UNIVERSITÉ DE TOURS

ÉCOLE DOCTORALE : Santé, Sciences Biologiques et Chimie du Vivant

ÉQUIPES de RECHERCHE :

UMR U1253, iBrain, Université de Tours, Inserm, Tours, France

EA6295 Nanomédicament et Nanosondes, Université de Tours, Tours, France

THÈSE présentée par :

Aline STELLA

Soutenue le : **17 Mai 2022**

pour obtenir le grade de : **Docteur de l'université de Tours**

Discipline/ Spécialité : Sciences de la vie et de la Santé

Analyse multivariée des images hyperspectrales Raman : application à l'étude des dynamiques moléculaires de la barrière cutanée

THÈSE dirigée par :

Mr TAUBER Clovis	Maitre de conférences – HDR, UMR U1253, iBrain, Univ. de Tours
Mr BONNIER Franck	Maitre de conférences – HDR, EA6295 NMNS, Univ. de Tours

RAPPORTEURS :

Mr DOBIGEON Nicolas	Professeur, IRIT/INP-ENSEEIHT, Univ. de Toulouse
Mme LECOMTE Sophie	Directeur de recherche, Univ. de Bordeaux, CNRS/CBMN, UMR 5248

JURY :

Mr CARDOT Hubert	Professeur, LIFAT EA6300, Univ. de Tours, [Président]
Mr DOBIGEON Nicolas	Professeur, IRIT/INP-ENSEEIHT, Univ. de Toulouse, [Membre]
Mme LECOMTE Sophie	Directeur de recherche, Univ. de Bordeaux, CNRS/CBMN, UMR 5248, [Membre]
Mr GUT Yoann	Docteur, Laboratoires Servier, [Membre]
Mr TAUBER Clovis	Maitre de conférences – HDR, UMR U1253, iBrain, Univ. de Tours, [Membre]
Mr BONNIER Franck	Maitre de conférences – HDR, EA6295 NMNS, Univ. de Tours, [Membre]

MEMBRES INVITÉS :

Mme MUNNIER Emilie	Professeur, EA6295 NMNS, Univ. de Tours

TABLE DES MATIÈRES

REMERCIEMENTS 7
RÉSUMÉ 9
ABSTRACT 10
LISTE DES FIGURES 11
LISTE DES TABLEAUX 19
LISTE DES ABRÉVIATIONS 21

INTRODUCTION GÉNÉRALE 22
Contexte et objectifs de la thèse 22
Contributions 23
Structure du manuscrit 24

CHAPITRE I. L'INTÉRÊT DE L'UTILISATION DE LA SPECTROSCOPIE RAMAN DANS LA RECHERCHE COSMÉTIQUE 26
Introduction 28
1. La peau 28
1.1. Physiologie de la peau 28
1.1.1. L'hypoderme 29
1.1.2. Le derme 29
1.1.3. L'épiderme 29
1.1.3.1. La couche basale (*stratum germinativum*) 30
1.1.3.2. La couche épineuse (*stratum spinosum*) 30
1.1.3.3. La couche granuleuse (*stratum granulosum*) 30
1.1.3.4. La couche claire (*stratum lucidium*) 30
1.1.3.5. La couche cornée (*stratum corneum*) 30
1.2. Modèles de peaux humaines utilisés en recherche cosmétique 31
1.2.1. Modèle de peau humaine *ex vivo* 32
1.2.2. Modèle de peau humaine reconstruite *in vitro* 32
1.2.2.1. Modèles simulant la peau de pleine épaisseur 33
1.2.2.2. Modèles simulant l'épiderme humain 33
2. Techniques d'étude de la pénétration d'ingrédients cosmétiques actifs 34
2.1. Techniques de chimie analytique couplées aux cellules de Franz 35
2.2. L'Imagerie Hyperspectrale : outil d'analyse recherché par les industriels 36
2.3. Spectroscopie de diffusion Raman 38
2.3.1. Le principe de la spectroscopie Raman 38
2.3.1.1. L'effet Raman 38
2.3.1.2. Les modes vibrationnels 40
2.3.2. Micro-spectromètre Confocal Raman 42
2.3.2.1. Les constituants de l'appareillage 42
(i) Les sources lasers 42
(ii) Microscope confocal et résolution spatiale 43
(iii) Systèmes de filtres 44
(iv) Système dispersif, détecteur et résolution spectrale 44
2.3.2.2. Les spectres Raman 45
2.3.2.3. Les images Raman 46
2.3.3. Imagerie Raman de la peau humaine 47

Conclusion 49

CHAPITRE II. ÉTAT-DE-L'ART 51

Introduction 52

1. Analyse basée sur l'intensité des bandes Raman 52

1.1. Suivi d'intensité des bandes spectrales 52

1.2. Suivi des rapports de bandes spectrales 53

1.3. Conclusion 54

2. Méthodes de classification 55

2.1. K-means 56

2.1.1. Principe de la méthode des K-means 56

2.1.2. Interprétation des résultats de la méthode K-means 56

2.1.3. Exemples d'application de la méthode K-means sur des données spectrales Raman 57

2.1.4. Limites de la méthode K-means 58

2.2. Classification hiérarchique 58

2.2.1. Principe de la classification hiérarchique 58

2.2.2. Exemples d'application de la CHA 60

2.2.3. Limites de la CHA 60

2.3. Conclusion 60

3. Méthodes de démélange 61

3.1. Analyse en Composantes Principales (ACP) 61

3.1.1. Principe de l'ACP 61

3.1.2. Interprétation des résultats de l'ACP 63

3.1.3. Exemples d'application de l'ACP sur des données spectrales Raman 63

3.1.4. Limites de l'ACP 65

3.2. Régression linéaire au sens des moindres carrés (CLS) 65

3.2.1. Principe de la CLS 66

3.2.2. Exemples d'application de la CLS 67

3.2.3. Limites de la CLS 69

3.3. Conclusion 69

Conclusion 69

CHAPITRE III. ANALYSES COMPARATIVES DES MÉTHODES DE PRÉTRAITEMENTS DES DONNÉES RAMAN 71

Introduction 72

1. Réductions des dimensions : définition d'un sous-ensemble 72

2. Débruitage 74

2.1. Bruit gaussien 74

2.1.1. Filtre des moyennes glissantes 74

2.1.2. Algorithme Savitzky-Golay 75

2.1.3. Projection par ACP 77

2.1.4. Discussion 78

2.2. Bruit de fond : correction de la ligne de base 79

2.2.1. Méthode Lieber 81

2.2.2. Méthode Rubberband 82

2.2.3. Méthode par minimisation semi-quadratique 83

2.2.4. Discussion 86

2.3. Bruit des raies cosmiques 86

Étape 1 : Identification des spectres contaminés et localisation des raies cosmiques 87

Étape 2 : Correction des raies cosmiques 91

3. Calibration de la résolution spectrale et alignement spectral 92

4. Normalisation 94

4.1. Normalisation centrée-réduite 94
4.2. Normalisation Min-Max 94
4.3. Normalisation ℓp, $p = \{1, 2, \infty\}$ 95
5. Proposition d'une routine de prétraitement des données Raman 96
Conclusion 98

CHAPITRE IV. ANALYSE QUALITATIVE ET SEMI-QUANTITATIVE DE LA COMPOSITION SPECTRALE DES DONNÉES RAMAN 99
Introduction 101
1. Méthode de démélange 101
1.1. MCR-ALS 102
1.1.1. Résolution d'un problème d'optimisation 103
1.1.2. Estimations initiales 104
1.1.2.1. Le nombre de composants 104
1.1.2.2. Les matrices de spectres de composants purs ou de contribution relatives initiales 105
1.1.3. Contraintes 107
1.1.4. Évaluation du modèle estimé 107
1.1.4.1. Interprétabilité des spectres de composants purs 107
1.1.4.2. Analyse de la matrice résiduelle 108
1.1.4.3. Analyse des ambigüités potentielles 109
1.2. MVC-NMF 110
1.2.1. Calcul du volume du simplex 110
1.2.2. Résolution du problème d'optimisation 112
1.2.3. Initialisations, contraintes et évaluation de la qualité du modèle estimé 112
2. Application des méthodes MCR-ALS et MVC-NMF à des problématiques cosmétiques 113
2.1. Suivi du Delipidol® dans un modèle de peau reconstruite 113
2.1.1. Ensemble des données expérimentales 113
2.1.2. Prétraitement des données 114
2.1.3. Caractérisation spectrale du Delipidol® 115
2.1.4. Application de la MCR-ALS 115
2.1.4.1. MCR-ALS appliquée sans contrainte d'égalité spectrale sur le Delipidol® ($K = 4$) 115
2.1.4.2. Comparaison des nombres de composantes 117
2.1.4.3. Comparaison MCR-ALS avec et sans contrainte d'égalité spectrale sur le Delipidol® 121
2.1.5. Conclusion 122
2.2. Suivi de la caféine dans un modèle de peau reconstruite 123
2.2.1. Ensemble des données expérimentales 124
2.2.2. Prétraitement des données 124
2.2.3. Caractérisation spectrale de la caféine 125
2.2.4. Analyse de données 126
2.2.4.1. Estimation du nombre de composants 126
2.2.4.2. Application de la MCR-ALS 127
(i) Sans contrainte d'égalité spectrale sur la caféine 127
(ii) Avec contrainte d'égalité spectrale sur la caféine 131
(iii) Suivi de la caféine à l'aide des modèles avec $K = \{2, 3, 5\}$ et une contrainte d'égalité spectrale 135
2.2.4.3. Application de la MVC-NMF 137
(i) Sans contrainte d'égalité spectrale sur la caféine 137
(ii) Avec contrainte d'égalité spectrale sur la caféine 143
(iii) Suivi de la caféine à l'aide des modèles avec $K = 2, 3$ et une contrainte d'égalité spectrale 147
2.2.5. Conclusion 149
2.3. Étude des conditions de stockage des échantillons de peau reconstruite 150
2.3.1. Ensemble des données expérimentales 150

2.3.2. Prétraitements des données 151
2.3.3. Analyse des données 152
2.3.3.1. Estimation du nombre de composants 153
2.3.3.2. Application de la MCR-ALS avec une contrainte d'égalité spectrale sur le Delipidol® 154
2.3.3.3. Évaluation de la pertinence d'utilisation d'une contrainte d'égalité spectrale dans la MCR-ALS 159
2.3.4. Conclusion 162
Conclusion 162

CHAPITRE V. UNE VÉRITÉ DE TERRAIN POUR ÉVALUER L'EFFICACITÉ DES MÉTHODES DE DÉMÉLANGE 165
Introduction 166
1. Les films de polymères : un modèle de mélange simplifié 167
1.1. Principe des films de polymères 167
1.2. Formulations avec des films de polymères : un support d'évaluation de la capacité des méthodes d'analyse multivariée à démélanger des données spectrales Raman 168
2. Génération de données synthétiques 170
2.1. Génération de mélanges avec une diffusion homogène de l'actif 171
2.1.1. Les références spectrales 171
2.1.2. Processus de génération de données synthétiques 172
2.2. Comparaison des données synthétiques et des données expérimentales 173
3. Évaluation de l'efficacité des méthodes de démélange 175
3.1. Évaluation de la MCR-ALS 175
3.1.1. Mélanges PVA-PNM-DL expérimentaux et synthétiques 176
3.1.1.1. Sans contrainte d'égalité spectrale 176
3.1.1.2. Avec contrainte d'égalité spectrale sur PNM-DL 177
3.1.2. Mélanges PVA-GS-PNM-DL expérimentaux et synthétiques 180
3.1.2.1. Sans contrainte d'égalité spectrale 180
3.1.2.2. Avec contrainte d'égalité spectrale sur le PNM-DL 182
3.2. Évaluation de la MVC-NMF 184
3.2.1. Mélanges PVA-PNM-DL expérimentaux et synthétiques 184
3.2.1.1. Sans contrainte d'égalité spectrale 184
3.2.1.2. Avec contrainte d'égalité spectrale sur le PNM-DL 186
3.2.2. Mélanges PVA-GS-PNM-DL expérimentaux et synthétiques 188
3.2.2.1. Sans contrainte d'égalité spectrale 188
3.2.2.2. Avec contrainte d'égalité spectrale sur le PNM-DL 190
Conclusion 192

CHAPITRE VI. PERSPECTIVES D'AMÉLIORATION 193
Introduction 194
1. Amélioration des méthodes d'analyse multivariée adaptées aux problématiques cosmétiques 194
1.1. Évolution de la formulation du problème d'optimisation à résoudre 194
1.1.1. La sous/sur-décomposition spectrale 195
1.1.2. L'homogénéité du voisinage 196
1.1.3. La cohérence des spectres estimés avec les données analysées 197
1.1.4. Les perturbations spectrales 197
1.1.5. Changement de la fonction de coût dans le terme d'attache aux données 200
1.2. Choix de l'algorithme de minimisation à utiliser 202
1.3. Tests de la robustesse du nouvel algorithme 203
2. Création d'une interface Matlab 203

CONCLUSION GÉNÉRALE 205

LISTES DES PUBLICATIONS _____ **207**
Conférences nationales _____ **207**
Conférences internationales _____ **207**
Journaux internationaux _____ **207**

RÉFÉRENCES _____ **221**

REMERCIEMENTS

J'aimerais tout d'abord remercier chaleureusement mes directeurs de thèse, Dr Clovis Tauber et Dr Franck Bonnier, sans qui ces travaux de thèse n'auraient pu aboutir. Je suis ravie d'avoir pu travailler en votre compagnie et bénéficier de vos conseils et de vos savoirs. Merci pour votre disponibilité, votre confiance et votre soutien tout au long de ces travaux de thèse.

Je remercie Cosmétosciences et la région Centre Val de Loire pour avoir participé à la mise en place de ce projet de recherche et d'avoir financé ma thèse.

Je remercie bien évidemment Dr. Émilie Munnier d'avoir porté ce projet de recherche et assuré une bonne coordination entre les différents partenaires.

Je tiens à remercier la directrice de l'UMR Inserm U1253 iBrain, Mme Catherine Belzung et le responsable de l'équipe 3 « Imagerie, Biomarqueurs et Thérapie », Mr Patrick Emond, pour m'avoir accueillie au sein de l'unité.

Je tiens également à remercier le professeur et directeur de l'équipe EA 6295 Nanomédicaments et Nanosondes, Mr. Igor Chourpa, de m'avoir également accueillie au sein de son laboratoire.

Mes remerciements s'adressent au Pr. Hugh J.Byrne, directeur de l'Institut de Recherche Focas à Dublin pour sa participation à la correction de l'article de recherche publié.

Je remercie le Pr. Nicolas Dobigeon de l'Institut de Recherche en Informatique de Toulouse et le Dr. Sophie Lecomte, directeur de recherche CNRS de Bordeaux, d'avoir accepté d'être rapporteurs de mon travail de thèse. Je remercie également l'ensemble des personnes qui ont accepté de faire partie de mon jury de thèse.

Je tiens également à remercier les membres de l'équipe EA 6295 Nanomédicaments et Nanosondes, et notamment mes camarades Louise, Hichem et Kamilia qui ont réalisé de nombreuses acquisitions Raman pour pouvoir mener à bien ces travaux de thèse. Merci pour votre disponibilité et votre gentillesse.

Il m'est impossible d'oublier plusieurs personnes de l'équipe 3 de l'unité iBrain sans qui ces années de thèses n'auraient pas été aussi agréables. Merci à mes camarades thésardes, Camille, Cécile et Alexandra, pour votre bonne humeur et tous les bons moments partagés ensemble. Mention spéciale pour Bobo. Merci pour ta bonne humeur, ton humour et ton soutien tant sur le plan professionnel que personnel. Nos sessions CardioBike pour se défouler me manquent déjà. Je tiens également à remercier chaleureusement Zuhal, Julie, Sophie, Géraldine, Nadine, Claire sans qui les pauses déjeuners n'auraient pas été aussi fun. Merci pour votre gentillesse et votre bonne humeur quotidienne. Même si je ne comprenais pas toujours tout ce que vous faisiez dans vos manip' et inversement, j'ai partagé de très bons moments en votre compagnie. Je regrette déjà cette bonne ambiance quotidienne.

Enfin, je tiens à remercier mes proches pour le soutien indéniable qu'ils m'ont apporté.

En premier lieu, un grand merci à mes parents et mon frère pour le soutien infaillible et l'amour que vous m'avez porté pendant toutes ces années d'études. Je vous remercie d'avoir cru en moi et de m'avoir apporté toute votre aide quand j'en ai eu besoin. Ce manuscrit vous est dédié à 200%.

Merci à ma meilleure amie, Mélissa, d'avoir été présente dans les bons moments comme dans les plus durs, de m'avoir encouragée et soutenue depuis le début de nos années d'études supérieures.

Je tiens également à remercier ma belle-famille, qui malgré les milliers de kilomètres qui nous séparent, m'a soutenue et encouragée.

Pour terminer, je tiens tout particulièrement à remercier mon conjoint, Georges, qui a su me soutenir, me supporter, m'encourager… pendant toute la durée de ma thèse et plus particulièrement durant les derniers mois de rédaction qui n'ont pas toujours été des plus agréables. Merci pour ta patience, ta confiance, ta présence et ton amour.

RÉSUMÉ

L'imagerie confocale Raman permet une étude non destructive des cellules et tissus. Les images collectées associent à chaque pixel un spectre de valeurs qui constitue une empreinte moléculaire. Elles permettent de mettre en évidence et de suivre des modifications dans la composition chimique de la peau. Cependant, les images hyperspectrales sont des ensembles de données complexes dont la pleine exploitation nécessite l'application de méthodes statistiques robustes afin d'extraire une information pertinente. Dans cette thèse, des méthodes d'analyse multivariées sont présentées et évaluées pour exploiter les images vectorielles Raman. Le travail s'inscrit dans une démarche d'applicabilité avec i/ une étude comparative des prétraitements des données et ii/ l'application des méthodes présentées à des données réelles sur des problématiques importantes pour l'industrie du cosmétique. Les approches proposées sont en particulier appliquées pour établir les cinétiques de pénétration d'actifs cosmétiques dans la peau et pour étudier l'impact des conditions de stockage sur les modèles de peau. En parallèle, la création de données simulées réalistes vise à estimer la pertinence d'application de ces méthodes.

Mots clés : Imagerie hyperspectrale, Fouille de données, Traitement d'image, Analyses statistiques multivariées, Microspectroscopie confocale Raman.

ABSTRACT

Confocal Raman imaging allows non-destructive study of cells and tissues. The images collected associate with each pixel a spectrum of values that constitutes a molecular fingerprint. They allow to highlight and follow changes in the chemical composition of the skin. However, hyperpectral images are complex data sets whose full exploitation requires the application of robust statistical methods to extract relevant information. In this thesis, multivariate analysis methods are presented and evaluated to exploit Raman vector images. With the concern of the applicability of the proposed approaches we present i/ a comparative study of data preprocessing and ii/ the application of presented methods to real data on important issues for the cosmetic industry. In particular, the proposed approaches are used to establish the kinetics of penetration of active ingredients through the skin and to study the impact of storage conditions on skin models. A last part of the thesis is dedicated to the simulation of realistic Raman data to estimate the relevancy of the multivariate methods.

Keywords : Hyperspectral imaging, Data mining, Image processing, Multivariate statistical analysis, confocal Raman microspectroscopy.

LISTE DES FIGURES

Figure I-1 – Représentation de la structure de la peau (www.microbiologiemedicale.fr) 29

Figure I-2 – Schéma des différentes couches cellulaires de l'épiderme [4] 30

Figure I-3 – Plateau constitué de 12 puits contenant des cultures d'épiderme humain reconstruit Episkin® (www.episkin.com) 34

Figure I-4 – Cellules de Franz (www.bioganelys.com) 35

Figure I-5 – Estimation de la cinétique de perméation de la molécule de caféine à travers la peau par CLHP. Masse accumulée estimée en mg/cm². 35

Figure I-6 – Les différents phénomènes physiques lorsqu'un rayonnement incident rencontre une transition entre deux milieux de natures différentes : réflexion et/ou réfraction, diffusion et absorption [27]. 38

Figure I-7 – Diagramme de Jablonski et raies de diffusion 39

Figure I-8 – Phénomène de la diffusion de la lumière. 40

Figure I-9 – Modes de vibrations d'une molécule 41

Figure I-10 – Schéma simplifié d'un microspectromètre confocal Raman 42

Figure I-11 – Schéma simplifié du principe de confocalité [26] 44

Figure I-12 – Principe d'une image hyperspectrale Raman 47

Figure I-13 – Spectres Raman moyens d'une ligne de pixel situés dans la couche supérieure (courbe continue), la couche médiane (courbe discontinue), et la couche inférieure (courbe en pointillé) d'un épiderme humain reconstruit (Episkin®) 48

Figure II-1 – Détection des bandes associées au resvératrol et évolution de l'intensité intégrée de la bande située à 1003 cm^{-1} à 6, 27 et 73 µm de profondeur en fonction temps [36] 53

Figure II-2 – Profils de pénétration semi-quantitatifs de la caféine dans la peau de porc. Les profils sont obtenus en calculant les ratios d'intensités de bandes associées à la caféine (CAF, 556 cm^{-1}) et aux acides aminés (AA, 1004 cm^{-1}) ou à la bande amide I (AI, 1660 cm^{-1}) pour différents niveaux de profondeur [38]. 54

Figure II-3 – Distribution du diméthyl sulfoxyde (DMSO) dans des échantillons de couche cornée de paume de main. Rapport de bandes intégrées DMSO/protéines autour de 677 et 1450 cm^{-1} en fonction de la profondeur. (**A**) 15-45 min ; (**B**) 150-170 min ; (**C**) ∼4 h ; (**D**) ∼22 h ; (**E**) ∼48 h ; (**F**) ∼72 h après application [40] 55

Figure II-4 – (**A**) Image optique d'une section de tissu artificiel non exposée ; (**B**) Résultat de la méthode K-means de la carte Raman associée ; (**C**) Moyennes spectrales de tissu cutané artificiel des (**i**) cluster 4, stratum corneum, (**ii**) cluster 3, stratum granulosum, (**iii**) cluster 2, couche basale, (**iv**) cluster 5, derme [45]. 58

Figure II-5 – Illustration d'un dendogramme issu d'une CHA et d'une CHD [42] 59

Figure II-6 – (**A**) Nuage de points de l'ACP des spectres associés au stratum corneum de la section de tissu artificiel non exposée (rouge) et au stratum corneum de la section de tissu artificiel exposée pendant 240 minutes à des rayons UV (bleu). (**B**) Comparaison de la (**i**) CP1 par rapport au spectre Raman de (**ii**) L-α-phosphatidylcholine et (**iii**) de céramide. La ligne pointillée indique le point zéro dans le cas de la CP1 [45]. 65

Figure II-7 – Cartes de distribution reconstruites d'échantillons de peau exposés au Delipidol® après (**A**) 1 heure, (**B**) 2 heures et (**C**) 3 heures obtenues par analyse CLS avec une contrainte de non-négativité sur les contributions estimées ; Les barres d'échelles représentent la contribution relative du spectre Delipidol® dans les spectres Raman à chaque pixel. Comparaison avec les cartes reconstruites obtenues par analyse univariée de la forte bande caractéristique du Delipidol située à 1625 cm^{-1} après (**D**) 1 heure, (**E**) 2 heures et (**F**) 3 heures ; Les barres d'échelles représentent son intensité relative dans les spectres prétraités. 68

Figure III-1 – Spectres Raman bruts enregistrés à partir d'un modèle Episkin® sur les intervalles [400-3500]cm^{-1} et [400-1800]cm^{-1}. 73

Figure III-2 – Correction d'un spectre brut, enregistré à partir d'un modèle Episkin®, par la méthode des moyennes glissantes avec des demi-largeurs de fenêtre glissante différentes : $l = \{1, 2, 5, 7\}$. Les spectres sont décalés pour une meilleure visualisation. 75

Figure III-3 – Correction d'un spectre brut, enregistré à partir d'un modèle Episkin®, par l'algorithme de Savitzky-Golay avec des demi-largeurs de fenêtre glissante et des degrés polynomiaux différents. Les spectres sont décalés pour une meilleure visualisation. 76

Figure III-4 – Correction de spectres bruts (**A**), enregistrés à partir d'un modèle Episkin®, à l'aide d'une projection par ACP en utilisant (**B**) 10%, (**C**) 5%, (**D**) 1% et (**E**) 0.5% des premières composantes principales de l'ensemble de données (soit respectivement 38, 19, 4 et 2 composante sur 375 dans ce modèle) 78

Figure III-5 – Comparaison d'un spectre brut, enregistré à partir d'un modèle Episkin®, corrigé par la méthode des moyennes glissantes, la méthode de Savitzky-Golay et à l'aide d'une projection par ACP. 79

Figure III-6 – Spectre comportant du bruit de fond illustré par la courbe discontinue (en haut) et le résultat de sa correction par la méthode Rubberband (en bas). 80

Figure III-7 – Les différentes étapes de l'estimation de la ligne de base d'un spectre débruité, enregistré à partir d'un modèle Episkin®, par la méthode Lieber. L'ordre des polynômes estimés au sens des moindres carrés est $d = 4$ 81

Figure III-8 – Estimation de la ligne de base d'un spectre débruité, enregistré à partir d'un modèle Episkin®, par la méthode Lieber avec différents degrés polynomiaux $d = \{1, 2, 3, 4, 7, 10\}$. 82

Figure III-9 – Estimation de la ligne de base d'un spectre débruité, enregistré à partir d'un modèle Episkin®, par la méthode Rubberband. 83

Figure III-10 – Fonctions de coût non-quadratiques utilisées dans la méthode d'estimation de ligne de base par approximation polynomiale en réalisant une minimisation semi-quadratique. La fonction quadratique (ligne pleine) est illustrée pour comparaison des fonctions de coûts symétriques (**A**) et asymétriques (**B**) 84

Figure III-11 – Estimation de la ligne de base d'un spectre débruité, enregistré à partir d'un modèle Episkin®, par la méthode de minimisation semi-quadratique avec (**A**) un degré polynomial fixé à $d = 4$ et un seuil variant $t = \{0.001, 0.01, 0.1, 1, 10\}$ u.a et (**B**) un seuil fixé à $t = 0.01$ et un degré polynomial variant $d = \{1,2,3,4,7,10\}$. 85

Figure III-12 – Processus de réflexion d'une image Raman à deux coordonnées spatiales Nx et Ny et une coordonnée spectrale $N\lambda$ à l'aide d'une fenêtre de voisinage de taille $(wx, wy, w\lambda)$ où $wx = wy = w\lambda = 2l + 1$ (carré jaune centré sur le pixel jaune). 88

Figure III-13 – (**A**) Fenêtre de voisinage, où $l = 1$, centrée sur le pixel n°409 d'une image Raman obtenue à partir d'un échantillon d'épiderme humain reconstruit non traité et les pixels voisins; (**B**) Coefficients de corrélation de chaque pixel voisin avec le pixel central n°409. (**C**) Enregistrement du numéro du pixel voisin le plus fortement corrélé de chaque pixel. 89

Figure III-14 – (**A**) Spectre contaminé par un pic cosmique à 515 cm^{-1} (pixel n°409) ; (**B**) Spectre lissé par un filtre médian utilisant une fenêtre de voisinage avec $l = 3$; (**C**) Spectre corrigé de son pic cosmique par la méthode hybride ; (**D**) Profil des différences spectrales entre le spectre contaminé et son spectre voisin le plus corrélé (pixel n°446) ; $T = 365$: seuil défini de manière interactive en observant tous les profils des différences spectrales en même temps. Les spectres analysés sont issus d'un échantillon d'épiderme humain reconstruit (Episkin®) non traité. 90

Figure III-15 – Profils des différences spectrales d'une sélection de 3 spectres susceptibles de comporter un bruit cosmique avec leur spectre voisin le plus corrélé respectif. $T = 365$: seuil défini de manière interactive en observant tous les profils des différences spectrales en même temps. Les différences spectrales associées aux bandes spectrales situées à 1297 cm^{-1} et 1441 cm^{-1} sont supérieures au seuil T pour au moins deux des trois spectres, aucune raie cosmique n'est donc présente à ces nombres d'ondes. Les spectres analysés sont issus d'un échantillon d'épiderme humain reconstruit (Episkin®) non traité. 90

Figure III-16 – Illustration du processus de « wound healing » en 2D pour la correction d'une valeur aberrante [79] représentée par le pixel rouge. Les pixels jaunes représentent les valeurs voisines de la valeur centrale à corriger. À chaque étape, la médiane des pixels bleus situés autour d'un pixel vert est attribuée à ce dernier. Les pixels bleus ne sont affichés que pour certains pixels verts pour des raisons de clarté. 92

Figure III-17 – Comparaison d'un spectre contaminé par une raie cosmique à 1358 cm^{-1} (pixel n°511) avant (courbe noire) et après (courbe rouge) correction par la méthode hybride. Les courbes noires et rouges se superposent 92

Figure III-18 – Variations de l'espacement entre deux nombres d'ondes enregistrés d'un spectre issu d'une acquisition Raman d'un modèle Episkin®. 93

Figure III-19 – Translation d'un spectre enregistré à partir d'un modèle Episkin® de 3 cm^{-1} vers la droite. Alignement spectral sur la bande Raman située à 1004 cm^{-1}. 93

Figure III-20 – Normalisations d'un spectre débruité et corrigé de sa ligne de base, enregistré à partir d'un modèle Episkin® : (**A**) normalisation centrée-réduite,(**B**) normalisation Min-Max, normalisation ℓp avec (**C**) $p = 1$,(**D**) $p = 2$. 96

Figure III-21 – Schéma de la routine de prétraitements des images Raman proposée. Les étapes n°4, 6 et 7 sont appliquées au cas par cas. 97

Figure IV-1 – Modèle bilinéaire représenté (A) élément par élément, (B) composé par composé et (C) sous forme matricielle 102

Figure IV-2 – Restructuration d'une image Raman pour une analyse MCR-ALS 103

Figure IV-3 – Spectre résiduel moyen et son histogramme 109

Figure IV-4 – Représentation géométrique des données et des simplex formés par les composants purs estimés par les méthodes SIMPLISMA, MCR-ALS et MVC-NMF 111

Figure IV-5 – Comparaison du spectre moyen des images spectrales à chaque durée d'exposition de la peau au Delipidol® avec la référence spectrale du Delipidol®. (**A**) contrôle (peau non traitée) ; (**B**) 1 heure ; (**C**) 2 heures ; (**D**) 3 heures et (**E**) Delipidol® pur. Les spectres sont prétraités selon la stratégie décrite ci-dessus. 115

Figure IV-6 – Application de la MCR-ALS avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relative pour (**A**) $K = 4$ et (**B**, **C**) $K = \{3,8\}$, et (**D**) avec une contrainte d'égalité spectrale supplémentaire sur le Delipidol® pour $K = 4$, sur des données prétraitées selon la stratégie décrite ci-dessus ; (**A1**-**B1**-**C1**-**D1**) Spectres estimés des K composants initialisés par SIMPLISMA (décalés en intensité pour plus de clarté) ; (**A2**-**B2**-**C2**-**D2**) Cartes de distribution du Delipidol® dans les échantillons de peau exposés pendant 3 heures. Cartes obtenues

en estimant le spectre du Delipidol® $S1$ (**A2**-**B2**-**C2**) ou le spectre référentiel $S1$ (**D2**). SC, Stratum Corneum; EV, Épiderme Viable. [105] 116

Figure IV-7 – Cartes de distribution du Delipidol® après 3 heures d'exposition reconstruites à partir des composants purs $S2$, $S3$ et $S4$ obtenus à l'aide d'une MCR-ALS avec $K = 4$ composants et sans contrainte d'égalité spectrale. SC, Stratum Corneum; EV, Épiderme Viable. [105] 117

Figure IV-8 – Profils de pénétration moyens du Delipidol® pour 1, 2 et 3 heures de temps d'exposition obtenus par MCR-ALS appliquée avec (rouge) et sans (noir) contrainte d'égalité spectrale sur le Delipidol® pour $K = 4$. Les concentrations relatives moyennes obtenues par CLS avec contrainte de non-négativité sont illustrées par les points bleus. SC, Stratum Corneum; EV, Épiderme Viable. [105] 117

Figure IV-9 – (**A**) Variance expliquée cumulée obtenue par SVD appliquée sur les images Raman des échantillons exposés au Delipidol® ; (**B**) Critère de Durbin-Watson appliqué sur les résultats de SIMPLISMA ; (**C**) Indice KMO appliqué sur les résultats SIMPLISMA 118

Figure IV-10 – Histogrammes des résidus spectraux moyens obtenus par MCR-ALS appliquée (A-B-C) avec et (D-E-F) sans contrainte d'égalité spectrale sur le Delipidol® pour $K = \{3,4,8\}$ à partir des images Raman des échantillons de peau exposés pendant 1, 2 et 3 heures. 121

Figure IV-11 – Comparaison des spectres à chaque temps d'exposition aux deux formulations contenant la caféine avec la référence spectrale de la caféine. (**A**) contrôle ; (**B**) 8h, émulsion avec tensio-actifs ; (**C**) 8h, émulsion par Hydrofeeling® ; (**D**) 24h, émulsion avec tensio-actifs ; (**E**) 24h, émulsion par Hydrofeeling® ; (**F**) caféine. Les spectres sont prétraités selon la stratégie décrite ci-dessus. Les spectres noirs représentent la moyenne de tous les spectres représentés en gris. 125

Figure IV-12 – (**A**) Variance expliquée cumulée obtenue par SVD appliquée sur les images Raman des échantillons exposés à la caféine ; (**B**) Critère de Durbin-Watson appliqué sur les résultats de SIMPLISMA ; (**C**) Indice KMO appliqué sur les résultats SIMPLISMA 126

Figure IV-13 – Application de la MCR-ALS avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour (**A**) $K = 2$, (**B**) $K = 3$ et (**C**) $K = 5$ sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures ; (**A1**-**B1**-**C1**) Spectres estimés des K composants initialisés par SIMPLISMA (décalés en intensité pour plus de clarté) ; (**A2**-**B2**-**C2**) Histogrammes des résidus spectraux moyens des modèles obtenus avec les K composants estimés illustrés dans **A1**-**B1**-**C1** ; Les données sont prétraitées selon la stratégie décrite ci-dessus. 128

Figure IV-14 – Application de la MCR-ALS avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives et une contrainte d'égalité spectrale supplémentaire sur la caféine pour (**A**) $K = 2$, (**B**) $K = 3$ et (**C**) $K = 5$ sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures ; (**A1**-**B1**-**C1**) Spectres estimés des K composants initialisés par SIMPLISMA (décalés en intensité pour plus de clarté) ; (**A2**-**B2**-**C2**) Histogrammes des résidus spectraux moyens des modèles obtenus avec les K composants estimés illustrés dans **A1**-**B1**-**C1** ; Les données sont prétraitées selon la stratégie décrite ci-dessus. 132

Figure IV-15 – Cartes de distribution de la caféine formulée avec tensio-actifs (TA) et avec la technologie Hydrofeeling® (HF) après 8 et 24 heures d'exposition. Cartes reconstruites à partir des composants $S1$, obtenus à l'aide d'une MCR-ALS avec (**A**) $K = 2$, (**B**) $K = 3$ et (C) $K = 5$ composants et en utilisant une contrainte d'égalité spectrale sur la caféine. Les pixels noirs représentent le substrat et les axes x et y des cartes de distribution sont exprimées en micromètres. 136

Figure IV-16 – Profils de pénétration moyens de la caféine, formulée avec tensio-actifs (TA) et avec la technologie Hydrofeeling® (HF) après 8 et 24 heures d'exposition. Profils obtenus à partir des concentrations $Ci = 1, \dots, K$, obtenues à l'aide d'une MCR-ALS avec $K = 2$ (courbe pleine) et $K = 3$ (courbe pointillée) composants et en utilisant une contrainte d'égalité spectrale sur la caféine. 137

Figure IV-17 – Application de la MVC-NMF avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour (**A**) $K = 2,\ \tau = 0.5$, (**B**) $K = 3,\ \tau = 0.5$ et (**C**) $K = 5,\ \tau = \{0.25, 0.5, 0.75\}$ sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures; (**A1**-**B1**-**C1**) Spectres estimés des K composants initialisés par SIMPLISMA (décalés en intensité pour plus de clarté); (**A2**-**B2**-**C2**) Histogrammes des résidus spectraux moyens des modèles obtenus avec les K composants estimés illustrés dans **A1**-**B1**-**C1**; Les données sont prétraitées selon la stratégie décrite ci-dessus..... 138

Figure IV-18 – Application de la MVC-NMF avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives et une contrainte d'égalité spectrale supplémentaire sur la caféine pour (**A**) $K = 2,\ \tau = 0.5$, (**B**) $K = 3,\ \tau = 0.5$ et (**C**) $K = 5,\ \tau = \{0.25, 0.5, 0.75\}$ sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures ; (**A1**-**B1**-**C1**) Spectres estimés des K composants initialisés par SIMPLISMA (décalés en intensité pour plus de clarté) ; (**A2**-**B2**-**C2**) Histogrammes des résidus spectraux moyens des modèles obtenus avec les K composants estimés illustrés dans **A1**-**B1**-**C1** ; Les données sont prétraitées selon la stratégie décrite ci-dessus..... 144

Figure IV-19 – Cartes de distribution de la caféine formulée avec tensio-actifs (TA) et avec la technologie Hydrofeeling® (HF) après 8 et 24 heures d'exposition. Cartes reconstruites à partir des composants $Si = 1, \dots, K$, obtenus à l'aide d'une MVC-NMF avec (**A**) $K = 2, \tau = 0.5$ et (**B**) $K = 3, \tau = 0.5$ composants et en utilisant une contrainte d'égalité spectrale sur la caféine. Les axes x et y des cartes de distribution sont exprimées en micromètres. 148

Figure IV-20 – Profils de pénétration moyens de la caféine, formulée avec tensio-actifs (TA) et avec la technologie Hydrofeeling® (HF) après 8 (gauche) et 24 heures (droite) d'exposition. Profils obtenus à partir des concentrations $Ci = 1, \dots, K$, obtenues à l'aide d'une MVC-NMF avec $K = 3, \tau = 0.5$ composants et en utilisant une contrainte d'égalité spectrale sur la caféine (courbe continue). Les profils de pénétration obtenus précédemment par MCR-ALS avec $K = 2$ composants en utilisant une contrainte d'égalité spectrale sur la caféine sont également tracés pour comparaison (courbe discontinue). 149

Figure IV-21 – Comparaison des spectres enregistrés à partir d'échantillons contrôles (à gauche) et exposés au Delipidol® pendant 1 heure (au centre) et 3 heures (à droite) avec la référence spectrale du Delipidol® pour trois conditions de stockage : à température ambiante (rouge), à -20°C (violet) et à -80°C (bleu). Les spectres noirs représentent la moyenne de chaque image Raman..... 152

Figure IV-22 – (**A**) Variance expliquée cumulée obtenue par SVD appliquée sur les images Raman des échantillons conservées selon les 3 conditions de stockage et exposés aux Delipidol® ; (**B**) Critère de Durbin-Watson appliqué sur les résultats de SIMPLISMA ; (**C**) Indice KMO appliqué sur les résultats SIMPLISMA. 153

Figure IV-23 – Spectres de composants purs estimés par la MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le Delipidol® pour (**A**) $K = 4$ et (**B**) $K = 8$ sur les images Raman des échantillons de peau conservés selon les 3 conditions de stockage et exposés aux Delipidol® pendant 1 et 3 heures (décalés en intensité pour plus de clarté). Les spectres ont été initialisés par SIMPLISMA. 156

Figure IV-24 – Résidus des modèles estimés par la MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le Delipidol® pour $K = 4$ composants sur les images Raman des échantillons de peau conservés selon les 3 conditions de stockage et exposés aux Delipidol® pendant 1 et 3 heures ; (**A**) Histogramme du spectre résiduel moyen du modèle estimé; (**B**) Spectres résiduels de l'ensemble des images Raman (en gris) et spectre résiduel moyen du modèle estimé (en noir). 157

Figure IV-25 – Cartes de distribution du Delipidol® après 1 et 3 heures d'exposition sur des échantillons de peau conservés selon les 3 conditions de stockage : à température ambiante, à -20°C et à -80°C. Cartes reconstruites à partir du composant $S1$ contraint à être la signature du Delipidol® dans le

modèle obtenu par MCR-ALS appliquée avec une contrainte d'égalité spectrale et $K = 4$ composants. Les axes x et y des cartes de distribution sont exprimées en micromètres. 158

Figure IV-26 – Profils de pénétration moyens du Delipidol® après 1 (à gauche) et 3 heures (à droite) d'exposition dans des échantillons de peau conservés selon les 3 conditions de stockage : à température ambiante (ligne continue), à -20°C (ligne discontinue) et à -80°C (ligne pointillée). Profils obtenus à partir des concentrations relatives $C1$ estimées à l'aide d'une MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le Delipidol® avec $K = 4$ composants. 159

Figure IV-27 – Spectres de composants purs estimé par la MCR-ALS appliquée sans une contrainte d'égalité spectrale sur le Delipidol® pour (**A**) $K = 4$ et (**B**) $K = 8$ sur les images Raman des échantillons de peau conservés selon les 3 conditions de stockage et exposés aux Delipidol® pendant 1 et 3 heures (décalés en intensité pour plus de clarté). Les spectres ont été initialisés par SIMPLISMA. 160

Figure V-1 – Illustration de la formation d'un film contenant une solution (**A**) ou une dispersion de particules (**B**) [113] ; FFS pour Film formation of Film-forming Systems. 167

Figure V-2 – **A1**-**A2** : Images optiques et **B1**-**B2** : cartes de distribution du Delipidol® reconstruites après analyse NCLS des films de PVA. **B1** : Delipidol® libre ; **B2** : Delipidol® encapsulé. L'échelle de couleur représente la contribution relative du spectre de Delipidol® dans les spectres Raman à chaque pixel. Les images Raman sont de taille 25×50 µm² avec une résolution spatiale de 1 µm. 168

Figure V-3 Spectres Raman des composants de références utilisés dans la formulation des films simples et complexes. Les spectres sont volontairement décalés pour plus de lisibilité. PVA : Polyvinylic alcohol, GLY : Glycérol, SEPI : Sépicide™ HB, PNM-DL : Polymeric Nano-Micelles loaded with Delipidol®..... 171

Figure V-4 – Processus de génération de données synthétiques simulant les films PVA-PNM-DL et PVA-GS-PNM-DL ; $nPVA$, $nGLY$, $nSEPI$ et $nPNMDL$ les normes $\ell 2$ respectives des composants PVA, GLY, SEPI et PNMDL. 173

Figure V-5 – Comparaison des spectres des films (A) PVA-PNM-DL et (B) PVA-GS-PNM-DL obtenus expérimentalement (noir) et synthétiquement (rouge) avec une concentration progressive en PNM-DL de 1%, 2.5%, 5%, 7.5% et 10% (de haut en bas). 174

Figure V-6 – Ratios de bandes moyens du PNM-DL (1640 cm^{-1}) et du PVA (1440 cm^{-1}) pour les spectres expérimentaux (noir) et synthétiques (rouge) des films PVA-PNM-DL (courbes continues) et PVA-GS-PNM-DL (courbes discontinues). 174

Figure V-7 – (**A**) Coefficients de corrélation moyens et (**B**) erreurs quadratiques moyennes entre les spectres expérimentaux et synthétiques des films PVA-PNM-DL (courbe continue) et PVA-GS-PNMD-DL (courbe discontinue). 175

Figure V-8 – Décompositions spectrales par la méthode MCR-ALS appliquée avec uniquement des contraintes de non-négativité et de normalisation sur les spectres et les concentrations relatives pour $K = 2$ composants sur les images Raman des films PVA-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques. 176

Figure V-9 – Décompositions spectrales par la méthode MCR-ALS appliquée avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $K = 2$ composants et une contrainte d'égalité spectrale sur le PNM-DL sur les images Raman des films PVA-PNM-DL obtenues de manière (A) expérimentale et (B) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques. 178

Figure V-10 – Cartes de distribution du PNM-DL dans les images Raman de films PVA-PNM-DL (**A**) expérimentales et (**B**) synthétiques. Cartes reconstruites à partir du spectre $S1$ contraint à correspondre au spectre du PNM-DL dans l'application de la MCR-ALS avec $K = 2$ composants. Les axes x et y des cartes de distribution sont exprimées en pixels. 179

Figure V-11 – Comparaison des concentrations relatives moyennes du PNM-DL estimées par MCR-ALS sous contrainte d'égalité spectrale sur le PNM-DL à partir des données PVA-PNM-DL expérimentales et synthétiques ; EXP. : expérimentales, SYNT. : synthétiques. 179

Figure V-12 – Spectres résiduels (gris) et spectre résiduel moyen (noir) des décompositions spectrales obtenues par MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le PNM-DL sur les données PVA-PNM-DL expérimentales (haut) et synthétiques (bas). 179

Figure V-13 – Décompositions spectrales par la méthode MCR-ALS appliquée avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $K = 4$ composants sur les images Raman des films PVA-GS-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques. 180

Figure V-14 – Décompositions spectrales par la méthode MCR-ALS appliquée avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $K = 4$ composants et une contrainte d'égalité spectrale sur le PNM-DL sur les images Raman des films PVA-GS-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques. 183

Figure V-15 – Décompositions spectrales par la méthode MVC-NMF appliquée avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $\tau = 0.1$ et $K = 2$ composants sur les images Raman des films PVA-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques. 185

Figure V-16 – Décompositions spectrales par la méthode MVC-NMF appliquée avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $\tau = 0.1$ et $K = 2$ composants et une contrainte d'égalité spectrale sur le PNM-DL sur les images Raman des films PVA-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques. 186

Figure V-17 – Comparaison des concentrations relatives moyennes du PNM-DL estimées par MVC-NMF sous contrainte d'égalité spectrale sur le PNM-DL avec $\tau = 0.1$ à partir des données PVA-PNM-DL expérimentales et synthétiques ; EXP. : expérimentales, SYNT. : synthétiques. 188

Figure V-18 – Spectres résiduels (gris) et spectre résiduel moyen (noir) des décompositions spectrales obtenues par MVC-NMF appliquée avec une contrainte d'égalité spectrale sur le PNM-DL sur les données PVA-PNM-DL expérimentales (haut) et synthétiques (bas) pour $\tau = 0.1$. 188

Figure V-19 – Décompositions spectrales par la méthode MVC-NMF appliquée avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $\tau = 0.1$ et $K = 4$ composants sur les images Raman des films PVA-GS-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques. 189

Figure V-20 – Décompositions spectrales par la méthode MVC-NMF appliquée avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $\tau = 0.1$ et $K = 4$ composants et une contrainte d'égalité spectrale sur le PNM-DL sur les images Raman des films PVA-GS-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques. 191

Figure VI-1 – Représentation géométrique des données et du principe de « dictionnaire » de spectres de composants purs perturbés. 198

Figure VI-2 – Distribution du bruit présent dans les données Raman utilisées dans les études (**A**) du suivi du Delipidol® et (**B**) de la détection et du suivi de la caféine dans un épiderme humain reconstruit, puis (**C**) pour l'étude des conditions de stockage des échantillons de peau reconstruit in vitro ; Une loi

normale (courbe rouge) et une loi de Student (courbe bleue) sont ajustées à chaque histogramme ; Les histogrammes sont tracés à partir du spectre résiduel moyen de chaque image Raman..... 201

LISTE DES TABLEAUX

Tableau I-1 – Attribution des bandes de groupes fonctionnels en spectroscopie Raman [29]. υ : élongation ; δ : déformation symétrique dans le plan. 41

Tableau I-2 – Attribution des bandes Raman d'un échantillon d'épiderme humain reconstruit Episkin®. υ : élongation ; δ : déformation. 49

Tableau IV-1 – Concentrations relatives estimées du Delipidol® moyennes après 1, 2 et 3 heures d'exposition obtenues par MCR-ALS appliquée avec et sans contrainte d'égalité spectrale sur le Delipidol® ($K = \{3, 4, 8\}$). 119

Tableau IV-2 – Coefficients de corrélation entre les spectres estimés $Si = 1, \dots, 8$ obtenus par MCR-ALS appliquée avec et sans contrainte d'égalité spectrale sur le Delipidol® ($K = \{3, 4, 8\}$) sur les images Raman après 1,2 et 3h d'exposition. 120

Tableau IV-3 – Mesure des critères du manque d'ajustement LOF (Lack Of Fit en %), de variance expliquée R^2 (en %) et de l'ambiguïté rotationnelle des modèles estimés par MCR-ALS initialisée par SIMPLISMA avec $K = 2, 3, 5$ composants purs avec (c-CAF) et sans (nc-CAF) contrainte d'égalité spectrale sur la caféine. 128

Tableau IV-4 – Coefficients de corrélation (en noir) et indice de qualité HQI (en gris) entre les spectres estimés $Si = 1, \dots, 5$ obtenus par MCR-ALS appliquée avec (c-CAF) et sans (nc-CAF) contrainte d'égalité spectrale sur la caféine pour $K = \{2, 3, 5\}$ sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures. 130

Tableau IV-5 – Mesure des critères du manque d'ajustement LOF (Lack Of Fit en %), de variance expliquée R^2 (en %) et de l'ambiguïté rotationnelle des modèles estimés par MVC-NMF initialisée par SIMPLISMA avec $K = 2, 3, 5$ composants purs avec (c-CAF) et sans (nc-CAF) contrainte d'égalité spectrale sur la caféine. 139

Tableau IV-6 – Coefficients de corrélation (en noir) et indice de qualité HQI (en gris) entre les spectres estimés $Si = 1, \dots, 5$ obtenus par MVC-NMF appliquée avec (c-CAF) et sans (nc-CAF) contrainte d'égalité spectrale sur la caféine pour $K = \{2, 3, 5\}$ sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures. 141

Tableau IV-7 – Mesure des critères de manque d'ajustement LOF (Lack Of Fit en %), de variance expliquée R^2 (en %) et de l'ambiguïté rotationnelle des modèles estimés par MCR-ALS initialisée par SIMPLISMA en utilisant $K = \{4,8\}$ composants purs avec (c-DEL) et sans (nc-DEL) contrainte d'égalité spectrale sur le Delipidol®. 155

Tableau IV-8 – Coefficients de corrélation des spectres estimés $Si = 1, \dots, 8$ obtenus par MCR-ALS appliquée avec (c-DEL) et sans (nc-DEL) contrainte d'égalité spectrale sur le Delipidol® pour $K = \{4,8\}$ sur les images Raman des échantillons de peau conservées selon les 3 conditions de stockage et exposés aux Delipidol® pendant 1 et 3 heures avec le spectre de référence du Delipidol® et le spectre moyen des images Raman des peaux contrôles. 157

Tableau IV-9 – Concentrations relatives estimées du Delipidol® moyennes par niveau de profondeur après 1 et 3 heures d'exposition sur des échantillons conservés selon les 3 conditions de stockage (à température ambiante, à -20°C et à -80°C) obtenues par MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le Delipidol® et $K = 4$ composants. 158

Tableau IV-10 – Coefficients de corrélation entre les spectres estimés $Si = 1, \dots, 8$ obtenus par MCR-ALS appliquée avec (c-DEL) et sans (nc-DEL) contrainte d'égalité spectrale sur le Delipidol® pour $K = \{4,8\}$ sur les images Raman des échantillons de peau conservées selon les 3 conditions de stockage et exposés aux Delipidol® pendant 1 et 3 heures. 162

Tableau V-1 – Concentrations des composants utilisés dans la formulation de films de polymères « simples » avant et après évaporation du solvant..... 169

Tableau V-2 – Concentrations des composants utilisés dans la formulation de films de polymères « complexes » avant et après évaporation du solvant..... 170

Tableau V-3 – Coefficients de corrélation entre (**A**) les spectres estimés $Si = 1{,}2$ obtenus par MCR-ALS appliquée sans contrainte d'égalité spectrale sur les données PVA-PNM-DL expérimentales et synthétiques et (**B**) les spectres estimés $Si = 1{,}2$ et les références spectrales du PVA et du PNM-DL ; EXP. : expérimentales, SYNT. : synthétiques. 177

Tableau V-4 – Coefficients de corrélation entre (**A**) les spectres estimés $Si = 1{,}2$ obtenus par MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le PNM-DL sur les données PVA-PNM-DL expérimentales et synthétiques et (**B**) les spectres estimés $Si = 1{,}2$ et les références spectrales du PVA et du PNM-DL ; EXP. : expérimentales, SYNT. : synthétiques. 178

Tableau V-5 – Coefficients de corrélation entre les spectres estimés $Si = 1, \dots, 4$ obtenus par MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le PNM-DL sur les données PVA-GS-PNM-DL expérimentales et synthétiques avec les références spectrales du PVA, du PNM-DL, du GLY et du SEPI ; EXP. : expérimentales, SYNT. : synthétiques. 181

Tableau V-6 – Coefficients de corrélation entre les références spectrales du PVA, du PNM-DL, du GLY et du SEPI. 184

Tableau V-7 – Coefficients de corrélation entre (**A**) les spectres estimés $Si = 1{,}2$ obtenus par MVC-NMF appliquée sans contrainte d'égalité spectrale sur les données PVA-PNM-DL expérimentales et synthétiques et (**B**) les spectres estimés $Si = 1{,}2$ et les références spectrales du PVA et du PNM-DL ($\tau = 0.1$) ; EXP. : expérimentales, SYNT. : synthétiques. 185

Tableau V-8 – Coefficients de corrélation entre (**A**) les spectres estimés $Si = 1{,}2$ obtenus par MVC-NMF appliquée avec une contrainte d'égalité spectrale sur le PNM-DL sur les données PVA-PNM-DL expérimentales et synthétiques et (**B**) les spectres estimés $Si = 1{,}2$ et les références spectrales du PVA et du PNM-DL ($\tau = 0.1$); EXP. : expérimentales, SYNT. : synthétiques..... 186

Tableau V-9 – Coefficients de corrélation entre les spectres estimés $Si = 1, \dots, 4$ obtenus par MVC-NMF appliquée avec une contrainte d'égalité spectrale sur le PNM-DL sur les données PVA-GS-PNM-DL expérimentales et synthétiques avec les références spectrales du PVA, du PNM-DL, du GLY et du SEPI ; EXP. : expérimentales, SYNT. : synthétiques. 190

LISTE DES ABRÉVIATIONS

AC	Actif Cosmétique
ACP	Analyse en Composantes Principales
DDE	Derme dé-épidermisé
CHA	Classification Hiérarchique Ascendante
CLHP	Chromatographie Liquide à Haute Performance
CLS	Classic Least Square (régression linéaire au sens des moindres carrés)
CP	Composante principale
GLY	Glycérol
HF	Technologie HydroFeeling®
LOF	Manque d'ajustement ; « Lack Of Fit »
MCR-ALS	Multivariate Curve Resolution Alternated Least Square
MVC-NMF	Minimum Volume Constrained Non-negative Matrix Factorisation
NCLS	Non-negative Constrained Least Square
PVA	Alcool Polyvinylique
PVA-PNM-DL	Formulation de films de polymères composés de PVA et de Delipidol® encapsulé
PVA-GS-PNM-DL	Formulation de films de polymères composés de PVA, de glycérol, de Sépicide™ et de Delipidol® encapsulé
PNM-DL	Delipidol® encapsulé dans des nanoparticules ; « *Polymeric Nano-Micelles loaded with Delipidol® »*
RMSE	Erreur quadratique moyenne ; « *Root Mean Squared Error* »
SC	Stratum Corneum
SEPI	Sépicide™
SG	Savitzky-Golay
TA	Tensio-Actifs

INTRODUCTION GÉNÉRALE

CONTEXTE ET OBJECTIFS DE LA THÈSE

La Région Centre Val de Loire est fortement impliquée dans la recherche et le développement socio-économique régional de la filière cosmétique. Elle est même le lieu d'implantation initial de la Cosmetic Valley qui contribue activement au développement de la filière parfumerie-cosmétique en France et qui entreprend une démarche partenariale autour d'une dynamique d'innovation et de conquête des marchés internationaux. Dans un contexte partenarial avec la Cosmetic Valley, la Région Centre Val de Loire a encouragé le rapprochement de différents acteurs académiques et industriels du domaine cosmétique de la région en finançant le programme Ambition Recherche et Développement 2020 (ARD 2020) Cosmétosciences porté par l'université d'Orléans avec le soutien de l'Université de Tours, du CNRS et Le Studium. Ce programme collaboratif a plus précisément pour ambition de valoriser les partenariats entre académiques et industriels de la filière cosmétique de la région et de soutenir la recherche transversale, la formation et l'entreprenariat.

Le travail présenté dans ce manuscrit de thèse fait partie du projet MISTIC (Matériaux intelligents pour la libération stimulée de bioactifs cosmétiques) s'inscrivant dans le programme ARD 2020 Cosmétosciences. Ce projet regroupe six partenaires :

- l'EA6295 Nanomédicaments et Nanosondes (EA6295 NMNS, Université de Tours, Tours) et l'unité iBrain (UMR U1253, Université de Tours, INSERM, Tours),
- l'Institut de Chimie Organique et Analytique (ICOA UMR7311, CNRS et Université d'Orléans, Orléans),
- l'unité Interfaces, Confinement, Matériaux et Nanostructures (ICMN UMR7374, CNRS et Université d'Orléans, Orléans),
- la société BioEurope (Groupe Solabia, Anet)
- l'entreprise Transderma Systems (Tours).

Dans le besoin de répondre aux attentes du marché international mais aussi de cibler des molécules d'intérêt pour des stratégies innovantes de formulation cosmétique, le projet MISTIC a pour objectif principal de développer une nouvelle génération de produits cosmétiques permettant une libération stimulée des molécules actives dans les tissus. Une partie du projet consiste à concentrer et à protéger des bioactifs cosmétiques développés par l'entreprise Bioeurope dans des capsules ou des films de matériaux intelligents stimuli-sensibles. Une seconde partie est consacrée au développement de méthodes d'objectivation performantes et adaptées pour caractériser et évaluer biologiquement l'innocuité et l'efficacité des ingrédients et des nouvelles formes cosmétiques développées. Pour mener à bien cette appréciation, la proposition du consortium s'appuie sur le couplage de l'imagerie Raman hyperspectrale à des méthodes d'analyse mathématique multivariée.

La spectrosocpie vibrationnelle de diffusion Raman est une technique de caractérisation moléculaire sensible et non destructive qui collecte des informations à l'échelle micrométrique. Son utilisation présente donc un intérêt considérable pour des applications dans l'industrie cosmétique : les perturbations moléculaires engendrées par la libération d'actifs cosmétiques (AC) dans les tissus cutanés sont observables via la visualisation des empreintes moléculaires Raman locales enregistrées à partir des échantillons. Cependant, les données Raman hyperspectrales sont difficiles à analyser de par leur nombre et leurs expressions complexes. Elles nécessitent à ce jour la mise au point de méthodes d'analyse multivariée adaptées pour détecter et isoler les informations pertinentes et favoriser leur compréhension de façon plus aisée et rapide. Plus encore, les industriels et chercheurs

non experts du traitement de données hyperspectrales ont besoin d'outils d'analyse permettant d'étudier les cinétiques de pénétration des AC qui soient applicables sur des données réelles et simples d'utilisation afin d'évaluer l'innocuité et l'efficacité des formulations cosmétiques développées sans trop de difficulté.

Dans ce contexte scientifique, les objectifs principaux de ce travail de thèse sont de proposer des méthodes d'analyse multivariée pour estimer les profils de pénétration de molécule active dans la peau et de démontrer leur applicabilité en présentant des exemples d'études applicatives sur données réelles dans le cadre de thématiques cosmétiques ciblées. Cette notion d'applicabilité de ces méthodes est d'une importance considérable pour les industriels et est au cœur de nos préoccupations. En effet, cette proposition ne doit pas être uniquement basée sur des théories mathématiques. Ces méthodes doivent être accessibles aux industriels et applicables sur données réelles afin qu'ils puissent les utiliser sur leurs données rapidement, en autonomie et sans faire appel à des experts en traitement de données et qu'ils parviennent à tirer de leurs analyses des conclusions importantes et impactantes pour le secteur cosmétique. Dans cette démarche, un autre objectif de ce travail de thèse est de le présenter comme un guide pratique pour toute personne souhaitant réaliser ce type d'étude en donnant notamment des repères pratiques pour faciliter leur utilisation. Puis, un objectif supplémentaire de ces travaux de thèse est d'évaluer la pertinence des méthodes d'estimation proposées pour évaluer l'efficacité des produits. Cependant, les véritables cinétiques de pénétration des AC à travers les couches cutanées étant inaccessibles, cette appréciation des performances des méthodes proposées nécessite la simulation d'une vérité de terrain à partir de données générées synthétiquement qui soient réalistes.

CONTRIBUTIONS

Le travail de cette thèse a abouti à des contributions portant sur le prétraitement des données Raman hyperspectrales et leur analyse par des méthodes multivariées. Les éléments contributifs majeurs de ce travail sont les suivants :

- La réalisation d'une analyse comparative de plusieurs méthodes de prétraitement de données hyperspectrales régulièrement utilisées en spectroscopie Raman;
- La proposition d'une routine de prétraitement de données Raman hyperspectrales collectées à partir d'échantillons de peau humaine *in vitro* ;
- L'étude de l'adaptation et la validation de deux méthodes d'analyse multivariée sur quatre jeux de données Raman réelles qui sont collectées à partir d'échantillons de peau humaine reconstruite ou de films de polymères très utilisés comme matrice d'application d'actifs en cosmétique;
- Le suivi des cinétiques de pénétration de deux ingrédients cosmétiques actifs d'intérêts, tels que le Delipidol® et la caféine, à travers les différentes couches épidermiques à l'aide des deux méthodes d'analyse multivariée proposées ;
- Une étude applicative démontrant que le stockage des échantillons de peau humaine reconstruite *in vitro* à -80°C dégrade fortement la fonction barrière cutanée de l'épiderme et que le stockage à -20°C préserve mieux l'imperméabilité du tissu cutané ;
- La création d'une vérité de terrain à partir de données Raman collectées sur des films de polymères ;
- L'analyse des performances des deux méthodes d'analyse multivariée utilisées à l'aide de la vérité de terrain créée ;
- L'analyse des avantages et limitations des deux méthodes d'analyse multivariée, en particulier lorsqu'elles sont utilisées sur des données Raman hyperspectrales caractérisant une

composition moléculaire complexe pouvant être perturbée via l'introduction de nouvelles molécules ;

- La suggestion d'axes d'amélioration à apporter aux méthodes d'analyse multivariée afin de les rendre plus adaptées à l'analyse de données Raman pour la recherche cosmétique.

STRUCTURE DU MANUSCRIT

Chapitre I

La structure physiologique de la peau et les différents modèles de peau couramment utilisés pour des tests d'efficacité des produits cosmétiques y sont présentés. Ce chapitre présente ensuite l'intérêt de l'utilisation de l'imagerie de diffusion Raman sur des modèles de peau humaine reconstruite pour la recherche cosmétique. La complexité des informations de spectroscopie Raman y est soulignée.

Chapitre II

Ce chapitre constitue un état-de-l'art des méthodes d'analyse de données régulièrement appliquées sur des données Raman pour en extraire des informations pertinentes et faciliter leur compréhension. Une analyse des avantages et des limitations des méthodes de la littérature nous conduit à mettre en évidence l'intérêt que présentent les méthodes de démélange spectral pour suivre la progression d'un AC à travers les couches cutanées en comparaison aux traditionnelles méthodes d'analyse univariée ou de classification.

Chapitre III

Dans ce chapitre, nous réalisons une étude comparative des méthodes de prétraitement de données fréquemment utilisées en spectroscopie Raman afin de i/ identifier celles les plus adaptées à nos jeux de données et ii/ fournir des recommandations à la communauté. Nous proposons en particulier des étapes de prétraitement spectral supplémentaires pour rendre la correction des artefacts présents dans les données plus complet. Nous établissons un protocole de prétraitement global de données spectrales pour l'analyse et le suivi d'un principe actif à travers l'épiderme en imagerie Raman. L'analyse faite dans ce chapitre est un pré-requis à l'application des méthodes de démélange spectral proposées dans le Chapitre IV.

Chapitre IV

Deux méthodes d'analyse multivariée des images Raman sont présentées. Dans une première partie, nous présentons la théorie des deux méthodes provenant du domaine de l'analyse des données hyperspectrales, ainsi que leur adaptation au domaine de l'imagerie Raman. Nous accompagnons cette description de repères pratiques afin de faciliter leur application. La seconde partie est consacrée à l'application des méthodes présentées pour l'estimation des cinétiques de pénétration cutanée d'actifs cosmétique à partir d'images Raman. Trois études applicatives sur données Raman réelles sont décrites, par lesquelles nous mettons en évidence les avantages et les limitations des méthodes présentées pour la recherche cosmétique.

Chapitre V

Ce chapitre est consacré à la création d'une vérité de terrain pour évaluer la pertinence des décompositions spectrales estimées par les méthodes multivariées présentées dans le Chapitre IV. Des données Raman hyperspectrales enregistrées à partir de formulation de films de polymère sont utilisées comme point de départ pour générer de données Raman synthétiques réalistes. Nous appliquons les méthodes présentées dans le Chapitre IV sur les données générées synthétiquement et comparons les résultats obtenus avec la vérité de terrain créée. Cette analyse comparative permet de mettre en évidence les limitations des méthodes présentées pour le suivi d'AC et des perturbations spectrales induites par les interactions moléculaires qu'il occasionne.

Chapitre VI

Ce chapitre présente une liste de suggestions pour faire évoluer et adapter les méthodes de démélange spectral aux données spectrales enregistrées dans le cadre d'études de recherche cosmétique. Une reformulation du problème d'optimisation à résoudre y est proposée et une reconsidération de la méthode de résolution de ce nouveau problème y est suggérée. Pour finir, un projet de création d'application accessible au plus grand nombre est également mentionné pour pouvoir utiliser ces méthodes d'analyse de données facilement sur des données Raman dans un contexte d'étude cosmétique.

Chapitre I.
L'INTÉRÊT DE L'UTILISATION DE LA SPECTROSCOPIE RAMAN DANS LA RECHERCHE COSMÉTIQUE

Résumé
Après une présentation de la structure de la peau et des différents modèles de peau régulièrement utilisés dans les tests d'efficacité des produits cosmétiques, nous décrivons le principe de la spectroscopie Raman et son intérêt pour la compréhension de la composition moléculaire de la peau. Nous y illustrons la nécessité d'un traitement adapté des données Raman pour en extraire des informations pertinentes et faciliter leur compréhension.

Introduction 28
1. La peau 28
1.1. Physiologie de la peau 28
1.1.1. L'hypoderme 29
1.1.2. Le derme 29
1.1.3. L'épiderme 29
1.1.3.1. La couche basale (*stratum germinativum*) 30
1.1.3.2. La couche épineuse (*stratum spinosum*) 30
1.1.3.3. La couche granuleuse (*stratum granulosum*) 30
1.1.3.4. La couche claire (*stratum lucidium*) 30
1.1.3.5. La couche cornée (*stratum corneum*) 30
1.2. Modèles de peaux humaines utilisés en recherche cosmétique 31
1.2.1. Modèle de peau humaine *ex vivo* 32
1.2.2. Modèle de peau humaine reconstruite *in vitro* 32
1.2.2.1. Modèles simulant la peau de pleine épaisseur 33
1.2.2.2. Modèles simulant l'épiderme humain 33
2. Techniques d'étude de la pénétration d'ingrédients cosmétiques actifs 34
2.1. Techniques de chimie analytique couplées aux cellules de Franz 35
2.2. L'Imagerie Hyperspectrale : outil d'analyse recherché par les industriels 36
2.3. Spectroscopie de diffusion Raman 38
2.3.1. Le principe de la spectroscopie Raman 38
2.3.1.1. L'effet Raman 38
2.3.1.2. Les modes vibrationnels 40
2.3.2. Micro-spectromètre Confocal Raman 42
2.3.2.1. Les constituants de l'appareillage 42
(i) Les sources lasers 42
(ii) Microscope confocal et résolution spatiale 43
Objectif optique et résolution spatiale latérale 43
Mesure confocale 43
(iii) Systèmes de filtres 44
(iv) Système dispersif, détecteur et résolution spectrale 44
2.3.2.2. Les spectres Raman 45
2.3.2.3. Les images Raman 46

2.3.3. Imagerie Raman de la peau humaine _____ 47

INTRODUCTION

À ce jour, de nombreux modèles de peaux permettent d'accéder à une bonne compréhension de la structure, de la composition biochimique et de la physiologie de la peau. Cette compréhension est même primordiale pour mener à bien des études de cinétiques de pénétration de molécules exogènes à travers la peau. Ces modèles de peaux sont donc aussi utilisés pour réaliser des tests de validation de produits cosmétiques en amont des études *in vivo* pour s'affranchir du risque de toxicité et des contraintes éthiques. Cependant en Europe, l'utilisation des modèles de peaux d'origines animales en cosmétique est interdite depuis 2012. Pour leur part, les modèles de peaux d'origines humaines sont difficilement accessibles : provenant le plus souvent de la chirurgie plastique, la quantité de tissu disponible est limitée et les prix fixés par des sociétés spécialisées sont généralement très élevés. De plus, ces modèles de peaux souffrent d'une grande variabilité intra- et inter-individus et font l'objet d'une réglementation très pointilleuse concernant leur utilisation dans des études de recherche. Ces limitations ont alors renforcé le besoin de développer des modèles de peau *in vitro* qui soient à la fois reproductibles, facilement accessibles et moins onéreux et qui ne posent aucun problème d'éthique ou de variabilité. Le développement de modèles de peau humaine reconstruite *in vitro* a donc révolutionné de manière importante la recherche cosmétique et permet de réaliser aussi bien des tests d'innocuité que des test d'efficacité de produits cosmétiques.

De plus, les chercheurs ont recours à différentes méthodes analytiques pour étudier la pénétration cutanée de molécules exogènes dans la peau : des méthodes analytiques et depuis quelques années des techniques spectroscopiques (imagerie hyperspectrale). Contrairement aux premières, les secondes permettent un accès à des informations précises concernant les changements moléculaires induits et la distribution spatiale de la molécule dans la peau. Des techniques de spectroscopie vibrationnelle sont donc fréquemment employées sur les différents types de modèles de peau pour améliorer la compréhension de la composition et des processus physiologiques engagés suite à l'application de produits cosmétiques à l'échelle moléculaire.

Ce chapitre offre une vue d'ensemble sur le contexte général de ce projet de thèse. Une première partie est consacrée à la description de la composition de la peau et souligne l'importance de l'utilisation de modèles de peaux *in vitro* pour mener des études dans les domaines cosmétiques, pharmaceutiques et biologiques. Puis une seconde partie présente les différentes méthodes d'évaluation de la pénétration cutanée d'une molécule, dont celle de la spectroscopie Raman avec une description de son principe et des raisons pour lesquelles cette technique est utilisée pour suivre la progression de molécules exogènes à travers la peau

1. LA PEAU

La peau est le plus grand organe du corps humain, elle couvre une surface d'environ 2 m² chez l'adulte. Son épaisseur varie de 1 à 3 mm selon les différentes parties du corps et son poids représente 16% du poids corporel humain. Constituée de plusieurs couches de tissu, elle forme la première ligne de défense du corps contre l'environnement extérieur.

1.1. PHYSIOLOGIE DE LA PEAU

Trois couches de tissu composent la peau : l'hypoderme, le derme, et l'épiderme (Figure I-1). Chacune est elle-même constituée de sous-couches de tissus ayant des caractéristiques bien spécifiques.

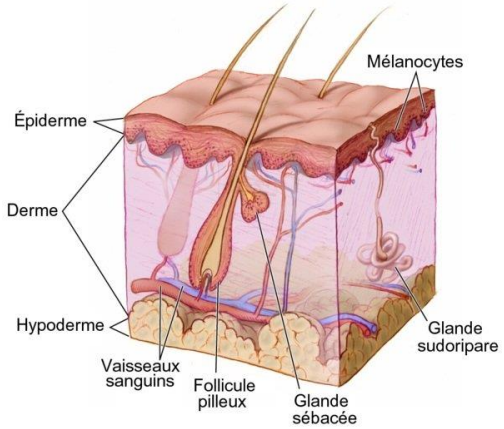


Figure I-1 – Représentation de la structure de la peau (www.microbiologiemedicale.fr)

1.1.1. L'hypoderme

L'hypoderme est la couche la plus profonde de la peau. Composé de tissu conjonctif lâche et de tissu adipeux plus ou moins épais selon la localisation, l'hypoderme est très vascularisé et innervé. Il forme une interface entre le derme et les cellules musculaires sous-jacentes. Il assure un rôle de protection des tissus profonds des chocs physiques extérieurs, d'isolant thermique et de réserve énergétique [1], [2].

1.1.2. Le derme

Le derme est une couche de tissu conjonctif fibreux et élastique se situant au-dessus de l'hypoderme. Il est composé de fibres de collagène à 70% et de fibres d'élastine dont le renouvellement est assuré par les fibroblastes qui sont des composants de la matrice extracellulaire [1]–[3]. Cette dernière est majoritairement composée par des glycosaminoglycanes tels que l'acide hyaluronique. Cette composition justifie des propriétés biomécaniques de la peau. Le derme, traversé par de nombreux vaisseaux sanguins et nerfs, contient également des glandes pilo-sébacées, des glandes sudoripares, des cellules adipeuses et des cellules de défenses immunitaires. Au final le derme, séparé de l'épiderme par une jonction dermo-épidermique, assure un rôle de support nutritionnel et structurel à l'épiderme [1]–[3].

1.1.3. L'épiderme

L'épiderme est la couche supérieure de la peau : il est en contact avec l'environnement extérieur. Il est constitué de quatre types de cellules : les kératinocytes représentant 80% des cellules, les mélanocytes, les cellules de Langerhans et les cellules de Merkel [1]–[3]. Ces différents types de cellules s'organisent en cinq couches cellulaires distinctes selon le processus de différentiation des kératinocytes: la couche basale (*stratum germinativum*), la couche épineuse (*stratum spinosum*), la couche granuleuse (*stratum granulosum*), la couche claire (*stratum lucidium*) et la couche cornée (*stratum corneum*). Ces différentes couches cellulaires sont illustrées dans la Figure I-2 ci-dessous.

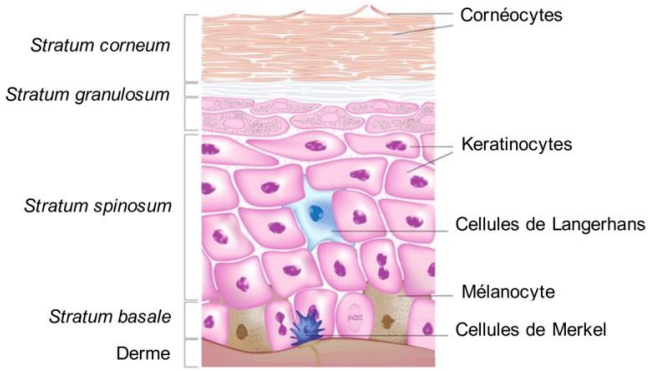


Figure I-2 – Schéma des différentes couches cellulaires de l'épiderme [4]

1.1.3.1. La couche basale (*stratum germinativum*)

La couche basale est la couche la plus profonde de l'épiderme. Les kératinocytes qui s'y produisent forment une couche monocellulaire et se multiplient par division cellulaire donnant naissance à deux cellules filles identiques : la première cellule migre vers la couche épineuse en débutant sa différentiation tandis que la deuxième cellule reste pour répéter cette division cellulaire [2]. Ce processus assure le renouvellement cellulaire de l'épiderme. Des mélanocytes sont aussi présents dans la couche basale. Ils permettent la synthétisation des mélanosomes présents dans le cytoplasme des kératinocytes en grains de mélanine pour que ces derniers migrent et se concentrent autour du noyau des kératinocytes. Ce mécanisme permet de protéger l'ADN des rayons ultraviolets [2]. En plus, des kératinocytes et des mélanocytes, la couche basale de l'épiderme est composée de cellules de Langerhans et de cellules de Merkel. Les cellules de Langerhans sont des cellules appartenant au système immunitaire, elles défendent l'organisme des agressions extérieures. Quant aux cellules de Merkel, ces cellules nerveuses jouent un rôle de récepteur sensoriel [1]–[3].

1.1.3.2. La couche épineuse (*stratum spinosum*)

La couche épineuse est située au-dessus de la couche basale. Elle est constituée de 5 à 6 couches de kératinocytes. Ces cellules sont de forme volumineuse et ont tendance à s'aplatir dans les régions les plus superficielles. La présence de nombreux desmosomes permet de lier les cellules entre elles créant ainsi une grande cohésion cellulaire responsable en partie de la très grande résistance mécanique de cette couche [1]–[3]. Des cellules de Langerhans sont également présentes dans la couche épineuse.

1.1.3.3. La couche granuleuse (*stratum granulosum*)

La couche granuleuse est la dernière couche de l'épiderme dont les cellules possèdent encore un noyau. Elle est composée de 3 couches de kératinocytes aplaties dont le noyau dense et ovale commence à se dégénérer. On distingue deux structures granulaires dans leur cytoplasme : les granulations de kératohyaline et les granulations de kératinosomes. Lors de leur migration vers la couche cornée, les grains de kératohyaline seront transformés en filaggrine, précurseur de la kératine, et les granules de kératinosomes libèreront des lipides qui formeront avec les desmosomes toujours nombreux un ciment intercellulaire entre les cellules mortes [1]–[3].

1.1.3.4. La couche claire (*stratum lucidium*)

La couche claire se situe au-dessus de la couche granuleuse. Elle est uniquement présente dans les peaux très épaisses [1]. Elle est composée de plusieurs couches de cellules sans noyau plates et claires avec un aspect homogène.

1.1.3.5. La couche cornée (*stratum corneum*)

La couche cornée, ou *stratum corneum* (SC), est la couche la plus superficielle de l'épiderme. Suivant sa localisation, elle est composée de 4 à 20 couches de cellules aplaties complètement kératinisées.

Ces cellules appelées cornéocytes proviennent de la différentiation des kératinocytes. Elles ne possèdent plus de noyau, ni de kératohyaline ni de kératinosomes et sont finalement remplies de kératine. La couche cornée se décomposent en deux sous-couches : la couche compacte (*stratum compactum*) et la couche desquamante (*stratum disjonctum*) [2].
La couche compacte fait suite à la couche granuleuse et est constituée de cornéocytes étroitement soudées par des cornéodesmosomes. Ces cornéocytes suivent un processus de différentiation au cours duquel une matrice fibreuse intracornécytaire est formée, des lipides intercellulaires sont produits, une enveloppe cornée apparait et la kératine est synthétisée. La matrice fibreuse intracornécytaire est créée au cours de la transition entre les kératinocytes granuleux et les cornéocytes, la production de la filaggrine permet l'agrégation de filaments de kératines. Les lipides contenus dans les kératinosomes de la couche granuleuse sont libérés dans l'espace intercellulaires et assurent la cohésion des cornéocytes entre eux. Cette cohésion intercellulaire participe à l'imperméabilité de la couche cornée et contribue à la régulation de la desquamation.
La couche desquamante est l'ultime couche de l'épiderme. Elle se situe au niveau de l'enveloppe cornée. Seuls les cornéocytes de surface desquament. Chaque cornéocyte desquamé est remplacé par un nouveau kératinocyte synthétisé dans la couche basale. Ainsi, l'épiderme se renouvelle de façon perpétuelle.
À noter que la fonction barrière de la peau est en grande partie assurée par l'organisation lipidique et la distribution des marqueurs de différentiation de la couche cornée.

1.2. MODÈLES DE PEAUX HUMAINES UTILISÉS EN RECHERCHE COSMÉTIQUE

La traversée du SC représente un premier défi pour les substances exogènes voulant atteindre des cellules cibles internes. En cosmétique, le choix des ingrédients actifs selon leurs caractéristiques physico-chimiques et du type de formulation est une étape très importante dans la création de nouveaux produits pour favoriser leur pénétration dans le SC. Une fois ces nouveaux produits développés, la clé du succès des industriels réside dans l'évaluation de leur efficacité, de leur pénétration cutanée et de leur innocuité. Seulement, pour mener à bien ces évaluations, l'utilisation de la peau humaine *in vivo* pose des problèmes éthique et économique [5]–[7]. En effet, le coût des essais sur l'Homme est élevé et l'application de substances ou matériaux ayant des effets potentiellement toxiques est au cœur des préoccupations [5]. De plus, les tests *in vivo* peuvent devenir difficiles à réaliser et à interpréter en raison d'une grande variabilité inter et intra-individuelle due à l'âge, l'hydratation, le site d'application ou l'état de santé [5].

Pour pallier à ces problèmes, le secteur pharmaceutique a utilisé de nombreux modèles de peau animale *ex vivo* comme alternatives à la peau humaine *in vivo* pour réaliser des tests de pénétration cutanée. Se comptent parmi eux la peau de porc, la peau de souris, la peau de rat ou encore la peau de serpent [5], [7], [8]. La peau de porc est le modèle de peau animale *ex vivo* le plus utilisé en raison de sa facilité d'accès mais surtout de sa grande similarité histologique avec la peau humaine. En effet, l'épaisseur de son SC, sa densité de follicules pileux dans la peau d'oreille ainsi que sa perméabilité sont comparables à ceux de la peau humaine [5]. Cependant, la transférabilité des résultats d'expérimentations animales à des problématiques humaines reste un défi de par les différences d'anatomie et de processus physiologiques [5], [9]. De plus, des préoccupations éthiques persistent: l'Union Européenne interdit depuis 2009 l'utilisation d'animaux pour la collecte de données toxicologiques pour les ingrédients cosmétiques [6].

De ce fait, l'utilisation de la peau humaine *in vivo* reste considérée comme la méthode de référence pour étudier la pénétration cutanée de formulations cosmétiques et constitue un élément marketing

pour encourager leur commercialisation. Cependant, pour éviter à des sujets d'être exposés à des risques toxicologiques mais aussi pour limiter le coût des expérimentations, les études de pénétration *in vivo* ne sont réalisées qu'en dernière étape de la démarche à suivre pour la mise sur le marché du produit cosmétique testé [10]. En effet, d'autres alternatives à l'utilisation de la peau humaine *in vivo*, à savoir la peau humaine *ex vivo* [8] et les modèles de peau artificielle [5] ou reconstruite *in vitro* [6], [9], sont couramment utilisées en amont pour évaluer la pénétration cutanée et l'innocuité des formulations développées. En plus de permettre le développement et l'optimisation des formulations cosmétiques sans danger, ces nouvelles alternatives offrent la possibilité de réaliser des tests à plus grande échelle.

