou même en fonction de la méthode de production des poudres, les propriétés physicochimiques et électrostatiques ci-dessus peuvent changer en partie ou en totalité dans des proportions plus ou moins importantes. Ceci a pour impact de modifier significativement tout l'équilibre des forces conduisant au collage. De plus, l'effet de l'ensemble de ces propriétés varie en fonction du type de revêtement de la surface des poinçons. Il peut donc être affirmé que le collage est spécifique aux propriétés des API et de la surface des poinçons : une combinaison particulière de ces propriétés est nécessaire à l'occurrence du collage. Outre l'API, d'autres composants de la formulation peuvent moduler le collage. Conformément au modèle conceptuel, une forte cohésion intrinsèque API-Excipient permet ainsi de réduire le collage. De même, la lubrification a généralement un effet bénéfique sur le collage, mais celui-ci peut varier en fonction du type de revêtement de surface des poinçons et la forme de leur bout. Le contenu en eau des poudres impacte également l'occurrence du collage, mais comme discuté plus haut, son effet est non linéaire.

Du point de vue des poinçons, ce sont les propriétés de surface qui jouent le plus grand rôle. Les propriétés adhésives par interaction électrostatique et l'hétérogénéité de la surface des poinçons favorisent ainsi le collage, tandis que son hydrophobicité le réduit substantiellement. La rugosité de la surface des poinçons influe également sur le collage, mais son effet dépend de la force de compression et des propriétés mécaniques des particules d'API, notamment la plasticité. Le changement du type de revêtement induit ainsi un grand changement dans l'équilibre des forces conduisant au collage, d'où sa désignation comme super-paramètre dans la carte mentale du collage. De plus, il est impliqué dans de nombreuses interactions,

notamment avec les propriétés des API, la lubrification, la force de compression ou encore l'usure des poinçons qui n'ont pas le même effet sur le collage selon le type de revêtement. Par ailleurs, la forme des poinçons peut également moduler l'occurrence du collage. Différentes formes de bouts de poinçons ou de gravures peuvent par exemple favoriser le collage. À l'opposé, la forme de la tête des poinçons module le temps de contact et par conséquent permet de réduire le collage.

Ensuite, du point de vue des paramètres opératoires, la vitesse et la durée de la compression aggravent le collage, d'où la prééminence du phénomène à l'échelle industrielle à la suite de longues séances de compression à haute fréquence. L'augmentation de la vitesse de compression s'accompagne d'une réduction du temps de contact, ce qui diminue la cohésion interne des comprimés et favorise donc le collage. Les longues séances de compression quant à elles contribuent à des augmentations importantes et localisées de température à l'interface poinçon et poudre, favorisant également le collage. À l'opposé, l'augmentation de la force de compression permet généralement de réduire le collage, mais de nombreuses interactions subsistent, notamment avec les propriétés du revêtement de surface des poinçons et les propriétés mécaniques des particules. Un autre paramètre opératoire très influent également dans l'occurrence du collage est la méthode de production des poudres, notamment les temps de mélange, de granulation et de séchage qui peuvent moduler plusieurs des propriétés physicochimiques et électrostatiques des API et de la formulation.

Finalement, d'un point de vue environnemental, l'humidité relative est le paramètre principal qui influe sur l'occurrence du collage, notamment en favorisant le contenu en eau des poudres et la formation de ponts capillaires à l'interface poinçon et poudre. Cet effet délétère de l'humidité relative dépend de plusieurs autres paramètres, notamment l'hydrophilicité de la surface des poinçons, leur rugosité et la forme des cristaux des particules. Ces paramètres conditionnent le développement plus ou moins important des interactions capillaires à l'interface entre les poinçons et la poudre, notamment à travers l'angle de contact et le rayon de kelvin. L'humidité relative étant indissociable de la température ambiante, elles agiraient de concert pour favoriser le collage.

De cette description globale et complète de l'état de l'art des causes du collage, nous pouvons conclure que celui-ci est un phénomène multifactoriel dont l'occurrence est gouvernée par un équilibre dynamique et fragile entre de très nombreux paramètres aux effets interdépendants (interactions). L'étude, la compréhension et la résolution d'un tel phénomène multivarié requièrent une approche intégrative et simultanée où les paramètres aux effets

interdépendants doivent être investigués de concert. La conduite de telles expérimentations simultanées à échelle laboratoire serait très lourde financièrement. De plus, les modèles d'étude du collage à échelle laboratoire souffrent de problèmes de reproductibilité et de représentativité à cause des faibles vitesses et durée de compression. Ils sont donc assez limités pour décrire de façon globale et complète le collage d'une formulation donnée. Cependant, une quantité importante de données est déjà générée lors de la fabrication des comprimés à l'échelle industrielle. Une exploitation effective de ces données permettrait d'ouvrir de nouvelles perspectives dans l'étude, la compréhension et la résolution du collage. La présente thèse s'intéresse à cette possibilité et se fixe donc comme objectif d'introduire, de définir, d'évaluer et de discuter une nouvelle approche d'étude du collage à travers l'analyse de données industrielles. Celle-ci est basée sur l'utilisation des techniques d'analyse multivariée. Une introduction à ces derniers et leur intérêt dans l'étude du collage à échelle industrielle sont présentés dans la prochaine section.

## **7. Les techniques d'analyse multivariée et leur intérêt dans l'étude du collage à échelle industrielle :**

La nouvelle approche d'étude du collage à échelle industrielle que nous introduisons dans la présente thèse est basée sur les techniques d'analyse multivariée. Ce choix est principalement dû à l'adéquation entre la nature spécifique des données générées à échelle industrielle et les propriétés spécifiques de ces techniques d'analyse de données. En effet, une quantité très importante de paramètres est mesurée lors de la production industrielle des comprimés et pour un phénomène aussi multifactoriel que le collage, il est essentiel d'analyser ces paramètres ensemble afin de tenir compte des multiples liens d'interdépendance. Les techniques d'analyse univariée comme la régression linéaire simple ou les analyses de la corrélation ne sont donc pas adaptées au contexte de la présente thèse. Nous recourons aux techniques d'analyse multivariée qui sont un ensemble de méthodes permettant l'exploration, la modélisation et l'analyse de jeux de données décrits par plusieurs variables qualitatives et/ou quantitatives. Elles peuvent être objectivement divisées en deux sous-groupes : les méthodes multivariées classiques et les méthodes multivariées à réduction dimensionnelle [77].

Les méthodes multivariées classiques sont les techniques d'analyse de données les plus connues et traditionnellement employées pour les grands jeux de données décrits par plusieurs variables. Il s'agit entre autres de la régression linéaire multiple, l'analyse discriminante linéaire, l'analyse des correspondances multiples, l'analyse canonique des corrélations et bien

d'autres algorithmes d'apprentissage machine. Les méthodes multivariées classiques utilisent directement les variables prédictives originelles des jeux de données pour l'exploration et le développement de modèles prédictifs. Elles sont bien adaptées à l'analyse de données industrielles, mais peuvent avoir des limites sous certaines conditions, notamment lorsque le nombre de variables prédictives dépasse largement le nombre d'observations, ce qui est très fréquent dans la production pharmaceutique industrielle.

En effet, en plus du nombre important de paramètres mesurés, la production industrielle consiste également en l'exécution d'un faible nombre de lots de grande taille. Cela produit des matrices de données rectangulaires avec beaucoup de variables et très peu d'observations. Dans ces cas, les méthodes multivariées classiques manquent d'assez de degrés de liberté pour estimer tous les coefficients des modèles. De plus, la variance de ces coefficients sera impactée par des problèmes de multicollinéarité [77,78]. Sur un plan physique, la multicollinéarité se traduit par l'existence dans un jeu de données de variables prédictives mesurant ou rapportant le même phénomène, c'est-à-dire des prédicteurs non indépendants. Cela est très courant dans de larges jeux de données de procédés chimiques où de nombreux paramètres interdépendants subsistent. L'exemple le plus simple et le plus frappant est la relation entre la température et l'humidité relative. Sur un plan purement mathématique, la multicollinéarité implique qu'une ou plusieurs variables prédictives sont fortement corrélées ou encore qu'une variable prédictive est potentiellement la combinaison linéaire de plusieurs autres. Les matrices de données rectangulaires de données sont propices à ce phénomène de multicollinéarité, car le nombre de paramètres croissant pour un nombre d'observations limité augmente les probabilités de collinéarité purement mathématique entre les variables. En présence d'une forte multicollinéarité, les matrices de variance-covariance des jeux de données ne sont plus de rang plein, ce qui conduit à une augmentation de la variance des coefficients des modèles les rendant instables et difficiles à interpréter [77,79]. Beaucoup de coefficients apparaîtront ainsi non significatifs de façon artificielle et les estimations changeront fortement d'un échantillon de données à l'autre. En plus de ces problèmes d'insuffisance de degré de liberté et de multicollinéarité, certaines techniques multivariées classiques comme les réseaux de neurones produisent généralement des modèles de type « boîte noire » à faible pouvoir explicatif, servant plus un but exclusivement prédictif.

Les méthodes multivariées à réduction dimensionnelle prennent le relai lorsque ces limites des méthodes classiques se font sentir à cause de la rectangularité croissante des jeux de données industrielles. Il s'agit principalement de l'ACP pour la classification non

supervisée, l'analyse discriminante par régression sur les moindres carrés partiels qui permet de faire de la classification supervisée, la régression sur composantes principales (PCR) et la régression sur les moindres carrés partiels (PLS) qui permettent de développer des modèles multilinéaires robustes. La force de ces méthodes réside dans leur capacité à la réduction de la dimensionnalité des données. Elles permettent de résumer avec quelques variables latentes la majeure partie de l'information capturée par les nombreuses variables originelles d'un jeu de données. Ces nouvelles variables, appelées composantes, sont construites selon différents critères d'optimisation. Par exemple, le critère utilisé pour la construction des composantes de l'ACP est la maximisation de la distance euclidienne des projections, c'est-à-dire les dimensions restituant la plus grande variance de la matrice de données [77]. L'annexe 2 de la présente thèse illustre de façon détaillée le principe de construction des composantes de l'ACP. Elle permet de comprendre de façon ludique l'ensemble des autres méthodes. Par exemple, la PCR est la combinaison de l'ACP et de la régression multilinéaire où les composantes principales extraites par ACP sont utilisées comme prédicteurs dans un processus de régression multilinéaire. Cette dernière est également décrite en détail dans l'annexe 2. La PLS est quant à elle similaire à la PCR, à la différence que le critère d'extraction des composantes est différent de celui de l'ACP. En plus de la variance de la matrice des prédicteurs utilisée dans l'ACP, l'extraction des composantes par PLS prend aussi en compte la covariance entre les prédicteurs et la ou les variables à prédire [77]. La PLS-DA est une variante de type classification de la PLS où la variable à prédire est qualitative. Dans ce cas précis, une variable binaire à prédire est générée pour chaque modalité de la variable qualitative originelle et la matrice résultante est utilisée comme jeu de variables à prédire dans une procédure PLS classique [77,80].

Les propriétés des composantes construites par les méthodes multivariées à réduction dimensionnelle permettent de prévenir les limites rencontrées avec les méthodes multivariées classiques. En effet, quelle que soit la méthode employée, les composantes sont toujours extraites par ordre croissant d'importance par rapport au critère d'optimisation spécifique, si bien qu'il en faut juste quelques-unes pour décrire la majeure partie de la variance du jeu de données étudié. Se faisant, bâtir des modèles de classification ou de régression en utilisant les composantes comme prédicteurs à la place des variables originelles permet d'éviter le problème d'insuffisance de degré de liberté pour estimer les paramètres des modèles. Les composantes extraites sont aussi deux à deux orthogonales permettant ainsi de prévenir également les problèmes de multicollinéarité. Les plans formés par ces composantes revêtent aussi un grand pouvoir explicatif. De plus, le caractère itératif des algorithmes d'extraction des

composantes permet une bonne gestion et une bonne imputation des données manquantes, très fréquentes dans les jeux de données industrielles (voir annexe 3).

En utilisant les techniques d'analyse multivariée, nous avons démontré dans cette thèse la pertinence d'une nouvelle approche d'étude industrielle du collage à travers deux études de cas. La première s'est intéressée à l'investigation du collage d'une formulation à base d'ibuprofène où le collage était décrit de façon qualitative, tandis que la seconde étude s'est attelée à l'investigation du collage d'une formation différente à base d'ibuprofène et de metocarbamol où le collage était décrit de façon quantitative. Les deux prochains chapitres de la présente thèse présentent et discutent ces deux études de cas respectivement.

# CHAPITRE 3 : ÉTUDE MULTIVARIÉE QUALITATIVE DU COLLAGE

## Avant-propos

### Auteurs et affiliation :

Mahamadou Dembélé : étudiant au doctorat, Université de Sherbrooke, Faculté de génie, Département de génie chimique et biotechnologique.

Sophie Hudon : ingénieure automatisation, Global Technology and Engineering, Process Automation, Pfizer Canada.

Jean-Sébastien Simard : directeur, GPPD - Digital Design & Process Analytics, Pfizer Worldwide Research, Development & Medical.

Nicolas Abatzoglou : professeur agrégé, Université de Sherbrooke, Faculté de génie, Département de génie chimique et biotechnologique.

Ryan Gosselin : professeur titulaire, Université de Sherbrooke, Faculté de génie, Département de génie chimique et biotechnologique.

**Date d'acceptation :** 27 novembre 2022

**État de l'acceptation :** Publié

**Journal :** Pharmaceutical Development and Technology

**Titre anglais :** A multivariate data analysis approach to tablet sticking on an industrial scale: A qualitative case study of an ibuprofen-based formulation

**Titre français :** Une nouvelle approche d'étude du collage par analyse multivariée des données de production industrielle : étude qualitative avec une formulation à base d'ibuprofène

**Contribution au document :** Cet article présente les résultats d'une étude de cas appliquant la nouvelle approche d'étude du collage proposé par la présente thèse de doctorat. Cette étude de cas a été réalisée avec une formulation à base d'ibuprofène où le collage est évalué de façon qualitative. L'article fait ainsi la preuve de concept de la nouvelle approche proposée dans cette thèse et il discute ses avantages, limites, bonnes pratiques et perspectives. Aussi, cet article introduit la méthodologie à employer dans le cas d'une description qualitative du collage.

## Résumé français

Une nouvelle approche d'étude du collage par analyse multivariée des données de production industrielle : étude qualitative avec une formulation à base d'ibuprofène

Le collage est l'un des problèmes les plus courants et les plus préjudiciables survenant lors de la fabrication des comprimés. C'est un phénomène qui consiste à l'adhésion de poudres pharmaceutiques sur la surface des outillages de compression. À cause du nombre important de paramètres impliqués dans son occurrence et leurs multiples interactions, l'étude et la compréhension du collage nécessitent une approche intégrative et simultanée où les paramètres aux effets interdépendants doivent être étudiés ensemble. La conduite de telles expérimentations simultanées à échelle laboratoire pourrait s'avérer très coûteuse. Dans la présente étude, nous évaluons une nouvelle approche d'étude du collage basée sur l'analyse des données déjà générées lors de la production industrielle des comprimés. Nous avons ainsi collecté et analysé les données de 71 lots industriels d'une formulation à base d'ibuprofène. Chaque lot est décrit par une centaine de variables, incluant un descripteur qualitatif du collage. En utilisant des techniques d'analyse multivariée, nous avons pu isoler un total de six variables prédictives pour leur implication dans le collage de la formulation étudiée. Il s'agit de la proportion de lubrifiant dans la formulation, la durée des arrêts en cours de compression ainsi que la température et l'humidité relative de certaines salles d'opération ou de stockage. Ces paramètres pointent le rôle de la lubrification et du contenu en eau des poudres ainsi que le faible point de fusion de l'ibuprofène dans sa propension au collage. En nous appuyant sur ces résultats, nous avons proposé et discuté une nouvelle approche d'étude du collage basé sur l'analyse de données industrielles, ainsi que sa méthodologie associée.

**Mots clés :** collage de poudres pharmaceutiques, compression des poudres, analyse multivariée des données, analyse en composantes principales, analyse discriminante linéaire et analyse discriminante par régression sur les moindres carrés partiels.



## Résumé en anglais

A multivariate data analysis approach to tablet sticking on an industrial scale: A qualitative case study of an ibuprofen-based formulation

Sticking is one of the most common and damaging issues that occur during tablet manufacturing. Sticking is the adhesion of powder onto tooling surfaces during compression. Because of the numerous factors involved in its occurrence, understanding tablet sticking requires the simultaneous investigation of these factors to clarify their possible interactions. However, conducting such a study experimentally can present a significant financial and technical burden. In this study, we aimed to leverage the large amount of data that is usually generated during industrial manufacturing to gain insights into sticking. This was achieved by collecting and analyzing a total of 71 historical batches that used an ibuprofen-based formulation. We associate each batch with a hundred parameters, including a qualitative descriptor of sticking, and employ a predefined methodology based primarily on multivariate data analysis. Our results highlight the role of lubrication, water content, and the low melting point of ibuprofen in its sticking tendency. Based on these findings, we propose and discuss an industrial manufacturing data analysis approach to sticking and its associated systematic methodology, consisting of collection, exploration, and data modeling.

**Keywords:** tablet sticking, powder compaction, multivariate data analysis, principal component analysis, linear discriminant analysis, and partial least squares-discriminant analysis.

## **1. Introduction :**

Tablets are the most prevalent form of medication [7,9]. This is primarily due to their physical and chemical stability under various conditions, as well as their ease of use and manufacturing [2,4,6]. Tablet production consists of compressing a defined quantity of powder formulation on a tablet press. The formulation is a powder blend of an active pharmaceutical ingredient (API) and excipients that are often dried or wet-granulated to enhance their compressibility and compactibility. Compression involves the application of a uniaxial force on the formulation powder that is embedded in the die cavity of the tablet press. Under pressure, the density of the powder blend is increased by expelling air from between the particles [5,8]. The powder then consolidates via a series of plastic and elastic deformations and particle fragmentations, leading to the formation of bonds between them [5,31].

One of the most common issues that arise during this process is the adhesion of powder onto the surfaces of the press tooling (die and punches); this is referred to as sticking [9]. This phenomenon can occur at any stage of drug manufacturing, but it is more prevalent and damaging during the industrial production phase, during which, possible changes to the formulation and process are limited [9]. When sticking appears at this stage, it causes tablet defects and reduces quality and yield. Operators generally stop the press and then remove, clean, and polish the tooling before continuing production. Thus, sticking significantly affects manufacturing costs and production times. It also limits the age of compression tools because of premature wear [5,31]. Because of these serious consequences for productivity, research has been performed on powder sticking, especially the adhesion of powder to the punch surface (referred to as punch sticking).

Sticking was first described as the result of two opposing forces: first, the adhesive forces between the formulation powder and punch surfaces, and second, the inner cohesive forces between the particles of the formulation powder [7,9]. This simple model states that particles detach from the powder bed and stick onto the punch surface when the strength of the adhesive forces between particles and the surface exceeds that of cohesive forces between the formulation powder particles.

In 2017, another model for sticking was introduced [46]. They assert that a model involving two opposing forces cannot explain several well-established observations about sticking: multiple layers of powder accumulate on punch surfaces and API particles make up a high proportion of the powder in these layers [36,37,46]. Based on their experimental results,

[46] propose a more complete model that involves two kinds of sticking and three forces. The first type of sticking (type I) corresponds to light sticking, in which only one layer of powder builds up on the punch surfaces, while the second type (type II) refers to major sticking, with several layers of accumulation. These two kinds of sticking arise from the interplay of three forces, designated as F1, F2, and F3. F1 represents the strength of adhesive forces between API particles and punch surfaces, while F2 is the strength of inner cohesive forces between API particles, and F3 denotes the strength of adhesive forces between API particles and the excipients matrix [46]. Sticking only occurs if F1 is greater than the sum of F2 and F3, when the adhesive forces of API particles to the punches predominates. The occurrence of light sticking (type I) or severe sticking (type II) would then depend on the strength ratio between F2 and F3. Accordingly, light sticking is observed if F2 is weaker than F3, in which case a single layer that is rich in API would cover punch surfaces. Severe sticking is observed when F2 is greater than F3. The intensity of these three forces is determined by various interactions, such as van der Waals forces, electrostatic interactions, capillary bonds, and mechanical locking, which all depend on the chemical and structural properties of the particles in the formulation (API and excipients) and the compression conditions, such as the properties of punches surfaces, the compression force, the compression speed, and the environmental conditions. Furthermore, the ratio between the three forces is dynamic and constantly evolving. Therefore, sticking is a complex phenomenon that involves various factors, particularly to evaluate the severity or level of sticking.

A broad range of factors have been investigated to elucidate the sticking mechanism, and clear effects have been found for some. Several studies with different methodologies and formulations have shown that sticking is primarily dependent on the API particles [37,46,57,62]. In many ways, such findings are predictable, as excipients are chosen for their suitability in manufacturing (mechanical, electrical, and chemical properties), whereas the API is chosen largely for its pharmacological properties. Paul, Taylor et al. (2017a) studied the sticking kinetics of 24 API formulations, finding that the stickiest formulations containing 10% API resulted in a stuck crust containing a 93–96 % concentration of API. Similar conclusions with other APIs and methodologies have been reported [35,37]. These observations were attributed primarily to the specific physical, chemical, and electrostatic properties of the API particles [35,37,46,54,57]. The results show that—due to their polarity, generally small size, and hydrophobic behavior—API particles are susceptible to triboelectric charging during handling, mixing, and compression [52,54]. This electrostatic behavior, especially the slow

dissipation of charge, has been identified as a key factor promoting sticking [52,55]. Because of their more suitable electrostatic behavior, excipients can enhance the electrostatic behavior of the whole formulation, thereby reducing sticking relative to API particles alone [53,54]. In addition to the central role of API particles in sticking, compression speed and duration can promote sticking [14,20,36,62,81].

Despite these clear findings, no root causes of sticking have been identified [9,10,26]. The reason is simple: sticking is multifactorial and involves several interactions between the parameters, which can also change over time. For instance, lubrication with magnesium stearate is known to help decrease sticking for numerous formulations (such as mannitol, pozanicline, and macrogol), but increases it for ibuprofen-based formulations [11,12,38,51]. Another example is the effect of compression force. An increase in the compression force leads to increased sticking for celecoxib or sorbitol-based formulations, but it decreases sticking in acetylsalicylic acid-based formulations [36,62,64]. In addition, for ibuprofen-based formulations, the effect of compression force also changes according to the properties of punch surfaces [13,14,37,51,63].

Therefore, recent studies have focused on the simultaneous study of several parameters to understand sticking [10,26]. Paul, Taylor et al. (2017a) simultaneously assessed the effect of powder properties and compression parameters on the sticking of 24 different API formulations. They identified the surface area of the API particles, tablet tensile strength, ejection force normalized by particle size, take-off force normalized by tablet tensile strength, and the die wall pressure. Studies by Roberts [11,12] and Al-Karawi [13,14] also illuminate the role of compression forces in sticking for ibuprofen-based formulations. These studies describe the significant interactions between the compression force and several properties of the punch surface. An increase in the compression force led to more sticking when the punch surface was smooth, while no effect on sticking was observed when the punch surface was rough [11,12]. Punch coating only appears to influence sticking for low compression forces, where uncoated or hard chromium-coated punches generated more sticking than punches coated with titanium nitride or chromium nitride, whereas no difference in their sticking propensity was observed for high compression forces [11–14,68,69]. This interaction was attributed to the increased internal cohesive forces induced by the high compression force, which subsequently overcome the specific adhesive forces between powder particles and each coating [11,12,20].

More multivariate studies that simultaneously investigate the effects of various parameters are necessary to fully understand tablet sticking. However, performing a dedicated experimental study can present a significant technical and financial burden. Indeed, increasing the number of parameters under study inevitably requires a greater number of experiments. Some parameters are also more difficult to obtain and test experimentally, such as punch age and wear, long compression times, high compression speeds, and low dwell times. Nevertheless, substantial data is usually generated during industrial-scale tablet manufacturing, where sticking is inherently prevalent. Accounting for all this data comprehensively may provide insights into the causes of tablet sticking.

In this study, we introduce an industrial data analysis approach to examine tablet sticking. We present a qualitative case study, followed by a deep discussion on the overall approach and methodology. In the case study, we investigate the sticking of an ibuprofen-based formulation by collecting and analyzing industrial manufacturing data. A total of 104 variables were gathered from the production of 71 historical batches, for which sticking behavior was assessed directly as a categorical variable (with either the presence or absence of sticking during compression). Our methodology is based primarily on multivariate data analysis (MVDA) techniques. The specific objectives of this work are expressed as follows:

- (1) Identifying the key parameters contributing to the occurrence of sticking using an ibuprofen-based formulation.
- (2) Developing a classification model to predict the occurrence of sticking on a batch during its compression.
- (3) Discussing the industrial data analysis approach and methodology to provide insights into tablet sticking.

## **2. Materials and methods:**

### **2.1 Formulation :**

The active pharmaceutical ingredient of the formulation is ibuprofen, a common non-steroidal anti-inflammatory drug. The other substances present in the formulation include aspartame, mannitol, microcrystalline cellulose, colloidal silicon dioxide, magnesium stearate, modified corn starch, natural and artificial flavors, sodium starch glycolate, and coloring agents. Tablets are manufactured through a direct compression process involving mixing and sieving operations before compression using a tablet press. Due to confidentiality concerns, no further details regarding the manufacturing process are provided.

### **2.2 Data collection and preparation:**

During the industrial manufacture of these confidential-formulation tablets, copious volumes of data regarding the processing steps are generated. For this study, we collected this data from various sources such as master batch record files, certificates of analysis for raw materials, and the records on temperature and relative humidity in the operating rooms and warehouses. A total of 104 variables were gathered for 71 historical batches.

The raw data were cleaned and prepared before analysis. First, some variables were found to be nearly constant or to present a high proportion of missing data ( $> 50\%$ ). Therefore, these were discarded. For quantitative variables, the coefficient of variation (ratio of the standard deviation to the mean) was employed to set a low variation criterion (1%). Thus, quantitative variables displaying a coefficient of variation of less than 1% were discarded from the data. For qualitative variables, the low variation criterion was set as the presence of a maximum of two observations among all batches with modalities different from the mode of the concerned variable.

Second, after the cleaning operations, only a few missing values remained in the data. Notably, the main analysis algorithms employed in this study—namely, the principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA)—can be adapted to account for the presence of missing data, and thus, the few missing values were imputed from the mean of the corresponding variable when necessary.

The final dataset consists of 71 batches described by a set of 46 variables, referred to herein as V1 to V45, plus a qualitative  $y$  variable describing the sticking behavior of each batch. This  $y$  variable separates the 71 batches into two classes consisting of batches that were not

affected by sticking and those that were affected. The first class is composed of batches for which no sticking was observed during compression, and the second class consists of batches for which sticking issues were observed during compression. Except for the qualitative  $y$  variable, all the remaining 45 variables were quantitative. These can be divided into five different groups:

- Group 1 : V1–8 are related to materials (API + excipients).
- Group 2 : V9–16 give information about the mixing step, including the cleaning methods used on machines and the durations of various mixing operations.
- Group 3 : V17–36 relate to the environmental conditions in different rooms where manufacturing operations, from dispensing to compression, take place.
- Group 4 : V37–39 represent the tablet dimensions and weight.
- Group 5 : V40–45 relate to the parameters of compression operations.

During the collection campaign (encompassing 71 batches), environmental conditions variables (V17–36) were unavailable for some batches (batches 1–46). To study the impact of these environmental variables, two datasets were generated: dataset1 and dataset2. Dataset1 includes all batches and all variables except those related to the environmental conditions while dataset2 only comprises batches for which data was available for all variables including environmental conditions. The resulting sizes of both datasets were 71 batches x 26 variables (dataset1) and 25 batches x 46 variables (dataset2). These were scaled and centered to unit variance prior to analysis.

### **2.3 Analysis methodology:**

Both dataset1 and dataset2 were analyzed in three main steps, using the Python programming language (version 3.7) along with MVDA software, namely SIMCA (version 16.0.2.10561) from Sartorius.

The first step involves univariate and bivariate analysis of both datasets to determine whether any single variable could significantly explain or predict the observed sticking behavior. For this purpose, scatter plots of input variables are colored according to the occurrence of sticking.

The second step consists of a multivariate investigation of both datasets using PCA to check for the presence of clusters of observations representing sticking. PCA is a well-established multivariate method employed to reduce the dimensionality of large datasets by

capturing the highest amount of variation in the datasets in a few orthogonal components. It facilitates the visualization and investigation of similarities between observations, following linear combinations of the input variables along the principal components. PCA was well-described and discussed by [77,82,83].

The third step consists of discriminant analysis (DA). Linear discriminant analysis (LDA) is applied to dataset1, while the DA version of partial least squares regression (PLS-DA) is used for dataset2. These analyses use a supervised approach to yield insights into sticking.

LDA is a classification technique that allows for the development of multivariate models to classify observations among modalities of a qualitative  $y$  variable. It is aimed at finding a linear combination of the original predictive variables that separate the different modalities of the  $y$  variable. This modeling technique is generally referred to in the literature as factorial DA or canonical DA (when the objective is strictly descriptive). For predictive purposes, LDA modeling provides a classification function for predicting the membership of a new batch among the modalities of the qualitative  $y$  variable. Both descriptive and predictive attributes of LDA have been reviewed and discussed previously [80,83–86].

For dataset2, there are more variables than the number of observations available, which precludes the use of LDA because the degree of freedom is too low to proceed, also owing to the risk of multicollinearity issues [78,79,87–89]. Instead of LDA, we employ PLS-DA, a different technique for discrimination. Broadly speaking, this technique has the same objective as LDA, but the extracted components for the descriptive model and the independent variables used for predictive modeling are different. PLS-DA employs latent variables extracted from the original predictor matrix by PLS, a well-known multivariate procedure. These latent variables facilitate the visualization and description of the discrimination as per descriptive outcomes of LDA. However, the latent variables are also used as predictors in PLS-DA predictive modeling, instead of the original variables for LDA. PLS-DA has been extensively described previously [77,80,83].



### **3. Results and discussions :**

#### **3.1 Sticking data :**

When sticking occurs during the compression of a batch, operators generally stop the press, note the cause of the interruption, assess the situation, and fix the issue before restarting. Since the operators did not specifically report the severity of individual sticking events, we utilized their notes to set up a qualitative  $y$  variable to assess the sticking behavior of batches. According to this qualitative  $y$  variable, batches were split into two classes: batches unaffected by sticking and batches that were affected by sticking (as assessed by the operators). Table 2 presents a breakdown of the number of observations in both datasets.

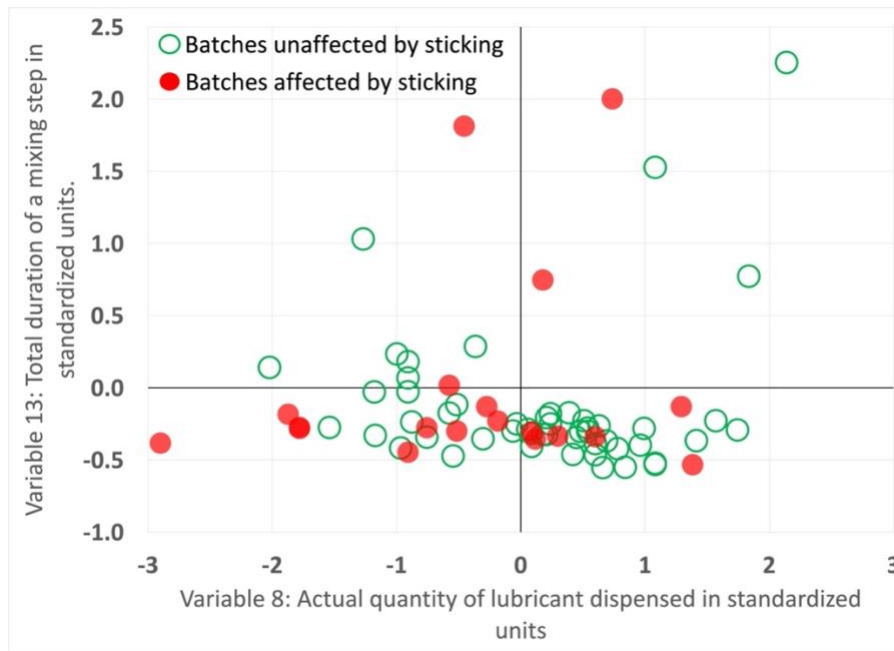
*Table 2 : Repartition of the number of batches following the qualitative  $y$  variable modalities.*

| Datasets | Modalities of the qualitative $y$ variable |       |                              |       | Total |
|----------|--|-------|------------------------------|-------|-------|
|          | Batches unaffected by sticking             |       | Batches affected by sticking |       |       |
| Dataset1 | 51   | 71.8% | 20                           | 28.2% | 71    |
| Dataset2 | 18   | 72%   | 7                            | 28%   | 25    |

#### **3.2 Univariate data analysis :**

##### **3.2.1 Dataset1 :**

Scatter plots of the input variables were color-coded according to the presence of sticking during the compression ( $y$  variable). An excerpt of this representation is given in Figure 6. These variables (V8 and V13) were chosen because they exhibit the highest correlation with the  $y$  variable.



*Figure 6 : Scatter plot of input variables in dataset1 colored according to the presence of sticking (y variable).*

*This excerpt corresponds to V8 and V13, which are the actual quantity of lubricant dispensed and mixing step duration, respectively.*

Examining the scatter plots of individual variables reveals no clear trend of discrimination between batches affected or unaffected by sticking (Figure 6). Thus, no single variable—nor any of the bivariate combinations assessed—can explain the occurrence of sticking. Furthermore, univariate data analysis is used to check for outliers from each variable. To this end, we identified several possible outliers. However, after evaluating different raw data sources, no data reporting error was noted, and all of the values were compliant with reglementary standards when needed. Therefore, no action has been taken to treat or remove them. This decision is also motivated by the fact that none of these outliers were physically unreasonable considering the reality of a day-to-day tablet manufacturing processes.

### **3.2.2 Dataset2:**

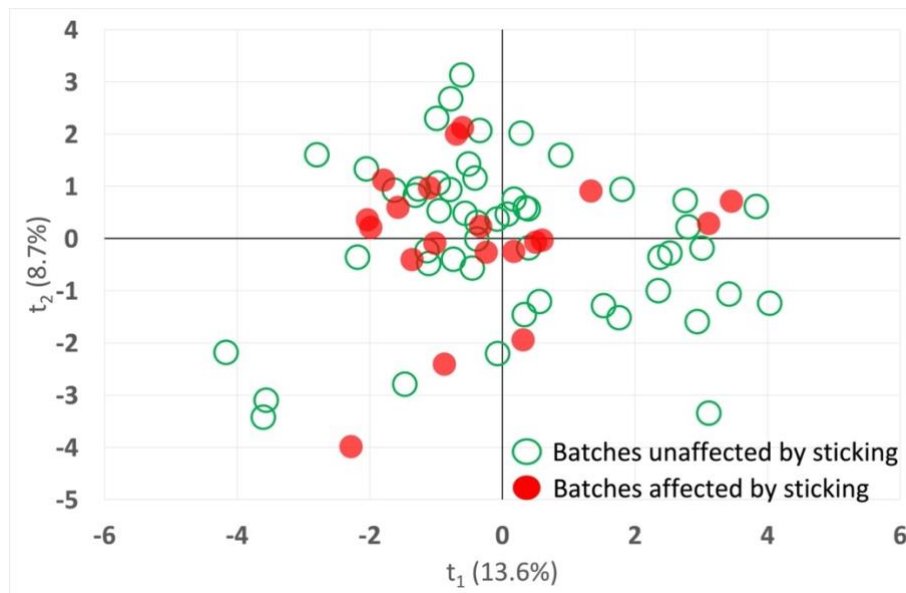
For dataset2, only scatter plots of variables representing environmental conditions (V17 to V36) were considered, since the remaining input variables were already assessed in dataset1. As with dataset1, no clear trend of discrimination was observed between the batches affected by sticking and those unaffected by sticking. Thus, no single variable related to environmental conditions can explain the occurrence of sticking. Several possible outliers were identified, but after verification of the raw data, they were also kept in data for the same reasons described for dataset1.

The results of the univariate analyses may be expected because sticking is known to be a highly multivariate phenomenon [9,10,13]. Consequently, combined information from several variables is required. However, it remains important to systematically use univariate data analysis to fully understand the data and to treat eventual outliers when they reflect errors or could induce noise in the data. To yield further insights regarding sticking, we employed MVDA techniques, beginning with PCA followed by DA.

### 3.3 Principal component analysis:

#### 3.3.1 Dataset1:

The first PCA model was built on dataset1. According to an analysis of the scree plot, a total of 10 components were exploited. These account for approximately 70% of the variance in dataset1. Figure 7 displays the score plot of the two first components of this PCA model.



*Figure 7 : The score plot of the PCA model built on dataset1.*

In the score plot, the batches affected and unaffected by sticking overlap (Figure 7). The same conclusion can be drawn from the analysis of the subsequent eight PCA components. Therefore, this first PCA model does not reveal any clear discrimination according to the observed sticking in dataset1. This does not mean that dataset1 cannot be used to explore sticking. It simply means that these first components explain other phenomena that account for more variance within the data than sticking.

### **3.3.2 Dataset2:**

A PCA model was also built on dataset2, including all available information. A total of eight components were retained according to the analysis of the scree plot, but no clear trend regarding sticking was found in dataset2.

The failure of PCA to explain the observed sticking is unsurprising. PCA is an unsupervised learning technique that tends to reveal spontaneous clusters for successive components that are decreasingly extracted following the most important dimension of variance in the data [77]. Therefore, PCA components describe the sticking phenomenon only if it represents one of the most important dimensions of variance retained. This was not the case for either dataset1 or dataset2. To overcome this limitation, we employ supervised learning techniques, particularly DA, in which the qualitative  $y$  variable (the occurrence of sticking) guides component extraction.

## **3.4 Linear discriminant analysis of dataset1:**

### **3.4.1 Descriptive outcomes:**

The first descriptive DA model was developed for all variables and observations in dataset1 (71 batches x 25 variables). The corresponding score, loading, and variable importance plots are depicted in Figure 8.

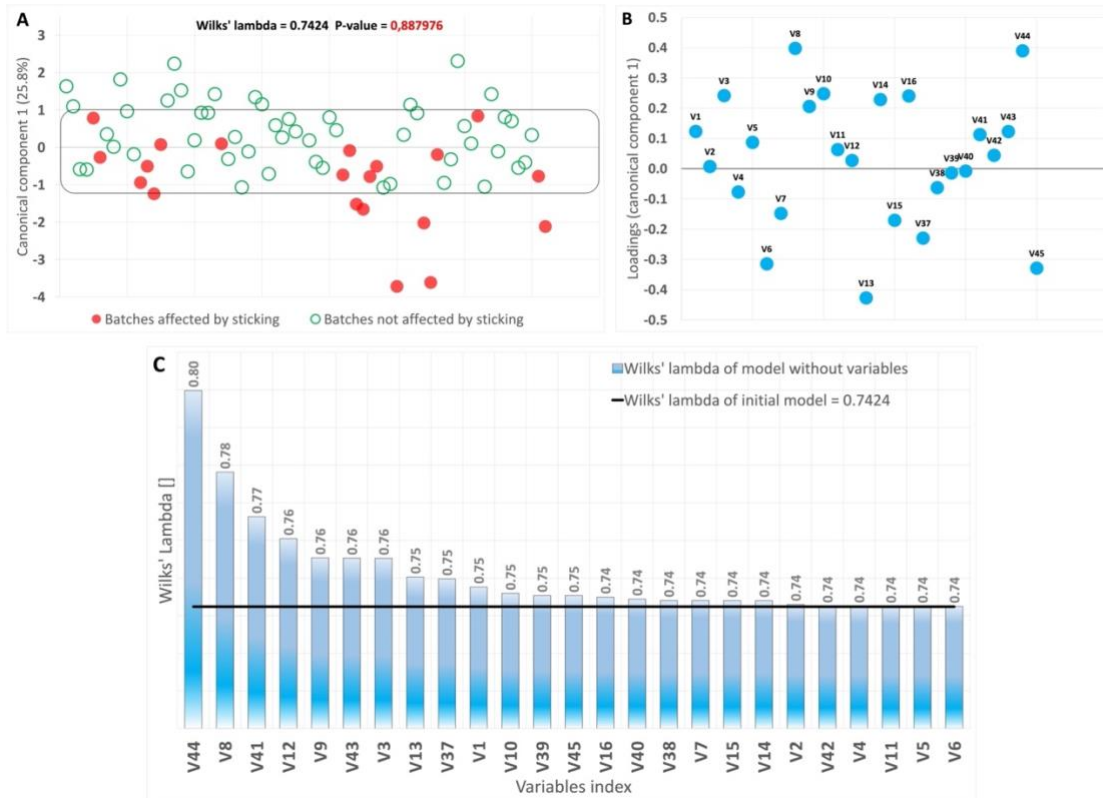


Figure 8 : Score (A), loading (B), and variable importance (C) plots of the first descriptive LDA model developed on dataset1.  
 The frame on the score plot (A) indicates the overlapping.

The significance of an LDA model is evaluated using Wilks' lambda, a well-known statistical parameter [83,84]. This statistical value represents the overall quality of discrimination results. It allows for the simultaneous comparison of several averages by assessing the part of intraclass inertia in the total variance. It corresponds to the ratio between the determinant of the intraclass variance-covariance matrix and the determinant of the total variance-covariance matrix. In this way, Wilks' lambda always varies between 0 and 1, where 0 represents more reliable discrimination because of a better interclass separation. A test with a null hypothesis, positing equality of the class averages, is performed to assess the significance of the observed Wilks' lambda value with regard to the number of observations, variables, and classes of the target variable. Wilks' lambda follows a complex distribution and, generally, a transformation is first computed to approximate more common distributions. The transformed statistic (RAO transformation and Bartlett transformation of Wilks' lambda) is then compared to a Fisher distribution and  $\chi^2$  distribution [80,83].

The first DA model built on dataset1 relies on one canonical component, with a Wilks' lambda value of 0.7424, which is insignificant with a p-value over 0.8 (Figure 8A). Therefore, the discriminating power highlighted by this first model does not lead to a significant separation

between the batches affected by sticking and those unaffected by sticking. Although the separation is insignificant, the batches affected by sticking exhibited a slightly lower score, along with component 1, than batches unaffected by sticking (Figure 8A). The corresponding loading plot (Figure 8B) indicates that the batches affected by the sticking cluster, aside from those unaffected by sticking, based on their relatively high values for V6, V13, and V45 and their relatively low values for V8 and V44. V3, V9, V10, V14, V16, and V37, may also be important.

The importance of each variable is given by the variable importance graphic (Figure 8C). This graphic displays the variation induced in the initial Wilks' lambda of the model by the removal of each variable, thereby preventing the misinterpretation that can sometimes arise with a direct lecture of low loadings, as observed in Figure 8B. As Wilks' lambda increases with the withdrawal of a specific variable, that variable becomes more relevant for discrimination. Following this graphic, some variables are relevant while many others account for nothing in the model. These results suggest that there is some information in dataset1 that is useful in distinguishing the two groups of batches, although its significance is probably masked by the presence of several variables that account for little or nothing in the discrimination. For instance, the Wilks' lambda value of the model does not change when some variables, such as V2 or V4, are discarded from the model; in contrast, it increases when V8, V41, or V44 are removed (Figure 8C).

Thus, we executed a well-known backward variable-selection strategy to optimize the LDA model [80,83,84]. We computed the first LDA model, including all of the variables, and calculated the initial Wilks' lambda value. The model was then optimized by recursively removing the variable contributing least to the Wilk's lambda value. Thus, seven variables were selected, and a second LDA model was then built with these variables. The corresponding score, loading, and variable importance plots are depicted in Figure 9.

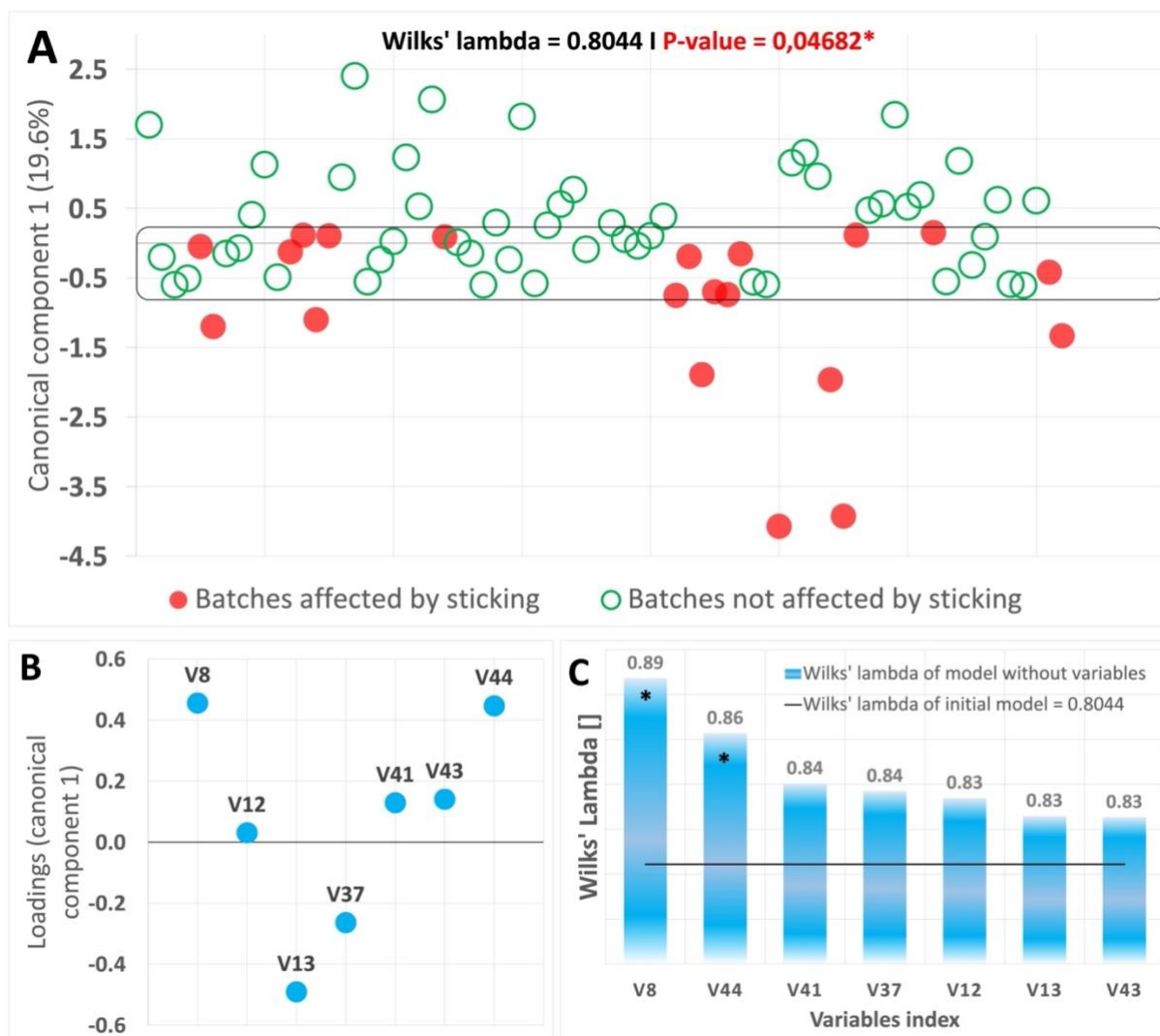


Figure 9 : Score (A), loading (B), and variable importance (C) plots of the optimized descriptive LDA model developed on dataset1.

The frame on the score plot (A) indicates the overlapping. (\*) indicates the statistical significance of the change induced in Wilks' lambda by the withdrawal of corresponding variables (C).

Unlike the first LDA model, a clear trend of discrimination is now observed in the score plot (Figure 9A), even if an overlap is still observed in the center of the graph, albeit to a lesser extent than in the first model (Figure 8A). Moreover, discrimination is observed using only seven variables, as opposed to using 25 variables in the first LDA model. Consequently, this second optimized model results with a slightly higher Wilks' lambda value of 0.8044, which is still significant with a p-value below 0.05. This indicates the existence of relevant information in dataset1 that distinguishes the batches affected by sticking from those unaffected by sticking. According to the loading plot in Figure 9B, the LDA component—along which several batches affected by sticking stand out from some unaffected by sticking on the score plot—is positively correlated with V8 and V44 and negatively correlated with V13 and V37. The variable importance plot in Figure 9C illustrates that all seven variables in the

optimized model are important for achieving the observed discrimination. Nevertheless, only two of them significantly contribute to the separation. These are V8 (corresponding to the total quantity of lubricant dispensed in the formulation) and V44 (the total duration of stops during the compression of the whole batch). The influences of V8 and V44 indicate that a batch is more likely to be affected by sticking when there is less lubricant dispensed (V8) in the formulation and when there are fewer stops during compression (V44). Similar results were obtained when using the PLS-DA method (Appendix 4). The effects of these two variables on sticking occurrences are discussed in detail in section 3.6 of this chapter. For the other variables that were identified as important but not significant, more data are necessary to confirm their role in sticking.

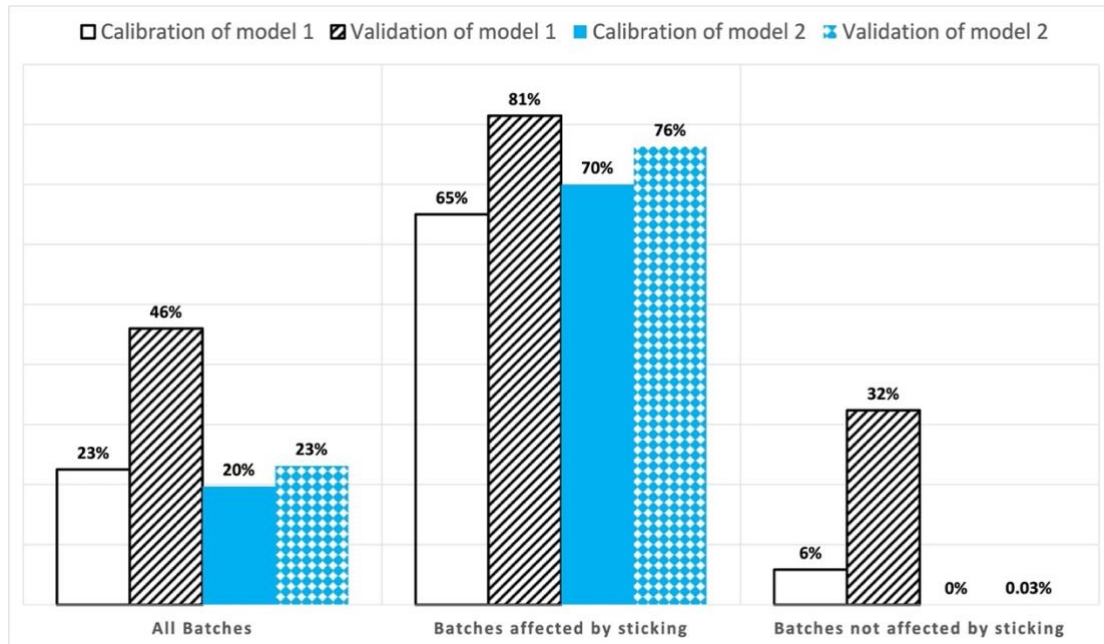
Although some insights regarding sticking were found, only a small portion of the difference between affected and unaffected batches is explained by the presented models, since the high values of Wilks' lambda (0.7424 and 0.8044) in both models indicate a low interclass variance in dataset1. This is probably because some important factors contributing to sticking are not included in dataset1. Furthermore, since the data were obtained from historical batches, data on some variables that are potentially important for the occurrence of sticking were unavailable. These variables include punch surface conditions, several specific properties of materials at defined steps of the process (such as the water content, the electrostatic properties, and particle size), and several compression parameters that are not systematically recorded, such as the actual compression force over time. Moreover, no variables related to environmental conditions at the manufacturing plant are present in dataset1, exacerbating the lack of potential information that can help explain the observed sticking. Data collection that addresses more variables may yield more complete and precise interpretations.

However, unlike the PCA model, the LDA model does provide insights about sticking, even if these do not account for the most important portion of the total variance in this specific data. This is due to the way that DA extracts its components. Instead of extracting the components with the greatest variance in data one after another, DA searches for the dimensions that maximize the separation between the centers of two classes. This has been achieved here, even if only a small portion of the phenomenon was captured.



### 3.4.2 Predictive outcomes:

A classification function is associated with the LDA models developed before and after optimization. The performances of these two classification models in predicting sticking during compression are presented in Figure 10.



*Figure 10 : Histogram of the classification error rate for the predictive LDA models. Model 1 and Model 2 represent the model before optimization (71 batches x 25 variables) and the optimized model (71 batches x 7 variables), respectively. The validation error rate corresponds to the result of a 7-folds cross-validation procedure.*

According to these results, the first LDA model (computed before optimization) classifies the 71 batches with error rates of 23% and 46% in calibration and validation, respectively (Figure 10). This indicates overfitting, as the model is better at classifying batches when they are included in its construction (calibration) than when they are excluded (validation). Strong overfitting is an indicator of variables in the dataset that contribute little or nothing to the actual explanation and prediction of the studied phenomenon, following the descriptive model discussion. After optimization, the new model led to a slightly lower error rate in calibration, with a 20% misclassification rate. However, the cross-validation error rate decreases significantly from 46% to 23%, which is closer to the new calibration error (Figure 10). Therefore, unlike the first model, the optimized model predicts both the batches used in its construction and new batches.

The optimized model possesses the necessary performance to predict sticking in production, or for providing an early indication. However, a detailed examination reveals that the prediction error rate of the model depends on the class of batches, indicating the existence

of bias. Moreover, batches affected by sticking are systematically the worst-classified batches, with an error rate of 65–81%, compared to an error rate of 0–30% for batches unaffected by sticking (see Figure 10).

In light of the results and raw data, the observed bias appears to be understandable. This is probably due to the very definition of batches affected by sticking. We found that 60% of these batches involve only one sticking event reported during their whole compression. Based on the importance of these single sticking events, some of these batches may be regarded as unaffected by sticking. However, without any other information on the importance of the single sticking events, sorting is not possible. Therefore, all of these were classified as affected by sticking, which probably prevents the models from correctly predicting batches affected by sticking. Thus, the developed models have trouble distinguishing some batches affected by sticking from those unaffected by sticking. The overlap observed in the center of the score plots in explicative models (Figure 8A and Figure 9A) is probably related to this issue.

This fine line between sticky and not-sticky batches constitutes a major limitation of our study on two levels. First, since some of the batches affected by sticking are difficult to separate from those unaffected by sticking, variables identified as significant contributors to sticking may just be artifacts caused by the outlying behavior of a few sticking batches. Fortunately, this can be assessed by analyzing the raw data on the identified variables. The results of this analysis are presented in a dedicated section of this paper (see section 3.6 of this chapter). Second, such a bias definitively prevents the developed model from determining whether a new batch is likely to have sticking issues, potentially making multiple false predictions for batches affected by sticking. To overcome this limitation, stickier batches can be gathered and included in the study, thereby preventing confusion between the two classes. Furthermore, a quantitative *y* variable with a large range of sticking behavior may be a more effective means for preventing the observed bias. Using several distinct profiles of sticking in the data may prevent eventual bias due to categorization. Moreover, it can be more informative because the relationships between input variables and sticking events can be described from multiple perspectives. Another means of preventing the observed bias is to perform planned data collection, where a systematic evaluation of sticking event intensity is conducted. This can be achieved in several ways. For instance, an intensity scoring methodology can be developed and adopted by manufacturers. Moreover, a procedure that allows operators to quickly take a photo of the punch surface before cleaning it—or even to preserve the stuck powder for subsequent analysis—could be implemented.

### **3.5 Partial least squares-discriminant analysis of dataset2:**

#### **3.5.1 Descriptive outcomes:**

The first PLS-DA model was built on dataset2, including all batches and variables available (25 batches x 45 variables). For simplicity, the outcomes of this first PLS-DA model are not presented here. Like the first LDA model, the first PLS-DA model consists of one latent component that was also subject to overfitting according to classification errors. Therefore, a variable-selection strategy was also conducted to optimize the model. The employed strategy is quite different from the one conducted for the LDA models with dataset1. The PLS-DA model is optimized based on the variable importance in the projection (VIP), which is a well-known statistic in partial least squares procedures [77,82,90]. VIP is a dimensionless parameter that enables estimation of the importance of each variable in the projection along with the latent variables extracted by the PLS procedures. A threshold value of 0.8 or 1 for VIP is generally accepted as heuristic of a particular variable's importance. This parameter is often used to perform variable selection when PLS-based models are overfitted [77,82,90]. The strategy excludes variables exhibiting a VIP value significantly lower than a predefined threshold. Because of their low contribution to the projection on latent structures, these variables are probably just modeling noise and data, rather than representing physics, leading to a good performance in calibration but poor performance in validation.

The VIP-based selection strategy was computed to optimize the first PLS-DA model. The first selection was operated by filtering all variables exhibiting a VIP score lower than 0.8. Then, a new PLS-DA model was built on the remaining variables. This model also exhibits overfitting issues. Thus, a second variable-selection operation was conducted by filtering out all of the variables that failed to display a VIP score significantly greater than 0.8. This led to the selection of a set of 10 variables (V8, V11, V13, V21, V22, V29, V33, V34, V42, and V44). Using these 10 variables and the 25 observations available in dataset2, we built an optimized PLS-DA model. The corresponding score, loading, and VIP plots are presented in Figure 11.

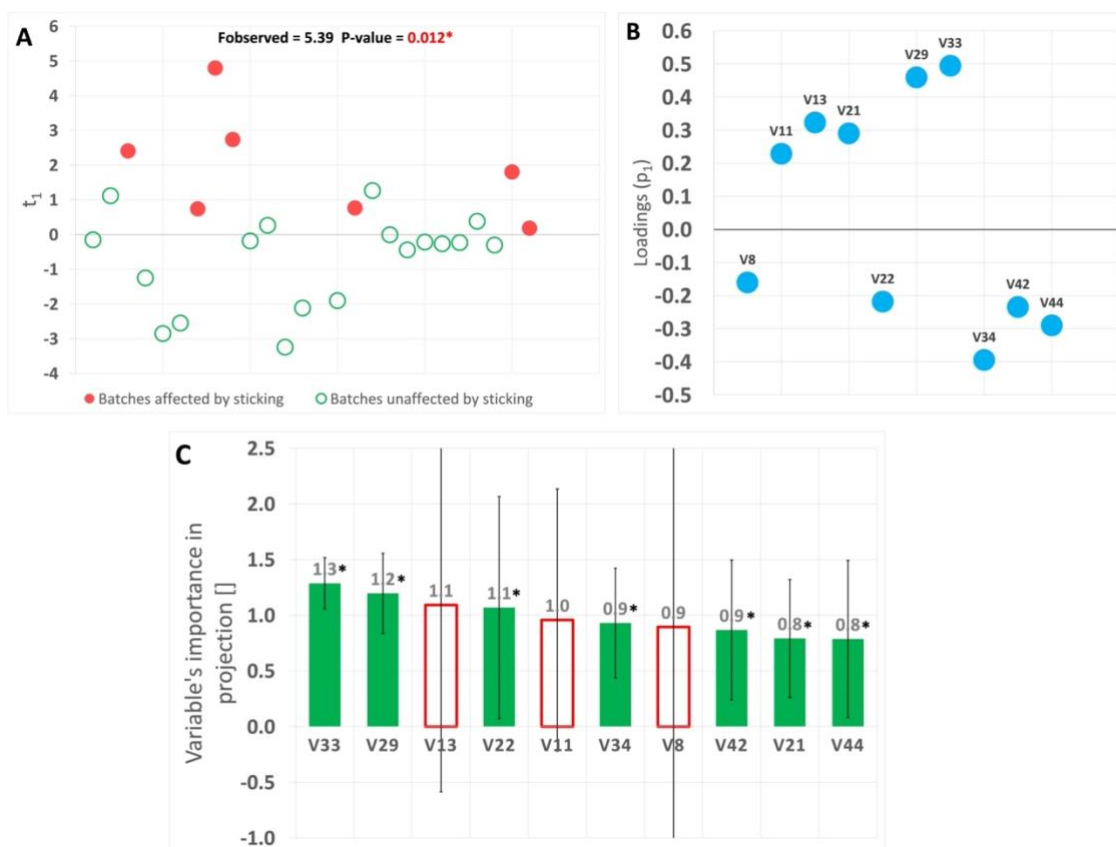


Figure 11 : Score (A), loading (B), and VIP (C) plots of the optimized PLS-DA model. (\*) indicates the statistical significance (95%) of the VIP scores (C).

The optimized PLS-DA model distinguishes batches affected by sticking from those unaffected by sticking, even if an overlap is observed in the center of the score plot (Figure 11A). This model allows for discrimination with an overall multiple analysis of variance p-value of 0.01. According to the score and loading plots given in Figure 11, some batches affected by sticking stand out from many batches unaffected by sticking, which is based on their relatively high values for V29 and V33, as well as their low values for V34. The VIP plot for more precision (Figure 11C) shows that all 10 variables are important for explaining the observed discrimination, although only seven of them, depicted in green, are identified as contributing significantly to sticking (in order of importance: V33, V29, V22, V34, V42, V21, and V44). Their effects are discussed in detail in section 3.6 of this chapter. For V8, V11, and V13, which were identified as important but not statistically significant, more data are needed to confirm their impact on sticking.

### 3.5.2 Predictive outcomes:

The predictive performances in the calibration and validation of the optimized PLS-DA model are presented in Figure 12.

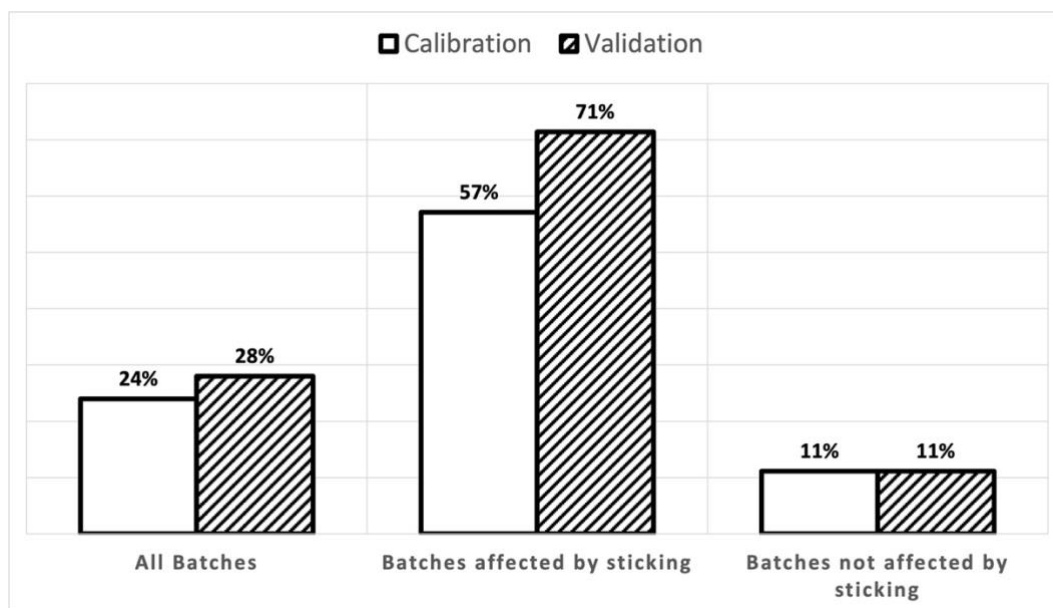


Figure 12 : Histogram of the classification error rate for optimized PLS-DA (25 batches x 10 variables).

The validation error rate corresponds to the result of a 7-folds cross-validation procedure.

The optimized PLS-DA model misclassifies 24% and 28% of batches in calibration and validation, respectively. These reasonable performances suggest that they can be useful in industrial production to predict sticking during the compression of a new batch, or at least to give an early indication of sticking. However, a detailed examination of the calibration and validation scores highlights the same bias observed in the prediction of affected batches by the LDA-optimized model. The PLS-DA model also made more errors predicting the affected batches versus unaffected batches, with error rates of 57% and 71% for affected batches in calibration and validation, respectively, versus an error rate of 11% for unaffected batches (Figure 12). This leads to the same bias as discussed with the optimized LDA model. Therefore, using the optimized PLS-DA model as a predictive tool carries a high risk of misprediction of batches affected by sticking. In addition, the sharp gap between the error rates in calibration (57%) and validation (71%) also suggests overfitting for these batches. This is probably due to the low number of batches affected by sticking in dataset2, potentially leading to a lack of generalization.

### 3.6 Discussion of the significant insights identified:

Our analyses of dataset1 and dataset2 have identified a total of eight variables (V8, V21, V22, V29, V33, V34, V42, and V44) as significantly contributing to sticking in the ibuprofen-based formulation. Due to the observed bias in the developed models, the importance of these variables may be artifacts, resulting from the presence of outliers in the data or the low number

of batches affected by sticking in dataset2. To evaluate these possibilities, the raw data for the identified variables are represented in Figures 13 – 16 to assess the differences more precisely between the two classes and detect the potential bias. In this section, the identified variables are also discussed in regard to findings reported in the literature.

### 3.6.1 The quantity of lubricant dispensed in the formulation (V8):

The quantity of lubricant dispensed in the formulation was identified through the analysis of dataset1 as a significant contributor to sticking. It was also identified as important in the analysis of dataset2, although this relationship was not significant ( $p > 0,05$ ). A box plot highlighting the mean comparison of this variable for affected batches versus unaffected batches is depicted in Figure 13.

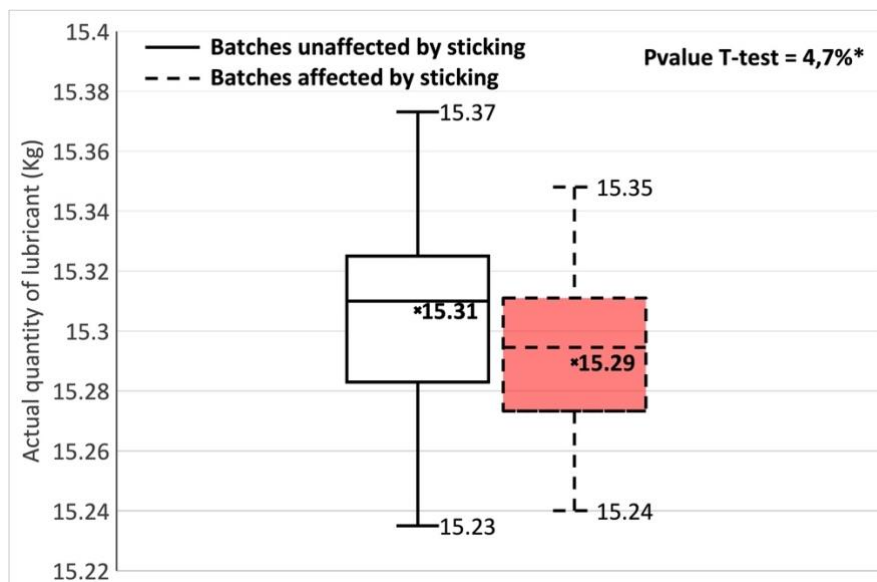


Figure 13 : Box plot of mean comparison between batches affected by sticking and those unaffected by sticking for V8.

V8 is the actual quantity of lubricant dispensed in the formulation in Kg. (\*) indicates the statistical significance (95%) of the Student's t-test.

The examination of the raw data (Figure 13) confirms that slightly less lubricant is dispensed in the formulations for batches affected by sticking, with a statistically significant difference of approximately 20 g of lubricant on average (p-value of 0.047). No outlying behavior appears in the box plot, indicating that V8 is not impacted by the observed bias. Therefore, the actual quantity of lubricant probably plays an important role in the occurrence of sticking in the ibuprofen-based formulation. Moreover, this effect is largely confirmed by other studies. Sendall et al. (1986), Toyoshima et al. (1988), and McDermott et al. (2011) reported on the benefits of lubricants in mitigating sticking issues with various kinds of formulations, such as pozanicline, acid acetylsalicylic, and macrogol [35,38,51]. It is generally

established that magnesium stearate, a common lubricant used in formulations, reduces sticking occurrences by covering API particles in the formulation, thus making them less available for adhesion to the punch surface [35,51]. However, some studies using ibuprofen-based formulations revealed the opposite effect, especially for magnesium stearate, where an increase in its proportion leads to more sticking [11,12,70]. It may be interesting to experimentally investigate why and how lubrication enhanced sticking behavior for the ibuprofen-based formulation studied in this work. Moreover, the average difference between the two classes of batches in this study was quite small. This represents ~0.1% of the targeted lubricant weight added to the formulation. Such levels of accuracy might not be addressable at an industrial scale and are practically irrelevant to prevent sticking on their own. The literature also revealed that the lubricant effect depends on punch surface properties [14]. Therefore, an experimental validation assessing the lubricant effect in coordination with the other parameters may be valuable to support further improvements or new formulation development efforts.

### **3.6.2 Average temperatures and relative humidity (V21, V22, V29, V33, and V34):**

Five of the variables identified as significantly contributing to sticking relate to the environmental conditions of manufacturing. V21 and V22 are the averages of the temperature and relative humidity, respectively, in the mixing room during mixing. V33 and V34 represent the same parameters in the compression room during compression. The final environmental variable is V29, the temperature average in the storage room during the storage of intermediate materials before compression. Figure 14 presents box plots displaying the comparison of these five variables for affected batches versus unaffected batches.

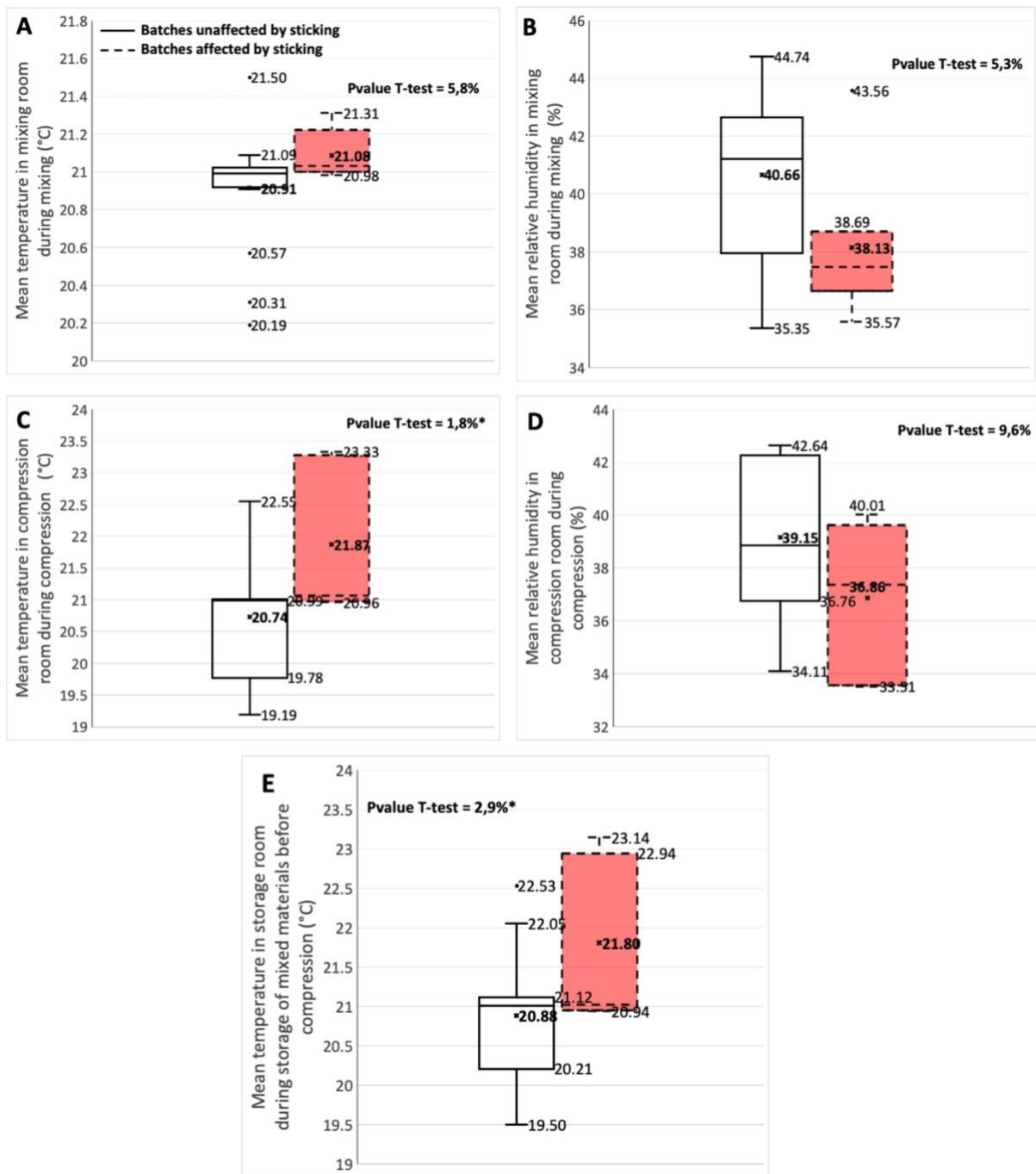


Figure 14 : Box plots of mean comparison between affected batches and unaffected batches for the environmental-related parameters.