1.2.1. Modèle de peau humaine *ex vivo*

Les modèles de peau humaine *ex vivo* sont obtenus à partir de chirurgie plastique ou sur des cadavres. Les principaux sites de prélèvement sont l'abdomen, la poitrine et le dos [5], [6], [8]. La durée de vie des peaux humaines est limitée après leur excision. Néanmoins, elles peuvent être conservées à -20°C. Bien que la congélation des tissus excisés engendre quelle que soit la température une dégradation de la viabilité cellulaire et de l'activité métabolique, le stockage à -80°C entraine une plus importante perte de la fonction barrière du SC que la congélation à -20°C [5], [11]. En effet, l'ajout de conservateurs tels que le glycérol ou le polyéthylène glycol avant le stockage à -20°C permet de limiter la perte des propriétés de barrière cutanée [5], [12]. Au final, l'utilisation des modèles de peau humaine *ex vivo* préalablement congelés à -20°C reste appropriée pour tester la perméation passive des produits cosmétiques, lorsque la viabilité de la peau et l'activité métabolique ne sont pas étudiées [5]. D'ailleurs, les modèles de peau humaine *ex vivo* sont très utilisés pour réaliser des tests de pénétration cutanée. Ils sont même considérés comme références absolues pour les études de pénétration [9].

En revanche, ces modèles ne peuvent pas être utilisés pour réaliser des tests de toxicité, d'irritabilité ou encore de corrosivité en raison de la perte d'activité métabolique et de la viabilité cellulaire. De plus, comme pour les tests sur la peau humaine *in vivo*, le problème de la grande variabilité entre donneurs complique l'analyse et l'interprétation des études de pénétration cutanée [8]. D'autre part, leur disponibilité est limitée et des préoccupations éthiques et économiques demeurent encore présentes [6], [9]. En effet, l'Union Européenne interdit le gain financier grâce à l'utilisation de tissu humain [6], [13], [14]. De nombreux documents faisant autorité stipulent que le corps humain et ses parties ne peuvent être un objet de commercialisation ou une source de profit. Les Articles 21 et 22 de *La Convention du Conseil de l'Europe pour la protection des droits de l'Homme et de la dignité de l'être humain à l'égard des applications de la biologie et de la médecine : Convention sur les droits de l'Homme et la biomédecine* indiquent précisément que « le corps humain et ses parties ne doivent pas être, en tant que tels, source de profit » et que « Lorsqu'une partie du corps humain a été prélevée au cours d'une intervention, elle ne peut être conservée et utilisée dans un but autre que celui pour lequel elle a été prélevée que conformément aux procédures d'information et de consentement appropriées » [14].

1.2.2. Modèle de peau humaine reconstruite *in vitro*

Depuis les années 1990, l'ingénierie *in vitro* de la peau humaine a commencé à se développer pour un accès plus rapide et une meilleure reproductibilité des méthodes d'analyses. L'instauration de nouvelles lois et réglementations durant ces dernières années concernant l'utilisation des modèles animaux et humains ainsi que les exigences des organisations de protection des animaux, des consommateurs et des scientifiques n'ont fait qu'encourager le développement de ces modèles de peaux *in vitro* innovants adaptés aux différentes expérimentations. Entre autres, l'évaluation de l'innocuité des produits cosmétiques nécessite de réaliser des tests sur des tissus vivants en toute

sécurité. Par conséquent, jusqu'à présent, un des défis consiste à proposer des modèles de peau humaine en survie pour pouvoir réaliser ces analyses sans danger et à grande échelle. De plus, ces modèles peuvent être un moyen d'améliorer et d'étendre les connaissances sur les processus biologiques de la peau [15]. Aujourd'hui, de nombreux modèles de peau humaine *in vitro* sont commercialisés, et peuvent être classés en deux catégories [7]:

1.2.2.1. Modèles simulant la peau de pleine épaisseur

Les modèles simulant la peau de pleine épaisseur sont constitués de deux couches : l'épiderme et le derme. Deux méthodes différentes sont utilisées pour créer ces modèles [7]. La première méthode consiste à utiliser un derme dé-épidermisé (DDE) issu d'une peau humaine excisée. Le DDE est ensuite peuplé de fibroblastes et cultivé en conditions immergées avant l'ensemencement de kératinocytes à sa surface. Le processus de différentiation est ensuite induit par l'exposition du DDE à l'air [7]. La seconde méthode consiste, quant à elle, à incorporer des fibroblastes dans un gel de collagène pouvant être considéré comme la partie dermique du modèle de peau. Des kératinocytes sont ensuite ensemencés à la surface de la matrice de collagène et la culture est maintenue en immersion pendant quelques jours afin d'engager la prolifération des kératinocytes. Pour finir, le processus de différentiation est induit par l'exposition de la culture à une interface air-liquide [7], [9].

Ces modèles sont très utilisés pour étudier les processus physiologiques mais aussi pour analyser l'absorption percutanée pour l'évaluation des risques tels que la photo-toxicité, la corrosivité ou encore l'irritabilité [9]. En effet, ils peuvent également être contraints à simuler des situations d'expérimentations particulières telles que l'infection, la présence d'une plaie ou encore de mélanomes [9]. Malgré une altération de la barrière cutanée les rendant plus perméables, les modèles de peau *in vitro* de pleine épaisseur sont employés de préférence lorsque la métabolisation du composé appliqué est affecté par le derme ou les interactions des fibroblastes dermiques/épidermiques [9].

1.2.2.2. Modèles simulant l'épiderme humain

Les modèles d'épiderme humain reconstruit étaient à l'origine utilisés pour des greffes de peau. Leur fabrication consiste à ensemencer des kératinocytes humains à la surface d'une membrane appropriée telle que l'acétate de cellulose, le polycarbonate ou encore une matrice de collagène [6], [7]. Après mise en culture immergée pour induire leur prolifération, les kératinocytes recouvrant la surface de la membrane sont exposés à l'air pour engager leur différentiation [7]. Un épiderme multicouche est ensuite obtenu après plusieurs semaines de différenciation. L'arrêt de la différentiation à différents stades de maturité permet d'obtenir des modèles adaptés à des études d'irritabilité ou de pénétration percutanée (Figure I-3) [6]. À noter aussi que des mélanocytes peuvent être introduits dans certains modèles et que le milieu de culture aqueux est supplémenté en vitamines, sels minéraux et hormones afin de réguler l'expression des marqueurs de différentiation.

Les modèles d'épiderme humain reconstruit sont les plus utilisés dans les tests de pénétration cutanée et d'innocuité. De nombreuses études comparent des modèles commercialisés, principalement les modèles Episkin®, Epiderm®, et SkinEthic®, en terme de morphologie, d'expression de marqueurs de différentiation, de composition et d'organisation lipidiques et de capacité à prédire la perméabilité des composés appliqués [5]–[7], [9], [15]–[20]. Toutes s'accordent à dire que ces modèles sont certes plus perméables que la peau humaine *ex vivo* mais restent des outils pertinents dans l'analyse de la photo-toxicité, de la corrosivité, de l'irritabilité ou de l'absorption percutanée. Plus précisément, l'altération de la barrière cutanée induisant une plus grande perméabilité de ces modèles ne constitue pas une limite à leur utilisation. En effet, les études de pénétration cutanée se portent généralement sur des comparaisons de formules cosmétiques différentes et les ordonnent selon leur capacité à pénétrer le tissu. Les résultats de ces études sont en réalité transposables à ceux obtenus lorsque les formules

cosmétiques sont utilisées sur des peaux humaines *ex vivo* [5]–[7]. D'autre part, les modèles d'épiderme humain reconstruit sont histologiquement similaires à l'épiderme humain *in vivo* et *ex vivo*. Les principaux marqueurs de différentiation sont présents et leur composition lipidique est d'un point de vue qualitatif comparable à celle de la peau humaine *in vivo* et *ex vivo*. Toutefois, l'organisation des lipides des modèles d'épiderme humain reconstruit *in vitro* diffère quantitativement de celle de la peau humaine *in vivo* et *ex vivo* et la fonction de desquamation est absente. L'augmentation constante de l'épaisseur de leur SC ne présente néanmoins en rien une limite à leur utilisation pour des études de pénétration cutanée puisque les processus industriels utilisés pour le développement de ces modèles permettent de maitriser leur stade de maturité.

Au final, les modèles d'épiderme humain reconstruit *in vitro* présentent plusieurs avantages :

- La conservation de leur viabilité cellulaire, autrement dit leur maintien en survie, permet de les utiliser pour étudier aussi bien la pénétration cutanée que l'innocuité des produits cosmétiques développés.
- Leur reproductibilité est assurée par la standardisation du processus de leur fabrication. En effet, en comparaison à la grande variabilité des peaux humaines *in vivo* ou *ex vivo* en fonction de leurs origines, la variabilité présente au sein et entre les modèles d'épiderme humain reconstruit *in vitro* est beaucoup plus faible.
- L'industrialisation de leur fabrication rend possible l'analyse de formulations cosmétiques à grande échelle.



Figure I-3 – Plateau constitué de 12 puits contenant des cultures d'épiderme humain reconstruit Episkin® (www.episkin.com)

2. TECHNIQUES D'ÉTUDE DE LA PÉNÉTRATION D'INGRÉDIENTS COSMÉTIQUES ACTIFS

Le développement de nouvelles formulations cosmétiques pour une action ciblée dans la peau nécessite une connaissance précise des mécanismes de diffusion des molécules qui les composent et de leur cinétique de pénétration à travers la peau. La diffusion des molécules à travers des tissus cutanés s'évalue traditionnellement à l'aide de techniques de chimie analytique. Parmi les techniques conventionnelles pour évaluer la perméation et le métabolisme de molécules actives appliquées localement sur la peau, les plus répandues sont la Chromatographie Liquide à Haute Performance (CLHP) et le « tape-stripping » couplées aux expérimentations sur des cellules de diffusion de Franz [21]–[23]. Cependant aujourd'hui, l'utilisation de la spectroscopie Raman devient de plus en plus répandue dans le monde de la recherche cosmétique pour étudier les perturbations moléculaires induites par l'introduction de molécules exogènes dans la peau. Cette partie est donc consacrée à la présentation des méthodes de chimie analytique couramment utilisées et de la spectroscopie Raman pour suivre la progression d'un AC à travers la peau.

2.1. TECHNIQUES DE CHIMIE ANALYTIQUE COUPLÉES AUX CELLULES DE FRANZ

Le principe des cellules de Franz repose sur l'application d'une molécule à la surface de la peau qui est située au milieu de deux compartiments, l'un receveur et l'autre accepteur, la face interne de la peau étant en contact avec le liquide récepteur (Figure I-4) [21], [23].

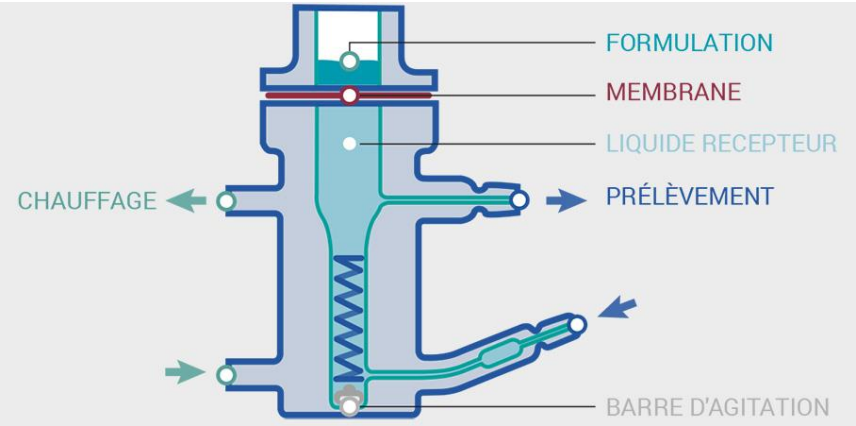


Figure I-4 – Cellules de Franz (www.bioganelys.com)

La diffusion de la molécule à travers la peau est ensuite évaluée en mesurant sa concentration dans le liquide récepteur en fonction du temps [21]. L'analyse de prélèvements à intervalles réguliers se fait le plus souvent par CLHP (Chromatographie Liquide à Haute Performance), le gold standard des techniques de chimie analytique pour ce type d'expérience. À partir de cette quantification, des informations sur la cinétique de diffusion et sur les coefficients de perméation des molécules peuvent être déterminées (Figure I-5). Les informations liées à la distribution spatiale des molécules à travers le tissu restent toutefois inaccessibles avec cette analyse [21], [22].

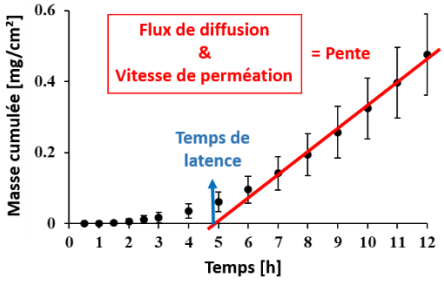


Figure I-5 – Estimation de la cinétique de perméation de la molécule de caféine à travers la peau par CLHP. Masse accumulée estimée en mg/cm².

L'utilisation de la CLHP se limite à l'étude des molécules traversant toutes les couches de la peau. En effet en pratique, alors que de petites molécules hydrophiles peuvent diffuser librement dans la peau et atteindre le liquide récepteur, d'autres molécules de par leur nature lipophile n'atteignent pas le liquide récepteur. En cosmétique, de nombreux actifs sont lipophiles et ont donc tendance à s'accumuler dans le SC et l'épiderme. Dans ce cas de figure, l'analyse du contenu du liquide récepteur par CLHP n'a aucun intérêt et l'utilisation de la technique du tape-stripping est alors privilégiée.

Concernant le couplage du tape-stripping avec l'utilisation des cellules de Franz, la mise en place de l'expérience reste inchangée mais la diffusion de la molécule à travers la peau est cette fois-ci évaluée en mesurant sa concentration à un temps arrêté en fonction de la profondeur cutanée. Plus précisément, la technique du tape-stripping consiste à retirer la couche cornée à l'aide de rubans adhésifs de manière successive [21], [22]. À savoir que l'utilisation du tape-stripping est uniquement

indiquée pour la séparation du SC et que la séparation de l'épiderme et du derme s'effectue par une autre méthode thermique. Les substances obtenues sur les bandes adhésives sont ensuite extraites avec un solvant de type méthanol puis analysées par CLHP. Les résultats sont généralement présentés sous forme de moyennes par profondeur et par couches : SC, épiderme et derme. En théorie, les bandes adhésives permettent de retirer des couches avec une précision de l'ordre du micromètre. Cependant bien que prometteuse, la technique du tape-stripping montre un manque de reproductibilité évident. En effet, la maitrise de l'épaisseur du SC retiré avec les différentes bandes adhésives dépend du type de peau mais aussi principalement de la pression et de la vitesse d'arrachage exercée par l'opérateur. De ce fait, le tape-stripping est aujourd'hui questionné par les instances internationales à cause de ce manque de fiabilité.

Au final, ces techniques chimiques analytiques sont certes fréquemment utilisées pour étudier la pénétration cutanée de principes actifs mais elles présentent de nombreuses limitations. Elles nécessitent un temps d'exécution important ainsi qu'une utilisation de consommables dont le coût financier est à terme non-négligeable. Leur emploi à très grande échelle est donc difficilement envisageable. De plus, alors que l'utilisation du tape-stripping dégrade l'intégrité des échantillons, l'emploi de la CLHP n'est pas transposable à des études cutanées *in vivo*. D'autre part, en plus d'être non résolues spatialement [21], [22], ces méthodes ne permettent pas non plus d'accéder aux informations concernant les perturbations induites telles que des changements de compositions moléculaires au sein du tissu cutané suite à l'application d'un principe actif à la surface de la peau : seule l'information sur la quantité d'actifs traversant l'épiderme ou présent dans le SC est récoltée. Par conséquent, un questionnement concernant la fiabilité des informations recueillies par ces méthodes subsiste puisque leur utilisation peut être source d'erreurs.

2.2. L'IMAGERIE HYPERSPECTRALE : OUTIL D'ANALYSE RECHERCHÉ PAR LES INDUSTRIELS

À l'heure actuelle, les industriels du monde cosmétique cherchent à obtenir le plus d'informations possibles sur les effets que peuvent engendrer des produits cosmétiques dans la peau. Leur demande ne cesse d'augmenter et ils exigent de multiples critères bien précis. Ils veulent des outils capables d'évaluer autant la pénétration que les effets induits par les produits cosmétiques dans le tissu cutané. Ces outils doivent être non invasifs, c'est-à-dire qu'ils ne doivent pas dégrader l'intégrité des échantillons. La préparation des échantillons doit être simple et le temps d'analyse rapide afin de pouvoir utiliser ces techniques d'analyse pour des études à grandes échelles. Et enfin, la technique d'analyse utilisée doit être transférable à des études *in vivo*.

Les industriels ont alors commencé à s'intéresser à un outil d'analyse respectant la plupart du temps tous ces critères : l'analyse par imagerie hyperspectrale. En effet, l'imagerie hyperspectrale combine la force de deux outils analytiques : la spectroscopie et la microscopie. De nombreuses techniques d'imagerie hyperspectrale donnent la possibilité d'identifier *in situ* les substances chimiques réparties au sein d'un échantillon sans avoir à recourir à un marquage spécifique [24]. Plus précisément, grâce à l'automatisation du traitement, leur utilisation ne nécessite que très peu d'étapes et d'intervenants. En effet, pour un bon nombre d'entre elles, seule la réalisation d'une coupe fine de l'échantillon est requise et aucun marquage spécifique préalable n'est nécessaire. De plus, la validation des observations peut se faire par un seul spécialiste via la visualisation d'images reconstruites en fausses couleurs obtenues suite à une interprétation de l'information biochimique contenue dans les spectres de l'image [24]. Celles-ci représentent la morphologie du matériau étudié et permettent une comparaison facile des résultats obtenus selon différentes conditions d'expérimentations. Ces cartes de distribution reconstruites en fausses couleurs sont donc un moyen de représentation très pertinent

et à fort impact dans le secteur industriel, et notamment en marketing. Par ailleurs, l'utilisation de l'imagerie hyperspectrale a également pour intérêt de nécessiter qu'un seul appareil d'imagerie pour pouvoir analyser de grands nombres d'échantillons.

Au final, l'utilisation de l'imagerie hyperspectrale a permis de réaliser des avancées majeures notamment dans le domaine de la biologie. Elle a été utilisée pour étudier la morphologie, les changements en réponse à un stress ou encore l'évolution vers un état pathologique ou un retour à l'état normal de tissus, d'organes ou de cellules biologiques tout en s'affranchissant de la nécessité d'utiliser des agents de marquage [24]. De même, dans le domaine pharmaceutique, elle a permis d'obtenir des informations immédiatement compréhensibles pour la mise au point et le suivi de la stabilité d'une formulation, ou bien pour la qualification et/ou le contrôle qualité d'un produit en observant la répartition et l'homogénéité de ses constituants. Ainsi, elle est devenue un outil d'identification de matières premières, un outil d'évaluation de comportement défectueux au sein d'un procédé de fabrication ou encore un outil de détection de falsifications produits [24]. L'imagerie hyperspectrale est donc toute assignée à devenir un outil d'analyse prometteur pour l'étude de la pénétration et des perturbations moléculaires induites suite à l'application d'un AC à la surface de la peau.

De nombreuses techniques d'imagerie hyperspectrales basées sur des techniques spectroscopiques existent et sont régulièrement utilisées dans les domaines médical, biologique et pharmaceutique. Certaines ne permettent néanmoins pas d'accéder à des informations concernant la composition moléculaire de l'échantillon :

- La spectrométrie de masse : elle permet de détecter et d'identifier des molécules d'intérêts par mesure de leur masse et de caractériser leur structure chimique et donc de suivre la quantité de molécules médicamenteuses et celle de métabolites dans l'organisme à un instant donné. Cette technique d'imagerie est particulièrement utilisée dans la recherche de biomarqueurs. Son plus gros inconvénient demeure dans la lourde préparation des échantillons avant analyse.
- La spectroscopie ultrasonore résonnante : elle utilise des propriétés sur les ondes acoustiques pour détecter des molécules au sein d'un échantillon de manière non invasive. Cependant, cette technique d'imagerie hyperspectrale ne permet pas d'identifier avec certitude la composition chimique de l'échantillon analysé. Seules des informations morphologiques sont disponibles. La quantification d'un AC dans la peau est alors impossible.
- La spectroscopie de fluorescence : elle nécessite que les molécules d'intérêts dans les échantillons analysés soient fluorescentes pour pouvoir les suivre.

Par ailleurs, deux autres techniques d'imagerie hyperspectrale sont régulièrement utilisées dans ces domaines de recherche et donnent accès à la composition moléculaire de l'échantillon : l'imagerie proche et moyen infrarouge et l'imagerie Raman. Elles consistent à utiliser la lumière, une onde électromagnétique, pour étudier respectivement l'absorption et la diffusion de la lumière par les molécules présentes dans l'échantillon. La quantité de lumière absorbée ou diffusée par les molécules permet de les distinguer et de les identifier. À noter que les informations sur la composition moléculaire des échantillons fournies par ces deux techniques d'imagerie hyperspectrales sont même complémentaires. De plus, ces deux techniques ne nécessitent aucune préparation d'échantillon telle qu'une coloration ou un marquage. En revanche, la résolution spatiale de l'imagerie infrarouge est beaucoup plus limitée que celle de l'imagerie Raman : elle est de l'ordre de 10-15 µm contre 1 µm. Par conséquent, la spectroscopie Raman s'est révélée ces dernières années être un outil d'analyse puissant dans de nombreux domaines de recherche : pharmaceutique, biologique et cosmétique.

2.3. SPECTROSCOPIE DE DIFFUSION RAMAN

Bien que de nouvelles techniques de recherche ont vu le jour ces dernières dizaines d'années et ont élargi le champ des possibles pour l'étude de la pénétration cutanée d'actifs cosmétiques et pharmaceutiques, la spectroscopie Raman a gagné du terrain dans les domaines des sciences biologiques et de la vie [25]. Elle a particulièrement été très utilisée pour la recherche sur la peau ces dernières années. Elle est même considérée dans le domaine de la recherche et du développement cosmétique comme un outil puissant pour confirmer la diffusion d'une molécule à travers la barrière cutanée sans avoir à accéder aux concentrations absolues. En effet, la reconstruction d'images de distribution en fausses couleurs permet une comparaison facile des résultats obtenus selon différentes conditions d'application qui est essentielle pour soutenir par exemple l'optimisation des formulations de soin de la peau.

Cette partie présente dans un premier temps l'interaction matière-rayonnement, et plus particulièrement la production de l'effet Raman à l'origine du principe de la spectroscopie Raman. Puis, les constituants et le processus d'acquisition d'un microspectromètre confocal Raman sont décrits. Enfin, une description des bandes caractéristiques de spectres issus d'une image hyperspectrale Raman d'un échantillon d'épiderme humain reconstruit, Episkin®, est illustrée.

2.3.1. Le principe de la spectroscopie Raman

2.3.1.1. L'effet Raman

En fonction de la nature de la matière et du type de rayonnement, il existe plusieurs types d'interaction rayonnement-matière comme la réflexion, l'absorption ou encore la diffusion (Figure I-6). Le phénomène de la diffusion Raman a été découvert dans les années 1920 d'un côté par C.V. Raman et K.S. Krishman (prix Nobel en 1930) et de l'autre par G.S. Landsberg et L.I. Mandelstam [26].

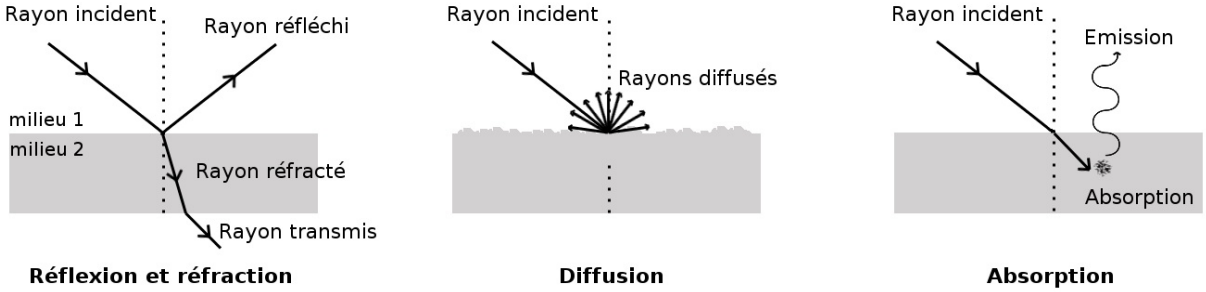


Figure I-6 – Les différents phénomènes physiques lorsqu'un rayonnement incident rencontre une transition entre deux milieux de natures différentes : réflexion et/ou réfraction, diffusion et absorption [27].

Après une irradiation de la matière par une source monochromatique, les molécules peuvent conserver leur niveau d'énergie vibrationnelle fondamental ou atteindre des niveaux énergétiques différents. Le diagramme de Jablonski, présenté sur la Figure I-7, illustre les différences d'énergie vibrationnelle à l'origine de la diffusion Raman. Ces différences d'énergie vibrationnelle sont aussi appelées décalages Raman qui sont présentés dans les spectres en nombres d'ondes $\bar{\nu}$ et exprimés en cm^{-1} (unité inverse à la longueur d'onde). Deux types de diffusion sont à l'origine de ces changements d'états vibrationnels :

- La diffusion élastique, appelée aussi diffusion Rayleigh, est majoritaire (Figure I-8). La molécule est excitée par un photon d'énergie $E_0 = h\,\nu_0$ (avec h la constante de Planck et ν_0 la fréquence de radiation de la source excitatrice) et atteint un niveau d'énergie virtuel avant de revenir à son niveau vibrationnel fondamental en émettant la même quantité d'énergie $E = h\,\nu_0$ (Figure I-7). Dans ce cas, il n'y a pas de changement de longueur d'onde au niveau des photons diffusés.

- La diffusion inélastique, quant à elle, se caractérise par une probabilité faible (1 photon sur 10^7 [26], [28]) d'observer des changements dans les longueurs d'onde des photons diffusés (Figure I-8). Cet écart de longueur d'onde peut être positif ou négatif et traduit respectivement une perte et un gain d'énergie moléculaire qui sont identifiés dans l'ordre comme la diffusion Stokes et la diffusion anti-Stokes :

  - La diffusion Stokes : la molécule est excitée par un photon d'énergie $E_0 = h\,v_0$ et atteint un niveau d'énergie virtuel avant de revenir au 1er état vibrationnel excité en émettant un photon d'énergie plus faible $E = h\,(v_0 - v_v)$ (Figure I-7). La longueur d'onde du photon diffusé est donc supérieure à celle de la source excitatrice qui se traduit par un décalage Raman avec des nombres d'onde en cm^{-1} positifs.

  - La diffusion anti-Stokes : avec une probabilité moindre, la molécule excitée se situe déjà au 1er niveau vibrationnel excité. Cette molécule qui absorbe un photon d'énergie $E_0 = h\,v_0$ revient à l'état vibrationnel fondamental en émettant un photon d'énergie supérieure $E = h\,(v_0 + v_v)$ (Figure I-7). Dans ce cas, la longueur d'onde du photon diffusé est inférieure à celle de la source excitatrice ce qui se traduit par un décalage Raman avec des nombres d'onde en cm^{-1} négatifs.

Les raies anti-Stockes sont cependant de plus faibles intensités que les raies Stockes. Par conséquent, l'analyse par spectroscopie Raman s'appuie généralement sur la mesure de la diffusion Raman Stokes.

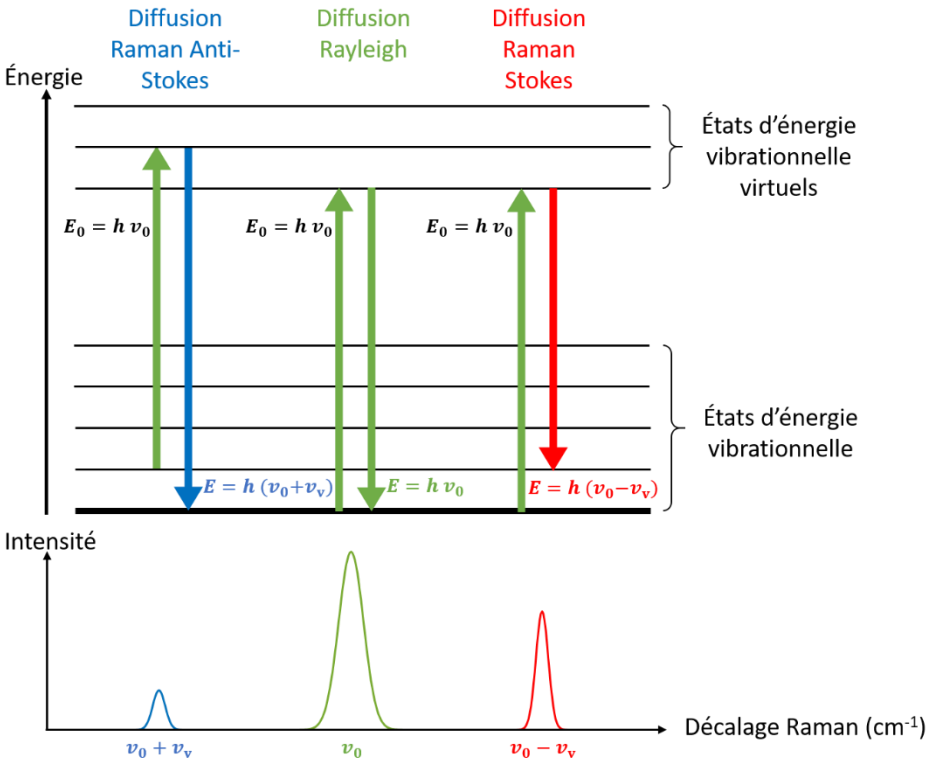


Figure I-7 – Diagramme de Jablonski et raies de diffusion

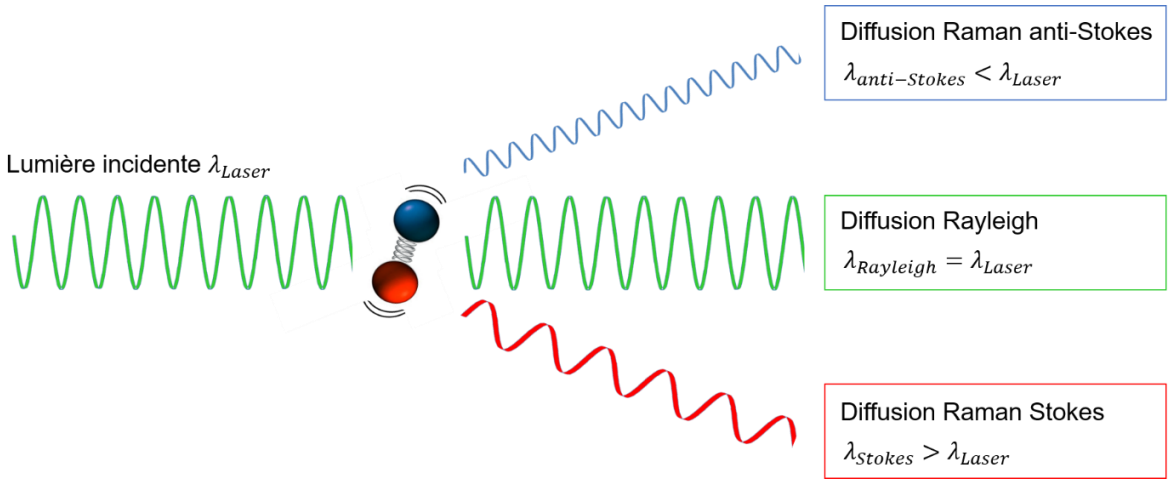


Figure I-8 – Phénomène de la diffusion de la lumière.

2.3.1.2. Les modes vibrationnels

Dans le cas de la spectroscopie Raman, seules les transitions vibrationnelles sont prises en considération. Les différences d'énergie vibrationnelle à l'origine de la diffusion Raman dépendent particulièrement de la masse des atomes qui composent la molécule et de la force de leurs liaisons. Cette relation est approximativement décrite par la formule suivante :

$$\bar{\nu} = \frac{1}{2\pi}\sqrt{\frac{\kappa}{\mu}} \qquad \text{(I-1)}$$

où κ est la mesure de la force de liaison et μ la masse réduite des atomes. Les décalages Raman sont donc source d'informations sur la composition moléculaire puisqu'ils sont plus ou moins différentiables et identifiables selon le type d'atomes et de liaisons des structures chimiques présents dans les molécules. Ainsi, l'observation de la position des raies Stokes et anti-Stokes définie par les décalages Raman permet de caractériser la composition des molécules irradiées.

Plus précisément, les décalages Raman sont définis par les différents modes vibrationnels induits par l'irradiation de la molécule. Les modes vibrationnels d'une molécule sont directement liés à sa structure chimique, c'est-à-dire au nombre d'atomes qui la composent, les ramifications et sa géométrie. De ce fait, une molécule contenant N_A atomes possède $3N_A$ degrés de liberté dont trois sont associés à une translation du centre de la molécule et trois autres correspondent à une rotation de la molécule entière. Dans le cas d'une molécule linéaire, seulement deux rotations d'ensemble existent parmi ces $3N_A$ degrés de liberté. Par conséquent, une molécule possède $3N_A - 6$ (et $3N_A - 5$ pour une molécule linéaire) degrés de liberté et autant de modes normaux de vibrations pour modifier sa structure chimique au voisinage de son état d'équilibre après avoir subi une perturbation via une source lumineuse excitatrice [26], [28]. Ces modes de vibrations peuvent être classés en deux classes : les modes de vibration d'élongation et les modes de vibration de déformation. Alors que les premiers modifient les longueurs des liaisons chimiques de manière symétrique ou asymétrique, les seconds modifient les angles entre les liaisons moléculaires adjacentes (Figure I-9). Ces modes vibrationnels sont nombreux au sein d'une seule molécule de par la complexité de sa structure chimique et peuvent être analysés selon la force des liaisons chimiques ou selon des groupements moléculaires fréquemment rencontrés (Tableau I-1).

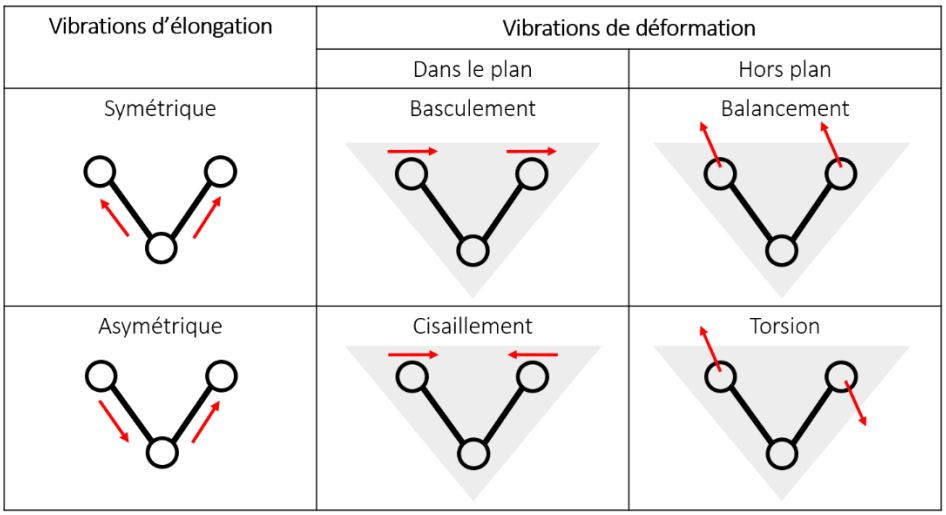


Figure I-9 – Modes de vibrations d'une molécule

Tableau I-1 – Attribution des bandes de groupes fonctionnels en spectroscopie Raman [29]. *υ : élongation ; δ : déformation symétrique dans le plan.*

Vibrations / Groupe fonctionnel	Décalage Raman (cm^{-1})	Intensité Raman
$\upsilon(O-O)$	$845-900$	forte
$\upsilon(C-O-C)$	$800-970$	moyenne
$\upsilon(C-O-C)$ asym	$1060-1150$	faible
$\delta(CH3)$	1380	moyenne
$\delta(CH2)/\delta(CH3)$ asym	$1400-1470$	moyenne
$\delta(H2O)$	~ 1640	faible
$\upsilon(C=N)$	$1610-1680$	forte
$\upsilon(C=C)$	$1500-1900$	forte
$\upsilon(C=O)$	$1680-1820$	moyenne
$\upsilon(C-H)$	$2800-3000$	forte
$\upsilon(N-H)$	$3300-3500$	moyenne
$\upsilon(O-H)$	$3100-3650$	faible

Par ailleurs, les changements de la polarisabilité d'une molécule induits par le champ électrique de la source excitatrice définissent la force du mode de vibration et impactent directement l'intensité de la diffusion Raman. Autrement dit, plus l'influence d'une vibration moléculaire sur la polarisabilité d'une molécule est importante et plus la réponse Raman est intense. Ainsi, l'intensité des vibrations actives en Raman pour une molécule est à nuancer (Tableau I-1, dernière colonne). De même, l'intensité de la diffusion Raman est linéairement dépendante de la concentration des molécules actives et facilement impactée par des changements de température. Tout compte fait, au même titre que leur position, l'intensité des raies Stokes et anti-Stokes fournit des informations supplémentaires aussi bien qualitatives que semi-quantitatives sur la composition des molécules irradiées.

En définitive, l'analyse de la position des raies Stokes combinée à l'analyse de leur intensité permet d'identifier et de différencier les modes de vibrations moléculaires induits par l'irradiation et par conséquent de caractériser les molécules et groupements fonctionnels présents. Par exemple, la vibration d'élongation symétrique de la molécule d'eau s'exprime par une bande de diffusions Stokes située vers 1640 cm^{-1} de faible intensité. A l'inverse la déformation de basculement de la double liaison carbone s'exprime dans la région spectrale [1500-1900] cm^{-1} par une forte bande de diffusion. Le Tableau I-1 concentre quelques exemples typiques d'expression de vibrations de groupements fonctionnels.

2.3.2. Micro-spectromètre Confocal Raman

L'intérêt grandissant pour la caractérisation moléculaire dans des domaines variés tels que les sciences du vivant, l'industrie agroalimentaire, pharmaceutique ou encore cosmétique, a encouragé le développement et l'amélioration de la technique Raman. Les progrès technologiques récents tels que le développement des sources monochromatiques puissantes comme les lasers et l'évolution des systèmes optiques ou encore des systèmes de détection ont permis d'en rendre son utilisation plus simple, plus fiable et plus rapide. D'ailleurs, le couplage de la spectroscopie Raman avec la microscopie confocale fait partie de ces évolutions et la microspectrométrie confocale Raman (MCR) est un outil de caractérisation moléculaire très plébiscité pour sa qualité et sa précision de mesure, notamment en termes d'imagerie.

2.3.2.1. Les constituants de l'appareillage

Un micro-spectromètre Raman est composé de cinq éléments principaux : une source lumineuse excitatrice telle qu'un laser, un microscope confocal, un système de réjection, un système dispersif et un système de détection (Figure I-10). De manière simplifiée, la mesure de la diffusion Raman s'obtient de la façon suivante : la source laser est focalisée sur l'échantillon via l'objectif du microscope, le rayonnement diffusé par la matière suite à des échanges énergétiques entre ses molécules est collecté à travers l'objectif du microscope et filtré avant d'être redirigé vers le système dispersif, le système dispersif sépare spatialement les longueurs d'ondes des photons diffusés par l'échantillon et les envoie sur le système de détection, enfin le détecteur convertit la lumière diffusée en un signal numérique. Les paragraphes suivants présentent une description et le rôle de chaque composant de façon plus détaillée.

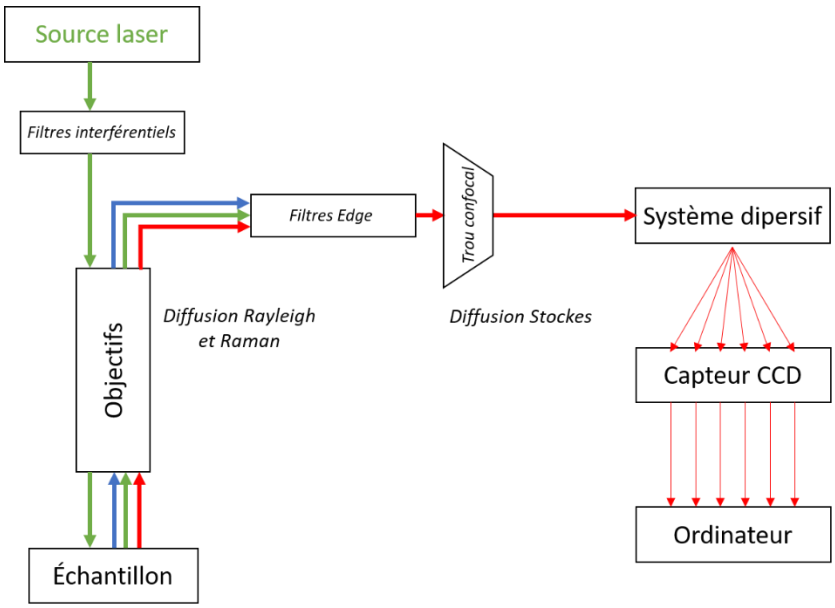


Figure I-10 – Schéma simplifié d'un microspectromètre confocal Raman

(i) Les sources lasers

Le choix de l'utilisation des lasers comme source de lumière monochromatique excitatrice repose sur leur capacité à émettre une importante densité de photons. Ainsi, les temps d'apparition de l'effet Raman sont réduits [26]. De plus, leur simplicité d'utilisation justifie leur emploi. Les lasers fréquemment utilisés dans les systèmes Raman actuels émettent à 532nm, 633nm et 785nm [28], [30]. Cependant pour certaines applications, notamment lorsque les échantillons irradiés sont quelques fois fluorescents il convient d'utiliser des lasers émettant à 830 nm et à 1064 nm.

(ii) Microscope confocal et résolution spatiale

Objectif optique et résolution spatiale latérale

L'objectif optique a plusieurs rôles : la focalisation du laser sur l'échantillon et la collecte de la lumière diffusée. La focalisation de la source lumineuse en un point de petite taille est assurée par l'utilisation d'un objectif possédant une grande ouverture numérique (ON ou NA pour *Numerical Aperture* en anglais) et un grossissement adapté permettant de percevoir les détails de l'échantillon.

L'optimisation des deux paramètres de l'objectif optique, à savoir l'ouverture numérique et le grossissement, influe directement sur la résolution spatiale de la mesure dans le plan (x, y) au même titre que la longueur d'onde de la source excitatrice utilisée. En effet, la résolution spatiale latérale est définie comme proportionnelle à la limite de diffraction de la lumière qui s'exprime comme un rapport de la longueur d'onde avec l'ouverture numérique de l'objectif [26], [28]:

$$\frac{0.61\ \lambda_{Laser}}{NA}$$

La limite de diffraction théorique de la lumière correspond à la distance minimale en dessous de laquelle il n'est plus possible de distinguer deux points voisins. La résolution spatiale latérale reste donc malgré tout limitée par la diffraction théorique du microscope. Il est généralement admis dans la littérature que les mesures obtenues par spectroscopie Raman ont une résolution spatiale latérale de l'ordre du micromètre [26], [28].

Toutefois, malgré une ouverte numérique de l'objectif maximale, il est encore possible d'augmenter la résolution spatiale latérale en choisissant une longueur d'onde plus faible. Toutefois, une diminution de la longueur d'onde doit être réalisée avec précaution puisque l'utilisation de longueurs d'onde plus courtes peut s'accompagner d'émission de fluorescence en cas de présence de molécules comportant des fluorochromes excités par la source laser [30]. Généralement, la diffusion Raman est masquée par la fluorescence qui est beaucoup plus intense. De même, une vigilance doit être apportée lors de l'utilisation d'une longueur d'onde courte puisque des phénomènes de photo-dégradation comme l'échauffement de la matière peuvent se produire.

Mesure confocale

L'utilisation d'un dispositif de confocalité permet de réaliser des sections optiques, c'est-à-dire des mesures à profondeur de champ contrôlée. Le but est d'effectuer un filtrage spatial du rayonnement diffusé par l'échantillon à l'aide d'un diaphragme ou système de caches placé dans le plan image du microscope, plus communément appelé trou confocal, pour enregistrer le signal à partir d'une tranche optique. Ainsi, la lumière parasite située à l'extérieur de la zone centrale est éliminée et seules les informations provenant de la zone de diffusion réduite de l'objectif sont prises en compte. De plus, la sélection de la partie centrale du rayonnement diffusé permet de contrôler la profondeur de la zone de mesure (Figure I-11). Le paramétrage de la fermeture du trou confocal permet alors d'ajuster à la fois la résolution spatiale latérale dans le plan (x, y) mais aussi la résolution spatiale axiale (en profondeur) dans la direction z. En définitive, le couplage de la spectroscopie Raman à la microscopie confocale permet de contrôler les dimensions de la zone de mesure aussi bien de manière latérale qu'axiale, c'est-à-dire de contrôler le volume de la zone de mesure.

En général, la résolution spatiale des sections optiques est de l'ordre du micromètre. Ainsi, des informations détaillées au niveau tissulaire et cellulaire sont accessibles. Cependant, la qualité de la confocalité a une incidence sur l'intensité de la mesure enregistrée puisque la diminution de la taille du trou confocal se fait au détriment du nombre de photons collectée et donc de l'intensité du signal.

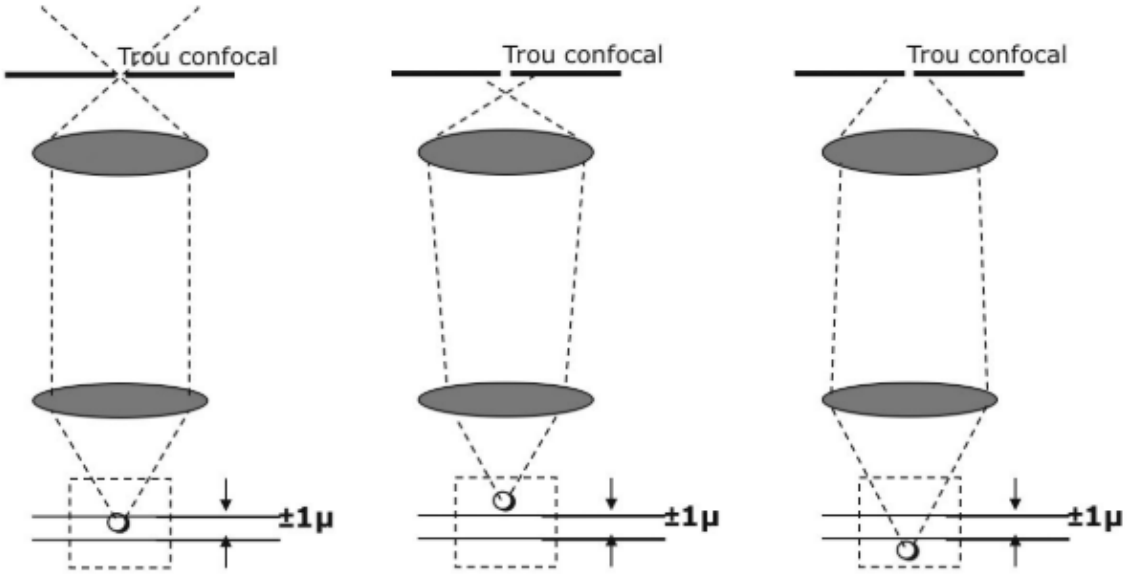


Figure I-11 – Schéma simplifié du principe de confocalité [26]

(iii) Systèmes de filtres

Les filtres intégrés au système de microspectromètre confocal Raman ont des rôles différents selon leur position dans le trajet optique [26], [28].

- Un premier filtre interférentiel dit passe-bande, positionné entre le laser et l'échantillon, garantit exclusivement le passage de la longueur d'onde de la source laser dans l'objectif en bloquant la lumière parasite hors champs (dont la lumière ambiante).

- D'autres filtres placés entre l'objectif et le détecteur ont pour but de rejeter la diffusion associée à la longueur d'onde excitatrice. En effet, après irradiation de l'échantillon par la source laser, la diffusion Rayleigh et la diffusion Raman sont toutes deux captées par l'objectif du microscope. La diffusion Raman étant six à sept ordres de grandeur moins intense que la diffusion Rayleigh [26], il est nécessaire de rejeter au maximum le signal associé à la source laser. Les filtres les plus communément utilisés pour la réjection de la raie excitatrice sont les filtres « Edge » et « Notch ». Le premier est généralement un filtre passe-haut par rapport à la longueur d'onde excitatrice. Autrement dit, seules les longueurs d'ondes supérieures à celle de la source laser sont autorisées à être redirigées vers le système dispersif. Ce filtre permet donc de rejeter la diffusion anti-Stokes et de sélectionner uniquement la diffusion Stokes. Quant au filtre Notch, il transmet toutes les longueurs d'ondes diffusées au système dispersif, exceptées celles avoisinant de trop près la longueur d'onde du laser. Ce filtre passe-bande a l'avantage de pouvoir enregistrer des spectres Raman Stokes et anti-Stokes. Cependant, ce filtre est plus coûteux et a une durée de vie plus faible que les filtres Edge. De plus, la diffusion anti-Stokes étant très faible est très rarement analysée sur des échantillons biologiques. Par conséquent, la majorité des systèmes sont actuellement équipés de filtres Edge et l'emploi d'un simple filtre Notch est beaucoup moins fréquent [26].

(iv) Système dispersif, détecteur et résolution spectrale

Le système dispersif a pour rôle de disperser géométriquement les longueurs d'ondes de la lumière diffusée sur le détecteur. En d'autres termes, la lumière diffusée est séparée spatialement afin de percevoir les informations détaillées des décalages Raman induits par l'irradiation de l'échantillon. Pour se faire, le réseau de diffraction est constitué d'une série de sillons étroitement espacés et gravés sur la surface, en général entre 150 et 4000 traits par millimètre [28], qui permettent de diffracter la lumière.

Les détecteurs sont généralement composés de matériaux électroniques semi-conducteurs, comme le silicium ou le germanium. Ils sont principalement de dimensions 1024 x 256 pixels ou 2048 x 512 pixels avec des pixels de taille identique de l'ordre de 10 à 20 μm^2 [26]. Ce caractère matriciel permet d'enregistrer une large gamme spectrale en des temps raisonnables (plus précisément, le nombre de pixels dans la première dimension du détecteur définit le nombre de variables spectrales pouvant être enregistrées) ce qui justifie leur efficacité. Les signaux électriques sont alors générés suite à la projection des photons redirigés par le système dispersif sur chacun de ces pixels. Les détecteurs les plus communément retrouvés dans les systèmes actuels sont les détecteurs à transfert de charges, notamment les capteurs CCD (Charge Coupled Device), en raison de leur très forte sensibilité et leur taux de conversion des photons en électrons élevé.

Enfin, le système dispersif et le détecteur jouent un rôle majeur dans la définition de la résolution spectrale que peut offrir un système d'acquisition : le nombre de rainures que dispose le réseau de diffraction et la distance qui le sépare du détecteur l'influencent grandement. En théorie, plus l'intervalle entre les nombres d'ondes consécutifs enregistrés est faible, meilleure est la résolution spectrale et plus l'analyse spectrale ultérieure est performante. À l'inverse, un intervalle entre deux mesures de longueurs d'ondes trop grand impacte principalement la finesse, la position mais aussi la forme des bandes Raman qui peuvent être écrêtées ce qui nuit fortement à l'interprétation et l'analyse spectrale en raison des nombreuses informations pouvant être délaissées. La résolution spectrale des spectres enregistrés est donc déterminée par la capacité d'un système d'acquisition à séparer les bandes Raman proches les unes des autres.
À noter aussi qu'en raison de l'agrandissement de la dispersion des photons et de l'augmentation de la résolution spectrale qu'elle engendre, l'augmentation du nombre de sillons du système dispersif implique également une baisse de l'intensité du signal envoyé au système de détection. Un juste équilibre entre une grande dispersion et une intensité du signal suffisante doit donc être trouvée. En pratique, des temps d'acquisitions plus longs permettent de compenser cette diminution de l'intensité du signal et des systèmes dispersifs, situés à une distance modulable par rapport au détecteur, sont généralement composés de 300 à 600 traits par millimètre pour travailler avec une résolution spectrale d'environ 1-2 cm^{-1} [28].

2.3.2.2. Les spectres Raman

Les progrès technologiques apportés aux différents composants des spectromètres Raman et le développement d'ordinateurs puissants ont permis d'acquérir des signaux numériques individuels qui contiennent les informations relatives à la diffusion Raman d'un échantillon en quelques secondes. Ces signaux sont appelés spectres Raman. Ils représentent l'expression des longueurs d'ondes associées aux vibrations des liaisons moléculaires produites suite à une exposition à une source laser excitatrice. Ces spectres Raman sont exprimés en nombre d'ondes par unité de longueur [cm^{-1}], appelé aussi décalage Raman, qui n'est autre que l'inverse de la longueur d'onde et qui est l'unité habituellement utilisée en spectroscopie vibrationnelle. Chaque bande spectrale est associée théoriquement à un mode vibrationnel spécifique et l'ensemble des bandes incarne l'empreinte moléculaire de l'échantillon. Cette signature spectrale est donc riche en informations sur la composition biochimique de l'échantillon.

Néanmoins, la complexité de la molécule peut être un frein à l'interprétation de sa signature spectrale. En effet comme mentionné précédemment dans ce chapitre, il existe de nombreux modes de vibration au sein d'une seule et même molécule. Chacun d'eux dépend de la masse des atomes, des forces de liaisons présentes et de l'environnement chimique voisin. Par conséquent, seule une analyse qualitative et quantitative approfondie des spectres Raman permet de comprendre la composition biochimique. Cette analyse passe par l'observation de la position, de l'intensité, du décalage et de la

largeur à mi-hauteur des bandes spectrales Raman [26]. La position d'une bande en nombre d'ondes fournit des informations sur la nature de la liaison chimique, et plus précisément sur les atomes et les forces de liaison présentes. L'intensité d'une bande quant à elle est proportionnelle à la concentration de la molécule. Concernant la largeur à mi-hauteur d'une bande, elle caractérise les désordres locaux du composant chimique. Autrement dit, elle rend compte de l'arrangement et de la force des liaisons chimiques : une configuration moléculaire organisée et régulière impliquera une bande plus étroite. Enfin, un décalage en nombre d'onde d'une bande Raman par rapport à sa position d'équilibre observée sur un spectre de référence peut être le résultat d'une variation de température ou de contraintes mécaniques perturbant les longueurs des liaisons à l'échelle moléculaire.

2.3.2.3. Les images Raman

Le concept de cartographie Raman proposé dans les années 1970 [26] repose sur l'association de la motorisation et du pilotage des platines sur lesquelles sont positionnés les échantillons et sur la résolution micrométrique offerte par un microscope confocal Raman.

Les images hyperspectrales Raman sont constituées d'un ensemble de pixels juxtaposés les uns à côté des autres et associés à des points de coordonnées définis sur l'échantillon (Figure I-12). Chaque pixel est composé d'un spectre exprimant l'intensité de la diffusion Raman pour de nombreux nombres d'ondes consécutifs. Au final, une image Raman n'est autre qu'un cube de données à trois dimensions : deux dimensions spatiales dans les plans (x, y) ou (x, z) à l'échelle micrométrique et une dimension spectrale dont l'unité est le nombre d'ondes. Par exemple, la Figure I-12 illustre un spectre défini sur la région spectrale [400-1800]cm^{-1} contenant N_λ variables spectrales et appartenant à un cube de données contenant $N_x \times N_y = N$ spectres. Néanmoins, les dimensions spatiales peuvent être remplacées ou complétées par une dimension associée à un facteur extérieur tel que le temps, la température ou une contrainte mécanique [26]. Cette mesure s'obtient en enregistrant une série de spectres sur un même point de coordonnées en fonction du paramètre externe. Les images Raman ainsi obtenues sont représentées comme des hypercubes à plus de deux dimensions : une dimension spectrale, une dimension externe et/ou éventuellement une ou deux dimensions spatiales.

Une image Raman peut être construite selon trois modes d'acquisition : la méthode point par point, la méthode en ligne ou la méthode globale [26], [28].

- La cartographie Raman point par point consiste à déplacer la platine à intervalle régulier sur les coordonnées de chaque point de la zone d'intérêt de l'échantillon prédéfinie et d'enregistrer un spectre pour chacune d'entre elles. Les spectres sont ainsi enregistrés les uns après les autres et la caractéristique de confocalité est conservée pour chaque acquisition. Cette méthode d'acquisition est celle la plus fréquemment utilisée.

- La cartographie Raman en ligne enregistre de façon simultanée des spectres associés à toute une ligne de points de coordonnées. Le principal avantage de ce procédé est le gain de temps dans la construction d'une image Raman assurée par l'utilisation des deux dimensions du capteurs CCD pour l'enregistrement des spectres. De plus, la faible exposition de l'échantillon à la lumière permet d'imager des matériaux thermiquement fragiles.

- La méthode d'imagerie globale est rarement utilisée dans le domaine de la recherche. Elle consiste à mesurer l'intensité de la diffusion Raman de la totalité de la surface illuminée en une seule fois sans déplacer la platine supportant l'échantillon. L'image Raman est alors formée directement sur le détecteur bidimensionnel, mais sa taille reste limitée à celle du détecteur.

Tandis que l'acquisition d'un seul spectre pour l'analyse d'un échantillon hétérogène composé de plusieurs composants peut être insuffisant pour informer sur la distribution locale de ses constituants

biochimiques, l'acquisition de nombreux spectres à des points de coordonnées maitrisés sur l'échantillon est un moyen d'en assurer une analyse plus complète. En effet, chaque pixel de l'image caractérise le mélange de différents composants présents à un point donné de l'échantillon et l'ensemble des pixels comportent des informations complexes concernant la composition locale de la matière. Autrement dit, l'imagerie Raman permet d'évaluer la distribution de chaque composant d'un échantillon. Elle est d'autant plus intéressante lorsque les dimensions de ces hétérogénéités sont comparables à la résolution spatiale du système d'acquisition. Cependant, à des fins d'identification et d'évaluation de la proportion des composants dans chaque mesure, l'application ultérieure de méthodes mathématiques est indispensable pour accéder aux informations qualitatives et semi-quantitatives locales de la composition d'un échantillon.

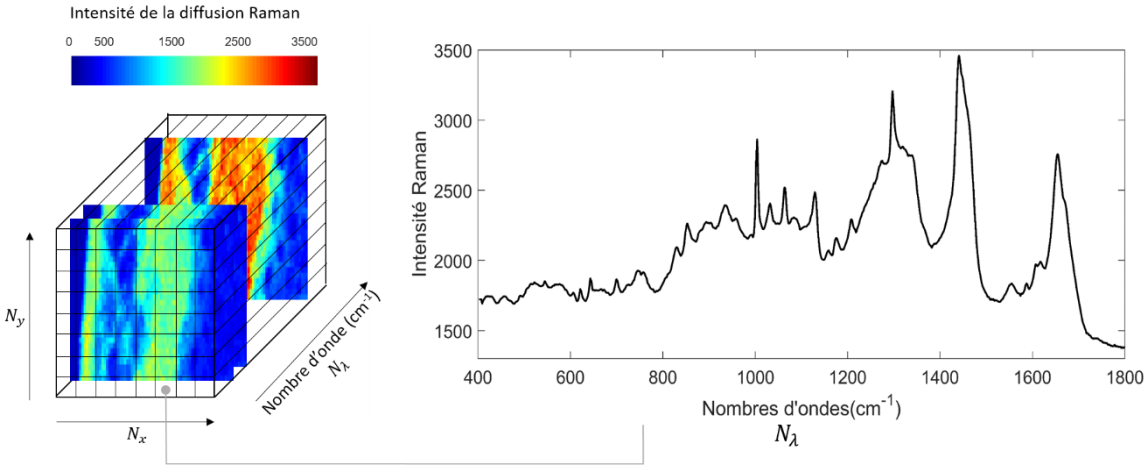


Figure I-12 – Principe d'une image hyperspectrale Raman

2.3.3. Imagerie Raman de la peau humaine

Ces dernières années, la spectroscopie Raman a été très utilisée pour étudier la composition moléculaire de la peau en raison des nombreux avantages qu'elle offre. Plus encore, le choix de l'imagerie Raman se justifie par sa capacité à fournir des informations sur la spécificité moléculaire de la peau et plus précisément sur la composition et la structure biochimique de ses hétérogénéités locales avec une qualité spatiale de l'ordre du micromètre. Elle est notamment employée dans le but de diagnostiquer des pathologies cutanées et d'étudier leurs mécanismes physiopathologiques mais aussi pour étudier les effets d'agents exogènes (UV, pollution, etc.) ou de produits pharmaceutiques et cosmétiques sur la peau. D'autre part, cette technique d'imagerie ne nécessite aucun marquage chimique supplémentaire de l'échantillon avant de l'irradier et est non-destructrice : l'intégrité de l'échantillon est conservée à condition néanmoins de contrôler la puissance et la longueur d'onde du laser utilisé afin d'éviter d'éventuelles dégradations à l'instar de l'échauffement de la matière.

La peau étant un épithélium squameux stratifié avec différentes couches, sa composition moléculaire est variable selon les couches. Par conséquent les signatures spectrales contenues dans les pixels de l'image Raman varient localement et représentent des mélanges différents de multiples composants biochimiques. Cette diversité est illustrée par la Figure I-13 où les spectres Raman moyens de différents niveaux de profondeur d'un échantillon d'épiderme humain reconstruit (Episkin®), dans le SC et dans une couche intermédiaire et plus profonde de l'épiderme viable, sont représentés. Plusieurs variations de bandes sont observées en fonction de la profondeur cutanée. Elles sont le résultat de mélanges composés principalement de protéines, des lipides, des acides nucléiques ou encore d'autres composants organiques qui sont distribués de manière différente selon la profondeur cutanée. De plus, ces signatures spectrales sont particulièrement complexes à interpréter. A. Tfayli *et al.* [31] ont fourni une interprétation détaillée des bandes caractéristiques de ces constituants dans les spectres

Raman enregistrés à partir d'un modèle d'épiderme humain reconstitué, le modèle Episkin®. Un résumé sommaire de ces bandes Raman est présenté dans le Tableau I-2 [31]–[35]. À noter que seuls des spectres Raman enregistrés à partir d'un échantillon d'épiderme humain reconstruit sont ci-après illustrés car les traitements de données appliqués dans les chapitres suivant ne concerne que des expériences réalisées sur ces modèles *in vitro*.

De nombreuses bandes spectrales sont associées à des acides aminés provenant des protéines contenues dans les cellules de l'épiderme et participant à l'hydratation naturelle de la peau. Parmi ces acides aminés, la tyrosine intervenant dans la synthèse de la mélanine s'exprime à 643, 830 et 853 cm^{-1}. La phénylalanine, précurseur de la tyrosine, et la cystéine et ses dérivés intervenant dans la synthèse de la mélanine et de la kératine quant à elles s'expriment respectivement à 621, 1004, 1032 et 1339 cm^{-1} et à 643, 670, 700 et 722 cm^{-1}. Les changements dans la composition de ces protéines et dans leur conformation se reflètent aussi dans la position et la forme de la bande amide I située à 1655 cm^{-1}. Pour ce qui est de la composition lipidique de l'épiderme reconstruit, on retrouve principalement des bandes Raman associées aux céramides à 1064 et 1129 cm^{-1}, au cholestérol à 545, 699, 1441 et 1669 cm^{-1} et aux acides gras à 1297 cm^{-1}. L'intensité de ces bandes est plus forte dans les couches supérieures de la peau. Enfin, des bandes spectrales d'ADN et d'acides nucléiques sont également observées à 722, 1062, 1324 et 1340 cm^{-1}. Au final, cette interprétation des spectres Raman moyens issus de différentes couches épidermiques montre la richesse d'informations disponibles mais aussi la complexité de leur analyse.

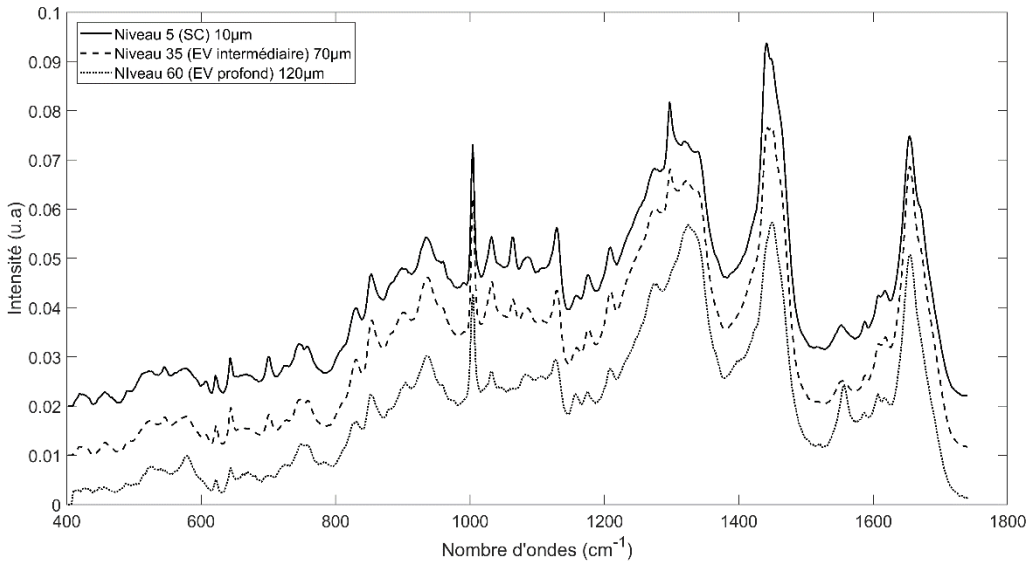


Figure I-13 – Spectres Raman moyens d'une ligne de pixel situés dans la couche supérieure (courbe continue), la couche médiane (courbe discontinue), et la couche inférieure (courbe en pointillé) d'un épiderme humain reconstruit (Episkin®).

Tableau I-2 – Attribution des bandes Raman d'un échantillon d'épiderme humain reconstruit Episkin®. υ : élongation ; δ : déformation

Nombre d'ondes (cm^{-1})	Attribution des bandes
545	Cholestérol
621	δ(CC) phenylalanine
643	δ(CC) tyrosine
670	υ(CS) cystosine
699	Cholestérol, Cholestérol ester
700	υ(CS) cystéine
722	υ(CS) cystine ADN
830	Doublet de Fermi de la tyrosine (cycle) Respiration hors plan du cycle de la tyrosine (OPO) étirement de l'ADN O-P-O des phospholipides
853	Respiration du cycle de la tyrosine et υ(CC)
1004	Phenylalanine υ(CC) phenylalanine
1032	δ(CH) phenylalanine
1062	Lipides, céramides $\upsilon(CO)$ ADN ARN
1064	Lipides, céramides
1129	Lipides, céramides
1297	Amide III Phospholipides, céramides
1324	Acides nucléiques ADN
1339	Phenylalanine
1340	Acides nucléiques ADN, ARN
1441	δ(CH) lipides (céramides, cholestérol)
1451	δ(CH) protéines et lipides
1557	Tryptophane
1655	Amide I $\upsilon(C=O)$ Amide I $\upsilon(C=C)$ lipides
1669	Cholestérol ester $\upsilon(C=O)$ Amide I ν(C=O) ARN ADN

CONCLUSION

Aujourd'hui, l'utilisation des modèles de peau humaine reconstruite *in vitro* permet de réaliser des tests de produits cosmétiques à grande échelle de manière reproductible et sans nocuité. L'application de l'imagerie Raman sur ces échantillons de peau s'avère être un outil puissant pour décrypter les processus physiologiques engagés suite à l'application de produits cosmétiques. En effet, elle rend possible l'étude des changements moléculaires induits par ces produits cosmétiques à l'échelle micrométrique de façon simple et rapide puisqu'elle fournit des empreintes moléculaires des échantillons de peaux sous forme de spectre Raman. Cependant, ces spectres Raman sont complexes

et difficiles à analyser en raison de leurs très nombreuses variables spectrales ayant un sens chimique. C'est pourquoi il est nécessaire d'utiliser des méthodes mathématiques adaptées relevant de la chimiométrie pour extraire un maximum d'informations qualitatives et quantitatives des images Raman [26]. Dans le chapitre suivant nous présentons l'état de l'art de ces méthodes.

Chapitre II.
ÉTAT-DE-L'ART

Résumé
Ce chapitre présente un état-de-l'art des méthodes d'analyse univariée et multivariée régulièrement utilisées sur des données spectrales Raman dans les domaines cosmétiques, pharmaceutiques et biologiques. Trois catégories sont dégagées selon que les méthodes comparent directement les intensités, opérent une classification, ou un démélange des spectres Raman. Des exemples d'applications sur données spectrales Raman y sont illustrés. Nous soulignons les limitations de chacune de ces méthodes pour notre application.

Introduction **52**
1. Analyse basée sur l'intensité des bandes Raman **52**
1.1. Suivi d'intensité des bandes spectrales 52
1.2. Suivi des rapports de bandes spectrales 53
1.3. Conclusion 54
2. Méthodes de classification **55**
2.1. K-means 56
2.1.1. Principe de la méthode des K-means 56
2.1.2. Interprétation des résultats de la méthode K-means 56
2.1.3. Exemples d'application de la méthode K-means sur des données spectrales Raman 57
2.1.4. Limites de la méthode K-means 58
2.2. Classification hiérarchique 58
2.2.1. Principe de la classification hiérarchique 58
2.2.2. Exemples d'application de la CHA 60
2.2.3. Limites de la CHA 60
2.3. Conclusion 60
3. Méthodes de démélange **61**
3.1. Analyse en Composantes Principales (ACP) 61
3.1.1. Principe de l'ACP 61
3.1.2. Interprétation des résultats de l'ACP 63
3.1.3. Exemples d'application de l'ACP sur des données spectrales Raman 63
3.1.4. Limites de l'ACP 65
3.2. Régression linéaire au sens des moindres carrés (CLS) 65
3.2.1. Principe de la CLS 66
3.2.2. Exemples d'application de la CLS 67
3.2.3. Limites de la CLS 69
3.3. Conclusion 69
Conclusion **69**

INTRODUCTION

Comme cela a été présenté dans le Chapitre I, les spectres Raman constituent une mine d'informations remarquable au niveau cellulaire et moléculaire. Toutefois, ils sont complexes et difficiles à analyser en raison de leurs très nombreuses variables spectrales ayant un sens chimique.

En spectroscopie conventionnelle, il est courant d'analyser ces données spectrales de manière univariée. L'évolution d'une intensité spectrale ou d'un rapport de bandes permet de suivre l'évolution de la composition et de la structure moléculaire d'un tissu biologique. Néanmoins, cette manière d'analyser les données n'exploite pas tout le potentiel des données spectrales à disposition des scientifiques. En effet, les spectres sont par définition des données dites multivariées. Par conséquent, une analyse simultanée de la totalité des informations contenues dans ces multiples variables ne peut que perfectionner le suivi de la pénétration d'une molécule à travers la peau. Ce type d'analyse est alors dite multivariée au sens large et chimiométrique si les données sont chimiquement sensées. D'ailleurs, les progrès apportés dans l'instrumentation spectroscopique offrant de grandes quantités de données de plus en plus complexes en un temps relativement rapide a suscité le développement d'analyses chimiométriques pour l'extraction d'informations biochimiques pertinentes [26]. La chimiométrie emprunte aux analyses multivariées classiques l'utilisation de méthodes mathématiques et statistiques pour surveiller les variations biochimiques au sein des données et pour révéler les similarités qui les lient entre elles. De nombreuses méthodes d'analyses multivariées permettant la classification ou la modélisation des données ont été utilisées pour suivre la progression d'une molécule à travers les différentes couches de la peau par spectroscopie Raman.

Ce Chapitre II constitue un état-de-l'art des méthodes utilisées pour le suivi des modifications moléculaires en spectroscopie Raman dans des tissus biologiques et des matériaux chimiques. Une première partie est consacrée aux études ayant employé des analyses de données univariées. La deuxième et troisième partie, quant à elles, sont dédiées aux études ayant utilisé des analyses de données multivariées. La deuxième partie se concentre tout particulièrement sur les méthodes de classification tandis que la troisième partie se focalise sur les méthodes de modélisation ou de démélange spectral. Dans chacune de ces trois parties, le principe des méthodes les plus couramment utilisées en spectroscopie Raman est détaillé et leurs limitations sont mentionnées.

1. ANALYSE BASÉE SUR L'INTENSITÉ DES BANDES RAMAN

1.1. SUIVI D'INTENSITÉ DES BANDES SPECTRALES

Lorsque la nature des changements de composition chimique est connue, c'est-à-dire quand la variable spectrale permettant l'identification du composé à l'origine du changement est identifiée, la génération de la carte ou du profil d'intensités spectrales peut fournir des informations sur la distribution spatiale ou la temporalité de la perturbation en question. Néanmoins, il est indispensable que la variation d'intensité induite soit dépendante uniquement de la concentration relative du ou des composés responsables de cette perturbation. Elle ne doit pas être due à un quelconque chevauchement de bandes spectrales, auquel cas les informations semi-quantitatives analysées seraient biaisées.

De nombreux exemples de suivi d'intensités de variables spectrales existent dans la littérature de la spectroscopie Raman. Pour exemple dans le domaine de la recherche sur la peau, A. Tfayli *et al.* [21] ont suivi en temps réel l'expression des bandes Raman caractéristiques du métronidazole, un agent thérapeutique contre la rosacée, à travers les couches de la peau humaine pour étudier sa pénétration.

La détermination des profils de variations d'intensité des bandes situées à 1191 cm^{-1} et 1368 cm^{-1} en fonction de la profondeur a permis de confirmer une augmentation de la teneur en métronidazole dans les couches inférieures de l'épiderme deux heures après son application locale. Toutefois, la sélection d'une seule variable spectrale pour le suivi d'un changement biochimique peut causer une perte d'informations sur l'évolution de sa largeur de bande. C'est pourquoi il peut être préférable dans certains cas de suivre la contribution d'une plage spectrale par le biais d'un calcul des intensités intégrées, c'est-à-dire en calculant l'aire sous la courbe du spectre au niveau des bandes Raman sélectionnées. D'ailleurs, S. Tfaili *et al.* [36] ont, à l'aide de la microscopie confocale Raman et en appliquant ce procédé, suivi les modifications produites dans la peau humaine par des solutions faiblement concentrées en resvératrol et en caféine à l'instar des quantités utilisées dans certaines formulations dermatologiques. Pour évaluer leur progression à travers la peau, les bandes spectrales autour de la caféine ([550-558] cm^{-1}) et du resvératrol ([992-1006] cm^{-1}) ont été intégrées et ont permis de déterminer la cinétique de pénétration cutanée de ces molécules qui démontre ainsi la complexité de leur diffusion (Figure II-1).

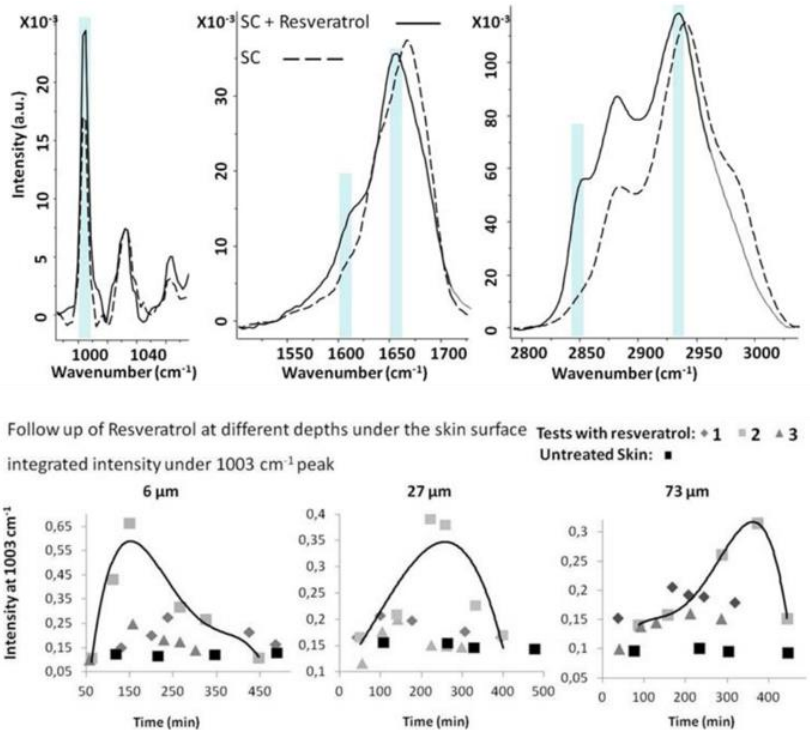


Figure II-1 – Détection des bandes associées au resvératrol et évolution de l'intensité intégrée de la bande située à 1003 cm^{-1} à 6, 27 et 73 µm de profondeur en fonction temps [36]

1.2. SUIVI DES RAPPORTS DE BANDES SPECTRALES

Une autre façon d'analyser un changement moléculaire spécifique est d'étudier le comportement de sa bande diagnostic par rapport à une valeur de référence interne. Cela consiste à analyser l'évolution du rapport de bandes entre la bande de l'actif ou du marqueur biochimique et une bande pseudo-constante spécifique à l'environnement étudié. Par exemple, en comparant la peau de porc à la peau humaine à trois niveaux différents qui sont le SC, l'épiderme et la jonction SC/épiderme, S. Tfaili *et al.* [34] ont montré que le ratio des bandes constituées de lipides et de protéines situées à 2850 cm^{-1} et à 2880cm^{-1} est le meilleur marqueur les discriminant. De la même manière, M. Janssens *et al.* [37] ont montré que la rapport lipides/protéines est un indicateur de l'affaiblissement de la fonction barrière cutanée. Plus précisément, l'étude révèle que la réduction du rapport lipides/protéines est plus prononcée dans les tissus cutanés lésés et non lésés de patients atteints d'eczéma atopique que dans

le SC de tissus cutanés de patients témoins et que celle-ci est corrélée avec la gravité de la maladie et la fonction barrière de la peau. Enfin, C. Alonso *et al.* [38] ont à leur tour analysé la pénétration cutanée de la caféine. Seulement cette fois, les profils de pénétration semi-quantitatifs ont été déterminés en calculant des rapports de bandes. En effet, la progression de la caféine dans les couches plus profondes de la peau a été évaluée en analysant son comportant par rapport à des bandes de référence spécifiques à la peau telles que la bande des acides aminés et la bande amide I située respectivement à 1004 cm^{-1} et 1660 cm^{-1} (Figure II-2).

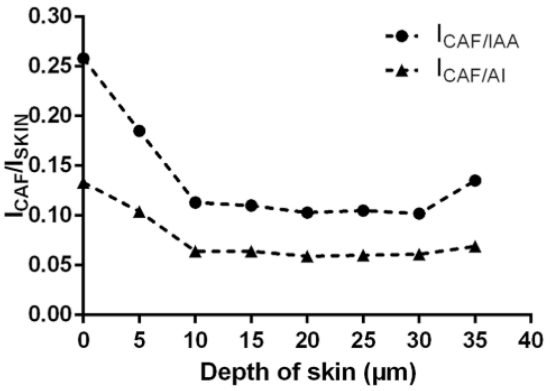


Figure II-2 – Profils de pénétration semi-quantitatifs de la caféine dans la peau de porc. Les profils sont obtenus en calculant les ratios d'intensités de bandes associées à la caféine (CAF, 556 cm^{-1}) et aux acides aminés (AA, 1004 cm^{-1}) ou à la bande amide I (AI, 1660 cm^{-1}) pour différents niveaux de profondeur [38].

Par ailleurs, il est également possible de calculer des rapports de bandes en utilisant des régions spectrales plutôt qu'une variable spectrale spécifique. Ainsi, l'évolution des largeurs de bandes dans le suivi des perturbations moléculaires est à nouveau prise en compte. Par exemple, l'article de P. J. Caspers *et al.* [39] est une référence très populaire pour la détermination de la concentration en eau dans la peau par le calcul du rapport de la bande de l'eau et des protéines situées respectivement à 3390 cm^{-1} et 2935 cm^{-1}, et ceci en utilisant les régions spectrales suivantes : [3350-3550] cm^{-1} et [2910-2965] cm^{-1}. De la même façon, P. J. Caspers *et al.* [40] réutilise ce procédé pour suivre un activateur de pénétration dans la peau, le diméthylsulfoxyde. Les profils de pénétration déterminés par les rapports d'intensités spectrales intégrées autour de la bande de cette molécule avec une bande de protéine révèlent la progression de l'actif à travers le SC (Figure II-3). De manière similaire, pour déterminer l'influence des paramètres de taille, de charge et d'élasticité des liposomes multi-lamellaires sur leur pénétration dans et à travers des modèles de peaux *in vitro* et *ex vivo,* R. Touti *et al.* [41] ont estimé leur profil de pénétration en calculant le ratio des bandes associées aux lipides et des bandes pseudo-constantes spécifiques aux modèles de peau. Pour finir, G. Zhang *et al.*[22] ont suivi la perméation d'un dérivé du resvératrol en calculant sa concentration relative en fonction du temps et de la profondeur sur la base de ratios resvératrol/phénylalanine. Ces ratios ont été calculés à partir des intensités intégrées des demi-intervalles autour de leurs bandes caractéristiques en raison de la présence d'un chevauchement de bandes. La construction de leur carte de distribution a permis de vérifier l'homogénéité de la diffusion du resvératrol à travers le SC et l'épiderme viable et d'observer la dépendance temporelle de sa perméation.