V21, V22, V33, V34, and V29 are presented in A, B, C, D, and E, respectively. (\*) indicates the statistical significance (95%) of the Student's t-test.

As shown in the DA models, high temperatures in the mixing, compression, and storage rooms (V21, V33, and V29, respectively) are associated with more sticking, and the opposite relationship is observed for relative humidity in the mixing and compression rooms (V22 and V34, respectively) (Figure 14). Some moderate outliers are identified regarding V21, V22, and V29, although these may only induce misinterpretation for V21, where 3 batches appear to contribute to the direction of discrimination (Figure 14). For this variable, more data or an



experimental study are necessary to confirm that the relationship with sticking is not an artifact induced by the three outlying batches, which would not be representative of general observations. For V22 and V29, the observed outliers even tend to oppose the mean difference, so they are not problematic in terms of bias (Figure 9). Therefore, except for V21, all the other environmental variables identified (V22, V29, V33, and V34) probably play important roles in sticking, even if some are not statistically different on average—especially V34, with a p-value of 0.096. This simply indicates that the effect of V34 alone is insufficient to explain a significant part of the difference between sticky and not-sticky batches, highlighting the power of MVDA. Moreover, V34 probably interacts with other important parameters to better explain sticking.

Notably, temperature and humidity are interdependent parameters. Previous reports clearly state that exposition of powders, even to low relative humidity values, can favor the increase of its water content, which was subsequently correlated with sticking [20,39,72–74]. Danjo et al. showed that sticking increases with the water content of pharmaceutical powder up to 3% (w/w) and then decreases with further increases in water content, where it starts to behave as a lubricant at the interface between particles and punch surfaces [74]. For the formulation studied here, it is then probable that the slightly higher temperature, favoring a slightly lower relative humidity induces the water uptake necessary to give a specific level of water content, which promotes sticking according to processing times. Experiments are necessary to elucidate the postulated mechanism of action. Herein, the slight difference in temperature and relative humidity observed between the affected and unaffected batches may not be exploitable to mitigate sticking issues at the industrial scale. Therefore, clarifying the mechanism may help identify a practical level of environmental conditions addressable by HVAC systems that can subsequently mitigate sticking of the formulation.

Additionally, environmental conditions measured during compression can be more of a consequence than a cause. For example, the low melting point of ibuprofen promotes sticking if an important increase in temperature occurs in the powder bed during compression [11–13,70]. Thus, the slightly higher temperature measured during compression, translating into the slightly lower relative humidity may be indicative of a high cadence of operations inducing a local increase of temperature above the melting point of ibuprofen, promoting sticking. This hypothesis is corroborated by the total duration of stops during compression. For instance, less time was spent stopping during the compression of batches that were affected by sticking,

which may also indicate a higher operational cadence with a higher probability of localized temperature increase at the interface between the powder bed and punch surface.

### 3.6.3 Average disintegration time (V42):

Although it was included in dataset1, the average tablet disintegration time (V42) was identified as an important contributor to sticking only in the analysis of dataset2. Thus, the observation of this effect may be due to the lower number of observations in dataset2, especially for batches affected by sticking, which are also subject to the bias observed in the developed models. To assess this aspect, two box plots displaying the comparison of the average disintegration time between affected and unaffected batches are presented in Figure 15. The first plot utilizes data present in dataset2 and the second plot presents the same comparison for dataset1.

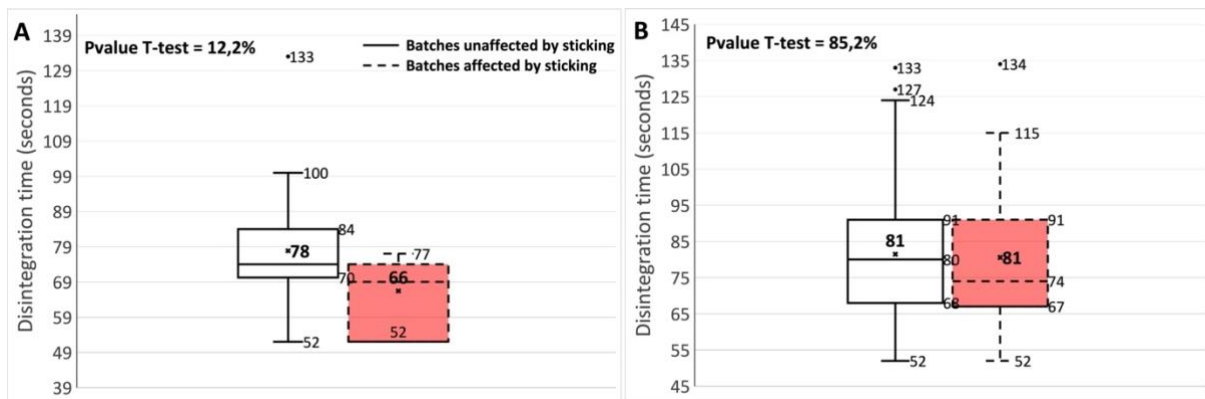


Figure 15 : Box plots of the mean comparison between affected and unaffected batches for the average disintegration time (V42).

Figures 10A and 10B use the information from dataset2 and dataset1, respectively.

As indicated by the DA models, the average disintegration time for dataset2 appears to be associated with sticking (Figure 15A). However, it may be a consequence, rather than a cause, because batches affected by sticking may yield tablets with lower internal cohesion, thus lowering their disintegration time. But an examination of this variable with a greater number of observations in dataset1 reveals a potential bias. Distributions for the two classes are quite similar in dataset1 (Figure 15B), indicating that the average disintegration time is probably identified as a factor in sticking due to the small number of observations in dataset2. Moreover, even fewer batches are affected by sticking in dataset2. This is highlighted by the change in the average values between the two datasets. The average disintegration time for batches unaffected by sticking is similar between dataset1 (81,47 seconds) and dataset2 (77,83 seconds), while it is significantly different for batches affected by sticking—81,55 seconds

in dataset1 versus 66,43 seconds in dataset2 (Figure 15). Therefore, the average disintegration time may be an artifact, induced by both the subsampling operated to study environmental conditions variables and the inherent observed bias in the models. An experimental study may help to clarify the effect of sticking on the tablet disintegration time.

### 3.6.4 The total duration of stops during compression (V44):

The total duration of stops during compression was identified as a significant contributor to sticking, which was apparent in the analyses of dataset1 and dataset2. Figure 16 presents two box plots displaying the comparison of the total duration of stops during compression between affected and unaffected batches for the two datasets.

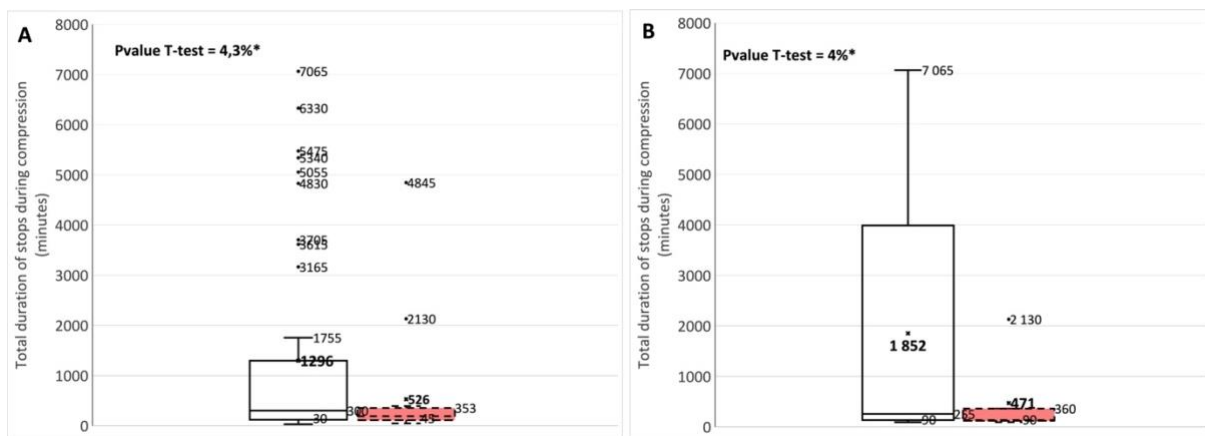


Figure 16 : Box plots of the mean comparison between affected and unaffected batches for the total duration of stops during compression, in minutes (V44).

Figures 11A and 11B use the information from dataset1 and dataset2, respectively. (\*) indicates the statistical significance (95%) of the Student's t-test.

Our examination of the raw data concerning the total duration of stops during compression confirms the relevance of this variable's relationship to sticking, both in dataset1 and dataset2 (Figure 16). The batches are not affected by sticking when the tablet press spends more time stopped (1296 and 1852 minutes on average) and the batches are affected when less time is spent stopping (526 and 471 minutes, on average). Notably, the total duration of stops during compression does not include the stop time due to sticking. It only accounts for sporadic stops due to different breaks of operators and other activities necessary for monitoring the manufacturing process. Because this condition differs significantly from one batch to another, several outliers in its distribution are identified for the affected and unaffected batches (Figure 16). Considering the nature of V44, the presence of outliers cannot be considered an indicator of artifacts. In addition, we observe a highly similar distribution for V44 across both datasets (Figure 16). Furthermore, the difference in the average for V44 between affected and

unaffected batches remains significant, with a p-value always below 0.05 in dataset1 and dataset2. This difference even persists when all batches that appear to be outliers are removed. Therefore, the total duration of stops due to sticking probably plays an important role in sticking.

No work in the literature has reported such a direct link. A potential explanation may be an indirect correlation involving the influence of temperature in sticking for the ibuprofen-based formulation. Several authors have reported on the low melting point of ibuprofen particles and a potentially sharp temperature increase in the powder bed during compression, which could lead to the sticking of particles onto punch surfaces [11–13,70]. The high cadence and duration of a continued compression would result in temperature increases in the powder bed, and more time stopped during compression would have a cooling effect on tooling, thus mitigating sticking occurrences on ibuprofen-based formulations. This postulated mechanism aligns well with the discussions about temperature and relative humidity measures during compression. A detailed experiment is necessary to confirm the hypothesis. If this is supported, it is worth mentioning that the intentional introduction of more stops during compression is not a viable solution from an industrial perspective, but cooling down the tablet press (especially the die cavity and the surfaces of punches) may be an interesting approach to mitigate sticking.

### **3.7 A new approach to studying tablet sticking and its related systematic methodology:**

This study validates a new approach to evaluate tablet sticking, consisting of collecting and analyzing industrial manufacturing data. This approach presents two main advantages. First, it allows the simultaneous assessment of numerous factors for their implications in sticking, to a greater extent than experimental studies. For instance, this study involves up to approximately 50 variables after data cleaning, even though only six of these were found to significantly contribute to sticking. Since sticking is known to be a multifactorial phenomenon, this approach may greatly enhance our understanding of sticking, or at least enable the informed selection of a few factors to assess in subsequent experimental studies. Second, another benefit of the industrial data analysis approach is the emphasis on actual factors that primarily contribute to sticking. One challenge in sticking studies is observing or reproducing this phenomenon on an experimental scale [9,10,26]. Many factors usually do not appear until the formulation is scaled up to the industrial stage [5]. Thus, an analysis based on this scale is likely to be more

informative. Moreover, many factors have tight tolerances and are strictly controlled during industrial manufacturing, allowing for a focus on variables for which sticking is very sensitive.

Based on our experiences conducting this study, we propose a systematic methodology to proceed with the new approach. This methodology is based on MVDA techniques. These are adequate tools for the new approach's purpose because of the limited number of batches performed at an industrial scale and the commonly large number of factors. Such datasets generally lack degrees of freedom to facilitate the correct use of common regression and classification techniques directly with original variables. Furthermore, they often lack enough observations to perform more complex modeling, such as artificial intelligence models like neural networks. Therefore, MVDA is the best fit, particularly for dimensionality reduction techniques such as PCA and partial least squares regression.

The proposed methodology involves three major steps: data collection and preparation, followed by an exploratory step, and then explicative and predictive modeling. The first step involves gathering, cleaning, and preparing relevant variables regarding the manufacturing process and tablet sticking. Special attention must be devoted to the choice of the  $y$  variable(s) for assessing the sticking behavior of the batches because this predetermines all eventual outputs. For this first step, well-planned and detailed data collection can be more informative (when it is possible) than a historical study. Regardless of the collection mode, data must be cleaned and prepared prior to analysis. This involves handling missing data, feature engineering, evaluating multicollinearity, and standardizing methods.

The resulting dataset from the first step is then assessed using univariate data analysis methods and PCA. Univariate data analysis reveals whether sticking is dependent on any single variable, although it also helps in finding, investigating, and processing eventual outliers, thereby completing data cleaning. PCA finalizes the investigation on a multivariate stage. The existence of spontaneous clusters according to the observed sticking is assessed, and the most important dimensions of variance in the data are analyzed to detect eventual multivariate outliers and ensure that there is no other phenomenon captured in the data that could impact sticking. If clear clustering with sticking is observed on the PCA components, there is effectively no need to proceed to the third step of the proposed methodology. The loading and contribution plots of these components will be sufficient to identify the important contributors to sticking. They can also be used in conjunction with multivariate linear regression to develop a predictive model. In contrast, if no clear trend regarding sticking is apparent after PCA, the final step of the proposed methodology must be performed, consisting of supervised learning

techniques. These modeling techniques will definitively provide explicative and predictive models that may assist in understanding the observed sticking, even if they rely on a limited part of the data. This systematic methodology can be adapted and employed for a broad range of cases by integrating any other necessary analysis techniques at each stage.

However, two major limits related to this proposed approach remain. Since this entire approach is based on data analysis, there must be relevant and sufficient data. This is the most important limit because when data do not contain relevant information about sticking, no analysis method would reveal helpful information. This occurred partially in this study because only a small portion of the observed sticking can be explained, probably due to missing data regarding important variables. The proposal of a well-planned and detailed data collection may be useful for mitigating this limitation. The second limitation is related to the inability to fundamentally describe the observed relationship between sticking and the identified variables. Future experimental studies should be considered to address this deficiency. For instance, researchers can implement a design that includes the identified variables along with potentially important variables that were not included in this data analysis approach. Such a study may yield a more complete and useful multivariate understanding of tablet sticking. Furthermore, integrating the results of several such studies involving different formulations may further assist the effort to generalize results and establish the root causes of tablet sticking.

## 4. Conclusion

The purposes of this study were to introduce, assess, and discuss a manufacturing data analysis approach to gain insights into tablet sticking on the industrial scale.

In the first step, we obtained industrial manufacturing data on approximately 100 variables, encompassing information about raw materials, processing machines, tablet properties, environmental conditions during production, and processing parameters.

Second, the relevance of the proposed approach in studying tablet sticking by directly assessing industrial data was determined. Using a predefined methodology based largely on MVDA, a total of six variables were identified as significant contributors to sticking for the proprietary ibuprofen-based formulation. Among these variables, the average temperatures during compression and storage of intermediate materials before compression were found to be positively correlated with sticking. Specifically, higher temperatures are associated with enhanced sticking. The actual quantity of lubricant dispensed in the formulation, the relative humidity in the mixing and compression rooms during these steps, and the total duration of stops during compression were found to be negatively correlated with sticking. Specifically, increases in these parameters mitigated sticking. These insights point out the role of lubrication, water content, and the low melting point of ibuprofen in its sticking tendencies. Therefore, the proposed approach is suitable to identify the impact of variations within closely defined tolerances, as observed in industrial pharmaceutical manufacturing processes.

Third, because of a large bias in the developed models for batches affected by sticking, these models were not suitable for predicting sticking. The descriptive outcomes exploited to identify important variables were also affected, since only a small portion of interclass separation could be explained, and several variables turned out to be artifacts. A study including batches with more severe sticking may prevent confusion between the two classes and address this limitation. Furthermore, a continuous  $y$  variable relying on a large range of sticking behavior may definitively prevent this kind of bias. Moreover, this can help describe the effects of variables on sticking, allowing for further determination of an optimal operating range for the identified parameters.

Finally, based on our experiences in conducting this work, we have proposed and discussed a systematic methodology for industrial manufacturing data analysis. We suggested a plan for data collection and an experimental design after data analysis to create a more complete description and multivariate understanding of tablet sticking.

## **5. Acknowledgments:**

The authors would like to thank the production and engineering teams from the plant in Saint-Laurent for their support in accessing the processes and data used in this work.

## **6. Declaration of competing interest:**

The authors report no conflicts of interest.



# CHAPITRE 4 : ETUDE MULTIVARIEE QUANTITATIVE DU COLLAGE

## Avant-propos

### Auteurs et affiliation :

Mahamadou Dembélé : étudiant au doctorat, Université de Sherbrooke, Faculté de génie, Département de génie chimique et biotechnologique.

Sophie Hudon : ingénieure automatisation, Global Technology and Engineering, Process Automation, Pfizer Canada.

Jean-Sébastien Simard : directeur, GPPD - Digital Design & Process Analytics, Pfizer Worldwide Research, Development & Medical.

Nicolas Abatzoglou : professeur agrégé, Université de Sherbrooke, Faculté de génie, Département de génie chimique et biotechnologique.

Ryan Gosselin : professeur titulaire, Université de Sherbrooke, Faculté de génie, Département de génie chimique et biotechnologique.

**Date d'acceptation :** 20 décembre 2022

**État de l'acceptation :** Publié

**Journal :** Pharmaceutical Development and Technology

**Titre anglais :** Insights into tablet sticking: a quantitative case study with an ibuprofen and methocarbamol-based formulation.

**Titre français :** Étude multivariée du collage à échelle industrielle : étude quantitative avec une formulation à base d'ibuprofène et de metocarbamol.

**Contribution au document :** Cet article présente les résultats d'une étude de cas appliquant la nouvelle approche d'étude du collage proposé par la présente thèse de doctorat. Cette étude de cas a été réalisée avec une formulation à base d'ibuprofène et de metocarbamol où le collage est évalué de façon quantitative. De façon complémentaire au chapitre 3, l'article confirme la pertinence de la nouvelle approche proposée et il introduit, démontre et discute les avantages ainsi que la méthodologie à employer dans le cas d'une description quantitative du collage.

## Résumé français

Étude multivariée du collage à échelle industrielle : étude quantitative avec une formulation à base d'ibuprofène et de metocarbamol.

Le collage consiste en l'accumulation continue de poudres de formulation sur la surface des outillages au cours de la compression. C'est un phénomène qui impacte gravement la productivité industrielle, la qualité des comprimés et la durée de vie des poinçons. Son occurrence est gouvernée par un équilibre dynamique et fragile de nombreux paramètres et leurs multiples interactions. Pour étudier un tel phénomène multifactoriel, nous avons proposé dans le chapitre 3 une nouvelle approche d'étude basée sur l'analyse des données industrielles. L'objectif était de faire la preuve de concept de la nouvelle approche lorsque le collage est décrit par un indicateur qualitatif. Cette étude a conclu que l'intégration de plus de paramètres prédictifs et un indicateur quantitatif de la sévérité du collage permettraient une meilleure explication et prédiction du collage. Dans la suite de cet effort, la présente étude s'attache alors à généraliser la nouvelle approche d'étude du collage en l'étendant aux cas où le collage est décrit par un indicateur quantitatif. Nous avons collecté et analysé un total de 345 paramètres sur 28 lots industriels d'une formulation à base d'ibuprofène et de metocarbamol où le collage était décrit quantitativement. En utilisant des modèles de régression PLS, nous avons montré que l'association de la durée de granulation et de la force de compression permettait d'expliquer 60 % des fluctuations du collage de la formulation étudiée. De plus, contrairement aux modèles de classification développés dans le chapitre 3, les résidus des modèles prédictifs de régression de la présente étude suivent une distribution normale et sont indépendants de la variable prédite.

**Mots clés :** collage de poudres pharmaceutiques, compression des poudres, comprimés bicouches, analyse multivariée des données, analyse en composantes principales et régression sur les moindres carrés partiels.

## Résumé en anglais

Insights into tablet sticking: a quantitative case study with an ibuprofen and methocarbamol-based formulation.

Tablet sticking is a continuous accumulation of pharmaceutical powder onto tooling surfaces during compression. Its occurrence greatly impacts tablet productivity, quality attributes, and tooling age. In a previous study<sup>1</sup>, the authors proposed a multivariate data analysis approach to gain insights into tablet sticking directly on the industrial stage [91]. The objective was to determine the combination of factors that could help distinguish between batches affected and unaffected by sticking. The present study aims to generalize this approach by extending it to quantitative predictions of punch sticking intensity. A total of 345 variables was gathered on 28 industrial batches of an ibuprofen and methocarbamol-based formulation. Using PLS regression models, it was shown that the association of granulation duration and compression force allows to significantly explain ~60% of sticking variations of studied formulation. In addition, unlike the classification models developed in the earlier work, the validation residues in the present study were found to be normally distributed (Shapiro-Wilks p-value=0.96) and independent from the target variable ( $R^2=9.5\%$ ).

**Keywords:** tablet sticking, powder compaction, bilayer tablet, multivariate data analysis, principal component analysis, partial least squares regression.

---

<sup>1</sup> Dembélé M, Hudon S, Simard JS, Abatzoglou N, Gosselin R. 2022. A multivariate data analysis approach to tablet sticking on an industrial scale: a qualitative case study of an ibuprofen-based formulation. *Pharmaceutical Development and Technology* [Accepted].

## **1. Introduction :**

Tablet sticking is the adhesion of formulation powder materials onto the surfaces of press tooling (dies and punches) during the compression step of the tablet manufacturing process [7,9]. When it occurs, sticking induces tablet defects, leading to important quality and yield issues [5,8,9]. The compression tools are also more likely to undergo premature breakage [5,7,31]. The process has been studied and several parameters have been identified as playing a significant role in the occurrence of tablet sticking, especially the physical, chemical, mechanical [10,14,26,46,47,57,62,64,81], and electrostatic properties [52–55,92] of the active pharmaceutical ingredient (API) particles.

However, the multifactorial nature of the phenomenon restrains the emergence of knowledge regarding the root causes [9,10,13,14,26,91], thus limiting the efforts to mitigate its impacts. In fact, the occurrence of tablet sticking has been found to involve multiple parameters as well as the existence of many interactions between many of these parameters [10–14,26,42,64,81,91]. To better understand tablet sticking for a particular formulation, it is essential to simultaneously study the parameters involved in its occurrence [9,10,26,91]. Carrying out a simultaneous experimental investigation on such a multifactorial phenomenon as tablet sticking presents an important financial and technical challenge. Financially, the time and cost associated with such a study would quickly become excessive due to the number of parameters to be investigated. On a technical level, the study of some of the parameters known to be important in sticking occurrences would present a real challenge at the laboratory or pilot scale [5,9,26,37,91].

To overcome these challenges in gaining insights into tablet sticking, an investigation of industrial manufacturing data was introduced and discussed at length in our earlier work [91]. The investigation consisted of gathering and analyzing all the data generated during the industrial manufacturing of ibuprofen-based tablets in search for insights about tablet sticking, thus allowing the simultaneous assessment of many more parameters than experimental studies could practically cover. Moreover, it was observed that sticking is more prevalent and damaging during the industrial production phase, when modifications and investigations on the formulation and process are limited [5,9,14,35]. The proposed approach could therefore lead to a better understanding of the sticking phenomenon or allow a significant reduction in the number of parameters to be assessed in a subsequent experimental study. The advantages and limits of the industrial data analysis approach in studying the sticking phenomenon were

discussed at length in our previous work [91]. Furthermore, its relevance was demonstrated through recent publications in the literature [26,91].

In 2020, Nuno et al. investigated the tablet sticking issues of a benzodiazepine-based formulation by analyzing its industrial manufacturing data [26]. A total of around 230 variables were collected on 8 industrial-scale batches, and the compression yield was used as a proxy for stickiness. A partial least squares (PLS) regression model was calibrated and validated on the data by keeping separate 2 of the 8 batches, to be used for external validation observations. By implementing this PLS model, which gave a root mean squared errors for calibration and validation of 13.47% and 10.24%, respectively, the authors were able to identify the airflow rate in granulation and the temperature variation of the air through the drying operation as critical factors contributing to sticking [26]. Subsequent lab-scale experiments confirmed these observations, highlighting the importance of water content in granules due to insufficient drying as the root cause of sticking of the studied formulation to both dies and punches [26].

The relevance of the industrial data analysis approach was also demonstrated in our earlier work carried out on an ibuprofen-based formulation [91]. That study involved 100 variables gathered on a total of 71 batches. The stickiness of each batch was assessed by the tablet press operator. Batches were classified as either affected or unaffected by sticking. The level of sticking was not quantified. Using a methodology based on multivariate data analysis (discriminant analysis techniques), a total of 6 variables was identified as significantly contributing to the occurrence of sticking in the studied formulation [91]. The average room temperatures during compression and storage of intermediates before compression were found to be positively correlated with sticking, whereas the lubricant quantity dispensed in the formulation, the average relative humidity (in the mixing and compression rooms), and the total duration of stops during compression were found to be negatively correlated with sticking. These insights point out the role of lubrication, water content, and the low melting point of ibuprofen in its sticking tendencies [91]. That previous work has also introduced and discussed a systematic methodology to deploy the industrial data analysis approach to tablet sticking.

In line with that earlier work, the present study aimed to investigate industrial manufacturing data to find insights into tablet sticking and to introduce the data analysis methodology in cases where sticking is described by a quantitative variable. The present study used a different formulation based on ibuprofen and methocarbamol, and more data of about 345 variables were gathered on 28 industrial batches displaying a large range of sticking behavior. Data were also analyzed using a different methodology based on the multivariate

data analysis: partial least square regression. The specific objectives of the present study are as follows:

- (1) Identification of the key variables contributing to the occurrences of sticking in the studied ibuprofen and methocarbamol-based formulation and their associated optimal operating range.
- (2) Development of a regression model which can be employed to forecast sticking occurrences in a batch during the compression phase.

In addition to these specific objectives, the present study also assesses the relevance of a hypothesis suggested during our earlier work [91]. The classification errors of the predictive models developed during that previous study were found to be dependent on the modalities of the qualitative y-variable describing sticking. A simple binary classification of a batch with respect to sticking (i.e., affected/unaffected) was concluded to be insufficient to properly understand tablet sticking. The use of a quantitative descriptor to describe various sticking behaviors was proposed as a potential solution. By fulfilling these conditions, the present study also aims to clearer limitations identified in the earlier work.

## **2. Materials and methods:**

### **2.1 Formulation :**

Bilayer tablets of a proprietary formulation containing two APIs, ibuprofen, and methocarbamol were used in this study. The other ingredients in the formulation of both layers are microcrystalline cellulose, ethyl cellulose, povidone, sodium lauryl sulfate, corn starch, calcium sulfate, carboxymethylcellulose, silicon dioxide colloidal, stearate magnesium, and talc. The tablets were manufactured through a wet granulation process followed by compression on a bilayer tablet press. The composition and granulation processes for the two layers are slightly different. One of the layers is colored while the second is not. The granules of the non-colored layer are first pre-compressed in the die cavity with a predefined precompression force. Then, the granules of the colored layer flow into the die, and a main bilayer compression force is applied to form the tablets. Each batch is intended to produce 600K to 700K individual tablets.

### **2.2 Gathering, cleaning, and preparation of data:**

Data were collected primarily via the master batch record (MBR) files. All the data available in these files regarding the manufacturing process was interrogated and 345 variables were gathered on 28 historical batches. This raw data was cleaned and prepared prior to analysis. The same cleaning and preparation procedures as used in our previous work were employed here [91].

Cleaning operations led to the withdrawal of more than 60% of the variables gathered, decreasing the total number of variables from 345 to 122 usable variables. These remaining variables concern all aspects of the process and can be classified into 5 groups:

- Group 1 corresponds to 17 quantitative variables related to the materials used in the formulation of both layers (e.g., the quantity of API and excipients dispensed).
- Group 2 consists of 17 qualitative variables related to the machines employed during tablet manufacturing. These include the type of cleaning performed prior to production and the time between that cleaning and the production run.
- Group 3 covers data about tablet quality attributes. It comprises a total of 6 quantitative variables.
- Group 4 encompasses environmental conditions data during the manufacturing of each batch. This consists of 21 quantitative and 4 qualitative variables which report on

average temperature, relative humidity, and differential pressure in each operating room.

- Group 5 is related to the manufacturing method for each batch. It includes 57 quantitative variables such as mixing, granulation, or compression parameters.

The final dataset comprises 28 batches described by 122 variables, including 121 input variables, herein referred to as V1 to V121, plus a  $y$ -variable that is the target variable describing the sticking behavior of each batch. The input variables consist of 100 quantitative and 21 qualitative parameters considered to explain sticking occurrences captured by the remaining quantitative target variable ( $y$ ).

### **2.3 Analysis methodology:**

Following the methodology proposed in our earlier work, the data analysis involves three steps consisting of data exploration and modeling [91]. The exploration step starts with a univariate exploration intended to determine if any single variable can significantly explain or predict observed sticking behavior. For this purpose, the linear correlation ( $R$ ) and determination ( $R^2$ ) coefficients were exploited. A visual bivariate assessment was also performed using scatter plots of pair quantitative variables colored according to the severity of sticking ( $y$ ). A Student's  $t$ -test was employed for bimodal qualitative variables, whereas an analysis of variance (ANOVA) was performed for multimodal qualitative variables.

After univariate exploration, multivariate modeling techniques were employed. Data were centered and scaled to unit variance before the computation of the multivariate models. The second step of the analysis starts with principal component analysis (PCA). This unsupervised technique was employed to check the presence of spontaneous clusters of batches when tablet sticking was observed [77,82,83].

The third step of data analysis consisted of supervised multivariate modeling of sticking occurrences. A common multivariate regression technique of partial least squares (PLS) was employed [77,82,83].

### **2.4 Model validation:**

A cross-validation procedure was employed to validate the PLS models. This consisted of the random splitting of data into seven groups of observations. The root mean squared error resulting from this cross-validation procedure (RMSECV) was used to assess the goodness of fit of the validation model. The significance of this goodness of fit was then assessed by



performing ANOVA on the residues generated by the cross-validation procedure. The cross-validation procedure and the associated ANOVA to test the significance of the model were described in detail by Eriksson et al. [77,82].

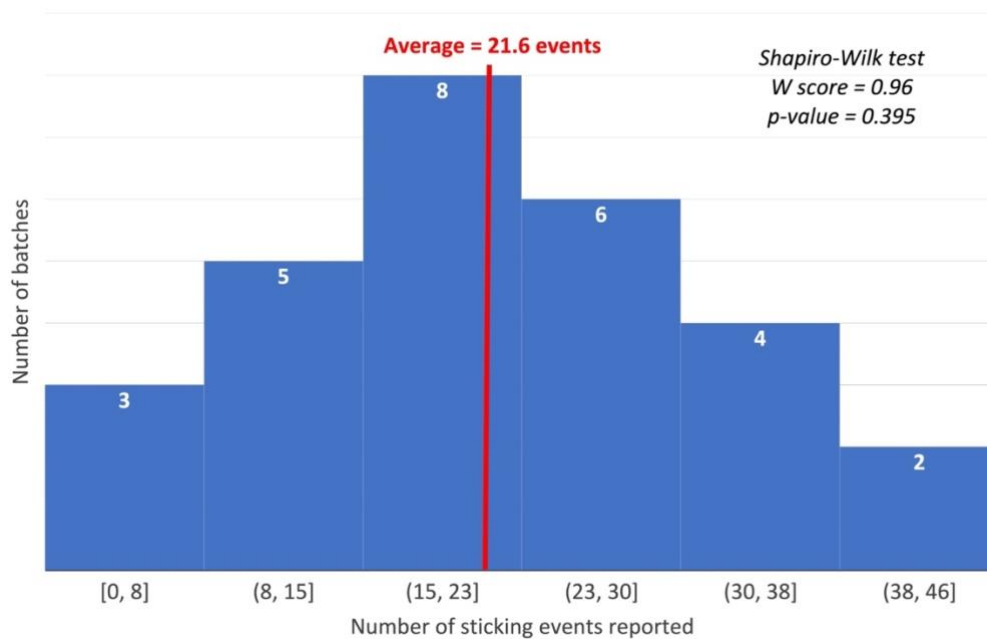
In short, the regression and residual sum of squares were computed through the results of the cross-validation. These were then divided by the corresponding degree of freedom to compute the regression and residual mean squares. The ratio of these two mean squares follows an F-distribution with the corresponding degrees of freedom. Thus, a comparison of the mean square ratios to the F-distribution makes it possible to validate the significance of the model in cross-validation. The procedure is similar to the classic ANOVA of a multilinear regression model, but since PLS components are employed as predictors instead of the original X matrix, the degrees of freedom (DoF) must be adjusted. Several methods to achieve this adjustment have been proposed in the literature [93,94]. In the present study, the number of DoFs was estimated using the common approximation method of doubling the number of PLS components retained [77,93,94].

In addition to the cross-validation procedure, PLS models were also checked for the presence of bias similar to that observed in our earlier work, where the predictive errors of the models were dependent on the modalities of the qualitative target variable [91]. To do so, the normality and homoscedasticity of normalized cross-validation residues were assessed by computing a common Shapiro–Wilks normality test and analyzing the significance of the correlation between validation residues and observed quantitative y values.

### 3. Results and discussions:

#### 3.1 Sticking data:

Batch stickiness was assessed by the number of sticking events reported by operators during the entire compression phase of each batch. Figure 17 depicts the distribution of this indicator across all batches.



*Figure 17 : Distribution of the number of sticking events (y-variable) reported across all gathered batches.*

*Red line highlights the average sticking events reported.*

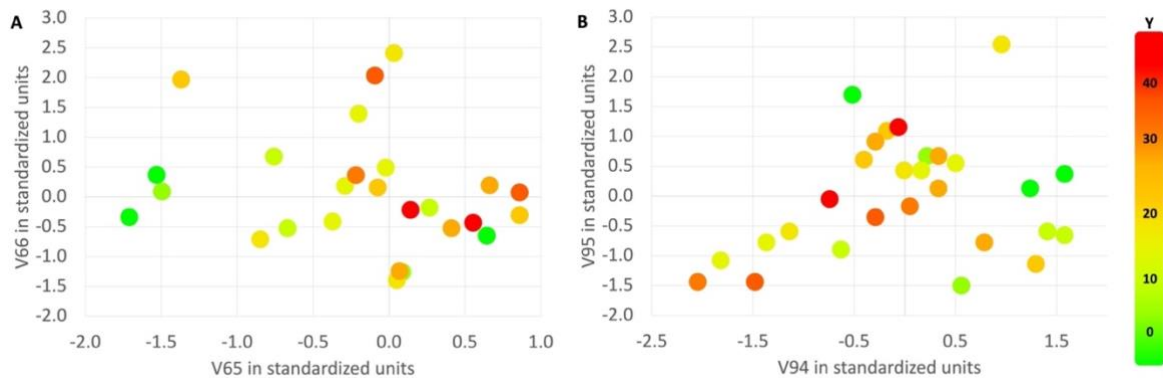
On average, 21.6 stopping events due to sticking were reported for each batch, resulting in an average loss of time of approximately 8 hours and 18 minutes during the compression phase. Therefore, sticking is very damaging for the studied formulation, as the time lost to stops due to sticking corresponds to almost 50% of the time needed to compress an entire batch. An average total duration of 27 hours and 21 minutes is necessary to complete the compression of a batch instead of the 19 hours and 3 minutes required if there were no stop events due to sticking. Analyzing the distribution plot, the number of sticking events across all gathered batches appears to follow a normal distribution trend; this was confirmed with a Shapiro–Wilks normality test displaying a p-value of 0.395. Thus, data employed in the present study exhibit a large range of sticking severity, with several batches unaffected or slightly affected by sticking and many others very affected by sticking. This kind of distribution is expected to help in preventing the limitations encountered during our previous work, where a fine line between

batches being affected or unaffected by sticking impacted the performances of the predictive models [91].

## 3.2 Univariate data analysis:

### 3.2.1 Quantitative input variables :

For the 100 quantitative input variables, correlation coefficients and scatter plots of variables colored according to the severity of sticking (y) were exploited to assess if any variable on its own could explain the observed sticking behavior. Using a 2D representation allowed for the assessment of bivariate combinations. For conciseness, only two excerpts of these plots are presented in Figure 18. Couples of input quantitative variables, (V65 and V66) and (V94 and V95), were chosen to highlight the power of subsequent multivariate data analysis (see section 3.5).



*Figure 18 : Scatter plots of quantitative input variables.*

Observations are colored from green to red according to the severity of sticking (y-variable). Two excerpts (A and B) are presented, corresponding to couples of input variables, (V65 and V66) and (V94 and V95), respectively.

All the univariate linear coefficients of correlation were less than 60% and the highest coefficients of determination (24% and 17%) were obtained for V65 and V94 respectively. Then, univariately, the best input parameter, V65, can only explain 24% of observed sticking variations. No clear trend of discrimination following sticking severity is also observed in bivariate assessments presented in Figure 18. Similar results are observed for the other quantitative parameters. Thus, no assessed single or bivariate combinations of quantitative input variables can substantially explain the observed fluctuations of sticking.

### 3.2.2 Qualitative input variables :

To assess the relationship between the 20 qualitative descriptors and the observed sticking, box plots highlighting the mean comparison of the number of sticking events between the different

classes of each qualitative input variable were exploited. For conciseness, only an excerpt of these representations is presented here, corresponding to qualitative variable V1, which displays the type of cleaning performed on a granulation machine before the granulation unit operation (Figure 19).

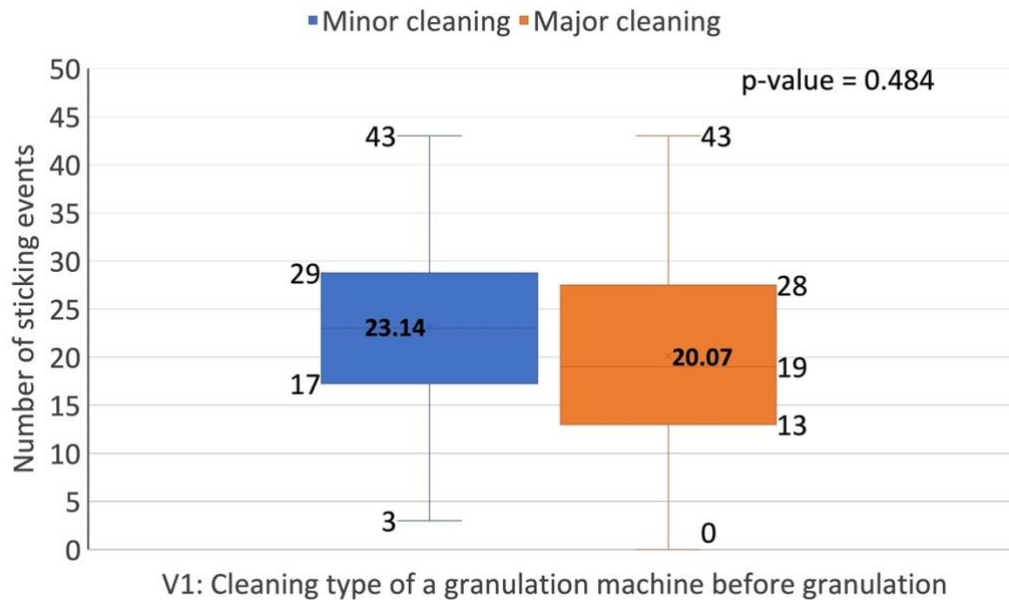


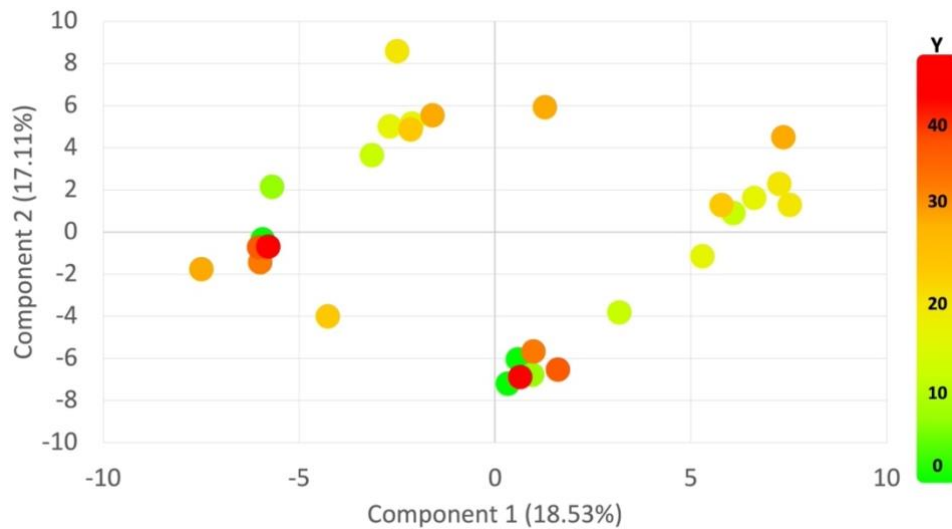
Figure 19 : Box plot displaying the mean comparison of the sticking events reported between different classes of qualitative input variables. The p-value corresponds to the result of a two-tailed student test.

Results show that an average of 23.14 sticking events are reported when minor cleaning is performed on the granulation machine, while 20.07 sticking events are observed when a major cleaning has been performed. These averages are not statistically different, with the p-value of the student’s t-test over 0.48. Similar conclusions are drawn from the remaining qualitative input variables; therefore, no single qualitative input variable can explain observed sticking.

The inability of any single variable, qualitative or quantitative, to sufficiently explain observed sticking is not unexpected, since sticking is known as a highly multivariate phenomenon. Therefore, combined information from several variables might be necessary to explain the fluctuations in its occurrence. To search for further insights into sticking, multivariate data analysis techniques were employed, starting with PCA, and followed by partial least squares regression for supervised modeling purposes.

### 3.3 Principal component analysis:

Figure 20 displays the score plot of the two first components of the PCA model built on data using all 121 variables and 28 observations.



*Figure 20 : Score plot of the two first PCA components.*

*Observations on the score plot are colored according to the sticking severity, from green for no sticky batches to red for severe sticky batches.*

The score plot shows two main clusters along both components 1 and 2. But these are not related to sticking severity, showing that another phenomenon explains better the most important source of variance in the dataset. Therefore, both components 1 and 2 cannot explain sticking. A total of eight components were retained and assessed, but no trend was identified. Since PCA is an unsupervised learning technique, it tends to describe and explain data following the most important dimension of variance. These dimensions are not necessarily the most interesting ones when studying a particular phenomenon. In terms of exploration, no outliers have been identified on the multivariate plans assessed.

### 3.4 Partial least squares regression:

#### 3.4.1 First PLS model

A first PLS model was built using all the variables and observations available. This model consisted of two components which are used as predictors to calibrate a regression model. Figure 21 displays the performances of the first PLS model both in calibration and validation. Table 3 shows the corresponding ANOVA results.

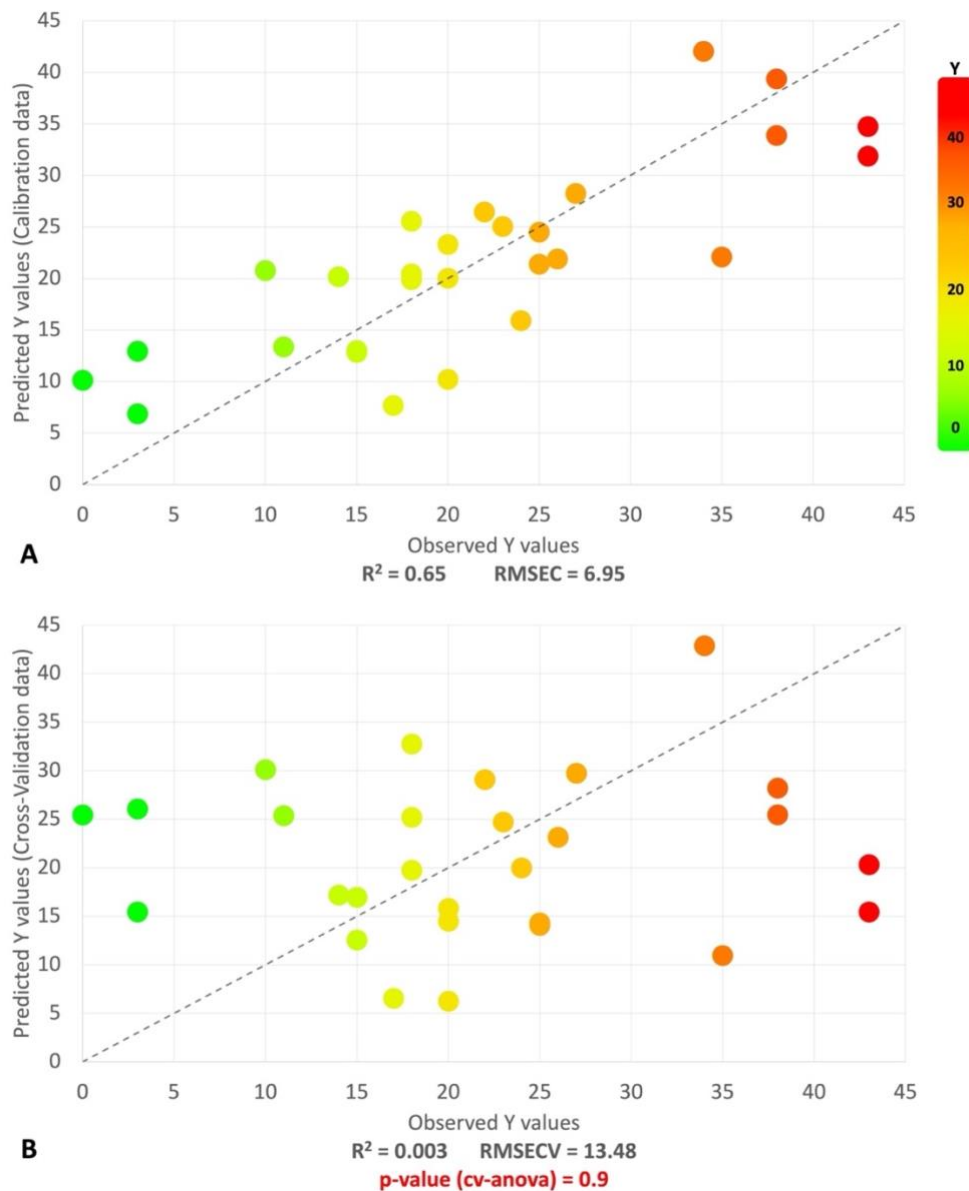


Figure 21 : Performances of the first PLS model in calibration and validation. Graphs A and B represent the scatter plots of observed y values versus predicted y values for calibration and cross-validation, respectively.

Table 3 : ANOVA results of the first PLS model

| Source of variation | Sum of squares | Degree of freedom | Mean squares | F-value | P-value |
|---------------------|----------------|-------------------|--------------|---------|---------|
| Regression          | 0.113          | 4                 | 0.03         | 0.0241  | 0.9988  |
| Residuals           | 26.89          | 23                | 1.17         |         |         |
| Total               | 27             | 27                | 1            |         |         |

As shown in Figure 21B, this first model fails in validation, with a RMSECV of 13.48. The root mean square error has the same unit as the y-variable. Thus, the first PLS model

predicts the number of sticking events with an average error of 13.48 events. Considering the maximum number of sticking events observed in the present study, which is approximately 43, this RMSECV leads to inadequate predictions in validation (Figure 21B). As a result, the ANOVA of the cross-validation procedure indicates the insignificance of the model, with a p-value significantly greater than the threshold of 0.05. The first PLS model is therefore more modeling noise than an underlying physical phenomenon.

However, when observations employed to calibrate the model are used as the validation set, a root mean square error of calibration (RMSEC) of 6.95 is obtained, allowing relatively good predictions with a coefficient of determination of 0.65 comparing to observed y values (Figure 21A). Therefore, the first PLS model is overfitted. Overfitting is generally due to the presence of many predictors which account for little or nothing in explaining the studied phenomenon. Such predictors favor good modeling of data instead of generalization, thus leading to better performance of the model in calibration than in validation. Therefore, before exploiting the PLS model, an optimization process was conducted to remove the overfitting.

### **3.4.2 Optimization of the PLS model:**

To overcome the overfitting of the first PLS model, a variable selection was performed. The same strategy was employed as was used in our earlier work to optimize the partial least squares discriminant analysis (PLS-DA) models [91]. This consisted of filtering the predictors' variables according to the variable importance in the projection (VIP) parameter. In the present study, the filtering VIP threshold was set to 1 and 3 filtering steps were performed. Figure 22 presents the evolution of the models' performances through the optimization process. Table 4 shows the corresponding ANOVA results.

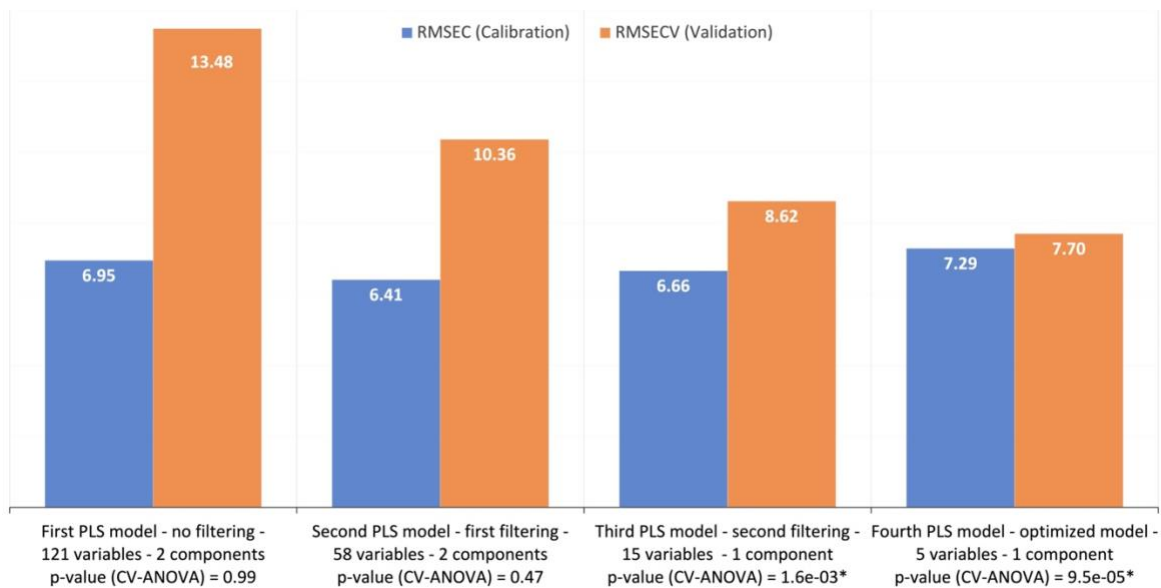


Figure 22 : Histogram of the evolution of RMSEC and RMSECV during the optimization procedure by filtering through VIP (threshold set to 1). The p-values displayed show the significance of the model as given by the ANOVA of the cross-validation results.

Table 4 : ANOVA results of PLS models developed through the optimization process.

| Models           | Source of variation | Sum of squares | Degree of freedom | Mean squares | F-value | P-value   |
|------------------|---------------------|----------------|-------------------|--------------|---------|-----------|
| First PLS model  | Regression          | 0.113          | 4                 | 0.03         | 0.0241  | 0.9988    |
|                  | Residuals           | 26.89          | 23                | 1.17         |         |           |
|                  | Total               | 27             | 27                | 1            |         |           |
| Second PLS model | Regression          | 3.713          | 4                 | 0.93         | 0.9167  | 0.4710    |
|                  | Residuals           | 23.29          | 23                | 1.01         |         |           |
|                  | Total               | 27             | 27                | 1            |         |           |
| Third PLS model  | Regression          | 10.87          | 2                 | 5.43         | 8.4192  | 0.001601* |
|                  | Residuals           | 16.13          | 25                | 0.65         |         |           |
|                  | Total               | 27             | 27                | 1            |         |           |
| Fourth PLS model | Regression          | 14.13          | 2                 | 7.06         | 13.723  | 0.000095* |
|                  | Residuals           | 12.87          | 25                | 0.51         |         |           |
|                  | Total               | 27             | 27                | 1            |         |           |

Successive filtering through VIP led to an optimization of the model. The values of the root mean squared error in calibration and validation converge after each filtering step

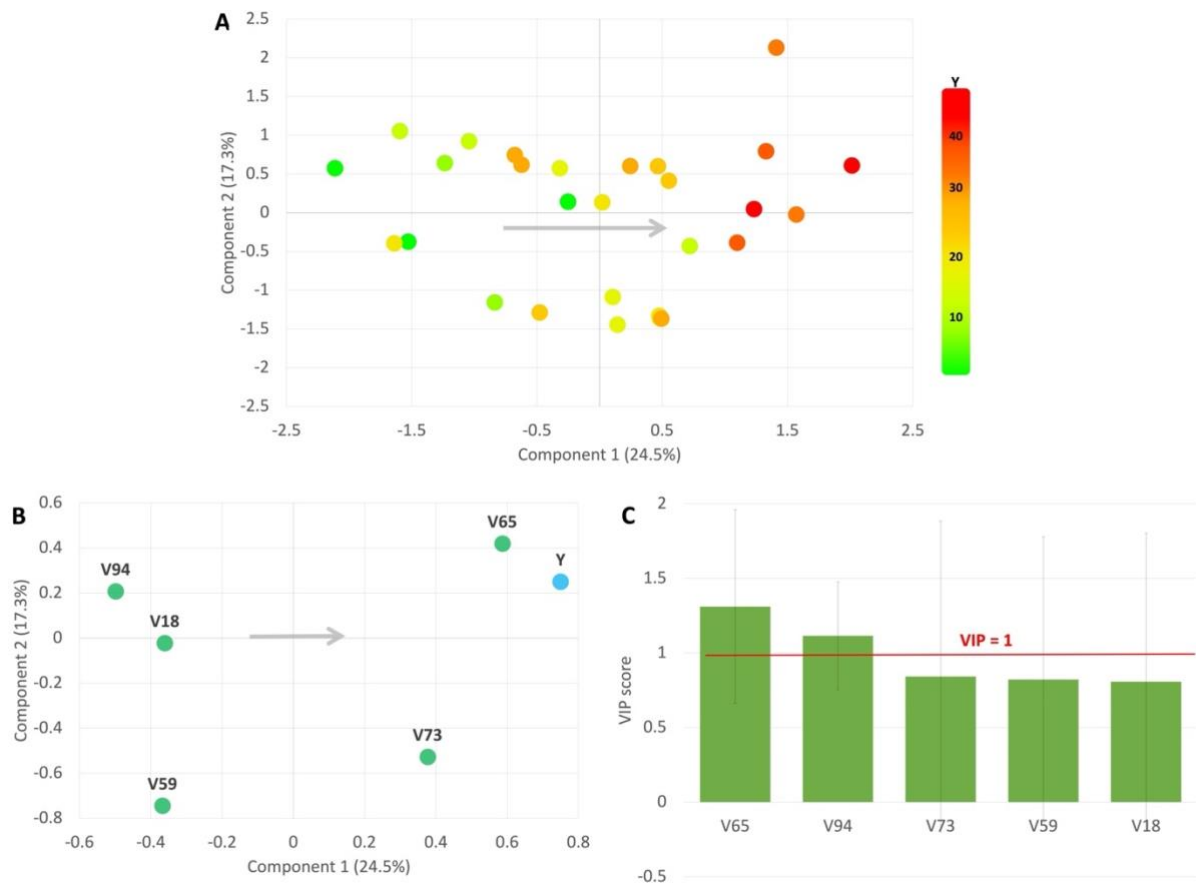


(Figure 22). Two filtering steps were necessary to obtain a satisfactory model ('Third PLS model') that is significant in validation (p-value CV-ANOVA < 0.01) with a RMSEC of 6.66 and a RMSECV of 8.62. A total of 15 variables were selected to compute this model. The Third PLS model consists of a single component which performs better than both the First and Second PLS models, each of which comprises two components. A third filtering of variables highlighting a VIP less than the threshold (VIP = 1) led to the selection of 5 predictors allowing the calibration of a fourth PLS model with the best predictive performances. From Third to Fourth PLS model, the RMSEC increases from 6.66 to 7.29 and the RMSECV decreases from 8.62 to 7.70, leading to closer scores in both calibration and validation for the Fourth PLS model, which was also tested significant with a p-value (CV-ANOVA) less than 0.01. These results confirm the use of VIP scores to optimize PLS models [77,90,91]. The resulting Fourth PLS model, the optimized PLS model, was exploited to gain insights into sticking of the studied ibuprofen and methocarbamol-based formulation.

### **3.4.3 Optimized PLS model:**

#### **- Explicative insights :**

The descriptive outcomes of the optimized PLS model are exploited to identify the most important contributors to sticking occurrences. Figure 23 depicts the score, loading, and VIP plots of the optimized PLS model.



*Figure 23 : Overview of the optimized PLS model.*

*Graphs A, B, and C display the model's score, loading, and VIP plots, respectively. Observations on score plot are colored according to the severity of sticking. The error bars on the VIP plot represent the confidence interval at 95% computed with the standard error obtained during cross-validation. Gray arrows in A and B indicate sticking severity direction.*

According to the optimization results, the first component of the optimized model is enough to explain and predict sticking behavior (Figure 23). Along this component on the score plot (Figure 23A), a clear trend of discrimination following the severity of sticking is observed. This first component is therefore linked to the observed sticking fluctuations.

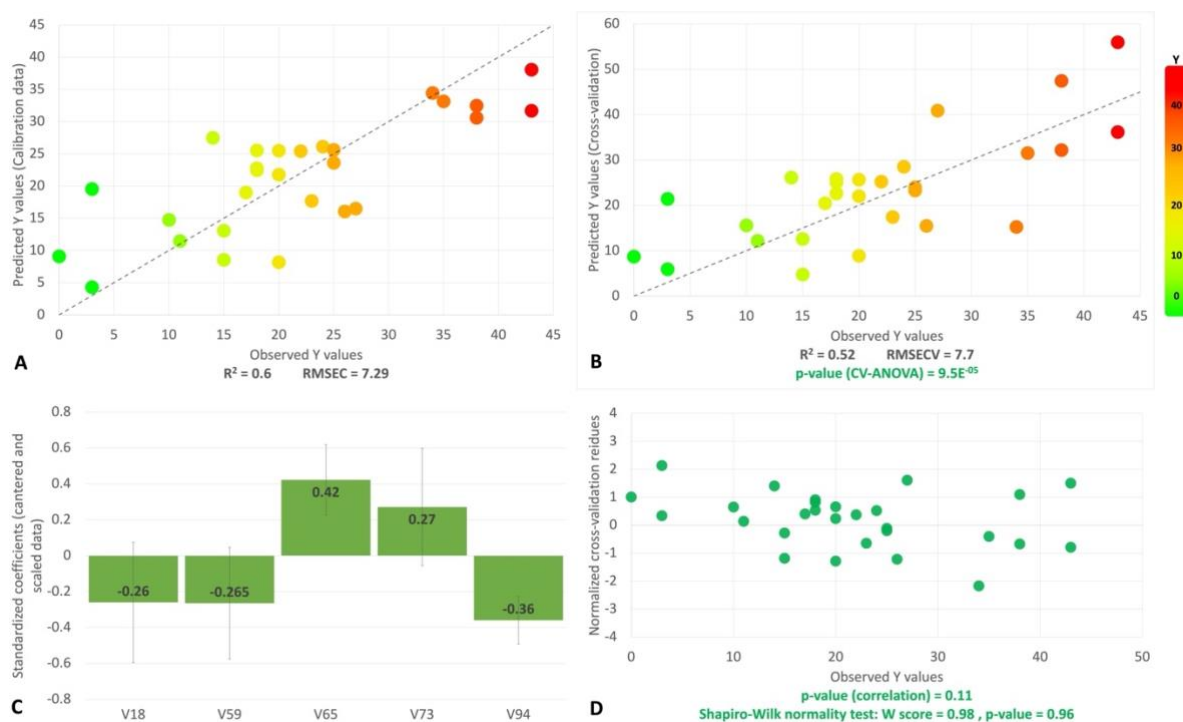
It appears that variables V65 and V94 are the most important contributors to the first component, thus explaining the sticking that was observed (Figure 23B). Variable V65 represents the duration of the colored layer granulation in seconds and variable V94 is the variation of bilayer compression force in percentage. The loading plot also suggests that variables V18, V59, and V73 may also be important. These variables represent the average temperature in the mixing room during the mixing operation of the non-colored layer (V18), the temperature in the Fitzmill (V59), and its feeding speed (V73) during the colored layer granulation. To clarify the implication of these variables, a PLS VIP plot was used. According to this graph, all 5 variables included in the model are important as they all exhibit VIP scores

greater than 0.8. However, following the VIP confidence interval at 95%, only variables V65 and V94 are significantly contributing to the observed sticking behavior. According to the loading plot, the opposition of these two variables governs the fluctuations of sticking on the studied product (Figure 23B). Sticking is less frequent during compression after a short time of granulation of the color layer and when the bilayer compression force applied to the powder bed deviates greatly from its reference value. The effect of these two variables (V65 and V94) is discussed in detail in section 3.5. More data are necessary to confirm the role of the remaining variables (V18, V59, and V73), which are identified as important but not significantly contributing to the observed sticking.

These results show that PLS helps provide insights into tablet sticking in cases where batch stickiness is described by a quantitative variable. Starting from several hundred variables, it leads to the identification of 5 variables for their importance in sticking fluctuations, including V65 and V94, which exhibit a significant effect on sticking behavior. The supervised nature of PLS makes it even more interesting in cases where unsupervised techniques such as PCA failed to highlight clusters about sticking, as was observed in the present study.

- **Predictive insights:**

The first component of the optimized PLS model which was found to explain sticking behavior was employed to compute a predictive model. The main outcomes of this multivariate regression model are presented in Figure 24.



*Figure 24 : Predictive performances of the optimized PLS model.*

Graphs A and B represent the scatter plots of observed y values versus predicted y values for the calibration and cross-validation, respectively. Observations are colored according to the severity of sticking. Graph C displays the standardized coefficients of the model when predictors are centered and scaled. The error bars on the coefficient plot represent the confidence interval at 95% computed with the standard error obtained during the cross-validation procedure. Graph D displays the relation between normalized cross-validation residues and observed y values.

Unlike the first PLS model, the optimized model presents similar behaviors in calibration and validation (Figure 24A and Figure 24B). The number of sticking events is predicted with an average error of 7.29 and 7.7 events of sticking in calibration and validation, respectively (Figure 24A and Figure 24B). The closeness of these error values shows the ability of the model to generalize in terms of the physics of the sticking phenomenon. An ANOVA of the validation residues confirms the significance of the model with a p-value of  $9.5e^{-05}$  (Figure 22). The standardized coefficients of the optimized PLS model can be employed to forecast the stickiness of new batches (Figure 24C).

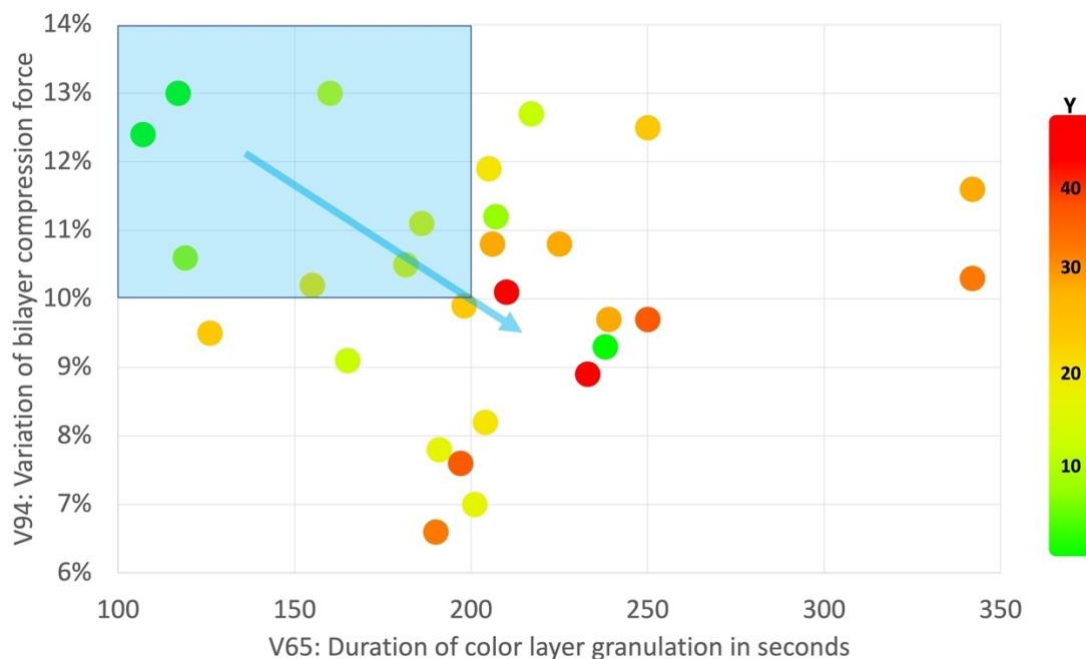
In our earlier study using a qualitative descriptor of sticking, significant classification models were also developed to forecast the sticking behavior of an ibuprofen-based formulation [91]. But, due to the dependence of the classification error on the target variable's classes, they were not exploitable [91]. The observations in the previous study were divided into two classes: batches affected and unaffected by sticking. The developed models were shown to systematically misclassify batches affected by sticking (65-81% of error) more than those unaffected by sticking (0-30% of error) [91]. This was attributed to an inaccurate

distinction between the two classes of batches, as the studied ibuprofen-based formulation was not very much affected by sticking and the severity was limited when the phenomenon occurs. To overcome this limitation, a quantitative target variable with a large range of sticking behavior was proposed [91]. Since the present study consists of such an approach, the relevance of the proposed solution was assessed by analyzing the validation residues of the optimized PLS model. This was conclusive, as the validation residues of the optimized PLS model were found to be normally distributed, with a p-value of Shapiro–Wilks’ normality test of 0.96, and they are also independent of the y-variable as it is shown in Figure 24D. This graph highlights a poor correlation between the normalized residues of the cross-validation procedure and the observed y-variable values, which appears insignificant with a p-value of 0.11. Therefore, unlike the classification models of the previous study, the optimized PLS model of the present study is free from any prediction bias, confirming the hypothesis suggested to explain the limitations encountered during earlier work [91].

Based on these performance indices, the optimized PLS model appears robust enough to be employed as a predictive tool to forecast sticking events on the studied ibuprofen and methocarbamol-based formulation. However, it is important to note that the average prediction errors remains quite important considering the range of the y-variable. Indeed, a root mean square error of 7.7 in validation could lead to inaccurate predictions. The model should then be improved first before being used in production. The persistent prediction error is probably due to the absence in the data of some important parameters relating to sticking. The present work was conducted on historical data. Data relating to some variables known to play a key role in the sticking phenomenon were not systematically gathered or archived and thus were not available to be included in the analysis. Such variables include punch surface information, particle properties at different steps of the process, and detailed information about the drying step.

### **3.5 Discussion of the significant insights:**

Data analysis led to the identification of 2 variables (V65 and V94) as significantly contributing to the explanation of observed sticking fluctuations. V65 represents the duration of colored layer granulation in seconds and V94 depicts the variation of the bilayer compression force in percentage. Figure 25 presents a scatter plot that shows how these two variables impact sticking.



*Figure 25 : Scatter plot of sticking behavior following the duration of colored layer granulation (V65) and the variation of bilayer compression force (V94). The blue arrow is indicating the direction of increasing severity. The blue rectangle in the upper left corner indicates the optimal operating range for V65 and V94.*

Figure 25 shows that variables V65 and V94 are enough to explain much of the observed sticking for the studied ibuprofen and methocarbamol-based formulation. Following the decrease of the variation of bilayer compression force and the increase of granulation duration for the colored layer, observations cluster according to the severity of sticking from green through yellow and orange to red. This highlights the power of multivariate data analysis in obtaining insights about sticking. Indeed, univariate data analysis (section 3.2) could also have led to the identification of V65 and V94. But this was not the case because the V65/V94 combination was not assessed. Instead of each variable on its own, it is their combination that could lead to a significant explanation of sticking. To have identified such insight during a univariate or bivariate analysis of data, it would have been necessary to visualize all possible two-dimensional representations, corresponding to the analysis of approximately 7140 graphs for the dataset of this study. One could have also used correlation or determination coefficients to guide the bivariate assessment and get to the most relevant combinations to visualize. But still, this would have remained only a visual assessment on a bivariate plot, which is somehow the simplest multivariate analysis. To qualify and understand entirely the relevance of the insight, multivariate modeling would have remained necessary. The univariate linear  $R^2$  of V65 and V94 are respectively 24% and 17%, which in a strictly cartesian way sum up to only 41%. Using multivariate data modeling, this study shows that the association of V65 and V94

explains around 60% of the observed sticking variations, which is far greater than their individual or simple additive contributions. In addition, if more than 2 variables were required to explain the observed sticking, no two-dimensional representations would have detected it. MVDA techniques avoid all these concerns by summarizing on a few dimensions the most valuable information captured in a large dataset either with or without supervision (PLS and PCA, respectively, in the present study).

The duration of the colored layer granulation (V65) is positively correlated with the sticking severity. The correlation coefficient is about 49%, which is tested significant with a p-value of 0.007. Longer durations of the colored layer granulation, therefore, led to more sticking. Such a relationship has not yet been studied in the literature. However, some work has discussed the importance of particle properties in sticking, especially API particle properties [20,38,39,46,54,57,63,64,73,81,92,95,96]. The duration of granulation probably impacts some of the formulation properties such as particle size distribution, surface area and energy, shape distribution, and water content, which could then lead to sticking.

Smaller particle sizes have been pointed out by several authors as one of the main contributors to sticking because of their physical, electrostatic, and chemical properties [10,47,52–55,92]. It would then be of interest to investigate how a longer granulation duration could impact the particle size distribution of the studied formulation after the subsequent milling step of the process. This could be achieved experimentally, but a planned collection of data at an industrial scale with systematic sampling and analysis of intermediates could also be a good method. Furthermore, research has shown that the effect of particle size generally depends on punch surface rugosity and coating [11–13,51,63], compression force [35,46,62,68,81], and tablet tensile strength [13,37,64,81]. Therefore, these parameters need to be taken into account in further investigations of the relationship between granulation duration and sticking through particle properties, especially particle size.

The water content of the formulation powder could also be affected by the duration of granulation and then leads to the sticking occurrence. It has been suggested by several authors that increasing water content of a formulation could lead to sticking through the creation of capillary bridges [9,20,26,39,72–74]. Since the studied formulation is wet granulated, it would be of interest to study how longer granulation could impact the final water content of the formulation after the subsequent drying steps of the manufacturing process.

Furthermore, since only the duration of the colored layer granulation contributes significantly to sticking, it could be suggested that the colored layer is more affected by sticking than the non-colored layer of the studied formulation. This hypothesis is also corroborated by the fact that, among the 5 variables isolated in the present study as being important in explaining sticking fluctuations, 4 of them are all related to the colored layer. These variables are the bilayer compression force (V94), as the colored layer is the last one to be compressed, the granulation duration (V65), the temperature in the Fitzmill (V59), and its feeding speed (V73) during the colored layer granulation and milling steps. Moreover, it was observed that most of the sticking events reported by operators occur on the upper punch, which is employed for the application of the bilayer compression force directly on the colored layer. Thus, further investigations on the studied formulation should pay particular attention to the particle properties of the granulated colored layer formulation powder.

The second variable identified as significantly contributing to sticking is the variation of bilayer compression force (V94), which was found to be negatively correlated with the sticking severity. The corresponding correlation coefficient is approximately  $-41.8\%$ , which is tested significant with a p-value of  $0.026\%$ . The variation of the bilayer compression force represents the proportion of deviation of the bilayer compression force from its reference value. It consists of an absolute value which is directly computed by the tablet press and the direction of the variation is not captured. The raw data generated by the tablet press are not stored and therefore could not be employed to identify the predominant direction of deviation. Since the present study was performed on historical data, it was impossible to determine the predominant direction of the variation. Therefore, it is either an increase or a decrease in the bilayer compression force which leads to the mitigation of sticking occurrences of the studied formulation. This finding is in accordance with the literature, as the compression force has been extensively studied and related to sticking by several authors [11–14,35,36,51,62–64,68].