1.3. CONCLUSION

Le calcul des rapports de bandes, correspondant à des variables spectrales seules ou à un petit nombre d'entre elles se succédant, permet de suivre l'évolution des changements moléculaires qu'elles caractérisent. Cependant, un pareil suivi n'est possible que si la nature du changement biochimique est connue et que la variable illustrant cette perturbation est identifiée, ce qui n'est pas toujours le cas par exemple dans le domaine de la recherche biomédicale. De plus, il n'est jamais absolument certain que la perturbation biochimique observée à la bande de diagnostic soit exclusivement due au composé qui en est à l'origine. Des chevauchements avec d'autres composés pouvant être inconnus ou inattendus peuvent survenir et des déplacements ou des élargissements de bandes peuvent être mal

interprétés, voire même considérés comme des changements d'intensité. Au final, ces éléments ne sont pas pris en compte dans le calcul des rapports de bandes ce qui mène à des résultats faussés. La sélection d'une seule variable ou un petit nombre de variables successives par rapport à la quantité de variables disponibles n'est en général pas suffisante pour un diagnostic précis des perturbations moléculaires. Par conséquent, il est préférable de considérer un ensemble constitué d'un grand nombre de variables pour comprendre leur évolution les unes par rapport aux autres. De nombreuses méthodes capables d'analyser un grand nombre de variables simultanément existent. Elles sont définies comme des méthodes d'analyse de données dites multivariées.

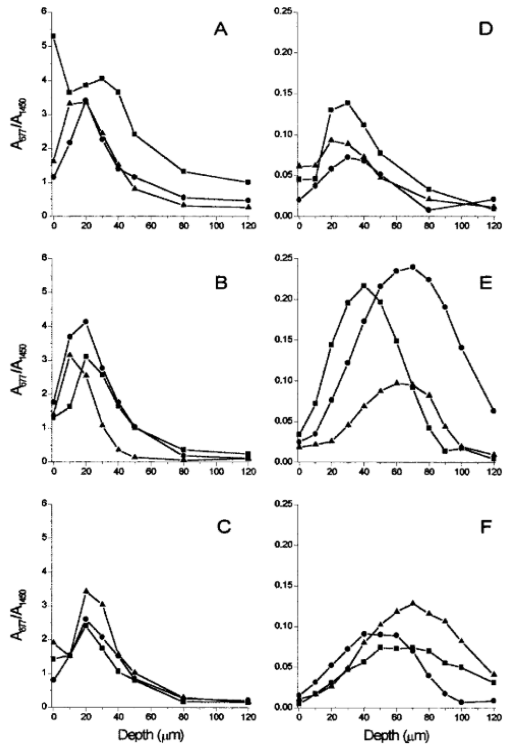


*Figure II-3 – Distribution du diméthyl sulfoxyde (DMSO) dans des échantillons de couche cornée de paume de main. Rapport de bandes intégrées DMSO/protéines autour de 677 et 1450 cm^{-1} en fonction de la profondeur. (**A**) 15-45 min ; (**B**) 150-170 min ; (**C**) ~4 h ; (**D**) ~22 h ; (**E**) ~48 h ; (**F**) ~72 h après application* [40].

2. MÉTHODES DE CLASSIFICATION

Le but des méthodes d'analyse de données multivariées est de décrire les variations des variables individuelles et les liens qui les unissent. En spectroscopie, cela consiste à décomplexifier les données spectrales pour extraire et interpréter des informations significatives. Plus particulièrement, ces méthodes permettent d'étudier les changements subtils présents dans les spectres. Certaines d'entre elles procèdent par classification et le couplage de la microspectroscopie confocale Raman avec des méthodes de classification semble être un outil prometteur pour la différenciation des modifications biochimiques produites dans les couches de la peau.

Les méthodes de classification ou de clustering sont des outils d'exploration de données. Elles ont pour objectif de déterminer les tendances de regroupements naturels des données sur la base de mesures de similitudes et de les classer selon leur proximité en différents groupes, appelés aussi clusters [26], [42]. Plus précisément, elles consistent à créer un partitionnement d'un ensemble de données en sous-ensembles de sorte que les différences entre les données au sein de chaque cluster soient minimisées et que les différences entre les clusters soient maximisées selon une distance définie [25]. D'un point de vue spectral, la construction d'une partition optimale d'un ensemble de spectres en un nombre de cluster défini repose sur l'identification des groupes de spectres ayant des caractéristiques spectrales proches témoignant de la similitude des propriétés moléculaires des données [42], [43].

Par ailleurs, il est souvent indispensable que des techniques de réduction de dimensions telle que l'ACP, présentée ci-après dans le Chapitre II.3.1, soient appliquées en amont d'une méthode de classification lorsque les données à analyser sont de grande dimension [26], ce qui est le cas des données spectroscopiques. Cette étape permet de simplifier les informations contenues dans les données et de faciliter leur classification. D'autre part, le clustering des données est dépendant de plusieurs paramètres. En effet, le nombre de clusters à trouver et leur initialisation, le choix de la mesure de distance utilisée pour évaluer les similitudes et les différences entre les données ou encore la définition du critère d'arrêt de classification déterminent le partitionnement des données final [26]. Ce dernier critère peut correspondre à des distances intragroupe et intergroupe ou un nombre maximal d'itérations à atteindre, ou bien encore à l'instant où plus aucun changement supplémentaire n'est opéré dans les clusters estimés. De plus, le clustering de données peut être réalisé de manière non supervisée ou supervisée. En effet, des connaissances préalables sur l'appartenance de données à différentes classes offrent la possibilité de construire un modèle d'apprentissage qui servira à automatiser l'affectation ou la non-affectation d'une nouvelle donnée à un de ces clusters. La validation de ces attributions prédictives est ensuite réalisée à l'aide de tests statistiques à l'exemple de la validation croisée.

De nombreuses méthodes de classification existent pour la classification de données spectrales. Les plus couramment utilisées en spectroscopie Raman sont la méthode K-means et la Classification Hiérarchique Ascendante (CHA). Les détails du fonctionnement de ces méthodes et des exemples de leur application sur des données Raman sont donnés dans les parties suivantes.

2.1. K-MEANS

2.1.1. Principe de la méthode des K-means

La méthode des K-means est une méthode de clustering non supervisée. Aucune connaissance *a priori* sur l'affectation des données à des classes n'est requise et la classification se fait uniquement en fonction des valeurs des données [26], [43]. Cette méthode est réputée pour être rapide et capable de traiter des données de grandes dimensions, raison pour laquelle elle est fréquemment utilisée en chimiométrie et plus particulièrement pour l'analyse d'images hyperspectrales [35], [44], [45]. Son principe est simple et repose sur une procédure itérative qui permet de trouver une partition optimale des données en K clusters, avec $K \in \mathbb{N}$ défini au préalable par l'opérateur [26], [42]–[44]. En effet, l'algorithme débute par l'initialisation aléatoire de K clusters de départ constitués chacun par un point de données choisi parmi l'ensemble des données. Une fois les K classes initialisées, chaque point de données est passé en revue et affecté au cluster le plus proche en terme de distance, qui est le plus couramment la distance euclidienne. Puis, après chaque affectation individuelle, le barycentre du cluster dans lequel le point de données a été affecté est recalculé. En spectroscopie, celui-ci s'apparente à la moyenne des spectres constituant le cluster en question. Cette moyenne est ensuite utilisée comme nouvel élément de comparaison dans le calcul des distances pour les affectations suivantes. Lorsque toutes les données ont été affectées à une classe, le processus est réitéré de nombreuses fois afin de vérifier si les données sont plus proches d'un autre cluster que celui qui leur a été attribué ou non. Enfin, l'algorithme s'arrête lorsque plus aucun changement d'affectation n'est réalisé ou quand le nombre maximum d'itérations permises de la procédure est atteint, ce qui témoigne l'atteinte d'une solution stable.

2.1.2. Interprétation des résultats de la méthode K-means

La méthode K-means a pour objectif final de résumer une grande quantité de données en un nombre défini de classes qui sont synthétisées par leur moyenne. En spectroscopie, les résultats s'interprètent comme le regroupement de spectres présentant les mêmes caractéristiques majeures, c'est-à-dire

partageant les mêmes propriétés moléculaires, et dont les clusters associés sont représentés respectivement par leur moyenne spectrale [35], [43], [45]. Quant aux informations sur l'appartenance des données aux clusters, chaque point de données se voit attribuer un label, un chiffre ou une couleur, spécifique à la classe à laquelle il appartient. Cette labellisation rend possible la localisation des différents groupes de données grâce à la construction de cartes d'appartenance aux clusters en fausses couleurs qui représentent la distribution spatiale des données de chacun d'entre eux [35], [43], [45].

2.1.3. Exemples d'application de la méthode K-means sur des données spectrales Raman

Dans la littérature de la spectroscopie Raman, la méthode K-means a été utilisée pour analyser et classifier des produits pharmaceutiques inconnus. L'enjeu derrière l'utilisation de cette méthode de clustering est la lutte contre la contrefaçon de produits médicamenteux et la vente de produits illégaux. L'étude menée par P.Y. Sacré *et al.* [46] montre que cette méthode peut être employée pour discriminer les images spectrales Raman enregistrées à partir de comprimés de Viagra® authentiques de celles obtenues à partir de comprimés falsifiés. De plus, le taux de réussite de classification de ces données par validation croisée est de 100%. Par conséquent, la méthode K-means peut être considérée comme un outil de détection fiable de produits pharmaceutiques contrefaits.

La méthode K-means a également été souvent utilisée dans le domaine de la biologie et de la recherche biomédicale. Son application sur des données spectrales Raman de cellules biologiques permet de comprendre leur morphologie et d'identifier les caractéristiques des différents éléments qui la composent. Par exemple, C. Krafft *et al.* [43] ont étudié la composition chimique et la structure moléculaire de fibroblastes pulmonaires soumis ou non à des conditions de stress oxydatif provoquant la mort cellulaire. L'application de la méthode K-means a permis de générer des images pseudo-colorées de la morphologie des cellules au cours de leur cycle de vie et d'en observer les changements moléculaires et structurels caractéristiques. De même, F. Bonnier et H. J. Byrne [47] ont utilisé la méthode des K-means pour analyser et identifier les différents composants d'une cellule pulmonaire cancéreuse, à savoir le cytoplasme, le noyau et le nucléole. Pour aller plus loin, Dorney *et al.* [48] ont employé la méthode K-means pour localiser et suivre des nanoparticules dans des cellules pulmonaires cancéreuses.

Les résultats obtenus en utilisant la méthode K-means dans les domaines pharmaceutiques et de la recherche biomédicale ont suscité l'intérêt de son application pour étudier la composition moléculaire et la morphologie des tissus cutanés et analyser le comportement de la peau lorsque des molécules exogènes lui sont appliquées à la surface. Il a été démontré dans un premier temps dans les études réalisées par S.M. Ali *et al.* [35], [45] et L. Miloudi *et al.* [44] que la méthode K-means est capable d'identifier les différentes structures présentes dans la peau, à savoir les couches dermiques, la couche cornée et les autres sous-structures de l'épiderme, sur la base des variations moléculaires caractérisées dans les spectres Raman (Figure II-4). Puis, la méthode K-means a également été utilisée pour l'analyse et le suivi de molécules exogènes à travers la peau à des fins cosmétiques et de traitements médicamenteux par voie cutanée. Par exemple, D. Lunter *et al.* [49] ont étudié l'influence de deux activateurs de pénétration cutanée sur la distribution d'un médicament, un anesthésique local qui ne pénètre pas facilement dans la peau connu sous le nom de la procaïne, à travers la peau de porc. Dans cette étude, la méthode K-means a été sensible aux différentes teneurs en ingrédient actif à une échelle thérapeutiquement attendue dans l'administration cutanée puisqu'elle a été capable de localiser la procaïne dans les différentes couches de la peau selon l'activateur de pénétration utilisé. En effet, en l'absence d'exhausteur de pénétration ou en solution avec du propylène glycol, la procaïne stagne dans la couche cornée alors qu'en solution avec du polyoxyéthylène-23-lauryl éther la procaïne progresse dans les couches plus profondes de la peau.

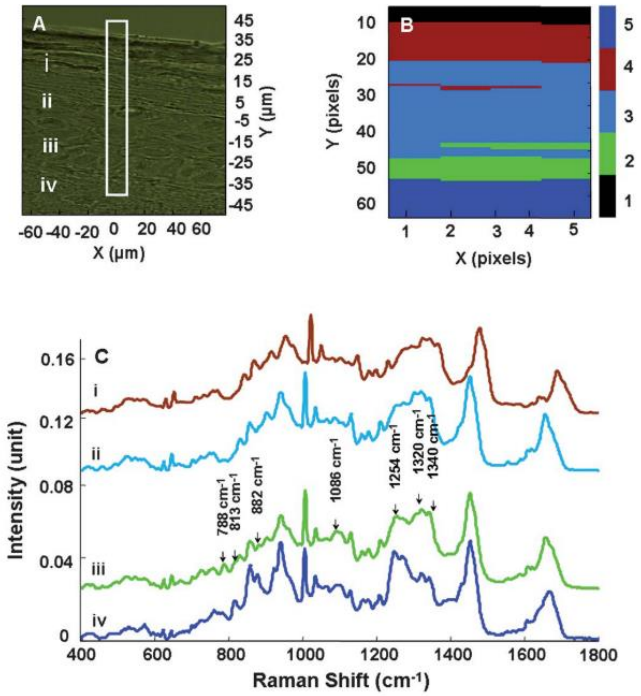


Figure II-4 – (**A**) *Image optique d'une section de tissu artificiel non exposée ;* (**B**) *Résultat de la méthode K-means de la carte Raman associée ;* (**C**) *Moyennes spectrales de tissu cutané artificiel des* (**i**) *cluster 4, stratum corneum,* (**ii**) *cluster 3, stratum granulosum,* (**iii**) *cluster 2, couche basale,* (**iv**) *cluster 5, derme* [45].

2.1.4. Limites de la méthode K-means

La méthode K-means permet de classer des données selon leurs similitudes et leurs différences. Elle simplifie et résume l'information sous forme de moyennes caractéristiques et d'appartenance à des clusters représentés par ces moyennes. Cependant, la représentation d'un cluster par la moyenne de tous les points qui le composent est problématique puisqu'elle ne laisse pas de place à une intravariabilté au sein du groupe [42], [44]. Par exemple, des données spectrales peuvent être classées dans des groupes distincts alors qu'elles présentent chacune la même propriété moléculaire mais à des quantités différentes. D'autre part, les résultats obtenus par cette méthode sont très dépendants du nombre de groupes choisi *a priori* et de leur initialisation [26], [42]. En effet, l'initialisation pose problème au niveau de la reproductibilité de la méthode en raison de son caractère aléatoire. C'est pourquoi il est courant d'appliquer la méthode plusieurs fois de suite afin de s'assurer qu'elle converge vers la même solution malgré des initialisations différentes. Pour finir, la méthode K-means suppose qu'un point de donnée ne peut appartenir qu'à un seul cluster à la fois [42]. Cependant en spectroscopie, un spectre peut être le reflet d'un certain nombre d'entités biologiques différentes en quantités différentes. Il peut par conséquent appartenir à différents clusters qui sont chacun caractéristique d'un composant biochimique tel que les lipides ou encore les protéines. L'utilisation de la méthode K-means peut donc être responsable d'une perte d'informations importante en raison de son algorithme trop rigide.

2.2. CLASSIFICATION HIÉRARCHIQUE

2.2.1. Principe de la classification hiérarchique

La méthode de classification hiérarchique est une méthode de clustering non supervisée régulièrement utilisée en chimiométrie. Elle nécessite aucune connaissance *a priori* sur le nombre de groupes à définir et peut être appliquée sans avoir aucune référence sur l'appartenance des données à des classes particulières. Comme toute méthode de classification, son objectif est de regrouper des données en clusters en fonction de leurs similitudes ou de leurs différences. Pour cela, elle opère selon deux

modes : le mode ascendant et le mode descendant [26], [42]. Toutefois, ces deux modes ne sont pas équivalents [26] et la Classification Hiérarchique Descendante (CHD) est informatiquement plus intense que la Classification Hiérarchique Ascendante (CHA) expliquant pourquoi cette dernière est la plus employée des deux. Le principe de la CHA repose sur un processus agglomérant. Chaque point de données est attribué à un cluster différent, autrement dit il y a au départ autant de clusters qu'il y a de points de données (nombre égal à $N \in \mathbb{N}$). Les points de données sont ensuite regroupés successivement en différents clusters jusqu'à ce qu'il ne reste plus qu'un seul groupe contenant l'ensemble des points de données [26], [42]. À l'inverse, la CHD agit en divisant. Au début du processus, tous les points sont regroupés dans un seul et même cluster. Les points sont ensuite séparés de manière itérative en différents clusters jusqu'à ce qu'ils soient tous dissociés les uns des autres [26], [42].

Les résultats obtenus par classification hiérarchique se représentent sous la forme d'un dendogramme bidimensionnel où les regroupements de données itératifs sont illustrés en fonction des distances qui séparent les groupes de données [25]. Ces dendogrammes sont construits en fonction de la manière dont les classes sont liées entre elles pour le mode ascendant et de la façon dont les classes sont séparées les unes des autres pour le mode descendant (Figure II-5) [42]. Le choix final du nombre de classes différentes est défini en sectionnant l'arbre du dendogramme en deux parties à un stade précis. Comme la procédure de clustering se réalise en $N - 1$ itérations au total, fractionner le dendogramme au niveau de la $(N - K)^{\text{ème}}$ itération pour le mode ascendant permet d'obtenir le partitionnement de l'ensemble des données en K classes.

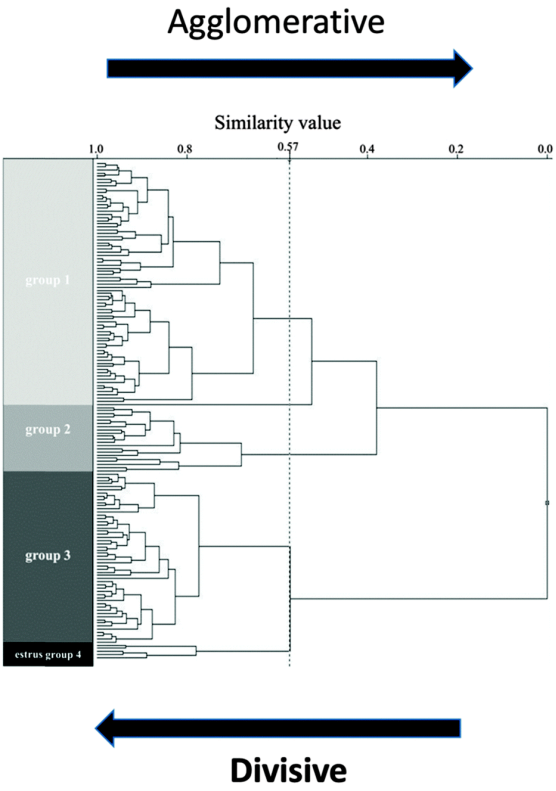


Figure II-5 – Illustration d'un dendogramme issu d'une CHA et d'une CHD [42].

Le regroupement ou la séparation des données en clusters est basé sur la mesure de leur similarité. Cette similarité est généralement exprimée sous la forme d'une matrice symétrique définissant la distance ou la corrélation entre les données deux à deux [25], [42]. Ensuite, dans un mode ascendant, le rassemblement des données en cluster se réalisent selon un critère prédéfini. Ce critère correspond généralement à la minimisation de la distance, le plus souvent euclidienne, entre deux points de

données. Les points de données les plus proches sont donc regroupés dans une même classe et la matrice de similarité par paire entre chacun des points de données restants et les clusters déjà déterminés est recalculée [25], [26]. L'avantage de cette façon de procéder réside dans le fait qu'un point est assigné à un cluster une seule fois [42], contrairement à la méthode K-means où le cluster d'appartenance d'une donnée peut être modifié plusieurs fois. Par conséquent, la CHA bénéficie d'un temps de calcul faible lorsque les données ne sont pas de trop grande dimension.

2.2.2. Exemples d'application de la CHA

En spectroscopie Raman, la CHA a été particulièrement utilisée dans le domaine pharmaceutique. Dans le cadre de la lutte contre la falsification de médicaments et la vente de produits illégaux, l'utilisation de la CHA a permis de discriminer des produits authentiques de ceux contrefaits. Par exemple, dans les études réalisées par M. de Veij *et al.* [50], [51], des comprimés antipaludiques falsifiés et des comprimés de Viagra® ont été différenciés de leurs versions authentiques à l'aide de la CHA.

D'autre part, la CHA a également été appliquée sur des données spectroscopiques Raman de tissus biologiques dans le but de diagnostiquer des pathologies. Par exemple, O. Piot *et al.* [52] ont utilisé la CHA sur des données Raman enregistrées à partir d'échantillons provenant d'aorte humaine *ex vivo*. Deux principaux groupes correspondant aux deux parties anatomiques de l'aorte, la partie ascendante et l'arcade, ont été distingués. Il a aussi été démontré que le groupe correspondant à la partie ascendante de l'aorte peut être divisé en deux sous-groupes. Ces deux sous-groupes pourraient correspondre à des compositions et des structures tissulaires différentes dont la caractérisation pourrait aider à évaluer les risques d'anévrisme pendant une intervention chirurgicale. Cette étude a également montré l'efficacité de la CHA pour différencier des peaux normales et des peaux cancéreuses.

2.2.3. Limites de la CHA

La CHA permet de partitionner un ensemble de données en différents clusters sans aucun *a priori* sur leur nombre et en un nombre d'itérations proportionnel à la quantité de données à classer. Cependant, les résultats obtenus sont, comme toute méthode de classification, dépendants du critère de regroupement choisi. De plus, la CHA suppose qu'un point de donnée ne peut être attribué qu'à un seul cluster [42] ce qui conteste la possibilité qu'un spectre puisse être composé de plusieurs entités biologiques en des quantités différentes. L'utilisation de la CHA expose donc à un risque de perte d'informations.

2.3. CONCLUSION

En spectroscopie Raman, les méthodes de classification ont pour objectifs de regrouper les données spectrales selon leur similarité et d'identifier les variables spectrales permettant la discrimination des différents groupes. En ce qui concerne les recherches scientifiques menées sur la peau, l'analyse des clusters obtenus par des méthodes de classification a permis de comprendre sa composition et sa structure biochimique. De plus, la localisation de certaines molécules exogènes a été possible grâce au partitionnement des données sur la base de l'expression des modifications induites dans les spectres par l'application locale de ces molécules. Cependant, les méthodes de clustering décrites ci-dessus n'offrent pas la possibilité à un spectre Raman d'appartenir à plusieurs groupes de données en même temps. Cette idée est contraire au principe qu'un spectre représente un mélange de plusieurs composants biochimiques. Théoriquement, un spectre devrait pouvoir être affecté à plusieurs clusters simultanément. D'autre part, les méthodes de clustering cherchent par définition à minimiser la variance intragroupe et maximiser la variance intergroupe. Toutefois, cette exigence empêche deux spectres contenant le même composant à des concentrations très différentes à appartenir au même groupe en raison de cette variabilité. Par conséquent, les méthodes de classification sont en règle

générale jugées trop strictes puisqu'elles ne laissent pas de place à la variabilité intragroupe. Elles ne sont donc pas adaptées au suivi de la cinétique de pénétration de molécules exogènes. En revanche, des méthodes d'analyse multivariée de démélange sembleraient être plus appropriées pour suivre les cinétiques de pénétration de molécules exogènes.

3. MÉTHODES DE DÉMÉLANGE

En spectroscopie Raman, les méthodes d'analyse de données multivariées sont idéales pour décrypter les informations liées aux changements de la composition chimique exprimés dans les spectres et pour révéler les corrélations insoupçonnées entre variables spectrales. Parmi ces méthodes, certaines reproduisent les données à partir d'un modèle estimé plus simple. Ce sont les méthodes de modélisation ou de démélange spectral. Ces méthodes sembleraient être plus appropriées pour suivre les cinétiques de pénétration de molécules exogènes telles que les actifs dermocosmétiques. En effet, en trouvant un modèle simplifié capable de reproduire les données, elles donnent la possibilité à un spectre d'être expliqué par différentes variations spectrales et laissent ainsi la place à l'expression de nombreuses variations liant plusieurs spectres entre eux. Le couplage de la microspectroscopie confocale Raman avec des méthodes d'analyse multivariée dite de démélange représente donc un outil prometteur pour l'étude des modifications moléculaires et le suivi de la pénétration d'actifs dermocosmétiques dans la peau.

Les méthodes de démélange spectral les plus fréquemment utilisées en spectroscopie Raman sont l'Analyse en Composante Principale (ACP) et la régression linéaire au sens des moindres carrés (CLS). Les détails de leur fonctionnement et des exemples de leur application sur des données Raman sont donnés dans les parties suivantes.

3.1. ANALYSE EN COMPOSANTES PRINCIPALES (ACP)

L'analyse en composantes principales (ACP) est une approche statistique multivariée descriptive. Elle est la technique la plus utilisée en chimiométrie et est souvent la première méthode appliquée à un ensemble de données pour explorer leurs caractéristiques et en savoir plus sur leur répartition [26], [53]. Elle a l'avantage de pouvoir être appliquée de manière non supervisée, c'est-à-dire être appliquée sans avoir aucune connaissance préalable sur la nature des données ou même sur les relations qui les lient ou les différencient [25].
La description de la distribution des données et la connaissance de leurs différences et leurs similitudes sont les raisons principales de l'application de l'ACP [26], [28], [53]–[55]. L'ACP est également employée pour la compression de données en sélectionnant les variables pertinentes pour leur description [26], [28], [42], [45], usage aussi connu sous l'expression de la réduction de la dimensionnalité des données. D'autre part, l'application de l'ACP est couramment utilisée dans le prétraitement des données spectrales [26], [28] puisqu'elle offre la possibilité de réduire le bruit et de supprimer des valeurs jugées comme aberrantes au sein des données spectrales (plus de détails dans le Chapitre III). Enfin, les résultats d'une ACP sont souvent combinés à des méthodes de classification pour simplifier l'information de données complexes à dimension multiples et faciliter le processus de discrimination. Ils sont aussi fréquemment utilisés dans d'autres méthodes d'analyse multivariée en tant qu'étape d'initialisation ou de critère d'arrêt au regard du pourcentage de variance exprimée par une modélisation des données. À noter que d'un point de vue spectral, l'ACP révèle des variations d'ordre physique et biochimique.

3.1.1. Principe de l'ACP

L'objectif commun à ces nombreuses applications de l'ACP est la synthétisation et la simplification de l'information disponible dans les données avec le moins de perte d'informations possible. D'un point de vue géométrique, l'ACP consiste à considérer les données comme un nuage de points dans un plan

dont les axes de coordonnées sont la dimension et l'intensité spectrales puis à projeter leurs coordonnées dans un nouvel ensemble d'axes, appelées composantes principales (CP) [25], [26], [28], [55]. Ces axes sont construits de façon à maximiser la variance des données, autrement dit à maximiser la propagation des points de données dans le nouvel espace de dimension réduite. Cela signifie que les CP construites correspondent aux directions de plus grande dispersion des données expliquant la variance maximale. Cependant, il est indispensable que les CP soient orthogonales les unes aux autres [25], [26], [28]. Cette orthogonalité garantit la représentation de phénomènes indépendants et non corrélés entre eux par les CP. En d'autres mots, elle permet d'éviter la redondance des informations présentes dans les données au sein des CP. Au final, les CP fournissent des informations sur la source de variabilité au sein de l'ensemble des données et décrivent leur répartition. En spectroscopie, les CP sont assimilées à des spectres contenant des pics positifs et négatifs qui expliquent la variation spectrale dans l'ensemble des données et dont la signifiance est jugée par l'observation de leurs pics et du bruit [42].

Mathématiquement, le calcul d'une matrice variance-covariance V permet d'évaluer les liens entre les données [26]. La construction des CP décorrélées est obtenue par la diagonalisation de la matrice de variance-covariance des données dont l'expression est la suivante :

$$V_{n,n} = \frac{\widetilde{D}_{n,N}{}^{T}\,\widetilde{D}_{N,n}}{N-1} \tag{II-1}$$

avec $\widetilde{D}$ la matrice de données centrée à la moyenne, N le nombre de points de données, n l'index d'un point de données et l'exposant ^{T} la transposée.
La diagonalisation de cette matrice V simplifie l'information qu'elle contient en la décomposant en valeurs et vecteurs propres. Ces vecteurs propres sont une combinaison linéaire des variables d'origine et constituent un nouvel espace dont les axes de coordonnées ont des directions orthogonales les unes aux autres. Cela signifie qu'ils définissent des axes de dispersion maximale pour les données. Les CP sont alors définies comme les vecteurs propres non corrélés associés à la matrice de variance-covariance des données. Quant aux valeurs propres calculées, elles sont directement liées à la variance expliquée par leur vecteur propre associé, c'est-à-dire qu'elles reflètent la diversité d'information contenue dans les CP. Une valeur propre élevée traduit une forte variabilité contenue dans sa CP associée et témoigne de la capacité de cette CP à modéliser une grande partie de la variabilité des données [26]. A l'inverse, une valeur propre proche de zéro signifie que la CP qui lui est associée ne contient que très peu d'informations sur les tendances de variabilité expliquée par les données. Autrement dit sa CP associée exprime seulement des informations non-structurées et le bruit contenu dans les données. En pratique, les valeurs propres sont rangées par ordre décroissant. Ainsi la première CP, notée CP1, correspond à la direction du maximum de variance des données. Quant à la CP2, elle définit la direction de la deuxième plus grande variation des données sur la variance résiduelle, c'est-à-dire sur la variance non prise en compte par la précédente CP, la CP1. Ce processus est réitéré pour obtenir les CP suivantes. Enfin, les dernières CP caractérisent les faibles variations présentes dans les données qui sont souvent assimilées au bruit et peuvent être considérées comme négligeables.

Au final, cette décomposition en valeurs et vecteurs propres de la matrice de variance-covariance permet de modéliser une matrice de données par le produit de deux autres matrices composées par les CP pour la première et par les coordonnées des points de données dans ce nouvel espace de dispersion maximale défini par ces CP pour la seconde. Ces matrices sont communément appelées de manière respective matrice des *loadings* et matrice des *scores*. La différence entre la matrice des données et le produit des *loadings* et des *scores* permet de juger de la qualité de reproduction des données.

3.1.2. Interprétation des résultats de l'ACP

Les *loadings* représentent le lien entre les variables des données et leurs contributions dans le nouvel espace à dimension réduite. Plus précisément, ils sont considérés comme l'expression des variations communes aux données. Ainsi, la visualisation de ces *loadings* est source d'informations sur la répartition des données et les variables qui les différencient ou les unissent [26]. L'analyse des graphiques des *loadings* permet d'identifier les variables de diagnostic utiles à la classification des données dans des groupes distincts et la représentation des *loadings* les uns en fonction des autres permet d'établir des relations entre ces groupes [26], [54]. Quant aux *scores*, ils représentent la contribution des *loadings* pour chaque point de données et leurs graphiques peuvent révéler des tendances entre les points de données.

En spectroscopie, l'application d'une ACP sur un ensemble de spectres donne accès à des informations sur la répartition de leurs variables spectrales et sur leurs différences et similitudes d'un point de vue physique et chimique. Les graphiques des *loadings* mettent en avant les variables spectrales responsables des variations au sein des données. Ces dernières peuvent se rapporter à des régions spectrales de diagnostic utiles à la différenciation des sous-ensembles de spectres. De plus, les graphiques des *loadings* aident à la compréhension des relations présentes entre les sous-ensembles de spectres au sein des données. Ces graphiques se révèlent donc être un support à la classification de groupes de données spectrales et à la réduction de la dimensionnalité des données de par la sélection des régions spectrales pertinentes. Des informations complémentaires peuvent aussi être apportées en visualisant les graphiques des *scores* associés aux *loadings* qui, dans un cadre d'imagerie hyperspectrale, s'interprètent comme des cartes de distribution spatiale des variations spectrales représentées dans les *loadings*.
D'autre part, les informations d'intérêt en spectroscopie sont essentiellement contenues dans très peu de CP puisque la sélection des premières CP suffit à la compréhension de la variabilité spectrale générale des données. De ce fait, l'ACP peut être utilisée comme outil de réduction de bruit spectral et de diagnostic de données atypiques. En effet, le bruit correspondant à la partie non modélisée de la variance expliquée par les premières CP et ses variations étant contenues dans les dernières CP, sa correction peut s'effectuer par reconstruction des données en utilisant un nombre optimal de *loadings* comptant parmi les premiers. Cette utilisation de l'ACP est abordée de façon plus détaillée dans le Chapitre III. De même, les spectres jugés comme très différents de l'ensemble des données spectrales se caractérisent par une distance au modèle reconstruit très élevée. Un calcul des distances de chaque spectre par rapport au modèle des données reconstruites permet d'identifier les spectres aberrants, aussi appelés *outliers*, qui seront par la suite corrigés ou supprimés des données.

3.1.3. Exemples d'application de l'ACP sur des données spectrales Raman

Dans la littérature de la spectroscopie Raman, l'ACP a en premier lieu été utilisée dans le domaine pharmaceutique en combinaison à des méthodes de classification pour analyser des comprimés pharmaceutiques [46], [50], [51]. L'objectif était de caractériser la composition moléculaire de produits pharmaceutiques inconnus et de les discriminer selon leur contenu. Cette caractérisation passe par l'identification des ingrédients présents (principe actif et excipients), l'obtention d'informations sur la structure chimique de ces ingrédients et une éventuelle vérification de la teneur et de la répartition du principe actif à l'intérieur des produits. De telles informations sont d'une utilité précieuse non seulement pour le contrôle qualité des produits mais aussi dans la lutte contre les produits illégaux et la contrefaçon médicamenteuse. L'ACP a donc été combinée à des méthodes de classification telles que les méthodes CHA [50], [51] ou K-means [46] dans le but de décomplexifier les informations spectrales Raman ce qui, par conséquent, a facilité le processus de classification. Des exemples d'une

telle utilisation de l'ACP sont la distinction de comprimés antipaludiques [50] ou de comprimés de Viagra® [46], [51] authentiques et contrefaits.

L'ACP a également été souvent utilisée dans le domaine de la biologie et de la recherche biomédicale. Son application sur des données spectrales Raman permet d'étudier les variations au niveau moléculaire de cellules biologiques et de comprendre leur métabolisme. Par exemple, F. Bonnier et H. J. Byrne [47] ont appliqué l'ACP sur des images Raman de cellules pulmonaires cancéreuses pour identifier les changements biomoléculaires locaux des différents composants subcellulaires : le cytoplasme, le noyau et le nucléole. Cette étude souligne donc la capacité de l'ACP à identifier les variations spectrales à l'origine des changements biomoléculaires locaux dans l'environnement cellulaire et confirme l'intérêt de son utilisation dans la recherche biomédicale. Par conséquent, l'ACP peut être présentée comme un outil d'analyse du métabolisme cellulaire éventuellement modifié suite à l'interaction des cellules avec des agents externes. Un exemple d'application du suivi des changements métaboliques cellulaires causés par les interactions des cellules avec des agents externes est la localisation de nanoparticules dans des cellules pulmonaires cancéreuses sur la base de variations de signatures spectroscopiques moléculaire Raman [48]. Les études menées par Z. Farhane [56], [57] sont d'autres exemples de la mise en évidence de l'évolution métabolique des cellules pulmonaires cancéreuses suite à l'absorption d'une molécule externe, la doxorubicine qui est utilisée dans le traitement du cancer du poumon par chimiothérapie, par l'ACP. Semblablement, F. Draux *et al.* [58] ont employé l'ACP pour surveiller les effets toxiques de la gemcitabine, un médicament anti-tumoral aussi utilisé dans le traitement du cancer du poumon par chimiothérapie, au niveau cellulaire. Ces études montrent que l'ACP permet d'évaluer l'efficacité, les mécanismes d'action et de suivre l'interaction d'un médicament dans un environnement intracellulaire en élucidant les changements biomoléculaires locaux induits par son absorption via une analyse des variations des données spectroscopiques.

L'emploi de l'ACP dans les domaines pharmaceutiques et de la recherche biomédicale a motivé son utilisation pour étudier la composition moléculaire et la structure de la peau et comprendre sa réaction face à des facteurs endogènes et exogènes tels que l'âge, l'environnement extérieur ou encore des traitements par voie cutanée. Il a été démontré dans un premier temps que les variations biomoléculaires des différentes couches cutanées peuvent être mises en évidence par une ACP. En effet, l'étude proposée par S. Tfaili *et al.* [34] montre que la peau de porc est très proche de la peau humaine d'un point de vue biomoléculaire et structurel. Le but de cette étude était d'identifier les variations moléculaires communes aux couches cutanées des deux types de peau mais aussi de déterminer celles responsables de leurs différences. L'interprétation des résultats d'une ACP appliquée sur des données spectrales Raman enregistrées dans les régions épidermiques, dermiques et leur jonction a démontré à la fois le caractère alternatif de la peau de porc à l'utilisation de la peau humaine dans la recherche scientifique et le fait qu'une variabilité intragroupe et intergroupe de ces deux types de peaux persiste. De même, L. Miloudi *et al.* [44] ont utilisé l'ACP pour mettre en évidence les changements biochimiques entre les couches cutanées d'un épiderme humain reconstruit (Episkin®) et pour exposer l'intravariabilité des spectres provenant de la couche cornée qui est directement liée aux informations quantitatives sur les protéines et les lipides qu'ils contiennent. Ces études variées montrent donc la faculté de l'ACP à identifier les variations spectrales à l'origine des changements biomoléculaires locaux dans l'environnement cutané et justifient l'intérêt de son application pour l'analyse de la réaction de la peau face à des facteurs endogènes et exogènes. Par exemple, F.J. Gonzalez *et al.* [59] ont couplé l'ACP à la spectroscopie Raman pour mettre en évidence les causes du vieillissement cutané et caractériser plus particulièrement les effets chronologiques et les dommages photo-induits sur la structure protéique de la peau, témoin de son niveau d'hydratation. Les résultats de l'ACP ont précisément montré qu'il existe une corrélation entre les peaux exposées au soleil et le

pourcentage de collagène et d'élastine contenu dans la peau mais qu'aucun lien ne peut être établi entre ces mêmes peaux exposées au soleil et l'âge des sujets. Les effets de l'irradiation solaire sur la peau ont également été observées grâce à une ACP appliquée sur des données Raman d'échantillons de peaux artificielles exposés à des rayons UV dans l'étude menée par S.M. Ali [45]. La mise en évidence des variations biologiques dues à l'exposition de la peau aux rayonnements UV par l'application de l'ACP (Figure II-6) fait de cette méthode d'analyse multivariée un outil puissant pour le diagnostic précoce de la mort cellulaire induite par le rayonnement solaire dans la peau humaine. Enfin, une étude discriminant des spectres Raman enregistrés sur des peaux humaines avant et après application de produits de protection solaire a été réalisé en employant une ACP [54].

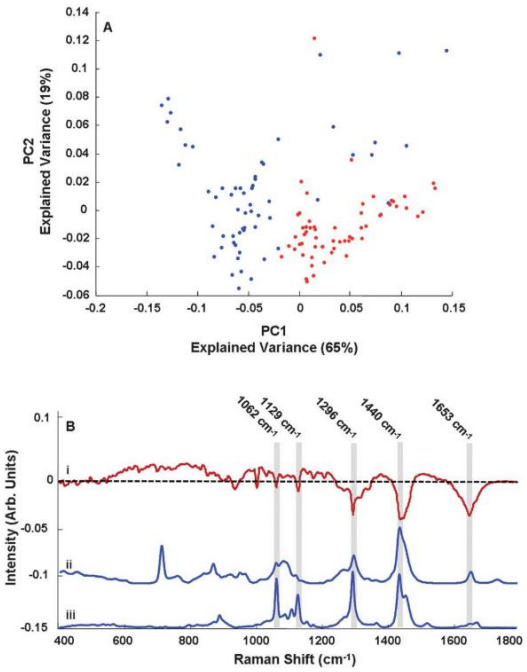


Figure II-6 – (**A**) *Nuage de points de l'ACP des spectres associés au stratum corneum de la section de tissu artificiel non exposée (rouge) et au stratum corneum de la section de tissu artificiel exposée pendant 240 minutes à des rayons UV (bleu).* (**B**) *Comparaison de la* (**i**) *CP1 par rapport au spectre Raman de* (**ii**) *L-α-phosphatidylcholine et* (**iii**) *de céramide. La ligne pointillée indique le point zéro dans le cas de la CP1* [45].

3.1.4. Limites de l'ACP

L'application de l'ACP permet d'identifier les variations biomoléculaires caractéristiques d'un ensemble de données spectrales. Dans le domaine de la recherche spectroscopique sur la peau, l'ACP permet d'identifier certains phénomènes physico-chimiques liés au vieillissement cutané ou à l'environnement extérieur. Pour ce faire, l'ACP fournit des axes de dispersion maximale des données afin de simplifier la compréhension des données et de mieux visualiser les différences spectrales. Cependant, les CP obtenues par une ACP soulignant les variables responsables de ces changements sont par définition supposées non-corrélées entre elles. Ces CP sont mathématiquement bien définies, mais n'ont pas de sens ou de signification physique particulière et leur interprétation d'un point de vue chimique peut être difficile [53], [55]. En effet, elles correspondent à une combinaison de caractéristiques spectrales provenant de plusieurs composés chimiques dont les distributions spectrales sont possiblement corrélées entre elles, ce qui rend la comparaison avec une référence spectrale difficile. Au final, les multiples modifications de la composition moléculaire des différents composants de la peau suite à son exposition à une molécule complexe rend l'identification de cette molécule à travers la peau délicate et son suivi à l'aide d'une ACP compliqués.

3.2. RÉGRESSION LINÉAIRE AU SENS DES MOINDRES CARRÉS (CLS)

La régression au sens des moindres carrés classique, traduit en anglais par Classical Least Square (CLS), est une méthode d'analyse multivariée très populaire en chimiométrie. Par définition, elle est capable

d'analyser de manière simultanée un ensemble de variables communes à des données. D'un point de vue spectroscopique, cela signifie qu'elle peut analyser les intensités de l'ensemble des variables spectrales de nombreux spectres en même temps.

La CLS a pour objectif d'extraire des informations semi-quantitatives sur les composants d'un mélange. Pour cela, elle requiert la connaissance *a priori* des composants biochimiques purs qui composent le mélange pour estimer leur contribution respective dans celui-ci. Cela fait d'elle une méthode multivariée supervisée [42], [44], [60]. À partir des connaissances préalables sur les composants du mélange étudié, la CLS estime leurs concentrations respectives sur la base d'un ajustement d'une combinaison linéaire de ces références sur le mélange expérimental [42], [60]. Cet ajustement consiste à estimer les concentrations de chaque composant de telle manière que le mélange modélisé soit aussi proche que possible du mélange expérimental avec un résidu minimal. Cette estimation est réalisée plus précisément par approximation au sens des moindres carrés [44], [55], [61]. Dans le cadre de l'imagerie hyperspectrale, les coefficients de contribution estimés correspondent aux concentrations relatives locales de chaque référence spectrale et ceux-ci peuvent être représentés sous forme de carte de distribution. Au final, la CLS permet de déterminer la distribution spatiale des concentrations des composants purs d'un mélange. De plus, à l'inverse de l'ACP où les concentrations relatives estimées sont associées aux CP qui ne peuvent correspondre à des composants moléculaires de référence, la solution proposée par la CLS sur des données spectrales est compatible avec la physique et la chimie du mélange [55]. Par ailleurs, la CLS est considérée comme la méthode la plus fiable dans le cas où les spectres de composants purs du mélange sont connus [60]. Elle est donc régulièrement utilisée comme modèle de référence pour la détermination de la distribution spatiale des concentrations des composants purs lorsque des références spectrales réelles sont utilisées [62].

3.2.1. Principe de la CLS

En spectroscopie, un spectre est considéré comme un mélange de nombreux autres spectres. Il peut donc s'exprimer sous la forme d'une somme pondérée de spectres de composants purs où les coefficients de pondération correspondent aux concentrations relatives de chacun d'entre eux [54], [63]. Comme cela a été mentionné dans le paragraphe précédent, le principe de la CLS consiste à estimer la concentration relative de chaque spectre de composants purs de référence par approximation des moindres carrés pour un spectre donné [55], [63]. Les spectres de référence utilisés dans cette reconstitution spectrale sont généralement obtenus en réalisant des acquisitions spectrales des composants purs en question de manière indépendante [42]. De plus, cette reconstitution spectrale peut être réalisée en simultanée sur de nombreux spectres formant par exemple une image hyperspectrale. Dans ce cas, le problème de démélange résolu par la CLS peut se formuler sous forme matricielle où la matrice de données se décompose comme le produit d'une matrice de spectres de références et d'une matrice contenant les concentrations relatives de chacun d'entre eux et dont l'expression est la suivante :

$$D = \sum_{i=1}^{N} C_i S_i^T + E_i = CS^T + E \qquad \text{(II-2)}$$

où D est la matrice des données de taille (N, N_λ) avec N le nombre de spectres et N_λ le nombre d'ondes, S la matrice contenant les spectres de références de taille (N_λ, K) avec $K \in \mathbb{N}^+$ le nombre de spectres de références, C la matrice contenant les concentrations relatives des spectres de références de taille (N, K) et E la matrice conentant les résidus d'ajustement de taille (N, N_λ).
La CLS obtient les concentrations relatives de chaque composant de référence par le simple calcul matriciel suivant :

$$C = DS(S^T S)^{-1} \quad \text{(II-3)}$$

Par ailleurs, pour que les concentrations estimées aient un sens, il est courant d'ajouter une contrainte de non-négativité dans le modèle bilinéaire, on parle alors de Non-negative Constrained Least Square (NCLS). Ainsi, si un coefficient de contribution est estimé négatif, une estimation proposant une erreur d'ajustement plus grande mais vérifiant les contraintes de non-négativité est préférée.

3.2.2. Exemples d'application de la CLS

En spectroscopie Raman, la méthode CLS a été utilisée dans le domaine pharmaceutique pour caractériser la composition et la structure des produits pharmaceutiques. Dans la lutte contre les produits illégaux et la contrefaçon, la connaissance de la composition chimique et la distribution des composants au sein des produits est un élément essentiel pour juger de la qualité des produits. En effet, l'identification de la présence d'un composé qui n'a pas lieu d'être dans un produit ou encore une modification des quantités des composants au sein des produits peuvent être considérés comme une fraude dans une certaine mesure, voire même comme un danger. Dans l'étude menée par B. Vajna *et al.* [62], la CLS est utilisée comme méthode de référence pour déterminer la distribution spatiale de l'ingrédient pharmaceutique actif et des excipients de comprimés contraceptifs. Les cartes de distributions contiennent les concentrations respectives de ces composants et servent de référence pour comparer les résultats obtenus par d'autres méthodes chimiométriques. De même, P. A. Bonnet *et al.* [55] ont utilisé la CLS sur des images Raman de comprimés de Viagra® et de Plavix® authentiques et falsifiés. Elle a permis la quantification directe des substances actives et des excipients de ces produits et d'analyser leur répartition spatiale. Les informations quantitatives récoltées par CLS ont permis de réaliser une vérification de la composition des produits par rapport à la norme en vigueur pour l'autorisation de mise sur le marché et de distinguer les produits contrefaits de ceux authentiques.

La CLS a également été très utilisée dans le domaine de la recherche sur la peau. En effet, son application rend possible l'étude de la composition moléculaire et l'analyse de la morphologie de chacun de ces composants. Son utilisation a été initiée par P. J. Caspers *et al.* [39] comme alternative aux calculs de rapport de bandes de spectres Raman pour déterminer les profils de concentration semi-quantitatifs des composants majeurs de la peau humaine *ex vivo*. De leur côté, S. Guo *et al.* [60] ont appliqué la CLS sur des images Raman enregistrées à partir d'échantillons de peau porcine excisée dans le but de définir la distribution spatiale des composants de la peau et des follicules pileux. Ils ont pu ainsi localiser clairement la couche cornée, l'épiderme viable, le derme et le poil. Cette visualisation fournie par la CLS est intéressante d'un point de vue pharmaceutique et cosmétique, puisqu'elle montre son potentiel pour analyser ultérieurement les possibles voies de pénétration de substances appliquées localement autour et sur un follicule pileux considéré comme un réservoir pour l'absorption de substances. En effet, la CLS peut être aussi appliquer dans le but de suivre la progression de molécules exogènes à travers la peau. Par exemple, F. D. Fleischli *et al.* [63] ont utilisé la CLS pour étudier la répartition au niveau cellulaire des constituants d'un échantillon de peau reconstruite et de deux activateurs de pénétration appliqués localement. Pour se faire, des spectres de kératine, de lipides, d'eau, d'éthanol et des deux produits administrés ont servi de base spectrale à l'estimation faite par la CLS. Au final, les cartes de distribution des deux actifs résultantes montrent une corrélation entre la distribution locale des constituants de la peau et le caractère hydrophile ou lipophile des produits appliqués. De même, M. Essendoubi *et al.* [61] ont employé la CLS pour surveiller la pénétration cutanée des acides hyaluronique de différents poids moléculaires . Ces derniers jouent un rôle important dans l'hydrodynamisme des tissus et leurs cartes de concentrations locales, estimées en utilisant comme références spectrales des spectres d'une peau non traitée dite de contrôle et les spectres des molécules d'acides hyaluroniques, ont finalement montré leur perméabilité dans la

couche cornée lorsque leur poids moléculaire est faible et leur imperméabilité lorsque leur poids moléculaire est élevé. Les premiers assureront donc l'hydratation de la peau en profondeur alors que les seconds la maintiendront et protègeront la peau. Puis, V. K. Tippavajhala *et al.* [54] se sont intéressés à la pénétration de produits solaires à travers la peau humaine et ont appliqué la CLS pour déterminer la quantité de produits ayant pénétrée la peau. L'étude a confirmé que la pénétration des deux produits commerciaux testés se limitait à la couche cornée. Et plus encore, M. E. Darvin *et al.* [64] ont utilisé la CLS dans le but de calculer les contributions semi-quantitatives de différentes encres de tatouage en fonction de la profondeur dans le derme. Les informations contenues dans les profils de pénétration résultants peuvent ainsi fournir des explications sur les effets indésirables (dermatoses, réactions allergiques, inflammations, infections) que peuvent provoquer les pigments présents dans les encres et sur leur phototoxicité. Quant à eux, L. Miloudi *et al.* [44] ont visualisé la distribution des composants biochimiques d'un épiderme humain reconstruit et ont suivi la progression d'une molécule cosmétique, le Delipidol®, à travers les tissus après plusieurs temps d'exposition à l'aide de la CLS. Dans un premier temps, la CLS a été appliquée en utilisant des spectres de références représentant des protéines, des lipides, de l'ADN présents dans la peau et le spectre de l'AC. Cette approche n'a pas fourni de résultats concluants car les spectres de références n'étaient pas assez représentatifs de la composition biochimique des différentes couches épidermiques. Ces références spectrales ont donc ensuite été remplacées par des spectres de peau non traitée indépendants et la signature spectrale de référence de la molécule a été conservée. Avec cette nouvelle base spectrale, la CLS a permis de collecter des informations sur la distribution de l'AC dans les deux directions spatiales et en fonction du temps et de mieux apprécier l'homogénéité de sa diffusion (Figure II-7). Enfin, une dernière étude réalisée par L. dos Santos *et al.* [65] utilise la CLS dans le but de mieux comprendre les altérations intrinsèques à la couche cornée liées à l'âge qui affectent la pénétration de deux vitamines. Les références spectrales utilisées dans l'estimation réalisée par la CLS correspondent aux spectres de la kératine, du facteur d'hydratation naturel, des céramides, du cholestérol et des deux vitamines étudiées. Les cartes de distribution locales résultantes ont montré que les deux vitamines pénètrent significativement dans la couche cornée mais que le manque de lipides dans les peaux plus âgées affecte négativement la pénétration de la vitamine la moins hydrophile des deux.

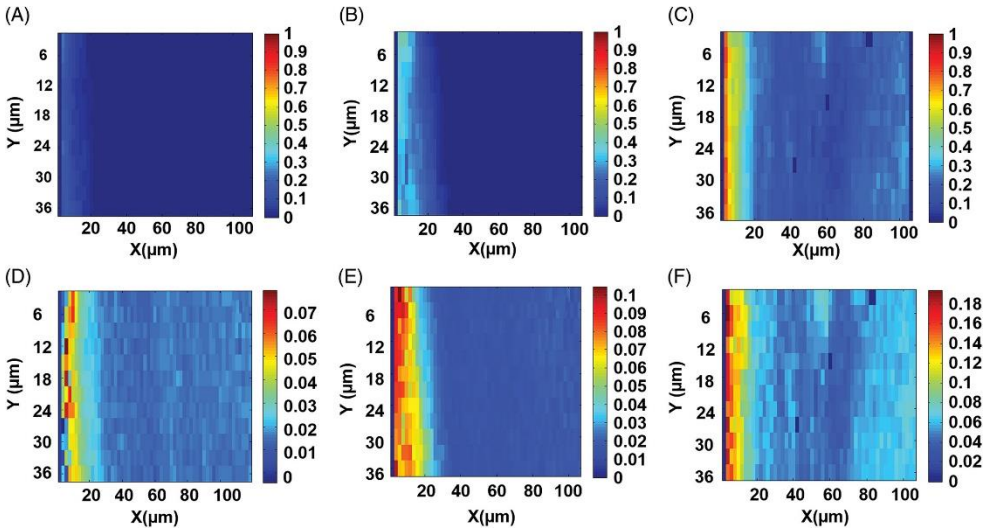


Figure II-7 – Cartes de distribution reconstruites d'échantillons de peau exposés au Delipidol® après (**A**) *1 heure,* (**B**) *2 heures et* (**C**) *3 heures obtenues par analyse CLS avec une contrainte de non-négativité sur les contributions estimées ; Les barres d'échelles représentent la contribution relative du spectre Delipidol® dans les spectres Raman à chaque pixel. Comparaison avec les cartes reconstruites obtenues par analyse univariée de la forte bande caractéristique du Delipidol située à 1625 cm^{-1} après* (**D**) *1 heure,* (**E**) *2 heures et* (**F**) *3 heures ; Les barres d'échelles représentent son intensité relative dans les spectres prétraités.*

3.2.3. Limites de la CLS

La CLS a le pouvoir d'estimer semi-quantitativement les concentrations des différents composants d'un mélange. Cependant, en l'absence des références de tous les composants, la CLS est difficilement applicable [62]. Leur disponibilité n'est pas toujours facile d'accès lorsqu'il s'agit de tissus biologiques ou de cellules et bien souvent seul un nombre limité parmi la grande quantité de composants est disponible [60]. Dans ces conditions, un ajustement d'un modèle par CLS serait imparfait et de nombreuses informations ne seraient pas modélisées dans les cartes de distributions mais probablement contenues dans les résidus du modèle. Cette limite a notamment été pointé du doigt par L. Miloudi *et al.* [44] lorsqu'un nombre restreint de composants de la peau a été utilisé dans la CLS. D'autre part, l'efficacité de la CLS est réduite en présence de décalages de position ou de changements de largeurs de bande qui peuvent se produire si les composants purs réagissent les uns avec les autres [60]. En effet, l'expression référentielle d'un composant d'un milieu à un autre peut être différente et n'est pas toujours accessible. Par conséquent, l'utilisation d'une référence spectrale qui ne correspond pas à l'expression du composant dans son nouvel environnement peut fausser le calcul de ses concentrations locales.

3.3. CONCLUSION

En spectroscopie Raman, les méthodes de démélange ou de modélisation spectrale actuellement utilisées ont pour objectif principal de simplifier l'information contenue dans les données spectrales. En raison de l'assimilation des données spectrales à une combinaison linéaire de spectres de composants purs et de leurs contributions relatives respectives, les méthodes de démélange sont plus souples que les méthodes de classification : la présence d'une variation spectrale dans un spectre qui est également exprimée dans d'autres spectres peut être nuancée par les concentrations estimées. Ainsi, dans le domaine de la recherche sur la peau, les méthodes d'analyse multivariée – telles que l'ACP et la CLS – ont permis notamment de comprendre la composition et la structure de la peau d'un point de vue autant qualitatif que semi-quantitatif. De plus, des molécules exogènes ont pu être localisées et relativement quantifiées grâce aux modifications spectrales induites par leur application sur la peau. Cependant, bien que ces méthodes de démélange soient plus flexibles quant à la description d'un spectre, celles les plus régulièrement utilisées – à savoir l'ACP et la CLS – nécessitent la connaissance a priori de l'expression Raman d'un ou plusieurs composants pour pouvoir les identifier ou les localiser. Elles ne permettent également pas de prendre en compte les éventuelles modifications spectrales telles que les décalages en nombre d'ondes ou les élargissements de bandes induites par les réactions moléculaires. Au final, ces méthodes manquent encore de liberté quant à l'identification des phénomènes moléculaires provoquées par l'interaction des molécules.

CONCLUSION

En résumé, il a été observé dans ce chapitre que le couplage de la microspectroscopie confocale Raman avec des méthodes d'analyse multivariée s'avère être un outil puissant pour comprendre les changements moléculaires induits par la pénétration de molécules exogènes dans la peau. Des méthodes de démélange spectral telles que l'ACP et la CLS ont aidé à identifier l'expression de ces variations moléculaires et à localiser dans certaines études les composants biochimiques qui en sont responsables. Cependant, ces méthodes présentent certaines limites. En effet, l'ACP fournit des résultats difficiles à interpréter et les composantes principales estimées n'ont pas de correspondance claire avec des références spectrales de composé chimique pur. De même, malgré une intention de faciliter l'interprétation des résultats issus d'une quantification des composants purs présents d'un mélange donné, la CLS requiert les références spectrales de ces composants purs, à savoir les composants de la peau et des molécules exogènes étudiées. Toutefois, ces références spectrales peuvent dans certains cas être indisponibles ou non conforme à leur expression dans le milieu cutané.

Ainsi, pour s'affranchir du besoin de connaissances *a priori* sur les spectres de composants purs et obtenir une décomposition spectrale pouvant être comparée à des références spectrales Raman, deux autres méthodes de démélange spectral non-supervisées sont présentées dans le Chapitre IV. Cependant avant même d'appliquer n'importe quelle méthode d'analyse multivariée, un prétraitement des données Raman est nécessaire afin de les rendre comparables et analysables. Un tel protocole de prétraitement est proposé dans le Chapitre III.

Chapitre III.
ANALYSES COMPARATIVES DES MÉTHODES DE PRÉTRAITEMENTS DES DONNÉES RAMAN

Résumé
Ce chapitre est consacré à une étude comparative de multiples méthodes de prétraitements de données régulièrement utilisées en spectroscopie Raman. Dans ce travail, nous avons considéré des méthodes de débruitage spectral, de réduction de dimensions, de correction de la résolution spectrale et de normalisation. Cette étude comparative a permis d'identifier quelles étaient les méthodes les plus adaptées aux jeux de données utilisés, ainsi que d'établir une routine ordonnancée de prétraitement de données spectrales pour l'analyse et le suivi d'un principe actif à travers l'épiderme en imagerie Raman.

Introduction 72
1. Réductions des dimensions : définition d'un sous-ensemble 72
2. Débruitage 74
2.1. Bruit gaussien 74
2.1.1. Filtre des moyennes glissantes 74
2.1.2. Algorithme Savitzky-Golay 75
2.1.3. Projection par ACP 77
2.1.4. Discussion 78
2.2. Bruit de fond : correction de la ligne de base 79
2.2.1. Méthode Lieber 81
2.2.2. Méthode Rubberband 82
2.2.3. Méthode par minimisation semi-quadratique 83
2.2.4. Discussion 86
2.3. Bruit des raies cosmiques 86
Étape 1 : Identification des spectres contaminés et localisation des raies cosmiques 87
Étape 2 : Correction des raies cosmiques 91
3. Calibration de la résolution spectrale et alignement spectral 92
4. Normalisation 94
4.1. Normalisation centrée-réduite 94
4.2. Normalisation Min-Max 94
4.3. Normalisation ℓp, $p = \{1, 2, \infty\}$ 95
5. Proposition d'une routine de prétraitement des données Raman 96
Conclusion 98

INTRODUCTION

Comme présenté dans le Chapitre I, la peau est un tissu complexe. Son hétérogénéité physique et chimique influencent le signal Raman et donnent notamment lieu à un certain nombre de perturbations et d'artefacts qui peuvent déformer et réduire considérablement la précision de la mesure spectrale [42]. De même que la composition des échantillons étudiés, les protocoles de mesures influent sur les signaux Raman enregistrés [30]. Bien que l'optimisation des substrats, des paramètres de l'appareil ou encore du temps d'acquisition ont permis d'adapter les protocoles de mesure Raman pour réduire leur influence sur les signaux et améliorer la reproductibilité des spectres enregistrés, du bruit et des artefacts inhérents au microspectromètre Raman restent présents dans les spectres. En effet en plus d'une empreinte moléculaire, les spectres présentent du bruit et une ligne de base qui se matérialise comme un décalage fonctionnel continu du spectre. L'origine de cette ligne de base est d'ailleurs la plupart du temps associée dans la littérature à une fluorescence intrinsèque de certaines molécules. Puis, les intensités Raman des spectres ne sont pas normalisées. Cependant, comme mentionné dans le Chapitre II, les données doivent être comparables pour pouvoir les analyser. Par conséquent, les influences de l'environnement tissulaire et du microspectromètre Raman sur les mesures spectrales doivent être corrigées pour pouvoir appliquer des méthodes d'analyse multivariée. Une étape de prétraitement spectral doit donc être réalisée sur l'ensemble des données afin de les rendre lisibles, comparables et analysables.

De nombreuses méthodes de prétraitement de données spectrales ont jusqu'à présent été développées. Chacune d'elles améliore d'une façon ou d'une autre l'interprétation des informations spectrales et elles sont régulièrement utilisées en spectroscopie Raman. Cependant, il n'existe à ce jour aucun consensus formulé par les instances internationales telles que la Conférence Internationale d'Harmonisation (CIH, *International Conference for Harmonisation ICH*, en anglais), l'Agence fédérale américaine des produits alimentaires et médicamenteux (*Food and Drug Administration FDA*, en anglais) et l'Agence Européenne du Médicament (EMA, *European Medicines Agency EMA*, en anglais) ni même d'étude formelle établissant quelles méthodes ou quelle combinaison de méthodes il est préférable d'utiliser [42]. Ces grandes instances internationales recommandent seulement d'adapter le protocole de prétraitement en fonction des données à analyser et d'en préciser les méthodes utilisées et l'ordre dans lequel celles-ci sont appliquées [66], [67].

Face à ce constat, ce Chapitre III présente une analyse comparative des méthodes de prétraitement de données Raman régulièrement utilisées dans la littérature. Les avantages et inconvénients de différents algorithmes de correction de bruit, de correction de la ligne de base et de normalisation spectrale sont présentés et comparés. Nous présentons aussi des techniques de réduction des dimensions, d'ajustement de la résolution spectrale et de réalignement spectral puis proposons une méthode de correction des raies cosmiques hybride pour rendre le prétraitement des données Raman plus complet. En d'autre mots, le but de ce chapitre est de définir les méthodes de prétraitement les plus adaptées à la correction des jeux de données utilisés dans les études de pénétration cutanée présentées dans les prochains chapitres. Pour finir, un protocole de prétraitement global des données Raman est proposé.

1. RÉDUCTIONS DES DIMENSIONS : DÉFINITION D'UN SOUS-ENSEMBLE

La spectroscopie Raman est souvent utilisée pour collecter des séries de spectres issus d'un seul et même échantillon, représentées sous la forme d'une liste ou d'une image Raman. Ces spectres sont acquis sur une large plage de nombre d'ondes pouvant aller de 0 à 3500cm^{-1}. L'acquisition de nombreux spectres est souvent utile pour des études statistiques, mais nombre d'entre eux

peuvent s'avérer être dépourvus d'informations pertinentes sur certaines plages de nombre d'ondes. Ces informations déficientes peuvent être concentrées dans un ou plusieurs sous-ensembles de spectres ou de nombre d'ondes, et sont non nécessaires à l'analyse et l'interprétation ultérieure des données.

Dans le but spécifique de se concentrer uniquement sur des données spectrales riches en informations sur les propriétés et la composition moléculaire des échantillons de peau et des principes actifs étudiés, il convient de réduire la plage de nombre d'ondes des spectres à un sous-ensemble de caractéristiques spectrales d'intérêts [68]. Cette réduction de la dimension spectrale des données apporte quelques avantages supplémentaires : les coûts de calcul des méthodes d'analyse appliquées ultérieurement sont diminués et la probabilité de trouver le meilleur modèle représentant les données est augmentée [69]. En pratique, la sélection des caractéristiques spectrales d'intérêts se réalise habituellement par troncature du nombre d'ondes à la région spectrale essentielle [69]. Pour la spectroscopie Raman de la peau, la région d'empreinte digitale la plus pertinente à analyser se situe sur l' intervalle [400-1800] cm^{-1} (Figure III-1).

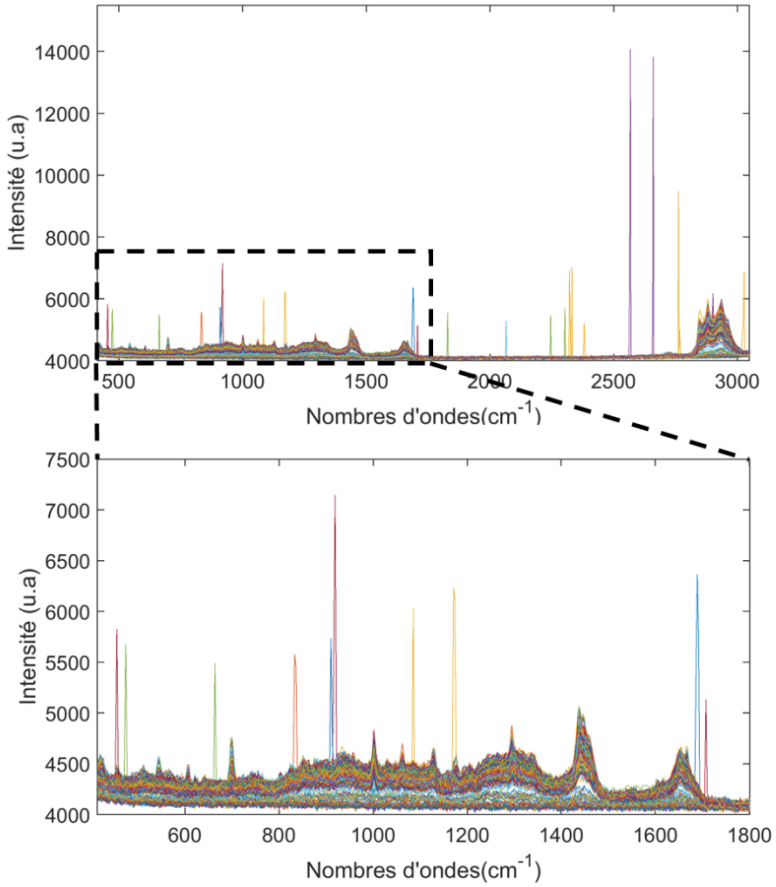


Figure III-1 – Spectres Raman bruts enregistrés à partir d'un modèle Episkin® sur les intervalles [400-3500]cm^{-1} et [400-1800]cm^{-1}.

Par ailleurs, lorsque des groupes de spectres manquent d'informations pertinentes sur l'intégralité de la plage de nombre d'ondes, leur prise en compte dans les méthodes d'analyse ultérieure est sans intérêt. L'ACP, qui est la méthode de réduction de dimension d'un ensemble de données la plus connue tout domaine d'application confondu, est alors fréquemment utilisée en spectroscopie Raman [69] pour exclure ces spectres non informatifs de l'ensemble des données Raman. En pratique, une fois l'ACP appliquée, l'analyse des données se fait :

- soit sur les composantes principales obtenues elles-mêmes. Dans le cas où les données sont issues d'une image Raman, comme dans nos études de pénétration cutanée d'AC, la prise en compte uniquement des premières composantes principales dans l'analyse ne permet plus d'accéder aux éléments d'information spatiale sur les caractéristiques moléculaires de l'échantillon. Par conséquent, nous avons choisi de ne pas retenir cette façon de procédé pour écarter de l'analyse les spectres dépourvus d'informations pertinentes.

- soit sur les données reconstruites à l'aide des premières composantes. Dans ce cas, l'analyse se fait sur des données spectrales corrigées du bruit. Cette correction du bruit est expliquée plus en détails dans la partie suivante.

2. DÉBRUITAGE

En spectroscopie Raman, différents types de bruit impactent la précision spectrale des données : le bruit électronique, le bruit de fond, et les raies cosmiques.

- Le bruit électronique est un mélange de plusieurs types de bruit provenant des instruments: le bruit de scintillement généré en présence d'électrons parasites dans le matériau irradié, le bruit grenaille causé par la discontinuité du flux lumineux et le bruit thermique produit par agitation thermique des électrons en équilibre thermique présents dans le matériau. Ce bruit électronique est imprévisible et constant en spectroscopie Raman [42]. Il est souvent supposé gaussien puisqu'il provient de processus non corrélés [69].

- Le bruit de fond, quant à lui, est intense et souvent causé par la fluorescence de certaines molécules ou la diffusion de base du capteur CCD [69]. Il correspond en réalité à la ligne de base visible sur les spectres. Ce signal de fond peut donc être traité indépendamment du bruit électronique.

- Les raies cosmiques proviennent de particules à haute énergie frappant le capteur CCD [68], [69] et peuvent aussi être traitées de façon indépendante.

De nombreux algorithmes atténuant séparément chacun de ces types de bruit existent dans la littérature. Les sous-sections suivantes présentent une étude comparative de ceux les plus communément utilisés en spectroscopie Raman. À noter qu'il n'existe actuellement aucun moyen d'obtenir une vérité de terrain concernant le bruit à corriger. La comparaison des approches de correction de bruit ne peut donc être que qualitative et repose seulement sur notre expertise.

2.1. BRUIT GAUSSIEN

Pour appréhender le bruit gaussien, deux types d'approches sont typiquement utilisées dans la littérature : d'une part, des méthodes de filtrage comme la méthode des moyennes glissantes et le filtre de Savitzky-Golay [68], [69] et d'autre part, la projection des données par ACP comme alternative à ces deux dernières.

2.1.1. Filtre des moyennes glissantes

Le filtre des moyennes glissantes, aussi appelé filtre des moyennes mobiles, est une approche de convolution. Il consiste à attribuer à chaque point du spectre la valeur moyenne de celles de ses voisins avec la sienne. Autrement dit, pour une demi-largeur de fenêtre l définie, la valeur associée à une variable spectrale est la moyenne des valeurs des variables spectrales sur l'intervalle $[-l: l]$. Ce type de filtrage réduit bien le bruit mais il est en même temps assez agressif. Selon la taille de la fenêtre choisie par l'utilisateur, les fortes variations spectrales peuvent être plus ou moins atténuées et certains pics Raman peuvent être écrêtés. Les résultats d'un filtrage

par moyennes glissantes utilisant différentes tailles de fenêtre glissante $l = \{1, 2, 5, 7\}$ sur un spectre brut issu d'un échantillon d'épiderme humain reconstruit (Episkin®) non traité sont présentés dans la Figure III-2. Plus la demi-largeur de fenêtre augmente et plus l'information biochimique exprimée par les petites variations spectrales s'estompe dans le spectre filtré. Notamment lorsque $l = \{5, 7\}$, les variations spectrales présentes initialement dans l'intervalle [800-1200] cm^{-1} sont totalement effacées dans les spectres filtrés. Plus encore, la bande caractéristique des systèmes biologiques située à 1004 cm^{-1} a même totalement disparu. Au final, le filtrage par moyennes mobiles pour réduire le bruit gaussien présent dans les données Raman doit être utilisé avec prudence et l'utilisateur doit être conscient du risque de perte d'information encouru lorsque la taille de la fenêtre glissante augmente.

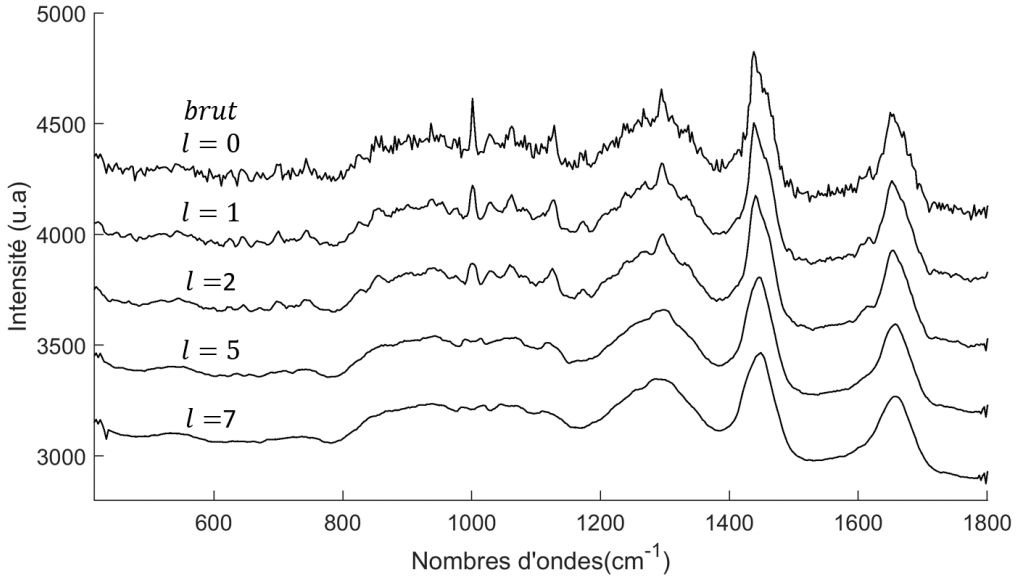


Figure III-2 – Correction d'un spectre brut, enregistré à partir d'un modèle Episkin®, par la méthode des moyennes glissantes avec des demi-largeurs de fenêtre glissante différentes : $l = \{1, 2, 5, 7\}$. Les spectres sont décalés pour une meilleure visualisation.

2.1.2. Algorithme Savitzky-Golay

L'algorithme de Savitzky-Golay (SG) est très populaire et est souvent incorporé dans les logiciels d'instruments commerciaux pour supprimer le bruit dit électronique ou gaussien [42]. Le principe de ce filtre de lissage est le même que celui de la méthode des moyennes mobiles. À chaque valeur spectrale est attribuée une nouvelle valeur définie en fonction de son voisinage. Alors que la méthode des moyennes glissantes remplace chaque valeur d'un spectre par la valeur moyenne du voisinage, la méthode SG estime pour chaque variable du spectre un polynôme de degré d au sens des moindres carrés passant par les variables spectrales voisines et remplace sa valeur par celle du polynôme estimé [68]–[70]. La méthode SG nécessite donc la définition d'un paramètre supplémentaire à celui de la demi-largeur l de la fenêtre glissante déjà utilisée dans le filtrage par moyennes glissantes : l'ordre polynomial d. Lorsque $d = 0$, la méthode SG n'est autre que la méthode des moyennes glissantes. En revanche lorsque $d > 0$, la taille de la fenêtre glissante doit être suffisamment grande devant l'ordre du polynôme pour que la réduction de bruit soit effective : si la taille de la fenêtre est égale au degré polynomial, le polynôme estimé n'est autre que le spectre lui-même sur cet intervalle et aucun lissage n'est effectué. Plus précisément, le degré polynomial doit vérifier l'équation $d < 2l$. À noter que si un ordre polynomial nul est associé à un nombre élevé de points de lissage (l grand), alors la méthode SG permet de corriger le

décalage du spectre par rapport à l'axe des abscisses et devient une méthode de correction de ligne de base.

Le choix de l'ordre du polynôme à estimer et de la taille de la fenêtre mobile utilisée dans l'estimation polynomiale impacte considérablement la qualité du spectre débruité [42]. L'influence de ces deux paramètres dans le débruitage d'un spectre Raman est illustrée la Figure III-3. Lorsque le degré polynomial est fixe ($d = 1$), l'augmentation de la demi-largeur l de la fenêtre glissante engendre de la même manière qu'avec la méthode des moyennes glissantes une forte atténuation des faibles variations spectrales. Cette perte d'informations est particulièrement visible sur l'intervalle [800-1200] cm^{-1} où la bande caractéristique de la peau située à 1004 cm^{-1} a à nouveau disparu. En revanche lorsque le degré polynomial augmente pendant que la demi-largeur reste fixe ($l = 5$), la perte d'informations spectrales est réduite. Les faibles variations spectrales sont préservées en raison de la prise en considération des différentes courbures que peut prendre le spectre. Généralement, un ordre fixé au minimum à 3 est requis pour que les points d'inflexion présents dans le spectre soient pris en compte.

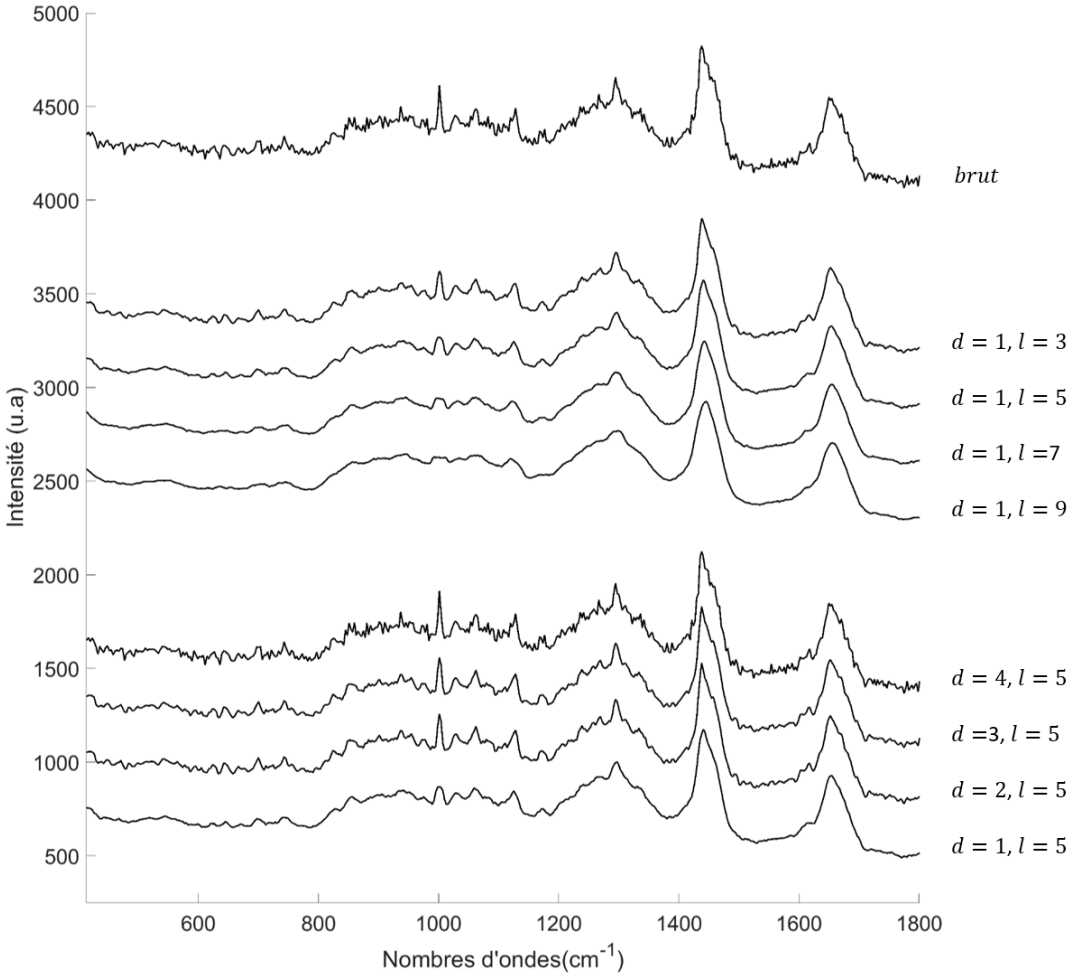


Figure III-3 – Correction d'un spectre brut, enregistré à partir d'un modèle Episkin®, par l'algorithme de Savitzky-Golay avec des demi-largeurs de fenêtre glissante et des degrés polynomiaux différents. Les spectres sont décalés pour une meilleure visualisation.

Au final l'optimisation des deux paramètres, la demi-largeur de la fenêtre et le degré polynomial, est nécessaire pour ne pas dénaturer et déformer les caractéristiques spectrales et relève simplement de l'appréciation qualitative de l'utilisateur. Bien que l'optimisation de ces deux paramètres soit plus complexe que celle de l'unique paramètre utilisé dans la méthode des

moyennes glissantes, elle permet d'assurer une perte d'informations moins importante. Par conséquent, l'application de la méthode SG pour réduire le bruit gaussien présent dans les données est préférable à celle du filtrage par moyennes glissantes.

2.1.3. Projection par ACP

La correction du bruit gaussien est souvent réalisée spectre par spectre, comme c'est le cas pour les filtres SG et des moyennes glissantes. Cependant, lorsqu'un ensemble de spectres est à disposition, une autre possibilité de prétraitement collectif est envisageable. Il s'agit d'utiliser une projection des données par ACP. Comme mentionné dans le Chapitre II, l'ACP est une technique de transformation linéaire orthogonale visant à décorréler les spectres les uns des autres [42]. Pour rappel, les composantes principales obtenues par ACP sont rangées par ordre décroissant en fonction de la variance qu'elles expliquent. Ainsi, le bruit gaussien provenant d'un processus décorrélé est alors supposé être exprimé dans les dernières composantes principales contenant la variance non expliquée. Le principe de la réduction de bruit par ACP repose donc sur i/ une projection des données sur les composantes principales, ii/ la modification des dernières composantes principales contenant le bruit et iii/ une rétroprojection dans l'espace d'origine. Généralement, les dernières composantes principales contenant le bruit sont soit négligées soit lissées lors de la reconstruction des données [68]. Pour mener à bien la réduction de bruit via la projection des données par ACP, il convient de définir le nombre de composantes principales à considérer dans la reconstruction des données. Ce nombre est souvent exprimé en terme de pourcentage par rapport au nombre de spectres dans l'ensemble de données.