Analysis of literature reports has shown that the role of compression force in sticking depends on the nature of the formulation and the properties of the punches' surfaces. On one hand, the compression force was reported to be proportional to the sticking importance for several formulations such as celecoxib, sorbitol, mefenamic acid, glyburide, or tofacitinib-based formulations [11,64]. This positive correlation was mainly attributed to the mechanical properties of API, especially their ductility, where more plastic deformations occur and lead to a larger coverage of punch surface by API particles [11,64]. On the other hand, it was shown that an increase in compression force led to the mitigation of sticking propensity for



formulations based on acetylsalicylic acid, lactose, n-butyl p-hydroxybenzoate, ABT279, or pozanicline [13,35,36,51,62,63]. This beneficial effect on sticking was generally explained by a strong internal cohesion in the resulting tablet induced by strong compression force. Therefore, knowledge of the physical properties of particles in a particular formulation is important to understand the effect of compression force on its sticking propensity. Furthermore, for ibuprofen-based formulations, which are the most studied in the literature, the effect of compression force was reported to change according to the properties of punches' surfaces [10–14,16,37,42,51,63,97]. Literature results show the existence of significant interactions between compression force and several properties of punches' surfaces such as rugosity, homogeneity, hydrophilic behavior, and the nature of coating [10–14,16,42,68,69,97]. Since these properties were not investigated in the present study, it would be of interest to investigate their role in the sticking of the studied formulation, especially the existence of an eventual interaction with the bilayer compression force.

Finally, based on the main insights drawn out from the present study, an optimal operating range could be proposed to mitigate sticking occurrences on the studied ibuprofen and methocarbamol-based formulation. The duration of the colored layer granulation (V65) should be set between 100 and 200 seconds and the bilayer compression force should be increased or decreased by about 10 to 14% from the current reference according to the observed predominant direction of the variation (see Figure 25). The predominant direction of the compression force variation could be determined by collecting and analyzing data from upcoming batches or by studying experimentally the effect of the compression force on the occurrence of sticking.

#### **4. Conclusion :**

The present work was intended to introduce an industrial data analysis approach to gain insights into tablet sticking in cases where sticking is described by a quantitative variable. The sticking behavior of an ibuprofen and methocarbamol-based formulation was investigated.

First, using an analysis methodology based on PLS regression modeling, the relevance of the proposed approach was confirmed. The granulation duration of the colored layer of the tablet and the bilayer compression force were identified as significantly governing 60% of the sticking variations observed in the studied formulation. Based on the results of this study, an optimal operating range was proposed for both identified variables. The duration of the colored layer granulation should be set between 100 and 200 seconds and according to the predominant direction of variation of the bilayer compression force, it should be increased or decreased by about 10 to 14% from its current reference. These recommendations could be tested experimentally to acquire more knowledge on the effects of identified variables and then implemented on an industrial scale to mitigate the occurrences of sticking in the studied formulation.

Second, a predictive regression model to forecast sticking occurrences was also calibrated and validated through a cross-validation procedure. This model consists of 5 input variables and predicts the number of sticking events during the compression of a batch, with an average error of 7.7 events of sticking. The accuracy of this model could be improved by gathering more observations or by studying the variables that were not assessed in the present study.

Finally, it was shown that the PLS model developed in the present study is free from any prediction bias compared to the classification models developed in our earlier work, where the classification error was dependent on the qualitative y-variable. This finding validates the hypothesis of a fine line between the two modalities of the qualitative descriptor as the root cause explaining the limitations encountered during the earlier work.

## **5. Declaration of competing interest:**

The authors report no conflicts of interest.

## **6. Acknowledgments:**

The authors would like to thank Jose M. Montenegro Alvarado as well as the production and engineering team of the Pfizer plant at Toluca in Mexico, especially both Carmen Manjarrez and Martha Alexia Vela for their technical support and all their efforts to help accessing to the process and provision of data used in this work.

## CHAPITRE 5 : CONCLUSIONS GENERALES ET PERSPECTIVES

L'objectif de la présente thèse de doctorat était d'introduire, d'évaluer et de discuter une nouvelle approche d'étude du collage de poudres pharmaceutiques. Celle-ci consiste en l'exploitation des données industrielles de fabrication des comprimés pour développer une compréhension multivariée du phénomène. Pour ce faire, nous avons mené deux études de cas impliquant deux formulations différentes et deux façons différentes d'évaluer le collage.

La première étude de cas a investigué une formulation à base d'ibuprofène où le collage était décrit par une variable qualitative statuant l'observation ou non du collage lors de la compression d'un lot. Nous avons collecté et analysé une centaine de variables prédictives sur 71 lots industriels de cette formulation en utilisant des techniques d'analyse multivariée de type discrimination. Nous avons ainsi pu isoler un total de six variables prédictives comme contributeurs significatifs au collage de la formulation étudiée. Il s'agit de la proportion de lubrifiant de la formulation, la durée des arrêts en cours de compression ainsi que la température et l'humidité de certaines salles d'opération ou de stockage. En accord avec les données de la littérature, ces résultats confortent le rôle de la lubrification, du contenu en eau des poudres et de la faible température de fusion de l'ibuprofène dans sa propension au collage. Cette étude de cas confirme ainsi la pertinence de la nouvelle approche d'étude du collage à échelle industrielle. Cependant, les modèles de classification développés et exploités dans cette première étude de cas avaient de faibles variances interclasses, mais surtout, ils souffraient d'importants biais de prédiction. Nous avons postulé qu'une description élaborée de la sévérité du collage et l'intégration de plus de paramètres explicatifs permettraient de pallier ces limites.

La deuxième étude de cas s'est ainsi attelée à investiguer trois fois plus de paramètres prédictifs sur une formulation différente à base d'ibuprofène et de metocarbamol où le collage était décrit de façon plus détaillée avec une variable quantitative se rapportant à sa sévérité. Cette formulation est utilisée pour produire des comprimés bicouches : une couche blanche et une couche colorée. Nous avons collecté et analysé 28 lots industriels de cette formulation en utilisant la technique d'analyse multivariée de régression sur les moindres carrés partiels. Nous avons ainsi pu montrer que l'association de deux paramètres opératoires permet d'expliquer 60 % des fluctuations du collage de la formulation étudiée. Il s'agit de la durée de la granulation de la couche colorée et de la force de compression bicouche. Nous avons aussi confirmé avec

cette deuxième étude que l'utilisation d'un descripteur quantitatif de la sévérité du collage et l'analyse de plus de paramètres permettaient de prévenir les biais et limites observées dans les modèles de classification de la première étude.

De façon complémentaire, les résultats de ces deux études de cas font la preuve de concept d'une nouvelle approche d'étude du collage basé sur l'exploitation des données industrielles de production des comprimés. Les bases d'un déploiement efficace de cette nouvelle approche ont aussi été élaborées, notamment une méthodologie systématique basée principalement sur les techniques d'analyse multivariée. Les deux études de cas montrent une application concrète de cette méthodologie lorsque le collage est décrit par une variable qualitative et quantitative respectivement. Nous avons également discuté les avantages, limites et perspectives de la nouvelle approche d'étude du collage. Elle permet notamment à la communauté scientifique d'entrevoir une nouvelle perspective peu coûteuse, mais surtout plus efficace dans l'étude, la compréhension et la résolution du collage.

L'étude du collage avec une approche d'analyse des données industrielles permet tout d'abord d'investiguer simultanément à moindre coût beaucoup plus de paramètres qu'une approche expérimentale. En utilisant un plan factoriel complet de seulement deux niveaux par paramètre, il faut un minimum de 32 essais différents pour évaluer expérimentalement de façon simultanée le rôle de seulement cinq paramètres dans le collage d'une formulation donnée. En face, comme démontré dans les études de cas, la nouvelle approche introduite par cette thèse ne requiert aucune expérimentation pour évaluer simultanément plusieurs dizaines de paramètres. Ces paramètres ont trait à tous les aspects de la fabrication des comprimés et peuvent notamment inclure des éléments particuliers qui sont économiquement ou techniquement difficiles à reproduire et à étudier au laboratoire, comme l'âge et l'usure des poinçons, les longues séances de compression à grande vitesse, les faibles temps de contact ou encore les nombreux paramètres opératoires des unités d'opération en amont de la compression. L'étude du collage directement à échelle industrielle permet aussi de prévenir les problématiques de représentativité et de reproductibilité qui handicapent les modèles d'étude expérimentale à échelle laboratoire de décrire une compréhension globale du phénomène. De plus, les faibles tolérances de variation des paramètres à l'échelle industrielle permettent de mettre l'accent sur les variables auxquelles le collage est vraiment sensible, ce qui peut substantiellement aider dans la recherche de solutions.

Les limites au déploiement de cette nouvelle approche ont aussi été discutées dans cette thèse. Il s'agit principalement de la génération et de la disponibilité de données pertinentes en

quantité et en qualité pour l'étude du collage. En effet, la fréquence des productions industrielles peut rallonger les délais de disponibilité d'un nombre pertinent d'observations. Aussi, à cause des réglementations strictes entourant la fabrication de produits pharmaceutiques, l'accès aux données détaillées sur certains paramètres peut être plus ou moins ardu en fonction des technologies utilisées par l'industriel. De même, un descripteur pertinent du collage est nécessaire pour obtenir de bons résultats. Une autre limite de la nouvelle approche d'étude du collage est son incapacité à décrire et expliquer fondamentalement la compréhension multivariée acquise.

Pour juguler ces limites, nous avons proposé une collecte planifiée des données industrielles qui s'efforcerait à définir et collecter des indicateurs pertinents à la sévérité du collage, mais aussi à collecter les données de tous les paramètres jugés utiles et nécessaires autant pour expliquer le collage, mais surtout pour décrire fondamentalement la compréhension multivariée développée. Ce dernier aspect nécessite par exemple de récupérer des échantillons de formulation à des moments précis de la production pour de futures analyses. Un autre moyen pour pallier les différentes limites observées est de compléter la collecte et l'analyse de données historiques par des essais au laboratoire en utilisant la méthodologie des plans d'expérience.

Au-delà de ces options de mitigation des limites, il existe également d'autres perspectives intéressantes permettant de bonifier les avantages de la nouvelle approche d'étude du collage, notamment dans le cadre de la prédiction et de la résolution du collage. Il serait par exemple intéressant d'investiguer à l'avenir la modélisation de la dynamique conduisant au collage au cours de la compression. Cette alternative consiste à utiliser des lectures temporelles comme observations au lieu d'un lot en entier, générant une quantité importante de données qui permettront une analyse fine de la dynamique conduisant au collage. Cela fournirait une compréhension plus détaillée du phénomène et pourrait surtout s'avérer plus pertinent à sa prédiction et à sa mitigation en cours de compression. Aussi, sur la base de ces modèles, il serait tout à fait envisageable de produire des systèmes de contrôle de procédés en ligne. Enfin, la nouvelle approche introduite dans cette thèse pourrait surtout s'appliquer de façon efficace à d'autres problématiques rencontrées dans l'industrie pharmaceutique et bien au-delà dans d'autres industries. Cette thèse de doctorat montre qu'une solide stratégie de gestion et d'analyse de données permettrait aux industriels et aux académiques de travailler en concert pour développer des connaissances et des outils prédictifs pertinents à de nombreuses problématiques d'intérêt social et industriel.

# RÉFÉRENCES

- [1] G.D. Tovey, T.M. Jones CBE, S. Stegemann, S. Tindal, P. Kendal, B. Carlin, M. Whiteman, C. Seiler, S. Wicks, M. Orlu, T. Arvinte, D. Adair, C. Wilson, *Pharmaceutical Formulation: The Science and Technology of Dosage Forms*, 1st ed., The Royal Society of Chemistry, London, 2018.
- [2] B. Boussaoud, [Diagnosis in pharmaceutical engineering], MSc Thesis, LAGEP, Université de Lyon 1, 2003. <https://dumas.ccsd.cnrs.fr/dumas-00353848v1>.
- [3] L. Meeus, Direct Compression Versus Granulation, *Pharm Technol Eur.* 23 (2011).
- [4] H.M.M. Santos, J.J.M.S. Sousa, Tablet Compression, in: S.C. Gad (Ed.), *Pharmaceutical Manufacturing Handbook: Production and Processes*, John Wiley & Sons, Inc., Hoboken, USA, 2008: pp. 1133–1163. <https://doi.org/10.1002/9780470259818.ch28>.
- [5] I. Saniocki, New Insights into Tablet Sticking: Characterization and Quantification of Sticking to Punch Surfaces during Tablet Manufacture by Direct Compaction, PhD Thesis, Hamburg University, 2014. <https://ediss.sub.uni-hamburg.de/handle/ediss/5620>.
- [6] M.K. Iqbal, P.K. Singh, M. Shuaib, A. Iqbal, M. Singh, Recent advances in direct compression technique for pharmaceutical tablet formulation, *Int J Pharm Res Dev.* 6 (2014) 49–57.
- [7] A. Tita-Goldstein, Mise en forme des poudres par compression : influence du procédé et de la formulation pour la maîtrise des propriétés d’usage, PhD Thesis, Université de Lorraine, 2013. <https://hal.univ-lorraine.fr/tel-01750532/document>.
- [8] J. v. Thomas, Evaluation and study on the adhesion of powder onto punch faces during tablet compaction, MSc Thesis, Drexel University, 2015. <https://core.ac.uk/download/pdf/190324118.pdf>.
- [9] S. Chattoraj, P. Daugherty, T. McDermott, A. Olsofsky, W.J. Roth, M. Tobyn, Sticking and Picking in Pharmaceutical Tablet Compression: An IQ Consortium Review, *J Pharm Sci.* 107 (2018) 2267–2282. <https://doi.org/10.1016/j.xphs.2018.04.029>.
- [10] C. Al-Karawi, Multifactorial analyses of the sticking tendency of ibuprofen and ibuprofen sodium dihydrate tablet formulations, Thesis, Hamburg University, 2018. <https://ediss.sub.uni-hamburg.de/handle/ediss/7860>.

- [11] M. Roberts, J.L. Ford, G.S. MacLeod, J.T. Fell, G.W. Smith, P.H. Rowe, Effects of surface roughness and chrome plating of punch tips on the sticking tendencies of model ibuprofen formulations, *J Pharm Pharmacol.* 55 (2003) 1223–1228. <https://doi.org/10.1211/0022357021684>.
- [12] M. Roberts, J.L. Ford, P.H. Rowe, A.M. Dyas, G.S. MacLeod, J.T. Fell, G.W. Smith, Effect of lubricant type and concentration on the punch tip adherence of model ibuprofen formulations, *J Pharm Pharmacol.* 56 (2004) 299–305. <https://doi.org/10.1211/0022357022827>.
- [13] C. Al-Karawi, C.S. Leopold, A comparative study on the sticking tendency of ibuprofen and ibuprofen sodium dihydrate to differently coated tablet punches, *Eur J Pharm Biopharm.* 128 (2018) 107–118. <https://doi.org/10.1016/j.ejpb.2018.04.004>.
- [14] C. Al-Karawi, I. Lukášová, A. Sakmann, C.S. Leopold, Novel aspects on the direct compaction of ibuprofen with special focus on sticking, *Powder Technol.* 317 (2017) 370–380. <https://doi.org/10.1016/j.powtec.2017.05.014>.
- [15] Y. Takeuchi, Y. Murase, K. Tahara, H. Takeuchi, Impact of surface roughness of pre-treated punches and powder properties on prevention of sticking during pharmaceutical tableting, *J Drug Deliv Sci Technol.* 60 (2020) 101999. <https://doi.org/10.1016/J.JDDST.2020.101999>.
- [16] S. Swaminathan, B. Ramey, J. Hilden, C. Wassgren, Characterizing the powder punch-face adhesive interaction during the unloading phase of powder compaction, *Powder Technol.* 315 (2017) 410–421. <https://doi.org/10.1016/j.powtec.2017.04.003>.
- [17] D.B. Patel, R.H. Dave, Predicting lubricants effect on tablet sticking using ketoprofen as model drug and evaluating sticking propensity using different metals and powder rheology, *Int J Pharm.* 606 (2021) 120913. <https://doi.org/10.1016/j.ijpharm.2021.120913>.
- [18] T. Osamura, Y. Takeuchi, R. Onodera, M. Kitamura, Y. Takahashi, K. Tahara, H. Takeuchi, Prediction of effects of punch shapes on tableting failure by using a multi-functional single-punch tablet press, *Asian J Pharm Sci.* 12 (2017) 412–417. <https://doi.org/10.1016/j.ajps.2017.05.001>.
- [19] K. Reed, C. Davies, K. Kelly, Tablet sticking: Using a ‘compression toolbox’ to assess multiple tooling coatings options, *Powder Technol.* 285 (2015) 103–109. <https://doi.org/10.1016/j.powtec.2015.05.005>.



- [20] M. Bunker, J. Zhang, R. Blanchard, C.J. Roberts, Characterising the surface adhesive behavior of tablet tooling components by atomic force microscopy, *Drug Dev Ind Pharm.* 37 (2011) 875–885. <https://doi.org/10.3109/03639045.2010.546402>.
- [21] M. Badal Tejedor, N. Nordgren, M. Schuleit, A. Millqvist-Fureby, M.W. Rutland, AFM Colloidal Probe Measurements Implicate Capillary Condensation in Punch–Particle Surface Interactions during Tableting, *Langmuir.* 33 (2017) 13180–13188. <https://doi.org/10.1021/acs.langmuir.7b02189>.
- [22] J.S. Saddik, R.H. Dave, Evaluation of powder rheology as a potential tool to predict tablet sticking, *Powder Technol.* 386 (2021) 298–306. <https://doi.org/10.1016/J.POWTEC.2021.03.062>.
- [23] H. Ando, Y. Hara, K. Sato, M. Dohi, T. Hakomori, E. Yonemochi, New approach to optimizing risk management of the sticking problem using scale-independent critical material attributes and the quantitative process parameter, *Int J Pharm.* 577 (2020). <https://doi.org/10.1016/J.IJPHARM.2020.119032>.
- [24] H. Juvonen, O. Antikainen, M. Lemmens, H. Ehlers, A. Juppo, The effect of relative humidity and formulation variables on chewable xylitol-sorbitol tablets, *Int J Pharm.* 601 (2021) 120 573. <https://doi.org/10.1016/j.ijpharm.2021.120573>.
- [25] S. Paul, C. Wang, K. Wang, C.C. Sun, Reduced Punch Sticking Propensity of Acesulfame by Salt Formation: Role of Crystal Mechanical Property and Surface Chemistry, *Mol Pharm.* 16 (2019) 2700–2707. <https://doi.org/10.1021/acs.molpharmaceut.9b00247>.
- [26] N.F. Costa, M.G. Paulo, H.P. Diogo, J.F. Pinto, Solving a sticking related tablet problem by multivariate statistics and computational tomographic analysis, *Powder Technol.* 367 (2020) 456–463. <https://doi.org/10.1016/j.powtec.2020.03.063>.
- [27] K.D. Khedekar, To study the characteristics of sticking and non-sticking formulations and evaluate powder layer sticking while conducting compression using punches with different head flats, MSc Thesis, LONG ISLAND UNIVERSITY, 2017. <https://www.proquest.com/openview/dd151e8a15a063176a747cb179013d99/1?pq-origsite=gscholar&cbl=18750>.
- [28] G.K. Bolhuis, N. Anthony Armstrong, Excipients for Direct Compaction—an Update, *Pharm Dev Technol.* 11 (2006) 111–124. <https://doi.org/10.1080/10837450500464255>.
- [29] Z. Li, L.J. Zhao, X. Lin, L. Shen, Y. Feng, Direct compaction: An update of materials, trouble-shooting, and application, *Int J Pharm.* 529 (2017) 543–556. <https://doi.org/10.1016/J.IJPHARM.2017.07.035>.

- [30] S. Shanmugam, Granulation techniques and technologies: recent progresses, *BioImpacts*. 5 (2017) 55–63. <http://www.ncbi.nlm.nih.gov/pubmed/25901297>.
- [31] M.S. Kadiri, Compression de poudres pharmaceutiques et interaction avec l’outillage. Analyse expérimentale et modélisation numérique, PhD Thesis, INPT, 2004. <http://ethesis.inp-toulouse.fr/archive/00000096/>.
- [32] P. Zámotný, Principles of tablet compression, Prague, 2017. <https://fr.scribd.com/document/406418372/Petr-Zamostny-tablet-compression-doc>.
- [33] J. Saddik, Investigating and Predicting Tablet Sticking Using Powder Rheology, Thermal Analysis. and Molecular Simulation, PhD Thesis, Long Island University, 2020. <https://www.proquest.com/openview/cc10316e91486350f2ffce57d11922e1/1?pq-origsite=gscholar&cbl=18750&diss=y>.
- [34] S.G. Begum, A.S. Bai, G. Kalpana, P. Mounika, J.A. Chandini, Review on tablet manufacturing machines and tablet manufacturing defects, *Indian Research Journal of Pharmacy and Science*. 5 (2018) 1479–1490. <https://doi.org/10.21276/IRJPS.2018.5.2.11>.
- [35] T.S. McDermott, J. Farrenkopf, A. Hlinak, J.P. Neilly, D. Sauer, A material sparing method for quantitatively measuring tablet sticking, *Powder Technol.* 212 (2011) 240–252.
- [36] F. Waimer, M. Krumme, P. Danz, U. Tenter, P.C. Schmidt, A Novel Method for the Detection of Sticking of Tablets, *Pharm Dev Technol.* 4 (1999) 359–367.
- [37] I. Saniocki, A. Sakmann, C.S. Leopold, How suitable is the measurement of take-off forces for detection of sticking during direct compression of various ibuprofen tablet formulations?, *Pharm Dev Technol.* 18 (2013) 257–265. <https://doi.org/10.3109/10837450.2012.712538>.
- [38] F.E. Sendall, J.N. Staniforth, A study of powder adhesion to metal surfaces during compression of effervescent pharmaceutical tablets, *J Pharm Pharmacol.* 38 (1986) 489–93. <https://doi.org/10.1111/j.2042-7158.1986.tb04620.x>.
- [39] M.P. Mullarney, B.C. MacDonald, A. Hutchins, Assessing Tablet-Sticking Propensity, *Pharm Technol.* 36 (2012) 57–62. <https://www.pharmtech.com/view/assessing-tablet-sticking-propensity>.
- [40] Z. Wang, U. v. Shah, D. Olusanmi, A.S. Narang, M.A. Hussain, J.F. Gamble, M.J. Tobyn, J.Y.Y. Heng, Measuring the sticking of mefenamic acid powders on stainless

- steel surface, *Int J Pharm.* 496 (2015) 407–413. <https://doi.org/10.1016/j.ijpharm.2015.09.067>.
- [41] H. Tsosie, J. Thomas, J. Strong, A. Zavaliangos, Scanning Electron Microscope Observations of Powder Sticking on Punches during a Limited Number ( $N < 5$ ) of Compactions of Acetylsalicylic Acid, *Pharm Res.* 34 (2017) 2012–2024. <https://doi.org/10.1007/s11095-017-2186-3>.
- [42] C. Al-Karawi, T. Kaiser, C.S. Leopold, A novel technique for the visualization of tablet punch surfaces: Characterization of surface modification, wear and sticking, *Int J Pharm.* 530 (2017) 440–454. <https://doi.org/10.1016/j.ijpharm.2017.08.006>.
- [43] J. Thomas, A. Zavaliangos, An In-line, High Sensitivity, Non-contact Sensor for the Detection of Initiation of Sticking, *J Pharm Innov.* 15 (2020) 66–72. <https://doi.org/10.1007/s12247-018-9367-4>.
- [44] K. Chaturvedi, H.S. Shah, K.R. Morris, R.H. Dave, Modeling of Adhesion in Tablet Compression at the Molecular Level Using Thermal Analysis and Molecular Simulations, *Mol Pharm.* 19 (2022) 26–34. <https://doi.org/10.1021/acs.molpharmaceut.1c00527>.
- [45] H. Chen, S. Paul, H. Xu, K. Wang, M.K. Mahanthappa, C.C. Sun, Reduction of Punch-Sticking Propensity of Celecoxib by Spherical Crystallization via Polymer Assisted Quasi-Emulsion Solvent Diffusion, *Mol Pharm.* 17 (2020) 1387–1396. [https://doi.org/10.1021/ACS.MOLPHARMACEUT.0C00086/SUPPL\\_FILE/MP0C00086\\_SI\\_001.PDF](https://doi.org/10.1021/ACS.MOLPHARMACEUT.0C00086/SUPPL_FILE/MP0C00086_SI_001.PDF).
- [46] S. Paul, L.J. Taylor, B. Murphy, J. Krzyzaniak, N. Dawson, M.P. Mullarney, P. Meenan, C.C. Sun, Mechanism and Kinetics of Punch Sticking of Pharmaceuticals, *J Pharm Sci.* 106 (2017) 151–158. <https://doi.org/10.1016/j.xphs.2016.07.015>.
- [47] J.P. Neilly, A.D. Vogt, W. Dziki, Characterization of Sticking Residue on Tablet Punch Faces by Scanning Electron Microscopy and X-Ray Mapping, *Microscopy and Microanalysis.* 15 (2009) 18–19. <https://doi.org/10.1017/S1431927609097165>.
- [48] D. Hooper, F.C. Clarke, R. Docherty, J.C. Mitchell, M.J. Snowden, Effects of crystal habit on the sticking propensity of ibuprofen—A case study, *Int J Pharm.* 531 (2017) 266–275. <https://doi.org/10.1016/j.ijpharm.2017.08.091>.
- [49] C. Wang, S. Paul, D.J. Sun, S.O. Nilsson Lill, C.C. Sun, Mitigating Punch Sticking Propensity of Celecoxib by Cocrystallization: An Integrated Computational and Experimental Approach, *Cryst Growth Des.* 20 (2020) 4217–4223. <https://doi.org/10.1021/acs.cgd.0c00492>.

- [50] S. Paul, L.J. Taylor, B. Murphy, J.F. Krzyzaniak, N. Dawson, M.P. Mullarney, P. Meenan, C.C. Sun, Toward a Molecular Understanding of the Impact of Crystal Size and Shape on Punch Sticking, *Mol Pharm.* 17 (2020) 1148–1158. <https://doi.org/10.1021/acs.molpharmaceut.9b01185>.
- [51] K. Toyoshima, M. Yasumura, N. Ohnishi, Y. Ueda, Quantitative evaluation of tablet sticking by surface roughness measurement, *Int J Pharm.* 46 (1988) 211–215.
- [52] M.U. Ghorji, E. Šupuk, B.R. Conway, Tribo-electric charging and adhesion of cellulose ethers and their mixtures with flurbiprofen, *Eur J Pharm Sci.* 65 (2014) 1–8. <https://doi.org/10.1016/j.ejps.2014.08.010>.
- [53] E. Šupuk, C. Seiler, M. Ghadiri, Analysis of a Simple Test Device for Tribo-Electric Charging of Bulk Powders, *Part Part Syst Char.* 26 (2009) 7–16. <https://doi.org/10.1002/ppsc.200800015>.
- [54] E. Šupuk, A. Zarrebini, J.P. Reddy, H. Hughes, M.M. Leane, M.J. Tobyn, P. Timmins, M. Ghadiri, Tribo-electrification of active pharmaceutical ingredients and excipients, *Powder Technol.* 217 (2012) 427–434. <https://doi.org/10.1016/j.powtec.2011.10.059>.
- [55] L. Samiei, K. Kelly, L. Taylor, B. Forbes, E. Collins, M. Rowland, The influence of electrostatic properties on the punch sticking propensity of pharmaceutical blends, *Powder Technol.* 305 (2017) 509–517. <https://doi.org/10.1016/j.powtec.2016.10.018>.
- [56] M. Capece, The Role of Particle Surface Area and Adhesion Force in the Sticking Behavior of Pharmaceutical Powders, *J Pharm Sci.* 108 (2019) 3803–3813. <https://doi.org/10.1016/j.xphs.2019.08.019>.
- [57] S. Paul, C.C. Sun, Modulating Sticking Propensity of Pharmaceuticals Through Excipient Selection in a Direct Compression Tablet Formulation, *Pharm Res.* 35 (2018) 113. <https://doi.org/10.1007/s11095-018-2396-3>.
- [58] S. Nakamura, N. Otsuka, Y. Yoshino, T. Sakamoto, H. Yuasa, Predicting the Occurrence of Sticking during Tablet Production by Shear Testing of a Pharmaceutical Powder, *Chem Pharm Bull (Tokyo)*. 64 (2016) 512–516. <https://doi.org/10.1248/CPB.C15-00992>.
- [59] G. Mollereau, V. Mazel, V. Busignies, P. Tchoreloff, F. Mouveaux, P. Rivière, Image Analysis Quantification of Sticking and Picking Events of Pharmaceutical Powders Compressed on a Rotary Tablet Press Simulator, *Pharm Res.* 30 (2013) 2303–2314. <https://doi.org/10.1007/s11095-013-1074-8>.
- [60] M. Ito, S. Aoki, J. Uchiyama, K. Yamato, Development of a New Punch Head Shape to Replicate Scale-Up Issues on a Laboratory Tablet Press III: Replicating Sticking

- Phenomenon Using the SAS Punch and Evaluation by Checking the Tablet Surface Using 3-D Laser Scanning Microscope, *J Pharm Sci.* 107 (2018) 2144–2151. <https://doi.org/10.1016/j.xphs.2018.04.012>.
- [61] N.K. Bejugam, S.K. Mutyam, G.N. Shankar, Tablet formulation of an active pharmaceutical ingredient with a sticking and filming problem: direct compression and dry granulation evaluations, *Drug Dev Ind Pharm.* 41 (2013) 333–341.
- [62] K. Kakimi, T. Niwa, K. Danjo, Influence of Compression Pressure and Velocity on Tablet Sticking, *Chem Pharm Bull (Tokyo).* 58 (2010) 1565–1568.
- [63] F. Waimer, M. Krumme, P. Danz, U. Tenter, P.C. Schmidt, The Influence of Engravings on the Sticking of Tablets. Investigations with an Instrumented Upper Punch, *Pharm Dev Technol.* 4 (1999) 369–375. <https://doi.org/10.1081/PDT-100101372>.
- [64] S. Paul, K. Wang, L.J. Taylor, B. Murphy, J. Krzyzaniak, N. Dawson, M.P. Mullarney, P. Meenan, C.C. Sun, Dependence of Punch Sticking on Compaction Pressure—Roles of Particle Deformability and Tablet Tensile Strength, *J Pharm Sci.* 106 (2017) 2060–2067. <https://doi.org/10.1016/j.xphs.2017.04.059>.
- [65] S. Naito, K. Nakamichi, Studies on Techniques of Manufacturing Pharmacy. I. Prediction of Tableting Troubles such as Capping and Sticking. (I), *Chem Pharm Bull (Tokyo).* 17 (1969) 2507–2514. <https://doi.org/10.1248/cpb.17.2507>.
- [66] A. Mitrove, L.L. Augsburger, Adhesion of Tablets in a Rotary Tablet Press I. Instrumentation and Preliminary Study of Variables Affecting Adhesion, *Drug Dev Ind Pharm.* 6 (2008) 331–377. <https://doi.org/10.3109/03639048009068709>.
- [67] T. Uchimoto, Y. Iwao, T. Yamamoto, K. Sawaguchi, T. Moriuchi, S. Noguchi, S. Itai, Newly developed surface modification punches treated with alloying techniques reduce sticking during the manufacture of ibuprofen tablets, *Int J Pharm.* 441 (2013) 128–134.
- [68] S. Schumann, G.D. Searle, The Effects of Chromium Nitride ION Bombardment Treatment of Tablet Tooling on Tablet Adherence, *Drug Dev Ind Pharm.* 18 (1992) 1037–1061. <https://doi.org/10.3109/03639049209069314>.
- [69] T. Uemura, K. Sawaguchi, T. Taketani, Y. Koike, T. Kobayashi, H. Ichikawa, Y. Fukumori, Y. Mizushima, Investigation of essential characteristics of metal-hardening and matt-surfaced punches for reducing tablet sticking, *J Jpn Soc Pharm Mach Eng.* 16 (2007) 14–20.

- [70] R.E. Gordon, C.L. VanKoeveering, D.J. Reits, Utilization of differential scanning calorimetry in the compatibility screening of ibuprofen with the stearate lubricants and construction of phase diagrams, *Int J Pharm.* 21 (1984) 99–105.
- [71] S. Abdel-Hamid, G. Betz, A novel tool for the prediction of tablet sticking during high speed compaction, *Pharm Dev Technol.* 17 (2012) 747–754. <https://doi.org/10.3109/10837450.2011.580761>.
- [72] S. Aoki, K. Danjo, [Effect of tableting conditions on the sticking of tablet using ibuprofen]., *Yakushigaku Zasshi.* 118 (1998) 511–8. Japanese.
- [73] Y. Shimada, Y. Yonezawa, H. Sunada, Measurement and Evaluation of the Adhesive Force between Particles by the Direct Separation Method, *J Pharm Sci.* 92 (2003) 560–568.
- [74] K. Danjo, S. Kojima, C.Y. Chen, H. Sunada, A. Otsuka, Effect of Water Content on Sticking during Compression., *Chem Pharm Bull (Tokyo).* 45 (1997) 706–709. <https://doi.org/10.1248/cpb.45.706>.
- [75] J.C. Hooton, C.S. German, S. Allen, M.C. Davies, C.J. Roberts, S.J.B. Tendler, P.M. Williams, Characterization of particle-interactions by atomic force microscopy: effect of contact area, *Pharm Res.* 20 (2003) 508–514. <https://doi.org/10.1023/A:1022684911383>.
- [76] J.C. Hooton, C.S. German, S. Allen, M.C. Davies, C.J. Roberts, S.J.B. Tendler, P.M. Williams, An atomic force microscopy study of the effect of nanoscale contact geometry and surface chemistry on the adhesion of pharmaceutical particles, *Pharm Res.* 21 (2004) 953–961. <https://doi.org/10.1023/B:PHAM.0000029283.47643.9C>.
- [77] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, *Multi- and Megavariate Data Analysis Principles and Applications*, 1st ed., Umetrics, Umea, Sweden, 2001.
- [78] C.F. Dormann, J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J.R.G. Marquéz, B. Gruber, B. Lafourcade, P.J. Leitão, T. Münkemüller, C. McClean, P.E. Osborne, B. Reineking, B. Schröder, A.K. Skidmore, D. Zurell, S. Lautenbach, Collinearity: A review of methods to deal with it and a simulation study evaluating their performance, *Ecography.* 36 (2013) 27–46.
- [79] T. Næs, B.-H. Mevik, Understanding the collinearity problem in regression and discriminant analysis, *J Chemom.* 15 (2001) 413–426.
- [80] C.J. Huberty, Stephen. Olejnik, *Applied MANOVA and discriminant analysis.*, 2nd ed., Wiley Series in Probability and Statistics, Hoboken, New Jersey, États-Unis, 2006.

- [81] S. Paul, L.J. Taylor, B. Murphy, J. Krzyzaniak, N. Dawson, M.P. Mullarney, P. Meenan, C.C. Sun, Powder properties and compaction parameters that influence punch sticking propensity of pharmaceuticals, *Int J Pharm.* 521 (2017) 374–383. <https://doi.org/10.1016/j.ijpharm.2017.02.053>.
- [82] R.E. Shaffer, Book review: Multi- and Megavariate Data Analysis. Principles and Applications, I. Eriksson, E. Johansson, N. Kettaneh-Wold and S. Wold, Umetrics Academy, Umeå, 2001, ISBN 91-973730-1-X, 533pp., *J Chemom.* 16 (2002) 261–262. <https://doi.org/10.1002/cem.713>.
- [83] R. Johnson, D. Wichern, Applied multivariate statistical analysis, 6th ed., Prentice Hall, New Jersey, 2002.
- [84] R. Rakotomalala, *Pratique de l'Analyse Discriminante Linéaire*, Lyon, 2020. [http://eric.univ-lyon2.fr/~ricco/cours/cours/Pratique\\_Analyse\\_Discriminante\\_Lineaire.pdf](http://eric.univ-lyon2.fr/~ricco/cours/cours/Pratique_Analyse_Discriminante_Lineaire.pdf).
- [85] J. Lammertyn, E.A. Veraverbeke, J. Irudayaraj, zNose™ technology for the classification of honey based on rapid aroma profiling, *Sens Actuators B Chem.* 98 (2004) 54–62. <https://doi.org/10.1016/j.snb.2003.09.012>.
- [86] J. Xing, S. Landahl, J. Lammertyn, E. Vrindts, J. de Baerdemaeker, Effects of bruise type on discrimination of bruised and non-bruised ‘Golden Delicious’ apples by VIS/NIR spectroscopy, *Postharvest Biol Technol.* 30 (2003) 249–258. [https://doi.org/10.1016/S0925-5214\(03\)00117-0](https://doi.org/10.1016/S0925-5214(03)00117-0).
- [87] L. Nørgaard, R. Bro, F. Westad, S.B. Engelsen, A modification of canonical variates analysis to handle highly collinear multivariate data, *J Chemom.* 20 (2006) 425–435. <https://doi.org/10.1002/cem.1017>.
- [88] P.W. Yendle, H.J.H. MacFie, Discriminant principal components analysis, *J Chemom.* 3 (1989) 589–600. <https://doi.org/10.1002/cem.1180030407>.
- [89] S. Rännar, P. Geladi, F. Lindgren, S. Wold, A PLS kernel algorithm for data sets with many variables and few objects. Part II: Cross-validation, missing data and examples, *J Chemom.* 9 (1995) 459–470. <https://doi.org/10.1002/cem.1180090604>.
- [90] T. Mehmood, S. Sæbø, K.H. Liland, Comparison of variable selection methods in partial least squares regression, *J Chemom.* 34 (2020) e3226.
- [91] M. Dembélé, S. Hudon, J.-S. Simard, N. Abatzoglou, R. Gosselin, A multivariate data analysis approach to tablet sticking on an industrial scale: A qualitative case study of an ibuprofen-based formulation, *Pharm Dev Technol.* (2022) [Accepted]. <https://doi.org/10.1080/10837450.2022.2153866>.

- [92] H. Watanabe, M. Ghadiri, T. Matsuyama, Y.L. Ding, K.G. Pitt, H. Maruyama, S. Matsusaka, H. Masuda, Triboelectrification of pharmaceutical powders by particle impact, *Int J Pharm.* 334 (2007) 149–155.
- [93] L. Zhang, S. Garcia-Munoz, A comparison of different methods to estimate prediction uncertainty using Partial Least Squares (PLS): A practitioner’s perspective, *Chemom Intell Lab Sys.* 97 (2009) 152–158. <https://doi.org/10.1016/j.chemolab.2009.03.007>.
- [94] N. Krämer, M. Sugiyama, The Degrees of Freedom of Partial Least Squares Regression, *J Am Stat Assoc.* 106 (2011) 697–705. <https://doi.org/10.1198/jasa.2011.tm10107>.
- [95] D.M. Simmons, D.S. Gierer, A material sparing test to predict punch sticking during formulation development, *Drug Dev Ind Pharm.* 38 (2012) 1054–60. <https://doi.org/10.3109/03639045.2011.637933>.
- [96] T. Kamiya, H. Kondo, H. Hiroma, K. Yamashita, T. Hakomori, K. Sako, Y. Iwao, S. Noguchi, S. Itai, Impact of process parameters on Mg–St content and tablet surface wettability in the external lubrication method for a rotary tablet press, *Advanced Powder Technol.* 27 (2016) 193–198. <https://doi.org/10.1016/J.APT.2015.11.012>.
- [97] V. Mazel, V. Busignies, H. Diarra, I. Reiche, P. Tchoreloff, The surface layer of pharmaceutical compacts: The role of the punch surface and its impact on the mechanical properties of the compacts, *Int J Pharm.* 442 (2013) 42–48. <https://doi.org/10.1016/j.ijpharm.2012.08.005>.
- [98] J. Mosig, P. Kleinebudde, Evaluation of lubrication methods: How to generate a comparable lubrication for dry granules and powder material for tableting processes, *Powder Technol.* 266 (2014) 156–166. <https://doi.org/10.1016/j.powtec.2014.06.022>.
- [99] M.K. Papp, G. Venkatesh, T. Harmon, Specialised techniques for developing ODT dosage forms, *Biology (Basel)*. (2010).
- [100] T. Yamamura, T. Ohta, T. Taira, Y. Ogawa, Y. Sakai, K. Moribe, K. Yamamoto, Effects of automated external lubrication on tablet properties and the stability of eprazinone hydrochloride, *Int J Pharm.* 370 (2009) 1–7. <https://doi.org/10.1016/j.ijpharm.2008.11.007>.
- [101] H. Kondo, H. Toyota, T. Kamiya, K. Yamashita, T. Hakomori, J. Imoto, S.-I. ichiro Kimura, Y. Iwao, S. Itai, Effect of the External Lubrication Method for a Rotary Tablet Press on the Adhesion of the Film Coating Layer, *Chem Pharm Bull (Tokyo)*, 2017. <https://doi.org/10.1248/CPB.C17-00376>.
- [102] M. Otsuka, M. Sato, Y. Matsuda, Comparative evaluation of tableting compression behaviors by methods of internal and external lubricant addition: Inhibition of



- enzymatic activity of trypsin preparation by using external lubricant addition during the tableting compression process, *AAPS Pharm Sci.* 3 (2001) 32–42. <https://doi.org/10.1208/ps030320>.
- [103] T. Jahn, K.-J.J. Steffens, Press Chamber Coating as External Lubrication for High Speed Rotary Presses: Lubricant Spray Rate Optimization, *Drug Dev Ind Pharm.* 31 (2005) 951–957. <https://doi.org/10.1080/03639040500306161>.
- [104] M. Jabir, Comparaison de méthodes d'imputation des données manquantes appliquées à la base nationale sur les collisions, MSc Thesis, HEC Montréal, 2018.
- [105] R.J.A. Little, D.B. Rubin, *Statistical analysis with missing data*, 3rd editio, Wiley, Hoboken, 2019. <https://doi.org/10.1002/9781119482260>.
- [106] M. Takahashi, *Statistical Inference in Missing Data by MCMC and Non-MCMC Multiple Imputation Algorithms: Assessing the Effects of Between-Imputation Iterations*, *Data Sci J.* 16 (2017) 37. <https://doi.org/10.5334/dsj-2017-037>.
- [107] M. Glasson-Cicognani, A. Berchtold, Imputation des données manquantes : comparaison de différentes approches, in : INRIA (Ed.), 42èmes Journées de Statistique, HAL archives ouvertes, Marseille, 2010 : p. 6. <https://hal.inria.fr/inria-00494698>.
- [108] J.L. Schafer, *Analysis of Incomplete Multivariate Data*, 1st ed., Chapman and Hall/CRC, New York, 1997. <https://www.taylorfrancis.com/books/9780367803025>.
- [109] N.K.A. Aurélien, K.K. Steve, Gestion des Données manquantes dans les bases de donnees en sciences sociales : algorithme NIPALS ou imputation Multiple ?, *Eur Sci J.* 12 (2016) 390. <https://doi.org/10.19044/esj.2016.v12n35p390>.
- [110] C.K. Enders, *Applied missing data analysis*, 1st ed., Guilford Press, New York, 2010.
- [111] S. Paquin, Comparaison de quatre méthodes pour le traitement des données manquantes au sein d'un modèle multiniveau paramétrique visant l'estimation de l'effet d'une intervention, MSc Thesis, Université de Montréal, 2010. [https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/4599/Paquin\\_Stephane\\_2010\\_memoire.pdf?sequence=2&isAllowed=y](https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/4599/Paquin_Stephane_2010_memoire.pdf?sequence=2&isAllowed=y).
- [112] J.L. Schafer, J.W. Graham, Missing data : Our view of the state of the art., *Psychol Methods.* 7 (2002) 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>.

# ANNEXES

## Annexe 1 : la lubrification externe

### 1. Généralités sur les lubrifiants :

Les lubrifiants sont des substances solides poudreuses ou liquides qui sont généralement ajoutées en petites quantités dans les formulations pharmaceutiques pour améliorer certaines de leurs caractéristiques physiques, notamment l'écoulement et la friction. Les lubrifiants ont trois rôles principaux qui motivent leur application dans les formulations pharmaceutiques [96] :

- Un rôle de lubrifiant où ils permettent de réduire la friction aux parois de la matrice lors de l'éjection et sur la surface du poinçon du bas lors du dégagement du comprimé. Cela permet d'éviter les forces d'éjection et détachement excessives, préservant ainsi les outillages de compression contre l'usure et la cassure. C'est la raison principale de l'emploi de lubrifiants dans les formulations pharmaceutiques ;
- Un rôle d'antiadhérent qui permet de prévenir le collage aux surfaces des matrices et poinçons ;
- Un rôle de glissant consistant à améliorer les propriétés d'écoulement de la formulation.

La quasi-totalité des produits requiert un minimum de lubrification pour réduire les frictions aux parois des outillages afin que ceux-ci puissent être compressés et éjectés [98]. Seulement quelques rares ingrédients actifs et excipients font exception à la règle comme l'acide acétyl salicylique, l'amidon et la cellulose microcristalline. On peut distinguer deux types de lubrifiants : les lubrifiants hydrophiles qui sont essentiellement liquides et les lubrifiants hydrophobes qui sont des solides divisés. Les lubrifiants hydrophiles sont inefficaces à l'anti-adhésion et n'ont généralement pas de propriété de glissant. Ce sont des huiles qui n'ont également de bonnes propriétés lubrifiantes que s'ils sont utilisés en grande quantité. À l'opposé, les lubrifiants hydrophobes ont de bonnes propriétés et sont efficaces même à faible concentration. Ce sont les plus employés dans l'industrie pharmaceutique avec des proportions massiques ne dépassant pas généralement 3 %(m/m) de la formulation. L'approche traditionnelle d'utilisation des lubrifiants consiste à mélanger le lubrifiant avec les autres composants de la formulation lors d'une dernière étape de mélange. Cette approche est

appelée la lubrification interne et est sujette à de nombreux problèmes, notamment une faible cohésion des comprimés. C'est pour cette raison qu'une autre approche de lubrification appelée lubrification externe a été introduite. Elle consiste à délivrer le lubrifiant à l'interface poinçon-poudre lors de la compression des poudres.

## **2. Description de l'approche de lubrification externe :**

Pour bien cerner les tenants et les aboutissants de la lubrification externe, il est essentiel de le mettre dans le contexte d'une suite de la lubrification interne. En effet, le besoin particulier ayant conduit à l'invention de la technique de la lubrification externe provient des lacunes de la lubrification interne. En adoptant cette démarche, on met rapidement en exergue le principe, le mécanisme ainsi que les avantages et limites pouvant être associés à la lubrification externe. C'est pourquoi, dans cette section, nous allons prendre le temps de bien décrire d'abord les lacunes de la lubrification interne. À la suite de quoi, nous introduirons et présenterons la lubrification externe.

### **2.1 La lubrification interne : entre nécessité et désavantages**

La première lacune de la lubrification interne porte sur son effet délétère sur les propriétés du produit final : la réduction de la résistance mécanique, le rallongement du temps de dissolution et de désintégration. En réalité, ces effets délétères sont directement reliés au mécanisme d'action du lubrifiant, sa quantité et sa façon d'incorporation. Prenons l'exemple avec le lubrifiant le plus couramment utilisé dans les formulations pharmaceutiques : le MgSt. Ce dernier est une substance poudreuse hydrophobe qui possède une structure lamellaire. Ces lamelles sont empilées ensemble comme un jeu de cartes. Au cours de l'incorporation à la formulation, les lamelles se séparent sous le travail du mélange et revêtissent les particules de poudre adjacentes [96]. Il va y avoir formation d'un film de lubrifiant tout autour des particules constitutives de la formulation (ingrédients actifs et excipients). C'est ce film, qui grâce au faible coefficient de friction du lubrifiant, permet de faciliter la compression et l'éjection, mais aussi prévient l'adhésion aux surfaces des outillages de compression. Le pouvoir de glissant est aussi assuré par ce mécanisme, car le film de lubrifiant va réduire également les frictions interparticulaires au sein de la poudre et améliorer les propriétés d'écoulement de celle-ci. Le problème principal avec un tel mécanisme est que la quantité mais aussi la distribution et la qualité du mélange peuvent tous impacter la performance du lubrifiant [96,98]. En fait, ce sont des paramètres critiques pour une bonne lubrification.

Lorsque la concentration de lubrifiant, la distribution (l'intensité du mélange) et le temps de mélange sont insuffisants (trop faibles), les propriétés lubrifiantes, antiadhérentes et glissantes du lubrifiant sont compromises. On parle de sous-lubrification. Le film de lubrifiant ne revêt pas adéquatement et suffisamment les particules. De nombreuses études rapportent sous ces conditions des événements de collage, laminage, décapage, mais surtout des difficultés à comprimer la poudre et l'éjecter de la matrice. Dans des cas limites, de grandes forces de compression, d'éjection ou de détachement nécessaire pour dépasser la friction aux parois de la matrice et sur la surface du poinçon inférieur peuvent entraîner la cassure prématurée des poinçons, mais surtout présenter des risques de santé et sécurité.

À l'inverse, lorsque la concentration de lubrifiant, l'intensité ou la durée de mélange sont trop importantes, les propriétés lubrifiantes, antiadhérentes et glissantes du lubrifiant sont grandement améliorées, mais les attributs de qualité du produit final sont affectés. En effet, les poudres s'écoulent bien et l'éjection est facilitée. Aucun collage ni autre problème relié n'est également rapporté. En fait, l'augmentation de la quantité de lubrifiant est même l'une des techniques couramment employées pour prévenir le collage. Par ailleurs, sous les mêmes conditions, la résistance mécanique donc la dureté du produit fini est compromise et les temps de désintégration et de dissolution sont rallongés [96,98]. Lorsque la quantité de lubrifiant, l'intensité et le temps de mélange sont importants, le film de lubrifiant recouvre la quasi-totalité des particules. Se faisant, peu de liaisons interparticulaires sont induites par la compression et la cohésion du comprimé résultant est affaiblie. D'un autre côté, la mise sous forme de comprimés de tels mélanges nécessite de grandes forces de compression qui peuvent induire l'usure prématurée des poinçons. Aussi, le caractère hydrophobe du lubrifiant recouvrant les particules rallonge le temps de dissolution et de désintégration du comprimé. Les particules à déformation plastique et élastique sont les plus affectées par ces effets délétères de la lubrification, car celles-ci restent enveloppées au cours de la compression. Les granules produits par la granulation humide ainsi que les particules à déformation fragmentaire sont moins affectés, car elles forment de nouvelles surfaces libres de films lubrifiants lors de la compression. Toutefois, les particules ne se déforment pas selon un seul mécanisme et ces effets délétères sont rapportés pour la majorité des modèles de formulation ayant fait l'objet d'étude.

Alors, pour bénéficier des propriétés des lubrifiants sans les désavantages causés sur les attributs qualité du comprimé résultant, il est donc crucial de trouver la quantité, l'intensité et le temps de mélange optimaux induisant assez de films lubrifiants pour d'une part réduire les

frictions et le collage aux parois des outillages et d'autre part juste assez pour ne pas trop affaiblir la cohésion interne et les propriétés de dissolution du comprimé. C'est ce principe fondamental d'utiliser le minimum nécessaire qui est difficile à adresser dans l'approche de la lubrification interne, car la distribution du lubrifiant est sensible aux différentes opérations unitaires de fabrication, les équipements de production, les échelles de production ainsi que la nature des formulations. À ces problèmes sur les attributs qualité du comprimé difficiles à adresser s'ajoutent les interactions diverses que peut avoir le lubrifiant avec les autres agents de la formulation. Plusieurs travaux ont rapporté par exemple la formation de mélanges eutectiques entre le MgSt et l'IBU [11,12]. Ces mélanges abaissent le point de fusion de l'ibuprofène et le rendent plus collant.

En résumé, la lubrification est essentielle dans la production d'un comprimé pour réduire la friction et l'adhésion aux parois des outillages durant la compression. Elle achève ces missions en recouvrant les particules d'un film lubrifiant hydrophobe. L'approche traditionnelle de la lubrification interne permet de réaliser ces missions, mais nécessite de trouver des optimums opérationnels, car le film de lubrifiant hydrophobe peut limiter la formation de liens entre les particules pendant la compression, mais aussi rendre le comprimé insoluble.

Lorsque ce problème est pris sous un autre angle, on peut voir que les avantages de l'utilisation des lubrifiants sont liés à leur capacité à réduire la friction et l'adhésion aux interfaces particules-outillages, mais surtout que c'est sa présence importante au sein du milieu particulaire qui induit tous les problèmes ci-dessus détaillés. C'est ce constat qui a amené l'industrie pharmaceutique à se poser la question suivante : pourrions-nous avoir de la lubrification juste à l'interface entre les parois des outillages et des particules, sans en mettre à l'intérieur de la formulation ? C'est à cette question que répond la lubrification externe. En effet, le principe de la lubrification externe consiste à délivrer la juste quantité nécessaire à l'interface entre les particules et les parois des outillages. Cela permet ainsi de réduire la friction et l'adhésion au niveau des parois sans avoir les désavantages de la présence de trop de lubrifiant dans la formulation (baisse de la cohésion interne, insolubilité et interactions avec les ingrédients actifs). Nous allons maintenant décrire cette technique.

## **2.2 La lubrification externe : Délivrer du lubrifiant là où nécessaire ?**

Historiquement, c'est dans les années 1990 que l'idée de la lubrification externe a commencé à être mise en pratique. Les premières applications étaient manuelles, c'est-à-dire que les

poinçons et les matrices étaient enduits de poudre de lubrifiant avant la compression. Cette technique est toujours d'usage comme astuce des opérateurs lorsqu'ils font face à du collage. La lubrification externe de façon manuelle n'était utilisée que dans les laboratoires de recherche pour faire les premières investigations. Des résultats encourageants ont ensuite conduit au développement de nombreux systèmes de lubrification et plusieurs brevets ont été déposés. Il existe aujourd'hui d'innombrables systèmes de lubrification externe qui sont proposés par plusieurs entreprises. Nous pouvons citer par exemple les systèmes PKB-1, 2 et 3 de la compagnie Fette ou encore les systèmes ELS P1 et P2 du japonais Yokoham et ExLub de Kyowa Hakko Kogyo [99]. Cependant, le principe de fonctionnement de tous ces systèmes reste identique. Il consiste à suspendre dans de l'air comprimé une substance lubrifiante, généralement de la poudre de MgSt. Celle-ci est ensuite acheminée et pulvérisée de façon continue sur les surfaces des outillages via une buse dédiée qui est montée juste après la barre de dégagement du comprimé sur les presses rotatives (voir figure 20 ci-dessous). La pulvérisation a lieu entre la fin de l'éjection et le remplissage de la matrice. Les excès de lubrifiant sont aussi récoltés continuellement à l'aide d'un aspirateur collecteur également monté à côté de la buse de pulvérisation.

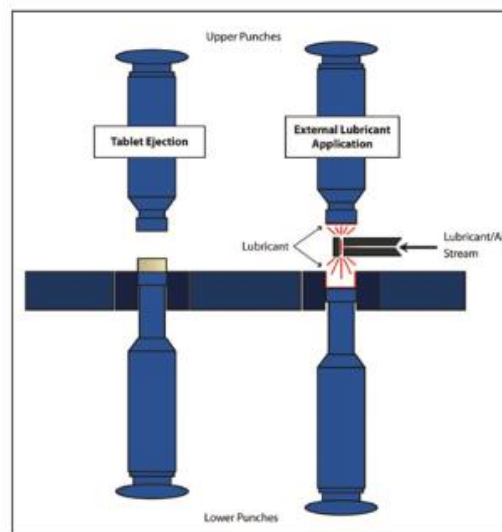


Figure 26 : Représentation schématique d'un système de lubrification externe [99].

Pour faire de la lubrification externe, il faut tout un système complet monté autour de la presse. Quatre parties essentielles composent tout dispositif de lubrification externe : un système d'alimentation de lubrifiant, un système de suspension et de convoiement avec de l'air comprimé, des buses de pulvérisation montées sur la presse et un système d'aspiration sous vide de l'excès de lubrifiant. Le principe est unique à tous les systèmes existants aujourd'hui.

En fait, les différents systèmes de lubrification se démarquent plus dans la façon d'adresser et de suivre un débit de pulvérisation précis, aspect crucial dans la validation du montage dans un environnement de production BPF (Bonnes Pratiques de Fabrication).

### **3. Les avantages de la lubrification externe :**

Le principe de la lubrification externe est de délivrer une quantité minimale de lubrifiant à l'interface entre les particules de poudre et les outillages de compression. Cela permet de laisser les surfaces de poudre libres de lubrifiant. Se faisant, la lubrification externe préserve ses fonctions principales de réduction de la friction et d'adhésion, mais surtout permet d'améliorer les attributs qualité des comprimés résultant par rapport à l'approche traditionnelle de lubrification interne.

Tout d'abord, en termes de collage, toutes les études réalisées montrent une performance supérieure de la lubrification externe par rapport à la lubrification interne quant à la réduction du collage. Yamamura et collab. (2009), ont étudié les avantages de la lubrification externe dans la compression de l'hydrochloride d'éprazinone [100]. Ils ont notamment évalué la propension au collage en utilisant une lubrification interne et externe ainsi que différentes proportions de lubrifiant lors de séances de compression. Leur résultat a montré que pour des concentrations de lubrifiant inférieur à 0,55 %, du collage est observé pour les comprimés fabriqués avec la lubrification interne, tandis que pour une lubrification externe, il faut descendre à des concentrations de moins de 0,08 % pour observer du collage lors de la compression de l'hydrochloride d'éprazinone [100]. Cela correspond à 1/13 de la quantité de lubrifiant requise dans le cas de la lubrification interne.

On peut se poser la question de savoir pourquoi avec le même lubrifiant, nous avons des écarts aussi importants entre la lubrification interne et externe. Selon le mécanisme d'action des lubrifiants, c'est un film de lubrifiant qui revêt les particules et limiterait leur adhésion aux surfaces des outillages. En fait, ce mécanisme reste vrai pour la lubrification externe également où une fine couche uniforme de lubrifiant revêt les particules de surface des poudres contrairement à la lubrification interne où la distribution en surface est non uniforme et ponctuée d'agglomérats de lubrifiant [96,101]. Un autre contributeur très important à la performance anti-collage de la lubrification externe est l'amélioration de la cohésion interne du comprimé. En effet, différentes sources concordent pour montrer que les comprimés produits avec une lubrification externe sont 30 % à 50 % plus durs [100,102]. Cela est certainement dû au fait que le lubrifiant n'est pas distribué à l'intérieur du mélange de poudre.

Les surfaces internes ainsi libres de films lubrifiants, les particules établissent des liens plus forts et assurent une forte cohésion. À l'inverse, le mélange de lubrifiant à l'intérieur du comprimé dans le cas de la lubrification interne affaiblit les liens de cohésion et les particules adhèrent plus facilement à la surface des poinçons. Ces explications sont corroborées également par la théorie la plus avancée du modèle conceptuel de collage présenté plus haut (Figure 4). Pour continuer sur les raisons de la meilleure efficacité de la lubrification externe à la réduction du collage par rapport à la lubrification interne, nous pouvons aussi noter la tendance du MgSt de former des mélanges eutectiques avec certains ingrédients actifs, comme l'IBU contribuant à abaisser leur point de fusion et exacerber le collage [11,12,20,68]. Lors de la lubrification externe, cet aspect est également écarté.

L'amélioration de la performance vis-à-vis du collage n'est pas le seul avantage de la lubrification externe. S'agissant de son second rôle en tant que lubrifiant, la lubrification externe permet également de réduire la friction aux interfaces des poinçons et matrices. Elle contribue ainsi à la prévention de l'usure ou la cassure prématurée des outillages. Aussi, le fait de ne pas nécessiter de grandes forces de compression pour former un compact solide permet à la lubrification externe d'offrir une plus grande longévité aux outillages.

Au-delà des points relatifs à ses rôles, la lubrification externe se démarque aussi surtout par les meilleurs attributs de qualité du produit fini qu'elle permet d'adresser. En effet, contrairement à la lubrification interne où la capacité à adresser moins de collage et de friction rimait avec la production de comprimés peu soluble et mécaniquement faible, la lubrification externe produit des comprimés avec une meilleure résistance mécanique, mais aussi des temps de dissolution et de désintégration plus courts [99]. Mosig et collab. (2014) ont étudié l'influence de la lubrification sur les propriétés des comprimés fabriqués par lubrification interne et externe [98]. Les auteurs ont montré que la lubrification externe permettait d'obtenir des comprimés plus résistants. Pour des particules comme le MCC et les poudres de cellulose qui ont un comportement de déformation à dominance plastique et élastique, la lubrification externe produisait des comprimés 30 à 40 % plus solides comparés à la lubrification interne. En revanche, pour les particules à tendance de déformation fragmentaire comme le carbonate de magnésium et le lactose, les auteurs reportent un gain de seulement 5 % dans la résistance [98]. En effet, les particules fragmentaires sont moins sensibles à la lubrification interne en raison des nouvelles surfaces libres de films de lubrifiant qu'elles forment lors de la compression. Des conclusions similaires où la lubrification externe permettait un gain de résistance mécanique de 30 à 50 % ont été rapportées par plusieurs autres auteurs [99,102].



Deux autres attributs de qualité du produit fini que la lubrification externe améliore sont les temps de dissolution et de désintégration. Otsuka et collab. (2001) ont étudié l'impact de la technique de lubrification sur la libération de la trypsine et ont rapporté que la lubrification externe permettait une libération immédiate de la trypsine (100 % libéré en 7 minutes) comparée à la lubrification interne où seulement 20 % de la trypsine était libérée au bout de 20 minutes [102]. Ces effets sont attribués à la faible présence d'éléments hydrophobiques favorisant une meilleure pénétration de l'eau. En effet, d'autres auteurs s'étant intéressés à la question ont montré que la quantité de lubrifiant par comprimé et l'angle de contact en surface étaient plus faibles pour des comprimés fabriqués avec lubrification externe comparée à ceux fabriqués par lubrification interne.

Les avantages de la lubrification externe ne se limitent pas aux seuls attributs de qualité des produits finis, elle améliore également la stabilité chimique des comprimés produits. En effet, dans leur étude, Otsuka et collab. (2001) ont également investigué l'impact de la lubrification sur l'activité de la trypsine et ont rapporté que la lubrification externe permettait à la trypsine de conserver une meilleure activité dans le temps par rapport à une lubrification interne [102]. Cette stabilité enzymatique de la trypsine a été reliée à la faible force de compression requise pour achever la même résistance mécanique que dans le cas d'une lubrification interne. Les comprimés sont exposés à une plus faible pression et chaleur. Une meilleure stabilité d'autres ingrédients actifs incompatibles avec le MgSt a été rapportée et corrélée à la faible quantité de lubrifiant par comprimé adressée par la lubrification externe [100].

Enfin, nous pouvons citer d'autres avantages de la lubrification externe en termes de procédé global comme la réduction du temps de production par l'élimination de l'étape de mélange finale avec le lubrifiant [96], son application à la granulation sous rouleaux compactionnaires où elle permet de prévenir le collage également, mais aussi la propriété de faciliter d'autres opérations unitaires en aval de la compression comme l'enrobage [102]. Kondo et collab. (2017) ont étudié l'effet de la stratégie de lubrification sur l'adhésion du film d'enrobage sur la surface des comprimés [101]. Ils ont montré que lorsque la quantité de lubrifiant externe dispensée sur la surface des comprimés augmentait, aucun changement dans l'adhésion du film d'enrobage n'était observé. Par contre, ce paramètre diminuait dans le cas de la lubrification interne. Les auteurs ont attribué cet effet à une meilleure distribution uniforme du lubrifiant sur la surface des comprimés produits avec la lubrification externe

comparativement à la lubrification interne où des agglomérats sont retrouvés en surface et affecterait l'adhésion par des pelages par endroit [101].

#### **4. Les inconvénients de la lubrification externe :**

Des nombreux avantages que nous avons détaillés, la lubrification peut paraître comme la solution miracle pour résoudre le collage et les différents problèmes occasionnés par le mélange direct de la lubrification dans la poudre. Bien qu'il faille reconnaître ces excellentes performances, plusieurs points limitent encore le développement de la lubrification externe, notamment le coût et la complexité associés à son implémentation, les requis réglementaires qu'elle peut nécessiter, mais surtout le manque de données scientifiques sur certains de ces aspects.

Tout d'abord, en s'intéressant à son pouvoir de lubrifiant, on retrouve dans la littérature des travaux relatant la mauvaise performance de la lubrification externe à réduire la force d'éjection comparée à la lubrification interne. En effet, Otsuka et collab. (2001) ont montré que la lubrification externe contribuait à réduire les frictions aux parois lors de la compression et de l'éjection, mais que son influence était bien moindre que pour la lubrification interne, toutefois plus élevée que dans le cas où aucune lubrification n'était ajoutée [102]. En fait, le mélange préalable du lubrifiant avec la poudre de formulation dans le cas de la lubrification interne permet un délaminage des cristaux de MgSt et donc une meilleure couverture de la surface des particules de poudre. En revanche, la rapidité du processus de compression et de pulvérisation dans la lubrification externe ne permet pas un délaminage important du lubrifiant sur la surface des particules, ce qui diminue son efficacité [102]. Pour certains produits où le problème était plus prononcé, les auteurs de cette étude ont même proposé de combiner les deux approches en incorporant un peu de lubrifiant dans le mélange et réaliser la compression sous lubrification externe. Cependant, aucune étude n'a encore repris la suite de ces conclusions.

Pour continuer sur le manque d'informations disponibles, une autre limite de la lubrification externe est le peu d'études disponibles sur la technique. En effet, il n'existe pas énormément de publications sur le sujet. Cela peut certainement être imputé à la jeunesse du sujet, mais il y'a d'autres facteurs contributeurs également comme le coût des expérimentations. En effet des études qui existent, une moitié concerne des études à portée limitée où des équipes de recherche étudient la lubrification externe en appliquant manuellement le lubrifiant sur les poinçons et les matrices. Il est très difficile dans ces

conditions de réaliser des investigations poussées et représentatives. Pourtant, des systèmes sophistiqués existent et clament leur robustesse, mais ce sont des équipements très chers. L'autre moitié des travaux publiés concernent les publications et études réalisées par les industriels proposant les systèmes de lubrification où le but est plus de vendre les mérites de l'appareil. Les aspects pouvant présenter un challenge sérieux ne sont donc pas systématiquement étudiés et reportés. La cherté des expériences est aussi un frein à l'implémentation de la lubrification externe en production, car l'investissement est trop important pour une implémentation sans résultats primaires encourageants. Les coûts et la complexité de l'implémentation d'un système de lubrification externe sont aussi des freins à son déploiement. Il faut noter que chaque système de lubrification externe est taillé sur mesure pour une presse donnée. En fait, ce sera en fonction des dimensions et des caractéristiques opératoires des presses déjà existantes pour une salle de production donnée que les industriels adapteront le système de lubrification. Cela représente un investissement initial conséquent, notamment en ce qui concerne les modifications à opérer sur les équipements et leur validation santé, sécurité et réglementaire.

Sur le même plan, l'installation d'un système de lubrification externe nécessite également la mise en place de procédure de nettoyage ainsi que la validation de ces nettoyages. Les risques de santé-sécurité dus à la suspension de poudre de lubrifiant dans l'air doivent également être pris en compte. En effet, bien que l'aspiration soit mise en place pour permettre l'évacuation des poudres de lubrifiant excédentaire, de la matière résiduelle peut se suspendre dans l'air. Le MgSt est sans danger pour la santé, mais son inhalation excessive peut causer des irritations. Une évaluation de l'exposition des opérateurs et une maîtrise de la suspension du lubrifiant seront donc nécessaires.

Un autre frein à l'implémentation de la lubrification externe est la problématique de l'espace, car comme on a pu voir plus haut, un système de lubrification externe requiert l'installation d'une unité d'alimentation de lubrifiant, un système de convoyage par air comprimé et un système aspirateur de collecte des résidus de lubrifiant. Ces différents éléments peuvent être assez encombrants. Pourtant, en production pharmaceutique, le mètre carré de salle propre coûte cher à la construction et à l'entretien. Cette charge supplémentaire s'ajoute à la facture. L'alimentation continue en lubrifiant n'est pas en marge de la facture. En effet, celui-ci est pulvérisé en continu au cours de toute la compression à des débits très élevés.

Sur un tout autre plan, on peut également citer les contraintes techniques que posent les aspects réglementaires liés à la validation et la maîtrise du procédé. En effet, la

recommandation Q8 de l'ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use considerations) portant sur le développement pharmaceutique stipule que soit spécifiés et contrôlés de façon appropriée les paramètres critiques des procédés influant la quantité et la qualité des produits. Cela pose surtout un challenge dans la lubrification externe, car optimiser et contrôler la quantité de lubrifiant est un exercice assez complexe, mais surtout coûteux. Il a été démontré que le débit de pulvérisation, le volume d'air de l'aspirateur et la vitesse de rotation de la presse sont tous des paramètres critiques contrôlant la quantité de lubrifiant final dans les comprimés [103]. Il existe d'ailleurs une relation linéaire entre le débit de pulvérisation et la concentration de lubrifiant (Lactose, mannitol, sorbitol, et amidon). C'est pourquoi les arguments principaux des industriels proposant les systèmes de lubrification portent sur leur capacité à délivrer et suivre un débit constant. Lorsque la quantité de poudre dans la lubrification n'est pas maîtrisée, des phénomènes de surlubrification ou de variation significative de la quantité de lubrifiant par comprimé peuvent se manifester. Cela peut d'une part compromettre des attributs qualité comme la dissolution, le temps de désintégration, mais surtout l'uniformité de poids des comprimés. D'autre part, les lots résultants seront non conformes engendrant de la perte de temps et de matières. Le débit de pulvérisation doit en fait être optimisé en déterminant le seuil de concentration de lubrifiant sur la surface des comprimés ou en prenant le débit caractérisé par la plus faible force d'éjection [103]. Alors, dans un cas ou dans l'autre, la réalisation d'expérimentations coûteuses est nécessaire pour fixer et maîtriser ce débit optimal de pulvérisation.

## **5. Conclusions :**

Pour conclure, la lubrification externe est une technique qui permet d'adresser les avantages de la lubrification interne en réduisant le collage et les frictions aux parois des outillages sans les inconvénients sur les attributs qualité des comprimés en termes de dureté et des temps de désintégration et de dissolution. Le principe consiste à délivrer continuellement une quantité minimale de lubrifiant poudreux à l'aide d'air comprimé sur la surface des outillages durant la compression. La propriété principale résultante est l'absence de lubrifiant au sein de la formulation. Se faisant, l'ensemble de ces avantages découle ainsi de l'existence de surfaces libres de films de lubrifiant hydrophobe au cœur des poudres.

La lubrification externe permet notamment de limiter le collage en créant une barrière hydrophobe entre les particules de poudre et les parois des outillages. Mais surtout en

permettant une meilleure cohésion interne du comprimé. Les attributs qualité en termes de dureté et dissolution et la stabilité vis-à-vis du lubrifiant sont également améliorés et le temps de production est réduit grâce à la suppression du mélange final.

Toutefois, la lubrification externe ne réduit pas les frictions aux parois aussi bien que la lubrification interne et c'est un système assez complexe et très coûteux à mettre en place qui va impliquer des études techniques également coûteuses, mais aussi des requis réglementaires et de santé et sécurité pour son implémentation. Bien que prometteuse, l'existence de peu de données scientifiques sur des aspects pouvant présenter un challenge pour la lubrification externe décourage de l'important investissement initial nécessaire.

## **Annexe 2 : Les techniques d'analyse multivariée à réduction dimensionnelle**

Les techniques d'analyse multivariée à réduction dimensionnelle sont un ensemble de méthodes d'analyse de données statistiques qui permettent d'étudier de très grands jeux de données contenant plusieurs variables et plus ou moins d'individus. Nous illustrons ici leurs principes de base à travers l'ACP et la régression multilinéaire.

### **1. Analyse en composantes principales (ACP) :**

L'analyse en composantes principales est une technique d'analyse exploratoire multidimensionnelle qui s'applique aux jeux de données de  $n$  individus (observations) en ligne sur lesquels, sont mesurés  $k$  variables en colonne. Cette analyse permet d'explorer de grands jeux de données afin d'étudier les ressemblances entre individus en termes de profil et d'investiguer les liaisons entre les variables pour mieux faire ressortir les spécificités du jeu de données. L'idée de l'ACP est de résumer l'information contenue dans un jeu de données de grande ( $k$ ) dimension dans quelques variables latentes (2 à 4) de variabilité maximale qui sont aussi appelées composantes principales.

Lorsque les  $n$  individus d'un jeu de données sont décrits par une seule variable, leur représentation est aisée sur un seul vecteur représentant cette variable. À travers cette représentation, nous pouvons décrire et explorer la structure du jeu de données. De même, pour des individus d'un jeu de données décrits par 2 variables, la représentation et l'exploration sont aisées dans un plan formé par les deux variables. Par contre, lorsque les individus d'un jeu de données sont décrits par 3, 4... jusqu'à  $k$  variables, l'appréhension visuelle et l'exploration des similarités et des différences deviennent très difficiles, voire impossibles. Pourtant, une réelle représentation de ce jeu de données dans un espace à 3, 4... jusqu'à  $k$  dimensions existe et sa construction est possible, même si sa lecture nous est quelque peu inaccessible. L'objectif de l'ACP est de restituer le plus fidèlement possible la structure du jeu de données de cet espace multidimensionnel inaccessible sous la forme de représentations simples qui nous sont plus accessibles. C'est à dire une restitution de façon à respecter le plus possible la structure multidimensionnelle d'origine : c'est ce point de départ qui définit le critère de l'ACP. Il faut alors entendre par là que la perte d'informations lors de cette restitution est inévitable, l'objectif est d'en perdre le moins possible.