Théoriquement, peu de composantes principales suffisent pour décrire la quasi-totalité des variations spectrales. La Figure III-4 illustre la reconstruction des données spectrales d'une image Raman obtenue à partir d'un échantillon d'épiderme humain reconstruit (Episkin®) non traité en utilisant différents nombres de composantes principales : 10%, 5%, 1% et 0.5% du nombre total de spectres contenus dans l'image, soit 38, 19, 4 et 2 composantes principales prises en compte. Pour cette image Raman, les trois premières composantes principales expliquent respectivement 99.9887 %, 0.0075 % et $6.0356 \cdot 10^{-4}$ % de la variance des données. Celles restantes expriment des variations spectrales non informatives et plus particulièrement du bruit. Par conséquent, la reconstruction des données à l'aide de 10% et 5% des composantes principales inclut trop d'informations non significatives telles que de très faibles variations spectrales assimilées au bruit ou encore des raies cosmiques. Elle ne permet donc pas de réduire considérablement le bruit présent dans les données. Quant à elle, la reconstruction des données spectrales à l'aide de seulement 0.5% des composantes principales ne laisse place à aucune information spectrale de très faibles intensités Raman. Dans nos expérimentations, la projection des données spectrales sur seulement 1% du nombre total de composantes principales a constitué le meilleur compromis pour réduire le bruit présent.

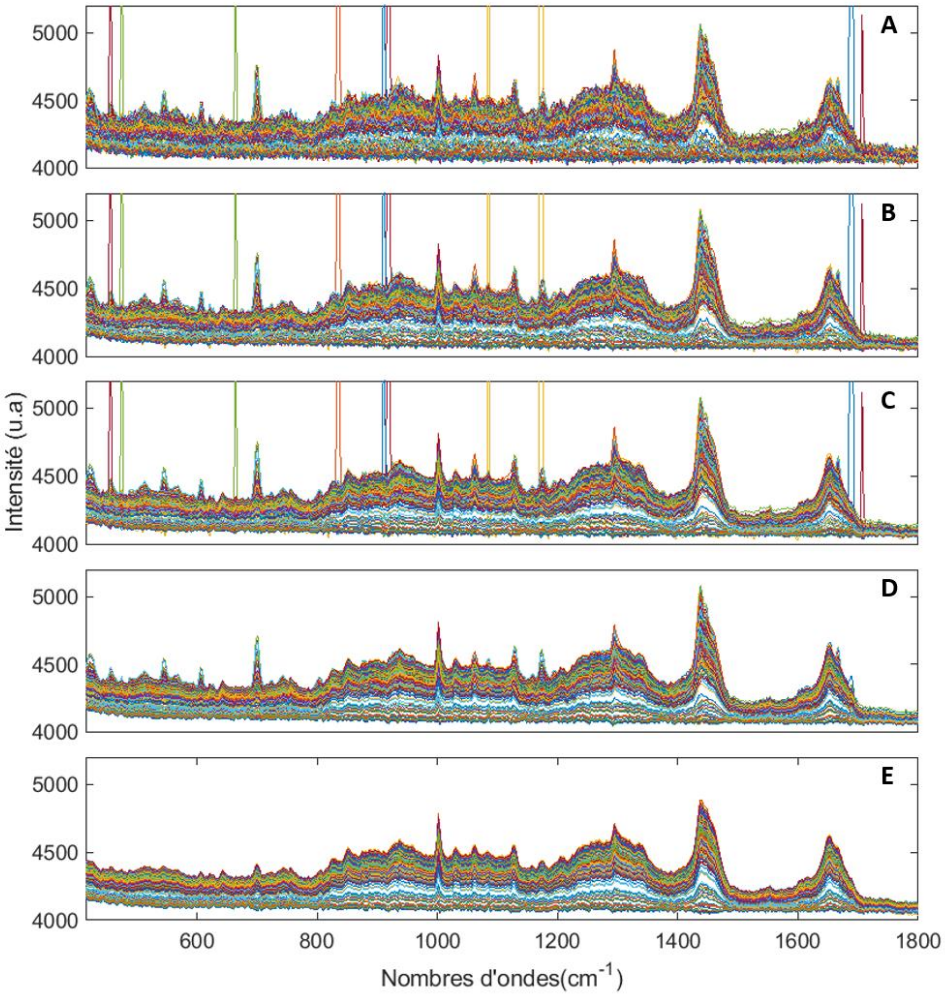


Figure III-4 – Correction de spectres bruts (**A**)*, enregistrés à partir d'un modèle Episkin®, à l'aide d'une projection par ACP en utilisant* (**B**) *10%,* (**C**) *5%,* (**D**) *1% et* (**E**) *0.5% des premières composantes principales de l'ensemble de données (soit respectivement 38, 19, 4 et 2 composante sur 375 dans ce modèle)*

2.1.4. Discussion

La réduction de bruit par projection par ACP présente de nombreux avantages. À l'inverse de la méthode SG qui doit être utilisée spectre par spectre, une seule application suffit pour corriger l'ensemble des spectres. En effet, les paramètres utilisés dans la méthode SG peuvent être ajustés autant de fois qu'il y a de spectres dans l'image tandis que celui de la méthode de projection par ACP peut l'être en une seule fois. De plus, alors que la méthode SG nécessite d'optimiser deux paramètres pour pouvoir être appliquée sans craindre une perte d'informations considérable, la méthode de projection par ACP ne nécessite l'optimisation que d'un seul paramètre : le nombre de composantes principales nécessaires à la reconstruction des données sans bruit. Il est cependant important de garder à l'esprit que la sélection du nombre optimal de composantes principales pour reconstruire les données sans bruit reste un défi et est soumis à la seule appréciation de l'utilisateur. Par ailleurs, la méthode de projection par ACP évite l'apparition de phénomènes d'atténuations des intensités spectrales comme cela peut être observé sur les résultats de la méthode SG lorsque le paramètre de demi-largeur de la fenêtre glissante est trop élevé. Enfin, bien que certaines peuvent rester présentes, les raies cosmiques dont les variations spectrales sont souvent contenues dans les composantes principales considérées comme insignifiantes sont effacées des données. À l'inverse, la suppression des raies cosmiques par la

méthode SG n'est possible que si la demi-largeur de la raie cosmique est inférieure ou égale à la demi-largeur de la fenêtre glissante utilisée.

En définitive, bien que les résultats de la correction du bruit obtenus par les trois approches étudiées soient similaires pour des paramètres bien choisis (Figure III-5), l'application de la méthode de réduction de bruit gaussien à l'aide d'une projection par ACP est préférable à celle de la méthode SG et du filtrage par moyennes glissantes en raison de sa plus simple optimisation qualitative et de sa capacité à corriger l'ensemble des spectres en une seule application. C'est celle que nous préconisons et que nous avons retenue pour toutes nos expérimentations.

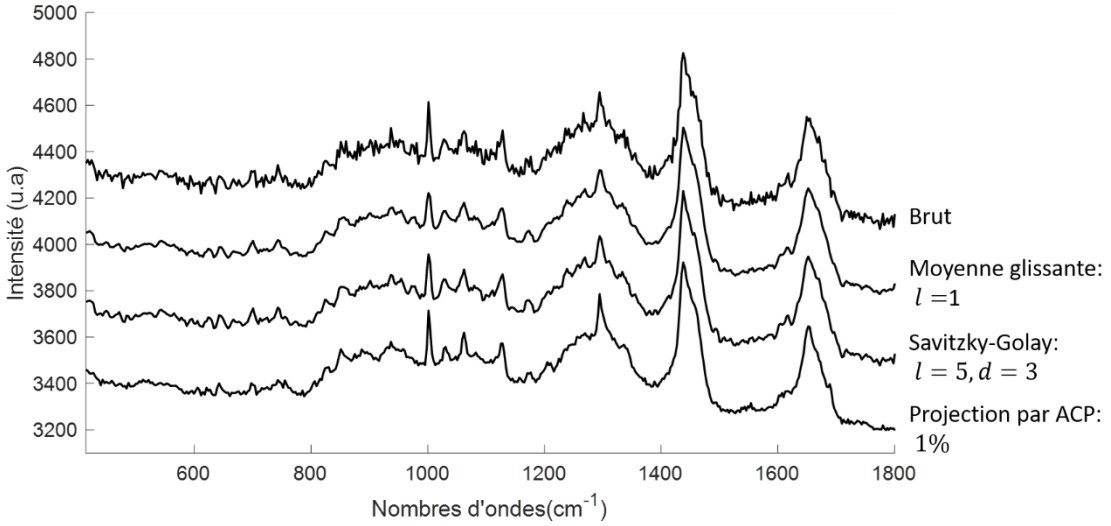


Figure III-5 – Comparaison d'un spectre brut, enregistré à partir d'un modèle Episkin®, corrigé par la méthode des moyennes glissantes, la méthode de Savitzky-Golay et à l'aide d'une projection par ACP.

2.2. BRUIT DE FOND : CORRECTION DE LA LIGNE DE BASE

Durant l'acquisition Raman, certains phénomènes instrumentaux ou une fluorescence intrinsèque de certaines molécules introduisent un signal de fond dans les spectres sous l'excitation du laser [71], [72]. Ce bruit de fond se caractérise par l'émission d'une large bande se produisant dans le même intervalle de longueur d'onde que le signal Raman (courbe discontinue dans la Figure III-6) et peut être jusqu'à 10^6 à 10^8 plus intense que la diffusion Raman [68]. La superposition de ce signal de fond aux bandes Raman caractéristiques d'un échantillon peut cependant masquer des changements d'intérêt spectral subtils. Il est alors nécessaire de supprimer ces tendances apparaissant sous différents profils de courbe pour améliorer la lisibilité et l'évaluation des données. Selon [71], les bandes spectrales d'un signal Raman idéal doivent se situer sur une ligne de base droite. Par conséquent, le bruit de fond doit être corrigé avant toute analyse de données afin d'en faciliter leur observation, leur interprétation et leur analyse. En effet, la correction de ligne de base permet une détermination des nombres d'onde et une mesure de la surface et de l'amplitude des bandes Raman plus précises.

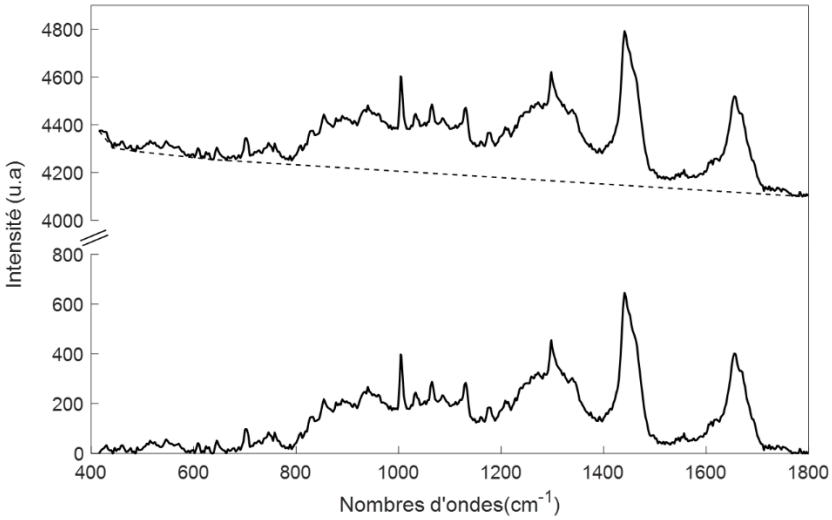


Figure III-6 – Spectre comportant du bruit de fond illustré par la courbe discontinue (en haut) et le résultat de sa correction par la méthode Rubberband (en bas).

Avant de recourir à des méthodes purement mathématiques pour supprimer ce bruit de fond, la présentation de l'échantillon et les protocoles de mesures peuvent être ajustés [30], [42]. En effet, le choix du substrat, de la longueur d'onde et de la puissance de la source laser, du temps d'acquisition et des conditions environnementales de l'échantillon peuvent influencer la qualité des spectres et leur optimisation peut particulièrement réduire la ligne de base des spectres [30]. Malgré une adaptation du protocole de collecte des spectres Raman, la qualité des spectres peut ne pas être optimale. Pour parvenir à une correction de la ligne de base satisfaisante, l'utilisation de méthodes mathématiques où aucun composant optique ou électronique supplémentaire ou autres modifications matérielles complexes ne sont requis est inévitable.

De nombreux algorithmes de correction de ligne de base existent dans la littérature. Un des plus utilisés est l'algorithme SG [68], [69]. Comme mentionné dans la section 2.1.2 de ce Chapitre III, cet algorithme capable de corriger le bruit gaussien présent dans les données est aussi capable de corriger la ligne de base des spectres. En effet, les courbes de la ligne de base peuvent être obtenues en utilisant un degré polynomial nul avec un nombre élevé de points de lissage [68]. De nombreux autres algorithmes reposent sur l'approximation polynomiale, une méthode simple mais jugée efficace. Elle peut être réalisée directement sur l'ensemble des valeurs d'intensité d'un spectre, comme pour la méthode Lieber [73], mais aussi sur un ensemble de points supports prédéfinis [68]. La détermination de ces points supports peut se révéler être une opération fastidieuse et chronophage lorsque le nombre de spectres à corriger est élevé. C'est pourquoi des méthodes incluant une sélection automatique de points supports telles que la méthode Rubberband [71], [72], [74] ou des méthodes sensibles à l'importance des bandes Raman telles que celle utilisant la minimisation semi-quadratique [75], [76] ont été développées. Les résultats des trois méthodes, Lieber, Rubberband et minimisation semi-quadratique sont comparées ci-après. À noter à nouveau qu'il n'existe actuellement aucun moyen d'obtenir une vérité de terrain concernant le bruit de fond introduit dans la mesure spectrale. La comparaison des approches de correction de ligne de base ne peut donc être que qualitative et repose seulement sur notre expertise.

2.2.1. Méthode Lieber

La méthode Lieber estime la ligne de base d'un spectre à l'aide d'approximations polynomiales par méthode des moindres carrés [73]. Le principe de cette méthode consiste à réaliser un lissage spectral de telle sorte que les bandes Raman soient automatiquement éliminées, ne laissant que la ligne de base à soustraire du spectre original (Figure III-7, courbe continue grise) [73]. Pour cela, une première approximation polynomiale d'ordre préalablement défini par l'utilisateur est réalisée sur le spectre Raman enregistré (Figure III-7, courbe discontinue noire). Les coefficients du polynôme ajusté au spectre sont estimés de sorte que l'erreur au sens des moindres carrés soit minimale. Ensuite, tous les points de données du spectre ayant une valeur d'intensité Raman supérieure à leur valeur respective dans le polynôme approximé sont automatiquement réaffectés à l'intensité du polynôme ajusté (Figure III-7, courbe continue rouge) [73]. Ces étapes d'approximation polynomiale et de réaffections des valeurs d'intensité sont réitérées un grand nombre de fois (les itérations n°2 et n°48 sont illustrées sur la Figure III-7) et le processus de filtrage prend fin lorsque plus aucune affectation n'est possible, autrement dit lorsque plus aucun changement entre deux approximations polynomiales n'est observé [73].

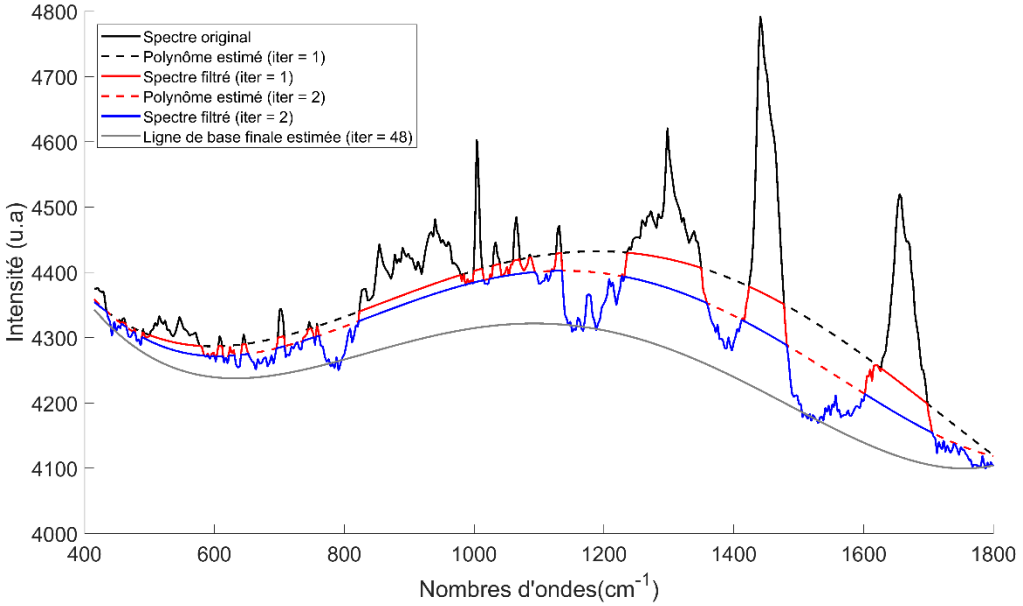


Figure III-7 – Les différentes étapes de l'estimation de la ligne de base d'un spectre débruité, enregistré à partir d'un modèle Episkin®, par la méthode Lieber. L'ordre des polynômes estimés au sens des moindres carrés est $d = 4$.

Bien qu'elle soit simple, cette technique d'ajustement polynomiale est sujette à la variabilité : le polynôme estimé est influencé par les bandes spectrales les plus intenses d'autant plus que la forme de la ligne de base dépend du degré polynomial choisi par l'utilisateur. Par exemple, une ligne de base estimée à partir d'une approximation polynomiale d'ordre élevé a tendance à s'ajuster plus facilement au bas du spectre ($d = 10$ sur la Figure III-8) tandis que celle estimée à partir d'une approximation polynomiale d'ordre plus faible tend à rester plus grossière ($d = 1$ sur la Figure III-8). Plus généralement, il est important de choisir un ordre polynomial bas, souvent entre 4 et 6, pour réduire la complexité du modèle ajusté et éviter le sur-ajustement (Figure III-8) [68]. Enfin, d'après les spectroscopistes, il est admis que le début et la fin d'un spectre doivent être proches de l'axe abscisse. Cependant, selon le degré polynomial choisi, un spectre corrigé de sa ligne de base par la méthode Lieber ne vérifie pas toujours ce consensus. La méthode Lieber est alors délicate à utiliser en raison de la forte dépendance de la forme de la ligne de base estimée au choix du degré polynomial d. Néanmoins de manière globale et contrairement à la méthode

SG utilisée pour corriger la ligne de base, la méthode Lieber a l'avantage de conserver la forme traditionnelle d'un spectre Raman pour faciliter son interprétation en terme de composition chimique et structurelle [68] et est tout de même capable d'extraire les caractéristiques spectrales Raman de faibles intensités à partir d'un spectre ayant un signal de fond de très élevé [73].

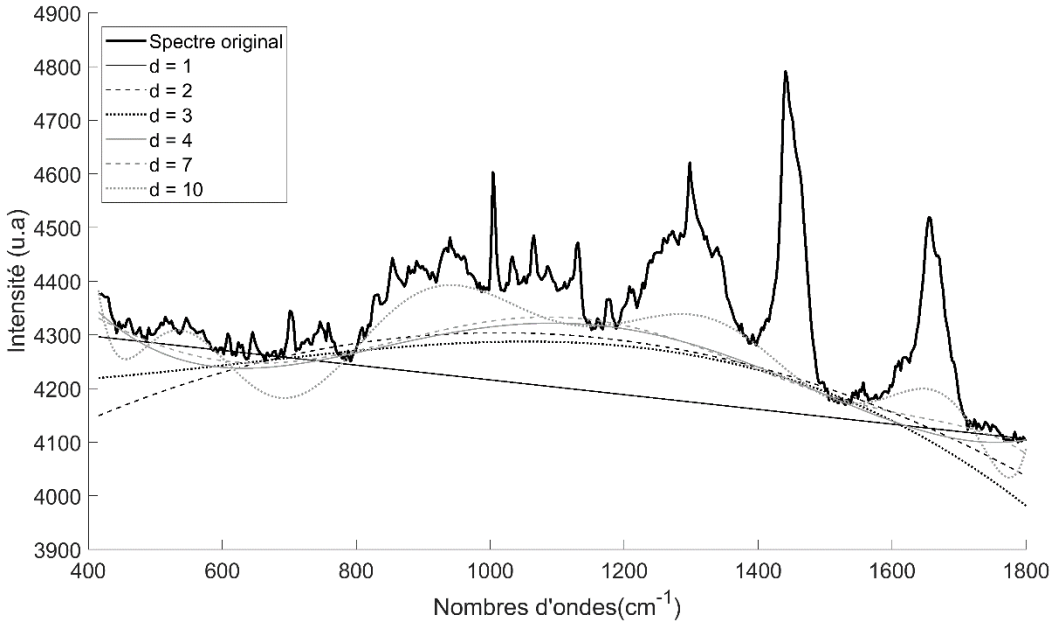


Figure III-8 – Estimation de la ligne de base d'un spectre débruité, enregistré à partir d'un modèle Episkin®, par la méthode Lieber avec différents degrés polynomiaux $d = \{1, 2, 3, 4, 7, 10\}$.

2.2.2. Méthode Rubberband

La méthode dite « Rubberband » est une méthode semi-automatique. Elle estime de façon autonome la ligne de base d'un spectre par morceaux. De cette façon, aucune restriction sur sa forme globale n'est imposée contrairement à la méthode Lieber. En revanche, le choix du degré des polynômes à ajuster au spectre par morceaux qui est préalablement défini par l'utilisateur [71] conditionne la forme de la ligne de base sur chaque morceaux.

Son principe consiste à déterminer une enveloppe convexe sous le spectre et la lui soustraire. Cette enveloppe convexe correspond à une région limitée par des points supports et l'axe des abscisses [71]. Ces points supports sont définis comme étant les minima locaux du spectre de façon à conserver l'hypothèse de convexité de l'enveloppe [74]. La recherche de ces points supports se fait en deux étapes. D'abord, le minimum global est choisi comme premier point support. Ensuite, une recherche des minima locaux de chaque côté du minimum global permet d'identifier les autres points supports (Figure III-9). Sur chaque nouvelle fenêtre spectrale, la soustraction d'une droite, de pente bien définie et toujours inférieure au spectre, permet de faire apparaitre un minimum différent du précédent point support. Une fois tous les points supports définis, la méthode Rubberband ajuste un polynôme entre chacun d'eux. L'ensemble de ces polynômes estimés constitue alors la ligne de base estimée. La plupart du temps, le degré polynomial est fixé à 1 [42], [69], [74]. De cette façon, une ligne de base brute est façonnée de manière totalement autonome et une sur-correction spectrale peut être évitée. Par exemple, du fait qu'aucun point support n'ait été identifié dans l'intervalle [630-1800] cm^{-1}, l'estimation de la ligne de base du spectre illustré dans la Figure III-9 est grossière dans cette zone spectrale mais n'est pas sur-ajustée. En définitive, à la différence de la méthode Lieber qui n'exclut pas la présence de valeurs négatives après correction, la méthode Rubberband est capable de supprimer les contributions de lignes bases aussi bien linéaires que concaves sans créer de valeurs négatives

et de conserver toutes les caractéristiques à leur emplacement spectral d'origine [74]. En effet, une telle conservation des caractéristiques spectrales après correction de la ligne de base a son importance puisque la diminution du bruit de fond sans altération spectrale permet une amélioration des capacités prédictives des analyses réalisées ultérieurement, comme cela est souligné dans [72]. Toutefois, le temps de calcul de la méthode Rubberband peut être lourd en présence de données volumineuses. En effet, en comparaison à la méthode Lieber qui nécessite une simple approximation polynomiale pour chaque spectre, la méthode Rubberband doit préalablement définir les points supports de chaque spectre avant d'ajuster un polynôme à chacun d'eux. Ces calculs supplémentaires la rendent plus coûteuse en temps ce qui représente un possible frein à son utilisation.

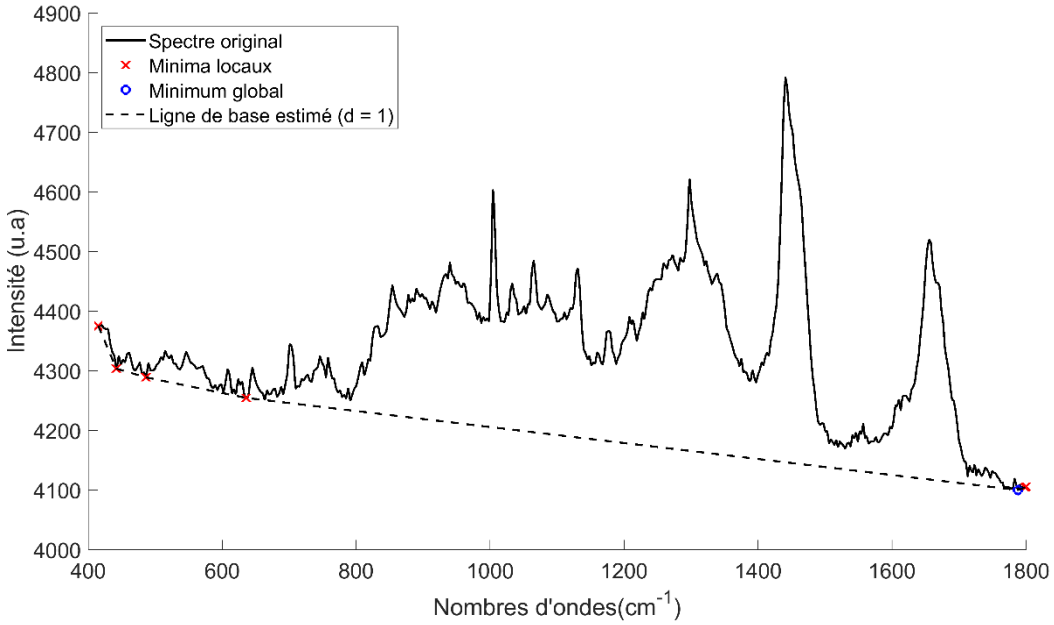


Figure III-9 – Estimation de la ligne de base d'un spectre débruité, enregistré à partir d'un modèle Episkin®, par la méthode Rubberband.

2.2.3. Méthode par minimisation semi-quadratique

La méthode de correction de ligne de base par minimisation semi-quadratique utilise des fonctions de coût non quadratiques, issues de la littérature des M-estimateurs robustes, adaptées pour l'approximation polynomiale. Habituellement lorsque l'approximation polynomiale se fait au sens des moindres carrés comme dans les méthodes Lieber et Rubberband, des valeurs de coût quadratique sont attribuées aux valeurs spectrales Raman de grandes intensités. L'utilisation de fonctions non quadratiques se justifie donc par l'attribution de faibles valeurs de coût à ces grandes intensités Raman. De cette façon, l'impact des bandes Raman de grandes intensités sur l'estimation des coefficients polynomiaux est réduit [75], [76].

Ces fonctions non-quadratiques sont construites en deux parties. La première partie est constituée d'une fonction quadratique au voisinage de zéro, modélisant le bruit de fond du spectre supposé Gaussien. La seconde partie est une fonction de croissance plus lente qu'une quadratique à partir d'un seuil t, modélisant la distribution des bandes Raman [75], [76]. La fonction Hubber et la fonction quadratique tronquée sont deux exemples de fonctions semi-quadratiques très utilisées. Elles sont respectivement constituées d'une fonction linéaire et d'une fonction constante pour leur seconde partie non-quadratique (Figure III-10, **A**). La partie linéaire de la fonction Hubber réduit l'influence des bandes de forte intensité sur l'estimation des coefficients polynomiaux utilisés dans l'estimation de la ligne de base tandis que la partie

constante de la fonction quadratique tronquée inhibe totalement leur impact. En effet, avec la fonction quadratique tronquée, une bande Raman d'intensité élevée affecte autant l'estimation de la ligne de base que celle d'une bande Raman de plus faible intensité *t* et ainsi seules les bandes Raman correspondant au signal de fond sont prises en compte dans l'estimation [76].
Par ailleurs, les fonctions non-quadratiques sont généralement définies de manière symétrique : un faible coût est associé aux grandes valeurs spectrales aussi bien positives que négatives. Cependant, les spectres Raman étant définis positifs, un ajustement asymétrique de ces fonctions non-quadratiques est préférable (Figure III-10, **B**): une faible valeur de coût est uniquement associée aux larges intensités spectrales positives tandis qu'un coût quadratique est associé aux larges intensités spectrales négatives qui, dans le cadre de la spectroscopie Raman, correspondent à du bruit Gaussien.

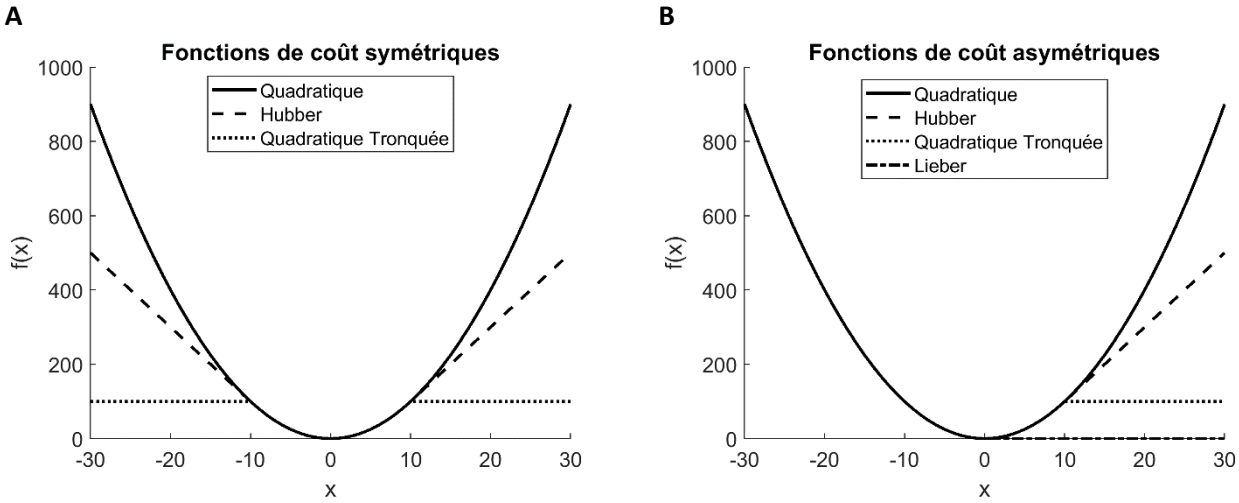


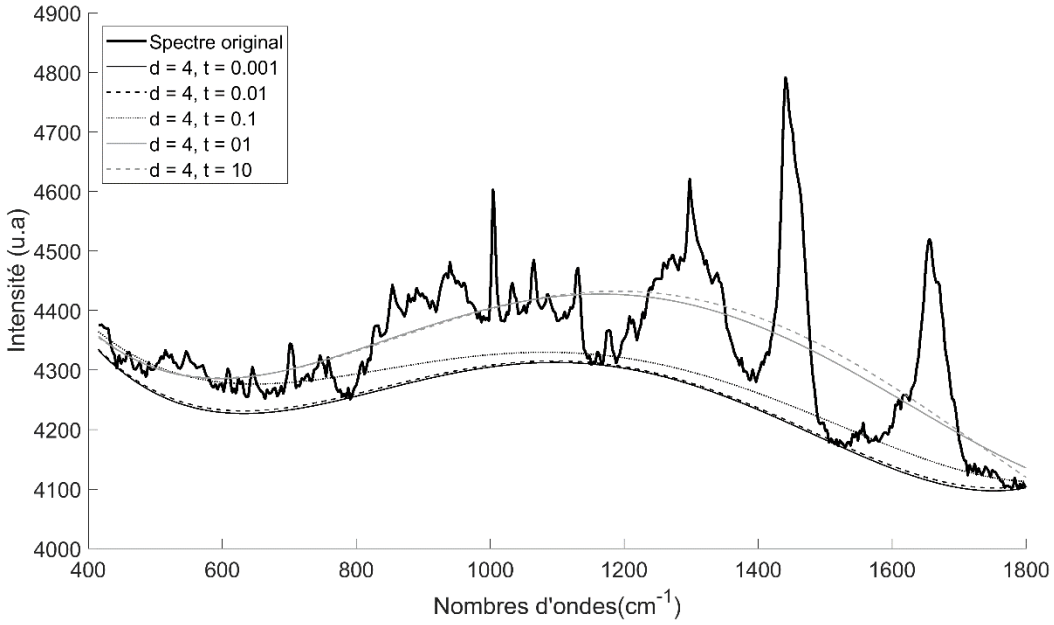
*Figure III-10 – Fonctions de coût non-quadratiques utilisées dans la méthode d'estimation de ligne de base par approximation polynomiale en réalisant une minimisation semi-quadratique. La fonction quadratique (ligne pleine) est illustrée pour comparaison des fonctions de coûts symétriques (**A**) et asymétriques (**B**).*

La méthode estime ensuite la ligne de base de façon itérative sur le même principe que la méthode Lieber. Cependant, la minimisation d'une fonction de coût non quadratique n'est pas aussi facile et explicite qu'une minimisation au sens des moindres carrés [76]. Pour simplifier cette opération, la technique de minimisation semi-quadratique consiste à ajuster un polynôme en minimisant une fonction quadratique tronquée possédant le même minimum global que la fonction de coût non-quadratique choisie [77]. À noter que le temps de calcul de cette méthode est quasiment proportionnel à la longueur des spectres et reste raisonnable pour de grands signaux [76].

Au final, en plus du choix de la fonction non-quadratique évaluant la proximité du polynôme estimé avec le spectre, deux autres paramètres restent à définir par l'utilisateur pour pouvoir estimer la ligne de base par minimisation semi-quadratique.

- À l'instar des méthodes Lieber et Rubberband, l'utilisateur doit préalablement définir le degré du polynôme d à estimer. Il permet d'ajuster le lissage du bruit de fond à estimer. Comme pour la méthode Lieber, un ordre polynomial trop grand fournit une ligne de base fortement influencée par les bandes Raman présentes dans le spectre et qui s'ajuste au bas du spectre ($d = 10$ dans la Figure III-11, **B**). Mais l'utilisateur doit définir un ordre suffisamment grand pour que le polynôme s'adapte aux irrégularités de la ligne de base, auquel cas la ligne de base sera une fois de plus grossière ($d = 1$ dans la Figure III-11, **B**). En définitif, l'ordre polynomial doit être fixé en fonction du signal de fond à ajuster mais aussi au besoin de l'utilisateur [68], [76].

A



B

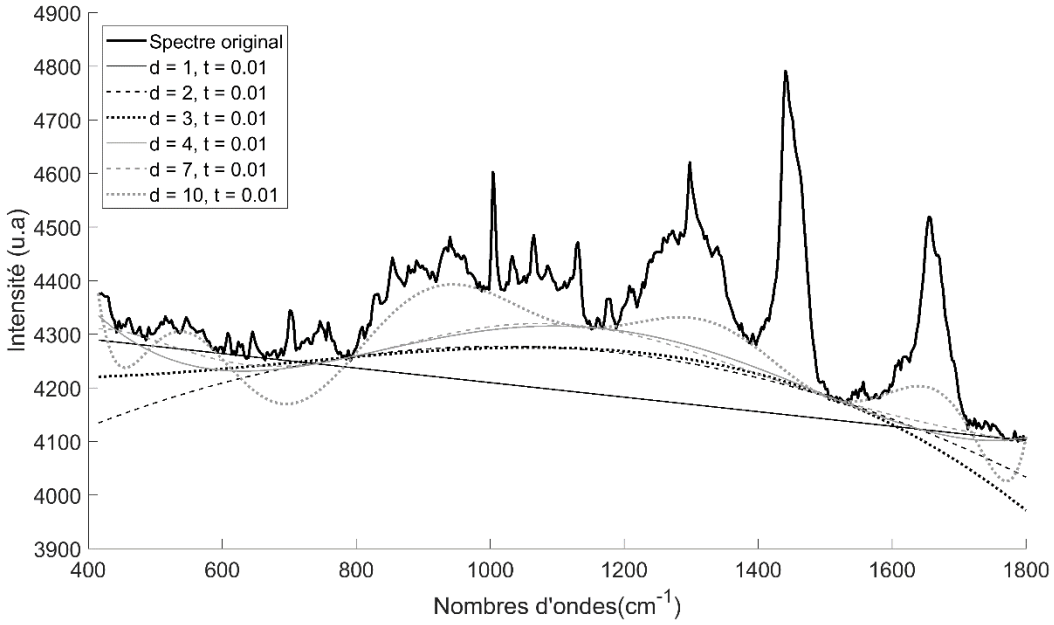


Figure III-11 – Estimation de la ligne de base d'un spectre débruité, enregistré à partir d'un modèle Episkin®, par la méthode de minimisation semi-quadratique avec (**A**) *un degré polynomial fixé à* $d = 4$ *et un seuil variant* $t = \{0.001, 0.01, 0.1, 1, 10\}$ *u.a et* (**B**) *un seuil fixé à* $t = 0.01$ *et un degré polynomial variant* $d = \{1,2,3,4,7,10\}$.

- Puis, l'utilisateur doit définir le seuil t définissant la frontière entre la partie quadratique de la fonction de coût correspondant au bruit de fond et la partie linéaire ou constante correspondant au poids des bandes Raman dans l'approximation polynomiale [75]. Pour une valeur de seuil trop grande, la fonction de coût tend vers une fonction quadratique et l'approximation polynomial revient à réaliser une estimation au sens des moindres carrés. Dans ce cas de figure, le polynôme estimé ne s'ajuste pas au bas du spectre ($s = 10$ dans la Figure III-11, **A**), ce qui par définition ne correspond pas véritablement à la ligne de base du spectre à corriger. À l'inverse, pour une faible valeur de seuil, la valeur de coût associée au bruit de fond est linéairement proportionnelle ou équivalente à celle associée aux bandes Raman de plus grande intensité, et par conséquent la ligne de base estimée tend à s'adapter au bas du spectre ($s = 0.001$ dans la Figure III-11, **A**). À

noter que la méthode Lieber n'est autre que la minimisation d'une fonction quadratique tronquée asymétrique dont le seuil est fixé à 0 à gauche (Figure III-10, **B**) [76]. Plus généralement, la valeur idéale du seuil dépend de la variance du bruit gaussien présent dans les données [75] et le seuil est défini comme optimal lorsqu'il est deux fois plus grand que l'écart-type du bruit [76]. Toutefois, le seuil optimal pour la fonction Hubber est inférieur à celui de la fonction quadratique tronquée pour réduire l'influence des bandes Raman [76].

2.2.4. Discussion

En conclusion, les lignes de base estimées par les méthodes Lieber et par minimisation semi-quadratique sont plus sujettes à des variabilités en raison des différents paramètres à estimer pour chaque spectre d'une image Raman. En effet, un compromis entre un ajustement par rapport au bruit de fond et la conservation de la forme traditionnelle d'un spectre Raman doit être trouvé et il est difficile d'établir un jeu unique de paramètres pour un ensemble de spectres, ce qui rend leur comparaison plus délicate. Bien qu'elle soit plus coûteuse en temps et qu'elle nécessite également la définition d'un ordre polynomial fixé la plupart du temps à 1, la méthode Rubberband assure une estimation de ligne de base spectrale épousant le bas du spectre grâce à son estimation par morceaux et la conservation de la forme traditionnelle d'un spectre une fois la correction effectuée. Cette technique d'approximation par morceaux permet à l'utilisateur de ne pas se soucier de la situation de la ligne de base estimée par rapport au spectre à corriger et elle assure une meilleure comparabilité des spectres. Par conséquent, nous préconisons l'utilisation de la méthode Rubberband pour corriger la ligne de base des spectres Raman par rapport à celle des méthodes Lieber et par minimisation semi-quadratique.

2.3. BRUIT DES RAIES COSMIQUES

Lors de l'acquisition Raman, des particules cosmiques de haute énergie peuvent frapper le détecteur CCD et perturber le bon enregistrement des spectres Raman. Ces évènements, connus sous le nom de rayons cosmiques, se produisent de manière aléatoire dans le temps, dans l'espace délimité par la surface du détecteur et en fréquence d'occurrence [78]. Ils provoquent des modifications importantes sur certains des spectres enregistrés bien que tous ne soient pas contaminés par ces rayons cosmiques. Ces perturbations se manifestent particulièrement par l'apparition d'une ou plusieurs bandes de forte intensité et de très faible largeur positionnées à des nombres d'ondes aléatoires (à 515 cm^{-1} dans le spectre **A** de la Figure III-14). Ces bandes d'intensité Raman très élevées sont alors considérées comme des valeurs aberrantes qui altèrent considérablement la présentation spectrale des informations de la composition moléculaire du matériau étudié. En effet, elles peuvent donner lieu à des distorsions spectrales et chevaucher potentiellement d'autres bandes d'intérêts masquant de ce fait des informations Raman utiles. Par conséquent, une pareille dégradation de la qualité de l'image Raman peut compliquer l'interprétation spectrale et induire des erreurs de calcul dans les analyses ultérieures [78]. Bien que la correction du bruit gaussien puisse supprimer des raies cosmiques indésirables, certaines peuvent rester présentes après cette correction. La correction de ces raies cosmiques est alors indispensable pour éliminer leur potentielle influence dans les futures analyses et doit être réalisée avant l'étape de prétraitement de normalisation faute de quoi leurs intensités risqueraient d'être mal interprétées. Pour les supprimer, nous proposons une méthode hybride combinant une méthode de localisation et une méthode de correction chacune régulièrement utilisée dans la littérature.

Dans la littérature, la correction des raies cosmiques présentes au sein des spectres se réalise généralement en deux étapes.

- La première étape consiste à identifier dans un ensemble de données les spectres contaminés par des raies cosmiques et obtenir la position de ces dernières, c'est-à-dire trouver les nombres d'ondes auxquels elles apparaissent. La méthode de détection la plus simple se base sur le critère aléatoire spatial et temporel des rayons cosmiques. Elle s'appuie donc sur la comparaison de mesures spectrales voisines [78]. Plus précisément, la présence de raies cosmiques s'évalue à travers les similitudes locales entre les spectres des pixels voisins. Quant à la localisation de ces artefacts, elle repose sur l'observation de changements d'intensité brutaux qu'ils peuvent provoquer et s'appuie donc sur leur caractéristique d'intensité très élevée.
- La seconde étape consiste à les supprimer. Cette suppression peut être effectuée de deux façons: l'exclusion des valeurs associées aux nombres d'ondes où se situent les raies cosmiques ou bien le remplacement de ces valeurs par une estimation locale. Toutefois, une valeur exclue peut être un problème pour l'application de méthodes d'analyses multivariées à causes des données manquantes. Par conséquent, la correction des bandes cosmiques par des estimations locales est préférable.

Par exemple, C. Behrend *et al.* [78] ont proposé une méthode de détection et de localisation des raies cosmiques semi-automatique et simple d'utilisation. Cette méthode repose sur les principes exposés ci-dessus : la similarité des spectres adjacents et les intensités très élevées des raies cosmiques. Concernant la suppression des raies cosmiques, ils utilisent une interpolation polynomiale du second ordre pour remplacer les valeurs des intensités spectrales de ces artefacts. De leur côté, O. Ryabchykov *et al.* [79] ont opté pour une correction des raies cosmiques en utilisant un procédé de « wound healing » (cicatrisation de plaie en anglais). Cette technique de suppression prend en compte les informations contenues dans les spectres voisins. Cela permet de mieux appréhender la correction de ces artefacts et d'augmenter la plausibilité de l'estimation locale qui en découle en comparaison à celles obtenues par interpolation polynomiale. Pour éliminer les raies cosmiques d'un ensemble de données Raman, nous proposons donc une méthode hybride s'appuyant sur ces deux modèles : une réadaptation de la méthode de détection et de localisation semi-automatique proposée par C. Behrend *et al.* à des images Raman à trois dimensions et un remplacement de l'étape de correction des raies cosmiques par la méthode de « wound healing » utilisée par O. Ryabchykov *et al.*

La méthode hybride proposée, restant semi-automatique, nécessite toutefois la définition de deux paramètres qui doivent être ajustés à chaque ensemble de données pour être appliquée. Le premier paramètre à définir est la taille de la fenêtre de voisinage $w = 2l + 1$ ($l \in \mathbb{N}$ est la demi-largeur) qui est utilisée à la fois pour analyser la similarité des spectres adjacents et pour estimer localement la nouvelle valeur associée au nombre d'onde contaminé par une raie cosmique. Le second paramètre à définir de façon interactive avec l'algorithme est le seuil T. Il permet à la fois d'identifier les spectres contaminés par les raies cosmiques et de trouver leur position dans les spectres estimés comme contaminés. À noter que le choix du seuil T dépend du jeu de données à analyser et est défini uniquement selon l'appréciation de l'utilisateur. Les détails de chaque étape de la méthode réadaptée sont présentés ci-dessous.

Étape 1 : Identification des spectres contaminés et localisation des raies cosmiques

Le processus commence par l'application d'un filtre médian sur l'image Raman 3D à corriger. Le principe de ce filtre consiste à remplacer chaque valeur spectrale par la médiane des valeurs spectrales contenues dans une fenêtre de voisinage de taille $w_x \times w_y \times w_\lambda$ préalablement

définie. La taille w_λ de cette fenêtre de voisinage détermine la largeur des variations spectrales qui sont effacées des données. Plus exactement, elle détermine le degré de lissage des données spectrales qui sont alors grossièrement dépourvues du bruit et des éventuelles raies cosmiques de largeur inférieure à elle (Figure III-14, **B**).

Toutefois, l'utilisation d'une fenêtre de voisinage restreint la zone de traitement de l'image. En effet, la taille de la zone traitée de l'image est réduite de $2l$ dans chacune des dimensions, soit égale à $(N_x - 2l) \times (N_y - 2l) \times (N_\lambda - 2l)$. Pour que tous les spectres de l'image Raman soient filtrés et analysés par la suite, des conditions aux bords de Neumann sont adoptées. La Figure III-12 illustre l'utilisation de ces conditions, qui se matérialisent par une « réflexion » : les valeurs des l premières et dernières variables de chaque dimension spatiale et spectrale sont recopiées en miroir autour de l'image traitée. La taille de l'image Raman à corriger est alors augmentée et devient $(N_x + 2l) \times \big(N_y + 2l\big) \times (N_\lambda + 2l)$. Ainsi, la fenêtre de voisinage du filtre médian, illustrée par le cadre jaune centré sur le pixel jaune dans la Figure III-12, peut être centrée sur le premier pixel de l'image et tous les spectres de l'image Raman peuvent être alors lissés et analysés.

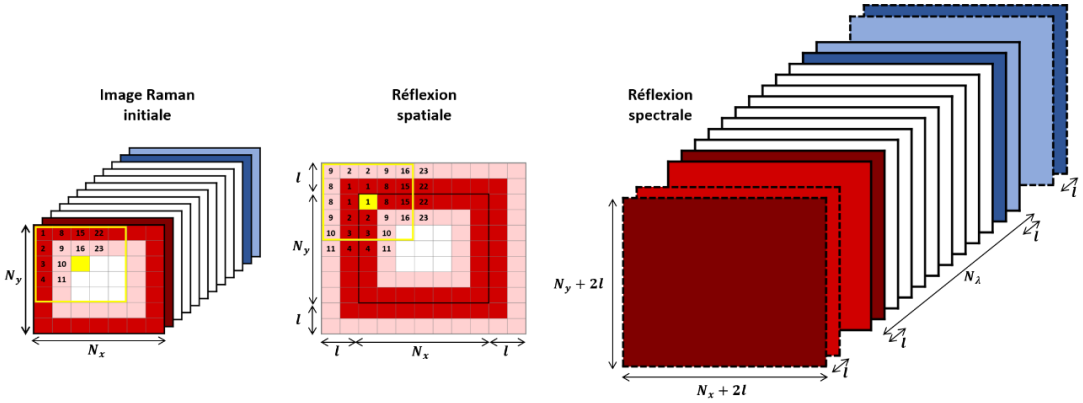


Figure III-12 – Processus de réflexion d'une image Raman à deux coordonnées spatiales N_x et N_y et une coordonnée spectrale N_λ à l'aide d'une fenêtre de voisinage de taille (w_x, w_y, w_λ) où $w_x = w_y = w_\lambda = 2l + 1$ (carré jaune centré sur le pixel jaune).

L'image obtenue après application du filtre médian sur l'image Raman élargie n'est utilisée que dans le but d'évaluer les similitudes locales et plus précisément d'identifier les meilleures correspondances spectrales entre spectres voisins. Comme illustré dans la Figure III-13, **A**-**B**, les coefficients de corrélation entre chaque spectre lissé et ses spectres voisins contenus dans une fenêtre de voisinage de taille $w_x \times w_y$ sont calculés. Le spectre avec le coefficient de corrélation le plus élevé dans le voisinage local est localisé et son emplacement dans l'image Raman est enregistré (Figure III-13, **C**). Celui-ci sera par la suite utilisé comme point de comparaison pour identifier les spectres contaminés par des raies cosmiques.

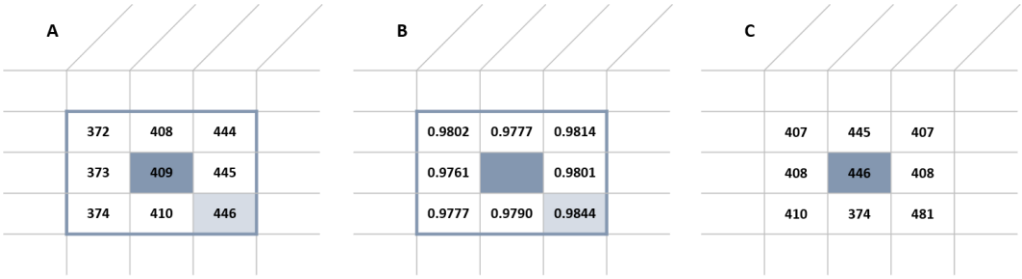


*Figure III-13 – (**A**) Fenêtre de voisinage, où $l = 1$, centrée sur le pixel n°409 d'une image Raman obtenue à partir d'un échantillon d'épiderme humain reconstruit non traité et les pixels voisins; (**B**) Coefficients de corrélation de chaque pixel voisin avec le pixel central n°409. (**C**) Enregistrement du numéro du pixel voisin le plus fortement corrélé de chaque pixel.*

L'identification des spectres contaminés est ensuite réalisée en comparant l'intensité de chaque spectre non filtré avec celle de son spectre voisin le plus fortement corrélé non filtré. Cependant pour éviter que cette comparaison ne soit erronée, les spectres sont préalablement corrigés de leur ligne de base. De plus, une raie cosmique étant une caractéristique spectrale unidirectionnelle, distinguable au-dessus du bruit du détecteur [1], les variations spectrales dues au bruit aléatoire doivent être prises en compte dans cette comparaison. Par conséquent, la similitude entre un spectre et son voisin le plus fortement corrélé est évaluée en calculant la différence suivante:

$$\Delta_i = S_i - S_{i+x} - \sigma\varepsilon \qquad \text{(III-1)}$$

où S_{i+x} est le spectre voisin le plus corrélé avec le spectre S_i et $\sigma\varepsilon$ représente une estimation du bruit aléatoire présent dans les données avec σ l'écart-type et ε le bruit contenu dans les spectres estimé par [1] :

$$\varepsilon = E_{max} - E_{min}$$

Une différence spectrale élevée indique que le spectre exprime vraisemblablement une raie cosmique à cette position. Par exemple, la différence spectrale significativement très élevée autour de 515 cm^{-1} sur le profil **D** de la Figure III-14 révèle très certainement la raie cosmique observée dans le spectre contaminé **A** en cette position. En somme, seuls les spectres présentant au moins une différence spectrale significativement grande et supérieure à un seuil T ($T = 365$ dans la Figure III-14) défini par l'utilisateur suite à l'observation simultanée de tous les profils de différences spectrales sont susceptibles d'être contaminés.

Toutefois, de par le positionnement spatial et la fréquence d'apparition aléatoires des raies cosmiques, la probabilité que plusieurs spectres aient une raie cosmique à un même nombre d'onde est quasi nulle. Ainsi, une différence spectrale significativement grande à un nombre d'onde donné est considérée comme révélatrice de la présence d'une raie cosmique seulement si elle n'est présente à cette position que pour un seul spectre parmi les autres. Autrement dit, parmi les spectres considérés comme susceptibles d'être contaminés à l'aide du seuillage, seuls ceux dont la différence spectrale supérieure au seuil T n'est pas présente pour les autres spectres sont conservés et identifiés comme contaminés par une raie cosmique. Pour exemple, les profils des différences spectrales de trois autres spectres issus de la même image Raman sont illustrés dans la Figure III-15. Chacun présente des différences spectrales supérieures au seuil T sur les mêmes plages de nombre d'ondes, à savoir [1293-1301] cm^{-1} et [1433-1449] cm^{-1}. De plus, leur allure aux alentours de ces régions spectrales ne concorde pas avec celle d'un changement d'intensité Raman brutal que peut provoquer l'apparition d'une raie cosmique. Par conséquent, ces spectres

auxquels sont associés ces profils de différences spectrales ne sont pas retenus comme contaminés par des raies cosmiques.

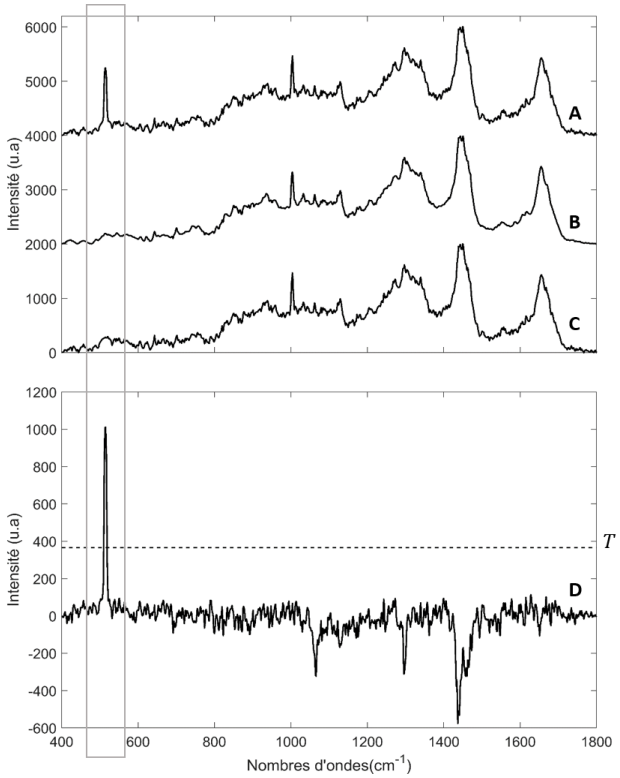


Figure III-14 – (**A**) *Spectre contaminé par un pic cosmique à 515 cm*$^{-1}$ *(pixel n°409) ;* (**B**) *Spectre lissé par un filtre médian utilisant une fenêtre de voisinage avec* $l = 3$ *;* (**C**) *Spectre corrigé de son pic cosmique par la méthode* hybride *;* (**D**) *Profil des différences spectrales entre le spectre contaminé et son spectre voisin le plus corrélé (pixel n°446) ;* $T = 365$ *: seuil défini de manière interactive en observant tous les profils des différences spectrales en même temps. Les spectres analysés sont issus d'un échantillon d'épiderme humain reconstruit (Episkin®) non traité.*

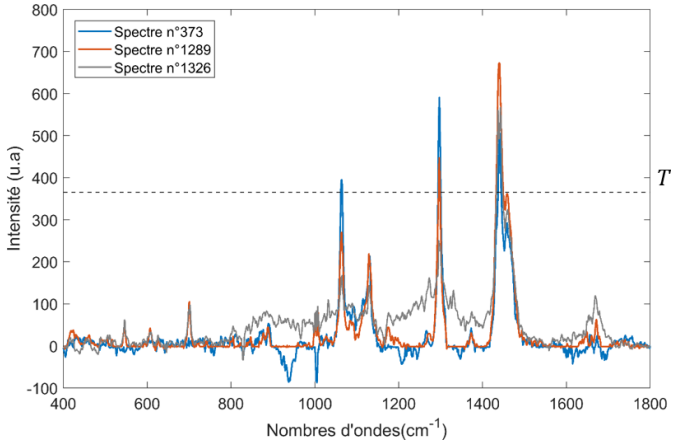


Figure III-15 – Profils des différences spectrales d'une sélection de 3 spectres susceptibles de comporter *un bruit cosmique avec leur spectre voisin le plus corrélé respectif.* $T = 365$ *: seuil défini de manière interactive en observant tous les profils des différences spectrales en même temps.* Les différences spectrales associées aux bandes spectrales situées à 1297 cm^{-1} et 1441 cm^{-1} sont supérieures au seuil T pour au moins deux des trois spectres, aucune raie cosmique n'est donc présente à ces nombres d'ondes. *Les spectres analysés sont issus d'un échantillon d'épiderme humain reconstruit (Episkin®) non traité.*

En résumé, le processus de seuillage des différences spectrales permet de localiser les raies cosmiques tant bien dans l'espace qu'au niveau de la dimension spectrale. Plus exactement, la localisation spectrale des éventuelles raies cosmiques indique en même temps quels sont les spectres susceptibles d'être contaminés par ces artefacts et donc leur localisation spatiale. Les nombres d'ondes pour lesquelles des différences spectrales sont observées comme supérieures au seuil T représentent la position de ces éventuels artefacts dans les spectres susceptibles d'être contaminés. Alors que la présence d'une différence spectrale significativement grande pour des nombres d'ondes dans plusieurs spectres permet d'affirmer que ces nombres d'ondes ne désignent pas la position des raies cosmiques, l'unicité de cette différence spectrale dans l'ensemble des spectres permet de considérer que ces nombres d'ondes définissent la position spectrale des raies cosmiques. Ces nombres d'ondes correspondent généralement à des nombres d'ondes consécutifs. Cette consécution s'explique par le fait que la variation brutale d'intensité Raman que provoque l'apparition d'un pic cosmique à un nombre d'onde aléatoire se répercute sur les intensités spectrales des nombres d'ondes voisins. Celles-ci accusent aussi une augmentation et les différences d'intensité entre les spectres et leur spectre voisin le plus corrélé en sont par conséquent impactées. Par exemple, concernant le spectre contaminé par un pic cosmique à 515 cm^{-1} (Figure III-14, **A**), tous les nombres d'ondes compris entre 511 et 518 cm^{-1} présentent une différence spectrale supérieure au seuil T (Figure III-14, **D**) . Ils sont donc retenus comme révélateur de la position du pic cosmique. Plus encore, la quantité de nombre d'ondes reflète en partie le nombre d'informations spectrales erronées par le pic cosmique dans le spectre contaminé. Ainsi, pour que la correction des raies cosmiques soit optimale, l'ensemble des intensités Raman associées aux nombres d'ondes estimés comme affectés par une modification spectrale significative suite à la manifestation d'un pic cosmique doit être corrigé.

Étape 2 : Correction des raies cosmiques

La correction des raies cosmiques à l'aide du processus de « wound healing » proposé par O. Ryabchykov [79] est un schéma d'interpolation par définition similaire à un processus de cicatrisation. La comparaison avec une cicatrisation résulte du fait que les données spectrales endommagées ne sont pas interpolées en une seule étape, mais que la zone de l'image impactée par une raie cosmique est réduite de manière itérative étape par étape. La méthode s'apparente à l'exécution locale d'un filtre médian sur la zone de contamination à la différence que toutes les valeurs du voisinage ne sont pas prises en compte dans le calcul. Plus précisément, la méthode consiste à remplacer itérativement les intensités spectrales affectées par une raie cosmique par les médianes de celles des spectres voisins les plus proches qui sont non contaminés ou déjà corrigés aux itérations précédentes par le processus. Les différentes étapes de ce processus sont illustrées dans la Figure III-16 pour la correction d'une image à deux dimensions avec une fenêtre de voisinage définie par $l = 2$. La sous-image n°0 montre une image possédant une raie cosmique marquée par un pixel rouge avant la première itération. La première itération, illustrée par la sous-image n°1, débute la correction par celle des pixels verts situés aux coins de la zone contaminée. Leurs valeurs sont interpolées à l'aide de celles des pixels bleus voisins. Pour plus de clarté, les pixels bleus ne sont affichés que pour une partie des pixels verts. À la dernière itération, soit la sous-image n°7, chaque valeur de la zone de contamination est corrigée à l'aide des valeurs voisines. Une réadaptation de cet algorithme a été réalisée pour corriger des artefacts présents dans des images à trois dimensions avec une fenêtre de voisinage de taille $w_x \times w_y \times w_\lambda$.

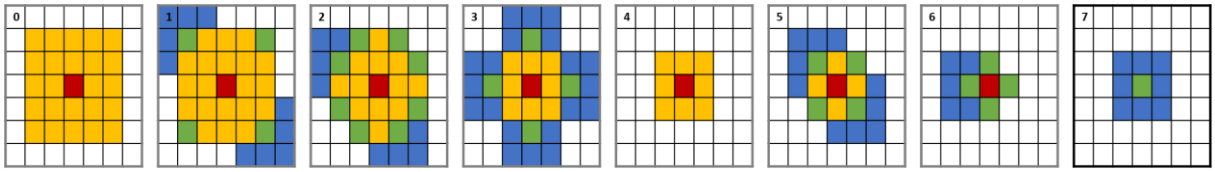


Figure III-16 – Illustration du processus de « wound healing » en 2D pour la correction d'une valeur aberrante [79] *représentée par le pixel rouge. Les pixels jaunes représentent les valeurs voisines de la valeur centrale à corriger. À chaque étape, la médiane des pixels bleus situés autour d'un pixel vert est attribuée à ce dernier. Les pixels bleus ne sont affichés que pour certains pixels verts pour des raisons de clarté.*

Au final, le processus de correction de raies cosmiques par « wound healing » est uniquement appliqué sur les spectres identifiés comme contaminés au niveau de leurs régions spectrales endommagées. Cependant, la correction des intensités Raman d'un spectre contaminé par une raie cosmique sur la zone spectrale concernée implique également la correction des intensités Raman des spectres voisins sur cette même région spectrale même si ces dernières peuvent ne pas avoir été impactées par la raie cosmique. Par conséquent, seule la correction des intensités Raman du spectre considéré comme contaminé est prise en considération. Celle des intensités Raman des spectres voisins est quant à elle ignorée et les valeurs spectrales originales sont conservées.

En définitive, cette méthode de correction du bruit cosmique se montre adaptée au prétraitement d'une image Raman provenant d'échantillons de peau. En effet, l'homogénéité des couches cutanées justifient l'utilisation du voisinage local pour localiser les raies cosmiques dans les dimensions spatiales et spectrales et pour corriger les intensités Raman des nombres d'ondes endommagées par l'apparition d'une raie cosmique. D'autre part, la suppression de l'altération spectrale engendrée par l'introduction d'artefacts à l'aide de cette méthode est efficace (Figure III-17) et permet par conséquent d'éviter toute interprétation spectrale corrompue.

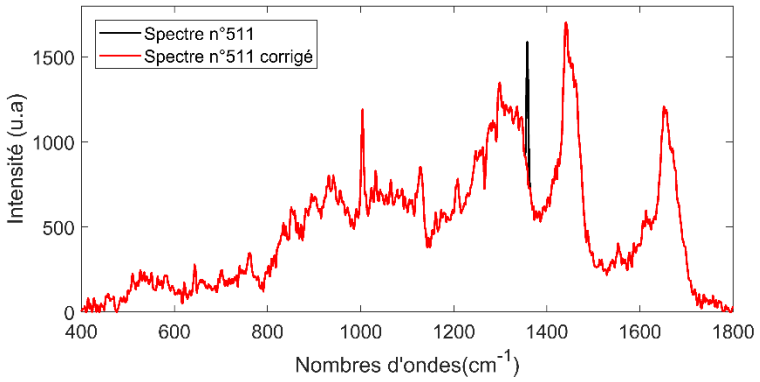


Figure III-17 – Comparaison d'un spectre contaminé par une raie cosmique à 1358 cm^{-1} (pixel n°511) avant (courbe noire) et après (courbe rouge) correction par la méthode hybride. Les courbes noires et rouges se superposent

3. CALIBRATION DE LA RÉSOLUTION SPECTRALE ET ALIGNEMENT SPECTRAL

Lors de l'acquisition de données avec un spectromètre Raman utilisant un capteur CCD, un espacement irrégulier entre les variables spectrales des données peut se produire en raison des différents paramètres d'étalonnage du système [42], [74]. Cette irrégularité s'illustre par une non-linéarité de l'axe des abscisses (Figure III-18). Ces variations d'intervalle d'abscisses sont potentiellement dues à des erreurs induites par la limitation du spectromètre et du sous-

échantillonnage latent du capteur CCD [74]. De même que les variations de température et/ou d'humidité d'un jour à l'autre, la variabilité de l'axe spectral des données impacte considérablement la position des bandes Raman. En effet, par manque de résolution spectrale suffisante pendant l'acquisition, certaines bandes caractéristiques peuvent être décalées de quelques nombres d'ondes.

Afin de rétablir le bon alignement des bandes caractéristiques spectrales, une calibration de l'axe spectral peut être nécessaire. En effet, le réalignement des bandes spectrales ne peut se faire que si la résolution spectrale des données est correctement calibrée. La calibration consiste à créer des intervalles entre nombres d'ondes réguliers. Plus précisément, de nouveaux points de données sont créés entre ceux déjà existants par interpolation linéaire de façon à ce que l'intervalle d'abscisse devienne constant [74]. Pour nos applications, la correction de la résolution spectrale des données est réalisée en utilisant l'algorithme de linéarisation spectrale de P. Knief [74]. Chaque point de données est espacé de 1 cm^{-1} et les nombres d'ondes sont réduits à des décimales complètes. Les spectres de peau avant et après linéarisation spectrale ne présentent aucun changement perceptible sur la forme du spectre. Finalement, la linéarisation spectrale permet de faciliter le réalignement spectral des données.

Enfin, des spectres de références ou des connaissances *a priori* peuvent être utilisés pour comparer et évaluer l'éventuel décalage existant. Dans le cas où l'ensemble des bandes spectrales sont décalées d'un même nombre d'onde, le réalignement du spectre est réalisé par une translation. Dans le cadre de nos applications cutanées, les bandes caractéristiques des spectres sont toutes décalées d'un même nombre d'onde et la bande Raman illustrant le mieux se décalage se situe à 1004 cm^{-1}. Une attention particulière lui est donc accordée. Les cartographies Raman acquises pour nos expériences révèlent cette bande très caractéristique de la peau autour de 1001 cm^{-1} (Figure III-19). Chaque spectre est donc translaté de façon à ce que cette bande soit centrée à 1004 cm^{-1} (Figure III-19). Au final, la linéarisation spectrale et le réalignement spectral des données facilitent l'identification et l'interprétation des caractéristiques d'un échantillon.

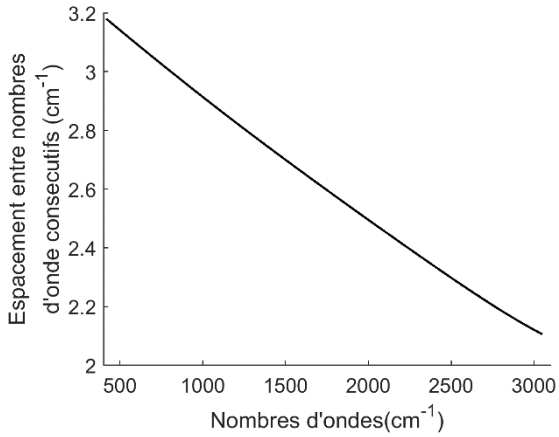


Figure III-18 – Variations de l'espacement entre deux nombres d'ondes enregistrés d'un spectre issu d'une acquisition Raman d'un modèle Episkin®.

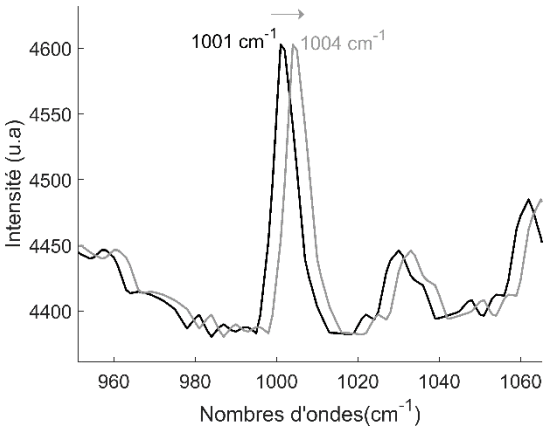


Figure III-19 – Translation d'un spectre enregistré à partir d'un modèle Episkin® de 3 cm^{-1} vers la droite. Alignement spectral sur la bande Raman située à 1004 cm^{-1}.

4. NORMALISATION

D'une acquisition Raman à une autre, la densité et la composition de l'échantillon ainsi que la focalisation de la source laser à sa surface changent. Elles peuvent donc être sources de variations d'intensité du signal Raman en raison des différentes quantités de photons dispersés sur le capteur. Pour minimiser et compenser ces effets de variations d'intensité à la source, une mise à l'échelle des données sur une plage de valeurs similaires est de rigueur. Cette normalisation permet de corriger l'analyse des différentes géométries des contributions spectrales Raman et rend possible une comparaison quantitative des spectres Raman obtenus avec différents spectromètres et sources laser [68], [69], [74]. Elle est donc une étape de prétraitement des plus importantes et doit être appliquée après que les contributions du bruit aient été supprimées [42]. La normalisation constitue généralement la dernière étape de prétraitement des données, et de nombreux algorithmes existent. Dans la littérature, l'algorithme employé est généralement défini selon les attentes de l'utilisateur. Parmi ces nombreux algorithmes, ceux utilisant uniquement les informations issues des spectres sont rapides et les plus simples à mettre en place. La normalisation centrée-réduite, la normalisation min-max et la normalisation $L_{p=\{1,2\}}$ en sont des exemples les plus couramment utilisés en spectroscopie Raman [42], [68], [69] et leur comparaison est présentée ci-dessous.

4.1. NORMALISATION CENTRÉE-RÉDUITE

La normalisation consistant à centrer et réduire une variable est très utilisée en statistiques. En spectroscopie, elle consiste à considérer chaque spectre comme un sous-ensemble de données à standardiser. La moyenne et l'écart-type de chaque spectre sont calculées. Le spectre normalisé est ensuite obtenu en soustrayant sa moyenne et divisant le résultat de cette différence par son écart-type (Figure III-20, **A**) [68], [69]. Ainsi, les intensités spectrales reflètent la distance par rapport à la moyenne en unités d'écart-type et la comparaison des variations spectrales avec d'autres spectres normalisés est facilitée. Cependant, le spectre ainsi normalisé ne vérifie plus toutes les propriétés propres à un spectre Raman, et plus particulièrement la non-négativité n'est plus respectée. Par conséquent, cette normalisation n'est pas toujours la plus adaptée selon l'objectif visé. Par exemple, la présence de valeurs négatives peut être un frein à une analyse ayant pour objectif le démélange des spectres Raman. De ce fait, cette normalisation ne sera pas utilisée pour les applications présentées par la suite.

4.2. NORMALISATION MIN-MAX

La normalisation min-max est également une méthode de normalisation simple et couramment utilisée en spectroscopie Raman. Elle répartit les intensités spectrales entre 0 et 1 afin de faciliter la comparaison de plusieurs spectres. Elle consiste à décaler les intensités spectrales de sorte que l'intensité minimale de l'ensemble du spectre ou d'une région spectrale prédéfinie soit nulle et que son intensité maximale soit égale à un (Figure III-20, **B**) [68], [69]. Plus précisément, l'opération appliquée à chaque intensité d'un spectre est la suivante :

$$\frac{D_{i,j} - \min(D_i)}{\max(D_i) - \min(D_i)} \tag{III-2}$$

où D_i est le spectre de données d'index $i \in [1, \dots, N]$ avec N le nombre de spectres dans l'image et $D_{i,j}$ l'intensité du spectre i au nombre d'ondes $j \in [1, \dots, N_\lambda]$.

Ainsi, contrairement à la normalisation centrée-réduite, la non-négativité spectrale est respectée. Cependant, lorsque cette normalisation est appliquée à un ensemble de spectres, leur

interprétation et leur comparaison peuvent en être affectées. En effet, l'intensité Raman d'une bande spectrale est liée à la quantité de molécules qu'elle caractérise et donc à la concentration du composant présent dans l'échantillon analysé. Si l'intensité d'une bande spectrale varie d'un spectre à un autre, autrement dit si les concentrations d'un composant évoluent, et si de plus cette bande est d'intensité maximale pour chaque spectre, alors la normalisation min-max échelonne chaque spectre de façon à ce que la contribution de cette bande soit égale à 1. De ce fait, la distinction entre les différentes variations d'intensité originales n'est plus aussi évidente et ces changements de concentrations ne sont alors distinguables qu'en analysant les ratios de bandes avec des bandes spectrales caractéristiques qui sont invariantes d'un spectre à un autre. Par conséquent, l'utilisation de la normalisation min-max n'est pas idéale pour analyser l'évolution des contributions des différents composants dans un échantillon et ne sera donc pas utilisée pour les applications suivantes.

4.3. NORMALISATION $\ell_p,\ p = \{1, 2, \infty\}$

La normalisation ℓ_p regroupe différentes façons d'harmoniser des spectres pour rendre leur comparaison plus aisée. De manière générale, elle consiste à diviser chaque intensité spectrale par la norme du spectre en question [69] de sorte qu'elle soit comprise entre 0 et 1. Ainsi, contrairement aux normalisations centrée-réduite et min-max, la non-négativité et la forme du spectre sont préservées. Cependant, de nombreuses normes existent, les plus courantes s'exprimant de façon générale de la façon suivante :

$$\|D_i\|_p = \frac{D_{i,j}}{\left(\sum_{j=1}^{N_\lambda} |D_{i,j}|^p\right)^{\frac{1}{p}}} \qquad \text{(III-3)}$$

où D_i est le spectre de données d'index $i \in [1, \dots, N]$ avec N le nombre de spectres dans l'image, $D_{i,j}$ l'intensité du spectre i au nombre d'ondes $j \in [1, \dots, N_\lambda]$ et p un entier positif.

Le choix de la norme utilisée détermine la méthode de normalisation appliquée [42], [68], [69]. En choisissant $p = 2$, soit la norme euclidienne appelée aussi norme vectorielle, la normalisation consiste à considérer que la somme de toutes les intensités spectrales au carré vaut 1 (Figure III-20, **D**). Pour $p = \infty$, la normalisation revient à échelonner chaque intensité spectrale selon le maximum du spectre et revient par conséquent à utiliser la normalisation min-max (Figure III-20, **B**). Enfin si $p = 1$, alors la somme de toutes les intensités spectrales en valeur absolue est égale à un et donc l'aire sous la courbe spectrale vaut un (Figure III-20, **C**). On constate que les intensités des spectres Raman normalisés par la norme ℓ_p deviennent plus faibles lorsque p diminue : elles sont d'ordre 10^{-1} pour la normalisation ℓ_∞, d'ordre 10^{-2} pour la normalisation ℓ_2 et d'ordre 10^{-3} pour la normalisation ℓ_1 (Figure III-20). Les faibles intensités spectrales obtenues par normalisation ℓ_1 peuvent être un frein à son utilisation. En effet, le calcul de l'inverse d'une matrice couramment utilisé dans des algorithmes de résolutions de courbes multivariées, comme ceux présentés dans le chapitre suivant, peut être problématique en présence de trop faibles valeurs d'intensités spectrales. Quant à la normalisation ℓ_∞ qui est équivalente à la normalisation min-max, son utilisation est comme précisé ci-dessus pas adaptée à des études d'évolution de contributions biochimiques des différents composants d'un échantillon. En revanche, à la différence de toutes les normalisations jusqu'ici présentées, la normalisation ℓ_2 ne présente aucun inconvénient majeur et c'est pourquoi elle est la normalisation la plus utilisée dans le domaine de la spectroscopie Raman. Par conséquent, toutes les données présentées ci-après seront normalisées par normalisation vectorielle.

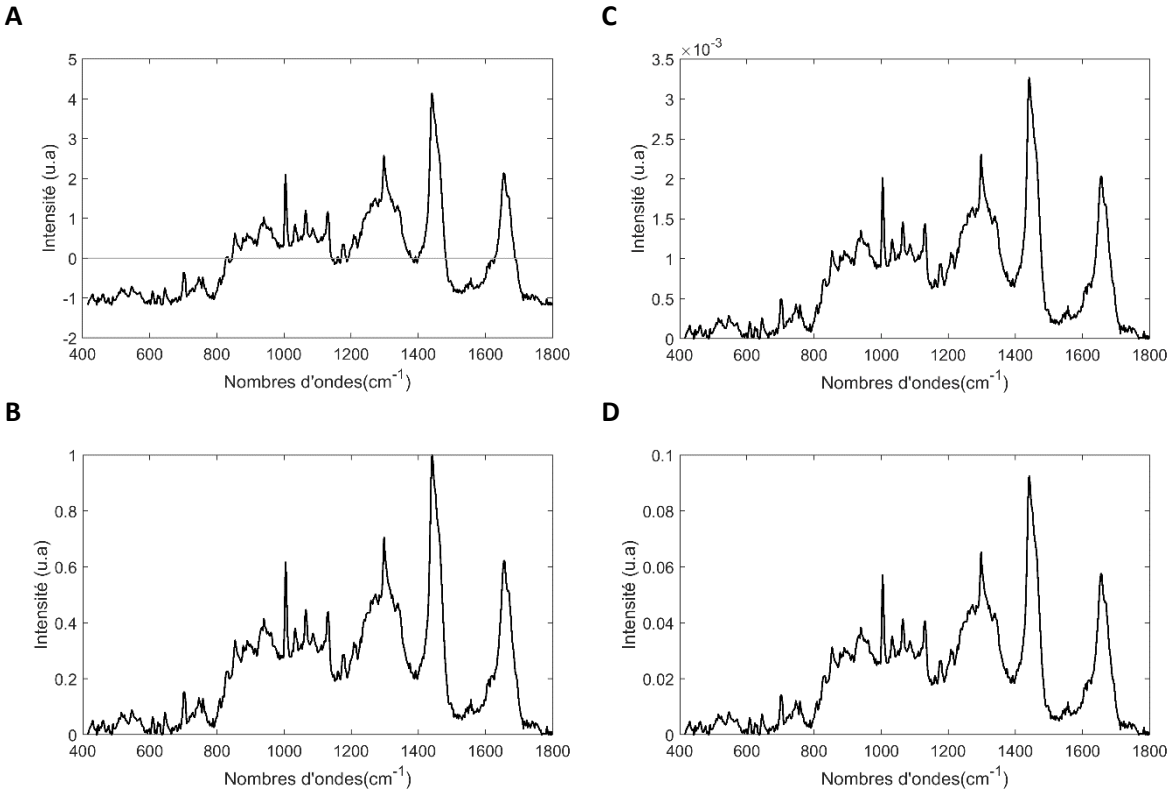


Figure III-20 – Normalisations d'un spectre débruité et corrigé de sa ligne de base, enregistré à partir d'un modèle Episkin® : (**A**) *normalisation centrée-réduite,*(**B**) *normalisation Min-Max, normalisation* ℓ_p *avec* (**C**) $p = 1$,(**D**) $p = 2$.

5. PROPOSITION D'UNE ROUTINE DE PRÉTRAITEMENT DES DONNÉES RAMAN

Aucune standardisation de protocoles de prétraitement de données spectrales Raman n'a jusqu'à présent été établie [42]. En effet, le protocole de prétraitement des données se base généralement sur des considérations pratiques plutôt que sur une analyse théorique et comparative préalable [68]. Plus précisément, l'ordre d'application des différentes méthodes de prétraitement repose souvent sur l'expérience et l'intuition de l'utilisateur et est établi à son appréciation. Toutefois, le choix de l'enchainement des différentes méthodes de prétraitements n'est pas sans importance puisqu'il peut amener à des résultats finaux très différents les uns des autres. Par exemple, la méthode de normalisation choisie doit absolument être appliquée en dernier, faute de quoi un faible bruit électronique ou un bruit de fond corrigé *a posteriori* pourrait complètement masquer des informations spectrales pertinentes. Par conséquent, l'enchainement des méthodes de prétraitement doit être pensé de manière logique et être spécifié pour chaque nouvelle application.

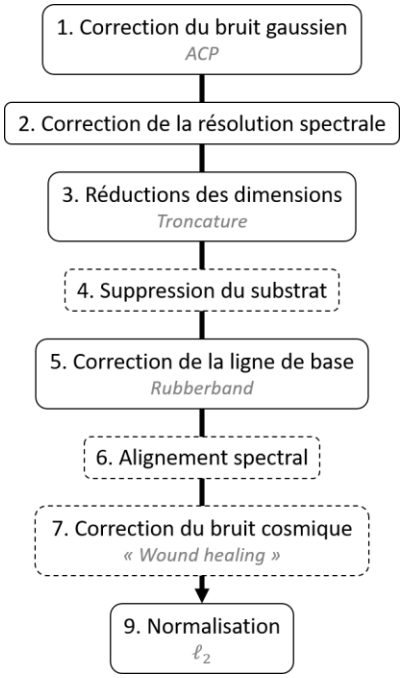


Figure III-21 – Schéma de la routine de prétraitements des images Raman proposée. Les étapes n°4, 6 et 7 sont appliquées au cas par cas.

Dans le cadre des études de pénétrations cutanées menées dans ce projet, l'ordre d'application des différentes méthodes de prétraitement estimées comme les plus adaptées aux jeux de données utilisés a été considéré de la manière suivante (Figure III-21) :

- De toute évidence, le bruit gaussien des données est corrigé en premier par la méthode de projection par ACP. Ainsi, le bruit ne subsiste pas dans les données tout au long du processus de prétraitement et risque moins par conséquent d'être amplifié par une autre méthode de correction (étape n°1 de la Figure III-21).

- La calibration de la résolution spectrale et la sélection des régions spectrales d'intérêts par troncature des données Raman sont ensuite appliquées (étapes n°2 et n°3 de la Figure III-21). Elles permettent de préparer les données de manière à ce qu'une interprétation des données spectrales soit plus efficace et précise.

- Dans le cas où l'échantillon étudié ne couvre pas la totalité du plan spatial de l'image Raman enregistrée, les spectres Raman associés au substrat sont annulés (étape n°4 de la Figure III-21). Ces spectres se distinguent par leurs très faibles variations spectrales. Pour les identifier, la méthode des K-means est appliquée sur la matrice contenant l'écart-type de chaque spectre. Puis les spectres associés au cluster dont l'écart-type moyen est le plus faible sont remplacés par des spectres nuls. Ainsi, les spectres associés au substrat contenant aucune information pertinente n'interfèreront pas dans les analyses multivariées appliquées ultérieurement.

- Puis, les spectres sont corrigés de leur ligne de base avec la méthode Rubberband (étape n°5 de la Figure III-21) afin d'améliorer leur lisibilité.

- Lorsqu'il est nécessaire, un réalignement spectral est appliqué sur la base d'une connaissance *a priori* sur la position d'une bande caractéristique de l'échantillon (par exemple, la bande située à 1004 cm^{-1} pour la peau) pour supprimer les éventuels décalages en nombres d'ondes présents dans les données (étape n°6 de la Figure III-21).

- Dans le cas où les étapes de correction de bruit électronique et de la ligne de base n'ont pas suffi à éliminer le bruit provenant des particules cosmiques, la méthode de correction du bruit cosmique de « wound healing » est conseillée pour les éliminer (étape n°7 de la Figure III-21).

- Enfin, les spectres sont normalisés à l'aide de la méthode de normalisation vectorielle (étape n°8 de la Figure III-21) pour qu'ils puissent être comparés entre eux mais aussi avec des spectres issus d'autres cartographies Raman et prétraités de la même manière.

CONCLUSION

En résumé, de nombreuses méthodes de prétraitement spectral existent dans la littérature et bon nombre d'entre elles sont régulièrement utilisées en spectroscopie Raman. Comme cela a été évoqué tout au long de ce chapitre, toutes jouent un rôle précis dans la préparation des données pour une analyse ultérieure: certaines corrigent le bruit présent dans les données ou améliorent la résolution spectrale alors que d'autres sélectionnent les données d'intérêts ou encore mettent les données à la même échelle. Cependant, il n'existait à ce jour aucun protocole de prétraitement standardisé pour traiter des données spectrales Raman. Dans ce contexte, l'étude comparative des avantages et des inconvénients de plusieurs de ces méthodes dans ce chapitre a permis de définir celles étant les plus adaptées à la correction des jeux de données utilisés dans les études de pénétration cutanée présentées dans les prochains chapitres (Chapitre IV et Chapitre V). Ainsi, nous proposons que les méthodes suivantes soient utilisées pour le prétraitement des données Raman : la correction du bruit gaussien à l'aide d'une projection des données par ACP, la réduction des dimensions spectrales par troncature, la correction de la ligne de base par la méthode Rubberband, la correction du bruit cosmique par une méthode semi-automatique réadaptée à des images à trois dimensions et enfin la normalisation des données par normalisation euclidienne.

Par ailleurs, le résultat final de la préparation des données étant fortement dépendant de l'ordre dans lequel ces différentes méthodes sont appliquées, nous avons également proposé un enchainement des différentes méthodes retenues pour la préparation des données. Cette proposition s'appuie sur un raisonnement logique basé sur les propriétés mathématiques des méthodes. Et en définitive, cette réflexion a abouti sur la mise en place d'une routine de prétraitement des données spectrales Raman utilisée en préalable des méthodes d'analyses multivariées présentées dans le chapitre suivant, ainsi que pour leur application dans les études de pénétration cutanée.

Chapitre IV.
ANALYSE QUALITATIVE ET SEMI-QUANTITATIVE DE LA COMPOSITION SPECTRALE DES DONNÉES RAMAN

Résumé

Ce chapitre est consacré à la présentation de deux méthodes d'analyse multivariée procédant à un démélange spectral des données sans nécessité d'avoir des connaissances *a priori* sur la composition du mélange : la MCR-ALS et la MVC-MNF. Les étapes de ces deux méthodes permettant d'obtenir un modèle caractérisant les données qualitativement et semi-quantitativement y sont décrites et de nombreux repères pratiques pour l'analyse des données spectrales y sont donnés.Trois études démontrant l'applicabilité de ces deux méthodes à des thématiques de recherche cosmétique sur des données réelles sont ensuite présentées. Elles portent sur des images Raman enregistrées à partir d'échantillons de peau humaine reconstruite dans un but de suivi des cinétiques de pénétration cutanée d'ICA. Ces études démontrent :

- La puissance et l'intérêt du caractère non supervisé de ces méthodes pour le suivi d'une molécule à travers la peau
- L'impact considérable que peuvent avoir les conclusions faites à partir des résultats obtenus par ces méthodes pour le secteur de la recherche et de l'industrie du cosmétique.

Nous terminons ce chapitre par une analyse des forces et faiblesses des méthodes présentées.

Introduction 101
1. Méthode de démélange 101
1.1. MCR-ALS 102
1.1.1. Résolution d'un problème d'optimisation 103
1.1.2. Estimations initiales 104
1.1.2.1. Le nombre de composants 104
1.1.2.2. Les matrices de spectres de composants purs ou de contribution relatives initiales 105
1.1.3. Contraintes 107
1.1.4. Évaluation du modèle estimé 107
1.1.4.1. Interprétabilité des spectres de composants purs 107
1.1.4.2. Analyse de la matrice résiduelle 108
1.1.4.3. Analyse des ambigüités potentielles 109
1.2. MVC-NMF 110
1.2.1. Calcul du volume du simplex 110
1.2.2. Résolution du problème d'optimisation 112
1.2.3. Initialisations, contraintes et évaluation de la qualité du modèle estimé 112
2. Application des méthodes MCR-ALS et MVC-NMF à des problématiques cosmétiques 113
2.1. Suivi du Delipidol® dans un modèle de peau reconstruite 113
2.1.1. Ensemble des données expérimentales 113
2.1.2. Prétraitement des données 114
2.1.3. Caractérisation spectrale du Delipidol® 115
2.1.4. Application de la MCR-ALS 115
2.1.4.1. MCR-ALS appliquée sans contrainte d'égalité spectrale sur le Delipidol® ($\boldsymbol{K = 4}$) 115

2.1.4.2. Comparaison des nombres de composantes 117
2.1.4.3. Comparaison MCR-ALS avec et sans contrainte d'égalité spectrale sur le Delipidol® 121
2.1.5. Conclusion 122
2.2. Suivi de la caféine dans un modèle de peau reconstruite 123
2.2.1. Ensemble des données expérimentales 124
2.2.2. Prétraitement des données 124
2.2.3. Caractérisation spectrale de la caféine 125
2.2.4. Analyse de données 126
2.2.4.1. Estimation du nombre de composants 126
2.2.4.2. Application de la MCR-ALS 127
(i) Sans contrainte d'égalité spectrale sur la caféine 127
Statistiques de modélisation 127
Interprétation spectrale 128
Erreurs de reproduction des données 130
Évaluation des modèles 130
(ii) Avec contrainte d'égalité spectrale sur la caféine 131
Statistique de modélisation 131
Interprétation spectrale 132
Erreurs de reproduction des données 134
Évaluation des modèles 135
(iii) Suivi de la caféine à l'aide des modèles avec $K = \{2, 3, 5\}$ et une contrainte d'égalité spectrale 135
2.2.4.3. Application de la MVC-NMF 137
(i) Sans contrainte d'égalité spectrale sur la caféine 137
Statistique de modélisation 139
Interprétation spectrale 139
Erreurs de reproduction des données 142
Évaluation des modèles 142
(ii) Avec contrainte d'égalité spectrale sur la caféine 143
Statistique de modélisation 143
Interprétation spectrale 145
Erreurs de reproduction des données 146
Évaluation des modèles 147
(iii) Suivi de la caféine à l'aide des modèles avec $K = 2, 3$ et une contrainte d'égalité spectrale 147
2.2.5. Conclusion 149
2.3. Étude des conditions de stockage des échantillons de peau reconstruite 150
2.3.1. Ensemble des données expérimentales 150
2.3.2. Prétraitements des données 151
2.3.3. Analyse des données 152
2.3.3.1. Estimation du nombre de composants 153
2.3.3.2. Application de la MCR-ALS avec une contrainte d'égalité spectrale sur le Delipidol® 154
Statistiques de modélisation 154
Interprétation spectrale des modèles estimés 155
Erreurs de reproduction des données 157
Cinétiques de pénétration du Delipidol® 158
2.3.3.3. Évaluation de la pertinence d'utilisation d'une contrainte d'égalité spectrale dans la MCR-ALS 159
Statistiques de modélisation 159
Interprétation spectrale des modèles estimés 160
2.3.4. Conclusion 162
Conclusion 162

INTRODUCTION

D'après l'état-de-l'art établi dans le Chapitre II, le couplage de la microscopie confocale Raman avec des méthodes d'analyse multivariée s'avère être un outil puissant pour étudier et mieux comprendre les changements moléculaires engendrés par la pénétration d'AC dans la peau. Cependant, les méthodes multivariées jusqu'alors couramment utilisées présentent certaines limitations dans leur utilisation. En particulier, elles requièrent des connaissances *a priori* sur la composition du mélange et ne tiennent pas compte des éventuelles évolutions des signatures spectrales des composants purs engendrées par certaines interactions moléculaires dans l'environnement cutané.

Pour outrepasser ces limitations, nous proposons l'utilisation de deux méthodes multivariées non-supervisées : la MCR-ALS (Multivariate Curve Resolution in Alterning Least Square) et la MVC-NMF (Minimum-Volume-Constrained Non-Negative Matrix Factorisation). Ces deux méthodes procèdent à un démélange spectral des données. Notre proposition s'appuie donc sur leur capacité à extraire les composants purs du mélange étudié et le sens chimique qu'ils expriment, sur l'estimation de leurs contributions relatives respectives mais aussi sur la possibilité d'utiliser ou non certaines informations préalables pour guider l'analyse. De plus, ces deux méthodes sont simples d'utilisations et sont des outils applicables sur des données réelles. Les industriels et chercheurs non experts du traitement de données peuvent donc les employer pour analyser leurs données spectrales et faire évoluer leur recherche.

Ce chapitre est divisé en deux parties. La première partie présente les détails théoriques et pratiques qui définissent les méthodes MCR-ALS et MVC-NMF. Elle constitue un guide pour toute personne souhaitant s'initier à l'utilisation de ces méthodes et fournit de nombreux repères pratiques pour l'analyse des données spectrales. Puis, la seconde partie est consacrée à l'application de ces méthodes sur des images hyperspectrales Raman réelles. Plus précisément, trois études sont menées pour démontrer leurs caractéristiques applicables et appliquées sur données réelles dans le cadre de thématiques de recherche cosmétique ciblées :

- Une première étude consiste à évaluer la capacité de la MCR-ALS à estimer la cinétique de pénétration d'un AC à travers les différentes couches de la peau sans aucune connaissance *a priori* sur la composition.
- Une deuxième étude se consacre à l'évaluation de la capacité des deux méthodes (MCR-ALS et MVC-NMF) à suivre la cinétique de pénétration d'une molécule exogène à travers la peau dont la signature spectrale Raman est difficilement détectable.
- Une dernière étude évalue l'influence des conditions de stockage des modèles de peaux sur la fonction barrière de la peau et le suivi d'un AC.

1. MÉTHODE DE DÉMÉLANGE

Les méthodes MCR-ALS et MVC-NMF sont très populaires dans le domaine de l'imagerie hyperspectrale. La MVC-NMF a par exemple été utilisée pour identifier et localiser des espèces minérales à partir d'images hyperspectrales de scènes réelles obtenues par un spectromètre visible/infrarouge aéroporté de la NASA [80], [81] mais aussi pour analyser la composition biochimique de plaques d'athéroscléroses coronaires humaines post-mortem à partir d'images multispectrales de fluorescence en temps réel [82]. Cependant, elle n'a pas encore été à ce jour appliquée sur des données spectroscopiques Raman. À l'inverse, la MCR-ALS a été appliquée aussi bien sur des données spectroscopiques Raman que sur des données spectroscopiques de natures différentes (par exemple en imagerie infrarouge ou MALDI, [53], [83]–[85]). En revanche, ses applications en spectroscopie Raman n'ont concerné que des études relevant de domaines

différents de celui de la recherche sur la peau. Elle a principalement été employée dans le domaine pharmaceutique pour identifier, quantifier et localiser les différents composants actifs et excipients contenus dans des produits pharmaceutiques [55], [62], [84]–[88]. Dans le domaine de la chimie, l'approche MCR-ALS a été utilisée comme outil de contrôle qualité de la composition de détergents [89]. Puis pour ce qui est du domaine de la biologie, la MCR-ALS a permis de caractériser la composition biochimique de tissus de natures différentes, à savoir des tissus végétaux [90] et des tissus animaux [91], [92]. Enfin, la MCR-ALS a été employée dans le domaine biomédicale à des fins de caractérisation de changements biochimiques dus à des pathologies dans des tissus humains [93]. Par conséquent, l'application de la MCR-ALS et de la MVC-NMF pour le suivi d'AC à travers la peau a donc semblé être pertinent et innovant pour le domaine de la recherche cosmétique. Nous commençons par présenter les principes théoriques de ces deux approches.

1.1. MCR-ALS

La méthode Multivariate Curve Resolution in Alterning Least Square (MCR-ALS) a été développée par Romà Tauler en 1993 [26], [53], [94], [95]. Comme la CLS, elle est basée sur l'hypothèse qu'un spectre est l'expression d'un mélange de composants biochimiques. Son objectif est d'extraire simultanément les composants purs du système chimique étudié et leurs contributions respectives dans le mélange à partir des données spectrales observées. Pour cela, la méthode considère que la matrice de données D, de taille (N, N_λ) où N et $N_\lambda \in \mathbb{N}^+$ sont respectivement le nombre de spectres dans l'ensemble de données et le nombre d'ondes, peut être exprimée comme une combinaison linéaire des spectres de composants purs pondérés par leur contribution associée dans le mélange. Cette décomposition spectrale peut être représentée élément par élément, composé par composé ou sous forme matricielle comme illustrée sur la Figure IV-1. Du point de vue matriciel, la matrice de données s'exprime mathématiquement comme le produit d'une matrice de spectres purs S, de taille (N_λ, K) où $K \in \mathbb{N}^+$ est le nombre de composants, multipliée par la matrice de leurs concentrations relatives respectives C, de taille $N \times K$:

$$D = CS^T + E \tag{IV-1}$$

où E est la matrice de taille (N, N_λ) contenant les résidus du modèle linéaire.

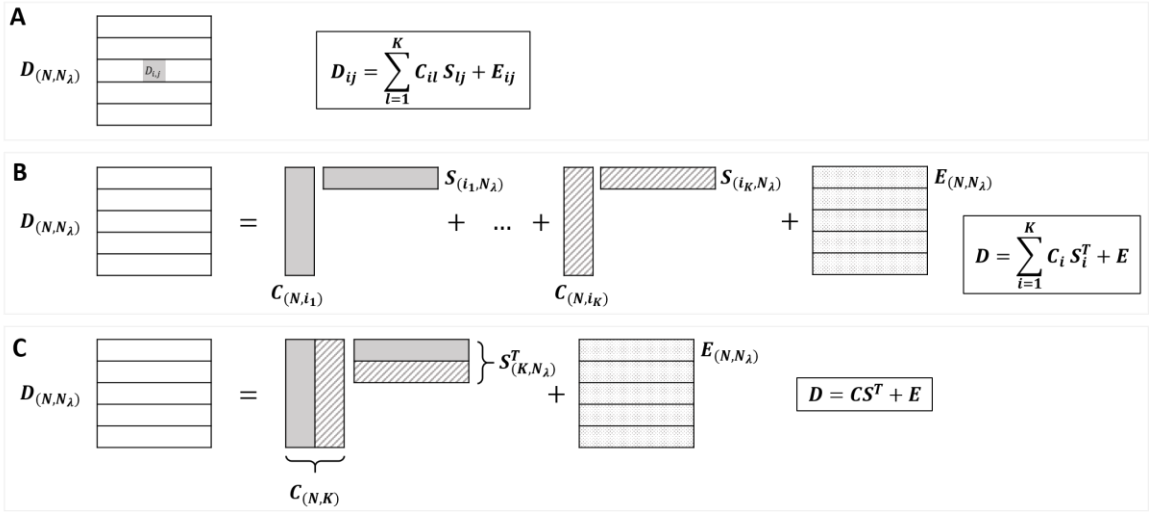


Figure IV-1 – Modèle bilinéaire représenté (A) élément par élément, (B) composé par composé et (C) sous forme matricielle

En imagerie hyperspectrale, un dépliage préalable du cube de données en une matrice 2D contenant les spectres de données sous forme de liste doit être réalisé pour pouvoir appliquer la MCR-ALS (Figure IV-2). Une fois le démélange spectral opéré, un repliement des concentrations relatives estimées est nécessaire pour retrouver la structure spatiale des concentrations associées aux composants purs estimés.

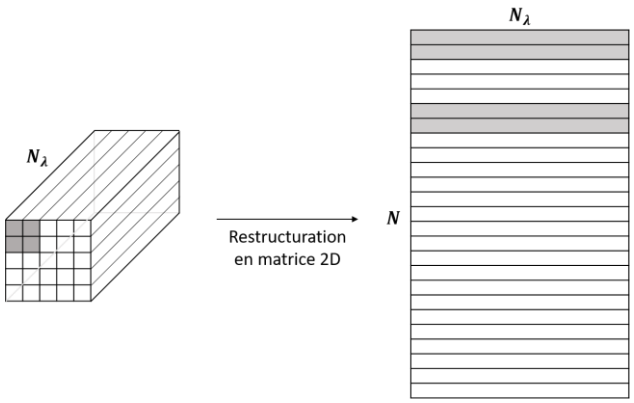


Figure IV-2 – Restructuration d'une image Raman pour une analyse MCR-ALS

1.1.1. Résolution d'un problème d'optimisation

La décomposition matricielle de la matrice de données résulte de la résolution d'un problème de minimisation des erreurs d'ajustement du modèle linéaire estimé par rapport aux données. L'expression du problème de minimisation à résoudre est la suivante :

$$\min\|E\| = \min\|D - CS^T\| \tag{IV-2}$$

Le principe de la MCR-ALS est de résoudre ce problème en minimisant alternativement les résidus par un ajustement du modèle linéaire au sens des moindres carrés, c'est-à-dire en fonction de la matrice des spectres de composants purs S puis en fonction de la matrice de concentrations relatives associées C. La norme utilisée pour la minimisation au sens des moindres carrés est donc par définition la norme euclidienne et l'expression du problème $(\text{IV-}2)$ se précise de la manière suivante :

$$\min\|E\|_2 = \min\|D - CS^T\|_2^2 \tag{IV-3}$$

D'un point de vue algorithmique, la régression du modèle linéaire alternée au sens des moindres carrés s'implémente ainsi :

$$C = DS(S^TS)^{-1} \tag{IV-4}$$

$$S^T = (C^TC)^{-1}C^TD \tag{IV-5}$$

L'algorithme s'arrête lorsqu'un nombre d'itérations maximum est atteint ou quand l'ajustement du modèle est jugé mathématiquement comme satisfaisant. Dans ce second cas, la similarité du modèle par rapport aux données est mesurée selon deux critères :

- le manque d'ajustement (LOF pour « Lack Of Fit ») qui exprime la fidélité du modèle aux données. Il s'exprime par :

$$\text{LOF} = 100\sqrt{\frac{\sum_{ij} E_{ij}^2}{\sum_{ij} D_{ij}^2}} \tag{IV-6}$$

  où E_{ij} et D_{ij} sont respectivement un élément des matrices de données E et D.

- le pourcentage de variance expliquée qui indique si le modèle décrit bien les données. Il s'exprime par :

$$\mathrm{R}^2 = 100\sqrt{\frac{\sum_{ij} D_{ij}^2 - \sum_{ij} E_{ij}^2}{\sum_{ij} D_{ij}^2}} \qquad \text{(IV-7)}$$

Aucune priorité n'est établie entre ces deux critères mathématiques, chacun d'eux est pris en compte dans l'évaluation de l'ajustement du modèle. Tant que ces deux critères augmentent d'une itération à une autre, l'algorithme de la MCR-ALS continue d'estimer les spectres de composants purs et les concentrations associées les plus proches des données observées. La MCR-ALS converge lorsque la différence de reproduction des données entre deux itérations est inférieure à un seuil de tolérance prédéfini. La différence ρ entre deux modèles estimés à une itération d'intervalle se calcule de la façon suivante :

$$\frac{\rho_{iter} - \rho_{iter+1}}{\rho_{iter+1}} \quad \text{avec} \quad \rho_{iter} = \sqrt{\frac{\sum_{ij}\left(E_{ij}^{iter}\right)^2}{N}} \qquad \text{(IV-8)}$$

Il n'existe pas de seuil de tolérance optimal, il doit être défini selon le bruit expérimental.

1.1.2. Estimations initiales

Les résultats fournis par la MCR-ALS dépendent des initialisations requises pour le lancement de l'algorithme : le nombre de composants présents dans le système biochimique étudié et l'initialisation d'une des deux matrices C et S, le plus souvent celle des spectres des composants.

1.1.2.1. Le nombre de composants

Le nombre de composants peut être connu à l'avance. Dans le cas contraire, son estimation à partir des données est nécessaire. Dans ce travail, nous avons considéré et comparé trois approches permettant de l'estimer:

- La Singular Value Decomposition (SVD) est la plus utilisée dans la littérature. Elle est basée sur le calcul de la matrice de covariance des données et permet d'identifier les dimensions nécessaires pour exprimer le maximum de variance expliquée qui reflète la majorité des informations pertinentes. Les plus grandes valeurs singulières exprimant la majorité des informations pertinentes sans redondance correspondent au nombre de composants nécessaires à la modélisation des données. Cependant, aucun seuil universel pour la variance expliquée minimale assurant un ajustement fiable des données ne peut être déterminé, en raison de la forte dépendance au type d'étude et de données analysées. Le SVD reste donc un critère très subjectif pour estimer le nombre de composants.

- Le critère de Durbin-Watson (DW) [96] est une statistique évaluant l'autocorrélation des résidus d'un modèle de régression linéaire. Plus précisément, le test de Durbin-Watson cherche à vérifier la significativité de la statistique DW, notée ϕ, dans l'expression suivante :

$$E_i = \phi E_{i-1} + \epsilon_i \qquad \text{(IV-9)}$$

où ϵ_i est un bruit blanc, E_i et E_{i-1} sont les résidus des spectres successifs dans le modèle estimé et ϕ est définie par :

$$\phi = \mathrm{DW} = \frac{\sum_{i=2}^{N}(E_i - E_{i-1})^2}{\sum_{i=1}^{N} E_i^2} \qquad \text{(IV-10)}$$

Le but est de vérifier si $\phi = 0$, c'est-à-dire « H_0 : aucune autocorrélation résiduelle n'existe » ou si $\phi \neq 0$ avec $|\phi| < 1$, autrement dit « H_1 : une autocorrélation résiduelle positive ou négative existe ». Le critère DW est borné par 0 et 4 : une valeur proche de 2 signifie que les

spectres résiduels sont non autocorrélés, c'est-à-dire qu'ils ne sont composés que par le bruit, et une valeur proche de 0 ou 4 signifie que les spectres résiduels sont respectivement corrélés positivement et négativement, autrement dit qu'il reste un ou plusieurs composants à extraire des données. Théoriquement, le nombre de composants à choisir devrait décrire un modèle avec un DW proche de 2. Cependant, en pratique K est estimé comme le nombre de composants à partir duquel l'autocorrélation résiduelle basée sur le modèle défini à l'aide des résultats SIMPLISMA diminue brutalement, c'est-à-dire lorsque le DW augmente significativement.

- L'indice de Kaiser-Meyer-Olkin (KMO) [97], [98] mesure la qualité de la modélisation des données sur la base des résidus. La corrélation entre les résidus de modélisation doit être aussi faible que possible. Cependant, deux spectres résiduels peuvent être corrélés si chacun a un lien avec un troisième spectre. Si le lien entre les deux premiers spectres n'est pas direct, leur corrélation partielle est nulle. Par conséquent, le KMO compare la corrélation brute à la corrélation partielle. Une corrélation partielle proche d'une corrélation brute confirme l'existence réelle d'un lien linéaire entre les spectres résiduels. Son expression est la suivante :

$$\mathrm{KMO} = \frac{\sum_i \sum_{i \neq j} W_{ij}^2}{\sum_i \sum_{i \neq j} W_{ij}^2 + \sum_i \sum_{i \neq j} p_{ij}^2} \qquad \text{(IV-11)}$$

où $W = [W_{i,j}] = \left[v_{i,j}\right]^{-1}$ est l'inverse de la matrice de corrélation correspondant en réalité dans l'algorithme à la matrice pseudo-inverse de Moore-Penrose (fonction *pinv()* dans Matlab) de la matrice de corrélation et $p_{ij} = -\frac{v_{ij}}{\sqrt{v_{ii} \times v_{jj}}}$ est le coefficient de la corrélation partielle. Le KMO varie entre 0 et 1. Deux spectres sont dits complètement décorrélés si leur KMO est proche de 1. L'interprétation de cette mesure reste très subjective. En pratique, un indice supérieur à 0.9 est considéré comme le meilleur résultat, un indice entre 0.8 et 0.9 est considéré comme très bon, un indice entre 0.7 et 0.8 est jugé comme bon, un indice entre 0.5 et 0.7 est estimé médiocre et un indice inférieur à 0.5 est jugé inacceptable [2]. Le nombre de composantes correspond au nombre de spectres résiduels ayant un KMO proche de 1, ce qui est généralement interprété comme supérieur à 0.7. En pratique, la baisse brutale du KMO estimé sur la base des résultats SIMPLISMA est prise en compte pour donner une idée du nombre de composant souhaité.

Les critères DW et KMO ont l'avantage d'être bornés, ce qui facilite leur interprétation et permet de conclure plus facilement. Néanmoins, malgré ces critères, l'évaluation du nombre de composants au sein d'un mélange reste difficile et dépend du modèle estimé. Qui plus est, dans un système biologique tel que celui de la peau, le nombre de composant est souvent inconnu et difficile à estimer. Il peut être alors tentant d'utiliser un nombre de composants élevé dans l'optique de sélectionner ceux qui semblent les plus pertinents. Cependant, le risque d'extraire de mauvais profils spectraux est augmenté. Par conséquent, il est courant d'analyser les résultats obtenus sur une fenêtre de nombre de composants pour définir la décomposition la plus pertinente. Dans les études présentées dans la partie suivante, ces trois méthodes seront utilisées pour guider la sélection d'une fenêtre de nombre de composants à tester pour la modélisation des données.

1.1.2.2. Les matrices de spectres de composants purs ou de contribution relatives initiales

Dans la MCR-ALS, l'identité chimique des composants purs du matériau biologique étudié et leurs concentrations dans les spectres ne sont pas requises. Néanmoins pour lancer le processus

itératif, une première estimation d'une des deux matrices à estimer est nécessaire. Le point de départ le plus souvent utilisé est la matrice de spectres de composants purs S car l'initialisation de la matrice C contenant les concentrations relatives de chaque composant est plus problématique. En effet, une initialisation des cartes de distributions relatives ou des spectres de composants purs trop éloignée de la solution ou aléatoire n'est pas judicieuse du fait qu'elle risque de conduire à des minima locaux irréalistes et aberrants. Une initialisation proche des données est donc préférable. Cependant, il est particulièrement compliqué de définir des cartes de distribution initiales pas trop éloignées de la solution pour favoriser la convergence de l'algorithme sans avoir aucune information *a priori* ni sur la distribution entre composants ni sur la répartition spatiale de chacun d'eux. Cette initialisation proche des données n'est alors possible qu'en utilisant les spectres de données disponibles. Différentes options d'initialisation de la matrice des spectres S sont disponibles.

L'estimation initiale par la méthode SIMPLISMA (SIMPLe-to-use-Interactive-Self-Modeling-Mixture-Analysis) est la plus couramment utilisée dans la littérature [96], [99]. Elle consiste à sélectionner les spectres considérés comme les plus « purs » parmi les données. La pureté d'un spectre est définie comme un coefficient de variation modifié dont l'expression est la suivante :

$$P_i = \frac{\sigma_i}{\overline{D_i} + \beta} \tag{IV-12}$$

où P_i est le coefficient de pureté pour un spectre donnée D_i de moyenne $\overline{D_i}$, σ_i est l'écart-type du spectre D_i et β est un coefficient représentant entre 1% et 5% du maximum des moyennes des intensités spectrales afin d'éviter une valeur aberrante dans le cas où $\overline{D_i} \approx 0$. Plus le coefficient de pureté P_i est élevé, plus la différence à la moyenne spectrale est significative et par conséquent, plus le spectre en question est considéré comme pur. Le premier spectre estimé comme pur correspond au spectre ayant le plus grand coefficient de pureté. Mais les $K-1$ spectres purs suivants ne se déterminent pas uniquement en sélectionnant les profils ayant les $K-1$ plus grands coefficients de pureté. Pour obtenir les spectres les plus dissemblables possibles, une propriété d'orthogonalité doit être respectée. On considère deux spectres comme orthogonaux s'ils sont non corrélés entre eux, c'est-à-dire si leur coefficient de corrélation est proche de 0. Avec une suite logique, le j-ème spectre pur doit être non corrélé avec les $1, \dots, j-1$-ème spectres purs déjà définis. Pour prendre en compte cette propriété d'orthogonalité entre les spectres, les coefficients de pureté sont pondérés selon la corrélation entre les potentiels spectres purs et les $j-1$ premiers spectres purs déjà retenus de la façon suivante :

- Si la corrélation est faible, un poids proche de 1 est attribué au coefficient de pureté,
- Si la corrélation est forte, un poids proche de 0 est attribué au coefficient de pureté,

Ainsi, le profil ayant la plus haute valeur calculée de cette façon correspond au prochain spectre pur. La pondération est réalisée à chaque nouvelle recherche de spectre considéré comme l'un des plus purs. Les poids ω attribués aux coefficients de pureté se déterminent par :

$$\omega_{ij} = \det\begin{pmatrix} W_{i,i} & W_{i,p_1} & \cdots & W_{i,p_j} \\ W_{p_1,i} & W_{p_1,p_1} & \cdots & W_{p_1,p_j} \\ \vdots & \vdots & \ddots & \vdots \\ W_{p_j,i} & W_{p_j,p_1} & \cdots & W_{p_j,p_j} \end{pmatrix} \tag{IV-13}$$

où ω_{ij} est le poids attribué au i-ème profil pour le calcul du j-ème spectre pur et W est la matrice de corrélation entre les $j-1$-ème premiers spectres purs et le potentiel j-ème spectre. Dans le cas où tous les profils sont parfaitement non corrélés entre eux, le déterminant calculé est celui

de la matrice identité qui vaut 1. Dans le cas contraire, si les profils purs sont corrélés entre eux, le déterminant tend vers 0.

D'autre part, il est courant d'employer le résultat d'une classification par K-means ou des références spectrales connues comme initialisation spectrale de la MCR-ALS. Cependant, dans les applications présentées ci-après, nous proposons d'utiliser uniquement la méthode SIMPLISMA pour initialiser la matrice de spectres initiales. En effet, bien que la méthode K-means fournit des spectres représentant la moyenne de différents clusters identifiés dans les données, ces spectres moyens ne sont pas toujours assez décorrélés entre eux pour représenter un point de départ menant plus facilement à une convergence de l'algorithme. Quant à l'initialisation spectrale à l'aide de références spectrales de composants connus, la difficulté réside à sélectionner parmi tous les composants de la peau un nombre réduit de composant à estimer suffisant pour modéliser le maximum les données de manière simple. De plus, l'accès aux références spectrales de ces composants n'est pas toujours facile.

1.1.3. Contraintes

Pour transformer le processus d'optimisation de la MCR-ALS en outil d'analyse de données centré sur le système chimique et assurer la pertinence de ses extractions, des contraintes sur les matrices C et S peuvent être intégrées au problème d'optimisation. Leur application est optionnelle mais doit être justifiée par des propriétés naturelles. Les contraintes les plus couramment utilisées sont :

- La non-négativité des spectres et des concentrations : $C \geq 0, S \geq 0$. Les propriétés de la spectroscopie Raman et le sens chimique des contributions des composés purs justifient son utilisation. Le problème à résoudre devient alors :
$$\min_{S\geq 0, C\geq 0} \|D - CS^T\|_2^2 \qquad \text{(IV-14)}$$
- La normalisation des spectres et des concentrations : $\|S_i\|_2 = 1,\ \sum_{i=1}^{K} C_i = \mathbf{1}$. Cette propriété assure que seules des contributions relatives seront extraites.
- L'égalité spectrale : lorsqu'une connaissance *a priori* sur la signature spectrale d'un des composants purs est disponible, il est possible de l'intégrer au modèle. Dans ce cas, à chaque itération de l'algorithme, l'égalité spectrale entre un des composants purs et la référence connue est vérifiée. Cette contrainte sera notamment employée à des fins d'évaluation de la qualité des modèles estimés dans les applications présentées ultérieurement lorsque l'AC sera connu.

1.1.4. Évaluation du modèle estimé

Les résultats obtenus par MCR-ALS sont issus d'une convergence mathématique basée sur des critères évaluant leur fidélité aux données. Cependant, des évaluations supplémentaires peuvent être nécessaires pour évaluer la pertinence de la décomposition. La signification biochimique et spectroscopique des matrices résultantes doit être vérifiée et les informations résiduelles et les ambiguïtés potentielles doivent être examinées.

1.1.4.1. Interprétabilité des spectres de composants purs

Les contraintes appliquées aux spectres de composants purs doivent être respectées. En plus d'un sens chimique, il est essentiel que les spectres estimés soient cohérents avec les caractéristiques spectrales présentes dans les données spectrales expérimentales. Il est aussi fréquent de recourir à la correspondance spectrale pour identifier la nature des composants purs estimés. Pour cela, chacun d'eux est comparé à des références spectrales de composants purs enregistrées dans les

mêmes conditions que les données spectrales. Cette comparaison peut être réalisée selon des critères différents. Les plus courants sont :

- La distance euclidienne mesure la différence entre chaque variable spectrale des spectres estimés et de référence. Néanmoins, elle ne rend pas compte de la position des bandes ce qui peut engendrer des correspondances faussement positives. Par conséquent, elle ne sera pas utilisée dans les études comparatives qui suivent.

- La corrélation mesure la force de la relation linéaire qui lie le spectre d'un composant pur estimé et le spectre de référence. Elle s'exprime par :

$$\mathbf{r} = \frac{\overline{S_i * S_{REF}} - \overline{S_i}\,\overline{S_{REF}}}{\sigma_{S_i}\sigma_{S_{ref}}} \qquad \text{(IV-15)}$$

où $\overline{S_i * S_{REF}}$, $\overline{S_i}$, $\overline{S_{REF}}$ sont respectivement les valeurs moyennes des spectres $S_i * S_{REF}$, S_i et S_{REF} avec '$*$' l'opérateur de multiplication éléments par éléments et σ_{S_i} et $\sigma_{S_{REF}}$ sont les écarts-types des spectres S_i et S_{REF}. Un coefficient de corrélation proche de 1 ou -1 traduit une corrélation positive ou négative entre les deux spectres. Toutefois, un coefficient de corrélation nul n'implique pas l'indépendance des spectres, il indique seulement une indépendance linéaire. De plus, le coefficient de corrélation au carré, dit coefficient de détermination, correspond au pourcentage de recouvrement des bandes spectrales entre les deux spectres comparés. En général, un coefficient de corrélation supérieur à 0,9 en valeur absolue est statistiquement considéré comme une corrélation très élevée, et donc un tel coefficient de corrélation est considéré comme satisfaisant, puisque plus de 90 % des variations des valeurs spectrales seront interdépendantes [100]. Cette mesure considérée comme fiable sera donc employée pour analyser les correspondances spectrales entre les spectres de composants purs et les données de référence dans les études ultérieures.

- L'indice de qualité HQI (Hit Quality Index) est une mesure statistique fournissant la corrélation spectrale entre une référence et un échantillon inconnu. Elle s'exprime par :

$$0 < \mathrm{HQI} = \frac{(S_i\, S_{REF}^T)^2}{(S_i)^2\,(S_{REF}^T)^2} < 1 \qquad \text{(IV-16)}$$

où S_{REF} est le spectre de référence comparé. Un indice proche de 1 témoigne d'une forte corrélation entre les spectres comparés. Estimé comme fiable et reconnu dans le domaine spectroscopique, cet indice sera également utilisé pour évaluer les modèles estimés dans les applications suivantes.

1.1.4.2. Analyse de la matrice résiduelle

L'analyse de la matrice résiduelle permet d'évaluer qualitativement l'ajustement du modèle aux données. La visualisation des spectres résiduels permet de vérifier que toutes les informations pertinentes ont été extraites des données et qu'aucune bande caractéristique importante n'a été délaissée. Par exemple, nous avons vérifié que les bandes positives et négatives du spectre résiduel moyen illustré dans la Figure IV-3 ne correspondent à aucune bande caractéristique de la peau ou de l'ingrédient actif délaissée dans la modélisation des données (Figure IV-3). Une autre façon de vérifier qu'aucune information n'a été oubliée dans le modèle estimé consiste à observer la répartition des résidus sous la forme d'un histogramme (Figure IV-3). Dans cet histogramme, toutes les valeurs résiduelles du modèle sont organisées en une série d'intervalles et le nombre de variables spectrales ayant une intensité incluse dans chacun d'eux leur est associé. La forme de l'histogramme fournit des informations sur la distribution des valeurs et en particulier, une forme gaussienne signifie que le bruit résiduel est gaussien.

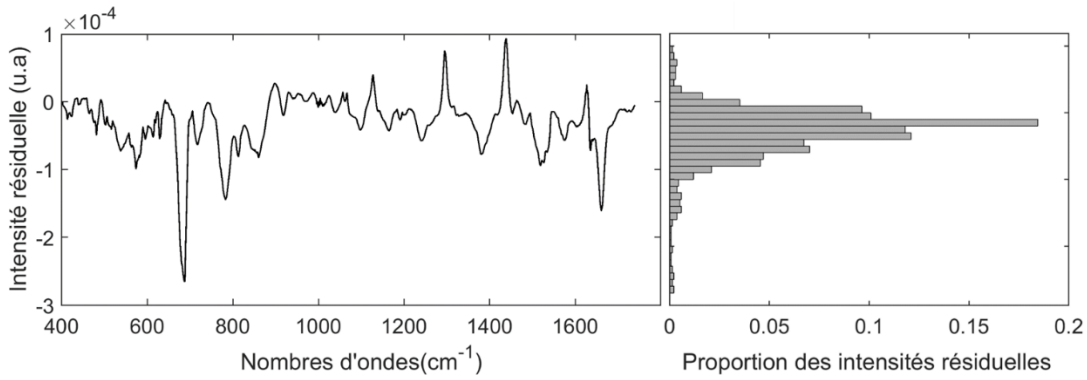


Figure IV-3 – Spectre résiduel moyen et son histogramme

1.1.4.3. Analyse des ambigüités potentielles

Enfin, l'unicité de la solution obtenue par la MCR-ALS est vérifiée en évaluant l'ambiguïté présente dans les données. Cette ambiguïté peut être exprimée sous deux formes :

- L'ambigüité d'intensité traduit la possibilité d'insérer une constante m dans la décomposition qui ne changera pas sa qualité :

$$D = \sum_{i=1}^{K} (m_i\, C_i)\left(S_i^T \frac{1}{m_i}\right) + E \qquad \text{(IV-17)}$$

  L'ajout de contraintes telles que la non-négativité et la normalisation des spectres et des concentrations permet d'obtenir des informations semi-quantitatives et donc de s'affranchir de l'ambigüité d'intensité.

- L'ambiguïté rotationnelle exprime la possibilité d'insérer une matrice de rotation inversible Q qui ne changera pas la qualité de la décomposition mais transformera potentiellement les matrices C et S :

$$D = (CQ)(Q^{-1}S^T) + E = \tilde{C}\tilde{S}^T + E \qquad \text{(IV-18)}$$

  où $\tilde{C}$ et $\tilde{S}$ correspondent aux matrices transformées de C et S. L'ajout de contraintes telles que la non-négativité et la normalisation des spectres et des concentrations ainsi qu'une égalité spectrale permettent de réduire l'ambiguïté rotationnelle tout en guidant la décomposition vers une solution unique. De même, l'utilisation d'une approche dite multi-set consistant à analyser différents ensembles de données de conditions différentes de manière simultanée permet de prendre en considération les cinétiques des composants purs au sein des données et donc d'augmenter la cohérence des extractions ce qui en conséquence réduit l'apparition de potentielles ambiguïtés.

Le degré de l'ambiguïté rotationnelle dans le modèle estimé par MCR-ALS est estimé à l'aide d'un algorithme, nommé MCR-Bands. Il consiste à calculer la contribution relative du spectre de chaque composant pur f_i exprimée par :

$$f_i(Q) = \frac{\left\|C_i S_i^T\right\|_2}{\|CS^T\|_2} \qquad \text{(IV-19)}$$

Alors que le terme $\|CS^T\|_2$ est constant pour toute matrice de rotation Q, le terme $\left\|C_i S_i^T\right\|_2$ donnera des valeurs différentes selon la matrice de rotation Q. L'algorithme a alors pour but d'estimer les matrices de rotation Q maximisant et minimisant la contribution f_i (IV-19) parmi les solutions réalisables [101], [102]. Dans le cas où il n'y a pas d'ambiguïté de rotation, les limites de bande maximale et minimale estimées sont proches de la solution et leur différence est proche

de 0, traduisant ainsi que le modèle estimé est fiable. Les algorithmes MCR-ALS et MCR-Bands ont été appliqués en utilisant les scripts du logiciel MCR-ALS GUI 2.0 dans Matlab [95].

1.2. MVC-NMF

Tout comme la MCR-ALS, la Minimum-Volume-Constrained Non-Negative Matrix Factorisation (MVC-NMF) est une méthode visant à décomposer une matrice de données en un produit de deux matrices, l'une composée de spectres de composants purs S et l'autre contenant leurs concentrations relatives respectives C [80]. Cette factorisation matricielle est réalisée en résolvant un problème d'optimisation qui consiste à minimiser l'erreur d'ajustement d'un modèle linéaire défini par les matrices C et S sous la contrainte qu'elles soient toutes deux non-négatives (Équation (IV-14)). Cette propriété de non-négativité se justifie par les propriétés physicochimiques que doivent respecter à la fois les spectres Raman et les contributions chimiques attendues.

À l'instar de la MCR-ALS, la MVC-NMF a pour objectif d'estimer les spectres des composants purs en écartant l'hypothèse que ces composants soient présents à l'état pur au sein même du mélange. D'un point de vue géométrique où les données définissent un nuage de points, cette estimation se résume à définir les K points représentant les spectres de composants purs de façon à ce que le simplex[1] qu'ils composent englobent tous les points de données [80], [103]. Ainsi, chaque spectre s'exprime comme un barycentre des spectres de composants purs, autrement dit chaque spectre est considéré comme une combinaison linéaire de ces derniers. La Figure IV-4 illustre cette représentation des données.

Cependant, il existe une infinité de possibilité de déterminer K spectres formant une enveloppe convexe circonscrite aux données. Le simplex le plus simple à définir est celui présentant un volume minimal. Contrairement à la MCR-ALS qui n'impose pas de contraintes sur la position des spectres de composants purs dans une représentation géométrique, la MVC-NMF intègre cette contrainte de volume minimal du simplex formé par les K spectres de composants purs dans la résolution du problème d'optimisation. Autrement dit, la MVC-NMF considère qu'un compromis entre une erreur minimale du modèle bilinéaire estimé par rapport aux données et le volume minimal de l'enveloppe convexe formée par les spectres de composants purs estimés peut être pris en compte. Au final, la MVC-NMF consiste à résoudre le problème d'optimisation suivant :

$$\min_{S\geq 0, C\geq 0} \|D - CS^T\|_2^2 + 2\,\eta\,\boldsymbol{\mathcal{V}}(S^T) \qquad \text{(IV-20)}$$

où $\boldsymbol{\mathcal{V}}(S^T) \in \mathbb{R}$ est le volume du simplex formé par les spectres contenus dans la matrice S et $\eta \in \mathbb{R}$ permettant de contrôler le compromis.

1.2.1. Calcul du volume du simplex

D'après J. R. Munkres [103], le volume du simplex $\boldsymbol{\mathcal{V}}(S^T)$ se détermine en calculant le déterminant de la matrice des éléments S^T qui le compose. Par soucis de dimensionnalité, la matrice S^T doit être réduite à une matrice carrée $\hat{S}$ pour pouvoir calculer son déterminant. Ainsi, une ACP est appliquée sur la matrice S^T centrée sur les données pour réduire sa dimension spectrale à $K-1$ variables puisque le volume du simplex formé par les composants purs s'exprime de la façon suivante :

[1] Simplex : généralisation du triangle à une dimension quelconque, enveloppe convexe d'un ensemble de N points utilisée pour former un repère affine dans un repère affine de dimension $N-1$.

$$\boldsymbol{\nu}(S^T) = \frac{1}{2\,(K-1)!}\det\left(\begin{bmatrix}\mathbf{1^T}\\ \hat{S}\end{bmatrix}\right) \qquad \text{(IV-21)}$$

où $\hat{S} = U_D^T(S^T - \mathbf{1}\,\mu_D)$ avec U_D composée des $(K-1)$ loadings les plus significatifs définis à partir de la matrice de données D, μ_D le spectre moyen des données et $\mathbf{1}$ un vecteur unitaire. Le vecteur unitaire $\mathbf{1}$ dans le calcul du déterminant assure que les éléments sont bien linéairement indépendants. L'expression de la matrice $M = \begin{bmatrix}\mathbf{1^T}\\ \hat{S}\end{bmatrix}$ apparaissant dans le calcul du déterminant peut être reformulée en fonction de la matrice réduite $\hat{S}$ comme suit:

$$M = \begin{bmatrix}\mathbf{1^T}\\ \hat{S}\end{bmatrix} = \begin{bmatrix}\mathbf{1}^T\\ \tilde{0}\end{bmatrix} + \begin{bmatrix}\mathbf{0}^T\\ \tilde{I}\end{bmatrix}\hat{S} = M_1 + M_2 U_D^T(S^T - 1\mu_D) \qquad \text{(IV-22)}$$

où $\tilde{0}$ est une matrice nulle de dimension $(K-1, K)$, $\tilde{I}$ la matrice identité de dimension $(K-1, K-1)$ et $\mathbf{0}$ un vecteur nul.

Finalement, le problème d'optimisation (IV-20) se réécrit de la manière suivante :

$$\min_{S\geq 0, C\geq 0} \|D - CS^T\|_2^2 + \tau\,\det(M_1 + M_2\,U_D^T(S^T - 1\mu_D)) \qquad \text{(IV-23)}$$

avec $\tau = \frac{\eta}{(K-1)!} \in \mathbb{R}$ et les matrices M_1, M_2, U_D et le vecteur μ_D des constantes définies à partir de la matrice de données D. Le problème d'optimisation exprime donc qu'un compromis entre une reproduction fidèle des données originales et une proximité maximale de ces spectres par rapport aux données doit être trouvé. Ce compromis est ajusté à l'aide d'un paramètre supplémentaire τ : un faible τ donne l'avantage au terme d'attache aux données dans la formulation du problème d'optimisation et donc privilégie un ajustement fidèle aux données tandis qu'un τ élevé priorise la proximité des spectres estimés aux données en donnant l'avantage au terme associé au volume qu'ils forment. L'intégration de ce paramètre supplémentaire est un élément à prendre en compte dans l'analyse de démélange spectrale réalisée par MVC-NMF puisque l'insertion de cette contrainte de volume minimal supplémentaire peut impacter la décomposition spectrale.

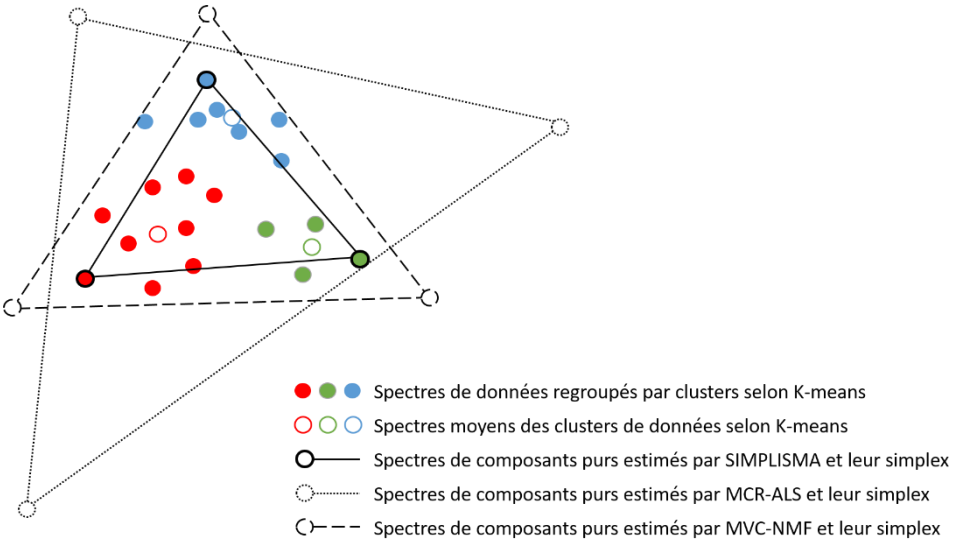


Figure IV-4 – Représentation géométrique des données et des simplex formés par les composants purs estimés par les méthodes SIMPLISMA, MCR-ALS et MVC-NMF

1.2.2. Résolution du problème d'optimisation

Les matrices contenant les spectres de composants purs et leurs contributions relatives associées étant toutes deux inconnues, la résolution du problème d'optimisation se réalise en résolvant deux sous-problèmes de manière alternée :

$$\begin{cases} \underset{C\geq 0}{min}\|D - CS^T\|_2^2 + \tau \det(M_1 + M_2\, U_D^T(S^T - 1\mu_D)) \\ \underset{S\geq 0}{min}\|D - CS^T\|_2^2 + \tau \det(M_1 + M_2\, U_D^T(S^T - 1\mu_D)) \end{cases} \quad \text{(IV-24)}$$

Pour respecter la contrainte de non-négativité des matrices, la solution de chaque sous-problème est calculée en utilisant la méthode de descente du gradient projeté, c'est-à-dire :

$$\begin{cases} S = \max(0, S - \alpha_1 \nabla_S J(S, C) \\ C = \max\ (0, C - \alpha_2 \nabla_C J(S, C) \end{cases} \quad \text{(IV-25)}$$

où $J(S, C) = \|D - CS^T\|_2^2 + \tau \det(M_1 + M_2\, U_D^T(S^T - 1\mu_D))$ et $\alpha_1, \alpha_2 \in \mathbb{R}$.

L'algorithme s'arrête lorsqu'un nombre d'itérations maximum est atteint ou quand plus aucun changement n'est observé entre deux itérations. De la même manière que la MCR-ALS, les critères mesurant l'ajustement mathématique du modèle aux données tels que le LOF et le R^2 sont utilisés à des fins d'évaluation d'une similarité satisfaisante. L'algorithme de la MVC-NMF a été appliqué en utilisant le script de la toolbox Matlab créée par M. Arngren, M. Schmidt et J. Larsen [104].

1.2.3. Initialisations, contraintes et évaluation de la qualité du modèle estimé

Comme la plupart des algorithmes de factorisation matricielle, la MVC-NMF est sensible aux initialisations qu'elle requiert en raison de l'existence de plusieurs minima locaux que présente la fonction à optimiser. En raison de l'utilisation de la méthode du gradient projeté dans l'algorithme, les matrices initiales des spectres de composants purs et de leurs concentrations relatives respectives dans l'échantillon biochimique étudié doivent être aussi proches que possible des données. De même, le nombre de composants utilisé dans la décomposition influe sur le résultat fourni par la MVC-NMF. Au final, en vue d'une étude comparative sur les résultats fournis par la MCR-ALS et la MVC-NMF, les méthodes utilisées pour choisir le nombre de composants et les spectres initiaux dans la MVC-NMF sont les mêmes que celles employées dans la MCR-ALS.

En ce qui concerne les contraintes que doivent vérifier les spectres des composants purs et les concentrations relatives, la non-négativité et la normalisation des spectres des composants purs et des concentrations associées est directement intégrée dans la minimisation des sous-problèmes d'optimisation via l'utilisation de la méthode du gradient projeté. Quant à l'incorporation d'une contrainte d'égalité spectrale, elle est intégrée à la MVC-NMF de la même manière qu'elle l'est dans la MCR-ALS grâce à une modification home-made du script Matlab utilisé.

Enfin, les résultats obtenus par la MVC-NMF peuvent être analysés de la même façon que ceux issus de la MCR-ALS. L'interprétabilité spectrale, les résidus du modèle estimé ainsi que les ambiguïtés rotationnelles sont évaluées pour juger de la qualité du modèle estimé.

2. APPLICATION DES MÉTHODES MCR-ALS ET MVC-NMF À DES PROBLÉMATIQUES COSMÉTIQUES

2.1. SUIVI DU DELIPIDOL® DANS UN MODÈLE DE PEAU RECONSTRUITE

L'étude présentée ci-après a fait l'objet d'une publication dans le Journal of Biophotonics [105]. Étant bien admis que la fonction barrière de la peau peut fortement compromettre l'efficacité des principes actifs cosmétiques, beaucoup d'entre eux étant incorporés dans des produits commercialisés doivent passer outre cette barrière et doivent atteindre leurs cellules cibles situées dans les couches plus profondes de la peau. De ce fait, l'évaluation des produits cosmétiques pour valider la diffusion des AC dans la peau est indispensable. Une approche économique, respectueuse de l'environnement et simple à mettre en œuvre pour définir les cinétiques de pénétration des AC à travers la peau a été proposée par L. Miloudi *et al.* [44]. Elle consiste à coupler l'imagerie confocale Raman avec des méthodes d'analyse multivariée telles que la CLS. Pour suivre la molécule du Delipidol® à travers la peau, la CLS a été employée en utilisant comme références spectrales un grand nombre de spectres Raman mesurés sur des peaux non traitées ainsi que la signature spectrale du Delipidol®, puis une contrainte de non-négativité sur les concentrations à estimer a été imposée. Ainsi, des cartes de distributions spatiales et des profils de pénétration du Delipidol® ont pu être établis. Cependant, comme cela a été mentionné dans le Chapitre II, l'utilisation de la NCLS présente quelques inconvénients. En effet, dans ce cas d'étude, la faisabilité de collecter des spectres de référence fiables de tous les composants purs présents dans la peau est réfutée et la performance de la NCLS est limitée par la disponibilité d'un grand nombre de spectres Raman acquis à partir d'échantillons de peau témoin.

L'étude présentée ci-après propose d'établir un protocole d'analyse non supervisé complet en revisitant et en optimisant l'analyse de données réalisée dans l'article de L. Miloudi *et al.*. Pour surpasser les limitations de la méthode NCLS, la MCR-ALS a été appliquée sur le même jeu de données afin d'évaluer la diffusion du Delipidol® à travers la peau. Le choix d'utiliser la MCR-ALS se justifie par sa capacité à estimer les composants purs présents dans les échantillons de peau sans connaissance *a priori* sur la composition et à prendre en compte les éventuels décalages de bandes survenus suites aux nombreuses interactions moléculaires induites. De plus, la MCR-ALS offre la possibilité d'intégrer une contrainte d'égalité spectrale pour guider l'analyse lorsque des références spectrales d'éléments de la composition sont connues. La signature spectrale du Delipidol® étant connue, les modèles obtenus par MCR-ALS appliquée avec et sans contrainte d'égalité spectrale sur le Delipidol® sont comparés entre eux ainsi qu'avec ceux obtenus par NCLS. Au final, dans cette étude, l'algorithme MCR-ALS est évalué sur son aptitude à suivre le Delipidol® à travers la peau en spectroscopie Raman. L'étude se concentre particulièrement sur la nécessité d'ajouter une contrainte d'égalité spectrale sur la signature du Delipidol® et l'évaluation des modifications de performances que cela peut engendrer. À noter que l'impact du choix du nombre de composants utilisés dans la MCR-ALS sur les modèles estimés est également analysé dans cette étude.

2.1.1. Ensemble des données expérimentales

Le Delipidol®, scientifiquement connu sous le nom de *punica granatum seed oil hydroxyphenethyl esters* et dont la formule brevetée est utilisée comme actif cosmétique anticellulite, a été fourni par le partenaire industriel du projet MISTIC (ARD 2020, Programme Cosmétosciences), à savoir

BioEurope Solabia, BioEurope France (Pantin, France). Dans cette étude, l'innocuité et l'efficacité de cette molécule sont considérées comme préalablement testées par le fournisseur et ne font donc pas l'objet d'une évaluation particulière. Seule la méthode d'estimation de sa cinétique de pénétration à travers la peau est étudiée. Pour cela, une solution éthanolique de Delipidol® à 5% est préparée. L'éthanol étant un activateur de pénétration, son utilisation conduit à une diffusion supérieure de la molécule qui permet d'analyser les variations de ses concentrations et d'augmenter sa sensibilité de détection par une analyse multivariée.

La solution contenant le Delipidol® est ensuite appliquée localement sur des échantillons d'épiderme humain reconstruit de surface 1.07 cm² commercialisés par Episkin® (Lyon, France). Ceux-ci ont été placés 30 minutes à température ambiante avant l'application de la solution. Après 1, 2 et 3 heures d'exposition, la surface des échantillons est rapidement nettoyée à l'aide d'une solution éthanolique afin d'y éliminer toute trace résiduelle de la solution. Puis des sections de 20 µm sur des échantillons traités et non traités sont réalisées à l'aide d'un cryo-microtome et placées sur un substrat de CaF_2 (fluorure de calcium qui est très utilisé en spectroscopie Raman en raisons de sa capacité à transmettre les longueurs d'ondes d'une large plage, de sa faible réponse Raman ne pouvant pas être confondue avec les résultats spectraux et de sa fluorescence minimale). Les coupes issues d'échantillons non traités servent d'éléments dits de contrôle.

Enfin, les coupes transversales fines d'échantillons traités et non traités sont analysées à l'aide d'un microspectromètre confocal Raman, un LabRam (Horiba Jobin-Yvon, Villeneuve-d'Ascq, France). Les images de diffusion Raman sont obtenues en utilisant un laser de 691 nm de puissance 8mW au niveau de l'échantillon et un objectif LMPlan x50 (Olympus, NA 0.75, Rungis, France). Le trou confocal est fixé à 250 µm pour toutes les mesures spectrales Raman et un réseau dispersif composés de 300 lignes/mm est utilisé pour disperser la diffusion Raman des échantillons sur un capteur CCD. Compte tenu de ces paramètres d'acquisition, la résolution spatiale des images Raman enregistrées est de 2 µm dans la direction x et 3 µm dans la direction y. L'acquisition de ces images Raman a pour objectif de définir la profondeur de pénétration du Delipidol® et d'évaluer l'éventuelle variabilité latérale de sa diffusion. À noter que toutes ces images Raman affichent le SC orienté vers la gauche et sont étudiées sur une zone d'environ 36×106 µm².

2.1.2. Prétraitement des données

Selon le protocole proposé dans le Chapitre III, tous les spectres composants les images Raman sont soumis à un prétraitement comprenant un débruitage spectral, une correction de la ligne de base et une normalisation. Le débruitage des spectres est d'abord réalisé à l'aide d'une ACP. Seul 1% du nombre total de spectres contenus dans les images Raman est utilisé pour décrire leurs caractéristiques spectrales sans bruit. Les informations sont ensuite concentrées dans la fenêtre spectrale [400-1800] cm^{-1} qui est la région spectrale d'intérêt des constituants majeurs de la peau et du Delipidol®. Enfin, les spectres sont corrigés de leurs lignes de bases respectives en utilisant la méthode Rubberband avant d'être normalisés avec la norme ℓ_2.

2.1.3. Caractérisation spectrale du Delipidol®

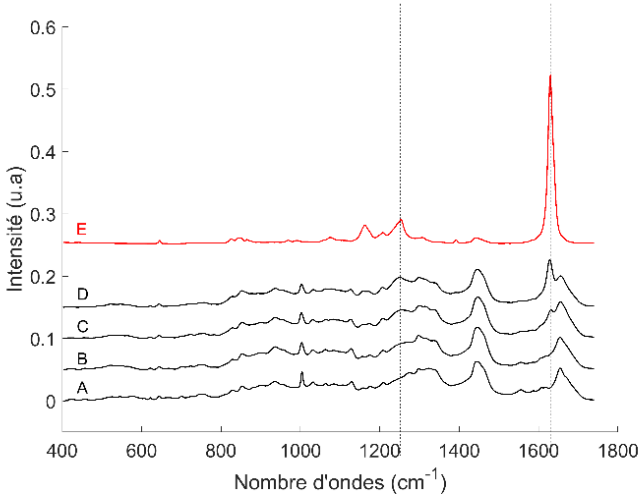


Figure IV-5 – Comparaison du spectre moyen des images spectrales à chaque durée d'exposition de la peau au Delipidol® avec la référence spectrale du Delipidol®. (**A**) *contrôle (peau non traitée) ;* (**B**) *1 heure ;* (**C**) *2 heures ;* (**D**) *3 heures et* (**E**) *Delipidol® pur. Les spectres sont prétraités selon la stratégie décrite ci-dessus.*

Le Delipidol® a une signature Raman caractéristique qui présente une combinaison de pics spécifiques intenses facilement identifiables et des bandes spectrales plus faibles chevauchant partiellement les caractéristiques de la peau (Figure IV-5(**E**)). Cette caractérisation fait de lui un candidat intéressant pour évaluer des approches de démélange telle que la MCR-ALS. En observant la signature spectrale du Delipidol® sur la Figure IV-5(**E**), des bandes situées à 1630 cm^{-1} (étirement C=C du cycle aromatique), 1254 cm^{-1} (C–C–C) et 1162 cm^{-1} dominent et d'autres bandes plus faibles à 1441, 1393, 1307, 1251, 1207, 1076, 991, 969, 865, 843 et 827 cm^{-1} se distinguent.

La Figure IV-5 illustre également les spectres moyens des cartes Raman d'échantillons de contrôle et d'échantillons après 1, 2 et 3 heures d'exposition au Delipidol®. Une description des bandes spectrales caractéristiques de la peau étant déjà donnée dans le Chapitre I, seule la diffusion du Delipidol® à travers la peau après 1, 2 et 3 heures d'exposition est caractérisée ici. Les bandes caractéristiques du Delipidol® se superposent à celles de la peau. Plus précisément, ses bandes dominantes, à savoir 1630, 1254 et 1162 cm^{-1}, sont très facilement distinguables (Figure IV-5 (**B-D**)). Cette distinction se révèle être évolutive en fonction du temps d'exposition des échantillons de peau au Delipidol®. Bien que la quantité croissante de Delipidol® dans la peau au cours du temps soit observable grâce à l'augmentation de l'intensité de ses bandes spectrales dans les spectres moyens des coupes de peau, les informations sur sa distribution latérale et sa profondeur de pénétration ne sont pas perceptibles avec cette représentation. L'application de la MCR-ALS sur ces images Raman a pour objectif de fournir des informations plus précises sur sa répartition dans les différentes couches de l'épiderme et permettre ainsi d'établir des cinétiques de pénétration.

2.1.4. Application de la MCR-ALS

2.1.4.1. MCR-ALS appliquée sans contrainte d'égalité spectrale sur le Delipidol® ($K = 4$)

Dans un premier temps, la MCR-ALS est appliquée en utilisant uniquement les contraintes de non-négativité et de normalisation sur les spectres des composants purs et les concentrations relatives à estimer. La signature spectrale du Delipidol® n'est pas utilisée pour guider l'algorithme et le nombre de spectres de composants purs à estimer est fixé à $K = 4$. La Figure IV-6 (**A**) illustre les

$K = 4$ composants du modèle estimé, identifiés par $S_{i=1,\dots,4}$ et ordonnés de façon à faciliter leur lecture et leur comparaison, ainsi que la carte de distribution du Delipidol® après 3 heures d'exposition.

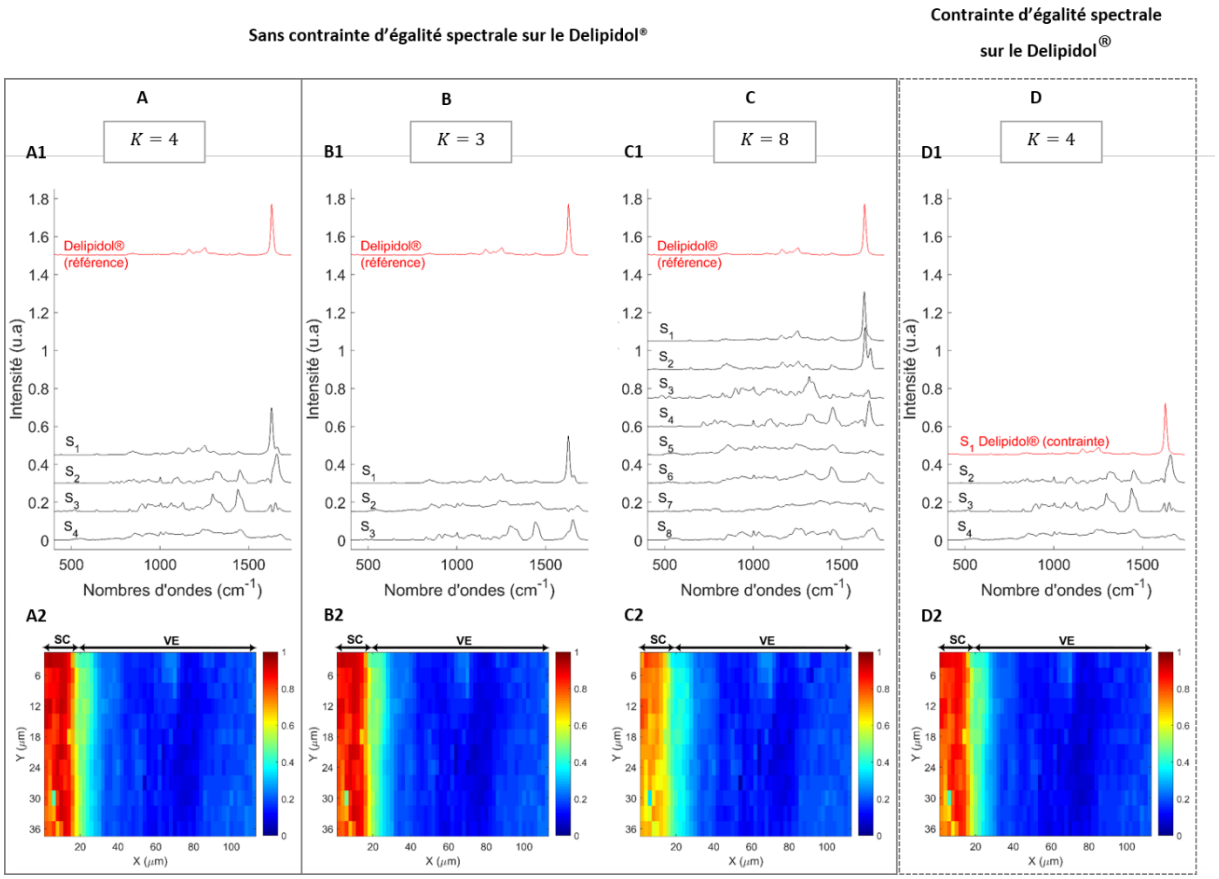


Figure IV-6 – Application de la MCR-ALS avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relative pour (**A**) $K = 4$ *et* (**B**, **C**) $K = \{3,8\}$, *et* (**D**) *avec une contrainte d'égalité spectrale supplémentaire sur le Delipidol® pour* $K = 4$, *sur des données prétraitées selon la stratégie décrite ci-dessus ;* (**A1**-**B1**-**C1**-**D1**) *Spectres estimés des* K *composants initialisés par SIMPLISMA (décalés en intensité pour plus de clarté) ;* (**A2**-**B2**-**C2**-**D2**) *Cartes de distribution du Delipidol® dans les échantillons de peau exposés pendant 3 heures. Cartes obtenues en estimant le spectre du Delipidol®* S_1 (**A2**-**B2**-**C2**) *ou le spectre référentiel* S_1 (**D2**). *SC, Stratum Corneum; EV, Épiderme Viable.* [105]

Par comparaison au spectre de référence du Delipidol®, le spectre S_1 lui est très similaire. La différence principale s'exprime par un épaulement situé à 1661 cm^{-1} qui correspond à un résidu caractéristique de la bande amide I. Néanmoins malgré cette variation, il est démontré que la MCR-ALS est capable d'identifier une composante spectrale fortement corrélée (0.9746) avec la référence spectrale du Delipidol® et que cette dernière n'est donc pas nécessaire pour obtenir sa carte de distribution spatiale. De plus, les spectres des 3 autres composants purs estimés, dont les contributions relatives sont illustrées dans la Figure IV-7, présentent des caractéristiques liées aux constituants biochimiques de la peau tels que les lipides, les protéines ou encore les acides nucléiques. Leur cohérence avec les spectres Raman d'origines est alors vérifiée.

Pour vérifier que la solution proposée par l'application de la MCR-ALS sans connaissance *a priori* sur la nature de l'AC, à savoir le Delipidol®, avec $K = 4$ composants purs est singulière, l'algorithme MCR-Bands est utilisé pour évaluer les potentielles ambiguïtés rotationnelles existantes. Les différences des contributions relatives des 4 composants aux limites maximales et

minimales sont estimées dans l'ordre à 1.16×10^{-3}, 8.00×10^{-3}, -5.81×10^{-3} et -25.61×10^{-3} et sont négligeables attestant ainsi l'unicité du modèle estimé.

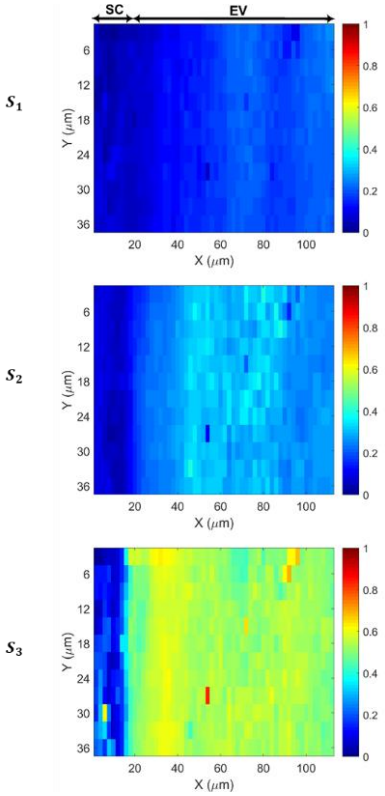


Figure IV-7 – Cartes de distribution du Delipidol® après 3 heures d'exposition reconstruites à partir des composants purs S_2, S_3 et S_4 obtenus à l'aide d'une MCR-ALS avec $K = 4$ composants et sans contrainte d'égalité spectrale. SC, Stratum Corneum; EV, Épiderme Viable. [105]

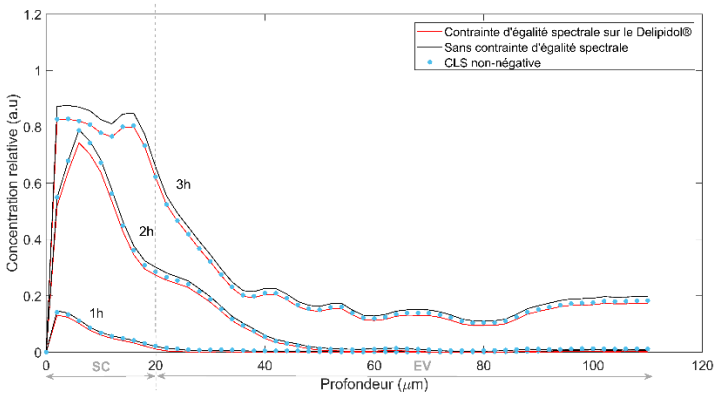


Figure IV-8 – Profils de pénétration moyens du Delipidol® pour 1, 2 et 3 heures de temps d'exposition obtenus par MCR-ALS appliquée avec (rouge) et sans (noir) contrainte d'égalité spectrale sur le Delipidol® pour $K = 4$. Les concentrations relatives moyennes obtenues par CLS avec contrainte de non-négativité sont illustrées par les points bleus. SC, Stratum Corneum; EV, Épiderme Viable. [105]

Au final, à partir des cartes de distribution du Delipidol® reconstruites pour chaque temps d'exposition, des profils de pénétration ont pu être définis en calculant la moyenne des concentrations relatives estimées par niveau de profondeur cutanée. Ces profils de pénétrations sont illustrés dans la Figure IV-8. Ils montrent que le Delipidol® a commencé à pénétrer le SC après 1h d'exposition, et qu'il continue de se diffuser progressivement dans les couches plus profondes de l'épiderme après 2 et 3h d'exposition tout en conservant une concentration relative conséquente dans les 20 premiers micromètres, soit le SC. Par ailleurs, ces estimations de cinétique de pénétration sont proches de celles obtenus par l'approche plus contraignante de la NCLS.

2.1.4.2. Comparaison des nombres de composantes

Le choix de $K = 4$ composants purs à estimer pour modéliser les images Raman dans la précédente partie a été basé sur la similarité du spectre estimé S_1 avec la référence spectrale du Delipidol® et reste jusqu'ici subjectif. La peau étant un tissu biologique complexe en terme de composition moléculaire et structurelle, ces composants biochimiques sont assurément nombreux. Le but de la MCR-ALS est de simplifier les informations pertinentes contenues dans les données pour en faciliter leur interprétation. Cependant, comme mentionné dans la partie 1.1 de ce Chapitre IV, le nombre de composants purs à estimer influence grandement les performances et les résultats de la MCR-ALS. Par conséquent, la sélection du nombre optimal de composants pour capturer le maximum d'informations pertinentes contenues dans les images Raman des échantillons de peau exposés ou non au Delipidol® est présumée difficile mais est indispensable pour modéliser au mieux les données.

Comme détaillés précédemment dans ce Chapitre IV, trois critères ont été utilisés pour guider la sélection du nombre idéal de composants purs à estimer : la SVD, le critère DW et l'indice KMO (Figure IV-9).

- La variance expliquée cumulée obtenue par SVD sur les images Raman originales, dont le graphique est illustré sur la Figure IV-9 (**A**), montre que la quantité d'informations prises en compte dans le modèle estimé augmente avec le nombre de composants purs utilisés pour la décomposition spectrale des images Raman. Ces résultats suggèrent que 11 composants sont nécessaires pour atteindre 95% de la variance expliquée. Cependant, les 4 premiers composants suffisent à exprimer les variations les plus importantes, à savoir 85% de la variance expliquée.
- Le critère DW est calculé à partir des modèles résultants de SIMPLISMA en utilisant les 20 premiers spectres considérés comme les plus purs. La Figure IV-9 (**B**) affiche des coefficients de DW tous plus proche de 0 que de 2, ce qui signifie que les spectres résiduels sont corrélés entre eux. Autrement dit, les spectres de composants purs n'ont pas tous été extraits des données. Une augmentation brutale du coefficient de DW indique que l'autocorrélation entre spectres résiduels diminue. Cette diminution fait suite à l'identification et l'extraction d'un spectre de composants fortement contributif dans les données. De telles augmentations du critère de DW s'observent lorsque $K = 4$ et $K = 8$.
- Les indices KMO calculés à partir des modèles obtenus par SIMPLISMA en utilisant les 20 premiers composants considérés comme les plus purs sont illustrés sur la Figure IV-9 (**C**). Ils diminuent au fur et à mesure que le nombre de composants dans le modèle augmente. Cependant, une baisse significative de l'indice est observée pour un modèle avec $K = 4$ composants et un léger changement apparait lorsque $K = 8$ composants sont utilisés. Ces informations traduisent une baisse considérable de la corrélation résiduelle et témoignent de l'extraction d'un composant très contributif dans les données.

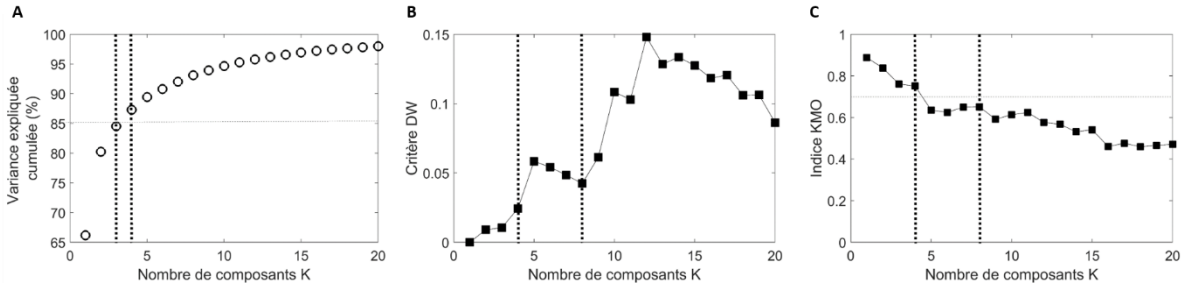


Figure IV-9 – (**A**) *Variance expliquée cumulée obtenue par SVD appliquée sur les images Raman des échantillons exposés au Delipidol® ;* (**B**) *Critère de Durbin-Watson appliqué sur les résultats de SIMPLISMA ;* (**C**) *Indice KMO appliqué sur les résultats SIMPLISMA.*

Cette analyse montre que la sélection du nombre optimal de composants à estimer pour ajuster au mieux un modèle aux données est très ambiguë et confuse. En effet, ces critères ne reflètent pas nécessairement directement le nombre optimal de composants pour détecter et localiser précisément l'AC dans des images biochimiques puisque tous s'accordent à dire qu'un nombre de composants élevé serait nécessaire pour englober l'ensemble de la variabilité spectrale biochimique. Cependant, l'utilisation de $K = \{4,8\}$ composantes a été suggérée par ces 3 critères.

Pour évaluer l'impact que peut avoir le nombre de composants purs dans le démélange spectral et dans la détermination des profils de pénétration du Delipidol® dans la peau, la MCR-ALS est appliquée à nouveau sans contrainte d'égalité spectrale sur le Delipidol® mais cette fois-ci en utilisant $K = 3$ et $K = 8$ composants purs. Les spectres de composants purs estimés et les cartes de distribution du Delipidol® après 3 heures d'exposition pour ces deux modèles sont illustrés

dans les Figure IV-6 (**B**-**C**). Comme attendu dans une étude de pénétration cutanée d'une molécule exogène, les cartes de distributions du Delipidol® venant des modèles estimés avec $K = \{3, 4, 8\}$ sans intégrer les informations connues sur cette molécule affichent toutes des gradients de diffusion allant de la surface de la peau jusqu'aux couches plus profondes de l'épiderme viable après 3 heures d'exposition. Les cartes de distribution du Delipidol® des modèles avec $K = 3$ et $K = 4$ sont très similaires. En revanche, celle du modèle avec $K = 8$ diverge. En effet, une contribution de Delipidol® nettement plus faible dans les 20 premiers micromètres est observée. Cette observation se confirme par le calcul des valeurs de contribution moyenne en fonction de la profondeur résumées dans le Tableau IV-1. Les valeurs des concentrations relatives moyennes sont sensiblement identiques pour $K = 3$ et $K = 4$ pour les 3 temps d'exposition. Après 3 heures d'expositions, la concentration relative moyenne estimée dans ces modèles est respectivement de 0.8050 et 0.8098 de 0 à 20 µm tandis que celle estimée dans le modèle avec $K = 8$ est de 0.6455. Cette différence apparait seulement dans les premiers micromètres de la peau pour chaque temps d'exposition. Autrement, les trois modèles s'accordent sur la diffusion progressive du Delipidol® dans les couches inférieures de l'épiderme au cours du temps.

Tableau IV-1 – Concentrations relatives estimées du Delipidol® moyennes après 1, 2 et 3 heures d'exposition obtenues par MCR-ALS appliquée avec et sans contrainte d'égalité spectrale sur le Delipidol® ($K = \{3, 4, 8\}$).

		$K = 3$			$K = 4$			$K = 8$		
		1h	**2h**	**3h**	**1h**	**2h**	**3h**	**1h**	**2h**	**3h**
Sans contrainte d'égalité spectrale	0 – 20 µm	0.0635	0.5439	**0.8050**	0.0691	0.5364	**0.8098**	0.0319	0.1986	**0.6455**
	20 – 40 µm	0.0012	0.1737	0.3477	0.0067	0.1753	0.3498	0.0028	0.1100	0.3064
	40 – 110 µm	0.0058	0.0129	0.1532	0.0036	0.0115	0.1623	0.0038	0.0110	0.1616
Contrainte d'égalité spectrale sur le Delipidol®	0 – 20 µm	0.0534	0.5072	**0.7643**	0.0586	0.4995	**0.7638**	0.0026	0.1220	**0.4904**
	20 – 40 µm	0.0001	0.1553	0.3246	0.0013	0.1567	0.3247	0.0000	0.0148	0.1423
	40 – 110 µm	0.0016	0.0043	0.1343	0.0009	0.0049	0.1426	0.0066	0.0083	0.0244

L'observation des spectres de composants purs estimés dans les 3 modèles permet également d'expliquer les différences de contribution relative du Delipidol® estimée. Une double expression des bandes caractéristiques du Delipidol® dans les spectres estimés du modèle avec $K = 8$ peut être par exemple observée. Bien que la quasi-totalité des bandes caractéristiques du Delipidol® soient présentes dans le spectre S_1, ces dernières sont également présentes dans le spectre S_2 et sont mélangées à des caractéristiques spectrales de la peau. Cette situation illustre que le Delipidol® n'est plus considéré comme un seul composant pur dans ce modèle et cette sur-décomposition du spectre du composant pur d'intérêt est finalement la cause de la différence de contribution avec les modèles $K = 3$ et $K = 4$.

Enfin, la fiabilité des modèles estimés est évaluée sur la base des coefficients de corrélation entre les spectres de composants purs et la signature de référence du Delipidol® mais aussi sur la distribution des spectres résiduels.

- Les coefficients de corrélation entre les composants estimés de chaque modèle et la référence du Delipidol®, présentés dans le **Erreur ! Source du renvoi introuvable.**Tableau IV-2, permettent de mesurer leur degré de similarité. Par exemple, les spectres S_1 des modèles avec $K = 3$ et $K = 4$ composants purs sont fortement corrélés à la signature du Delipidol® de référence (0.9753 et 0.9746). En revanche, les spectres S_2 et S_3 ne lui correspondent pas puisque leurs coefficients de corrélation sont faibles : 0.0054 et 0.2664 pour $K = 3$ et 0.2518 et 0.0681 pour $K = 4$. Ces informations permettent de considérer les spectres S_1 comme proche de la pureté spectroscopique. À l'inverse, pour $K = 8$, les spectres S_1 et S_2 présentent

tous deux une forte corrélation avec le spectre référentiel du Delipidol®, à savoir 0.9596 et 0.8801, prouvant une fois de plus la sur-décomposition spectrale des données.