## 1.1 Le critère de l'ACP :

Soit un jeu de données  $X$  de  $n$  individus classés en ligne et  $k$  variables mesurés en colonne. Les  $n$  individus sont alors représentables dans un espace à  $k$  dimensions. Le critère utilisé dans l'ACP pour restituer le plus fidèlement cet espace multidimensionnel de  $k$  dimensions dans un ou quelques plans est la maximisation de la distance euclidienne des points projetés dans le plan restitué. Soit l'observation  $i$  des  $n$  individus du jeu de données, représentée par le point  $M_i$  dans l'espace multidimensionnel originel de centre  $O$ . Soit  $H_i(I)$ , la projection de  $M_i$  sur le premier axe construit par ACP (composante  $I - U_s$  avec  $s = I$ ). Les plans restitués par l'ACP sont construits axe par axe de telle sorte que la somme des distances euclidiennes  $OH_i(I)$  soit maximale.

$$\frac{1}{n} \sum_{i=1}^n OH_i^2$$

Se faisant l'ACP propose résume les données à travers un axe qui maximise l'appréciation des similarités et différences entre les individus du jeu de données. Le second axe est choisi selon le même critère avec la contrainte d'être orthogonal au premier axe et ainsi de suite pour les autres axes qui sont aussi construits sur le même critère avec la contrainte d'être orthogonal aux précédents axes. Le problème de l'ACP revient alors à trouver les vecteurs  $U_s$  tel que la quantité  $\frac{1}{n} \sum_{i=1}^n OH_i^2$  soit maximale avec la contrainte de  $U_s$  orthogonal aux  $U_{s-1}$ .

Lors de la réalisation de l'analyse, nous centrons les données, ce qui fait que maximiser la quantité ci-dessus revient à maximiser la variance, la variabilité contenue dans le jeu de données : on parle d'inertie. L'ACP restitue donc, par ordre décroissant, une suite d'axes orthogonaux d'inertie maximum ( $U_s$ ).

## 1.2 La solution :

$U_s$  est un vecteur unitaire de la matrice de corrélation/variance-covariance associée à la valeur propre de rang  $s$  ( $\lambda_s$ ), c'est à dire les vecteurs propres associés à la diagonalisation de la matrice de corrélation/variance-covariance représentent les vecteurs  $U_s$  avec les valeurs propres associées déterminant l'inertie restituée par le vecteur correspondant.

L'analyse en composantes principales consiste alors à récupérer ces éléments propres et explorer la structure du jeu de données (individus et variables) dans les nouveaux axes. La première étape de l'analyse consiste au centrage et la réduction des données, puis on établit la matrice de variance-covariance, puis par différents moyens nous récupérerons les éléments

propres de cette matrice. Enfin, nous pourrions accéder aux nouveaux axes et représentations pour étudier la structure des données.

### **1.3 Le centrage des données :**

Cette étape consiste à ramener le centre de gravité au centre du nuage de points en retirant à chaque valeur d'une variable, la moyenne de l'ensemble de ses observations. Cela permet de conserver les mêmes inerties pour les composantes indépendamment de la position et de la forme du nuage de points.

### **1.4 La réduction des données :**

Les données sont réduites lorsque les différentes variables décrivant les individus n'ont pas les mêmes unités, car cela pourrait grandement influencer le choix des nouveaux axes en conférant des poids plus élevés aux variables ayant de plus grandes variances à cause de leur unité. Donc c'est un processus de normalisation, en se plaçant dans un système de mesure par unité de variance. La réduction consiste à diviser une donnée par l'écartype de la variable concernée.

### **1.5 La matrice de corrélation / la matrice de variance-covariance**

La matrice de corrélation lorsque les données n'ont pas été réduites et variance-covariance dans le cas de données réduites est construite respectivement à l'aide des formules usuelles de calcul de la corrélation et de la covariance.

### **1.6 Les éléments propres**

Comme discuté précédemment, les vecteurs propres représentent les vecteurs unitaires de base des nouveaux plans et les valeurs propres associées, leurs inerties (variances expliquées). Les vecteurs permettent alors de passer d'un espace à l'autre (loadings) et donc d'accéder aux scores sur les nouveaux axes.

### **1.7 Le graphique des scores et le cercle de corrélation**

Le graphique des scores est une représentation plane des coefficients des individus sur les composantes formées. Tandis que celle-ci permet d'observer les ressemblances et disparités entre individus dans les nouveaux plans, les liaisons entre variables sont étudiées en calculant et représentant les coefficients de corrélation entre chaque variable et chaque composante. Cela permet également de donner un sens à chaque composante et ainsi d'expliquer les éventuels regroupements observés sur le graphique des scores.



Cette description de l'ACP illustre le principe de base de la réduction dimensionnelle. Elle permet ainsi de comprendre de façon ludique la plupart des autres méthodes d'analyse multivariée à réduction dimensionnelle. Par exemple, la PCR est la combinaison de l'ACP et de la régression multilinéaire où les composantes principales extraites par ACP sont utilisées comme prédicteurs dans un processus de régression multilinéaire. La PLS est quant à elle similaire à la PCR, à la différence que le critère d'extraction des composantes est différent de celui de l'ACP. En plus de la variance de la matrice des prédicteurs utilisée dans l'ACP, l'extraction des composantes par PLS prend aussi en compte la covariance entre les prédicteurs et la ou les variables à prédire [82]. La PLS-DA est une variante de type classification de la PLS ou la variable à prédire est qualitative. Dans ce cas précis, une variable binaire à prédire est générée pour chaque modalité de la variable qualitative originelle et la matrice résultante est utilisée comme jeu de variables à prédire dans une procédure PLS classique [82], [85]. Pour compléter le tour d'horizon, nous proposons dans la section suivante une introduction à la régression multilinéaire.

## **2. La régression multilinéaire :**

### **2.1 Estimation des paramètres du modèle :**

La régression linéaire multiple est une technique d'analyse multivariée utilisée pour expliquer ou prédire une variable ou des variables réponses ( $y$ ) à partir d'un ensemble de variables prédictives ( $x$ ). Soit  $X$ , la matrice des prédicteurs (en ligne les  $n$  observations et en colonne les différents  $k$  variables prédictives). Soit  $Y$ , la matrice des réponses à prédire (en ligne, les réponses dans les  $n$  observations). L'équation de la régression multilinéaire s'écrit :

$$Y = \beta * X$$

Avec  $\beta$ , la matrice des coefficients associés aux  $k$  paramètres + une constante ( $\beta_0, \beta_1, \beta_2 \dots \beta_k$ ). Sous forme matricielle, cette équation s'écrirait :

$$\begin{array}{rcccccc} y_1 & \beta_0 & \mathbf{1} & x_{11} & \dots & x_{k1} \\ y_2 & \beta_1 & * \mathbf{1} & x_{12} & \dots & x_{k2} \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ y_n & \beta_k & \mathbf{1} & x_{1n} & \dots & x_{kn} \end{array}$$

Tout le travail dans la régression multilinéaire consiste à estimer les paramètres bêta. Cet **estimateur**  $\hat{\beta}$ , nous pouvons le déduire ainsi qu'il suit :

$$\hat{\beta} * X = Y \Rightarrow$$

$$\hat{\beta} * X' . X = X' . Y \Rightarrow$$

$$\hat{\beta} * (X' . X)^{-1} . X' . X = (X' . X)^{-1} . X' . Y \Rightarrow$$

$$\boxed{\hat{\beta} = (X' . X)^{-1} . X' . Y}$$

Après avoir défini les coefficients du modèle de prédiction, nous testons d'abord la significativité du modèle global et cela est effectué grâce au calcul du coefficient de détermination  $R^2$  et à une analyse de la variance expliquée par rapport à la variance résiduelle du modèle.

## 2.2 Significativité du modèle :

### 2.2.1 Coefficient de détermination du modèle $R^2$ :

$$R^2 = \frac{\text{Somme des carrés des écarts au modèle (SCEM)}}{\text{Somme des carrés des écarts totaux (SCET)}}$$

$$= \left( \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \right)$$

$$= \left( 1 - \frac{\text{Somme des carrés des écarts résiduels (SCER)}}{\text{Somme des carrés des écarts totaux (SCET)}} \right)$$

$$= \left( 1 - \frac{\sum(\hat{y}_i - y_i)^2}{\sum(y_i - \bar{y})^2} \right)$$

$$= (\text{corrcoef.}(\hat{y}_i, y_i))^2$$

### 2.2.2 Analyse de la variance :

Analyse de la variance avec  $H_0$  : le modèle n'explique rien de significatif par rapport aux résidus.

$$F_{obs} = \frac{\text{Carré Moyen dû au Modèle (CMM = SCEM/ddl associé (k - 1))}}{\text{Carré Moyen Résiduel (CMR = SCER/ddl associée (n - k))}}$$

$$\rightarrow \text{Fischer snedecor}(0,95 ; k - 1 ; n - k)$$

Si le modèle est testé non significatif : soit le choix des prédicteurs n'est pas bon (ils ne permettent pas d'expliquer, de prédire la réponse), soit la variance résiduelle est trop grande (expériences difficiles à réaliser) ou il y'a peu de données pour permettre de bâtir un modèle prédictif significatif. À l'inverse, si le modèle global est testé significatif, nous nous

intéresserons maintenant de plus près à chaque paramètre et cela en déterminant l'erreur sur chacune d'elles, test de la significativité et leur intervalle de confiance :

- **Erreur standard sur chaque paramètre  $\hat{\beta}_j$**

Cette erreur ( $\widehat{\sigma_{\beta_j}}$ ) est estimée grâce à l'erreur standard du modèle global ( $\sigma$ ) :

$$\widehat{(\sigma_{\beta_j})}^2 = (X' \cdot X)_{jj}^{-1} \cdot \sigma^2 \text{ avec } \sigma^2 = \frac{SCER}{(n - k)}$$

- **Test de la significativité de chaque paramètre  $\hat{\beta}_j$**

On réalise un test de Student (0,95 ;  $n-p-1$  [ $k=p+1$  pour la constante]) avec  $H_0 : \beta_j = 0$  (pas

d'effet significatif) :  $L\left(\frac{\hat{\beta}_j}{\widehat{(\sigma_{\beta_j})}}\right) \rightarrow T_{(0,95 ; n-k)}$

- **Intervalle de confiance associé à chaque paramètre  $\hat{\beta}_j$**

$$IC(\hat{\beta}_j) = \hat{\beta}_j \pm t\left(1 - \frac{\alpha}{2}; n - k\right) * \widehat{(\sigma_{\beta_j})} \text{ avec un risque } \alpha$$

En utilisant la régression multilinéaire, la PLS, la PLS-DA et la RCP permettent de bâtir des modèles sur les grands jeux de données. Ces modèles sont généralement validés par la technique de la validation croisée ou validation externe et permettent d'expliquer les relations entre les variables originelles et la réponse.

## **Annexe 3 : la gestion des données manquantes**

### **1. Introduction :**

L'analyse de données au sens large est une discipline scientifique et technique visant à faire émerger l'information (connaissance) existante dans la structure des données. Ce processus peut s'apparenter à celui de la traduction linguistique où les données constituent en soi un langage d'expression à part entière de l'information utile. L'analyse serait alors la stratégie de traduction de ce langage en un autre plus explicite et exploitable pour formaliser des conclusions, des théories, des lois et bâtir des modèles théoriques ou expérimentaux.

Aujourd'hui, la science moderne nous permet de disposer d'innombrables stratégies de traduction, c'est-à-dire de nombreuses techniques d'analyse de données. La plupart de ces techniques ne fonctionnent que si le jeu de données à analyser est complet, c'est-à-dire qu'elle n'accuse d'aucune donnée manquante. Dans un monde parfait, les données à analyser et exploiter seraient alors complètes, mais hélas ce n'est pas le cas. Dans le monde réel, nous avons très généralement à faire à des jeux de données incomplets, et ce dans toutes les disciplines scientifiques.

Une donnée manquante dans un jeu de données est toute observation dont la valeur n'est pas renseignée. Le jeu de données est alors qualifié d'incomplet et l'analyse doit être menée adéquatement pour éviter de fausses conclusions. Les données peuvent manquer pour diverses raisons : l'incapacité d'accéder aux informations, pannes ou indisponibilités d'appareil de mesures, oubli de report d'un résultat... Le manque de données est souvent cité comme plutôt la règle qu'une exception [104]. Aussi, selon différents auteurs, il conviendrait de considérer également les données aberrantes qui peuvent être constatées dans les jeux de données comme des données manquantes lorsque la connaissance physique du phénomène étudié exclut l'occurrence de ces valeurs [105]. Ce qui étend encore plus l'envergure de l'occurrence des données manquantes. La gestion des données manquantes constitue alors un challenge incontournable dans toutes les disciplines scientifiques.

Les données étant le support de l'information, tout manque sera fort probablement dommageable. En gardant l'analogie de la traduction linguistique, l'existence de données manquantes dans un processus d'analyse s'apparenterait à la traduction d'un texte avec des mots manquants ou illisibles. Cela conduira inéluctablement dans le meilleur des cas à de la perte d'information. Dans le pire des cas, elle conduira à une fausse interprétation ou

l'incapacité totale d'exploiter le texte. Dans le cadre de l'analyse de données, l'existence de données manquantes induit alors une perturbation dans la structure et les relations limitant plus ou moins fortement l'exploitation de ces données. Il existe d'innombrables techniques de gestion des données manquantes et dépendamment de la façon dont elles sont prises en charge, elles peuvent conduire à des estimateurs biaisés et contribuer à réduire la puissance statistique des techniques d'analyse employées. Elles peuvent avoir également pour effet de réduire la représentativité des échantillons analysés et donc donner lieu à des inférences statistiques et conclusions erronées [105]. Toutefois, lorsqu'elles sont bien prises en charge, leur impact est minimisé et pris en considération dans la conclusion finale émise.

Dans ce document, nous allons nous pencher sur les techniques de gestion de données manquantes en analyse de données. Nous commencerons par un rapport de revue littéraire sur le sujet, où nous décrirons les différents mécanismes de données manquantes et leur structure avant de discuter des différentes techniques de traitement des données manquantes en mettant l'accent sur le principe, la méthode, les avantages et les inconvénients associés à chacun. Ce rapport littéraire sera suivi d'une section d'application où nous allons utiliser les techniques couramment employées sur un jeu de données complet et différentes de ses variantes auxquelles nous aurons retiré de façon aléatoire de 2 % à 30 % des données. Enfin, nous procéderons à la comparaison des différentes sorties et une conclusion.

## **2. Revue de littérature sur la gestion des données manquantes :**

La littérature sur la gestion des données manquantes remonte aux années 1930 [105]. Les articles traitant du sujet ont commencé à abonder à partir des années 1970 et continuent jusqu'à nos jours. Ceci témoigne de l'intérêt que porte la science moderne à ce domaine, mais surtout le rôle fondamental que celui-ci joue dans la validité des conclusions scientifiques.

Lorsque nous avons à travailler avec des jeux de données incomplets, l'activité préalable au bon choix de toute méthode de traitement est de décrire au mieux la nature des données manquantes. Cette activité préalable est conseillée par tous les écrits sur le sujet et remonte à l'établissement des concepts de base du traitement des données manquantes par Little et Rubin en 1987 [105]. La description de la nature des données manquantes permet de réfléchir en amont sur les conséquences que peut avoir la méthode de traitement sur les résultats de l'analyse. C'est alors une étape très importante pour pouvoir choisir une méthode adéquate qui minimise l'impact des manquants. Deux caractéristiques principales permettent de décrire la nature des manquants : la configuration et le mécanisme. La configuration a trait à

l'emplacement et la disposition des observations manquantes dans la structure du jeu de données. Elle est utile lorsqu'on travaille sur des données longitudinales. N'étant pas dans ce cas, la configuration reste peu pertinente pour nous et nous n'allons pas l'aborder dans cette revue de littérature. Cependant, le mécanisme de données manquantes est très important pour tout type de données. Il porte sur le processus ayant conduit à la « non-réponse ». Nous décrivons dans cette section cet aspect fondamental de la gestion des données manquantes.

## **2.1 Les mécanismes de données manquantes :**

On distingue trois mécanismes de données manquantes : les données manquantes complètement au hasard (MCAR pour « Missing Completely At Random »), les données manquantes au hasard (MAR pour « Missing At Random ») et les données non manquantes au hasard (MNAR pour « Missing Not At Random »).

### **2.1.1 Les données manquantes complètement au hasard (MCAR) :**

Soit un jeu de données rectangulaire de  $X_1$ ,  $X_2$  et  $X_3$  variables avec  $n$  individus. Lorsqu'une observation est manquante sur la variable  $X_1$ , celle-ci est dite MCAR si et seulement si la probabilité de sa « non-réponse » est indépendante des valeurs prises par les toutes les variables du jeu de données  $X_1$ ,  $X_2$  et  $X_3$  de l'observation qu'elles soient manquantes ou pas. En d'autres termes, la non-réponse ne tient qu'au pur hasard, c'est-à-dire que la probabilité de non-réponse est uniforme. On exprime très souvent cette probabilité avec l'expression mathématique :  $P(r | \phi)$  où  $r$  représente la réponse et  $\phi$  décrit la relation qui existe entre les données et  $r$ . Cette relation signifie que la probabilité de réponse (non-réponse) est conditionnée aux seuls paramètres décrits par  $\phi$ , excluant toutes les variables du jeu de données.

Pour illustrer ce mécanisme dans un cas concret, imaginons que nous réalisons une collecte de données industrielles en vue d'étudier le collage. Après avoir planifié les différents lots sur lesquels les données seront collectées, imaginons que l'enregistrement de la force de compression ait été omis pendant un certain temps durant la production d'un lot. Ces observations qui vont manquer dans notre base de données sont dits MCAR si et seulement si l'omission n'a aucun lien avec la force de compression elle-même ou les autres variables collectées. Par exemple, si elle est due à l'oubli d'un opérateur de production de lancer le programme d'enregistrement ou à un dysfonctionnement ponctuel de la presse à comprimer qui n'est aussi relié à aucune des autres variables collectées. Le mécanisme MCAR est le processus de manque qui influe le moins sur la représentativité de l'échantillon, car les données observées constituent toujours un échantillon aléatoire [104–106].

### **2.1.2 Les données manquantes au hasard (MAR) :**

Une donnée est dite manquante au hasard si et seulement si sa probabilité de non-réponse peut être corrélée à l'existence d'autres variables renseignées du même jeu de données, mais pas aux données manquantes elles-mêmes. En d'autres termes, la non-réponse sur une variable  $X$  du jeu de données peut dépendre des valeurs d'une ou de plusieurs autres variables du jeu de données excepté  $X$  (c'est-à-dire la valeur qu'elle aurait si elle n'était pas manquante). Le fait que  $X$  soit manquant n'a donc aucun lien avec sa valeur hypothétique. Par ailleurs, un lien subsiste entre d'autres variables observées du jeu de données et la probabilité de non-réponse de  $X$ . Le mécanisme MAR est décrit généralement avec l'expression mathématique :  $P(r | \phi, X_{\text{observés}})$  où  $r$  représente la réponse,  $X_{\text{observés}}$  représentent les variables observées et  $\phi$  décrit la relation qui existe entre les données et  $r$ . Cette relation signifie que la probabilité de réponse (non-réponse) peut être conditionnée aux paramètres décrits par  $\phi$  et les  $X_{\text{observés}}$ . Elle exclut la variable concernée par le manque de données.

Pour illustrer le mécanisme MAR, poursuivons avec notre exemple de la section précédente sur la collecte de données industrielles pour étudier le collage. Imaginons cette fois que l'enregistrement de la force de compression est manquant à cause du fait qu'à certaines vitesses de compression et de rotation, les temps d'acquisition et de réponse des capteurs de force peuvent faire que la donnée ne soit pas enregistrée. La probabilité que la force de compression soit alors manquante dans notre jeu de données ne dépend pas de l'intensité de cette force, mais elle peut être corrélée aux vitesses de compression et de rotation. Elle est alors dite manquante au hasard. En fait, ce terme peut prêter à confusion, car les données ne sont pas tout à fait manquantes au hasard au sens littéral, mais seulement manquant par un concours de circonstances qui peut être capturé par d'autres variables du jeu. Le mécanisme MAR est le processus le moins restrictif, car elle nécessite plus essentiellement que les données ne manquent juste pas à cause de leur propre valeur. De ce point de vue, le mécanisme MCAR peut être vu comme un cas particulier de MAR où la non-réponse n'a aussi aucune corrélation avec les autres variables observées du jeu de données [104,105].

### **2.1.3 Les données non manquantes au hasard (MNAR) :**

Une observation d'un jeu de donnée est dite non manquante au hasard si et seulement si sa probabilité de non-réponse est dépendante exclusivement de sa valeur même. En d'autres termes, la donnée manque à cause de la valeur qu'elle prendrait si elle était observée. On peut l'écrire avec l'expression mathématique suivante :  $P(r | \phi, X_{\text{manquant}})$  où  $r$  représente la

réponse,  $X_{\text{manquant}}$  représente la variable pour laquelle l'observation est manquante et  $\phi$  décrit la relation qui existe entre les données et  $r$ . Cette relation signifie que la probabilité de réponse (non-réponse) est conditionnée aux seuls paramètres décrits par  $\phi$  et la valeur manquante elle-même.

En reprenant notre exemple sur la collecte de données industrielles en vue d'étudier le collage, le mécanisme MNAR est illustrée lorsque l'enregistrement de la force de compression est manquant à cause d'un seuil de détection des capteurs qui a été dépassé lors de la compression. Dans ce cas, la non-réponse de la force de compression ne dépend que de sa valeur. Le mécanisme MNAR est le processus le plus délicat à traiter en gestion de données manquantes, car la non-réponse est due à une raison qu'il faut scrupuleusement comprendre pour mieux travailler sur la base de données incomplète.

La description de ces hypothèses trouve son importance dans la possibilité d'ignorer ou pas le motif d'absence. En effet, lorsque les données sont manquantes selon un processus MCAR ou MAR, le motif d'absence est dit ignorable. C'est-à-dire qu'il n'y a pas lieu de se pencher sur le pourquoi de l'absence de données puisque celle-ci est aléatoire [105]. En revanche, lorsque les données manquent par un processus MNAR, le motif d'absence est dit non ignorable et pour éviter de fausses conclusions d'emblée, nous avons tout intérêt à nous pencher sur le comportement d'absence des données. Reprenons notre exemple où la force de compression est manquante dans notre collecte de données à cause d'un seuil de détection des capteurs. Dans ce cas de données manquantes MNAR, on ne pas ignorer le processus à l'origine des non-réponses, car on sait d'avance que les valeurs manquantes sont des valeurs strictement supérieures au seuil de détection des capteurs de force. Dans les cas MNAR, ce genre de comportement doit être modélisé à l'aide des connaissances physiques du phénomène étudié avant le choix de toute technique de gestion des données manquantes. Cela permet de s'affranchir d'induire un énorme biais dans les données bien avant leur analyse.

Par exemple, si nous ne tenons pas compte du processus ayant généré les non-réponses de la force de compression et que nous remplaçons les données manquantes par la moyenne des données observées ou que nous décidions d'enlever les lots avec données manquantes de notre analyse (deux techniques de gestion de données manquantes couramment utilisées par défaut dans les logiciels de statistiques et par les scientifiques). Dans le premier cas, nous induisons un très fort biais dans les données, car les valeurs réelles sont bien supérieures à toutes les valeurs observées. Nous contribuons également à réduire la dispersion des données. Dans le cas où nous omettons les observations avec données manquantes, nous affaiblissons



considérablement notre échantillon, car tout lot au cours duquel la force de compression a pu dépasser une fois le seuil de détection des capteurs de force est abandonné. En plus, nous induisons également un biais, car le fait que la force de compression dépasse le seuil de détection peut avoir un intérêt dans l'étude réalisée. En omettant les observations avec données manquantes, nous nous privons de l'étude de ce lien d'intérêt.

Il est alors primordial d'observer et de comprendre le jeu de données pour en déterminer les mécanismes de non-réponse avant le choix de toute technique de traitement. Par ailleurs, l'analyse du processus d'absence de données est très compliquée et nécessite une réelle attention, connaissance du phénomène étudié et du processus de collecte, car, il n'existe pas de techniques pour tester et définir si les données sont (MCAR, MAR) ou MNAR. En effet, le plus souvent, c'est l'analyste qui émet une hypothèse sur le processus de manquants. Pour aider la formulation de cette hypothèse, la littérature suggère de chercher des travaux similaires caractérisés par des données manquantes qui peuvent informer sur le mécanisme ou d'inclure dans l'étude des variables qui pourront aider à l'explication des non-réponses pour rendre probable l'hypothèse MAR plutôt que MNAR [105].

## **2.2 Les méthodes de traitement des données manquantes :**

Après avoir postulé une hypothèse sur le mécanisme de données manquantes, nous allons nous focaliser maintenant sur les méthodes de traitement. Historiquement, la méthode classique de traitement des données manquantes était l'analyse des cas complets [104,105]. C'est-à-dire, l'analyse basée sur les individus où toutes les données étaient observées.

En raison de leur nombre, l'énumération de toutes les méthodes de traitement des données manquantes serait impossible. Mais, globalement nous pouvons distinguer quatre stratégies principales : la stratégie de traitement ignorant les données manquantes, l'imputation simple, l'imputation multiple et les méthodes de traitement intégrés aux algorithmes d'analyse [104,105]. La stratégie ignorant les données manquantes regroupe les techniques qui consistent à travailler uniquement sur les données observées. Les techniques d'imputation simple portent sur le remplacement des données manquantes par des valeurs estimées à partir de différentes méthodes pour disposer d'une base de données complète sur laquelle mener l'analyse. L'imputation multiple est une stratégie basée sur plusieurs imputations simples d'une même base et de mener l'analyse sur chacune des bases complètes générées à l'issue de ce processus. Se faisant, nous aurons non pas une seule estimation pour la valeur manquante imputée, mais plusieurs valeurs plausibles pour refléter l'incertitude due l'imputation elle-même. La dernière

stratégie concernant les méthodes de traitement intégrées aux algorithmes d'analyse ne fait qu'utiliser des variantes de techniques d'analyse pouvant tolérer des valeurs manquantes. Nous décrivons dans les sections qui suivent les techniques les plus couramment employées pour chacune de ces stratégies.

### **2.2.1 Stratégies de traitement ignorant les données manquantes :**

Deux techniques sont les plus couramment citées : l'analyse des cas complets et l'analyse des cas disponibles.

#### **- Analyse des cas complets :**

Cette technique consiste à supprimer de l'échantillon toutes les observations concernées par une donnée manquante. En fait, c'est la méthode la plus simple et courante employée par les scientifiques [104]. Le jeu de données à analyser ne comportera alors aucune donnée manquante. C'est une des solutions par défaut implémentées dans les logiciels statistiques. Pourtant l'analyse des cas complets accuse de nombreuses lacunes. Principalement, elle occasionne une perte d'informations plus ou moins importante et peut induire un biais important dans les données, mais aussi réduire la représentativité de l'échantillon [106].

Lorsque les données manquantes sont MNAR, ignorer les observations avec données manquantes revient à retirer un échantillon non aléatoire du jeu de données. Cela conduit alors à biaiser l'analyse si les données retirées traduisaient une partie du phénomène. Sous l'hypothèse MAR où les données sont manquantes pour des valeurs particulières d'autres variables observées dans le jeu de données, l'analyse des cas complets revient à exclure de l'échantillon toutes les observations pour les valeurs particulières des variables observées corrélées au mécanisme de réponse. Cela introduit également un important biais dans l'analyse. Enfin, si les données manquantes sont MCAR, retirer les observations avec des valeurs manquantes n'introduit pas forcément de biais dans l'analyse. Le mécanisme de non-réponse étant complètement aléatoire, le sous-échantillon complet restant est aussi censé être un échantillon aléatoire. Cependant, lorsque le nombre d'observations complètes n'est pas très élevé, la technique d'analyse des cas complets limite la qualité de l'analyse. En effet, dans un échantillon où nous disposons de différentes classes d'observation, les plus rares peuvent être réduits à un nombre insuffisant ou être complètement omis de l'échantillon occasionnant un biais. En somme, l'analyse des cas complets ne reste envisageable que pour de faibles proportions de données manquantes (moins de 5 % selon la littérature) où le mécanisme de manque est MCAR. La quantité de données à supprimer est donc la limite majeure de cette

technique. Afin de pallier cela, une autre technique ignorant les données manquantes a été proposée : l'analyse des cas disponibles.

- **Analyse des cas disponibles :**

Cette méthode n'exclut aucune ou peu d'observations. L'analyse des cas disponibles consiste tout simplement à l'étude du jeu de données tel qu'il est. En fait, différents aspects du phénomène étudié sont investigués avec divers sous-échantillons. Cela peut être réalisé en faisant des suppressions par pool de variables ou d'individus, communément appelé analyse par paires (« pairwise deletion »). La procédure entraîne alors moins de suppression de données. La suppression d'une variable à cause d'une grande quantité de données manquantes est aussi une procédure d'analyse des cas disponibles.

Le problème majeur de cette technique est que les conclusions tirées de différents sous-échantillons ne sont pas totalement compatibles. Aussi, des biais peuvent toujours subsister dans les différents sous-échantillons [107]. Dans les cas MNAR et MAR, cette technique est aussi déconseillée [104,105]. Par ailleurs, bien que pouvant être appliquée sous hypothèse MCAR, l'utilisation de l'analyse des cas disponibles reste limitée. Cela peut s'expliquer par la complexité de la technique quand elle peut être utile. En effet, c'est en présence d'une proportion importante de données manquantes que l'analyse des cas disponibles peut s'avérer plus utile que l'analyse des cas complets. Cependant, sous ces mêmes conditions, elle devient très complexe à opérer, car de nombreux sous échantillons seront nécessaires et les différents aspects du problème étudié seront décousus et compliqués à recouper.

L'utilisation des différentes stratégies de traitement ignorant les données manquantes est donc limitée à des mécanismes MCAR où la proportion de données manquantes est très faible. Autrement, elles peuvent induire un biais dans l'analyse ou s'avérer très complexes à opérer et tirer des conclusions fiables. Pour d'autres besoins, des techniques plus sophistiquées dites d'imputation sont employées.

## **2.2.2 Les stratégies d'imputation simple :**

L'imputation au sens général du terme est la méthode standard de gestion des données manquantes en analyse de données. Elle consiste à remplacer les valeurs manquantes par des estimations obtenues selon différentes procédures. Elle est dite simple lorsque chaque valeur manquante est remplacée par une seule valeur plausible pour générer une nouvelle base de données complète à analyser. Cette technique est bien adaptée aux cas de données MCAR et surtout MAR, car les valeurs manquantes sont estimées à partir de l'information disponible.

Cependant, dans le cas MNAR l'imputation simple est très difficilement justifiable, car la non-réponse est indépendante des variables observées et n'est pas aléatoire. Les méthodes d'imputation simple surclassent donc les stratégies d'analyse ignorant les données manquantes.

- **Imputation par la moyenne :**

L'imputation par la moyenne est sans doute la technique la plus populaire. C'est une technique très simple et efficace qui consiste à remplacer chaque valeur manquante par la moyenne des valeurs observées de la même variable. Bien que très employée encore aujourd'hui, plusieurs études ont pointé les limites de la technique. Principalement, l'imputation par la moyenne biaise tous les paramètres de la distribution du jeu de données excepté la moyenne [105]. En effet, les variances, covariances et corrélations des variables sont toutes diminuées. Ce qui a pour effet de réduire la dispersion de chaque variable et les éventuelles associations multidimensionnelles entre les variables [104]. L'emploi de cette méthode dans des analyses multivariées peut alors être préjudiciable.

- **Imputation par la médiane ou la mode :**

En présence de valeurs aberrantes, l'imputation par la moyenne peut vite devenir problématique. Pour parer à cela, l'imputation par la médiane pour des variables quantitatives ou par la mode pour les variables qualitatives a été proposée par plusieurs auteurs [105,108]. Toutefois, il a été rapporté que celle-ci accuse aussi d'effets nuisibles sur les estimations des paramètres de la distribution [104].

- **Imputation par régression :**

L'imputation par la régression consiste à estimer les paramètres d'un modèle de régression quelconque (linéaire, logistique...) à partir des informations disponibles et d'utiliser ces paramètres pour prédire les valeurs des observations manquantes. Pour mieux comprendre, soit trois variables  $X_1$ ,  $X_2$  et  $X_3$  constitutifs d'un jeu de données de  $n$  individus. Supposons que pour  $k$  individus, des données sont manquantes sous hypothèse MAR ou MCAR pour la variable  $X_1$ . La méthode de l'imputation par régression consiste à estimer les paramètres d'un modèle quelconque défini par l'analyste en utilisant les données des  $(n-k)$  individus de  $X_2$  et  $X_3$  comme variables explicatives et celles de  $X_1$  comme variables à prédire. Les paramètres de ce modèle sont ensuite utilisés pour prédire les  $k$  données manquantes de  $X_1$  avec les  $k$  données observées de  $X_2$  et  $X_3$ . Cette technique se prête bien par sa nature à l'imputation de données manquantes MAR où la non-réponse est corrélée aux variables observées. Elle est

applicable également en cas de données manquantes MCAR, car les variables ont tendance à être corrélées dans un jeu de donnée [109].

Bien que réputée meilleure par rapport à l'imputation par la moyenne, l'imputation par régression induit également des biais dans l'analyse. Les valeurs de remplacement seront fortement corrélées avec les autres variables, ce qui surestime les corrélations [110]. Dans un grand jeu de données avec plusieurs valeurs manquantes sur plusieurs variables, les surestimations des corrélations pourraient alors compromettre la structure multidimensionnelle du jeu de données. L'imputation par régression conduit également à une déformation de la distribution des données imputées, résultant en une sous-estimation de leur variance.

Pour pallier cet aspect, l'imputation par régression stochastique fut proposée. Elle consiste à réaliser l'imputation classique par la régression et d'ajouter aux valeurs prédites un résidu aléatoire pour rétablir la variabilité. Les valeurs de remplacement calculées par régression sont alors augmentées d'un résidu aléatoire tiré d'une distribution normale de moyenne nulle et de variance égale à la variance des résidus de la régression. Cette modification limite le biais introduit par l'estimation des valeurs manquantes [110].

D'autres variantes de l'imputation par régression existent avec les mêmes caractéristiques permettant de limiter le biais d'estimation comme dans l'imputation par régression stochastique. Nous pouvons citer notamment l'imputation par les forêts aléatoires qui est d'un tout autre registre, mais s'apparente parfaitement aux techniques d'imputation par régression. Le principe est le même qu'une imputation par régression classique, sauf que c'est un modèle de type forêts aléatoire qui est construit en lieu et place du modèle de régression et l'imputation s'effectue de façon itérative jusqu'à la convergence basée sur un critère fixé ou un nombre maximal d'itérations fixé d'avance. En effet, l'imputation par les forêts aléatoires est réalisée grâce à un algorithme itératif (Algorithme « randomforest ») où les données imputées sont mises à jour à chaque itération. Le critère de convergence est atteint lorsque la différence entre les valeurs de remplacement à l'étape  $t$  et  $(t-1)$  de l'itération augmente. Le principe des forêts aléatoires est de réduire l'erreur de prédiction en utilisant la moyenne des prédictions de plusieurs modèles indépendants [104]. Ces modèles sont construits par la sélection de plusieurs échantillons bootstrap. De ce fait, la méthode d'imputation à l'aide des forêts aléatoires limite aussi le biais d'estimation.

- **Imputation par la méthode des k plus proches voisins (K-NN) :**

L'imputation par les K plus proches voisins (K-NN pour « K-Nearest Neighbor ») est une technique d'estimation des valeurs manquantes très utilisée qui consiste à choisir les nouvelles valeurs à imputer parmi les observations semblables non manquantes de la même base. En fait, la procédure consiste à identifier les k observations les plus similaires à l'observation dont une valeur est manquante et d'imputer la valeur manquante par la moyenne de la valeur prise par les k observations similaires pour une variable continue. Pour une variable discrète ou qualitative, ce sera le mode qui sera utilisé. L'algorithme pour réaliser cette procédure peut s'écrire ainsi qu'il suit :

*Algorithme des K plus proches voisins (K-NN)*

- Choisir un entier  $k : 1 \leq k \leq n$  ;
- Calcul des distances  $d(Y_{\text{immanquant}}, Y_{\text{observés}})$  pour toutes les observations non manquantes ;
- Retenir les k observations pour lesquelles ces distances sont les plus faibles ;
- Affecter aux valeurs manquantes la moyenne/le mode des valeurs des k voisins.

Dépendamment de la façon de calculer la distance entre les observations, il existe différentes variantes de cette technique. Les distances euclidiennes et celle de Mahalanobis sont les plus couramment employées. L'imputation K-NN ne déforme pas la distribution des données, mais l'inconvénient principal de cette méthode porte sur le temps de calcul qui peut devenir relativement très important pour de grands jeux de données avec beaucoup de valeurs manquantes. Le choix du nombre de voisins K est aussi très important. Un K très élevé revient à quasiment réaliser une imputation par la moyenne par exemple. Malheureusement, il n'y a pas de règles scientifiques pour le choix du nombre de voisins K. Cela dépend fortement de la taille de la base et de la proportion de données manquantes.

- **Imputation par la méthode du maximum de vraisemblance :**

Les données manquantes peuvent aussi être imputées par la maximisation de la fonction de vraisemblance. Rappelons que la technique du maximum de vraisemblance est une méthode paramétrique qui permet de trouver les estimations de paramètres inconnus qualifiant la distribution des données observées. C'est-à-dire lorsqu'il n'y a aucune donnée manquante, les estimateurs fournis par cette méthode sont ceux qui maximisent la probabilité d'observer ce qui a réellement été observé. Cette probabilité est conditionnelle aux variables observées et un ensemble de paramètres pour toutes les données. En présence de données manquantes, deux

fonctions de vraisemblance conditionnent la probabilité (une fonction pour les données observées et une autre pour les données partiellement observées). L'estimation des paramètres se fait de façon itérative par des algorithmes de calcul numérique. En fait, le critère maximisé en pratique est le log-vraisemblance. Il fut proposé un algorithme EM (Estimation-Maximisation) permettant de réaliser le maximum de vraisemblance en présence de données manquantes [107,111]. Cet algorithme demeure celui utilisé pour l'imputation de données manquantes par la maximisation des fonctions de vraisemblances. Cela est dû à la robustesse de l'algorithme et les excellents résultats rapportés dans son application. Le principe est simple : l'algorithme itère l'étape d'estimation des fonctions de vraisemblance (étape E) et celle de maximisation du log-vraisemblance (étape M) jusqu'à la convergence. En sortie, nous obtenons une matrice de données complète où les données manquantes ont été imputées à la suite des étapes d'itérations.

Dans cette partie concernant les méthodes d'imputation simple, nous avons décrits les techniques couramment employées, leur principe, leurs avantages et inconvénients. Il en existe de nombreuses autres qui sont également utilisées par la communauté scientifique dans des cas particuliers ou pour leur propriété particulière. Nous pouvons citer notamment les méthodes d'imputation stationnaire employées très généralement dans les études longitudinales comme la méthode de la valeur la plus fréquemment rencontrée (CMCF pour « Concept Most Common Attribute Value Fitting ») ou plus simplement celle de la dernière valeur connue [LOCF pour « Last observation carried forward »]. On peut citer également la méthode du « hot-deck » où les valeurs manquantes sont remplacées par celle d'une observation provenant du même jeu de données, la méthode du « cold-deck » où les valeurs de remplacement sont choisies dans les observations d'un jeu de données similaire et bien d'autres techniques.

De ces nombreuses méthodes d'imputation simple, nous pouvons surtout retenir que la majorité d'entre elles ne fonctionne que sous hypothèse MCAR et/ou MAR. Elles fournissent pour certains des paramètres biaisés et pour d'autres des paramètres sans biais, mais aucune d'entre elles n'est adaptée à l'imputation de données manquantes MNAR.

### **2.2.3 Les méthodes d'imputation multiple :**

L'imputation multiple est la technique par excellence de gestion des données manquantes. C'est la méthode la plus aboutie qui existe aujourd'hui pour travailler avec des jeux de données incomplètes. Elle a été développée par Rubin en 1976. Le principe est simple. Il consiste à remplacer chaque valeur manquante de la base de données originale par une série de  $m$  valeurs

plausibles. Cela génère  $m$  versions de la même matrice dans lesquelles les valeurs manquantes sont imputées dans chaque cas par une valeur plausible différente. L'analyse à laquelle les données étaient destinées est réalisée sur chaque version de la matrice originale rendue complète et nous obtenons une série de  $m$  résultats qui seront combinés. Le principe général d'une analyse avec une imputation multiple des données manquantes peut alors être séparé en trois étapes distinctes : l'imputation, l'analyse et la combinaison des résultats.

- **L'étape d'imputation :**

La première étape est l'imputation proprement dite et consiste à générer les  $m$  copies complétées de la base de données originale. Celle-ci est réalisée en utilisant l'information disponible. C'est l'étape la plus importante de la procédure d'imputation multiple, car les valeurs manquantes vont y être estimées. En fonction de la méthode employée à cette étape d'imputation, nous pouvons distinguer 02 grandes approches à l'imputation multiple : l'imputation multiple basée sur une distribution jointe (distribution multivariée normale) et l'imputation multiple par équations chaînées décrites (MICE) [105,108].

L'imputation multiple basée sur la distribution jointe consiste à bâtir un modèle prédictif sur les données disponibles et déterminer les valeurs de remplacement en utilisant la technique de Monte-Carlo par chaîne de Markov [CMCM pour « Monte-Carlo Markov Chain »]. Différents rapports sur l'application de cette approche d'imputation multiple basée sur les CMCM rapportent qu'elle peut générer des valeurs non plausibles à cause de l'hypothèse de distribution normale multivariée [105,108]. En effet, la pratique concède que ce modèle ne convient pas à certains types de variables et le même modèle multivarié étant utilisé pour toutes les variables, des valeurs non plausibles peuvent être prédites. C'est pourquoi on lui préfère la seconde approche basée sur des équations chaînées.

Cette dernière est différente de l'imputation basée sur la distribution jointe par l'aspect conditionnel du modèle bâti en fonction de chaque variable. En effet, elle consiste à bâtir un modèle conditionnel pour chaque variable de façon itérative jusqu'à la convergence. Pour cette raison, elle est aussi appelée spécification conditionnelle pleine [FCS pour « Fully Conditional Specifications »]. L'approche MICE ou FCS est très flexible, car elle ne fait pas l'hypothèse d'une distribution normale multivariée. Ce sont des régressions séquentielles qui sont réalisées jusqu'à la convergence. Un modèle spécifique est bâti pour chaque variable avec des données manquantes et c'est ce modèle qui sera aussi utilisé pour imputer ces données manquantes. Ainsi, au lieu de bâtir un modèle prédictif unique basé sur une distribution jointe,



dans l'approche MICE ou FCS c'est une série de modèles spécifique à la nature de chaque variable qui est utilisée pour imputer les données.

De façon pratique, cela peut se décrire ainsi qu'il suit. Imaginons 10 variables ( $X_1, X_2, \dots, X_{10}$ ) constituant un jeu de données avec chacune plus ou moins de valeurs manquantes pour certaines observations. L'approche MICE consiste à imputer les données manquantes pour  $X_1$  en construisant un modèle prédictif spécifique à  $X_1$  à partir des données observées pour les autres variables ( $X_2, \dots, X_{10}$ ). Pour les données manquantes sur  $X_2$ , un autre modèle prédictif spécifique est bâti avec comme variables explicatives cette fois-ci le jeu ( $X_1, X_3, \dots, X_{10}$ ). Ainsi de suite, des régressions séquentielles spécifiques à chaque variable sont réalisées jusqu'à la convergence. Cela permet par exemple d'utiliser une régression stochastique pour des variables continues, une régression logistique pour une variable binaire et une régression multinomiale pour des variables catégorielles. Cette stratégie rend les prédictions plus plausibles pour chaque donnée manquante. L'approche MICE est réalisée suivant l'algorithme ci-dessous :

*Algorithme d'imputation par équations chaînées*

- *Initialisation : imputation naïve des valeurs manquantes par la moyenne*
- *Pour chaque variable avec des données manquantes (en commençant par celle avec le moins de données manquantes) :*
- *Remplacer les items manquants (celles qui avaient été remplacées par la moyenne) ;*
- *Bâtir un modèle spécifique pour la variable avec les autres variables comme prédicteurs (dans les implémentations de l'algorithme, on peut choisir les modèles et les prédicteurs à utiliser pour chaque variable) ;*
- *Prédire les données manquantes pour la variable ;*
- *Imputer les valeurs prédites pour les données manquantes pour la variable (tout usage futur de cette variable sera fait avec les données observées et les données imputées. C'est-à-dire comme si elle n'avait pas de données manquantes) ;*
- *Répéter les quatre séquences précédentes pour chaque variable avec des données manquantes ;*
- *Répéter les étapes précédentes jusqu'à la convergence des paramètres gouvernant l'imputation (stabilité) ou un nombre de cycles fixé par l'analyste (la littérature suggère 10 cycles);*

Les données manquantes sont ainsi imputées m-fois suivant ces différentes étapes et m-bases complètes avec les valeurs plausibles prédites pour chaque valeur manquante sont générées. Cela permet de prendre en compte l'erreur due au processus d'imputation dans l'analyse. C'est là toute la puissance de l'imputation multiple. En effet, au lieu de choisir une seule valeur plausible pour chaque valeur manquante, elle renseigne m-valeurs plausibles pour refléter l'erreur pouvant être due au fait que les données à analyser ont subi une imputation.

- **L'étape d'analyse :**

Les m-bases générées à l'étape d'imputation sont analysées indépendamment avec la méthode d'analyse qui était planifiée. Cela génère m-sorties d'analyse qu'il faudra combiner.

- **L'étape de compilation :**

Il s'agit de combiner les résultats des m-bases analysées. Précurseur de la technique d'imputation multiple, Rubin a établi les règles pour combiner les résultats des m-bases analysées. L'estimateur global des paramètres d'intérêt est approché par la moyenne des estimateurs de ces paramètres pour chaque base complétée [105]. En d'autres mots, les paramètres sont estimés par la moyenne de leur valeur locale pour chaque base. La variance associée à cet estimateur global pour chaque paramètre d'intérêt est constituée de deux parties : une variance intra-imputation qui sera la moyenne des variances des différentes bases et une variance inter-imputation qui est donnée par la variance par rapport à l'estimateur global. Soit  $\widehat{\theta}_m$ , les estimateurs des paramètres d'intérêt issus de l'analyse pour chaque base m analysée ( $\widehat{\theta}_1, \widehat{\theta}_2, \dots, \widehat{\theta}_m$ ) et  $s_m$  leur variance associée. L'estimateur global  $\bar{\theta}$  et sa variance associée  $S_m$  sont donnés respectivement par :

$$\bar{\theta} = \frac{1}{m} \sum_{i=1}^m \widehat{\theta}_i. \quad S_m = \frac{1}{m} \sum_{i=1}^m s_m + F * \frac{1}{m-1} \sum_{i=1}^m (\widehat{\theta}_i - \bar{\theta})^2$$

Où  $F = (1+1/D)$  est un facteur de correction pour la non-infinité des imputations. On peut alors voir ici que plus le paramètre m, donc le nombre de bases générées est grand, plus les paramètres estimés seront précis. En pratique, ce sont 5 à 10 bases qui sont générées [105]. Les données manquantes étant imputées en utilisant l'information disponible sur les autres variables, l'imputation multiple est bien adaptée aux données MCAR et MAR. Toutefois, les études récentes l'étendent au cas MNAR et rapportent des résultats satisfaisants [112]. En effet, en prenant en compte l'incertitude due au processus d'imputation des données manquantes, l'imputation multiple peut être vue comme une technique qui ne fait que compléter la base de données pour que celle-ci puisse être analysée. De ce fait, même sous condition MNAR même si les incertitudes peuvent être importantes pour les observations avec des données manquantes,

les paramètres estimés resteront sans biais. Beaucoup d'auteurs proposent même d'ajouter des variables auxiliaires dans la phase d'imputation pour réduire l'incertitude en rendant l'hypothèse MAR plus probable [104,105]. Ces variables auxiliaires apporteraient de l'information sur le mécanisme de manque des données et ne seront pas utilisées lors de l'analyse des données. Elles seront impliquées dans le seul modèle d'imputation. Par ailleurs, l'identification de ces variables auxiliaires peut nécessiter une réelle connaissance de métier.

L'imputation multiple est donc une technique de choix et l'une des seules valables et sans biais, lorsque nous sommes face à des données manquantes MNAR. Au-delà de ça, la conservation de l'incertitude relative à la présence de valeurs manquantes est un atout sous hypothèse MCAR et MAR également. L'imputation multiple préserve de surcroît la distribution des données. Toutefois la complexité de sa mise en œuvre et les nombreux choix à faire (nombre de bases à générer, nombre de cycles pour l'imputation, le choix des modèles d'imputation pour chaque variable avec des données manquantes, les variables auxiliaires...) constituent un frein à son utilisation systématique.

#### **2.2.4 Les méthodes de traitement intégrées aux algorithmes d'analyse :**

Les méthodes de traitement intégrées aux algorithmes d'analyse sont une autre classe assez spéciale de méthode de gestion de données manquantes. Elles sont basées pour la majorité sur les méthodes d'analyse factorielle et nous allons nous intéresser principalement à ceux ayant trait à l'analyse en composantes principales (ACP). Il existe de nombreuses variantes d'algorithme pour réaliser cette analyse en présence de données manquantes. Nous allons nous intéresser aux deux principales couramment employées : l'algorithme NIPALS pour « Nonlinear Iterative Partial Least Squares » et l'algorithme d'ACP itérative basée sur un modèle algorithmique d'estimation maximisation (EM pour « Expectation-Maximization ») comme pour l'imputation des données par maximum de vraisemblance.

##### **- L'algorithme NIPALS :**

C'est une méthode itérative proposée pour la première fois pour réaliser une méthode de régression spécifique : celle sur les moindres carrés partiels (PLS). Le principe de la méthode consiste à réaliser des régressions partielles alternées de façon itérative pour estimer les différents paramètres du modèle PLS. Tous basés sur la notion de réduction de dimensions, L'ACP et la PLS ont beaucoup de points communs. L'algorithme NIPALS a alors été adapté pour déterminer les axes et les composantes principales de l'ACP. Le principe est itératif et

peut être décrit ainsi qu'il suit. Pour la recherche de chaque axe et composante, nous itérons jusqu'à la convergence les étapes suivantes :

- *Initialisation : choix d'une variable quelconque du jeu de données comme étant la composante principale ;*
- *Déterminer l'axe correspondant par régression sur la composante choisie ;*
- *Normalisation de l'axe ;*
- *Recalculer la composante principale sachant cet axe ;*
- *Répéter jusqu'à la convergence (stabilité de la composante principale).*

Les axes sont alors déterminés de façon successive après une étape de déflation. L'algorithme NIPALS permet ainsi de déterminer les axes et les composantes principales de l'ACP. C'est une version de cet algorithme qui a été adapté à l'imputation de données manquantes. L'idée proposée consiste à utiliser une matrice de poids pour sauter à chaque étape de l'algorithme les occurrences de données manquantes dans le processus de construction des axes et composantes de l'ACP. Ces derniers sont finalement utilisés pour reconstruire les valeurs manquantes du jeu de données de départ. En fait, l'objectif principal de l'algorithme NIPALS adapté aux données manquantes est de réaliser de l'ACP même en présence de données manquante et non l'imputation, mais il peut être employé à cette fin. Cette technique d'imputation a l'avantage d'être simple et efficace, mais il faut prêter attention aux différents aspects qui s'y rattachent, notamment le fait qu'in fine les valeurs de remplacement seront celles qui minimisent leurs erreurs de reconstruction des axes et composantes ayant servi à leur imputation. De là, le nombre d'axes choisis peut grandement influencer les valeurs imputées. L'utilisation de l'algorithme NIPALS adapté aux données manquantes est surtout intéressante pour réaliser des ACP ou PLS en présence de données manquantes comme ce fut le cas dans la présente thèse de doctorat.

- **L'algorithme d'ACP itérative : EM-ACP (estimation et maximisation – ACP) :**

Lorsque la quantité de données et/ou la proportion de données manquantes est très importante, plusieurs travaux rapportent l'instabilité de l'algorithme NIPALS adapté aux données manquantes, mais surtout des temps de calcul très coûteuses [104,105]. En fait, cela est dû à la nature de l'algorithme adapté. En présence de données manquantes, les calculs ne sont pas faits axe par axe seulement, mais individu par individu pour construire les axes et les composantes.

Comme alternative, une nouvelle méthode permettant de réaliser l'ACP en présence de données manquantes a été proposée. Celle-ci est basée sur une approche itérative grâce à

l'algorithme EM. L'idée consiste à imputer les données manquantes par des valeurs quelconque de départ et de réaliser des ACP sur ce jeu de données rendu complet. Puis, d'imputer les valeurs manquantes grâce à la formule de reconstruction de l'ACP pour un nombre d'axes et de composantes déterminé à l'avance. Les étapes d'estimation de l'ACP et d'imputation par reconstruction sont répétées jusqu'à la convergence. De cette façon, les axes, composantes et données manquantes sont estimés de façon simultanée. Comme pour NIPALS, le but principal de cette technique n'est pas d'imputer des non-réponses, mais de réaliser l'ACP en leur présence. Elle a les mêmes avantages et inconvénients de la technique avec NIPALS, notamment la contrainte du choix du nombre d'axes avant l'imputation.

### **3. Étude d'application des techniques de gestion de données manquantes :**

Dans cette section, nous allons procéder à l'évaluation pratique des différentes techniques détaillées dans la revue de littérature. Pour cela, l'approche globale est d'utiliser un jeu de données complet auquel nous allons retirer de façon aléatoire plusieurs proportions de son contenu puis appliquer les techniques de gestion de données manquantes les plus courantes pour évaluer leur performance de reconstruction et d'analyse. La performance de reconstruction sera évaluée par la comparaison de la somme des carrés des écarts par rapport aux données réelles pour chaque technique. La performance d'analyse sera évaluée en comparant les conclusions pouvant être tirées d'une analyse d'ACP du jeu complet à celles de l'analyse du jeu reconstitué.

#### **3.1 Matériels et méthodes :**

##### **3.1.1 Les jeux de données :**

Pour cette étude, nous utilisons une partie d'un jeu de données provenant d'une collecte industrielle dans le cadre de la présente thèse de doctorat. Le jeu est constitué de 47 individus qui sont des lots de compression indépendants. Sur chaque lot, nous avons neuf variables qui décrivent la dureté (SCU), l'épaisseur (mm) et le poids (g) moyens, maximums et minimums observés au cours de la compression du lot.

Pour les analyses, nous avons simulé à partir de ce jeu de données original, cinq nouveaux jeux de données appelées « dataset1 à 5 » ayant respectivement une proportion de 2 %, 5 %, 10 %, 20 % et 30 % de données qui ont été retirées aléatoirement. Le mécanisme de manque de ces jeux de données est donc MCAR et la quasi-totalité des techniques de gestion

des données manquantes devrait pouvoir s'appliquer. Pour générer les jeux de données avec des proportions de données manquantes aléatoirement, nous avons codé et utilisé une fonction sur python « Missrandomgen ».

### **3.1.2 Les techniques de gestion des données manquantes :**

Nous avons appliqué aux 5 jeux de données incomplètes, neuf différentes stratégies de gestion qui sont : l'analyse des cas complets, l'imputation par la moyenne, l'imputation par médiane, l'imputation par régression stochastique, l'imputation par forêts aléatoires, l'imputation par les k plus proches voisins, l'imputation par la méthode du maximum de vraisemblance, l'imputation à l'aide de NIPALS pour données manquantes, l'imputation par ACP itérative et l'imputation multiple. L'ensemble des fonctions pour exécuter ces stratégies ont été codés avec le langage de programmation python.

### **3.1.3 Visualisation des données :**

Avant de traiter les données, nous les affichons pour voir la répartition des données manquantes dans la structure du jeu. Nous avons réalisé ces différents graphiques qui seront présentés dans la section analyse à l'aide du package « missingno » de python (peut nécessiter une installation avant de pouvoir l'importer et l'utiliser). Nous avons réalisé un graphique pour le jeu de données complet et tous les jeux incomplets.

### **3.1.4 Évaluation de la performance de reconstruction :**

Pour évaluer la performance de reconstruction de chaque technique, nous avons calculé la somme des carrés des écarts entre la matrice de données complètes et chacune des matrices reconstruites par les différentes techniques. Pour les résultats de la régression multiple, nous avons calculé les sommes des carrées des écarts de chacune des m bases complétés et nous avons représenté les cinq valeurs. Étant donné que la méthode d'analyse des cas complets ne reconstruit pas la matrice originelle, nous l'excluons de cette partie de l'évaluation. Cependant, elle a été évaluée séparément puisque son principe consiste à supprimer des observations.

### **3.1.5 Évaluation de la performance d'analyse :**

Pour évaluer les performances d'analyse, nous avons comparé les sorties (graphique des scores et des loading) des deux premières composantes d'une analyse ACP entre le jeu de données complet et les données reconstituées par différentes techniques.

## 3.2 Résultats et discussions :

### 3.2.1 Visualisation des données :

Nous voyons sur la Figure 27 ci-dessous la répartition des données manquantes dans la structure des jeux de données utilisés.

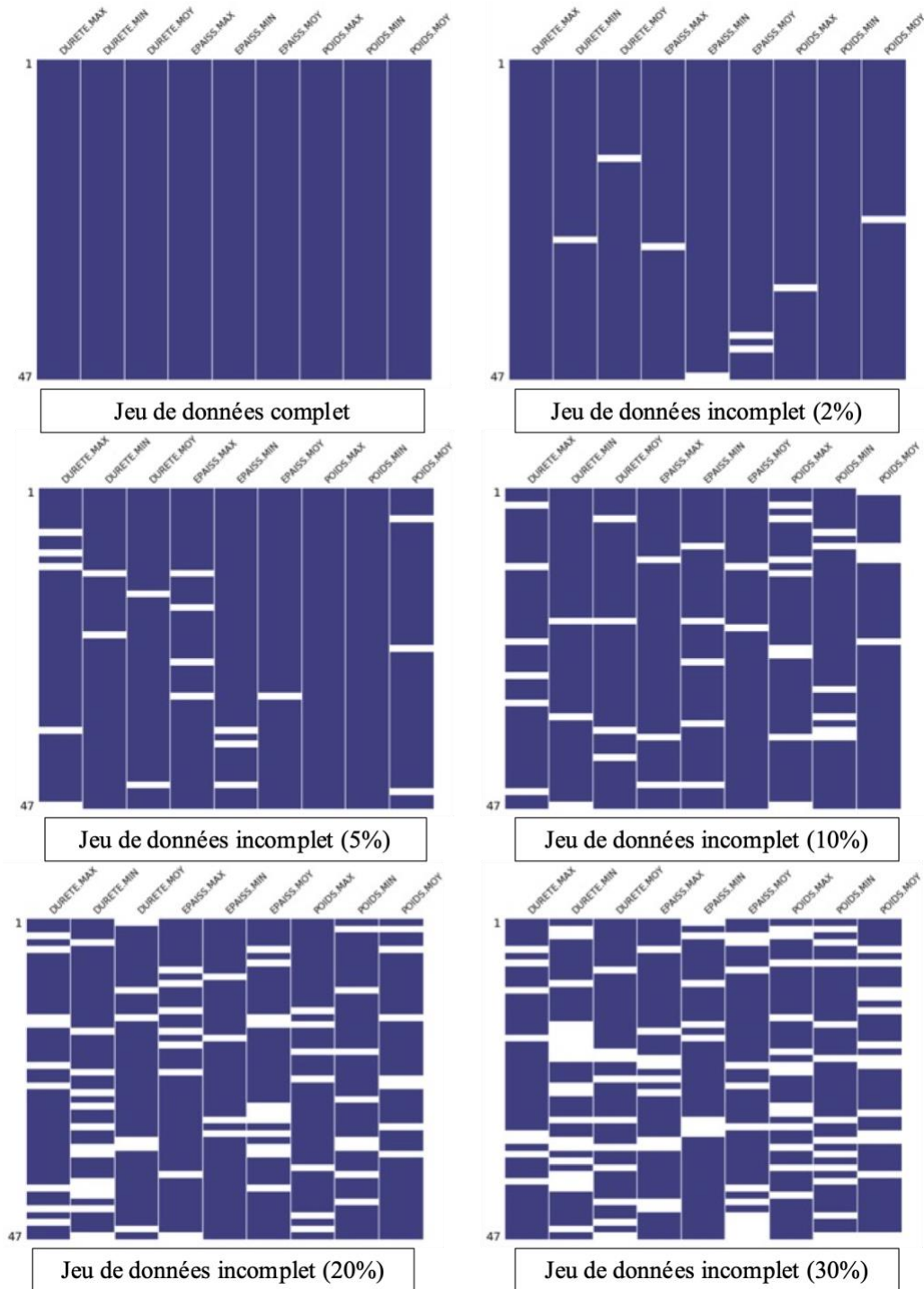


Figure 27 : carte de données manquantes des différents jeux de données utilisés.

### 3.2.2 Évaluation des performances de reconstruction :

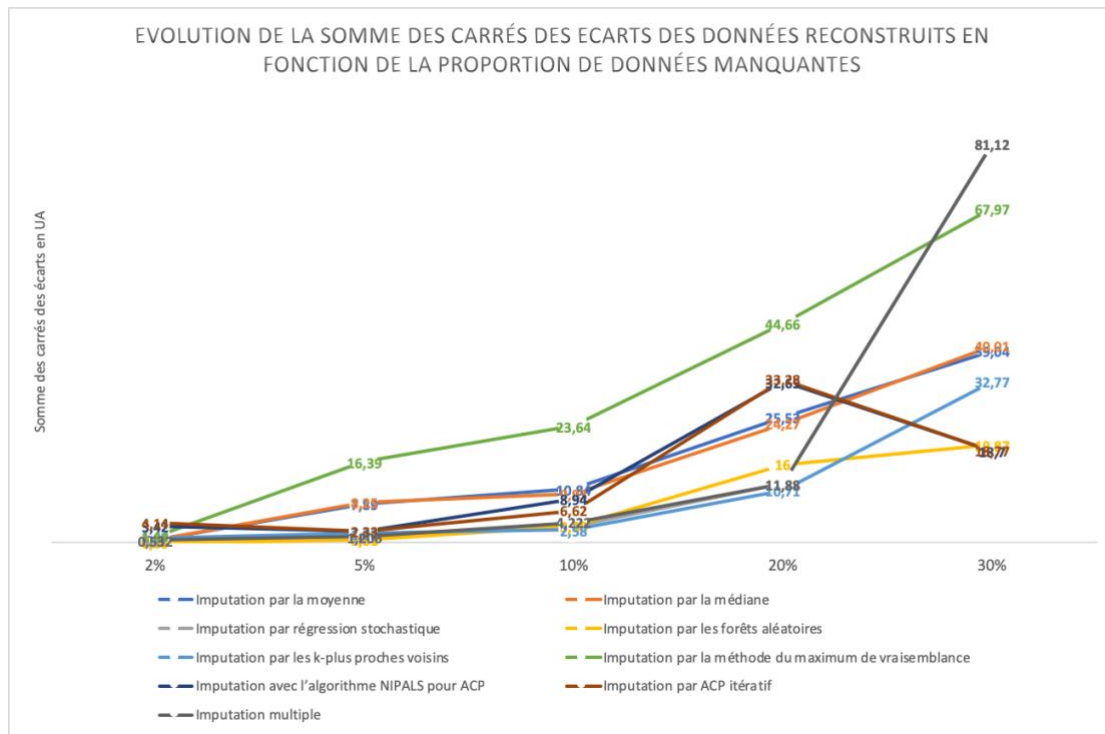
La Table 5 et la Figure 28 ci-dessous présentent les résultats de l'évaluation de la performance (somme des carrés des écarts) de reconstruction des données pour toutes les techniques de gestion excepté celle de l'analyse des cas complets.

*Table 5 : Résultats d'évaluation des performances à la reconstruction des données originales.*

*Résultats d'évaluation exprimés en unité arbitraire de la somme des carrés des écarts par rapport à la matrice originale complète.*

| Stratégies de gestion des données manquantes          | Proportion de données manquantes |       |       |       |       |
|---|----------------------------------|-------|-------|-------|-------|
|   | 2 %                              | 5 %   | 10 %  | 20 %  | 30 %  |
| Imputation par la moyenne                             | 0,60                             | 7,89  | 10,84 | 25,53 | 39,04 |
| Imputation par la médiane                             | 0,67                             | 8,25  | 9,93  | 24,27 | 40,01 |
| Imputation par régression stochastique                | 0,27                             | 1,22  | 3,48  | 11,88 | 81,12 |
| Imputation par les forêts aléatoires                  | 0,03                             | 0,61  | 3,38  | 16,00 | 19,87 |
| Imputation par les k-plus proches voisins             | 1,07                             | 1,92  | 2,58  | 10,71 | 32,77 |
| Imputation par la méthode du maximum de vraisemblance | 1,44                             | 16,39 | 23,64 | 44,66 | 67,97 |
| Imputation avec l'algorithme NIPALS pour ACP          | 3,42                             | 2,33  | 8,94  | 32,65 | 18,70 |
| Imputation par ACP itératif                           | 4,14                             | 2,33  | 6,62  | 33,28 | 18,77 |
| Imputation multiple m1                                | 0,06                             | 1,19  | 2,30  | 11,88 | 81,12 |
| Imputation multiple m2                                | 0,89                             | 1,09  | 3,54  | 11,88 | 81,12 |
| Imputation multiple m3                                | 0,65                             | 0,78  | 3,97  | 11,88 | 81,12 |
| Imputation multiple m4                                | 0,55                             | 1,84  | 5,93  | 11,88 | 81,12 |
| Imputation multiple m5                                | 0,51                             | 1,13  | 5,37  | 11,88 | 81,12 |





*Figure 28 : Résultats d'évaluation des performances à la reconstruction des données originales.*

*Résultats exprimés en unité arbitraire de la somme des carrés des écarts par rapport à la matrice originale complète.*

La tendance générale montre que la qualité de reconstruction diminue avec la proportion de données manquantes, ce qui était attendu. Ces résultats de simulation montrent que les techniques les plus performantes à l'imputation des données que nous avons aléatoirement omises sont l'imputation par les forêts aléatoires et la méthode des k plus proches voisins. Ces deux techniques sont suivies par la reconstruction avec les modèles d'analyse en composantes principales par l'algorithme NIPALS et d'ACP itérative, ainsi que les méthodes d'imputation par régression stochastique et l'imputation multiple. Ces résultats s'expliquent par le principe des différentes méthodes. En effet, les forêts aléatoires et les k plus proches voisins vont classifier les données avant de procéder et le jeu de données que nous avons testé se prête bien à ces classifications d'individus semblables. L'imputation avec l'algorithme NIPALS, l'ACP-itérative, la régression stochastique et l'imputation multiple sont aussi semblables par le principe de la reconstruction par régression. D'ailleurs, l'imputation multiple donne des résultats semblables à la régression stochastique par ce que nous n'avons spécifié qu'un seul modèle qui est également stochastique dans la fonction d'imputation multiple. L'imputation par le maximum de vraisemblance donne les plus mauvais résultats tandis que l'imputation par la moyenne et la médiane sont très proches.

### **3.2.3 Évaluation de l'analyse des cas complets :**

Pour évaluer les performances de l'analyse des cas complets, nous avons cherché à savoir le pourcentage d'observations supprimées à chaque proportion de données manquantes. Les résultats montrent qu'une proportion de 17 % d'observations est supprimée avec seulement 2 % de données manquantes tandis que plus de 90 % des observations sont supprimées dans le jeu de données avec 30 % de données manquantes.

### **3.2.4 Évaluation des performances à l'analyse :**

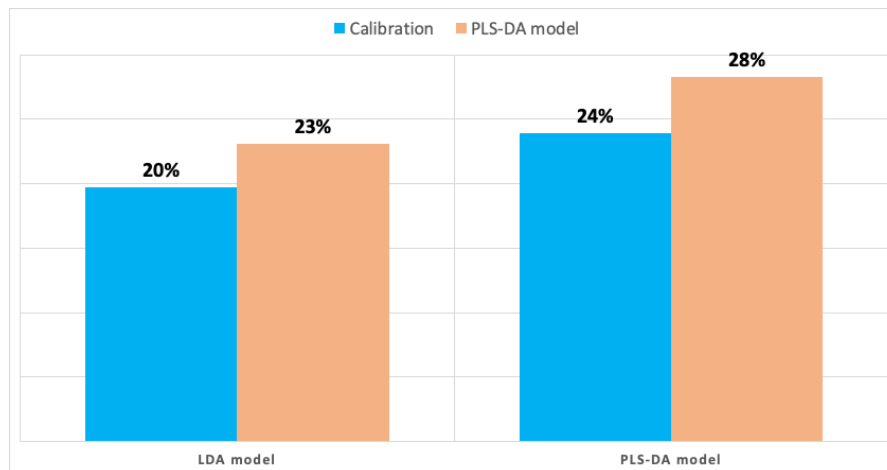
En analysant les sorties d'ACP pour tous les jeux de données reconstruits par rapport au jeu de données complet (résultats non présentés par souci de concision pour le présent document), nous avons noté que les techniques de gestion de données manquantes avec NIPALS et la méthode des forêts aléatoires arrivent à reconstruire à l'identique les deux premiers axes des composantes principales. Toutes les autres techniques de gestion des données manquantes induisent un biais plus ou moins important dans l'analyse. On remarquera toutefois qu'à par l'analyse des cas complets, les performances à l'analyse sont acceptables pour toutes les autres techniques évaluées.

## **4. Conclusion :**

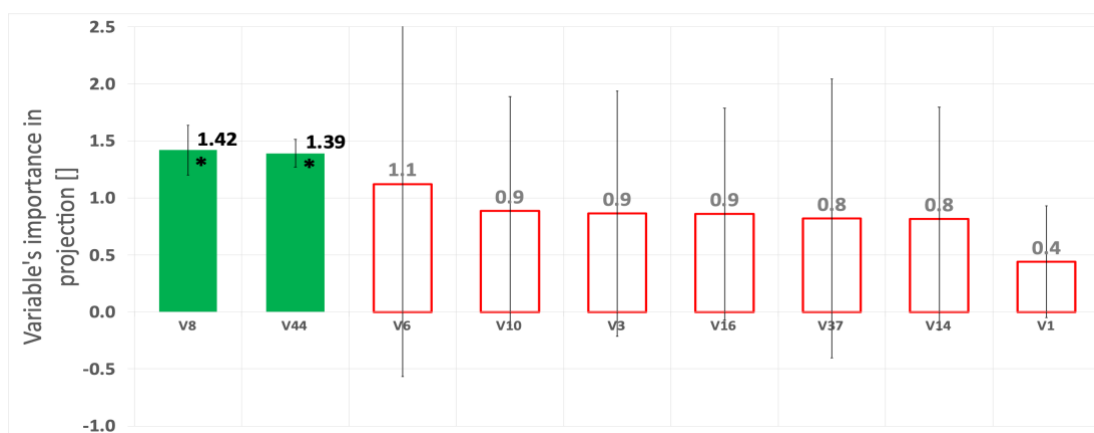
Avant de décider d'une méthode de traitement des données manquantes, il convient tout d'abord d'analyser leur mécanisme de manque qui peut être MCAR, MAR ou MNAR. Nous avons présenté la majorité des techniques de gestion des données manquantes en passant en revue leur principe, leur implémentation ainsi que les avantages et inconvénients associés. L'analyse des cas complets n'est utilisable que sous une hypothèse MCAR avec très peu de données manquantes, mais on leur préfère toujours des techniques d'imputation simple comme les forêts aléatoires, les k plus proches voisins, la régression stochastique qui donnent de très bons résultats sous hypothèse MAR. Lorsque les données manquent sur une grande base de données et que nous disposons d'informations pour analyser les mécanismes, une approche d'imputation multiple sera plus flexible. Nous avons aussi vu que des stratégies comme du NIPALS modifié permettant de prendre en charge les données manquantes produisaient des matrices reconstruites fidèles, mais surtout que l'analyse avec 30 % de données manquantes a donné des résultats satisfaisants. Cette stratégie a alors été privilégiée dans les cas où nous analysons des données via ACP ou PLS.

## Annexe 4 : Résumé comparatif des résultats du dataset1 avec LDA et PLS-DA

PLS-DA was also used to analyze dataset1 and similar results as per LDA was obtained. Below are the results.



Error rates of both LDA and PLS-DA models using dataset 1.



Variables importance plot for PLS-DA model

Both LDA and PLS-DA models provide similar results in terms of predictive and descriptive outcomes. LDA model has slightly better performance in classification of about 3-5%. This is probably due the difference in final set of variables selected throughout the two different optimization processes.