- La distribution des résidus du modèle estimé à l'aide d'histogramme permet également d'évaluer la performance de la MCR-ALS. Les histogrammes des spectres résiduels moyens des modèles avec $K = \{3, 4, 8\}$ composants purs estimés sans contrainte d'égalité spectrale sur le Delipidol® sont illustrés dans les Figure IV-10(**D**-**F**). Ils apportent des informations supplémentaires sur l'impact du nombre de composants K utilisés pour reconstruire les données. Par exemple, pour $K = 3$, l'histogramme des résidus moyens est formé par deux gaussiennes. Cela indique que les composants les plus contributifs aux données ne sont pas tous extraits. En effet, la gaussienne située à droite peut être considérée comme l'expression du bruit résiduel alors que la gaussienne située à gauche indique qu'une certaine variabilité des données n'est pas modélisée mathématiquement. En revanche, l'ajout d'un composant pur dans le modèle permet de restaurer une distribution gaussienne unimodale. En effet, la distribution presque normale des résidus moyens obtenus avec un modèle à $K = 4$ composants purs suggère qu'aucun composant fortement contributif ne reste dans les résidus. De plus, il est observé que la distribution gaussienne des résidus se recentre autour de zéro lorsque le nombre de composants K augmente. En effet, les gaussiennes des modèles $K = \{3, 4, 8\}$ sont centrées respectivement à -0.85×10^{-3}, -0.25×10^{-4} et 0.05×10^{-4}. Les décalages dans le négatif illustrent donc la surexpression mathématique d'une certaine variabilité des données dans le modèle lorsque le nombre de composant K est trop petit.

Tableau IV-2 – Coefficients de corrélation entre les spectres estimés $S_{i=1,\dots,8}$ obtenus par MCR-ALS appliquée avec et sans contrainte d'égalité spectrale sur le Delipidol® ($K = \{3, 4, 8\}$) sur les images Raman après 1,2 et 3h d'exposition.

	$K = 3$			$K = 4$				
	Contrainte d'égalité spectrale							
Sans contrainte d'égalité spectrale sur le Delipidol®	S_1 **Delipidol®**	S_2	S_3	S_1 **Delipidol®**	S_2	S_3	S_4	
S_1	**0.9753**	0.0512	0.3452	**0.9746**	0.3218	0.1393	0.1821	
S_2	0.0054	**0.9994**	0.4070	0.2518	**0.9983**	0.4214	0.3911	
S_3	0.2664	0.4425	**0.9971**	0.0681	0.4535	**0.9979**	0.7216	
S_4				0.1110	0.4206	0.6790	**0.9999**	
	$K = 8$							
	S_1 **Delipidol®**	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1	**0.9596**	**0.8348**	0.6598	0.1339	0.2040	0.2352	0.1259	0.1831
S_2	**0.8801**	**0.9935**	0.3895	0.3115	0.1540	0.3157	0.1011	0.3184
S_3	0.0812	0.1170	0.2006	0.3493	0.9519	0.6006	0.4535	0.5224
S_4	0.0883	0.4265	0.5153	**0.8862**	0.2018	0.6665	0.2074	0.6071
S_5	0.0739	0.2149	0.4273	0.1333	0.6773	0.7912	**0.9305**	**0.8007**
S_6	0.1320	0.3577	0.5838	0.4962	0.6287	**0.9993**	0.7647	**0.8698**
S_7	0.2024	0.1498	0.6264	0.5849	0.1092	0.4096	0.5617	0.3143
S_8	0.1705	0.3429	0.5124	0.4314	0.5010	0.8419	0.7302	**0.9703**

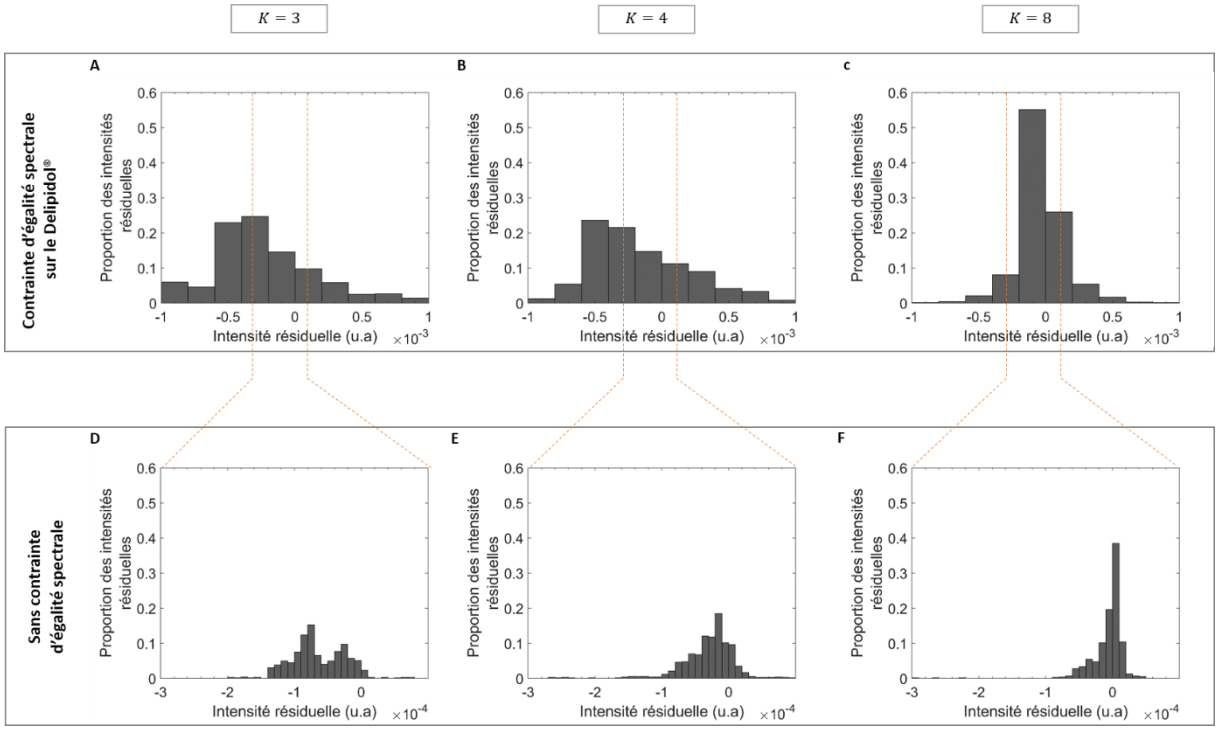


Figure IV-10 – Histogrammes des résidus spectraux moyens obtenus par MCR-ALS appliquée (A-B-C) avec et (D-E-F) sans contrainte d'égalité spectrale sur le Delipidol® pour $K = \{3,4,8\}$ à partir des images Raman des échantillons de peau exposés pendant 1, 2 et 3 heures.

Au final, l'étude comparative des modèles estimés avec différents nombres de composants montre que le choix du nombre optimal de composants purs pour la construction d'un modèle simplifié est difficile et que les critères couramment utilisés à ces fins ne sont pas suffisants. La performance de la décomposition spectrale doit être évaluée sur la cohérence des spectres estimés et la fidélité du modèle aux données. Il a donc été démontré qu'un nombre trop élevé de composants n'est pas une solution pour améliorer l'ajustement du modèle aux données puisque cette façon d'opérer peut engendrer une sur-décomposition des spectres de composants purs. À l'inverse, un nombre trop faible de composants purs ne permet pas un ajustement du modèle aux données optimal en terme de résidus. En définitive, pour cette étude d'évaluation de la performance de la MCR-ALS pour suivre la pénétration du Delipidol® à travers la peau, le nombre de composants purs $K = 4$ à estimer semble être un bon compromis entre une erreur d'ajustement aux données acceptable et une décomposition spectrale cohérente.

2.1.4.3. Comparaison MCR-ALS avec et sans contrainte d'égalité spectrale sur le Delipidol®

L'application de techniques spectroscopiques pour l'analyse des profils de pénétration des AC dans le domaine cosmétique signifie généralement que les produits eux-mêmes sont disponibles et que des spectres purs ou des références spectrales peuvent être collectés avant leur application sur la peau. En termes d'application de protocoles d'analyses multivariées, ces références peuvent être employées pour guider l'analyse en les utilisant comme contrainte d'égalité spectrale ou pour valider l'analyse, comme cela a été fait juste avant. Il convient de garder à l'esprit que l'insertion de contraintes supplémentaires conduit naturellement à un plus grand manque d'ajustement en raison d'un ensemble de conditions à respecter qui diminue la flexibilité du modèle à trouver la solution. En effet, l'utilisation d'un spectre de référence fixe dans le processus itératif de la MCR-ALS conduit à un degré de liberté amoindri par rapport à une approche n'utilisant que des

contraintes de non-négativité et de normalisation. Ce degré de flexibilité doit donc être pris en compte dans l'analyse des résidus des modèles estimés.

Dans la suite de cette étude, la référence spectrale du Delipidol® est utilisée pour guider l'analyse et aider la MCR-ALS à estimer ses contributions relatives locales. La MCR-ALS est alors appliquée en utilisant $K = 4$ spectres de composants purs, où le spectre S_1 est soumis à la contrainte d'égalité spectrale avec le Delipidol® tandis que les spectres S_2, S_3 et S_4 sont calculés librement par l'algorithme itératif. Comme attendu, l'ajout d'une contrainte d'égalité spectrale a conduit à une augmentation des résidus du modèle (Figure IV-10(**B**)). En effet, les distributions des résidus moyens sont plus proches d'une distribution normale mais les erreurs d'ajustement sont 10 fois plus élevées (ordre de 10^{-3}) que celles obtenues sans utiliser de contrainte d'égalité spectrale (ordre de 10^{-4}). Pour plus de précision, le modèle obtenu sans contrainte d'égalité spectrale a un manque d'ajustement de 5.21% et comme attendu, l'ajout d'une contrainte d'égalité spectrale sur le Delipidol® dans l'analyse accentue ce manque d'ajustement (7.39%).

Néanmoins, les spectres de composants purs estimés et les cartes de distribution du Delipidol® ne présentent aucune différence notable par rapport au modèle calculé sans contrainte d'égalité spectrale supplémentaire (Figure IV-6(**A**,**D**). En effet, les spectres de composants purs estimés avec et sans contrainte d'égalité spectrale sur le Delipidol® sont très similaires puisque les coefficients de corrélation entre les spectres S_2, S_3 et S_4 sont très proches de 1, avec des valeurs respectives de 0.9983, 0.9979 et 0.9999 (Tableau IV-2). Ce résultat n'est pas surprenant étant donné que sans contrainte d'égalité spectrale, la MCR-ALS est capable d'estimer un spectre S_1 très proche de celui de la référence du Delipidol® ce qui par conséquent conduit à des décompositions mathématiques des autres composants très similaires. La similarité des résultats des deux modèles estimés avec et sans contrainte d'égalité spectrale est également confirmée par la comparabilité des tendances des profils de pénétration du Delipidol® estimés : le Delipidol s'accumule dans le SC et se diffuse progressivement dans les couches plus profondes de la peau au cours du temps (Figure IV-8 et Tableau IV-1). De plus, les cinétiques de pénétration obtenues à l'aide d'une contrainte d'égalité spectrale supplémentaire sont, au même titre que celles obtenues sans cette contrainte, proches des cinétiques de pénétration estimée par NCLS. Cependant, une baisse légère de la concentration relative moyenne calculée à partir des résultats obtenus avec la contrainte d'égalité spectrale est mise en évidence dans le Tableau IV-1. En effet, le Delipidol® est estimé être présent dans les 20 premiers micromètres à hauteur de 6.91%, 53.64% et 80.98% après respectivement 1, 2 et 3h d'exposition lorsqu'aucune contrainte d'égalité spectrale n'est imposée dans l'analyse tandis qu'il est estimé être présent à hauteur de 5.86%, 49.95% et 76.38% lorsque cette contrainte est appliquée. Malgré cette légère différence de contribution relative, les profils de pénétration du Delipidol® des deux approches sont très similaires, ce qui soutient la possibilité d'utiliser la MCR-ALS sans contrainte d'égalité spectrale pour ce cas d'étude.

2.1.5. Conclusion

Au final, la MCR-ALS s'est révélée être plus adaptative et donc plus facile à utiliser que la CLS en raison de sa faible exigence sur la disponibilité de spectres de référence pour effectuer l'analyse. Elle se présente donc comme une alternative potentiellement puissante pour le démélange spectral et pour l'étude de la distribution d'un AC dans les sections transversales de la peau. L'utilisation de la MCR-ALS avec une contrainte d'égalité spectrale s'est avérée être pertinente lorsqu'une information sur un des composants purs est disponible car cette contrainte supplémentaire a permis de profiler sa distribution au sein des images Raman acquises. Dans le cas du Delipidol®, l'avantage de l'utilisation de cette contrainte d'égalité spectrale réside dans la

réduction des ambiguïtés rotationnelles malgré une légère perte de qualité d'ajustement. Néanmoins, la MCR-ALS appliquée avec des contraintes d'égalité spectrale pourrait aussi souffrir de limitations dans le cas où le spectre d'intérêt est modifié en raison des nombreuses interactions moléculaires induites dans l'environnement local. Les résultats obtenus avec la MCR-ALS non supervisée et sans contrainte sur l'AC confirment l'intérêt du protocole proposé pour l'analyse des images Raman.

2.2. SUIVI DE LA CAFÉINE DANS UN MODÈLE DE PEAU RECONSTRUITE

Depuis plusieurs dizaines d'années, la caféine est une molécule très utilisée dans les produits cosmétiques. Les raisons de son utilisation sont multiples. Elle est ajoutée dans des formulations cosmétiques dans le but de stimuler la pousse des cheveux, d'inhiber les rayons UVB du soleil, de favoriser la cicatrisation ou même encore de traiter la cellulite. Sa simplicité d'utilisation et son faible coût ont de plus participé à la production de nombreux articles scientifiques la concernant. En effet, plusieurs études spectroscopiques Raman ont été menées pour suivre la caféine à travers la peau et évaluer sa capacité à pénétrer la barrière cutanée. Par exemple, S. Tfaili *et al.* [36] ont suivi la caféine à faibles concentrations à travers des peaux humaines excisées. Cette étude a montré qu'il était possible de détecter par spectroscopie Raman la caféine respectant les concentrations utilisées dans des formulations dermatologiques. De même, C. Alonso *et al.*[38] et S.M. Ascensio *et al.* [106] ont montré que la spectroscopie Raman était un outil puissant pour la détection et le suivi de la caféine à travers des peaux de porc excisées.

Toutefois d'un point de vue chimiométrique, la complexité et l'intensité modérée de la signature spectrale de la caféine à l'état pur la rend plus difficilement détectable lorsqu'elle est appliquée à faible concentration comme cela est souvent le cas dans les produits cosmétiques présents sur le marché. Nombreuses de ses bandes spectrales caractéristiques chevauchent des bandes Raman spécifiques de la peau plus intenses. La caféine représente donc un modèle d'étude intéressant pour tester les méthodes d'analyse multivariée MCR-ALS et MVC-NMF dans des situations plus complexes, à savoir dans des situations où l'ingrédient actif est plus difficile à détecter.

Pour évaluer la capacité des méthodes MCR-ALS et MVC-NMF à détecter et suivre la progression de la caféine dans la peau, les analyses sont réalisées à partir de données spectrales Raman acquises dans le cadre d'une étude portant sur le développement de formulation complexe permettant à la caféine d'atteindre ces cellules cibles en profondeur. En effet, la caféine est une molécule hydrophile. Elle a par conséquent de grandes difficultés à atteindre les couches plus profondes de la peau. Les images Raman analysées proviennent donc d'épidermes humains reconstruits exposés à deux formulations de caféine :

- Une formulation réalisée à l'aide d'une émulsion classique comportant des tensio-actifs (TA)
- Une formulation réalisée à l'aide de la technologie Hydrofeeling® (HF) développée par l'entreprise GeniAlis. Cette technologie est une des alternatives innovantes permettant de simplifier les formules des produits dermo-cosmétiques en réduisant le nombre d'ingrédients dans les formulations et en limitant le nombre d'ingrédients issus de la synthèse chimique afin d'éviter des effets indésirables pour l'Homme et l'environnement comme le veut la tendance dans la recherche cosmétique actuelle. Le principe de cette technologie consiste à réaliser une émulsion à froid respectueuse des ingrédients actifs utilisés en éliminant l'utilisation de substances potentiellement irritantes ou allergisantes telles que des tensio-actifs. Ce procédé est basé sur une exposition du produit à des

ultrasons dans une gamme de fréquences bien définies pour stabiliser les solutions ou les émulsions sans ajouter des tensio-actifs solubilisants ou émulsionnants.

Nous proposons donc de mener une étude comparative sur les résultats obtenus par les deux méthodes MCR-ALS et MVC-NMF à partir de ces données afin de voir si l'une d'elle est plus performante pour suivre la pénétration d'un ingrédient actif plus difficilement détectable.

2.2.1. Ensemble des données expérimentales

Deux formulations contenant la molécule de caféine ont été préparées. Chacune d'elles est constituée de 2% de caféine. La première est une émulsion classique utilisant des tensio-actifs (TA). La seconde est réalisée selon le procédé Hydrofelling® sans ajout de tensio-actifs (HF).

Les deux formulations contenant la caféine sont appliquées localement sur des échantillons d'épiderme humain reconstruit provenant de chez Episkin® (Lyon, France). Ceux-ci ont été placés 30 minutes à température ambiante avant l'application des solutions. Après 8 et 24 heures d'exposition, des sections de coupes transversales ont été réalisées et placées sur un substrat de CaF_2.

Enfin, les coupes transversales fines d'échantillons de peaux exposées à la caféine sont analysées à l'aide d'un microspectromètre confocal Raman aplha 300R (WITec, Hulm, Allemagne). Les images de diffusion Raman sont obtenues en utilisant un laser de 633nm de puissance 25mW au niveau de l'échantillon et un objectif x50 (NA 0.8) et ont une dimension de 80×60 pixels incrémentés de 2 µm dans les deux directions x et y. Toutes présentent le SC orienté vers le haut de l'image.

2.2.2. Prétraitement des données

Les spectres des quatre images Raman acquises sont soumis à la routine de prétraitement présentée dans le Chapitre III. Plus précisément, le prétraitement comprend un débruitage spectral, une correction de la ligne de base, une annulation des spectres correspondant au substrat, une suppression des artefacts et une normalisation. Le débruitage des spectres est réalisé à l'aide d'une projection par ACP des spectres sur 1% de leur nombre total de composantes principales. Puis, un réalignement spectral sur la bande caractéristique de la peau située à 1004cm^{-1} est effectué et est suivi par la sélection de la région spectrale [400-1800] cm^{-1} pour appliquer les méthodes d'analyse multivariée puisqu'elle représente la région d'intérêt des constituants de la peau et de la caféine. Ensuite, pour que les spectres associés uniquement au substrat et ne correspondant pas au tissu cutané lui-même ne perturbent pas l'analyse multivariée, ceux-ci sont tous mis à zéro. Le calcul de l'écart-type des intensités Raman de chaque spectre permet de les identifier : les spectres présentant un très faible écart-type correspondent très probablement à des spectres ne comprenant que le bruit associé au substrat. La méthode K-means est alors appliquée sur les valeurs d'écart-type estimées afin de dissocier les régions associées au substrat de celles associées au tissu cutané. Enfin, les spectres sont corrigés de leurs lignes de bases respectives en utilisant la méthode Rubberband et de leurs artefacts avec la méthode hybride proposée utilisant la technique de « wound healing » avant d'être normalisés avec la norme ℓ_2. Les spectres des quatre images Raman ainsi prétraités sont ensuite concaténés pour ne former qu'une seule matrice de données qui sera soumise aux méthodes de décomposition spectrale MCR-ALS et MVC-NMF par une approche dite multi-set permettant de décomposer l'ensemble des données en même temps.

2.2.3. Caractérisation spectrale de la caféine

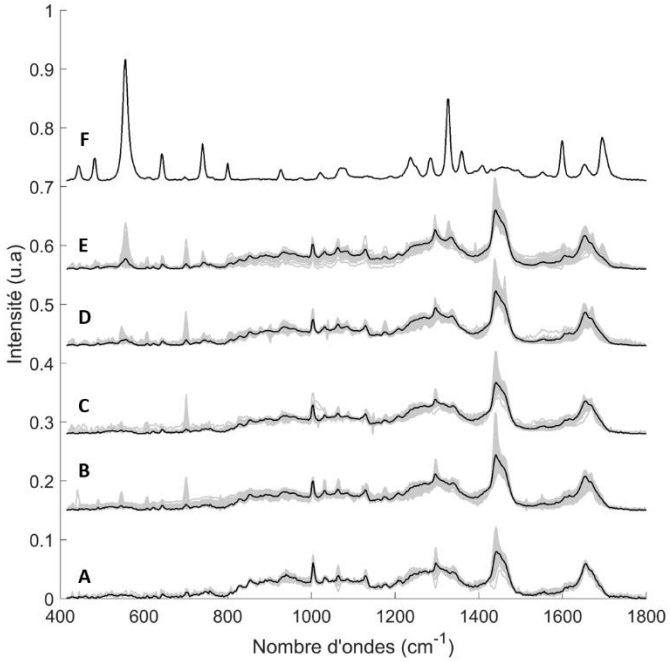


Figure IV-11 – Comparaison des spectres à chaque temps d'exposition aux deux formulations contenant la caféine avec la référence spectrale de la caféine. (**A**) *contrôle ;* (**B**) *8h, émulsion avec tensio-actifs ;* (**C**) *8h, émulsion par Hydrofeeling® ;* (**D**) *24h, émulsion avec tensio-actifs ;* (**E**) *24h, émulsion par Hydrofeeling® ;* (**F**) *caféine. Les spectres sont prétraités selon la stratégie décrite ci-dessus. Les spectres noirs représentent la moyenne de tous les spectres représentés en gris.*

La caféine a une signature Raman caractéristique qui présente une combinaison de pics spécifiques plus ou moins intenses (Figure IV-11(**F**)). Utilisée en faible concentration dans les produits cosmétiques, ses bandes spectrales sont parfois difficilement identifiables lorsqu'elles sont mélangées à celles de la peau en raison de leur chevauchement. Néanmoins, la caféine présente des bandes spectrales caractéristiques identifiables parmi les bandes associées à la peau (Figure IV-11). La bande spectrale de la caféine située à 556 cm^{-1}, caractérisant une déformation du groupe (C=O-N) (δ(C=O-N)), est particulièrement utilisée comme marqueur de sa présence en raison de son intensité plus importante et du fait qu'elle ne chevauche pas les bandes de la peau. Elle modifie particulièrement les spectres de la peau dans la région [505-595] cm^{-1}. De même, les bandes situées à 1328 cm^{-1} et 1360 cm^{-1} impliquent une modification de la structure secondaire des protéines qui se caractérise dans la bande amide III située dans la région spectrale [1250-1380] cm^{-1}. Enfin, les bandes situées à 1600 cm^{-1} et 1695 cm^{-1} caractéristiques respectivement d'une élongation CC (υ(C=C)) et d'une élongation CO (υ(C=O)) induisent des changements dans la région spectrale autour de la bande amide I de la peau [1590-1730] cm^{-1}.

Par l'observation des spectres issus des échantillons exposés aux deux formulations de la caféine à des temps différents (Figure IV-11(**B**-**E**)), la caféine dont la détection est non aisée semble pénétrer l'épiderme de manière plus efficace après 24 heures d'exposition pour les deux formulations. En revanche, la formulation HF semble améliorer la pénétration de la caféine en comparaison à la formulation TA. Ces observations restent à vérifier à l'aide des méthodes d'analyse multivariée telles que la MCR-ALS et la MVC-NMF.

2.2.4. Analyse de données

2.2.4.1. Estimation du nombre de composants

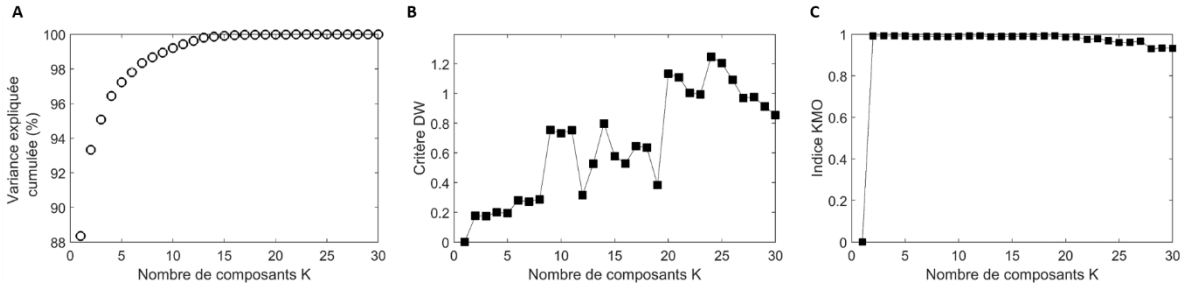


Figure IV-12 – (**A**) *Variance expliquée cumulée obtenue par SVD appliquée sur les images Raman des échantillons exposés à la caféine ;* (**B**) *Critère de Durbin-Watson appliqué sur les résultats de SIMPLISMA ;* (**C**) *Indice KMO appliqué sur les résultats SIMPLISMA.*

L'objectif fondamental des méthodes d'analyses multivariées MCR-ALS et MVC-NMF est de faciliter l'interprétation de la composition moléculaire et structurelle des données en simplifiant les informations spectrales à l'aide d'un nombre optimal de composants les décrivant au mieux. Cependant, étant donné la nature complexe du tissu cutané, l'estimation de ce nombre de composants purs nécessaire à la description la plus fidèle possible aux données reste difficile.

Comme détaillés précédemment dans ce Chapitre IV, trois critères ont été utilisés pour estimer le nombre de composants nécessaire à la modélisation la plus proche des données expérimentales: la SVD, le critère DW ainsi que l'indice KMO (Figure IV-12). Tous démontrent qu'un nombre de composants élevé serait nécessaire pour capturer un maximum de variabilité spectrale. Cependant, une analyse du comportement de ces critères peut aider à déterminer un nombre réduit de composants suffisant à la description des données.

- La variance expliquée cumulée obtenue par SVD sur les images Raman originales, dont le graphique est illustré sur la Figure IV-12 (**A**), montre que la quantité d'informations exprimées par les modèles estimés augmente avec le nombre de composants purs utilisés pour le démélange spectral des images Raman. Plus précisément, ces résultats montrent qu'au moins 10 composants purs sont nécessaires pour atteindre plus de 99% de variance expliquée. Cependant, seulement 2 composants purs suffisent pour décrire plus de 90% des données expérimentales, 3 composants purs peuvent décrire 95% des données et 5 composants purs sont nécessaires pour modéliser plus de 97.25% des données.

- Le critère DW est calculé sur les résidus des modèles définis à partir des 30 premiers spectres les plus purs estimés par SIMPLISMA. La Figure IV-12 (**B**) affiche des coefficients de DW tous compris entre 0 et 1, ce qui signifie que les spectres résiduels sont corrélés entre eux. L'augmentation brutale du coefficient de DW entre 1 et 2 composants indique que l'autocorrélation entre spectres résiduels diminue et qu'une deuxième composante fortement contributive dans les données est identifiée et extraite. De fortes augmentations du critère DW similaires sont également observées lorsque $K = \{3, 5, 8, 11\}$.

- Les indices KMO sont calculés sur les résidus des modèles définis à partir des 30 premiers spectres les plus purs estimés par SIMPLISMA. La Figure IV-12 (**C**) montre que les indices KMO estimés à l'aide de 1 à 30 composants sont tous supérieurs à 0.9, ce qui signifie que tous les composants utilisés semblent être décorrélés entre eux. Ce résultat ne permet

pas de définir une idée du nombre réduit de composants idéal à utiliser pour simplifier au mieux les données expérimentales.

Une fois de plus, l'analyse de ces trois critères montre que le choix du nombre optimal de composants purs à estimer pour modéliser au mieux les données est très équivoque. Cependant, par observation des critères SVD et DW, les nombres de composants retenus à utiliser dans les analyses MCR-ALS et MVC-NMF sont $K = \{2, 3, 5, 8, 11\}$. Dans la suite, seuls les nombres $K = \{2, 3, 5\}$ seront testés puisque $K = 5$ semble suffisant pour modéliser les données.

2.2.4.2. Application de la MCR-ALS

Une première étude concernant le choix de la méthode d'initialisation des spectres de composants purs à utiliser a été réalisée. Cette étude oppose les initialisations des spectres de composants purs par les résultats des méthodes K-means et SIMPLISMA. Après initialisations des spectres de composants purs, la méthode MCR-ALS a été appliquée en utilisant des contraintes de non-négativité et de normalisation sur les spectres de composants purs et leurs concentrations relatives associées avec $K = \{2,\ 3,\ 5\}$ composants. Il a été observé qu'avec ou sans contrainte d'égalité spectrale sur la caféine supplémentaire, les résultats obtenus par la MCR-ALS initialisée à l'aide de la méthode K-means sont non concluants : les spectres de composants purs estimés sont tous très similaires au spectre de données moyen, aucun démélange de données n'est effectué et l'entièreté de l'information concernant la caféine n'est pas dissociée de celle concernant la peau ce qui rend donc impossible le suivi de la caféine à travers les échantillons d'épiderme. À l'inverse, la MCR-ALS initialisée à l'aide de la méthode SIMPLISMA et appliquée avec ou sans contrainte d'égalité spectrale sur la caféine supplémentaire fournit des spectres de composants purs estimés différents les uns des autres ce qui montre qu'un démélange spectral a été effectué (les résultats sont présentés ci-après). De plus, l'initialisation par SIMPLISMA a déjà fait ses preuves à propos de la capacité de la MCR-ALS à démélanger les spectres dans l'étude concernant le suivi du Delipidol® à travers des échantillons d'épiderme humain reconstruit, présentée dans la partie 2.1 de ce Chapitre IV. [105]. Au final, l'utilisation d'une initialisation par SIMPLISMA nous parait comme la plus adaptée à notre objectif de décomposition spectrale des données pour le suivi de la caféine à travers les échantillons d'épiderme.

(i) Sans contrainte d'égalité spectrale sur la caféine

La MCR-ALS est d'abord appliquée en utilisant uniquement les contraintes de non-négativité et de normalisation sur les spectres des composants purs et les concentrations relatives à estimer. La référence spectrale de la caféine n'est pas utilisée pour guider l'algorithme. Pour initier l'algorithme, les spectres de composants purs sont, comme mentionnés ci-dessus, initialisés par la méthode SIMPLISMA. Les spectres purs résultants de la méthode MCR-ALS en utilisant $K = \{2, 3, 5\}$, identifiés par $S_{i=1,\dots,5}$, sont illustrés dans la Figure IV-13 (**A1**-**B1**-**C1**). À noter que leur ordre a été défini de façon à faciliter la lecture et leur comparaison.

Statistiques de modélisation

Les modèles estimés avec $K = \{2, 3, 5\}$ spectres de composants purs s'ajustent aux données expérimentales avec une erreur respective de 6.7503 %, 6.0467 % et 5.4924 % (cf. LOF dans le Tableau IV-3) et expliquent chacun 99.7719 %, 99.8170 % et 99.8491 % (cf. R^2 dans le Tableau IV-3) de leurs variations spectrales. L'augmentation du nombre de composants K implique donc une amélioration des statistiques d'évaluation de reproduction des données. Ainsi, le choix d'un nombre de composants K élevé semble être judicieux pour assurer une modélisation des données la plus satisfaisante possible.

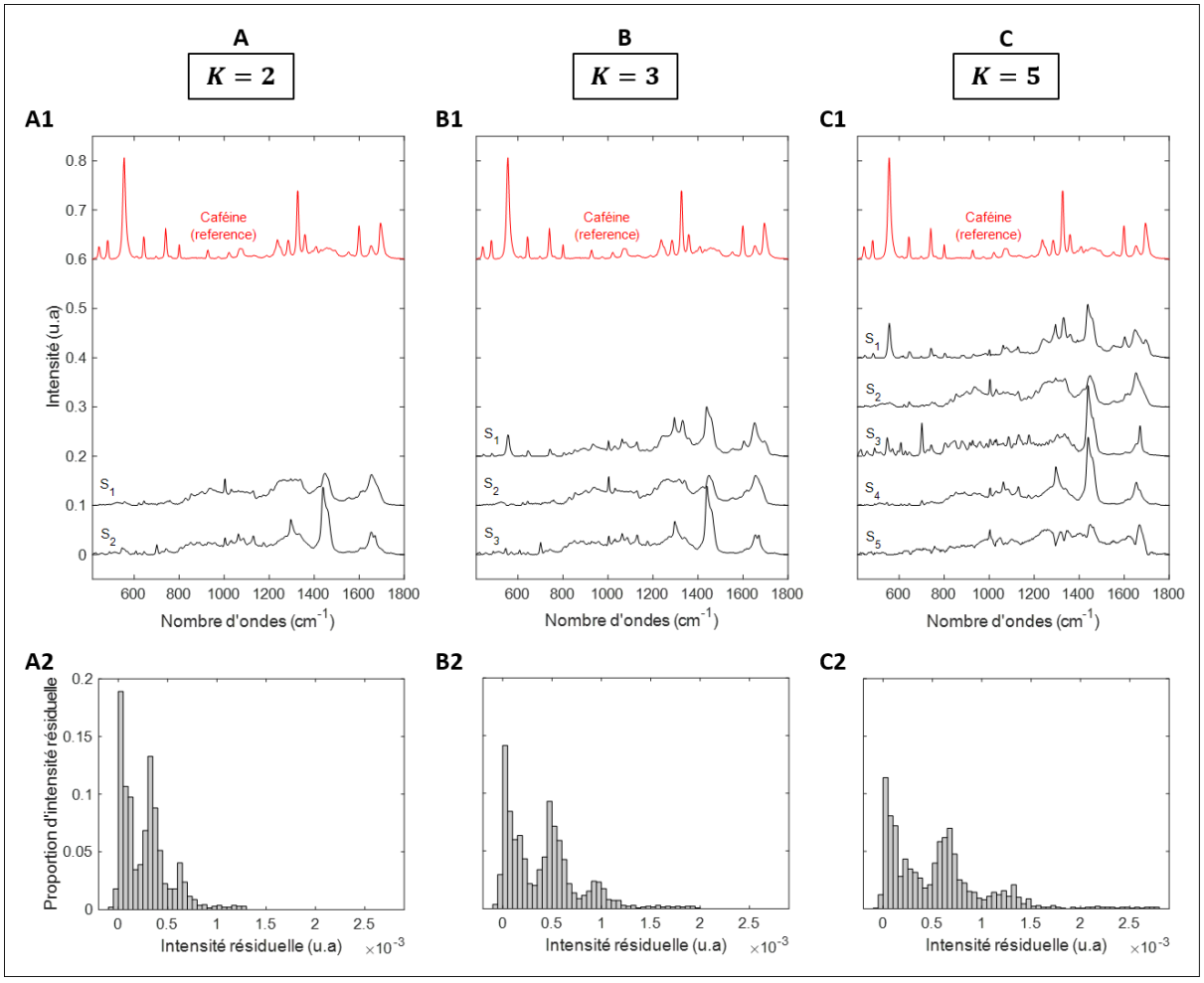


Figure IV-13 – Application de la MCR-ALS avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour (**A**) $K = 2$, (**B**) $K = 3$ *et* (**C**) $K = 5$ *sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures ;* (**A1**-**B1**-**C1**) *Spectres estimés des* K *composants initialisés par SIMPLISMA (décalés en intensité pour plus de clarté) ;* (**A2**-**B2**-**C2**) *Histogrammes des résidus spectraux moyens des modèles obtenus avec les* K *composants estimés illustrés dans* **A1**-**B1**-**C1** *; Les données sont prétraitées selon la stratégie décrite ci-dessus.*

Tableau IV-3 – Mesure des critères du manque d'ajustement LOF (Lack Of Fit en %), de variance expliquée R^2 (en %) et de l'ambiguïté rotationnelle des modèles estimés par MCR-ALS initialisée par SIMPLISMA avec $K = \{2, 3, 5\}$ *composants purs avec (c-CAF) et sans (nc-CAF) contrainte d'égalité spectrale sur la caféine.*

MCR-ALS		$K = 2$		$K = 3$		$K = 5$	
		nc-CAF	**c-CAF**	**nc-CAF**	**c-CAF**	**nc-CAF**	**c-CAF**
LOF		6.7503	10.9197	6.0467	8.5943	5.4924	6.1806
R^2		99.7719	99.4020	99.8170	99.6300	99.8491	99.8088
Ambiguïté rotationnelle	S_1	$1.743 \cdot 10^{-7}$	$-3.664 \cdot 10^{-8}$	$-1.069 \cdot 10^{-2}$	$6.651 \cdot 10^{-2}$	$-2.947 \cdot 10^{-4}$	$5.006 \cdot 10^{-3}$
	S_2	$-1.664 \cdot 10^{-7}$	$-2.318 \cdot 10^{-9}$	$-5.042 \cdot 10^{-2}$	$8.811 \cdot 10^{-5}$	$-6.973 \cdot 10^{-3}$	$-1.865 \cdot 10^{-2}$
	S_3			$3.380 \cdot 10^{-2}$	$-1.494 \cdot 10^{-2}$	$-3.027 \cdot 10^{-3}$	$-1.675 \cdot 10^{-3}$
	S_4					$2.475 \cdot 10^{-2}$	$-2.550 \cdot 10^{-2}$
	S_5					$-2.767 \cdot 10^{-4}$	$1.795 \cdot 10^{-3}$

Interprétation spectrale

Idéalement, il est attendu que les spectres estimés par les modèles comme les signatures Raman des composants purs nécessaires à la reproduction des données mènent à une interprétation biochimique sensée. Dans cette étude, il est espéré que la signature spectrale de la caféine soit

dissociée du reste des variations spectrales et soit considérée comme un composant pur contributif à part entière dans les données expérimentales. Cependant, les modèles estimés avec $K = \{2, 3, 5\}$ ne remplissent pas cette attente.

- Bien que les spectres S_1 et S_2 du modèle estimé avec $K = 2$ composants présentent des bandes spectrales pouvant être associées respectivement aux protéines et aux lipides de la peau, aucun d'eux ne correspond de près à la référence spectrale de la caféine. En effet, leur HQI et leur corrélation par rapport à la caféine valent respectivement 0.2467 et 0.2026 pour S_1 ainsi que 0.1559 et 0.1324 pour S_2 (Tableau IV-4). De ce fait, l'information concernant la caféine n'étant pas correctement capturée par les spectres estimés, son suivi à travers les échantillons par ce modèle est impossible.

- En revanche, le modèle estimé avec $K = 3$ composants est parvenu à intercepter des informations sur la caféine. Tandis que S_2 et S_3 présentent, pareillement au modèle avec $K = 2$ composants, des bandes spectrales pouvant être associées respectivement aux protéines et aux lipides de la peau, des bandes spectrales associées à la caféine sont observables dans S_1. En effet, le HQI et la corrélation entre la caféine et S_1 sont un peu plus élevés (0.3657 et 0.3841) (Tableau IV-4). À l'inverse, S_2 et S_3 ne contiennent aucune information liée à la caféine puisque leur corrélation avec sa référence spectrale est très faible : 0.0950 pour S_2 et 0.0470 pour S_3. S_1 est alors le seul spectre contenant l'information liée à la caféine et devient par conséquent le candidat potentiel pour la suivre dans les données. Cependant, S_1 ne contient pas uniquement les caractéristiques spectrales de la caféine à l'état pur. Des bandes Raman associées à la peau sont également observables et expliquent que sa corrélation avec la caféine soit éloignée de 1. Par conséquent, S_1 ne peut pas être utilisé pour localiser et quantifier la caféine à travers les échantillons d'épiderme, auquel cas un biais serait introduit dans les résultats.

- Quant aux spectres estimés par le modèle avec $K = 5$ composants, leur interprétation est difficile puisque des bandes Raman caractéristiques de la caféine et de la peau se mélangent dans plusieurs spectres estimés. Tandis que le HQI entre les spectres S_2, S_3, S_4 et S_5 et la caféine (0.2464, 0.2072, 0.1546 et 0.2110 ; Tableau IV-4) ne permet pas de dissocier correctement ceux contenant des informations sur la caféine, la corrélation montre que les spectres S_2, S_3 ont plus de points communs avec la signature de la caféine que les spectres S_4, S_5. En effet, les corrélations de S_2 et S_3 avec la caféine sont plus élevées que celles de S_4, S_5 (0.1649 et 0.1141 contre 0.0638 et 0.0535 ; Tableau IV-4) et sont non-négligeables. Néanmoins, ces deux statistiques s'accordent sur le fait que S_1 est le spectre le plus proche de la référence spectrale de la caféine (HQI = 0.4595 et r = 0.5132 ; Tableau IV-4). Mais S_1 ne correspond pas à la signature Raman de la caféine à l'état pur puisque des bandes caractéristiques de la peau sont aussi présentes. Par conséquent, les spectres S_1, S_2 et S_3 ne peuvent pas être utilisés seuls pour suivre la caféine de manière fiable car une partie de l'information la concernant est contenue dans les trois spectres à la fois. Celle-ci ne serait donc pas entièrement prise en compte dans l'estimation de sa contribution. De plus, le suivi de la caféine avec chacun de ces trois spectres n'est également pas possible sans qu'un biais lié à la contribution de la peau ne soit intégré dans l'estimation de sa concentration. Le résultat de cette décomposition spectrale avec $K = 5$ composants illustrent donc une sur-décomposition de l'information liée à la caféine empêchant le suivi de sa progression. Dans ce cas de figure, il est préconisé de revenir à une décomposition spectrale utilisant un nombre de composants purs plus faible pour espérer suivre de manière plus facile et sans biais la caféine.

Tableau IV-4 – Coefficients de corrélation (en noir) et indice de qualité HQI (en gris) entre les spectres estimés $S_{i=1,\dots,5}$ obtenus par MCR-ALS appliquée avec (c-CAF) et sans (nc-CAF) contrainte d'égalité spectrale sur la caféine pour $K = \{2, 3, 5\}$ sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures.

	$K = 2$		$K = 3$			$K = 5$				
	c-CAF									
nc-CAF	S_1 CAF	S_2	S_1 CAF	S_2	S_3	S_1 CAF	S_2	S_3	S_4	S_5
S_1	0.1559 0.2467	0.9401 0.9473	0.3841 0.3657	0.8201 0.8283	0.8368 0.8259	0.5132 0.4595	0.7473 0.7452	0.6772 0.6803	0.7532 0.7366	0.6361 0.6554
S_2	0.1324 0.2026	**0.9593** 0.9552	0.0950 0.2178	**0.9828** 0.9859	0.7232 0.7242	0.1649 0.2464	**0.9986** 0.9987	0.7026 0.7435	0.8043 0.7972	0.8017 0.8465
S_3			0.0470 0.1580	0.6434 0.6839	**0.9964** 0.9935	0.1141 0.2072	0.6773 0.7266	**0.9980** 0.9979	0.9006 0.8902	0.5819 0.6802
S_4						0.0638 0.1546	0.8070 0.8015	0.9116 0.9030	**0.9993** 0.9991	0.6540 0.6738
S_5						0.0535 0.2110	0.8228 0.8590	0.6130 0.6897	0.6513 0.6708	**0.9949** 0.9955
CAF	1.0000 1.0000	0.0801 0.1958	1.0000 1.0000	0.1531 0.2442	0.0662 0.1529	1.0000 1.0000	0.1888 0.2581	0.1168 0.2051	0.0872 0.1654	0.0402 0.2138

Erreurs de reproduction des données

Pour évaluer aux mieux l'ajustement des modèles estimés aux données expérimentales, les histogrammes de leurs résidus moyens sont présentés dans la Figure IV-13 (**A2**-**B2**-**C2**) pour $K = \{2, 3, 5\}$. Chacun d'eux est constitué de trois gaussiennes : une première centrée en zéro attestant de la présence de bruit résiduel et deux autres centrées en des valeurs résiduelles positives témoignant que la totalité des variations spectrales au sein des données n'est pas décrite par les modèles. Théoriquement, l'augmentation du nombre de composants utilisés dans les modèles devrait améliorer l'ajustement aux données, c'est-à-dire réduire de plus en plus les erreurs d'ajustement et les histogrammes des résidus moyens associés devraient tendre vers une unique gaussienne centrée en zéro. De manière inattendue, on observe que plus le nombre de composants utilisés dans le modèle est grand et plus l'intervalle d'erreurs d'ajustement des modèles s'élargit. De manière plus précise, les erreurs associées au bruit résiduel sont moins nombreuses alors que celles associées aux variations spectrales non expliquées par les modèles s'élèvent et sont plus fréquentes.

Évaluation des modèles

L'analyse des différentes statistiques calculées à partir des modèles estimés montrent que le manque d'ajustement (LOF) et la variance expliquée par le modèle (R^2) ne sont pas suffisants pour juger de la qualité du modèle. En effet, l'observation des résidus et des mesures de similarité permet de fournir une analyse plus complète de la pertinence du modèle estimé.

Par exemple, bien que parmi les trois modèles testés le modèle estimé avec $K = 2$ composants soit celui présentant une répartition des résidus plus proche d'un bruit résiduel, son manque d'ajustement global est le plus élevé et les mesures de similarité telles que le HQI et la corrélation ont montré qu'il ne contient aucune information précise sur la caféine au sein des échantillons d'épiderme. Par conséquent, ce modèle n'est pas approprié pour suivre la caféine à travers la peau.

A l'inverse, les statistiques de similarités ont révélé que des spectres de composants purs estimés dans les modèles avec $K = \{3, 5\}$ correspondent en parti avec la référence spectrale de la caféine. À savoir que la corrélation a été plus révélatrice de ces correspondances que la mesure HQI : l'étude des corrélations sera donc préférée dans les analyses suivantes. En effet, des bandes

Raman caractéristiques de la caféine sont présentes dans des spectres de composants purs estimés de ces modèles. Cependant, la caféine n'est pas exprimée à l'état pur dans ces signatures spectrales. Elle est mélangée avec des bandes Raman caractéristiques de la peau et peut même être présente dans plusieurs spectres à la fois pour le modèle avec $K = 5$ composants. Par conséquent, le suivi de la caféine à travers les échantillons d'épiderme humain à l'aide des spectres estimés par ces modèles est compromis en raison d'une probable introduction d'erreur d'estimation de sa contribution.

Au final, ces éléments d'analyse ont montré que l'augmentation du nombre de composants n'améliore pas forcément la qualité du modèle contrairement à ce que le LOF et le R^2 laissent supposer. Il a aussi été montré que cette augmentation n'assure pas non plus la détection de la totalité des composants purs présents dans le mélange, y compris les éventuels ingrédients actifs appliqués en faibles concentrations tels que la caféine. En définitive, l'application de la MCR-ALS sans aucune contrainte d'égalité spectrale sur la caféine ne permet pas de la détecter à l'état pur et par conséquent de suivre sa cinétique de pénétration dans les échantillons d'épiderme.

(ii) Avec contrainte d'égalité spectrale sur la caféine

Les résultats de la MCR-ALS appliquée en utilisant uniquement les contraintes de non-négativité et de normalisation sur les spectres des composants purs et les concentrations relatives à estimer n'ont pas été concluants pour suivre la caféine à travers les échantillons d'épiderme humain reconstruit exposés aux deux formulations pendant 8 et 24 heures. Pour guider l'algorithme dans l'estimation de la contribution de la caféine seule, une contrainte d'égalité spectrale sur la caféine est ajoutée. L'imposition de cette contrainte n'est possible qu'en raison de la disponibilité de sa référence spectrale. De la même manière que précédemment, les spectres de composants purs sont initialisés par la méthode SIMPLISMA. Les spectres purs résultants de la méthode MCR-ALS appliquée avec une contrainte d'égalité spectrale sur la caféine en utilisant $K = \{2, 3, 5\}$, identifiés par $S_{i=1,\dots,5}$, sont illustrés dans la Figure IV-14 (**A1**-**B1**-**C1**). À noter que leur ordre a été défini de façon à faciliter la lecture et leur comparaison.

Statistique de modélisation

Les modèles estimés avec $K = \{2, 3, 5\}$ spectres de composants purs s'ajustent aux données expérimentales avec une erreur respective de 10.9197 %, 8.5943 % et 6.1806 % (cf. LOF dans le Tableau IV-3) et expliquent chacun 99.4020 %, 99.6300 % et 99.8088 % (cf. R^2 dans le Tableau IV-3) de leurs variations spectrales. Avec la contrainte d'égalité spectrale sur la caféine ajoutée dans leur estimation, l'augmentation du nombre de composants K permet une nouvelle fois d'améliorer la capacité des modèles à reproduire la majeure partie des variations spectrales des données expérimentales. En revanche, comme attendu, cette contrainte supplémentaire mène de manière générale à un moins bon ajustement des modèles par rapport aux données. En effet, les estimations du LOF et de la variance expliquée par ces modèles sont respectivement plus élevées et plus faibles que celles associées aux modèles résultants de l'application de la MCR-ALS sans cette contrainte.

De plus, à l'exception du modèle avec $K = 2$ composants où elles sont diminuées d'un ordre de grandeur, les ambigüités rotationnelles estimées témoignant de l'unicité de la décomposition spectrale pour chacun des modèles sont légèrement augmentées par l'imposition de cette contrainte (Tableau IV-3). L'influence de cette contrainte supplémentaire ne remet pas en question la fiabilité des modèles estimés avec $K = \{3, 5\}$ composants en raison de la conservation de faibles valeurs associées aux ambiguïtés rotationnelles présentes. Par ailleurs, les ambiguïtés rotationnelles présentes dans le modèle estimé avec $K = 2$ sont d'ordre 4 à 5 fois inférieures à celles présentes dans les modèles estimés avec $K = \{3, 5\}$ composants qu'une contrainte

spectrale sur la caféine soit imposée ou non. La diminution supplémentaire d'un ordre de grandeur des ambiguïtés rotationnelles lorsqu'une contrainte spectrale sur la caféine est ajoutée au modèle estimé avec $K = 2$ fait de ce modèle le candidat potentiellement le plus approprié pour suivre la caféine à travers l'épiderme.

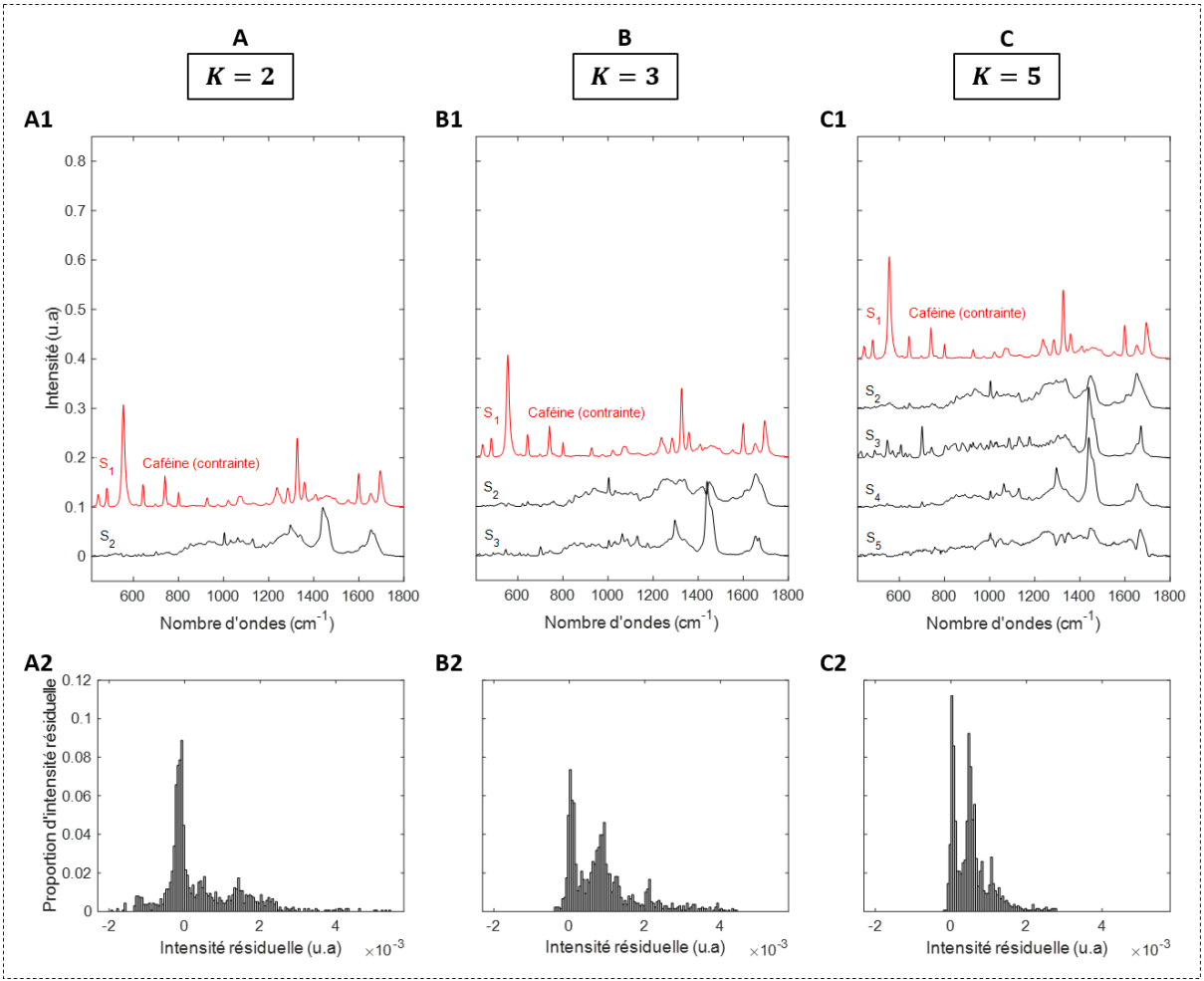


Figure IV-14 – Application de la MCR-ALS avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives et une contrainte d'égalité spectrale supplémentaire sur la caféine pour (**A**) $K = 2$, (**B**) $K = 3$ *et* (**C**) $K = 5$ *sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures ;* (**A1**-**B1**-**C1**) *Spectres estimés des* K *composants initialisés par SIMPLISMA (décalés en intensité pour plus de clarté) ;* (**A2**-**B2**-**C2**) *Histogrammes des résidus spectraux moyens des modèles obtenus avec les* K *composants estimés illustrés dans* **A1**-**B1**-**C1** *; Les données sont prétraitées selon la stratégie décrite ci-dessus.*

Interprétation spectrale

L'introduction d'une contrainte d'égalité spectrale sur la caféine dans l'algorithme de la MCR-ALS avait pour but de dissocier les informations spectrales la concernant du reste des variations spectrales associées aux caractéristiques de la peau et donc de la considérer comme un composant pur contributif à part entière dans les données expérimentales. Ainsi, pour chaque modèle utilisant $K = \{2, 3, 5\}$ composants, le spectre S_1 a été défini comme le spectre référentiel de la caféine et n'a pas été modifié par l'algorithme MCR-ALS.

En revanche, les spectres $S_{i=2,\dots,5}$ initialisés par la méthode SIMPLISMA ont été réévalués au fil des itérations et ont finalement été estimés comme très proches de ceux obtenus lorsqu'aucune contrainte d'égalité spectrale sur la caféine n'avait été imposée. En effet, l'insertion de cette contrainte supplémentaire n'a que très peu modifié l'estimation des spectres de composants purs

de chaque modèle $S_{i=2,..,5}$ en comparaison à ceux estimés sans elle dans les trois autres modèles précédents, comme en attestent leurs fortes corrélations deux à deux (Tableau IV-4). Pour le modèle avec $K = 2$, le spectre S_2 a une corrélation de 0.9593 avec le spectre S_2 du modèle estimé sans contrainte d'égalité spectrale sur la caféine. De même, pour les modèles avec $K = 3$, les corrélations entre les spectres S_2 et S_3 s'élèvent respectivement à 0.9828 et 0.9964. Quant aux spectres $S_{i=2,...,5}$ des modèles avec $K = 5$, leur corrélation deux à deux sont respectivement de 0.9986, 0.9980, 0.9993 et 0.9949. Au final, l'interprétation biochimique des spectres $S_{i=2,...,5}$ estimés dans les modèles utilisant la contrainte d'égalité spectrale est quasi similaire à celle présentée précédemment.

- Dans le modèle estimé avec $K = 2$ composants, alors que S_1 est contraint à être la signature de référence de la caféine, S_2 ne contient aucune bande caractéristique de celle-ci et affiche les bandes Raman d'un échantillon de peau non traitée. En effet, la corrélation entre S_2 et la caféine est très faible : 0.0801 (Tableau IV-4) et est comparable à celle entre la caféine et le spectre moyen de peau non traitée qui est de 0.0875. Par conséquent, le potentiel de ce modèle pour suivre la caféine à travers les échantillons se confirme et ce dernier peut être assuré en observant les concentrations relatives associées au spectre contraint S_1.

- Pour le modèle estimé avec $K = 3$ composants, tandis que S_1 correspond également par contrainte à la caféine, S_2 et S_3 présentent des bandes spectrales pouvant être principalement associées aux protéines et aux lipides de la peau de manière respective. Mais alors que S_3 ne contient aucune information spectrale associée à la caféine comme peut le confirmer sa corrélation avec elle (0.0662), S_2 présente une corrélation non-négligeable avec la caféine de 0.1531 (Tableau IV-4). Cependant, l'observation du spectre S_2 ne permet pas d'identifier de façon distincte des bandes Raman associées à la caféine responsables de cette corrélation non-négligeable. La présence de caféine dans le spectre S_2 doit être très subtile ou bien des bandes Raman communes à la caféine et à d'autres molécules dans le tissu peuvent subsister. Par conséquent, la localisation et la quantification relative de la caféine à travers les échantillons d'épiderme en utilisant S_1 est à interpréter avec prudence puisque les concentrations associées peuvent être potentiellement erronées en raison du fait qu'une partie de la contribution est éventuellement contenue dans S_2.

- Quant aux spectres estimés par le modèle avec $K = 5$ composants, à l'exception de S_1 qui est bien défini comme étant la signature spectrale de la caféine, leur interprétation biochimique reste difficile car des bandes Raman caractéristiques de la caféine et de la peau se mélangent dans plusieurs spectres estimés. En effet, bien que la distinction des bandes Raman qui lui sont associées reste compliquée, les spectres S_2 et S_3 ont une corrélation avec la caféine non-négligeable (0.1888 et 0.1168 ; Tableau IV-4). Quant aux spectres S_4 et S_5, leur corrélation quasi nulle avec la caféine (0.0872 et 0.0402 ; Tableau IV-4), et comparable à celle entre la caféine et le spectre moyen de peau non traitée (0.0875), assure qu'ils ne contiennent aucune information la concernant. Par conséquent, de la même manière que pour le modèle estimé avec $K = 3$ composants, la localisation et la quantification relative de la caféine à travers les échantillons d'épiderme en utilisant S_1 est à interpréter avec prudence puisque les concentrations associées peuvent être potentiellement erronées en raison du fait qu'une partie de la contribution est éventuellement contenue dans S_2 et S_3. En effet, malgré la contrainte d'égalité spectrale qui lui est imposée, S_1 ne peut être utilisé de manière totalement certaine pour suivre la caféine à travers les échantillons d'épiderme puisqu'une partie de sa contribution potentiellement contenue dans les spectres S_2 et S_3 risque de ne pas être prise en considération.

Pour cette application figurant les cas où l'AC est difficilement détectable, ces analyses mènent à considérer comme plus fiables les modèles utilisant la contrainte d'égalité spectrale sur la caféine par rapport à ceux estimés sans cette contrainte. Bien que les modèles utilisant ou non une contrainte d'égalité spectrale sur la caféine présentent des spectres de composants purs très proches, seuls les modèles estimés avec cette contrainte assurent une dissociation de la signature spectrale de la caféine nécessaire pour pouvoir estimer semi-quantitativement au moins une partie de sa contribution dans les échantillons d'épiderme. L'utilisation de la MCR-ALS avec une supervision spectrale s'avère être par conséquent plus appropriée au suivi d'un actif plus difficilement détectable. Cependant, cette contrainte supplémentaire ne garantit pas que toute l'information concernant la caféine soit attribuée uniquement au spectre contraint, ici S_1 : un nombre trop élevé de composants purs peut provoquer une sur-décomposition spectrale des données et des bandes Raman caractéristiques de la caféine peuvent être retrouvées dans plusieurs spectres de composants purs estimés comme c'est le cas pour le modèle avec $K = 5$.

Erreurs de reproduction des données

Pour évaluer aux mieux l'ajustement des modèles estimés en utilisant une contrainte d'égalité spectrale sur la caféine par rapport aux données expérimentales, les histogrammes de leurs résidus moyens sont présentés dans la Figure IV-14 (**A2**-**B2**-**C2**) pour $K = \{2, 3, 5\}$. Chacun d'eux est à nouveau constitué de trois gaussiennes : une première centrée en zéro attestant de la présence de bruit résiduel et deux autres centrées en des valeurs résiduelles positives révélant que de nombreuses variations spectrales au sein des données ne sont pas décrites par les modèles estimés.

De manière prévisible, l'ajout de la contrainte d'égalité spectrale sur la caféine dans l'algorithme de la MCR-ALS a occasionné des erreurs de modélisation plus élevées que celles obtenues sans imposer cette contrainte : la plage d'erreurs d'ajustement s'est élargie puisqu'elle est passée de [-0.0002 0.003] u.a à [-0.0025 0.006] u.a. Néanmoins, cette fois-ci l'augmentation du nombre de composants utilisés pour reproduire au mieux les données améliore l'ajustement des modèles aux données puisque les plages d'erreurs se réduisent au fur et à mesure que ce nombre devient grand (Figure IV-14 (**A2**-**B2**-**C2**)).

Cependant, l'augmentation du nombre de composants ne garantit pas que les variations spectrales non expliquées par les modèles correspondent uniquement au bruit expérimental associé aux mesures Raman qui est représenté théoriquement par une seule gaussienne. En effet, le modèle avec $K = 5$ composants fournit certes des erreurs globalement plus faibles que les modèles avec $K = \{2, 3\}$ composants, mais les trois gaussiennes de l'histogramme des résidus se distinguent chacune par des effectifs plus importants. En ce qui concerne le modèle avec $K = 3$ composants, deux gaussiennes parmi les trois se distinguent particulièrement : une première centrée en zéro d'écart-type faible et une seconde décalée dans les résidus positifs d'écart-type supérieur et d'effectifs cumulés plus élevés. Cela indique que les erreurs associées au bruit résiduel sont moins nombreuses que celles associées aux variations spectrales non expliquées par le modèle. En revanche, le modèle avec $K = 2$ composants présente plus d'erreurs d'ajustement associées au bruit résiduel que celles associées aux autres variations spectrales non expliquées par le modèle : la première gaussienne centrée en zéro d'écart-type faible a des effectifs plus importants que les deux autres gaussiennes à l'allure plus aplatie. Par conséquent, le modèle avec $K = 2$ composants semble être le plus fidèle aux données expérimentales et donc le plus approprié pour définir les cinétiques de pénétration de la caféine à travers les échantillons d'épiderme.

Évaluation des modèles

De la même manière que pour l'étude des modèles estimés sans contrainte d'égalité spectrale, l'étude de l'ensemble des différentes statistiques calculées telles que les résidus et les corrélations mais aussi des bandes Raman présentes dans chacun des spectres de composants purs estimés est nécessaire pour analyser de façon pertinente les décompositions spectrales obtenues par MCR-ALS : les critères LOF et R^2 ont été à nouveau considérés comme insuffisant pour cette analyse.

L'ajout d'une contrainte d'égalité spectrale sur la caféine dans l'algorithme MCR-ALS a permis d'isoler l'information la concernant dans les modèles utilisant $K = \{2, 3\}$ spectres de composants purs pour reconstruire les données. Malgré une mise en garde sur la validité des concentrations relatives estimées de la caféine obtenues par le modèle avec $K = 3$, l'absence de bandes caractéristiques de la caféine bien distinctes dans les deux autres spectres non contraints rend ce modèle tout aussi légitime que celui utilisant $K = 2$ composants pour établir les cinétiques de pénétration de la caféine à travers les échantillons d'épiderme exposés aux deux formulations pendant 8 et 24 heures. Par ailleurs, une fois de plus, l'augmentation du nombre de composants ($K = 5$) pour modéliser les données expérimentales n'a pas conduit à une meilleure décomposition spectrale des composants de la peau et isolation de la caféine. À l'inverse, elle a donné lieu à une sur-décomposition spectrale de la caféine qui toutefois n'est observable que par l'analyse des corrélations. Le suivi précis de la caféine dans l'épiderme est alors encore plus incertain que pour le modèle estimé avec $K = 3$ composants.

(iii) Suivi de la caféine à l'aide des modèles avec $K = \{2, 3, 5\}$ et une contrainte d'égalité spectrale

Les modèles résultants de l'application de la MCR-ALS utilisant $K = \{2, 3\}$ et une contrainte d'égalité spectrale sur la caféine ont été précédemment jugés comme les plus appropriés au suivi de la caféine dans les données expérimentales. Pour chacun de ces modèles, les concentrations relatives associées au spectre S_1 contraint à correspondre à la référence spectrale de la caféine ont été reconstituées sous forme de cartes de distribution comme illustré dans la Figure IV-15. Par volonté de mettre en évidence la pertinence de ces deux modèles, les cartes de distributions associées au spectre contraint à correspondre à la caféine S_1 dans le modèle estimé avec $K = 5$ composés sont également illustrés pour comparaison (Figure IV-15). De manière générale, les contributions relatives de la caféine pour l'ensemble des quatre images Raman acquises sont plus faibles dans le modèle utilisant $K = 5$ composants purs. Les concentrations relatives estimées par le modèle avec $K = 3$ composants sont certes supérieures à celles estimées dans le modèle avec $K = 5$ composants, mais elles restent malgré tout inférieures aux contributions relatives estimées dans le modèle n'utilisant que $K = 2$ composants purs. Ces différences étaient prévisibles au vu des mises en garde faites sur la possibilité que l'information concernant la caféine ne soit pas totalement contenue dans l'expression du spectre S_1 des modèles avec $K = \{3,5\}$. Ainsi, nous concluons une fois de plus que le modèle estimé à l'aide d'une supervision spectrale sur la caféine avec $K = 2$ composants s'avère être le plus approprié pour suivre la caféine dans cette étude de cas où la détection de l'actif est difficile.

Quant au comportement de la caféine estimée par le modèle avec $K = 2$, sa diffusion à travers les différents niveaux de profondeur cutanée est latéralement homogène quels que soient la formulation appliquée et le temps d'exposition. Toutefois, sa contribution relative diffère légèrement selon les conditions expérimentales soumises à l'échantillon. En effet, la caféine progresse légèrement dans les couches plus profondes de l'épiderme au fil du temps quelle que soit la formulation appliquée. Puis après 24 heures d'exposition, les quantités de caféine dans

l'épiderme sont plus élevées lorsqu'elle est formulée par HF. Les mêmes observations peuvent être faites sur les cartes de distribution de la caféine estimée par le modèle avec $K = 3$ composants. En effet, malgré une différence d'intensité notable par rapport aux résultats obtenus avec le modèle utilisant $K = 2$ composants, après 24 heures d'exposition la caféine pénètre de manière homogène plus profondément dans l'épiderme quelle que soit la formulation utilisée et est retrouvée en plus grande quantité dans le tissu cutané lorsqu'elle a été formulée par HF. Ces informations sont plus facilement lisibles sur les profils de pénétrations obtenus en calculant les concentrations moyennes par niveau de profondeur à partir des cartes de distribution (Figure IV-16).

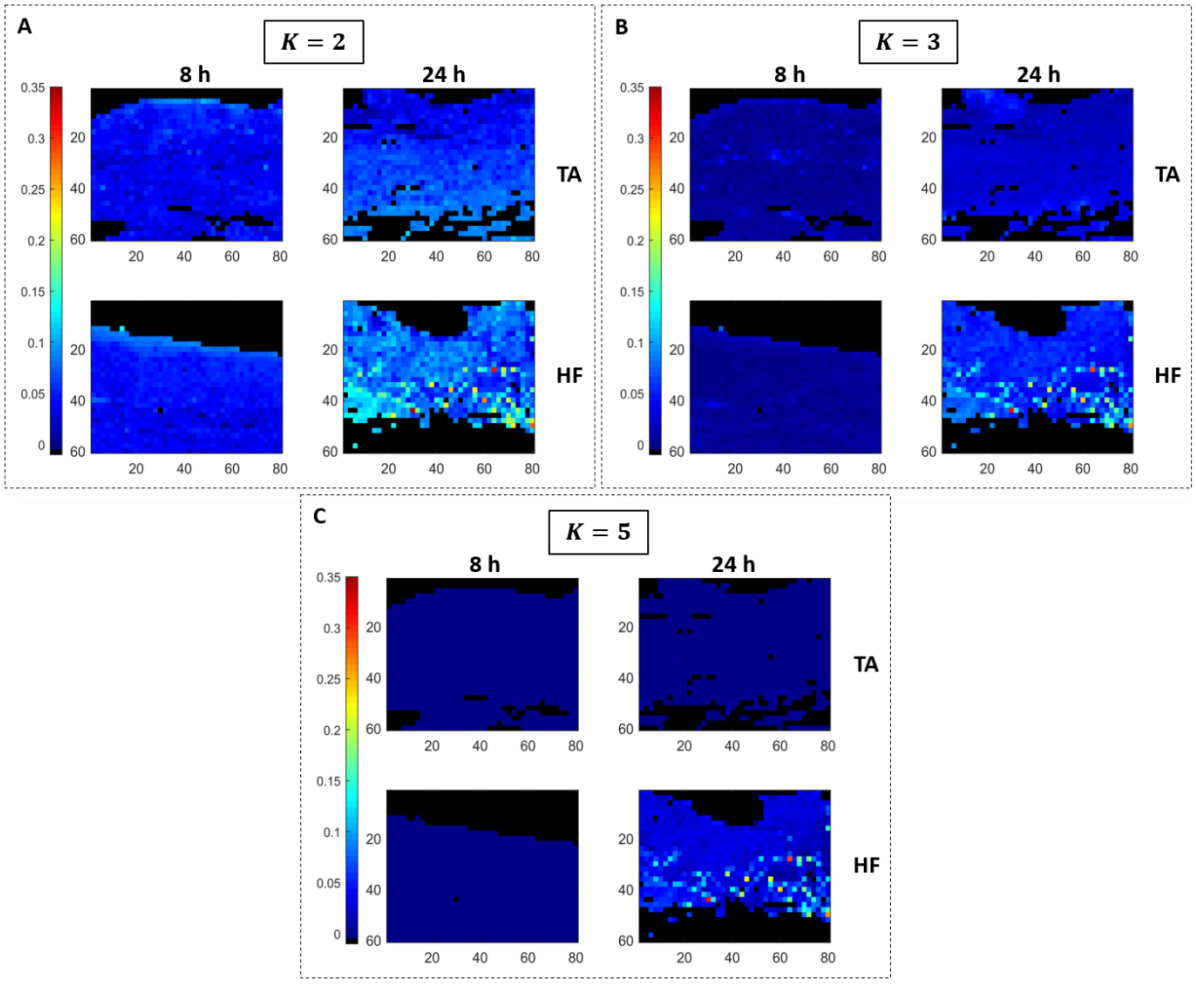


Figure IV-15 – Cartes de distribution de la caféine formulée avec tensio-actifs (TA) et avec la technologie Hydrofeeling® (HF) après 8 et 24 heures d'exposition. Cartes reconstruites à partir des composants S_1, obtenus à l'aide d'une MCR-ALS avec (**A**) $K = 2$, (**B**) $K = 3$ *et (C)* $K = 5$ *composants et en utilisant une contrainte d'égalité spectrale sur la caféine. Les pixels noirs représentent le substrat et les axes x et y des cartes de distribution sont exprimées en micromètres.*

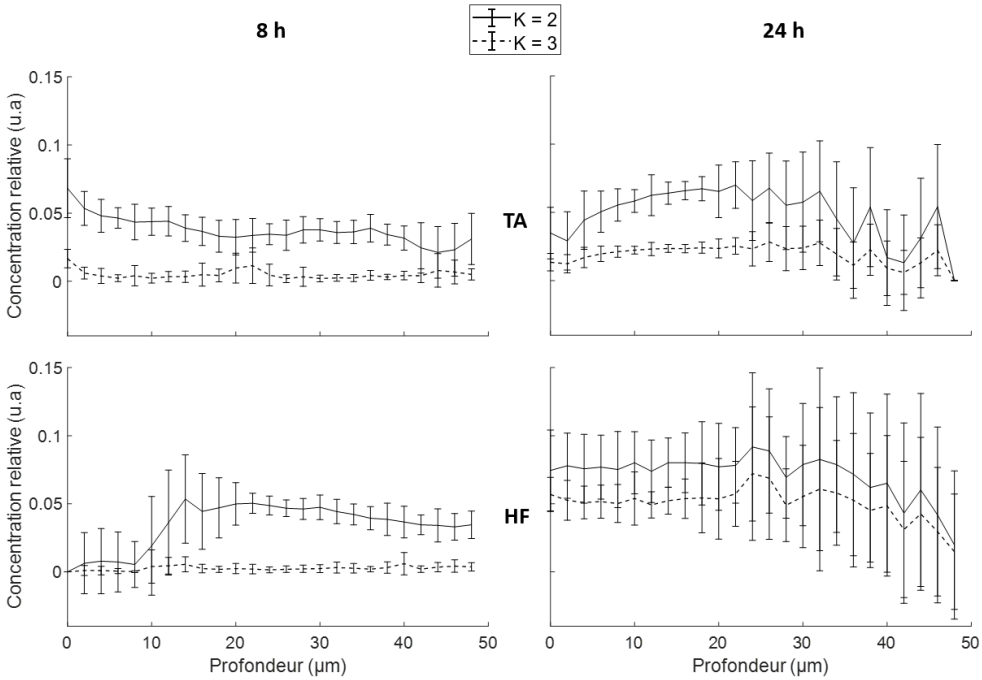


Figure IV-16 – Profils de pénétration moyens de la caféine, formulée avec tensio-actifs (TA) et avec la technologie Hydrofeeling® (HF) après 8 et 24 heures d'exposition. Profils obtenus à partir des concentrations $C_{i=1,\dots,K}$, obtenues à l'aide d'une MCR-ALS avec $K = 2$ (courbe pleine) et $K = 3$ (courbe pointillée) composants et en utilisant une contrainte d'égalité spectrale sur la caféine.

2.2.4.3. Application de la MVC-NMF

La MVC-NMF est une méthode reconnue comme capable de réaliser une décomposition spectrale plus contraignante que la méthode MCR-ALS. Elle consiste à estimer un modèle utilisant des spectres de composants purs de façon à ce qu'un compromis entre une reproduction fidèle des données originales et une proximité maximale de ces spectres par rapport aux données soit trouvé. La MVC-NMF est donc utilisée sur ce même jeu de données pour évaluer les éventuelles différences engendrées par la contrainte de volume minimal supplémentaire dans les modèles estimés. Les performances de la MVC-NMF appliquée sans et avec contrainte d'égalité spectrale sur la caféine pour suivre sa progression selon les deux formulations dans l'épiderme sont alors confrontées à celles de la MCR-ALS.

(i) Sans contrainte d'égalité spectrale sur la caféine

La MVC-NMF est d'abord appliquée en utilisant uniquement les contraintes de non-négativité et de normalisation sur les spectres des composants purs et les concentrations relatives à estimer. La référence spectrale de la caféine n'est pas utilisée pour guider l'algorithme. Pour initier l'algorithme, les spectres de composants purs sont initialisés par la méthode SIMPLISMA comme ils l'ont été pour l'application de la MCR-ALS. Le nombre de composants ayant déjà été estimé à $K = \{2, 3, 5\}$ précédemment, le paramètre τ contrôlant le compromis entre l'ajustement du modèle estimé aux données et la contrainte de volume minimal formé par les spectres de composants purs reste à déterminer. Trois τ différents ont été utilisés dans l'estimation des modèles : un paramètre privilégiant la minimisation de l'erreur d'ajustement aux données ($\tau = 0.25$), un autre permettant de trouver un équilibre entre le terme d'attache aux données et la contrainte de volume minimal ($\tau = 0.5$) et un dernier favorisant la contrainte de volume minimal formé par les spectres de composants purs estimés ($\tau = 0.75$). Au total, neufs modèles ont été estimés. Ces choix se justifient par la volonté d'évaluer l'influence du paramètre τ sur l'estimation des modèles.

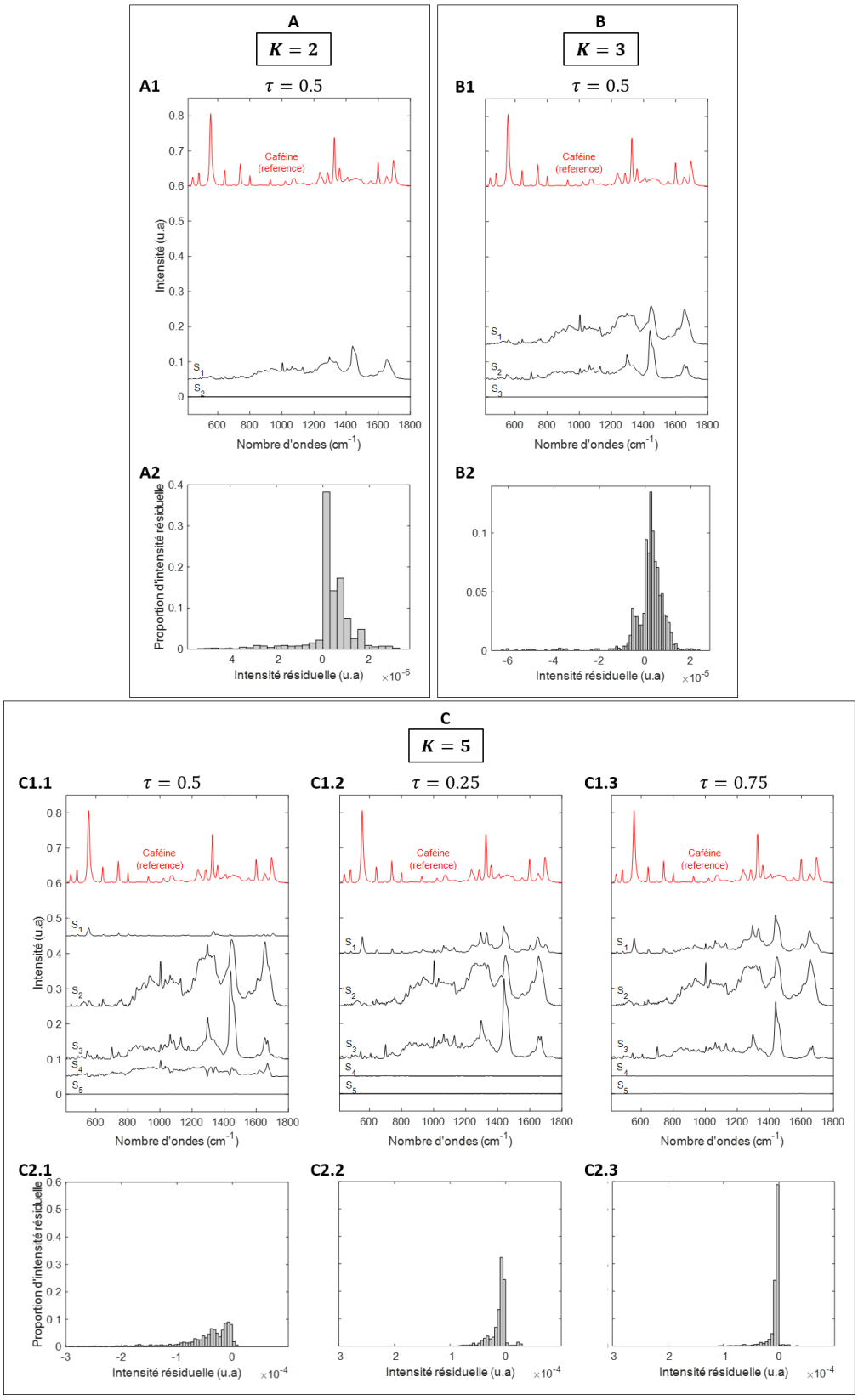


*Figure IV-17 – Application de la MVC-NMF avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour (**A**)* $K = 2$, $\tau = 0.5$, *(**B**)* $K = 3$, $\tau = 0.5$ *et (**C**)* $K = 5$, $\tau = \{0.25, 0.5, 0.75\}$ *sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures; (**A1**-**B1**-**C1**) Spectres estimés des* K *composants initialisés par SIMPLISMA (décalés en intensité pour plus de clarté); (**A2**-**B2**-**C2**) Histogrammes des résidus spectraux*

*moyens des modèles obtenus avec les K composants estimés illustrés dans **A1**-**B1**-**C1**; Les données sont prétraitées selon la stratégie décrite ci-dessus.*

En définitive, le paramètre τ n'a pas eu d'impact sur l'estimation des modèles utilisant $K = \{2, 3\}$ composants puisque la MVC-NMF a fourni une décomposition spectrale très similaire quel que soit le paramètre utilisé. En revanche, les décompositions spectrales obtenues avec $K = 5$ composants diffèrent selon le paramètre τ utilisé. Les spectres purs résultants de la méthode MVC-NMF en utilisant $K = \{2, 3\}$ avec $\tau = 0.5$ et $K = 5$ avec $\tau = \{0.25, 0.5, 075\}$ sont illustrés dans la Figure IV-17 (**A1**-**B1**-**C1**). À noter que leur ordre a été défini de façon à faciliter la lecture et leur comparaison.

Statistique de modélisation

Les modèles estimés avec $K = \{2, 3, 5\}$ spectres de composants purs s'ajustent aux données expérimentales avec une erreur respective de 9.7888 %, 6.4783 % et inférieure à 5.40 % (cf. LOF dans le Tableau IV-5) et expliquent chacun 99.5197 %, 99.7899 % et plus de 99.85 % (cf. R^2 dans le Tableau IV-5) de leurs variations spectrales. L'augmentation du nombre de composants K implique donc une amélioration des statistiques d'évaluation de reproduction des données. Cependant, en comparaison avec les valeurs obtenues par les modèles estimés par MCR-ALS sous les mêmes contraintes, les statistiques obtenues pour les modèles avec $K = \{2, 3\}$ sont plus mauvaises. Toutefois, celles obtenues pour les modèles avec $K = 5$ sont meilleures quel que soit τ.

Tableau IV-5 – Mesure des critères du manque d'ajustement LOF (Lack Of Fit en %), de variance expliquée R^2 (en %) et de l'ambiguïté rotationnelle des modèles estimés par MVC-NMF initialisée par SIMPLISMA avec $K = \{2, 3, 5\}$ composants purs avec (c-CAF) et sans (nc-CAF) contrainte d'égalité spectrale sur la caféine.

MVC-NMF		$K = 2$		$K = 3$		$K = 5$	
	τ	**nc-CAF**	**c-CAF**	**nc-CAF**	**c-CAF**	**nc-CAF**	**c-CAF**
LOF	0.25	48.9997	46.2340	6.4786	9.3979	5.3887	5.0350
	0.5	9.7888	46.2340	6.4783	9.3979	5.3188	5.0206
	0.75	48.9996	46.2340	6.4813	9.3979	5.3781	5.0560
R^2	0.25	87.1724	88.6703	99.7899	99.5574	99.8547	99.8732
	0.5	99.5197	88.6703	99.7899	99.5574	99.8585	99.8739
	0.75	87.1724	88.6703	99.7897	99.5574	99.8553	99.8721
Ambiguïté rotationnelle	0.25	$1.975 \cdot 10^{-1}$	$4.353 \cdot 10^{-9}$	$-4.203 \cdot 10^{-1}$	$7.928 \cdot 10^{-3}$	$2.434 \cdot 10^{-2}$	$-1.178 \cdot 10^{-5}$
		$3.953 \cdot 10^{-1}$	$-1.551 \cdot 10^{-9}$	$-3.528 \cdot 10^{-2}$	$3.546 \cdot 10^{-3}$	$4.562 \cdot 10^{-3}$	$-2.882 \cdot 10^{-2}$
				$1.949 \cdot 10^{-4}$	$9.005 \cdot 10^{-2}$	$9.618 \cdot 10^{-2}$	$8.582 \cdot 10^{-2}$
						$-1.965 \cdot 10^{-2}$	$-1.539 \cdot 10^{-3}$
						$-2.734 \cdot 10^{-2}$	$1.437 \cdot 10^{-6}$
	0.5	$3.887 \cdot 10^{-3}$	$4.362 \cdot 10^{-9}$	$-5.225 \cdot 10^{-3}$	$2.880 \cdot 10^{-6}$	$2.023 \cdot 10^{-2}$	$1.506 \cdot 10^{-2}$
		$2.858 \cdot 10^{-3}$	$-1.556 \cdot 10^{-9}$	$-2.367 \cdot 10^{-2}$	$6.121 \cdot 10^{-6}$	$9.300 \cdot 10^{-2}$	$2.709 \cdot 10^{-2}$
				$-2.117 \cdot 10^{-10}$	$-1.335 \cdot 10^{-6}$	$1.919 \cdot 10^{-2}$	$-2.701 \cdot 10^{-3}$
						$-7.728 \cdot 10^{-2}$	$3.203 \cdot 10^{-2}$
						$3.301 \cdot 10^{-3}$	$1.347 \cdot 10^{-3}$
	0.75	$1.903 \cdot 10^{-1}$	$4.378 \cdot 10^{-9}$	$-9.12 \cdot 10^{-8}$	$3.138 \cdot 10^{-6}$	$7.213 \cdot 10^{-3}$	$2.065 \cdot 10^{-2}$
		$3.857 \cdot 10^{-1}$	$-1.569 \cdot 10^{-9}$	$1.114 \cdot 10^{-7}$	$-6.371 \cdot 10^{-7}$	$1.438 \cdot 10^{-2}$	$1.512 \cdot 10^{-2}$
				$4.116 \cdot 10^{-9}$	$-8.571 \cdot 10^{-6}$	$-3.246 \cdot 10^{-2}$	$6.551 \cdot 10^{-3}$
						$3.892 \cdot 10^{-3}$	$-7.971 \cdot 10^{-2}$
						$3.892 \cdot 10^{-3}$	$1.348 \cdot 10^{-1}$

Interprétation spectrale

De manière générale, à la différence de la MCR-ALS, la MVC-NMF considère qu'un spectre quasi nul correspondant au substrat des échantillons préparés et analysés est un composant pur à part entière des données. En effet, au moins un spectre quasi nul est estimé comme composant pur

pour chaque modèle quel que soit le nombre de composants et le paramètre τ utilisés (Figure IV-17 (**A1**-**B1**-**C1.1**-**C1.2**-**C1.3**)). En revanche si plus d'un spectre estimé est quasi nul, cela signifie que le nombre de composants à estimer est jugé trop grand pour reproduire les données par la MVC-NMF. À noter également qu'aucun des modèles n'a dissocié la signature spectrale de la caféine comme composant pur contributif à part entière du reste des variations spectrales associés au tissu cutané et au substrat.

- Dans le modèle estimé avec $K = 2$ composants et $\tau = 0.5$, alors que S_2 est un spectre quasi nul pouvant être interprété comme correspondant au substrat des échantillons, le spectre S_1 est quant à lui principalement composé des bandes caractéristiques de la peau (Figure IV-17 (**A1**)). Il représente plus spécifiquement le spectre moyen des données expérimentales. En effet, des variations spectrales associées à celles de la caféine y sont également dissimulées et justifient de sa corrélation non-négligeable de 0.1522 (Tableau IV-6) avec la référence spectrale de la caféine. Bien que cette décomposition spectrale soit différente de celle obtenue par MCR-ALS dans les mêmes conditions, l'information concernant la caféine n'est pas non plus correctement isolée du reste des informations spectrales concernant le tissu cutané. Par conséquent, à l'instar de celui proposé par MCR-ALS, ce modèle n'est pas adéquat pour le suivi de la progression de la caféine à travers l'épiderme.

- Pour le modèle estimé avec $K = 3$ composants et $\tau = 0.5$, deux des spectres qui le composent (S_1 et S_2) sont proches des spectres S_2 et S_3 estimés par la MCR-ALS dans les mêmes conditions et proches des spectres S_1 et S_2 estimés par la MCR-ALS avec $K = 2$ composants. En revanche, le spectre S_1 obtenu par MCR-ALS avec $K = 3$ exprimant des bandes associées à la caféine et à la peau a laissé place avec la MVC-NMF à un spectre quasiment nul S_3. Ce dernier est alors associé au substrat des échantillons tandis que S_1 et S_2 présentent des bandes spectrales associées respectivement aux protéines et aux lipides de la peau. Cependant, ces derniers présentent en plus une corrélation avec la référence spectrale de la caféine non négligeable : 0.1563 et 0.1278 (Tableau IV-6). Ces corrélations révèlent que des informations la concernant sont localisables dans chacune de leur expression. Par conséquent, le mélange d'informations sur la caféine et le tissu cutané rend à nouveau délicat le suivi de la caféine seule à l'aide de la MVC-NMF utilisant $K = 3$ composants.

- Quant aux spectres des modèles à $K = 5$ composants, ceux obtenus avec les paramètres $\tau = \{0.25, 0.75\}$ s'apparentent à ceux obtenus dans le modèle estimé par MCR-ALS dans les mêmes conditions avec seulement $K = 3$ composants et auxquels deux autres spectres quasiment nuls ont été estimés en supplément. En effet, pour ces deux modèles, S_1 comporte de nombreuses bandes caractéristiques de la caféine bien distinctes, S_2 et S_3 présentent des bandes spectrales associées respectivement aux protéines et aux lipides de la peau et S_4 et S_5 peuvent être associés aux spectres correspondant au substrat des échantillons. Cependant, malgré une corrélation conséquente de 0.5057 pour $\tau = 0.25$ et 0.3737 pour $\tau = 0.75$ (Tableau IV-6) avec elle, S_1 ne correspond pas à la caféine à l'état pur puisque de nombreuses bandes spectrales caractéristiques de la peau sont également bien perceptibles. De même, S_2 ne comporte pas uniquement des informations liées à la composition protéique de la peau : sa corrélation avec la caféine s'élève à 0.1113 pour $\tau = 0.25$ et 0.1114 pour $\tau = 0.75$ (Tableau IV-6). Les modèles utilisant $\tau = \{0.25, 0.75\}$ sur-décomposent donc tous deux l'information relative à la caféine.
Pour leur part, les spectres estimés en utilisant $\tau = 0.5$ n'ont pas de correspondance probante avec les spectres estimés par MCR-ALS dans les mêmes conditions. Comme chaque modèle estimé par MVC-NMF, un spectre presque nul S_5 est identifié comme composant pur à part entière et est assimilé au substrat des échantillons. Puis, S_2, S_3 et S_4 présentent chacun des bandes spectrales

spécifiques au tissu cutané. Alors que S_3 et S_4 semblent totalement dépourvus d'informations relatives à la caféine (corrélations de -0.1458 et 0.0965 ; Tableau IV-6), les bandes Raman spécifiques à la peau de S_2 recouvrent discrètement des bandes caractéristiques de la caféine. En effet, sa corrélation avec la caféine atteint 0.1770 (Tableau IV-6). D'autre part, une grande partie de l'information concernant la caféine est exprimée par S_1 puisque son coefficient de corrélation s'élève à 0.8169 (Tableau IV-6). Cependant, la discordance de S_1 avec la signature spectrale de la caféine à l'état pur et la présence de caféine dans différents spectres de composants purs ne permet pas une fois de plus de suivre de manière précise la caféine à travers l'épiderme. Au final, bien qu'un équilibre entre une erreur d'ajustement aux données et un volume minimal permet d'exprimer une plus grande partie des bandes spectrales de la caféine dans un spectre de composant pur, chacun des trois modèles estimés avec $K = 5$ composants et $\tau = \{0.25, 0.5, 0.75\}$ n'est pas approprié pour l'estimation des cinétiques de pénétration de la caféine sans biais dans le tissu cutané.

Tableau IV-6 – Coefficients de corrélation (en noir) et indice de qualité HQI (en gris) entre les spectres estimés $S_{i=1,\dots,5}$ obtenus par MVC-NMF appliquée avec (c-CAF) et sans (nc-CAF) contrainte d'égalité spectrale sur la caféine pour $K = \{2, 3, 5\}$ sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures.

		$K = 2$		$K = 3$			$K = 5$				
		c-CAF									
nc-CAF	τ	S_1 CAF	S_2	S_1 CAF	S_2	S_3	S_1 CAF	S_2	S_3	S_4	S_5
S_1	0.25	0.1511 0.2377	0.9967 0.9966	0.1558 0.2467	0.9566 0.9617	-0.0001 0.1761	0.5057 0.4404	0.8867 0.8605	0.7205 0.6890	0.5912 0.5849	0.0982 0.0878
	0.5	0.1522 0.2380	0.9966 0.9966	0.1563 0.2463	0.9699 0.9730	0.0265 0.1841	0.8169 0.7340	-0.0623 0.0975	0.0130 0.1056	0.0506 0.0936	-0.1194 0.0090
	0.75	0.1525 0.2380	0.9965 0.9966	0.1563 0.2462	0.9717 0.9746	0.0416 0.1985	0.3737 0.3544	0.7940 0.7999	0.9337 0.9228	0.3837 0.4990	0.2780 0.1900
S_2	0.25	0.1533 0.2383	0.9964 0.9965	0.1317 0.2011	0.9408 0.9365	0.4999 0.4609	0.1113 0.2243	0.9373 0.9301	0.7008 0.7166	0.9841 0.9861	0.1301 0.1320
	0.5	0.1595 0.1680	0.3793 0.3880	0.1278 0.1949	0.9271 0.9202	0.5139 0.4675	0.1770 0.2442	0.9667 0.9670	0.7300 0.7553	0.8682 0.8338	0.0307 0.1100
	0.75	0.1519 0.2381	0.9966 0.9965	0.1264 0.1925	0.9217 0.9136	0.5301 0.4871	0.1114 0.2247	0.9823 0.9850	0.8181 0.7948	0.4343 0.6039	0.2125 0.1644
S_3	0.25			0.3038 0.3084	-0.2745 0.1435	0.2406 0.2725	0.0385 0.1516	0.8558 0.8533	0.9947 0.9938	0.6955 0.7234	0.0986 0.1025
	0.5			0.1389 0.1482	-0.0719 0.1262	-0.0302 0.0710	0.0965 0.1590	0.6215 0.6255	0.9767 0.9625	0.9542 0.9429	0.0480 0.0907
	0.75			0.2723 0.2425	-0.0939 0.1351	0.0183 0.1049	0.0078 0.1216	0.5339 0.5673	0.9461 0.9369	0.7640 0.7505	0.1742 01272
S_4	0.25						0.1243 0.1738	-0.3853 0.0463	-0.1812 0.0965	-0.3512 0.1106	-0.2611 0.0018
	0.5						-0.1458 0.1419	0.7747 0.8399	0.5583 0.6511	0.3798 0.4475	0.0854 0.1585
	0.75						0.3156 0.2518	0.2628 0.2839	0.0801 0.1408	0.2890 0.3055	-0.1315 0.0061
S_5	0.25						0.0077 0.0639	0.5083 0.4152	0.3136 0.2646	0.3448 0.3046	0.3219 0.1904
	0.5						-0.0535 0.0381	0.3287 0.2780	0.2975 0.2496	0.3417 0.2598	0.2505 0.1532
	0.75						-0.1147 0.0717	-0.3172 0.1059	-0.3152 0.0470	-0.2114 0.1599	-0.1203 0.0209
CAF	0.25	1.0000 1.0000	0.0726 0.1913	1.0000 1.0000	0.0955 0.2048	0.0865 0.1252	1.0000 1.0000	0.1892 0.2367	0.0389 0.1458	0.0764 0.2178	0.0136 0.0440
	0.5	1.0000 1.0000	0.0726 0.1913	1.0000 1.0000	0.0955 0.2048	0.1106 0.1357	1.0000 1.0000	0.1086 0.2245	0.0465 0.1656	0.1518 0.1808	0.0145 0.0578
	0.75	1.0000 1.0000	0.0726 0.1913	1.0000 1.0000	0.0955 0.2048	0.0989 0.1336	1.0000 1.0000	0.1342 0.2313	0.1044 0.1666	0.0022 0.1782	0.0953 0.0724

En définitive, semblablement à la MCR-ALS, l'application de la MVC-NMF sans contrainte d'égalité spectrale sur la caféine ne permet pas de dissocier correctement sa signature spectrale à l'état pur dans cette application où l'AC est difficile à détecter.

Erreurs de reproduction des données

Les histogrammes des résidus moyens associés aux modèles estimés par MVC-NMF sans contrainte d'égalité spectrale sur la caféine avec $K = \{2, 3, 5\}$ composants et $\tau = \{0.25, 0.55, 0.75\}$ sont illustrés dans la Figure IV-17 (**A2**-**B2**-**C2**). De manière générale, les résidus obtenus pour chacun de ces modèles sont distribués sous forme d'une seule gaussienne plus ou moins aplatie et sont 10 à 10^3 plus faibles que ceux obtenus par la MCR-ALS dans les mêmes conditions. Ainsi, de par cette observation, la MVC-NMF semble fournir des modèles décrivant mieux les variations spectrales exprimées par les données expérimentales en ne laissant inexpliquées que des variations relatives à du bruit résiduel. La qualité des cartes de distributions devrait alors être meilleure et les estimations des concentrations devraient être plus précises. Cependant, l'augmentation du nombre de composants engendre une surexpression des variations spectrales de ces données : des erreurs d'ajustements négatives deviennent plus fréquentes et sont multipliées par 10 voire 100. Par conséquent, les estimations des concentrations relatives risquent d'être encore erronées, mais cette fois-ci elles seraient revues à la hausse.
Quant à l'impact du paramètre τ (gérant le compromis entre l'ajustement du modèle aux données et la contrainte de volume minimal) sur les résidus moyens, on observe que la meilleure distribution des variations résiduelles pour un modèle avec $K = 5$ composants est obtenue avec $\tau = 0.75$. Les résidus obtenus avec $\tau = \{0.25, 0.75\}$ forment des gaussiennes centrées en zéro plus étroites que celle formée par les résidus obtenus avec $\tau = 0.5$ mais l'effectif des erreurs proches de 0 pour $\tau = 0.75$ est plus important que celui pour $\tau = 0.25$. Toutefois, avec des erreurs résiduelles de l'ordre de 10^{-6}, le modèle le plus fidèle aux données expérimentales en terme de fidélité est le modèle obtenu avec $K = 2$ composants et $\tau = 0.5$.

Évaluation des modèles

L'analyse des différentes statistiques calculées à partir des modèles estimés mène à des conclusions différentes quant à la pertinence de ces modèles. L'étude de la distribution des variations résiduelles non expliquées par les modèles suggère que le modèle utilisant $K = 2$ composants et $\tau = 0.5$ serait celui décrivant le mieux les données expérimentales alors que le modèle utilisant $K = 5$ composants et $\tau = 0.5$ serait celui qui s'ajusterait le moins bien en terme de fidélité aux données. À l'inverse les estimations du LOF et du R^2 montrent que le modèle utilisant $K = 5$ composants et $\tau = 0.5$ modéliserait globalement mieux les variations spectrales des données expérimentales que le modèle utilisant $K = 2$ composants et $\tau = 0.5$.

Toutefois, l'interprétation spectrale des modèles reste l'analyse la plus rationnelle pour évaluer le sens biochimique de la décomposition spectrale obtenue et s'assurer que l'estimation des contributions de chaque composant de l'échantillon soit recevable. Cette interprétation a permis de juger la MVC-NMF comme incapable de dissocier ne serait-ce qu'une partie de l'information sur la caféine à l'aide de $K = \{2, 3\}$ composants. En effet, aucun des spectres estimés de ces modèles ne présente de manière très distincte les bandes Raman caractéristiques de la caféine, ces dernières restent masquées derrières les bandes Raman spécifiques à la peau. Par conséquent, le suivi de la caféine à l'aide de ces modèles est impossible et ceci même de manière partielle. Par ailleurs, les modèles estimés par MVC-NMF avec $K = 5$ composants et $\tau = \{0.25, 0.5, 0.75\}$ ont réussi à dissocier une partie de l'information relative à la caféine dans un de leurs spectres de composants purs estimés S_1. Cependant, cette séparation n'est pas optimale puisque les spectres S_1 représentent chacun un mélange de bandes associées à la fois à la caféine et à la peau. De plus,

une seconde partie de l'information recherchée sur la caféine est contenue dans l'expression des autres spectres estimés par ces modèles. Par conséquent, le suivi de la caféine de manière précise par ces modèles n'est pas fiable non plus puisque des erreurs d'estimation de ses contributions locales seraient probablement introduites.

Au final, malgré la génération de résidus beaucoup plus faibles, la MVC-NMF appliquée sans contrainte d'égalité spectrale sur la caféine n'est pas plus efficace que la MCR-ALS appliquée dans les mêmes conditions pour dissocier la contribution de la caféine de celle des différents composants de l'épiderme dans cette étude.

(ii) Avec contrainte d'égalité spectrale sur la caféine

Les résultats de la MVC-NMF appliquée en utilisant uniquement les contraintes de non-négativité et de normalisation spectrale sur les spectres de composants purs et les contributions relatives à estimer n'ont pas permis de détecter efficacement la signature spectrale de la caféine parmi les variations spectrales spécifiques à la peau. Afin de guider l'algorithme MVC-NMF à estimer les concentrations semi-quantitatives de la caféine à travers les échantillons d'épiderme humain reconstruit exposés aux deux formulations pendant 8 et 24 heures, une contrainte d'égalité spectrale sur la caféine est ajoutée. De la même manière que précédemment, les spectres de composants purs sont initialisés par la méthode SIMPLISMA et la méthode MVC-NMF a été appliquée avec les mêmes paramètres $K = \{2, 3, 5\}$ et $\tau = \{0.25, 0.5, 0.75\}$. Bien que différents paramètres τ aient été utilisés, les décompositions spectrales des données réalisées avec $K = \{2, 3\}$ sont indépendamment pratiquement identiques. Par conséquent, seuls les spectres estimés par la méthode MVC-NMF avec $\tau = 0.5$ sont illustrés pour $K = \{2, 3\}$ composants dans la Figure IV-18 (**A1**-**B1**), alors que ceux obtenus en utilisant $\tau = \{0.25, 0.5, 0.75\}$ pour $K = 5$ sont tous présentés dans la Figure IV-18 (**C1**)). À noter que leur ordre a été défini de façon à faciliter la lecture et leur comparaison.

Statistique de modélisation

Les modèles estimés avec $K = \{2, 3, 5\}$ spectres de composants purs s'ajustent aux données expérimentales avec une erreur respective de 46.2340 %, 9.3979 % et inférieure à 5.06 % (cf. LOF dans le Tableau IV-5) et expliquent chacun 88.6703 %, 99.5574 % et plus de 99.87 % (cf. R^2 dans le Tableau IV-5) de leurs variations spectrales. À nouveau, l'augmentation du nombre de composants K permet d'améliorer la capacité des modèles à reproduire la majeure partie des variations spectrales des données expérimentales. Bien que l'ajout d'une contrainte d'égalité spectrale sur la caféine peut théoriquement générer de moins bonnes statistiques d'ajustement global dans l'estimation des modèles en comparaison à celles obtenues sans cette contrainte (modèles estimés avec $K = \{2, 3\}$ composants), une amélioration de ces critères peut également apparaitre (modèles estimés avec $K = 5$ composants et $\tau = \{0.25, 0.5, 0.75\}$). Les critères LOF et R^2 suggèrent même que le modèle reproduisant au mieux les données expérimentales est celui obtenu avec $K = 5$ composants, $\tau = 0.5$ et une contrainte d'égalité spectrale sur la caféine. Par ailleurs, les estimations du LOF et de la variance expliquée par les modèles obtenus avec $K = \{2, 3\}$ et $K = \{5\}$ sont respectivement plus élevées et plus faibles que celles associées aux modèles résultants de l'application de la MCR-ALS en utilisant cette contrainte. Cette dernière observation suggère ainsi que l'application de la MVC-NMF pourrait être plus pertinente pour suivre de manière précise la caféine à travers l'épiderme.

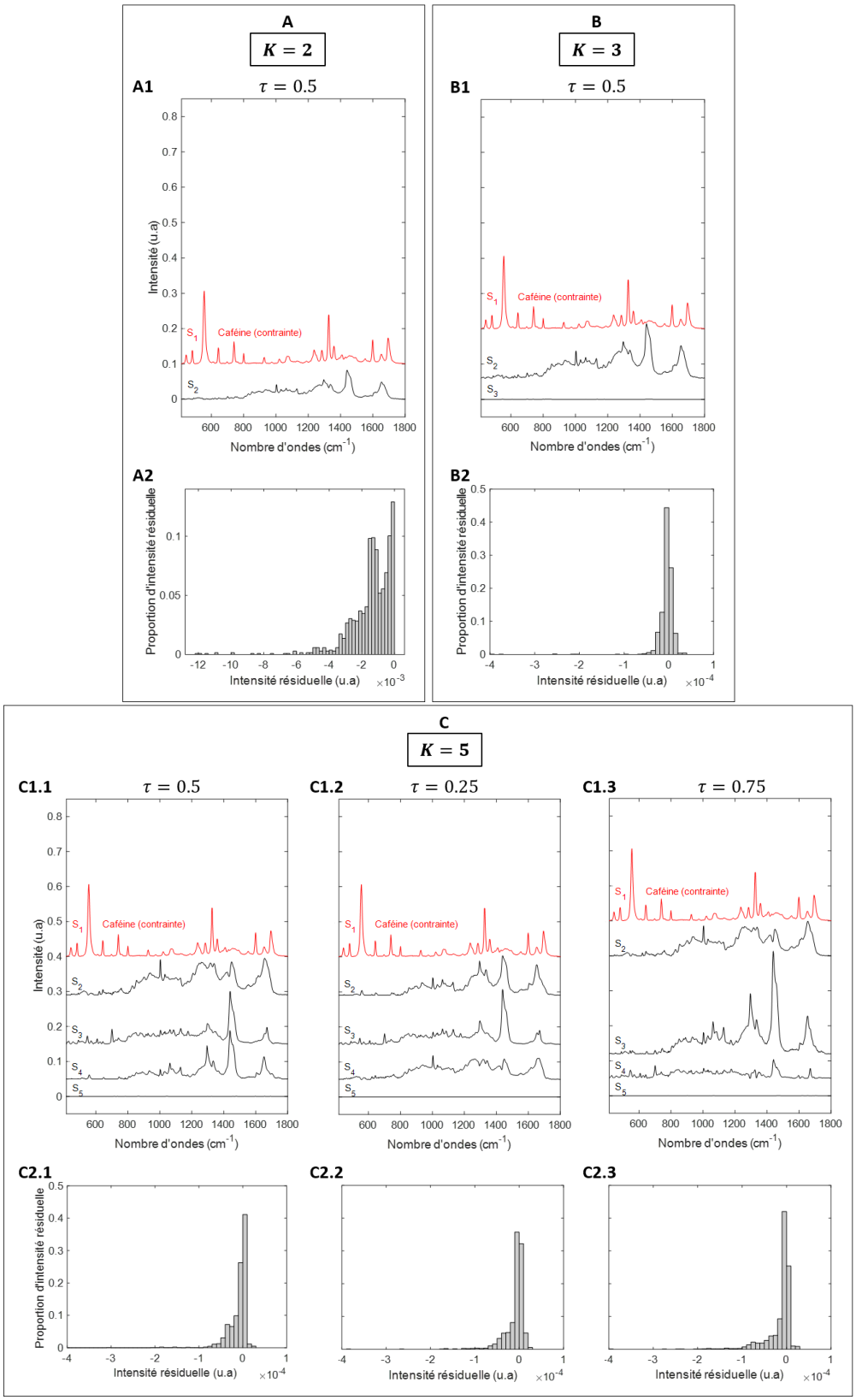


*Figure IV-18 – Application de la MVC-NMF avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives et une contrainte d'égalité spectrale supplémentaire sur la caféine pour (**A**) $K = 2,\ \tau = 0.5$, (**B**) $K = 3,\ \tau = 0.5$ et (**C**) $K = 5,\ \tau = \{0.25, 0.5, 0.75\}$ sur les images Raman des échantillons de peau exposés aux deux formulations de caféine pendant 8 et 24 heures ; (**A1**-**B1**-**C1**) Spectres estimés des K composants initialisés par SIMPLISMA (décalés en intensité pour plus de clarté) ; (**A2**-**B2**-**C2**) Histogrammes des résidus spectraux moyens des modèles obtenus avec les K composants estimés illustrés dans **A1**-**B1**-**C1** ; Les données sont prétraitées selon la stratégie décrite ci-dessus.*

D'autre part, les ambiguïtés rotationnelles estimées pour les modèles avec $K = 2$ composants sont fortement diminuées par l'ajout d'une contrainte d'égalité spectrale sur la caféine quel que soit τ (Tableau IV-5). Cette nette amélioration suggère que les décompositions spectrales ainsi obtenues sont très certainement uniques. Pour ce qui est de la fiabilité des modèles obtenus avec $K = \{3,5\}$, tandis qu'elles sont quelque peu réduites à l'ajout d'une contrainte d'égalité spectrale lorsque $\tau = \{0.25, 0.5\}$, les ambiguïtés rotationnelles sont à l'inverse globalement augmentées d'un à trois ordre de grandeur lorsque $\tau = 0.75$. Toutefois, la plupart des modèles estimés présentent des ambigüités de rotation faibles et d'ordre inférieure à 10^{-2}. Par conséquent, excepté pour les modèles obtenus sans contrainte d'égalité spectrale avec $K = 2, \tau = \{0.25, 0.75\}$ et pour qui les ambigüités rotationnelles sont de l'ordre de 10^{-1}, l'unicité des décompositions spectrales n'est pas remise en question.

Interprétation spectrale

De la même manière que pour l'application de la MCR-ALS sur ce jeu de données, une contrainte d'égalité spectrale sur la caféine a été introduite dans l'algorithme MVC-NMF dans le but de dissocier l'information relative à la caféine de celle des autres composants de la peau et de la considérer comme un composant pur contributif à part entière. Ainsi, pour chacune des estimations, à chaque itération le spectre S_1 a été fixé comme étant la signature spectrale de la caféine à l'état pur pendant que les autres spectres de composants purs $S_{i=2,\dots,5}$ étaient ajustés.

À l'inverse de la MCR-ALS où les spectres $S_{i=2,\dots,5}$ obtenus avec et sans contrainte d'égalité spectrale sur la caféine pour S_1 ont été estimés comme très similaires, ceux estimés par MVC-NMF avec une contrainte d'égalité spectrale sur la caféine pour S_1 diffèrent de ceux obtenus sans appliquer cette contrainte supplémentaire. En effet, tous modèles confondus, la moitié des coefficients de corrélation deux à deux est supérieure à 0.92 tandis que l'autre moitié est comprise entre -0.35 et 0.38 (Tableau IV-6). Par conséquent, l'interprétation de ces nouvelles décompositions est différente de celle des décompositions obtenues sans contrainte d'égalité spectrale supplémentaire.

- Dans le modèle estimé avec $K = 2$ composants et $\tau = 0.5$, alors que S_1 est contraint à être la signature de référence de la caféine, S_2 correspond quant à lui au spectre moyen des données contrôle. Cette assimilation est confirmée par sa corrélation très élevée avec le spectre moyen d'un échantillon d'épiderme non traitée : 0.9826. De plus, S_2 ne contient aucune information concernant la caféine comme en atteste sa faible corrélation de 0.0726 (Tableau IV-6). Ces éléments d'analyse rappellent ceux rapportés dans l'analyse du modèle obtenu par MCR-ALS en utilisant $K = 2$ composants et la contrainte d'égalité spectrale sur la caféine. En effet, les spectres S_2 sont très proches comme en témoigne leur très forte corrélation de 0.9973. Par conséquent, le suivi de la caféine de manière précise à travers les échantillons d'épiderme peut être réalisé en étudiant les contributions relatives associées au spectre contraint S_1 de ce modèle.

- Dans le modèle estimé avec $K = 3$ composants et $\tau = 0.5$, alors que S_1 correspond également par contrainte à la référence spectrale de la caféine, S_2 est très proche du spectre S_2 estimé par le modèle précédent obtenu par MVC-NMF avec $K = 2$ composants et $\tau = 0.5$ (corrélation s'élevant à 0.9995). En effet, de manière similaire, S_2 présente des bandes spectrales caractéristiques d'un échantillon d'épiderme non traité (corrélation de 0.9823) et ne contient aucune spécificité spectrale associée à la caféine comme en atteste sa faible corrélation de 0.0955 avec elle (Tableau IV-6). Pour ce qui est du 3ème et dernier spectre S_3 estimé dans ce modèle, ses valeurs spectrales presque nulles le lient directement au spectre de composant pur représentant le substrat des échantillons. Au final d'un point de vue théorique, l'ajout de ce dernier spectre S_3 aux deux autres spectres S_1 et S_2 ayant un sens biochimique bien distinct ne devrait pas perturber

l'estimation des concentrations relatives de la caféine à l'aide du spectre S_1. Par conséquent, ce modèle semble être bien approprié au suivi des cinétiques de pénétration de la caféine dans le tissu cutané de manière fiable.

- Quant aux spectres des modèles à $K = 5$ composants, ceux obtenus avec les paramètres $\tau = \{0.5, 0.75\}$ présentent chacun deux spectres composés à la fois de bandes Raman caractéristiques de la peau et de la caféine, un spectre présentant de nombreuses bandes spectrales spécifiques à la peau et un spectre presque nul pouvant être associé au substrat des échantillons. En effet, les spectres S_2 et S_4 du modèle obtenu avec $\tau = 0.5$ et les spectres S_2 et S_3 obtenu avec $\tau = 0.75$ ont une corrélation non-négligeable avec la caféine : respectivement 0.1086 et 0.1518 pour $\tau = 0.5$ et 0.1342 et 0.1044 pour $\tau = 0.75$ (Tableau IV-6). Pour leur part, les spectres S_3 obtenu avec $\tau = 0.5$ et S_4 obtenu avec $\tau = 0.75$ sont tous deux dépourvus d'informations concernant la caféine comme en atteste leur faible corrélation avec elle (respectivement 0.0465 et 0.0022 ; Tableau IV-6). Cependant, la présence des caractéristiques spectrales de la caféine dans deux spectres estimés autres que le spectre S_1 contraint à exprimer sa signature Raman à l'état pur indique que ces deux modèles ne sont pas appropriés pour suivre la progression de la caféine à travers l'épiderme sans que des erreurs d'estimation de ses contributions locales soient introduites.
En ce qui concerne les spectres du modèle estimé avec $K = 5$ composants et $\tau = 0.25$, S_1 est de nouveau défini comme le spectre de la caféine pure, S_2 présente un mélange de bandes associées à la fois à la peau et à la caféine (corrélation de 0.1892 avec la caféine ; Tableau IV-6), S_3 et S_4 exposent chacun des bandes exclusivement associées à la peau comme en résultent leur faible corrélation avec la caféine (respectivement 0.0389 et 0.0764 ; Tableau IV-6) et enfin S_5 est un spectre presque nul qui correspond au composant pur associé au substrat des échantillons. Une fois de plus, la sur-décomposition de la caféine dans les spectres S_1 et S_2 rend l'utilisation de ce modèle pour estimer les contributions relatives à la caféine de manière fiable impossible.

En définitive, pareillement à la méthode MCR-ALS, ces analyses mènent à considérer les modèles obtenus en appliquant la MVC-NMF avec $K = \{2, 3\}, \tau = 0.5$ et une contrainte d'égalité spectrale sur la caféine comme les plus fiables d'un point de vue spectral pour pouvoir estimer semi-quantitativement l'intégralité de la contribution de la caféine. En effet, à l'inverse des modèles obtenus avec $K = 5$ composants, ces modèles ne sur-décomposent pas la signature spectrale de la caféine, celle-ci est bien dissociée des autres variations spectrales de l'épiderme évitant ainsi l'introduction d'erreur d'estimation de sa contribution.

Erreurs de reproduction des données

Les histogrammes des résidus moyens associés aux modèles estimés par MVC-NMF avec une contrainte d'égalité spectrale sur la caféine avec $K = \{2, 3, 5\}$ composants et $\tau = \{0.25, 0.5, 0.75\}$ sont illustrés dans la Figure IV-18 (**A2**-**B2**-**C2**). De manière générale, les résidus obtenus pour chacun de ces modèles sont distribués sous forme d'une seule gaussienne.

De manière non surprenante, l'ajout d'une contrainte d'égalité spectrale a conduit à de moins bons ajustements des modèles estimés en terme d'erreurs de modélisation résiduelles. Par exemple, en comparaison aux erreurs d'ajustements des modèles obtenus par MVC-NMF sans cette contrainte, les erreurs de modélisation résiduelles obtenues par les modèles utilisant $K = \{2, 3\}$ composants et $\tau = 0.5$ sont respectivement 10^3 et 10 fois plus élevées. Néanmoins, le modèle estimé avec $K = 3$ composants semble modéliser plus de variations spectrales présentes dans les données expérimentales que le modèle estimé avec $K = 2$ composants. En effet, ses erreurs résiduelles sont distribuées sous la forme d'une gaussienne centrée en zéro d'écart-type plus de 10 fois plus faible pouvant être assimilée à du bruit résiduel. L'intégration du spectre

quasiment nul S_3 est probablement l'explication d'une meilleure reproduction des variations spectrales des données expérimentales par ce modèle en comparaison avec le modèle n'utilisant que $K = 2$ composants.
Quant aux modèles obtenus avec $K = 5$ composants et $\tau = \{0.25, 0.5, 0.75\}$, leurs erreurs résiduelles sont du même ordre de grandeur, à savoir 10^{-4}, que celles obtenues sans la contrainte d'égalité. Toutefois, l'évolution de la distribution des erreurs résiduelles avec l'ajout de la contrainte d'égalité spectrale sur la caféine dans l'algorithme de la MVC-NMF est dépendante du paramètre τ. En effet, alors qu'une nette amélioration est observable pour le modèle utilisant $\tau = 0.5$ avec une augmentation du nombre d'erreurs proches de zéro et une diminution de la surexpression des variations spectrales des données, seule une légère amélioration est perceptible pour le modèle utilisant $\tau = 0.25$ dont l'histogramme est sous forme de gaussienne plus étroite. En revanche, les erreurs d'ajustements associées au bruit résiduel obtenues lorsque $\tau = 0.75$ sont bien moins nombreuses que lorsqu'aucune contrainte d'égalité spectrale n'est imposée attestant d'une moins bonne reproduction des données.

Par comparaison avec les modèles estimés dans les mêmes conditions par la MCR-ALS, la MVC-NMF appliquée avec une contrainte d'égalité spectrale sur la caféine propose des modèles reproduisant mieux les données expérimentales. En effet, la distribution des erreurs résiduelles d'ajustement sous forme d'une seule gaussienne révèle qu'aucune variation spectrale associée à un composant particulièrement très contributif dans les données n'est délaissée par les modèles. D'autre part, les résidus obtenus par MVC-NMF sont plus faibles que ceux obtenus par MCR-ALS. Par exemple, pour le modèle utilisant $K = 2$ composants, l'intervalle d'erreurs résiduelles [-2 ; 4] $\cdot 10^{-3}$ obtenues par MCR-ALS est décalé sur la plage d'erreurs [-8 ; 0] $\cdot 10^{-3}$. L'augmentation du nombre de résidus négatifs montre que contrairement à la MCR-ALS qui délaisse la modélisation de certaines variations spectrales dans les données, la MVC-NMF les reproduit en totalité et a même tendance à les sur-exprimer. Quant aux modèles obtenus avec $K = \{3\}$ composants et $\tau = 0.5$ puis $K = 5$ composants et $\tau = \{0.25, 0.5, 0.75\}$, les erreurs résiduelles sont divisées par 10 par rapport à celles obtenues par MCR-ALS.

Évaluation des modèles

À l'instar de la MCR-ALS, l'ajout d'une contrainte d'égalité spectrale sur la caféine dans l'algorithme MVC-NMF a permis d'isoler l'information la concernant dans les modèles utilisant $K = \{2, 3\}$ spectres de composants purs et $\tau = 0.5$ pour reconstruire les données. Globalement, le modèle estimé avec $K = 2$ composants a des statistiques d'ajustements de données et des erreurs résiduelles moins avantageuses que celui estimé avec $K = 3$ composants. Néanmoins, l'interprétation spectrale de ces deux modèles affirment que ces deux modèles semblent être pertinents pour suivre la progression de la caféine à travers les échantillons d'épiderme exposés aux deux formulations pendant 8 et 24 heures. Par ailleurs, bien qu'une fois de plus l'augmentation du nombre de composants K semblait améliorer statistiquement les décompositions spectrales en terme d'ajustement de données et de variance expliquée, elle n'a à nouveau pas mené à de meilleures décompositions spectrales. Au contraire, elle a de nouveau conduit à une sur-décomposition spectrale de la caféine quel que soit le paramètre τ utilisé rendant ainsi impossible le suivi précis de la caféine dans l'épiderme.

(iii) Suivi de la caféine à l'aide des modèles avec $K = \{2, 3\}$ et une contrainte d'égalité spectrale

Les modèles résultants de l'application de la MVC-NMF utilisant $K = \{2, 3\}$ et $\tau = 0.5$ et une contrainte d'égalité spectrale sur la caféine ont été précédemment jugés comme les plus appropriés au suivi de la caféine dans les données expérimentales. Pour chacun de ces modèles,

les concentrations relatives associées au spectre S_1 contraint à correspondre à la référence spectrale de la caféine ont été reconstituées sous forme de cartes de distribution qui sont illustrées dans la Figure IV-19.

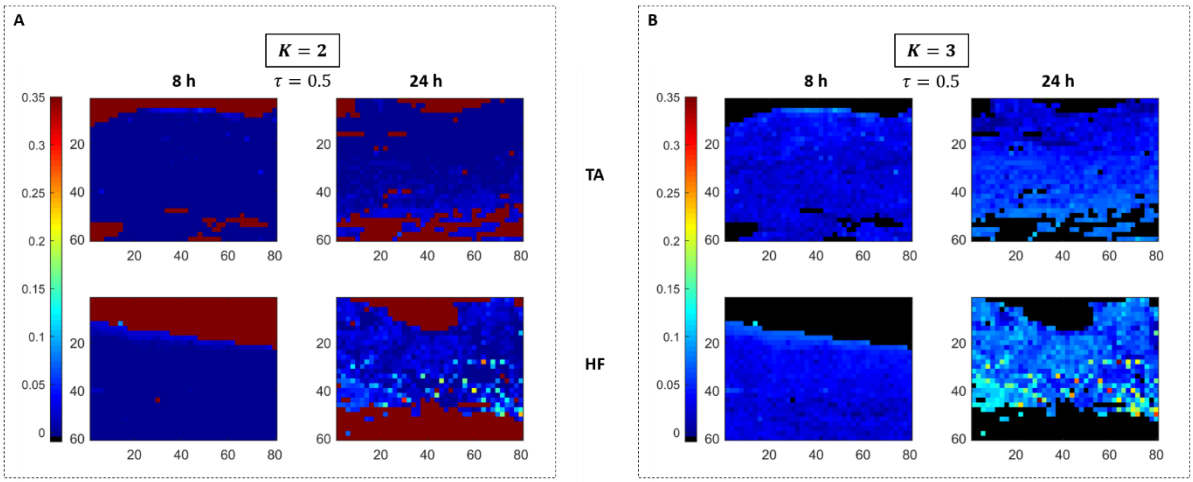


*Figure IV-19 – Cartes de distribution de la caféine formulée avec tensio-actifs (TA) et avec la technologie Hydrofeeling® (HF) après 8 et 24 heures d'exposition. Cartes reconstruites à partir des composants $S_{i=1,\dots,K}$, obtenus à l'aide d'une MVC-NMF avec (**A**) $K = 2, \tau = 0.5$ et (**B**) $K = 3, \tau = 0.5$ composants et en utilisant une contrainte d'égalité spectrale sur la caféine. Les axes x et y des cartes de distribution sont exprimées en micromètres.*

- Pour le modèle utilisant $K = 2$ composants, l'interprétation des cartes de distribution de la caféine est problématique. Les véritables contributions de la caféine sont masquées par celles des spectres correspondant au substrat des échantillons d'épiderme qui n'ont pas de spectre de composant pur pour les représenter dans ce modèle et qui s'expriment alors à travers le spectre de la caféine. Cette observation souligne alors l'importance de la présence du spectre presque nul représentant le substrat des échantillons dans la modélisation des données par MVC-NMF. Par conséquent, l'information relative de la caféine n'étant pas correctement dissociée de l'information concernant le substrat des échantillons d'épiderme, ces cartes de distributions ne sont pas exploitables pour évaluer les cinétiques de sa pénétration.

- Quant aux cartes de distribution de la caféine reconstituées à partir du modèle obtenu avec $\tau = 0.5$ et $K = 3$ composants purs représentant chacun de manière bien distincte l'information spectrale relative au tissu cutané, au substrat des échantillons et à la caféine, une distribution latérale homogène de la caféine est observée quels que soient la formulation et le temps d'exposition appliqué. Les concentrations relatives estimées ne dépassent pas 0.35 u.a.. Cependant, selon les conditions expérimentales soumises aux échantillons d'épiderme, la répartition de la caféine n'est pas exactement la même. En effet, que la caféine soit formulée avec des TA ou bien formulée par HF, la caféine progresse légèrement dans les couches plus profondes de l'épiderme au fil du temps. En revanche après 24 heures d'exposition, les quantités de caféine dans les couches inférieures de l'épiderme sont plus élevées lorsqu'elle est formulée par HF.

Les observations faites sur le modèle estimé par MVC-NMF avec $K = 3$ composants et une contrainte d'égalité spectrale sur la caféine font écho à celles faites sur les cartes de distributions de la caféine obtenues à partir du modèle estimé par MCR-ALS avec $K = 2$ composants avec cette même contrainte. En effet, la comparaison des profils de pénétrations obtenus en calculant les concentrations moyennes par niveau de profondeur à partir des cartes de distribution de la caféine de ces deux modèles (Figure IV-20) montre qu'ils estiment des cinétiques de pénétration de la caféine à travers la peau très proches. Plus précisément, malgré une différence de

concentration globale de 0.01 u.a., la caféine pénètre de manière homogène plus profondément dans l'épiderme après 24 heures d'exposition quelle que soit la formulation utilisée et elle est retrouvée en plus grande quantité dans le tissu cutané lorsqu'elle a été formulée par HF.

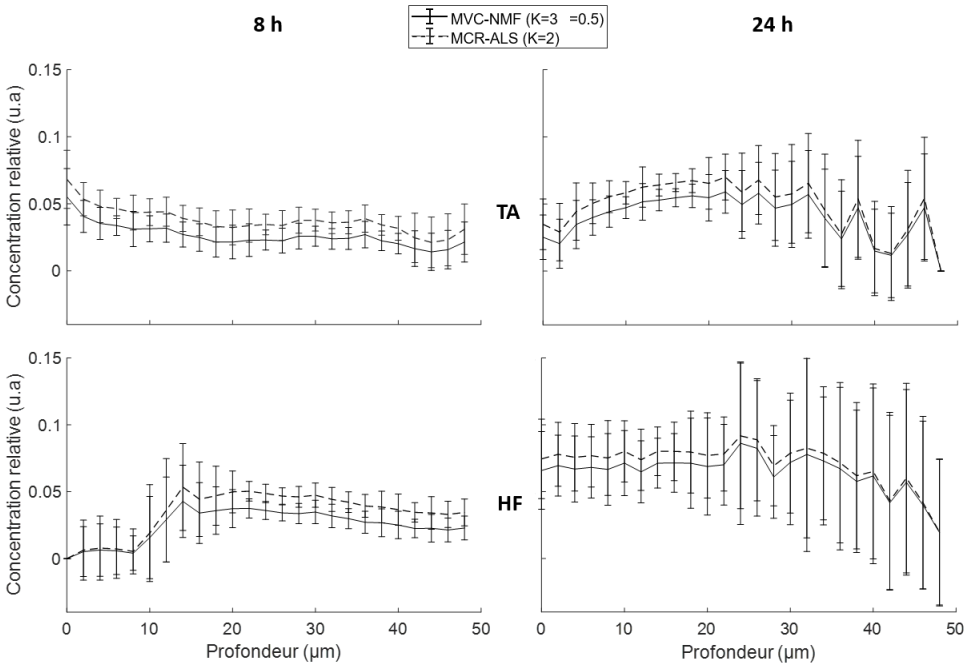


Figure IV-20 – Profils de pénétration moyens de la caféine, formulée avec tensio-actifs (TA) et avec la technologie Hydrofeeling® (HF) après 8 (gauche) et 24 heures (droite) d'exposition. Profils obtenus à partir des concentrations $C_{i=1,\dots,K}$, obtenues à l'aide d'une MVC-NMF avec $K = 3, \tau = 0.5$ composants et en utilisant une contrainte d'égalité spectrale sur la caféine (courbe continue). Les profils de pénétration obtenus précédemment par MCR-ALS avec $K = 2$ composants en utilisant une contrainte d'égalité spectrale sur la caféine sont également tracés pour comparaison (courbe discontinue).

2.2.5. Conclusion

Dans une application visant à illustrer les situations où l'AC est difficilement détectable, la MCR-ALS et la MVC-NMF n'ont l'une comme l'autre pas été capables de détecter de façon totalement non supervisée la signature spectrale de la caféine à l'état pur parmi les variations spectrales des données expérimentales. Cependant, pour les deux méthodes, l'application d'une contrainte d'égalité spectrale sur la caféine a permis de dissocier l'information qui lui est relative du reste des informations concernant le tissu cutané. Cette dissociation est pertinente lorsque seulement $K = 2$ composants pour la MCR-ALS et $K = 3$ composants avec $\tau = 0.5$ (qui équilibre la fidélité du modèle par rapport aux données avec la contrainte de volume minimal formé par les spectres de composants purs) pour la MVC-NMF sont utilisés. Ces modèles sont en conséquence considérés comme les plus plausibles et fiables pour suivre les cinétiques de pénétration de la caféine à travers les couches de l'épiderme. De plus, bien que la MVC-NMF fournisse une décomposition spectrale des données avec des erreurs d'ajustements plus faibles que celles obtenues par MCR-ALS, les profils de pénétration de la caféine obtenus à partir de ces deux modèles sont très similaires et présentent les mêmes cinétiques de pénétration. Au final, cette étude montre que dans un cas d'étude où l'actif est difficilement détectable il est recommandé d'utiliser les méthodes MCR-ALS et MVC-NMF avec une supervision spectrale sur sa signature Raman pour pouvoir suivre ses cinétiques de pénétration cutanée.

D'un point de vue cosmétique, l'application de ces deux méthodes a permis d'affirmer l'homogénéité de la pénétration de la caféine à travers les couches de l'épiderme au cours du

temps et quelle que soit la formulation utilisée. Plus encore, elle révèle que l'utilisation de la formulation par Hydrofeeling® permettrait une pénétration de la caféine en quantité plus importante dans les couches plus profondes de l'épiderme après 24 heures d'exposition. Ce résultat est encourageant puisque malgré une suppression des tensio-actifs, la technologie Hydrofeeling® pourrait permettre la délivrance d'AC en quantité plus importante et donc améliorer l'efficacité des produits cosmétiques développés. Il convient de nuancer que ces conclusions peuvent ne pas se généraliser à toutes les molécules. Des analyses comparatives similaires à partir d'autres molécules formulées avec des tensio-actifs et par Hydrofeeling® sont nécessaires pour conclure à une amélioration générale apportée par la technologie l'Hydrofeeling® sur la pénétration de la molécule dans la peau.

2.3. ÉTUDE DES CONDITIONS DE STOCKAGE DES ÉCHANTILLONS DE PEAU RECONSTRUITE

Dans le domaine de la recherche cosmétique, les modèles de peau reconstruite sont de plus en plus utilisés pour des études de pénétration. Cette alternative à l'utilisation de peaux humaines excisées permet d'outrepasser les problèmes d'inter-variabilité et de disponibilité puisque les échantillons de peau reconstruite sont commercialisés prêts à l'emploi dans des plaques multi-puits après que leurs cellules aient été différenciées sur une matrice de collagène. Ces modèles de peau sont maintenus en vie grâce à leur immersion dans une solution nutritive et leurs cellules continuent alors de se différencier. La fonction de desquamation cellulaire étant inexistante dans les cultures *in vitro*, le SC a donc tendance à s'épaissir ce qui engendre une augmentation de la fonction barrière cutanée. Par conséquent, les fournisseurs étant qualifiés dans la gestion des stades de maturité des tissus cultivés recommandent généralement de les utiliser dans les 24 heures pour des études de pénétration afin d'éviter les problèmes de comparabilité des résultats. Cette recommandation peut cependant parfois être difficile à respecter en raison d'un nombre élevé de conditions d'études à réaliser impliquant ainsi un devoir de gestion des échantillons bien réalisée, ou bien du fait que le nombre de conditions à tester soit inférieur à la quantité d'échantillons disponibles destinant les échantillons restant à être non utilisés dans l'immédiat.

De nombreuses études de pénétration ont été menées sur des biopsie de peaux humaines congelées. En effet, il a été démontré que leur stockage à -20°C n'affecte que très peu la barrière cutanée après 3 semaines [11]. En revanche, une congélation des échantillons à -80°C implique des dommages significatifs au niveau du SC qui assure principalement à lui seul la fonction barrière cutanée [11]. L'objet de cette étude est donc d'évaluer l'influence de telles conditions de stockage sur des modèles de peau reconstruite à l'aide de l'imagerie Raman et d'une méthode de démélange spectral. Le but final est de savoir si le stockage des échantillons de peau reconstruite à -20°C et -80°C a un quelconque impact sur la fonction barrière de la peau et s'il n'altère pas les cinétiques de pénétration de molécules exogènes qui, par conséquent, faussent les résultats des études de pénétration. Pour se faire, un actif cosmétique (Delipidol®) est traqué à travers des échantillons de peau reconstruite utilisés dès leur réception et après avoir été congelés. Les profils de pénétration du Delipidol® sont estimés par la méthode d'analyse multivariée MCR-ALS selon les différentes conditions de stockage pour pouvoir comparer leurs cinétiques de pénétration et éventuellement mettre en évidence les variations dans la perméabilité du SC.

2.3.1. Ensemble des données expérimentales

L'AC employé est le Delipidol®, qui est la molécule utilisée dans la précédente étude (Chapitre IV. 2.1) et fournie par le partenaire industriel du projet MISTIC (ARD 2020, Programme Cosmétosciences), à savoir BioEurope Solabia, BioEurope France (Pantin, France). Le choix de

cette molécule pour l'étude de l'influence des conditions de stockage sur les cinétiques de pénétration de molécules exogènes à la peau se justifie par sa faculté à diffuser dans les couches inférieures de l'épiderme et sa forte caractérisation spectrale Raman facilement identifiable. Pareillement à l'étude précédente, pour respecter les concentrations généralement retrouvées dans les produits cosmétiques une solution éthanolique de Delipidol® à 5% est préparée.

Après leur réception, les échantillons d'épiderme humain reconstruit provenant de chez Episkin® (Lyon, France) sont séparés en trois groupes pour tester trois conditions de stockage différentes. Le premier ensemble est utilisé directement pour l'étude de pénétration du Delipidol® tandis que les deux autres ensembles sont stockés à -20°C et -80°C pendant plusieurs jours avant d'être utilisés. Avant leur utilisation, les échantillons de peau congelés sont placés selon les instructions du fournisseur dans le milieu de culture composé d'une solution nutritive pendant 2 heures afin de les ramener à température ambiante, température de travail idéale. Ensuite, la solution éthanolique de Delipidol® est appliquée à la surface des échantillons d'épiderme et laissée en contact pendant 1 et 3 heures pour chaque condition de stockage. Une fois le temps d'exposition atteint, la surface des échantillons est nettoyée à l'aide d'une solution éthanolique afin d'y enlever toute trace résiduelle de la solution. Puis des coupes transversales de 20 µm sont réalisées à l'aide d'un cryo-microtome et sont placées sur un substrat de CaF_2. Des coupes de peaux non traitées sont également préparées pour chaque condition de stockage en tant qu'échantillons contrôles.

Enfin, les images Raman sont enregistrées à l'aide d'un microscope confocal Raman, un LabRam (Horiba Jobin-Yvon, Villeneuve-d'Ascq, France). Les images de diffusion Raman sont obtenues en utilisant un laser de 633 nm de puissance 8mW au niveau de l'échantillon et un objectif LMPlan x100 (Olympus, NA 0.75, Rungis, France). Le trou confocal est fixé à 300 µm pour toutes les mesures spectrales Raman et un réseau dispersif composé de 300 lignes/mm est utilisé pour disperser la diffusion Raman des échantillons sur un capteur CCD et obtenir une résolution spectrale de 4 cm^{-1}. Compte tenu de ces paramètres d'acquisition, la résolution spatiale des images Raman enregistrées est de 2 µm dans les deux directions x et y. À noter que deux images par condition de stockage, soit un total de 12 images Raman, sont utilisées dans les analyses multivariées suivantes et que toutes affichent le SC orienté vers la gauche et une profondeur cutanée comprise entre 40-100 µm. D'autre part, les images Raman acquises à partir d'échantillons exposés pendant 1 heure et 3 heures au Delipidol® ont été réalisées à des dates différentes : juin 2017 pour les premières et janvier 2020 pour les deuxièmes. L'acquisition de ces images Raman a pour objectif de définir la profondeur de pénétration du Delipidol® et d'évaluer l'éventuelle variabilité de sa diffusion selon les différentes conditions de stockage des échantillons de peau reconstruite. Ces informations permettront de définir si la barrière cutanée des épidermes reconstruits a été altérée ou non par les différentes conditions de stockage.

2.3.2. Prétraitements des données

Tous les spectres composants les images Raman sont soumis à la routine de prétraitement présentée dans le Chapitre III. Plus précisément, le prétraitement comprend un débruitage spectral, une correction de la ligne de base, une annulation des spectres correspondant au substrat, une suppression des artefacts et une normalisation. Le débruitage des spectres est d'abord réalisé à l'aide d'une ACP. Seulement 1% du nombre total de spectres contenus dans les images Raman est utilisé pour décrire leurs caractéristiques spectrales sans bruit. Les spectres sont ensuite réalignés spectralement sur la bande caractéristique située à 1004 cm^{-1} et les informations sont concentrées dans la fenêtre spectrale [400-1800] cm^{-1}, la région spectrale d'intérêt des constituants majeurs de la peau et du Delipidol®. Puis, les spectres appartenant au substrat sont annulés. Enfin, les spectres sont corrigés de leurs lignes de bases respectives en

utilisant la méthode Rubberband et de leurs artefacts avec la méthode hybride proposée utilisant la technique de « wound healing » avant d'être normalisés avec la norme ℓ_2.

2.3.3. Analyse des données

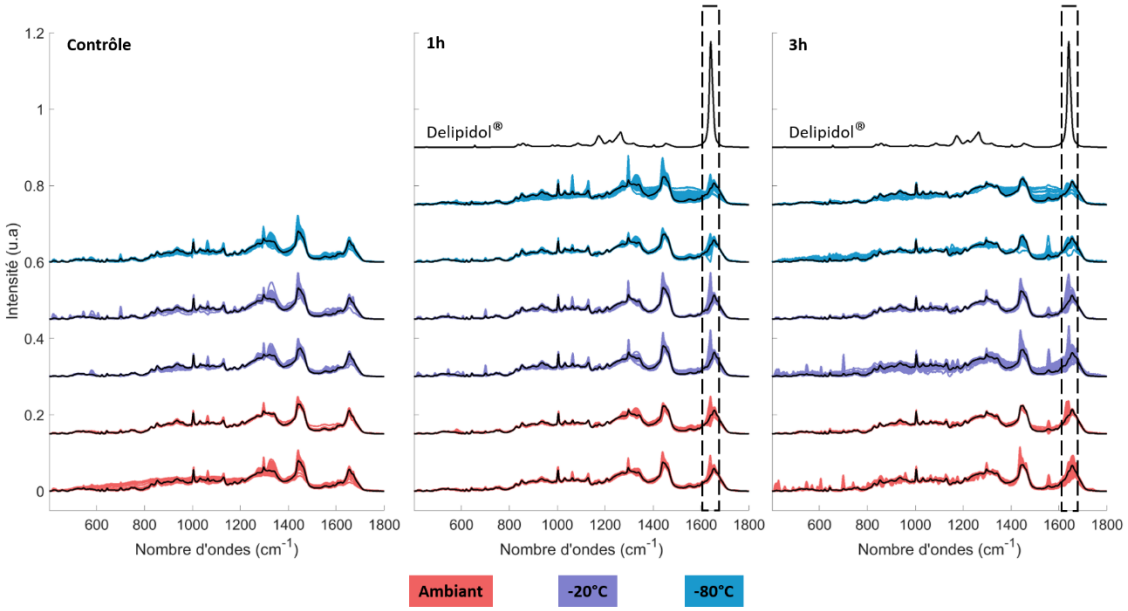


Figure IV-21 – Comparaison des spectres enregistrés à partir d'échantillons contrôles (à gauche) et exposés au Delipidol® pendant 1 heure (au centre) et 3 heures (à droite) avec la référence spectrale du Delipidol® pour trois conditions de stockage : à température ambiante (rouge), à -20°C (violet) et à -80°C (bleu). Les spectres noirs représentent la moyenne de chaque image Raman.

Les spectres des images Raman obtenues à partir des échantillons de peau témoins et exposés au Delipidol® pendant 1 et 3 heures après avoir été conservés selon trois conditions de stockage (température ambiante, -20°C et -80°C) sont illustrés dans la Figure IV-21. En se rapportant à la description des bandes caractéristiques de la peau et du Delipidol® réalisée dans le Chapitre I et le Chapitre IV. 2.1., la diffusion du Delipidol® à travers les échantillons d'épiderme apparait comme effective. En effet, les bandes caractéristiques du Delipidol® se superposent à celles de la peau. Plus précisément, sa bande dominante originellement située à 1630 cm^{-1} est exprimée à 1640 cm^{-1} quelles que soient les conditions de stockage de l'échantillon sur lequel il a été appliqué. L'expression de cette bande est cependant plus ou moins distinguable selon la température de stockage de l'échantillon et la durée de son exposition à l'actif. En particulier, lorsque l'échantillon a été stocké à -80°C, l'intensité de la bande caractéristique de l'actif après 1 heure ou 3 heures d'exposition est plus faible que celle observée lorsque l'échantillon a été préalablement conservé à -20°C ou à température ambiante. Plus encore, alors qu'aucune différence d'intensité considérable n'est observée entre les échantillons exposés pendant 1 heure et 3 heures au Delipidol® après avoir été stockés à -20°C ou à température ambiante, la bande caractéristique de l'actif observée dans les échantillons conservés à -80°C est d'intensité beaucoup plus faible après 3 heures d'exposition que celle observée après 1 heure d'exposition. D'après cette analyse, la conservation des échantillons d'épiderme à -80°C semblerait donc affecter la perméabilité de la peau et donc altérer sa fonction barrière. Cependant, cette simple observation générale des spectres issus des différentes images Raman enregistrées ne permet pas de définir de manière précise la progression du Delipidol® à travers la peau selon les différentes conditions de stockage. L'application de la MCR-ALS sur l'ensemble des images Raman enregistrées à partir des échantillons exposés pendant 1 et 3 heures au Delipidol® après avoir été stockés selon trois conditions de stockage a pour objectif de fournir des informations sur la répartition spatiale et la

profondeur de pénétration de l'actif et donc permettre d'établir des cinétiques de pénétration pour chacune des expériences. Ainsi, le conditionnement d'échantillons déterminant une cinétique de pénétration différente de celle obtenue lorsque les échantillons sont utilisés dès leur réception pourrait être identifié.

Pour ce qui est de l'utilisation de la MCR-ALS pour suivre la progression de l'actif dans les échantillons d'épiderme humain reconstruit, il est important de noter que les précédentes observations réalisées sur les spectres de chaque image Raman ont révélées de fortes modifications spectrales liées à l'application de Delipidol® à la surface des tissus selon les conditions de stockage et les temps d'exposition. Par conséquent, la détection et le suivi de l'actif dans les données spectrales risque d'être difficile dans certaines conditions expérimentales. Dans ce cas de figure, la précédente étude sur le suivi de la caféine à travers l'épiderme a conduit à une recommandation qui consiste à appliquer la MCR-ALS avec une contrainte d'égalité spectrale sur l'actif lorsque celui-ci est difficilement détectable à travers des données Raman (Chapitre IV.2.2). Forts de ces conclusions, nous proposons donc d'appliquer la MCR-ALS à l'aide d'une supervision spectrale sur le Delipidol® pour suivre sa progression à travers les tissus cutanés. Une comparaison des résultats avec ceux obtenus par application de la MCR-ALS sans contrainte d'égalité spectrale sur l'actif confirmeront la pertinence de l'utilisation de cette contrainte supplémentaire dans la MCR-ALS pour estimer ses cinétiques de pénétration.

2.3.3.1. Estimation du nombre de composants

L'objectif principal de la MCR-ALS est d'aider l'interprétation de la composition biochimique et structurelle des données en modélisant les informations spectrales par un nombre réduit de composants biochimiques K suffisants à leur description. Dans la précédente étude visant à suivre le Delipidol® à travers des échantillons d'épiderme (Chapitre IV. 2.1), le modèle estimé par la méthode MCR-ALS avec $K = 4$ composants a été jugé comme le plus approprié pour modéliser simplement les données expérimentales et donc pour suivre sa progression dans les différentes couches épidermiques. Pour cette étude de l'impact des conditions de stockage des échantillons d'épiderme sur la pénétration d'un AC, les mêmes modèles de peau et le même AC ont été utilisés. Par conséquent, la décomposition spectrale des données expérimentales à l'aide de $K = 4$ composants est intuitive.

Toutefois, pour guider la sélection du nombre optimal de composants purs à estimer pour reproduire le mieux possible les données expérimentales et potentiellement confirmer le choix de $K = 4$ composants déjà utilisé, les trois critères que sont la SVD, le critère DW et l'indice KMO ont été utilisés. À nouveau, tous s'accordent sur le fait qu'un nombre de composants élevé est nécessaire pour englober l'ensemble de la variabilité spectrale des données. Cependant, le tracé de chacun de ces trois critères aide à définir un nombre réduit de composants permettant une description d'une grande partie des variations spectrales des données.

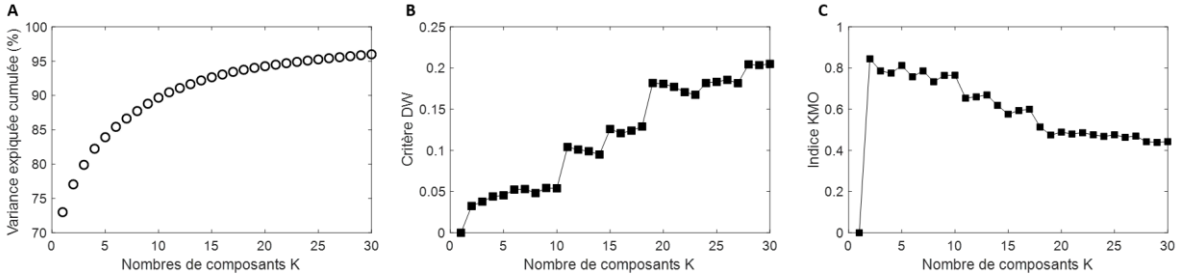


Figure IV-22 – (**A**) *Variance expliquée cumulée obtenue par SVD appliquée sur les images Raman des échantillons conservées selon les 3 conditions de stockage et exposés aux Delipidol® ;* (**B**) *Critère de Durbin-Watson appliqué sur les résultats de SIMPLISMA ;* (**C**) *Indice KMO appliqué sur les résultats SIMPLISMA.*

- La variance expliquée cumulée obtenue par SVD sur l'ensemble des images Raman enregistrées à partir des échantillons exposés aux Delipidol® montre, d'après la Figure IV-22 (**A**), que la quantité d'informations prise en compte dans le modèle estimé augmente avec le nombre de composants utilisés pour la décomposition spectrale. Plus précisément, au moins 24 composants sont nécessaires pour espérer décrire plus de 95% de la variabilité des données, 11 composants permettent de modéliser au moins 90% des données ou encore 6 composants permettent d'atteindre 85% de la variance expliquée. En revanche, 4 composants suffisent pour exprimer 80% des variations spectrales les plus importantes.

- Le critère DW est calculé sur les résidus des modèles définis à partir des 30 premiers spectres les plus purs estimés par SIMPLISMA. Les coefficients de DW, illustrés sur la Figure IV-22 (**B**), sont tous compris entre 0 et 0.25, ce qui signifie qu'une corrélation entre les spectres résiduels existe et donc que tous les composants purs n'ont pas été extraits des données. L'extraction d'un composant contribuant fortement aux données se traduit par une diminution de l'autocorrélation des spectres résiduels et s'observe donc par une augmentation brutale du critère DW. De tels changements brutaux dans le tracé du critère DW se distinguent lorsque $K = \{5, 8, 10, 14{,}18{,}23{,}27\}$.

- Les indices KMO calculés sur les résidus des modèles définis à partir des 30 premiers spectres les plus purs estimés par SIMPLISMA sont illustrés sur la Figure IV-22 (**C**). Lorsque le nombre de composants utilisés dans le modèle augmente, l'indice KMO diminue ce qui témoigne d'une baisse de la corrélation entre spectres résiduels et donc de l'extraction d'un composant contribuant fortement dans les données. De forts déclins sont observés dans le tracé des indices KMO lorsque $K = \{4, 6, 8, 13, 15{,}17\}$.

La sélection du nombre réduit de composants nécessaires à la modélisation de la plupart des variations spectrales par l'analyse de ces trois critères est à nouveau très partagée. Néanmoins, deux des critères suggèrent le nombre de composants $K = 4$ comme potentiellement approprié à la modélisation d'une grande partie de la variabilité des données expérimentales à l'exemple de l'étude précédente réalisée à partir des mêmes matériaux. D'autre part, l'analyse des trois critères mène à définir le nombre de composants $K = 8$ comme autre nombre de composants potentiellement approprié à la modélisation des données expérimentales. Par conséquent, seuls ces deux nombres $K = \{4, 8\}$ seront utilisés pour appliquer la MCR-ALS aux données dans la suite.

2.3.3.2. Application de la MCR-ALS avec une contrainte d'égalité spectrale sur le Delipidol®

La MCR-ALS est appliquée sur l'ensemble des données spectrales enregistrées à partir des échantillons de peau conservés selon les trois conditions de stockage et exposés pendant 1 et 3 heures au Delipidol® de manière simultanée. Elle est appliquée en utilisant des contraintes de non-négativité et de normalisation sur les spectres et leurs concentrations relatives associées avec $K = \{4, 8\}$ spectres de composants purs initialisés par la méthode SIMPLISMA mais aussi en imposant une égalité spectrale sur la référence spectrale du Delipidol®.

Statistiques de modélisation

Les statistiques d'ajustement global des deux modèles estimés par rapport aux données expérimentales, à savoir le manque d'ajustement global (LOF) et la variance expliquée (R^2), sont présentées dans le Tableau IV-7. Alors que le modèle estimé avec $K = 8$ composants et une contrainte d'égalité spectrale sur le Delipidol® s'ajuste aux données avec une erreur globale de

8.6676 % et explique 99.6237 % de la variabilité spectrale, celui estimé sous les mêmes contraintes avec seulement $K = 4$ composants manque d'ajustement aux données expérimentales à hauteur de 14.9191 % et explique 98.8808 % des variations spectrales. En résumé, l'augmentation du nombre de composants utilisés dans la décomposition spectrale permet de diminuer le manque d'ajustement global du modèle aux données expérimentales et de prendre en compte une plus grande partie de leurs variations spectrales. Autrement dit, un nombre de composants plus élevé dans un modèle rend la description des données plus précise et donc une meilleure reproduction des données.

Enfin, pour évaluer l'unicité des solutions fournies par la MCR-ALS appliquée avec une supervision spectrale sur l'actif, le degré d'ambiguïté rotationnelle de chaque composant des deux décompositions spectrales est évalué et les différences de contributions relatives de chacun d'eux aux limites maximales et minimales sont présentées dans le Tableau IV-7. Ces différences sont pour la plupart d'ordre compris entre 10^{-5} et 10^{-2} quel que soit le nombre de composants $K = \{4, 8\}$ et ne remettent donc pas en question la singularité des solutions.

Tableau IV-7 – Mesure des critères de manque d'ajustement LOF (Lack Of Fit en %), de variance expliquée R² (en %) et de l'ambiguïté rotationnelle des modèles estimés par MCR-ALS initialisée par SIMPLISMA en utilisant $K = \{4,8\}$ composants purs avec (c-DEL) et sans (nc-DEL) contrainte d'égalité spectrale sur le Delipidol®.

		$K = 4$		$K = 8$	
		nc-DEL	**c-DEL**	**nc-DEL**	**c-DEL**
LOF		56.4038	14.9191	6.9005	8.6676
R²		82.5749	98.8808	99.7616	99.6237
Ambiguïté rotationnelle	S_1	$3.365 \cdot 10^{-2}$	$-1.963 \cdot 10^{-4}$	$-1.787 \cdot 10^{-4}$	$7.897 \cdot 10^{-3}$
	S_2	$2.117 \cdot 10^{-3}$	$-2.037 \cdot 10^{-4}$	$1.486 \cdot 10^{-2}$	$1.305 \cdot 10^{-2}$
	S_3	$-1.178 \cdot 10^{-1}$	$-2.023 \cdot 10^{-4}$	$-2.347 \cdot 10^{-3}$	$-1.186 \cdot 10^{-4}$
	S_4	$-7.121 \cdot 10^{-2}$	$-5.868 \cdot 10^{-5}$	$3.379 \cdot 10^{-3}$	$1.887 \cdot 10^{-3}$
	S_5			$1.504 \cdot 10^{-2}$	$2.229 \cdot 10^{-3}$
	S_6			$1.39 \cdot 10^{-2}$	$-1.632 \cdot 10^{-3}$
	S_7			$1.219 \cdot 10^{-3}$	$-2.754 \cdot 10^{-4}$
	S_8			$4.785 \cdot 10^{-3}$	$3.701 \cdot 10^{-3}$

Interprétation spectrale des modèles estimés

L'ajout d'une contrainte d'égalité spectrale sur le Delipidol® dans l'application de la MCR-ALS sur les données expérimentales permet de s'assurer que les informations relatives à l'AC soient isolées des informations concernant la peau. Les spectres purs ainsi estimés en utilisant $K = \{4, 8\}$ composants et identifiés par $S_{i=1,\dots,8}$ sont illustrés dans la Figure IV-23 (**A**-**B**). À noter que leur ordre a été défini de façon à faciliter la lecture et leur comparaison. Alors que S_1 a été défini comme le spectre référentiel du Delipidol® non modifié par l'algorithme MCR-ALS, les autres spectres estimés $S_{i=2,\dots,8}$ ont été réévalués au fil des itérations.

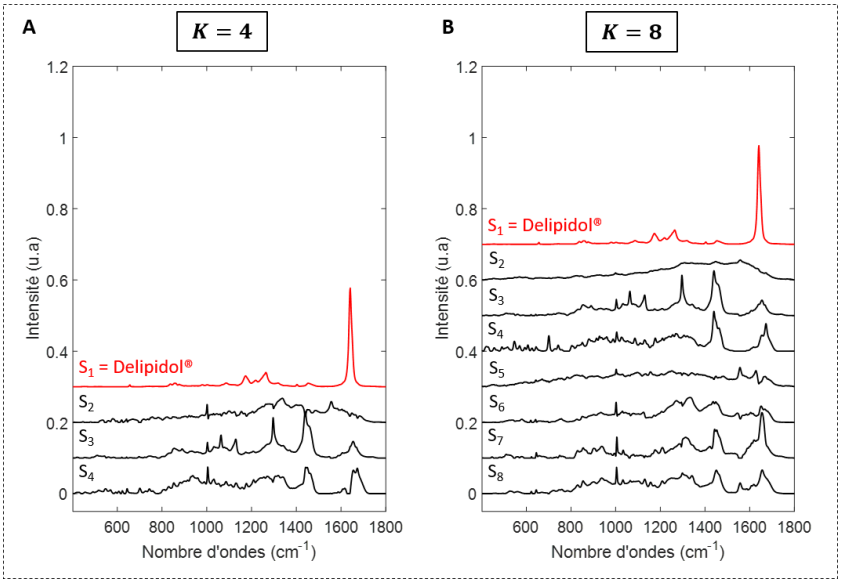


*Figure IV-23 – Spectres de composants purs estimés par la MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le Delipidol® pour (**A**) $K = 4$ et (**B**) $K = 8$ sur les images Raman des échantillons de peau conservés selon les 3 conditions de stockage et exposés aux Delipidol® pendant 1 et 3 heures (décalés en intensité pour plus de clarté). Les spectres ont été initialisés par SIMPLISMA.*

- Pour $K = 4$, alors que S_1 est contraint à être la signature de référence du Delipidol®, les spectres S_2, S_3 et S_4 ne contiennent aucune bande caractéristique de celui-ci et affichent les bandes Raman d'un échantillon de peau non exposé. En effet, leur corrélation avec le spectre moyen des données témoins s'élève à 0.7107, 0.9023 et 0.8992 (Tableau IV-8). Quant à leur corrélation avec la référence spectrale de l'actif, elle s'élève respectivement à 0.1304, 0.1860 et 0.1413 (Tableau IV-8) ce qui reste inférieur à la corrélation de référence entre l'actif et la moyenne des échantillons contrôles qui est de 0.2708 (Tableau IV-8). Au final, en raison de la dissociation effective des informations concernant l'actif de celles relatives à la peau, l'utilisation du spectre contraint S_1 pour suivre le Delipidol® à travers les échantillons cutanés est justifié et jugé comme pertinent.

- Pour $K = 8$, tandis que S_1 correspond également par contrainte au Delipidol®, $S_{i=2,\dots,8}$ présentent des bandes spectrales associées à la peau comme le confirme leur corrélation au spectre moyen des données contrôles qui s'élève dans l'ordre à 0.8912, 0.8898, 0.6418, 0.9238, 0.7993 et 0.9474 (Tableau IV-8). De plus, aucune bande du Delipidol® n'est observable de manière distincte dans ces spectres estimés. Cependant, S_7 et S_8 ont une corrélation avec la référence spectrale de l'actif de 0.4838 et 0.2825 (Tableau IV-8) qui est supérieure à la corrélation de référence entre l'actif et la moyenne des échantillons contrôles. Cela signifie que ces deux spectres contiennent très certainement des informations concernant le Delipidol® dissimulées par des données relatives à la peau. Par conséquent, malgré la contrainte d'égalité spectrale qui lui est imposée, S_1 ne peut pas être utilisé de manière fiable pour suivre le Delipidol® à travers les échantillons d'épiderme puisqu'une partie de sa contribution dissimulée dans les spectres S_7 et S_8 ne serait pas prise en considération.

En résumé, parmi les deux modèles obtenus, les contributions relatives locales du Delipidol® estimées par la MCR-ALS avec une contrainte d'égalité spectrale et $K = 4$ composants se montrent être les plus fiables pour étudier sa progression à travers l'épiderme.

Tableau IV-8 – Coefficients de corrélation des spectres estimés $S_{i=1,\dots,8}$ obtenus par MCR-ALS appliquée avec (c-DEL) et sans (nc-DEL) contrainte d'égalité spectrale sur le Delipidol® pour $K = \{4,8\}$ sur les images Raman des échantillons de peau conservés selon les 3 conditions de stockage et exposés aux Delipidol® pendant 1 et 3 heures avec le spectre de référence du Delipidol® et le spectre moyen des images Raman des peaux contrôles.

	$K = 4$				$K = 8$							
c-DEL	S_1 **DEL**	S_2	S_3	S_4	S_1 **DEL**	S_2	S_3	S_4	S_5	S_6	S_7	S_8
CTRL	**0.2708**	0.7107	0.9023	0.8992	**0.2708**	0.6388	0.8912	0.8898	0.6418	0.9238	0.7993	0.9474
DEL	1.0000	0.1304	0.1860	0.1413	1.0000	0.1473	0.1862	0.1608	0.0768	0.2058	**0.4838**	**0.2825**
nc-DEL												
CTRL	**0.8310**	0.8417	0.9182	0.9197	**0.5906**	0.6024	0.8947	0.8827	0.5904	0.9125	0.8086	0.9402

Erreurs de reproduction des données

Les erreurs de modélisation des données associées au modèle estimé par la MCR-ALS avec une contrainte d'égalité spectrale sur le Delipidol® et $K = 4$ composants ainsi que l'histogramme du spectre résiduel moyen sont illustrés dans la Figure IV-24. L'histogramme des résidus moyens (Figure IV-24 (**A**)) est formé de deux gaussiennes : une gaussienne étroite centrée en 0 représentant le bruit résiduel présent dans les données et une gaussienne plus aplatie centrée en 0.003 u.a représentant les autres variations spectrales non expliquées par le modèle. La présence de cette seconde gaussienne indique que toutes les variations spectrales présentes dans les données ne sont pas considérées dans les spectres de composants purs et donc que tous les composants des données expérimentales n'ont pas été extraits. L'observation du spectre résiduel moyen (en noir sur la Figure IV-24 (**B**)) permet de comprendre que la composante non extraite par le modèle présente principalement des bandes Raman associées à la peau. Plus encore, l'analyse de l'allure des erreurs résiduelles associées à chaque spectre des images Raman (en gris sur la Figure IV-24 (**B**)) permet d'assurer qu'aucune de ces erreurs ne soient liées aux bandes caractéristiques du Delipidol®. Ainsi par l'analyse de ces résidus, il a été vérifié que toute l'information relative à la progression du Delipidol® dans les différentes couches cutanées a bien été prise en compte dans le modèle. Au final, la pertinence de l'utilisation du modèle estimé par la MCR-ALS avec une contrainte d'égalité spectrale et $K = 4$ composants pour suivre la pénétration percutanée du Delipidol® est confirmée.

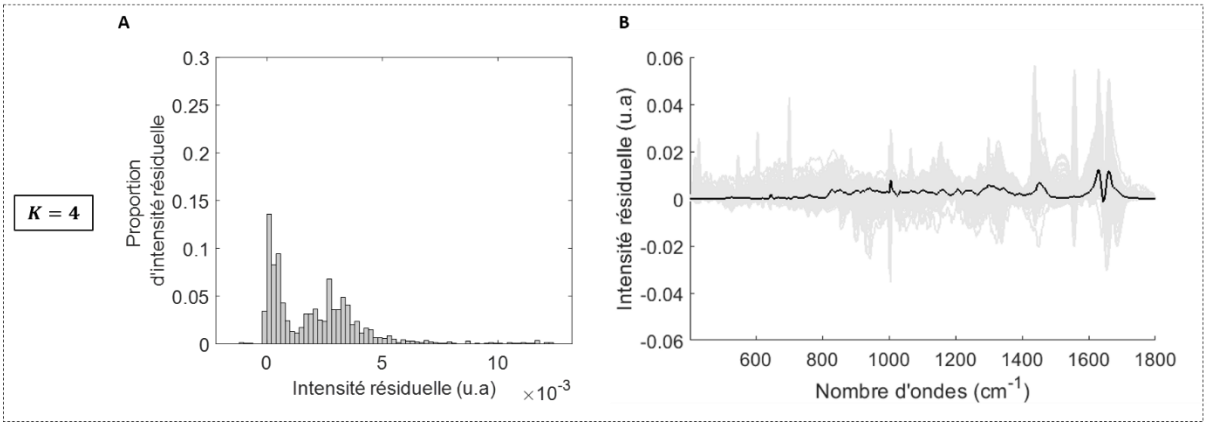


Figure IV-24 – Résidus des modèles estimés par la MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le Delipidol® pour $K = 4$ composants sur les images Raman des échantillons de peau conservés selon les 3 conditions de stockage et exposés aux Delipidol® pendant 1 et 3 heures ; (**A**) *Histogramme du spectre résiduel moyen du modèle estimé;* (**B**) Spectres résiduels de l'ensemble des images Raman (en gris) et spectre résiduel moyen du modèle estimé (en noir).

Cinétiques de pénétration du Delipidol®

Les concentrations relatives associées au spectre S_1 contraint à correspondre à la référence spectrale du Delipidol® dans le modèle estimé sous contrainte d'égalité spectrale avec $K = 4$ composants par la MCR-ALS ont été reconstituées sous forme de cartes de distribution qui sont illustrées dans la Figure IV-25. Pour chaque condition de stockage et de temps d'exposition, les cartes de distribution affichent un gradient allant du SC jusqu'aux couches épidermiques plus profondes : les contributions relatives du Delipidol® sont globalement et significativement plus importantes dans le SC que dans les autres couches cutanées. Cependant, la différence de concentrations entre le SC et l'épiderme viable est moins évidente sur les cartes associées aux échantillons stockés préalablement à -80°C et exposés pendant 3 heures au Delipidol®.

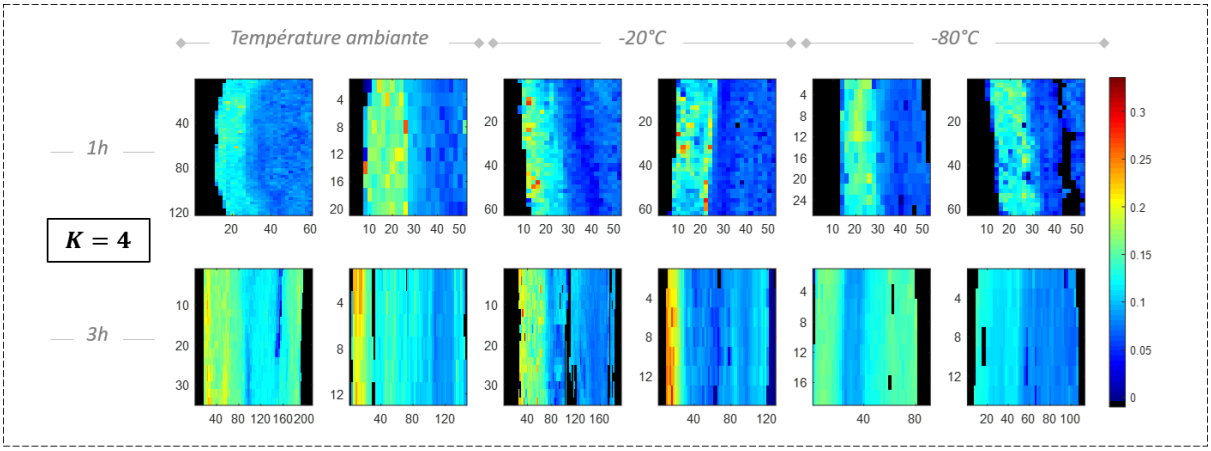


Figure IV-25 – Cartes de distribution du Delipidol® après 1 et 3 heures d'exposition sur des échantillons de peau conservés selon les 3 conditions de stockage : à température ambiante, à -20°C et à -80°C. Cartes reconstruites à partir du composant S_1 contraint à être la signature du Delipidol® dans le modèle obtenu par MCR-ALS appliquée avec une contrainte d'égalité spectrale et $K = 4$ composants. Les axes x et y des cartes de distribution sont exprimées en micromètres.

Cette observation est confirmée par le calcul des concentrations relatives moyennes du Delipidol® par niveau de profondeur pour chaque condition d'expérimentation présentées dans le Tableau IV-9. En effet, pour les conditions de stockage à température ambiante, à -20°C et à -80°C, la concentration relative du Delipidol® dans les 20 premiers micromètres après 1 heure d'exposition s'élève en moyenne à 0.1155 u.a, 0.1219 u.a et 0.1244 u.a et descend à 0.0781 u.a, 0.0607 u.a et 0.0609 u.a dans les 20 micromètres suivants. Puis après 3 heures d'exposition sur des échantillons conservés à température ambiante ou à -20°C, l'actif stagne dans le SC avec une contribution plus élevée (0.1807 et 0.1733 u.a) avant de pénétrer progressivement dans l'épiderme viable (0.1662 et 0.1591 u.a). En revanche, après 3 heures d'exposition sur des échantillons préalablement stockés à -80°C, la contribution relative du Delipidol® dans les 20 premiers micromètres n'est qu'en moyenne de 0.1466 u.a et de 0.1008 u.a dans les 20 micromètres suivants.

Tableau IV-9 – Concentrations relatives estimées du Delipidol® moyennes par niveau de profondeur après 1 et 3 heures d'exposition sur des échantillons conservés selon les 3 conditions de stockage (à température ambiante, à -20°C et à -80°C) obtenues par MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le Delipidol® et $K = 4$ composants.

$K = 4$	Ambiante		-20°C		-80°C	
c-DEL	**1h**	**3h**	**1h**	**3h**	**1h**	**3h**
0 – 20 µm	0.1155	0.1807	0.1219	0.1733	0.1244	0.1466
20 – 40 µm	0.0781	0.1662	0.0607	0.1591	0.0609	0.1008

Les différences de contribution de l'actif en fonction des conditions d'expérimentation sont plus facilement observables sur les profils de pénétration illustrés sur la Figure IV-26. Alors qu'après 1 heure d'exposition les profils de pénétration du Delipidol® pour chaque condition de stockage suivent la même tendance, le profil de pénétration obtenu après 3 heures d'exposition sur les échantillons stockés à -80°C illustre une tendance bien en dessous de ceux obtenus avec des échantillons conservés à température ambiante ou à -20°C. Cette infériorité est particulièrement plus prononcée sur les 20 premiers micromètres, à savoir dans le SC. Tout compte fait, la conservation des échantillons d'épiderme à -80°C s'avère provoquer une altération de la fonction barrière cutanée puisque la pénétration du Delipidol® après 3 heures d'exposition est affaiblie.

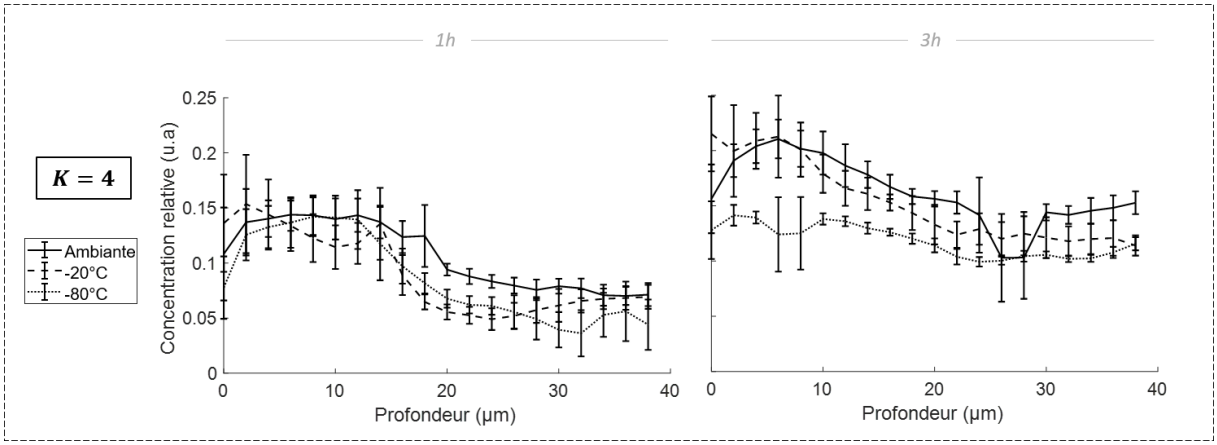


Figure IV-26 – Profils de pénétration moyens du Delipidol® après 1 (à gauche) et 3 heures (à droite) d'exposition dans des échantillons de peau conservés selon les 3 conditions de stockage : à température ambiante (ligne continue), à -20°C (ligne discontinue) et à -80°C (ligne pointillée). Profils obtenus à partir des concentrations relatives C_1 estimées à l'aide d'une MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le Delipidol® avec $K = 4$ composants.

2.3.3.3. Évaluation de la pertinence d'utilisation d'une contrainte d'égalité spectrale dans la MCR-ALS

Afin de confirmer la pertinence d'utiliser une contrainte d'égalité spectrale sur le Delipidol® supplémentaire dans la MCR-ALS pour cette étude, la MCR-ALS a également été appliquée en utilisant uniquement des contraintes de non-négativité et de normalisation sur les spectres et leurs concentrations relatives associées avec $K = \{4, 8\}$ spectres de composants purs initialisés par la méthode SIMPLISMA sur l'ensemble des données spectrales issues des différentes conditions expérimentales. Les modèles ainsi estimés sont comparés à ceux obtenus lorsqu'une supervision spectrale sur le Delipidol® est introduite dans l'analyse MCR-ALS des données en terme d'ajustement global aux données et d'interprétation de la décomposition spectrale.

Statistiques de modélisation

Les statistiques d'ajustement global des deux modèles estimés par rapport aux données expérimentales, à savoir le manque d'ajustement global (LOF) et la variance expliquée (R^2), sont présentées dans le Tableau IV-7. Alors que le modèle estimé avec $K = 8$ spectres de composants purs sans contrainte d'égalité spectrale sur le Delipidol® s'ajuste aux données avec une erreur globale de 6.9005 % et explique 99.7616% de la variabilité spectrale, celui estimé sous les mêmes contraintes avec seulement $K = 4$ composants manque d'ajustement aux données expérimentales à hauteur de 56.4038 % et explique 82.5749 % des variations spectrales. Ainsi, à l'instar des modèles estimés de façon semi-supervisée, l'augmentation du nombre de composants utilisés dans la décomposition spectrale rend la description des données plus précise et donc une meilleure reproduction des données. Par ailleurs, on observe que l'introduction d'une contrainte

d'égalité spectrale sur le Delipidol® dans la décomposition spectrale n'est pas aussi unanime quant à l'amélioration de la description de données expérimentales. En effet, la contrainte supplémentaire augmente la qualité globale du modèle estimé lorsque $K = 4$ composants (le LOF augmente et la variance expliquée diminue) mais la réduit lorsque $K = 8$ composants (le LOF diminue et la variance expliquée augmente).

Enfin, l'évaluation du degré d'ambiguïté rotationnelle de chaque composant des deux décompositions spectrales et le calcul des différences de contributions relatives de chacun d'eux aux limites maximales et minimales, présentées dans le Tableau IV-7, ne remettent pas en question la singularité des solutions estimées. En effet, ces différences sont pour la plupart d'ordre compris entre 10^{-4} et 10^{-2} quel que soit le nombre de composants $K = \{4, 8\}$. Toutefois, l'ajout d'une contrainte d'égalité spectrale dans l'analyse MCR-ALS utilisant $K = 4$ composants purs occasionne une forte diminution de l'ambigüité rotationnelle assurant un peu plus l'unicité et donc la fiabilité de ce modèle estimé.

Interprétation spectrale des modèles estimés

Les spectres purs résultants de la méthode MCR-ALS appliquée sans contrainte d'égalité spectrale sur le Delipidol® et en utilisant $K = \{4, 8\}$ composants, identifiés par $S_{i=1,\dots,8}$, sont illustrés dans la Figure IV-27 (**A**-**B**). À noter que leur ordre a été défini de façon à faciliter la lecture et leur comparaison. Ils ont été estimés comme proches de ceux obtenus lorsqu'une contrainte d'égalité spectrale sur le Delipidol® a été introduite dans la méthode MCR-ALS. En effet, les inter-corrélations spectrales s'élèvent dans l'ordre à 0.9616, 0.9861 et 0.9816 pour le modèle avec $K = 4$ et à 0.9366, 0.9926, 0.9853, 0.9652, 0.9969, 0.9869 et 0.9983 pour le modèle avec $K = 8$ (Tableau IV-10). Au final, l'interprétation de la composition moléculaire des spectres $S_{i=2,\dots,8}$ estimés dans les modèles n'utilisant pas la contrainte d'égalité spectrale est quasi similaire à celle présentée pour les modèles obtenus avec cette contrainte supplémentaire.

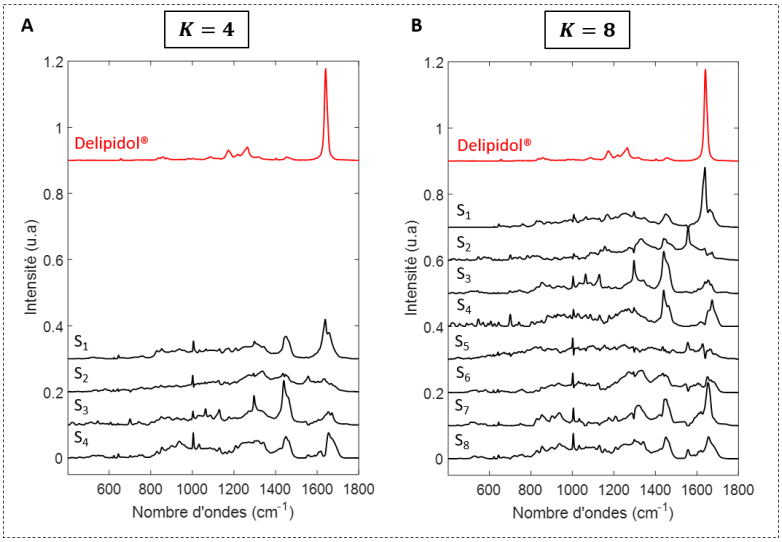


*Figure IV-27 – Spectres de composants purs estimé par la MCR-ALS appliquée sans une contrainte d'égalité spectrale sur le Delipidol® pour (**A**) $K = 4$ et (**B**) $K = 8$ sur les images Raman des échantillons de peau conservés selon les 3 conditions de stockage et exposés aux Delipidol® pendant 1 et 3 heures (décalés en intensité pour plus de clarté). Les spectres ont été initialisés par SIMPLISMA.*

- Pour $K = 4$, les spectres estimés S_2, S_3 et S_4 présentent principalement des bandes Raman caractéristiques de la peau. En effet, leur corrélation avec le spectre moyen des images Raman contrôles est élevée (respectivement 0.8417, 0.9182 et 0.9197, Tableau IV-8) et aucun de ces trois spectres ne présente la bande caractéristique du Delipidol®

exprimée dans les données à 1640 cm^{-1}. De plus, bien que leur corrélation avec la référence de cet AC est non négligeable (respectivement 0.2313, 0.1828 et 0.2183, Tableau IV-10), celle-ci reste inférieure à la corrélation entre le spectre moyen des échantillons témoins et la référence spectrale du Delipidol® qui est de 0.2708 (Tableau IV-8). Ces spectres ne contiennent donc pas d'informations concernant l'actif. Par contre, la bande Raman située à 1640 cm^{-1} relative à la présence de l'actif est fortement exprimée dans le spectre estimé S_1. La corrélation de S_1 avec la référence spectrale est considérablement grande : 0.6711 (Tableau IV-10). Toutefois, S_1 présente également de nombreuses bandes caractéristiques de la peau dont celles situées à 1004 cm^{-1}, 1452 cm^{-1} et 1655 cm^{-1}. Sa corrélation avec le spectre moyen des données contrôles s'élève à 0.8310 (Tableau IV-8). Au final, bien que S_1 soit le seul spectre estimé contenant des informations liées à l'actif, il ne correspond pas de près à sa référence spectrale à son état brut. Par conséquent, S_1 ne peut pas être utilisé pour estimer les concentrations relatives locales du Delipidol® à travers les échantillons d'épiderme, auquel cas un biais serait introduit dans les résultats rendant le suivi de sa progression imprécis.

- Pour $K = 8$, les bandes caractéristiques du Delipidol® s'observent de manière très distincte dans le spectre estimé S_1. En effet, S_1 est très similaire à la signature spectrale de cet actif comme en témoigne sa corrélation élevée de 0.7956 (Tableau IV-10). Cependant, S_1 exprime également un épaulement à 1655 cm^{-1} caractéristique de la peau. Sa corrélation de 0.5906 (Tableau IV-8) avec le spectre moyen des données témoins confirme la présence d'informations liées à la peau. Dans les autres spectres estimés $S_{2,\dots,8}$, aucune bande caractéristique du Delipidol® n'est observable. Toutefois, en comparaison à la corrélation de la référence de l'actif avec le spectre moyen des images Raman témoins, la corrélation des spectres S_7 et S_8 avec le spectre de l'actif (respectivement 0.4937 et 0.2800, Tableau IV-10) est non-négligeable et montre que ces spectres contiennent des informations qui lui sont relatives. Quant aux spectres $S_{i=2,\dots,6}$, leur corrélation avec l'actif à l'état pur est inférieure à cette corrélation de référence. De ce fait, ils ne contiennent pas d'informations concernant l'actif à suivre. Au final, bien qu'elles soient principalement contenues dans S_1, les informations relatives à l'actif sont également contenues dans S_7 et S_8. Par conséquent, cette sur-décomposition de l'information concernant le Delipidol® ne permet pas de suivre de manière précise sa progression à travers les échantillons d'épiderme à l'aide de S_1. Les estimations des contributions relatives locales du Delipidol® à l'aide de S_1 seraient erronées à la fois par l'introduction de contributions concernant la peau et par le manque d'informations exprimées dans d'autres spectres.

En résumé, aucun de ces deux modèles obtenus par application de la MCR-ALS sans contrainte d'égalité spectrale sur le Delipidol® n'a permis de détecter efficacement sa signature spectrale et de dissocier les informations qui lui sont liées de celles concernant la peau. Autrement dit, l'AC n'a pas été considéré comme un composant pur contributif à part entière dans les données expérimentales. En effet, bien qu'un spectre estimé possède ses bandes Raman caractéristiques dans chacun des modèles, celui-ci exprime en supplément des bandes Raman associées à la peau. Le suivi du Delipidol® à l'aide de ce spectre pour chaque modèle est alors fortement biaisé. Au final, la pertinence de l'utilisation du modèle estimé par la MCR-ALS avec une contrainte d'égalité spectrale et $K = 4$ composants pour suivre la pénétration percutanée du Delipidol® est validée.

Tableau IV-10 – Coefficients de corrélation entre les spectres estimés $S_{i=1,\dots,8}$ obtenus par MCR-ALS appliquée avec (c-DEL) et sans (nc-DEL) contrainte d'égalité spectrale sur le Delipidol® pour $K = \{4,8\}$ sur les images Raman des échantillons de peau conservés selon les 3 conditions de stockage et exposés aux Delipidol® pendant 1 et 3 heures.

	$K = 4$				$K = 8$							
	c-DEL											
nc-DEL	S_1 DEL	S_2	S_3	S_4	S_1 DEL	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1	**0.6711**	0.4935	0.7163	0.7175	**0.7956**	0.3915	0.4868	0.4657	0.4639	0.5053	0.6338	0.6053
S_2	0.2313	**0.9616**	0.7003	0.5990	0.1404	**0.9366**	0.4684	0.3749	0.5984	0.6932	0.4771	0.4889
S_3	0.1828	0.5531	**0.9861**	0.7604	0.1802	0.5101	**0.9926**	0.8172	0.4694	0.7294	0.6475	0.7919
S_4	0.2183	0.5469	0.7502	**0.9816**	0.1427	0.3953	0.8145	**0.9853**	0.4981	0.7495	0.6509	0.8461
S_5					0.0932	0.5661	0.4519	0.4462	**0.9652**	0.6041	0.3144	0.6569
S_6					0.2054	0.7522	0.7219	0.7388	0.6441	**0.9969**	0.6964	0.8276
S_7					**0.4937**	0.5034	0.6458	0.6684	0.3327	0.7118	**0.9869**	0.7647
S_8					**0.2800**	0.4785	0.7855	0.8411	0.7017	0.8292	0.7446	**0.9983**

2.3.4. Conclusion

L'application de la MCR-ALS sous contrainte d'égalité spectrale avec $K = 4$ composants a permis d'estimer les profils de pénétration du Delipidol® à travers les couches épidermiques d'échantillons conservés préalablement à température ambiante, à -20°C et à -80°C. Après 1 heure d'exposition, le Delipidol® présente des cinétiques de pénétration non significativement différentes selon la condition de stockage appliquée aux échantillons. Autrement dit, la congélation des échantillons n'a pas altéré les propriétés de barrière cutanée de la peau pour la pénétration cutanée d'une durée d'une heure. En revanche, après 3 heures d'exposition, les quantités de Delipidol® dans le SC et les couches épidermiques plus profondes des échantillons stockés préalablement à -80°C sont significativement plus faibles que celles trouvées dans les échantillons conservés à température ambiante ou à -20°C. En d'autres mots, la congélation des échantillons à -80°C a eu un impact considérable sur l'étude de pénétration du Delipidol® d'une durée de 3 heures. Au final, la différence de contributions relatives considérable dans le SC montre que la perméabilité de l'épiderme a été particulièrement endommagée par le stockage des échantillons à -80°C. Cette conclusion rejoint alors les résultats de J. Nielsen et al. [11]. Toutefois, cette étude a également permis de constater que le stockage des échantillons d'épiderme humain reconstruit à -20°C ne modifie par leurs propriétés de perméabilité qui est indispensable pour que les études de pénétration d'AC soient pertinentes.

D'autre part, cette étude a permis de démontrer l'intérêt de l'approche MCR-ALS pour analyser des images Raman en cosmétologie et plus précisément de souligner la pertinence d'utiliser l'option de supervision spectrale de la MCR-ALS lorsque la détection et le suivi de l'actif étudié peut être difficile dans certaines conditions expérimentales. Son utilisation a notamment permis de confirmer des études préliminaires concernant le stockage des échantillons d'épiderme humain reconstruit qui constitue un sujet incontournable dans le domaine cosmétique.

CONCLUSION

Dans ce chapitre, il a été proposé d'utiliser deux méthodes d'analyses multivariées, la MCR-ALS et la MVC-NMF, pour l'analyse des images Raman. Il a été particulièrement démontré que l'utilisation de ces deux méthodes permet de contourner les limitations des méthodes régulièrement utilisées dans la littérature et qui ont été mentionnées dans le Chapitre II : elles sont toutes les deux applicables avec ou sans connaissances *a priori* sur la composition du mélange étudié et prennent en compte des changements d'expression spectrale d'un composant pur lorsqu'aucune contrainte d'égalité spectrale n'est exercée sur lui. En effet, elles procèdent à un

démélange spectral des données et donc estiment à la fois les spectres de composants purs du mélange et leurs contributions relatives respectives. De plus, les spectres de composants purs estimés peuvent être assimilés à des références spectrales de composants chimiques. Les informations fournies par ces deux méthodes sont donc sensées chimiquement. D'autre part, la répartition de chaque composant pur estimé peut être observée en reconstruisant des cartes de distributions associées en fausses couleurs. Cette visualisation des composants estimés dans les données est facile d'interprétation et représente un argument marketing pour l'industrie cosmétique.

Par ailleurs, il a été démontré que ces deux méthodes de démélange sont applicables sur des données Raman réelles et peuvent être appliquées à des thématiques de recherche cosmétique. Plus encore, les résultats qu'elles fournissent peuvent avoir un fort impact dans le domaine de la recherche et du développement de l'industrie cosmétique. Par exemple, deux études (parties 2.1 et 2.2) ont souligné leur utilité pour le contrôle qualité et/ou la mise au point d'une formulation de produit cosmétique: elles ont montré qu'il était possible de suivre les cinétiques de pénétration d'AC en couplant la technique d'imagerie Raman avec ces deux méthodes d'analyses multivariées. Une troisième étude a permis de parvenir à une conclusion de taille pour l'industrie cosmétique qui réalise des tests à grande échelle et pour qui la question du stockage des échantillons est au cœur des préoccupations : le stockage à -80°C des échantillons de peau humaine reconstruite *in vitro* altère considérablement la fonction barrière cutanée de la peau et engendre une plus grande perméabilité de la peau que le stockage à -20°C, le stockage des échantillons à -20°C est alors préférable.

Toutefois, les trois études menées ont également démontré que les méthodes de démélange spectral appliquées (MCR-ALS et MVC-NMF) permettent de détecter et de suivre la progression d'un AC dans la peau dans une certaine mesure. En effet, cette capacité a été confirmée dans les études 2.1 et 2.3 où la signature de l'actif étudié (le Delipidol®) est forte et en partie détectable dans les données spectrales enregistrées. En revanche, l'étude 2.2 a révélé une incapacité de ces méthodes à suivre sans connaissance *a priori* un actif (la caféine) ayant une plus faible signature spectral et étant beaucoup plus difficilement détectable dans les données spectrales enregistrées. Par conséquent, la capacité de ces méthodes à identifier les différents composants purs dans un mélange semble discutable. Les méthodes ont toutefois démontré leur pertinence dans le cadre d'une analyse semi-supervisée, incorporant la connaissance du spectre de l'AC.

Une difficulté majeure pour le développement de méthodes d'analyses en spectrométrie Raman est qu'à ce jour, il n'existe aucun moyen de vérifier que la décomposition spectrale obtenue par ces méthodes reflète bien la réalité. Autrement dit, aucune vérité de terrain n'est disponible pour évaluer la capacité des deux méthodes à démélanger correctement les données. En effet, la validation des modèles estimés dans chacune de nos études n'a été possible qu'en raison de la connaissance *a priori* sur la nature des actifs cosmétiques présents dans le mélange et la disponibilité de leur référence spectrale avant l'analyse de données. Cependant, il est possible que la nature des actifs suivis ne soient pas toujours connues et/ou que leur référence spectrale ne soit pas disponible avant l'analyse des données, notamment lors des études consacrées à la détection de contrefaçons. De même, des interactions moléculaires peuvent donner lieu localement à l'apparition ou la suppression de composants et ainsi bouleverser la décomposition spectrale des données. Au final, il n'est possible actuellement de valider la décomposition spectrale des données fournies par ces méthodes que si la composition du mélange et les interactions moléculaires engendrées sont en partie connues à l'avance. C'est pourquoi nous

proposons dans le Chapitre V de générer des données de manière synthétique afin de créer une vérité de terrain et y confronter les résultats obtenus par les deux méthodes sur ces données.

Chapitre V.
UNE VÉRITÉ DE TERRAIN POUR ÉVALUER L'EFFICACITÉ DES MÉTHODES DE DÉMÉLANGE

Résumé

Ce chapitre est consacré au développement d'une simulation de données Raman réaliste pour la création de données accompagnées d'une vérité de terrain. L'objectif est de permettre une évaluation de la capacité des méthodes multivariées à estimer les composants purs présents dans un mélange et leurs contributions relatives locales respectives. Pour cela, nous nous basons sur des données spectrales enregistrées à partir de formulations de films de polymères dont la composition qualitative et quantitative est connue et la complexité du mélange spectral est plus simple à étudier. Nous présentons les différentes étapes de génération synthétique de données afin de reproduire au plus près celles enregistrées expérimentalement à partir des films de polymères formulés. Une analyse comparative des résultats obtenus par les deux méthodes d'analyse multivariée est présentée.

Introduction 166
1. Les films de polymères : un modèle de mélange simplifié 167
1.1. Principe des films de polymères 167
1.2. Formulations avec des films de polymères : un support d'évaluation de la capacité des méthodes d'analyse multivariée à démélanger des données spectrales Raman 168
2. Génération de données synthétiques 170
2.1. Génération de mélanges avec une diffusion homogène de l'actif 171
2.1.1. Les références spectrales 171
2.1.2. Processus de génération de données synthétiques 172
2.2. Comparaison des données synthétiques et des données expérimentales 173
3. Évaluation de l'efficacité des méthodes de démélange 175
3.1. Évaluation de la MCR-ALS 175
3.1.1. Mélanges PVA-PNM-DL expérimentaux et synthétiques 176
3.1.1.1. Sans contrainte d'égalité spectrale 176
3.1.1.2. Avec contrainte d'égalité spectrale sur PNM-DL 177
3.1.2. Mélanges PVA-GS-PNM-DL expérimentaux et synthétiques 180
3.1.2.1. Sans contrainte d'égalité spectrale 180
3.1.2.2. Avec contrainte d'égalité spectrale sur le PNM-DL 182
3.2. Évaluation de la MVC-NMF 184
3.2.1. Mélanges PVA-PNM-DL expérimentaux et synthétiques 184
3.2.1.1. Sans contrainte d'égalité spectrale 184
3.2.1.2. Avec contrainte d'égalité spectrale sur le PNM-DL 186
3.2.2. Mélanges PVA-GS-PNM-DL expérimentaux et synthétiques 188
3.2.2.1. Sans contrainte d'égalité spectrale 188
3.2.2.2. Avec contrainte d'égalité spectrale sur le PNM-DL 190
Conclusion 192

INTRODUCTION

Comme présenté dans le Chapitre IV, le couplage de l'imagerie Raman avec les méthodes de démélange spectral que sont la MCR-ALS et la MVC-NMF permet d'obtenir une décomposition qualitative et semi-quantitative des images Raman enregistrées à partir d'échantillons de peau humaine reconstruite exposés à des AC. Ce couplage s'est montré être un outil puissant pour suivre les cinétiques de pénétration cutanée d'ingrédients actifs. Cependant, il a été souligné qu'il n'existe à ce jour aucun moyen de vérifier que la décomposition spectrale obtenue par ces méthodes à l'aide d'une supervision spectrale ou non reflète la réalité, si ce n'est que d'avoir des connaissances *a priori* sur la composition du mélange avant l'application des méthodes sur les données.

Pour évaluer les performances d'une méthode multivariée, il est d'habitude de l'appliquer sur des données simulées. Généralement, plusieurs spectres de composants purs sont sélectionnés dans une base de données puis mélangés selon un processus aléatoire ou contrôlés à l'aide de contributions prédéfinies. La qualité de la décomposition spectrale résultante est ensuite jugée sur la proximité des spectres de composants purs et des concentrations relatives estimées avec ceux et celles utilisés dans la simulation de données. Cependant, une telle simulation de données spectrales d'échantillons de peau n'est pas envisageable. En effet, comme mentionné dans le Chapitre I et le Chapitre II, la peau est un mélange complexe de très nombreux composants dont l'accès à la signature Raman de chacun n'est pas toujours possible. De plus, les contributions relatives locales de la totalité des composants doivent être connues pour simuler des données qui soient réalistes. L'utilisation alternative d'un processus aléatoire pour mélanger les spectres de composants purs est donc inimaginable pour conserver le réalisme de la simulation de données.

Dans la recherche cosmétique, de nombreuses études portent sur la formulation de films de polymère à utiliser comme support pratique d'application d'AC prolongée sur une région cutanée spécifique. Ces formulations comportent généralement un nombre de composants limité et constituent des mélanges moins complexes à étudier que celui de la peau d'autant plus que leur composition qualitative et quantitative exacte est connue. Par conséquent, nous proposons de créer une vérité de terrain sur la base des informations qualitatives et quantitatives de ces formulations. En d'autres mots, nous proposons de reproduire synthétiquement et de manière réaliste des données spectrales enregistrées à partir de formulations de films de polymères. Le but final consiste à évaluer les performances des méthodes MCR-ALS et MVC-NMF pour une application cosmétique en confrontant les résultats obtenus à partir des données générées synthétiquement et expérimentales avec la vérité de terrain.

Ce chapitre est donc structuré de la façon suivante. Une première partie est consacrée à la présentation du principe et de l'intérêt d'utilisation des films de polymères en cosmétique. Elle explique également en détails pourquoi leurs données spectrales Raman associées ont suscité notre intérêt pour la création d'une vérité de terrain en vue d'évaluer les performances des méthodes d'analyse multivariée. Puis, une deuxième partie décrit les différentes étapes à suivre pour générer des données synthétiques qui soient proches des données expérimentales enregistrées à partir des films de polymère et qui challengent les méthodes de démélange spectral. Enfin, une troisième partie présente une analyse comparative des résultats obtenus par les méthodes MCR-ALS et MVC-NMF à partir des données expérimentales et synthétiques avec la vérité de terrain créée.

1. LES FILMS DE POLYMÈRES : UN MODÈLE DE MÉLANGE SIMPLIFIÉ

1.1. PRINCIPE DES FILMS DE POLYMÈRES

Ces dernières années, les films de polymères sont très utilisés dans les domaines de la recherche cosmétique, dermatologique et pharmaceutique. En effet, les polymères sont intégrés en tant qu'agent filmogène dans les formulations des produits dermo-cosmétiques et pharmaceutiques en raison de leurs diverses propriétés : ils ne pénètrent pas dans la peau et agissent seulement en surface. Plus précisément, les films de polymères constituent un réservoir de principes actifs adhésif. Au contact direct de la peau, ils libèrent de manière prolongée et contrôlée des agents actifs sur la zone cutanée traitée. Ils constituent donc un support pratique pour l'application de traitement par perméation cutanée maintenant les effets des principes actifs qu'ils contiennent sur une longue période. Par conséquent, les films de polymères représentent dans le domaine de la cosmétique un moyen de favoriser la pénétration cutanée des AC dans la peau. Ils sont utilisés par exemple pour développer des masques antioxydants contre le vieillissement cutanée [107]. Dans les domaines pharmaceutique et dermatologique, ils peuvent même être considérés comme une alternative aux traitements administrés par voie orale en évitant l'apparition d'effets secondaires qui leurs sont associés. Par exemple, ils sont employés pour traiter des maladies de la peau [108], [109], pour soigner et aider la cicatrisation des plaies [110], [111] ou même encore de manière préventive pour protéger l'organisme contre des agressions chimiques externes telles que l'intrusion de pesticides ou d'agents chimiques hautement toxiques de guerre par voie cutanée [112].

Les polymères les plus fréquemment utilisés pour obtenir des films de polymères sont l'alcool polyvinylique (PVA), le polyvinyl pyrrolidone (PVP), les polysaccharides ou encore l'hydroxylpropyl cellulose (HPC). Ces polymères filmogènes sont dissous dans un ou plusieurs solvants volatiles avec des principes actifs en solution ou en dispersion sous forme de microparticules/nanoparticules de molécules encapsulées. Au contact de la peau, le solvant s'évapore pour former un film chargé en principes actifs comme l'illustre la Figure V-1. Au final, un film de polymère se révèle être de manière simplifiée un mélange de trois composants principaux : un agent filmogène, un solvant et un principe actif, qui une fois au contact de la peau libère de manière contrôlée et prolongée le principe actif à travers elle.

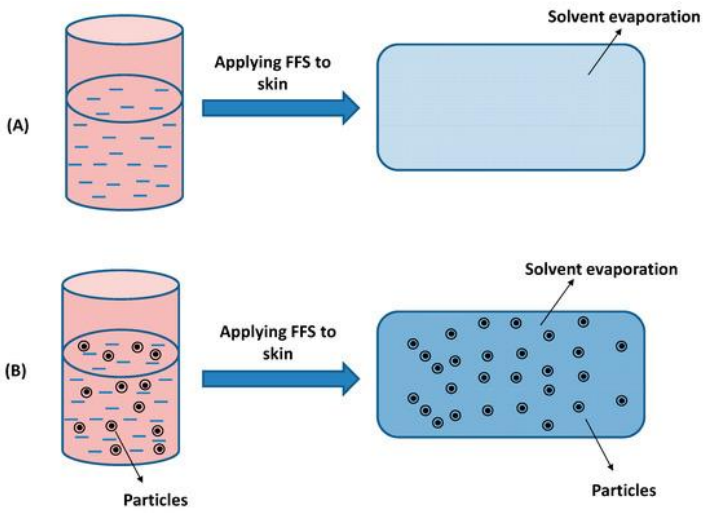


Figure V-1 – Illustration de la formation d'un film contenant une solution (**A**) *ou une dispersion de particules* (**B**) [113] *; FFS pour Film formation of Film-forming Systems.*

1.2. FORMULATIONS AVEC DES FILMS DE POLYMÈRES : UN SUPPORT D'ÉVALUATION DE LA CAPACITÉ DES MÉTHODES D'ANALYSE MULTIVARIÉE À DÉMÉLANGER DES DONNÉES SPECTRALES RAMAN

Dans le projet MISTIC, une partie du développement d'une nouvelle génération de produits cosmétiques se concentre sur la formulation de nano-dispersions d'AC à introduire en concentrations élevées dans des films de polymères pour libérer les molécules actives dans la peau par stimulation contrôlée. Par exemple, l'AC du Delipidol® a été encapsulé (noté PNM-DL pour *Polymeric Nano-Micelles loaded with Delipidol®*) en solution à 1%, 2.5%, 5%, 7.5% et 10% de concentration dans des films de polymères composés de PVA afin d'évaluer sa capacité de diffusion. L'homogénéité de la diffusion de l'actif encapsulé dans les films de polymères et sa capacité de diffusion ont ensuite été confirmés en appliquant la méthode NCLS (Figure V-2) sur les données Raman acquises à partir des échantillons de films de polymères. Les acquisitions Raman ont été obtenues en utilisant un microspectromètre Alpha300 R (WItec, Ulm, Allemagne) avec un laser de 532 nm à 20 mW au niveau de l'échantillon, un objectif x50 (Zeiss, LD Epiplan HD, N.A 0.5) et une résolution spectrale de 4 cm^{-1}.

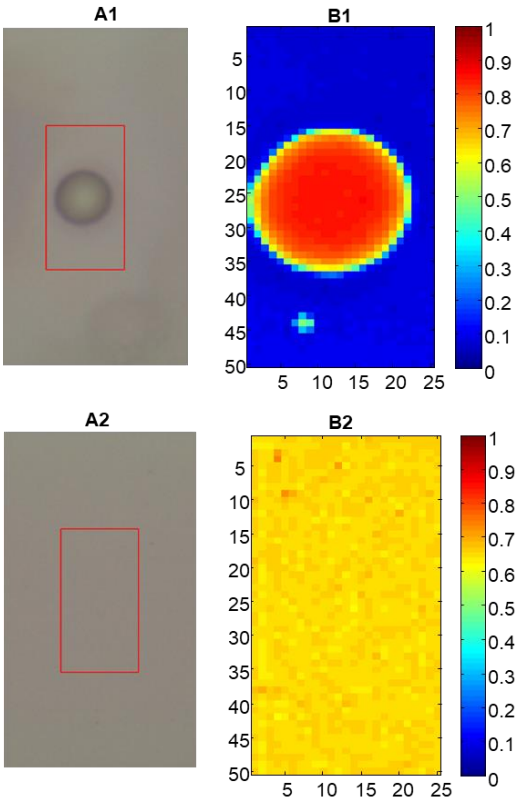


Figure V-2 – **A1**-**A2** *: Images optiques et* **B1**-**B2** *: cartes de distribution du Delipidol® reconstruites après analyse NCLS des films de PVA.* **B1** *: Delipidol® libre ;* **B2** *: Delipidol® encapsulé. L'échelle de couleur représente la contribution relative du spectre de Delipidol® dans les spectres Raman à chaque pixel. Les images Raman sont de taille* 25×50 *µm² avec une résolution spatiale de 1 µm.*

Ces films de polymères constituent un support d'étude intéressant pour l'évaluation de l'efficacité des méthodes de démélange spectral. En effet, ces films de polymères sont conçus à partir d'un mélange « simple » de trois composants : le PVA, le Delipidol® encapsulé et l'eau utilisée comme solvant. Le Tableau V-1 présente les concentrations respectives de chaque composant utilisé dans

la formulation des films de polymères. La formulation de ces films de polymères (renommés PVA-PNM-DL) n'est alors autre qu'une combinaison linéaire de ces trois composants pondérés par leurs concentrations associées. Au final, la composition tant qualitative que quantitative de ces mélanges est connue : la nature des composants et leur expression Raman mais aussi leurs concentrations après évaporation du solvant sont accessibles. Ces informations servent donc d'éléments de comparaison pour juger la qualité des décompositions spectrales obtenues par les méthodes multivariées MCR-ALS et MVC-NMF appliquées sur les images Raman enregistrées à partir des films de polymères. Cette comparaison permet par conséquent de mesurer la performance de ces méthodes de démélange pour l'évaluation de l'innocuité et de l'efficacité de ces nouvelles formes de cosmétiques développées.

Tableau V-1 – Concentrations des composants utilisés dans la formulation de films de polymères « simples » avant et après évaporation du solvant.

Films « simples »	**Avant évaporation du solvant**			**Après évaporation du solvant**		
Concentration du PNM-DL en solution	**PNM-DL (g)**	**PVA (g)**	**Eau (g)**	**PNM-DL (%)**	**PVA (%)**	Ratio PNM-DL/PVA
1%	0.1032	1.0045	8.9084	0.98	99.02	0.0098931
2.5%	0.2555	1.0034	8.7562	2.39	97.61	0.0245199
5%	0.4971	1.004	8.5053	4.55	95.45	0.0476774
7.5%	0.7510	1.0107	8.2503	6.68	93.32	0.0715517
10%	1.0053	1.0000	8.0100	8.83	91.17	0.0973169

Toutefois, la simplicité des films de polymères précédents formulés avec seulement deux composants autres que le solvant ne challenge pas assez les méthodes d'analyse multivariée utilisées pour pouvoir juger de leur efficacité convenablement. Par conséquent, une seconde campagne de formulations de films de polymères a été menée. Pour complexifier les mélanges, deux autres composants ont été ajoutés au PVA et au PNM-DL dans les formulations :

- Le glycérol (noté GLY) : un humectant qui confère aux produits une onctuosité et les préserve du dessèchement et qui est très utilisé dans les formulations cosmétiques.
- Le Sépicide^{TM} HB (noté SEPI) : un agent conservateur également très utilisé dans les formulations cosmétiques.

À nouveau, seule la concentration du PNM-DL varie d'une formulation à une autre : 1%, 2.5%, 5%, 7.5% et 10%. Le Tableau V-2 présente les quantités respectives de chaque composant utilisé dans la formulation des films de polymères « complexes ». La formulation de ces films de polymères complexes (renommés PVA-GS-PNM-DL) repose à nouveau sur un mélange linéaire des cinq composants (solvant inclus) pondérés par leurs concentrations associées où les références spectrales des composants sont connues et leurs quantités relatives après évaporation du solvant sont déduites. Par conséquent, ces informations peuvent être considérées comme point de comparaison pour évaluer la pertinence des décompositions spectrales proposées par les méthodes multivariées MCR-ALS et MVC-NMF.

Tableau V-2 – Concentrations des composants utilisés dans la formulation de films de polymères « complexes » avant et après évaporation du solvant.

Films « complexes »	Avant évaporation du solvant					Après évaporation du solvant				
Concentration du PNM-DL en solution	PNM-DL (g)	PVA (g)	GLY (g)	SEPI (g)	Eau (g)	PNM-DL (%)	PVA (%)	GLY (%)	SEPI (%)	Ratio PNM-DL / PVA
1%	0.1150	1.0060	1.0140	0.1027	7.8020	0.52	47.16	47.51	4.81	0.011
2.5%	0.2619	1.0010	0.9957	0.1023	7.6490	1.19	47.12	46.88	4.82	0.025
5%	0.5359	1.0020	1.0020	0.1010	7.3980	2.39	46.46	46.46	4.68	0.052
7.5%	0.7424	1.0010	1.0147	0.1017	7.1490	3.27	45.72	46.37	4.65	0.071
10%	1.0218	1.0030	1.0031	0.1020	6.8960	4.46	45.46	45.45	4.62	0.098

2. GÉNÉRATION DE DONNÉES SYNTHÉTIQUES

Comme la composition exacte de la formulation des films de polymères est connue, les concentrations relatives de chaque composant après évaporation du solvant sont déduites. Elles constituent donc une référence de composition basée sur la théorie de la formulation des films de polymères pour évaluer l'efficacité des méthodes de démélange. Toutefois, en imagerie Raman, la sensibilité Raman peut être différente d'un composant à un autre et des interactions moléculaires peuvent survenir suite au mélange des composants et aboutir à une modification de l'expression Raman d'un d'entre eux dans le mélange (par exemple, un décalage Raman ou une déformation d'une bande caractéristique). Par conséquent, d'un point de vue spectral, les compositions tant qualitative que quantitative des mélanges peuvent ne pas être aussi évidentes à dissocier à travers la réponse Raman de ces films. C'est pourquoi, pour assurer la validité des proportions et de la référence spectrale de chaque composant dans l'expression Raman du mélange, nous proposons dans ce chapitre de générer numériquement des données spectrales synthétiques sans se soucier des éventuelles perturbations spectrales pouvant survenir suite à des interactions moléculaires. De cette manière, les références spectrales des composants et leurs concentrations relatives respectives utilisées dans la génération de ces données synthétiques sous forme de mélange linéaire forment une vérité de terrain avérée face aux résultats obtenus par l'application des méthodes d'analyse multivariée MCR-ALS et MVC-NMF sur ces mêmes données. Ainsi, la comparaison des décompositions spectrales obtenues avec la vérité de terrain permet de mesurer la performance de ces méthodes multivariées dans l'évaluation de l'innocuité et de l'efficacité des nouvelles formes de cosmétiques développées.

Dans cette partie, nous proposons un algorithme simple qui permet de générer des données Raman synthétiques en reproduisant le mélange supposé linéaire des différents composants chimiques utilisés dans la formulation des films PVA-PNM-DL et PVA-GS-PNM-DL. Ces images Raman synthétiques sont ensuite comparées aux images Raman acquises à partir des films simples et complexes expérimentaux.

2.1. GÉNÉRATION DE MÉLANGES AVEC UNE DIFFUSION HOMOGÈNE DE L'ACTIF

2.1.1. Les références spectrales

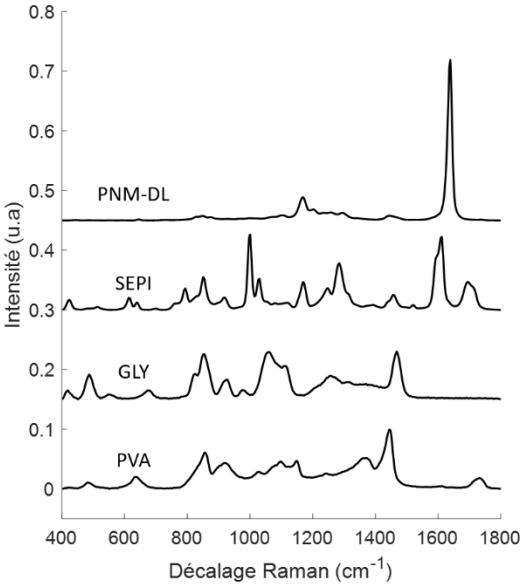


Figure V-3 Spectres Raman des composants de références utilisés dans la formulation des films simples et complexes. Les spectres sont volontairement décalés pour plus de lisibilité. PVA : Polyvinylic alcohol, GLY : Glycérol, SEPI : Sépicide™ HB, PNM-DL : Polymeric Nano-Micelles loaded with Delipidol®.

Les données synthétiques sont créées sur la base d'un mélange linéaire des 4 composants utilisés dans les formulations, à savoir le PVA, le GLY, le SEPI et enfin le PNM-DL. Pour chaque composant, une dizaine de spectres ont été acquis et moyennés. Leurs spectres de référence moyens sont illustrés dans la Figure V-3 [114]:

- les bandes spectrales principales du PVA apparaissent à 633 cm^{-1} (déformation O–H), 855 cm^{-1} (élongation C–C), 1128 cm^{-1} (élongation C–O et déformation H–O), 1360 cm^{-1} (déformations C–H et O–H) et 1440 cm^{-1} (déformations C–H et O–H) ;

- les caractéristiques spectrales principales du GLY s'expriment par deux larges bandes situées dans la région spectrale [1000-1500] cm^{-1} (élongation CCO, torsion CH_2 et élongation COH). D'autres bandes se situent à 1060 cm^{-1} et 1111 cm^{-1} (élongation symétrique des groupes –OH respectivement primaires et secondaires) et dans la région [1207-1310] cm^{-1} et à 1468 cm^{-1} (torsion CH_2) ;

- les bandes spectrales caractéristiques de SEPI apparaissent à 1001 cm^{-1} (cycles aromatiques du phenoxyéthanol et du parabène), à 1285 cm^{-1} (torsion CH_2, déformations C–H et du cycle aromatique du phenoxyéthanol), à 1612 cm^{-1} (déformations C–H et du cycle aromatique du phenoxyéthanol) et à 1694 cm^{-1} (élongation C=O du parabène) ;

- les bandes spectrales caractéristiques du PNM-DL sont principalement les mêmes que celles du Delipidol® libre : une bande intense à 1169 cm^{-1} (élongations C–O et C–H) et des bandes de plus faibles intensités sur la région spectrale [800-1450] cm^{-1}, à savoir 644 cm^{-1} et 846 cm^{-1} (élongation C–C), 1114 cm^{-1} (élongations C–O et C–C), 1202 cm^{-1} (élongation C–C), 1257 cm^{-1} (élongation C–C, déformation =CH), 1294 cm^{-1} (déformations C–OH et C–C–H, torsion CH_2), 1440 cm^{-1} (cisaillement

CH_2, déformation asymétrique CH_3). En revanche, la bande intense caractéristique de l'élongation C=C originellement située à 1636 cm^{-1} sur le spectre du Delipidol® libre est décalée à 1640 cm^{-1} ;

2.1.2. Processus de génération de données synthétiques

En plus de l'acquisition Raman des films simples et complexes, des acquisitions d'images Raman de films témoins (c'est-à-dire composés uniquement de PVA) ont également été enregistrées. Rappelons que toutes ces images Raman ont été obtenues en utilisant un microspectromètre Alpha300 R (WItec, Ulm, Allemagne) avec un laser de 532nm de puissance 20mW au niveau de l'échantillon et un objectif x50 (NA 0.5) (Zeiss, LD Epiplan HD, N.A 0.5). Elles ont une dimension de 100×100 µm² avec une résolution spatiale de 4 µm dans les deux directions x et y et une résolution spectrale d'environ 4 cm^{-1}.

Ces images Raman de films témoins constituent le point de départ pour générer les données synthétiques. En effet, chaque spectre de cette image qui est composé uniquement de PVA est mélangé aux 3 autres composants (GLY, SEPI, PNM-DL) selon leurs concentrations relatives déduites respectives utilisées dans la formulation des films PVA-PNM-DL et PVA-GS-PNM-DL (Tableaux V-1 et V-2).

- Les concentrations locales de chacun des composants sont supposées spatialement homogènes. Les mêmes concentrations de composants sont donc utilisées dans la génération de chaque pixel de l'image synthétique.

- Les mélanges sont réalisés à partir des spectres contrôles et des spectres de référence débruités et corrigés de leur ligne de base par les méthodes retenues dans le Chapitre III.

- Rappelons que chaque composant possède une section efficace de diffusion Raman différente : le PNM-DL a par exemple une plus forte réponse Raman que le PVA ce qui explique que la norme de son spectre de référence (1.5375×10^5) est supérieure à celle du spectre de référence du PVA (7.2928×10^3). Cette différence de sensibilité de la spectroscopie Raman est prise en compte dans la conception du mélange synthétique en faisant le choix d'utiliser les spectres corrigés mais non normalisés. Dans le cas où les spectres seraient utilisés sous leur forme normalisée, il serait nécessaire de les multiplier par leur norme respective.

- Enfin, un bruit gaussien et une ligne de base sont rajoutés à chaque spectre synthétique exprimant le mélange des composants.

Au final, chaque spectre $\widehat{D}_i$ de l'image générée $\widehat{D}$ est calculé de la sorte :

PVA-PNM-DL: $$\widehat{D}_i = C_{PVA,i} \times D_{CTRL,i} + C_{PNMDL,i} \times S_{PNMDL} + E_i + L_{B,i} \quad \text{(V-1)}$$

PVA-GS-PNM-DL: $$\widehat{D}_i = C_{PVA,i} \times D_{CTRL,i} + C_{PNMDL,i} \times S_{PNMDL} + C_{GLY,i} \times S_{GLY} + C_{SEPI,i} \times S_{SEPI} + E_i + L_{B,i} \quad \text{(V-2)}$$

où $D_{CTRL,i}$ est le spectre n°i de l'image Raman contrôle, S_{PNMDL}, S_{GLY} et S_{SEPI} sont les spectres de référence des autres composants, $C_{PVA,i}$, $C_{PNMDL,i}$, $C_{GLY,i}$ et $C_{SEPI,i}$ sont les concentrations de chacun des composants à la position n°i dans l'image générée, E_i un bruit gaussien et $L_{B,i}$ une ligne de base. Le processus de génération des données synthétiques est schématisé dans la Figure V-4.

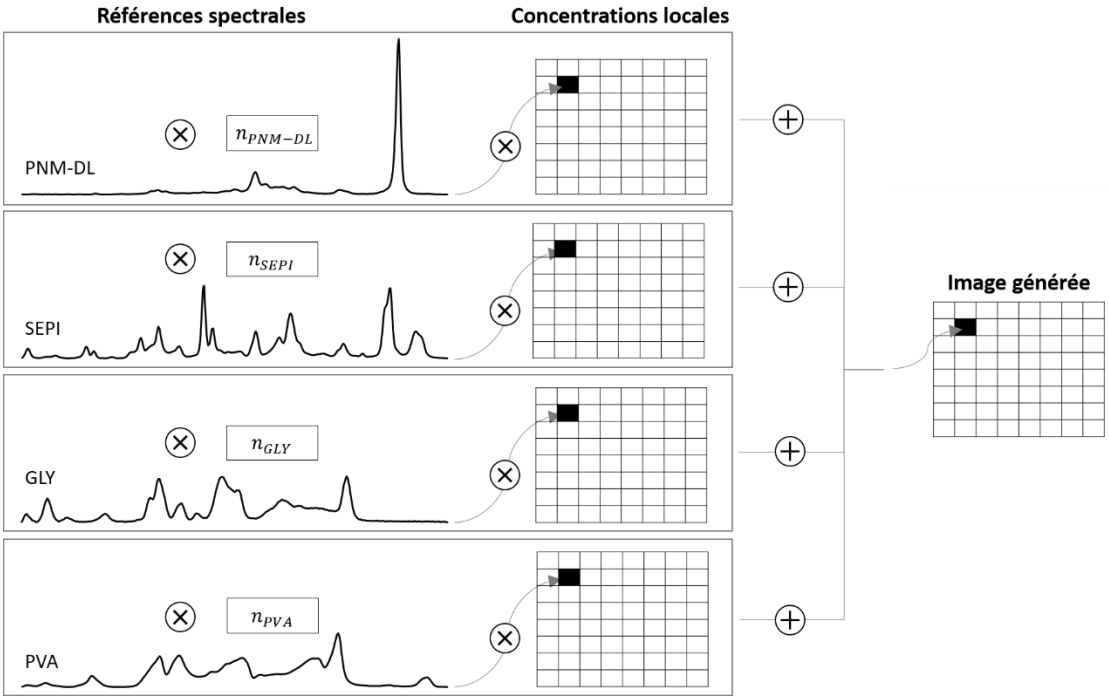


Figure V-4 – Processus de génération de données synthétiques simulant les films PVA-PNM-DL et PVA-GS-PNM-DL ; n_{PVA} , n_{GLY}, n_{SEPI} et n_{PNMDL} les normes ℓ_2 respectives des composants PVA, GLY, SEPI et PNMDL.

2.2. COMPARAISON DES DONNÉES SYNTHÉTIQUES ET DES DONNÉES EXPÉRIMENTALES

Pour pouvoir réaliser une comparaison, les données générées synthétiquement sont soumises aux mêmes prétraitements que les données spectrales obtenues expérimentalement. Les méthodes de prétraitement spectral utilisées sont celles retenues dans le Chapitre III. Les Figure V-5, **A** et Figure V-5, **B** illustrent la comparaison des spectres générés synthétiquement (rouge) avec les spectres acquis à partir des films formulés (noir) selon les cinq concentrations du PNM-DL pour les formulations PVA-PNM-DL et PVA-GS-PNM-DL. Alors que les spectres synthétiques apparaissent comme très proches des spectres expérimentaux quelle que soit la concentration du PNM-DL pour les mélanges PVA-PNM-DL, des différences spectrales notables sont observées entre ceux issus des mélanges PVA-GS-PNM-DL :

- Les bandes caractéristiques du SEPI situées à 1001 cm^{-1} et 1612 cm^{-1} et la bande caractéristique du PNM-DL située à 1640 cm^{-1} sont plus intenses dans les spectres générés synthétiquement pour chaque mélange PVA-GS-PNM-DL (Figure V-5, **B**). Des calculs de ratios de bandes, à savoir celles du PVA à 1440 cm^{-1} et du PNM-DL à 1640 cm^{-1}, permettent également de confirmer l'importance de la différence d'intensité de la bande caractéristique du PNM-DL (Figure V-6). En effet, bien que l'augmentation des ratios de bandes reste proportionnelle à l'augmentation de la concentration du PNM-DL dans chacun des mélanges PVA-PNM-DL/PVA-GS-PNM-DL expérimentaux ou synthétiques, la différence de ratios est plus importante entre les mélanges PVA-GS-PNM-DL expérimentaux et synthétiques qu'entre les mélanges PVA-PNM-DL expérimentaux et synthétiques (Figure V-6, courbes discontinues).

- D'autre part, l'épaulement marqué par la bande caractéristique du GLY à 1468 cm^{-1} est absent dans les spectres synthétiques des mélanges PVA-GS-PNM-DL (Figure V-5, **B**).

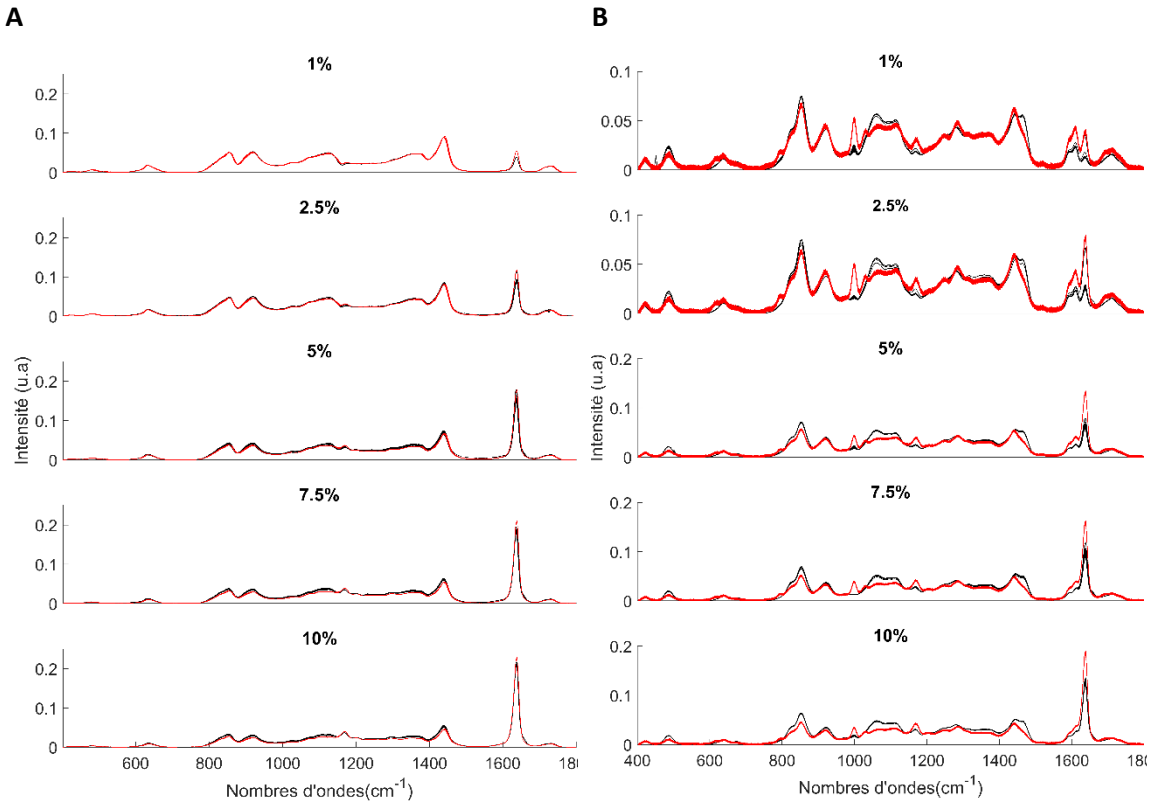


Figure V-5 – Comparaison des spectres des films (A) PVA-PNM-DL et (B) PVA-GS-PNM-DL obtenus expérimentalement (noir) et synthétiquement (rouge) avec une concentration progressive en PNM-DL de 1%, 2.5%, 5%, 7.5% et 10% (de haut en bas).

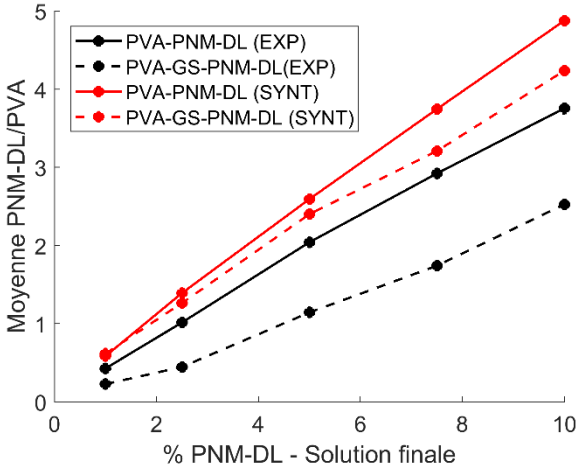


Figure V-6 – Ratios de bandes moyens du PNM-DL (1640 cm^{-1}) et du PVA (1440 cm^{-1}) pour les spectres expérimentaux (noir) et synthétiques (rouge) des films PVA-PNM-DL (courbes continues) et PVA-GS-PNM-DL (courbes discontinues).

D'autres critères statistiques tels que la corrélation ou l'erreur quadratique moyenne (RMSE pour *Root Mean Squared Error*) indiquent également que les spectres générés synthétiquement sont plus proches des spectres expérimentaux dans les mélanges PVA-PNM-DL que dans les mélanges PVA-GS-PNM-DL :

- La corrélation moyenne entre les images synthétiques et les images expérimentales des formulations PVA-PNM-DL étant supérieure à 0.98 quelle que soit la concentration de PNM-DL confirme leur forte proximité (Figure V-7, **A**). En revanche, bien qu'elle reste élevée (supérieure à 0.85), la corrélation entre les images synthétiques et les images expérimentales des formulations

PVA-GS-PNM-DL reflète une moins forte proximité (Figure V-7, **A**). Pour cause, les spectres synthétiques souffrent d'un manque d'informations sur le GLY et d'une surexpression de l'information concernant le SEPI et le PNM-DL.

- Le RMSE entre les images synthétiques et les images expérimentales est inférieur à 0.004 pour les formulations PVA-PNM-DL et inférieur à 0.01 pour les formulations PVA-GS-PNM-DL (Figure V-7, **B**). En d'autres termes, les différences spectrales dans les données PVA-PNM-DL ne représentent que 20% de la moyenne des données spectrales expérimentales alors que celles présentes dans les données PVA-GS-PNM-DL représentent 50% de la moyenne des données spectrales expérimentales.

Nous considérons que les différences spectrales entre les formulations PVA-GS-PNM-DL expérimentales et synthétiques proviennent non seulement de la difficulté à considérer pleinement les différences de section efficace des composants mais aussi du fait que les

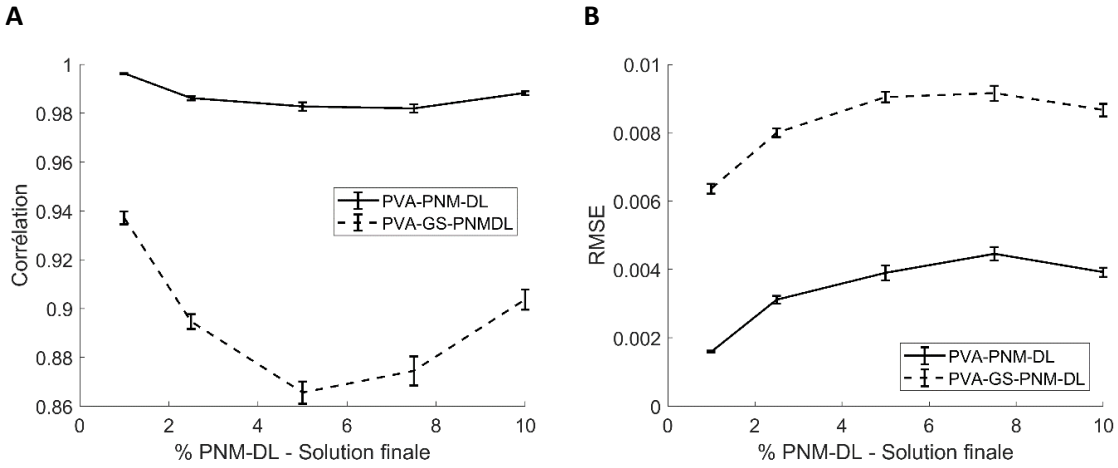


Figure V-7 – (**A**) *Coefficients de corrélation moyens et* (**B**) *erreurs quadratiques moyennes entre les spectres expérimentaux et synthétiques des films PVA-PNM-DL (courbe continue) et PVA-GS-PNMD-DL (courbe discontinue).*

interactions moléculaires entre composants n'ont pas été prises en compte dans la génération des données synthétiques. Au final, la génération de données synthétiques a montré que de nombreux paramètres doivent être pris en compte dans le démélange spectral : les interactions moléculaires, les différences de sensibilité Raman et la concentration de chaque composant. Toutefois, bien que les données synthétiques ne reproduisent pas exactement les données expérimentales, elles constituent une véritable vérité de terrain pour évaluer les performances des méthodes multivariées telles que MCR-ALS et MVC-NMF sur des modèles linéaires simples.

3. ÉVALUATION DE L'EFFICACITÉ DES MÉTHODES DE DÉMÉLANGE

3.1. ÉVALUATION DE LA MCR-ALS

Pour évaluer sa capacité à retrouver les signatures spectrales des composants purs présents dans un mélange et à estimer leurs concentrations relatives respectives, la méthode MCR-ALS est appliquée d'une part sur les images Raman acquises à partir des formulations des films PVA-PNM-DL et PVA-GS-PNM-DL et d'autre part sur les images Raman générées synthétiquement qui reproduisent les mélanges supposés linéaires des formulations des films PVA-PNM-DL et PVA-GS-PNM-DL. Une comparaison entre les spectres de composants purs estimés par la MCR-ALS et les

spectres de référence des composants utilisés dans les formulations et les générations de mélanges PVA-PNM-DL et PVA-GS-PNM-DL ainsi qu'une analyse des résidus des modèles estimés permettent d'examiner ses performances.

3.1.1. Mélanges PVA-PNM-DL expérimentaux et synthétiques

3.1.1.1. Sans contrainte d'égalité spectrale

Les décompositions spectrales obtenues en appliquant la MCR-ALS sans contrainte d'égalité spectrale avec $K = 2$ composants (PVA, PNM-DL) sur les données expérimentales et synthétiques des films PVA-PNM-DL sont présentées dans la Figure V-8.

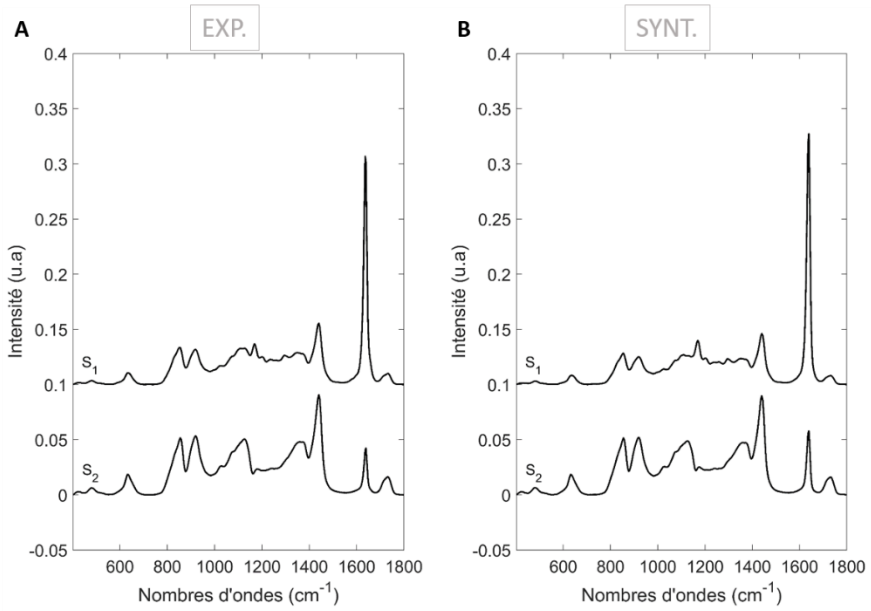


Figure V-8 – Décompositions spectrales par la méthode MCR-ALS appliquée avec uniquement des contraintes de non-négativité et de normalisation sur les spectres et les concentrations relatives pour $K = 2$ composants sur les images Raman des films PVA-PNM-DL obtenues de manière (**A**) *expérimentale et* (**B**) *synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques.*

- Les spectres de composants purs estimés à partir des données expérimentales ne correspondent pas aux spectres de références des composants utilisés dans la formulation des films (Figure V-8, **A**). Bien qu'il présente en majorité les caractéristiques spectrales du PNM-DL (corrélation de 0.8719, Tableau V-3, **B**), le spectre S_1 exprime aussi des bandes caractéristiques du PVA (corrélation de 0.4213, Tableau V-3, **B**). De même, le spectre S_2 présente aussi bien des bandes Raman associées au PVA (corrélation de 0.9783, Tableau V-3, **B**) qu'au PNM-DL (corrélation de 0.1361, Tableau V-3, **B**). La décomposition spectrale obtenue à partir des données expérimentales dans des conditions non supervisées est donc loin de la réalité concernant la formulation des films PVA-PNM-DL.

- Les spectres de composants purs estimés à partir des données synthétiques sont très proches de ceux obtenus à partir des données expérimentales (Figure V-8, **B**): la corrélation entre les premiers spectres de composants estimés S_1 s'élève à 0.9879 et celle entre les seconds spectres estimés S_2 est de 0.9960 (Tableau V-3, **A**). Par conséquent, leur proximité avec la vérité de terrain est toute aussi faible. En effet, à la fois S_1 et S_2 correspondent à des mélanges des composants de référence utilisés dans la formulation des données synthétiques des films PVA-PNM-DL: les corrélations de S_1 avec les spectres de référence du PNM-DL et du PVA sont respectivement de 0.9327 et 0.2935, et celles de S_2 avec ces mêmes références spectrales s'élèvent dans l'ordre à 0.2214 et 0.9570 (Tableau V-3, **B**). Au final, devant la vérité de terrain avérée des données synthétiques des films

PVA-PNM-DL, l'application de la MCR-ALS dans des conditions d'application non supervisée sur les spectres de composants purs à estimer n'est pas assez performante pour obtenir la véritable décomposition spectrale des données.

Tableau V-3 – Coefficients de corrélation entre (**A**) *les spectres estimés $S_{i=1,2}$ obtenus par MCR-ALS appliquée sans contrainte d'égalité spectrale sur les données PVA-PNM-DL expérimentales et synthétiques et* (**B**) *les spectres estimés $S_{i=1,2}$ et les références spectrales du PVA et du PNM-DL ; EXP. : expérimentales, SYNT. : synthétiques.*

A	**EXP.**	
SYNT.	S_1	S_2
S_1	0.9879	0.4840
S_2	0.6653	0.9960

B	**EXP.**		**SYNT.**	
Références	S_1	S_2	S_1	S_2
PVA	0.4213	0.9783	0.2935	0.9570
PNM-DL	0.8719	0.1361	0.9327	0.2214

En définitive, sans aucune contrainte d'égalité spectrale, la MCR-ALS est incapable de retrouver distinctement les spectres du PVA et du PNM-DL quelques soient les images Raman analysées. Alors que le mélange est ici constitué seulement de deux composants, les performances de la MCR-ALS sont incertaines.

3.1.1.2. Avec contrainte d'égalité spectrale sur PNM-DL

Afin de la guider dans le démélange des spectres de composants purs contenus dans les films PVA-PNM-DL à partir des données Raman expérimentales et synthétiques, une contrainte d'égalité spectrale sur le PNM-DL a été intégrée dans l'algorithme de la méthode MCR-ALS.

Les décompositions spectrales obtenues en appliquant la MCR-ALS avec cette contrainte d'égalité spectrale sur le PNM-DL et $K = 2$ composants (PVA, PNM-DL) sur les données expérimentales et synthétiques des films PVA-PNM-DL sont présentées dans la Figure V-9. Ces décompositions sont très similaires : alors que le spectre S_1 est contraint à être la référence spectrale du PNM-DL dans les deux modèles, le spectre S_2 est très proche de la signature spectrale du PVA dans les deux décompositions spectrales. En effet, sa corrélation avec le PVA s'élève respectivement à 0.9996 et 0.9999 pour les estimations obtenues à partir des données expérimentales et des données synthétiques (Tableau V-4). De plus, les spectres S_2 ne contiennent aucune information supplémentaire concernant le PNM-DL comme en atteste leur très faible corrélation: -0.0837 pour l'estimation réalisée sur les données expérimentales et -0.0822 pour celle obtenue à partir des données synthétiques (Tableau V-4). Celles-ci sont d'ailleurs comparables à la corrélation entre les références spectrales du PVA et du PNM-DL qui est de -0.0708. Au final, les spectres de composants purs estimés à partir des données expérimentales et des données synthétiques correspondent de très près aux spectres de référence des composants utilisés dans la formulation des films PVA-PNM-DL et dans la génération synthétique de leur expression spectrale Raman.

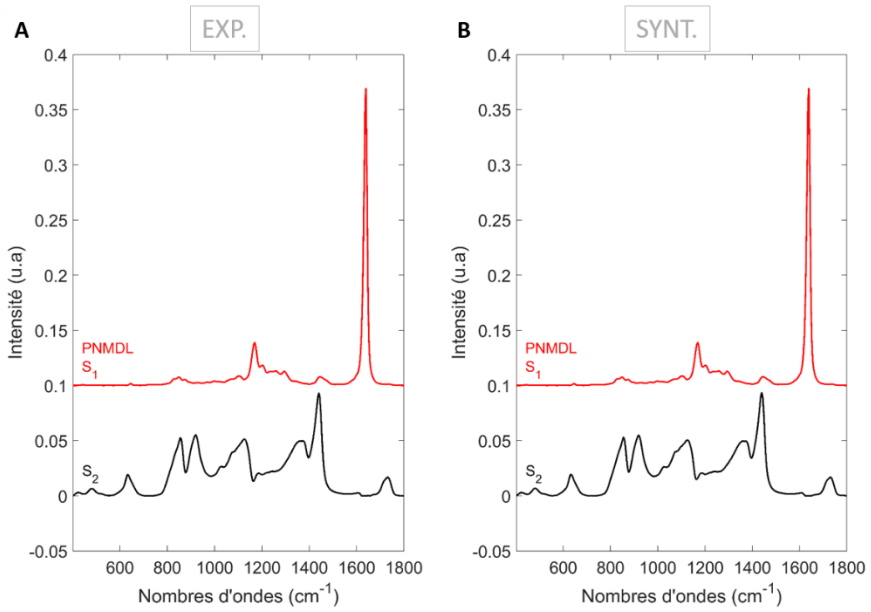


Figure V-9 – Décompositions spectrales par la méthode MCR-ALS appliquée avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $K = 2$ composants et une contrainte d'égalité spectrale sur le PNM-DL sur les images Raman des films PVA-PNM-DL obtenues de manière (A) expérimentale et (B) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques.

Tableau V-4 – Coefficients de corrélation entre (**A**) *les spectres estimés $S_{i=1,2}$ obtenus par MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le PNM-DL sur les données PVA-PNM-DL expérimentales et synthétiques et* (**B**) *les spectres estimés $S_{i=1,2}$ et les références spectrales du PVA et du PNM-DL ; EXP. : expérimentales, SYNT. : synthétiques.*

A	**EXP.**	
SYNT.	**S_1 PNM-DL**	**S_2**
S_1 PNM-DL	1.0000	-0.0837
S_2	-0.0822	0.9998

B	**EXP.**		**SYNT.**	
Références	**S_1 PNM-DL**	**S_2**	**S_1 PNM-DL**	**S_2**
PVA	-0.0708	0.9996	-0.0708	0.9999
PNM-DL	1.0000	-0.0837	1.0000	-0.0822

En raison de la forte correspondance entre les spectres de composants purs estimés avec les spectres de référence des composants utilisés quelles que soient les données analysées, l'évolution de la concentration du PNM-DL dans les films PVA-PNM-DL expérimentaux et synthétiques est observable en reconstruisant les cartes de distribution qui lui sont associées. Les cartes de distribution de la concentration relative estimée du PNM-DL dans les films PVA-PNM-DL expérimentaux et synthétiques sont illustrées dans la Figure V-10. Quelle que soit l'origine des données spectrales analysées, la répartition du PNM-DL est homogène et sa concentration relative estimée augmente progressivement avec le pourcentage de PNM-DL introduit dans les formulations ou la génération synthétique des mélanges PVA-PNM-DL. Cependant, les moyennes des concentrations relatives du PNM-DL estimées à partir des données expérimentales se montrent un peu moins élevées que celles estimées à partir des données synthétiques (Figure V-11). D'autre part, l'observation des moyennes des concentrations relatives du PNM-DL estimées permet de constater également une perte de linéarité dans leur estimation en comparaison aux quantités réellement utilisées quelles que soient les données analysées. Cette perte de linéarité dans l'estimation des concentrations relatives du PNM-DL s'explique notamment par la présence de la signature spectrale du PNM-DL dans les résidus des modèles estimés par la MCR-ALS (Figure V-12). Cette présence signifie qu'une partie de l'information concernant ce composant n'a pas été

reproduite par les modèles estimés et par conséquent que les estimations de concentrations relatives associées sont erronées.

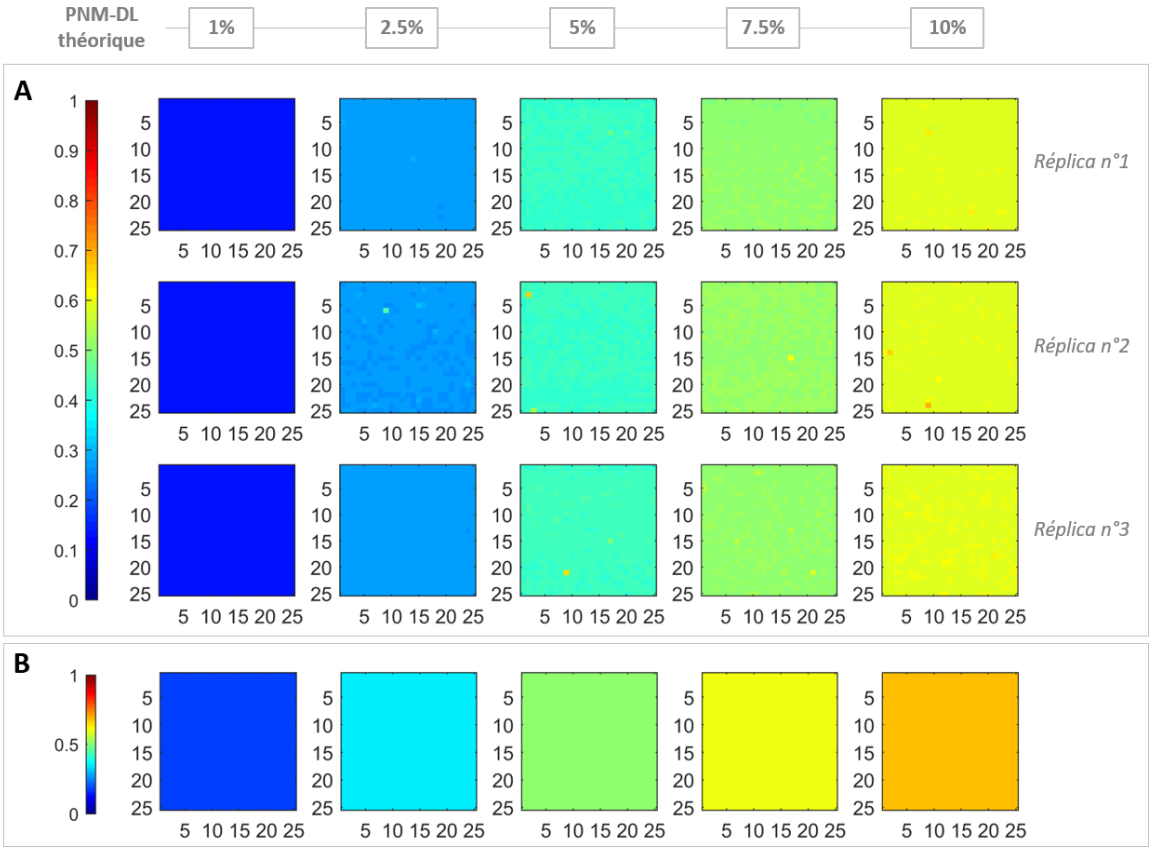


*Figure V-10 – Cartes de distribution du PNM-DL dans les images Raman de films PVA-PNM-DL (**A**) expérimentales et (**B**) synthétiques. Cartes reconstruites à partir du spectre S_1 contraint à correspondre au spectre du PNM-DL dans l'application de la MCR-ALS avec $K = 2$ composants. Les axes x et y des cartes de distribution sont exprimées en pixels.*

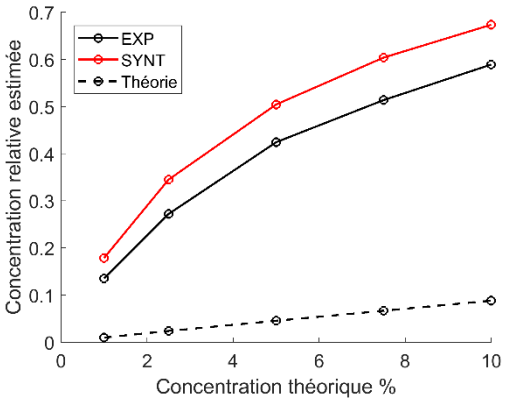


Figure V-11 – Comparaison des concentrations relatives moyennes du PNM-DL estimées par MCR-ALS sous contrainte d'égalité spectrale sur le PNM-DL à partir des données PVA-PNM-DL

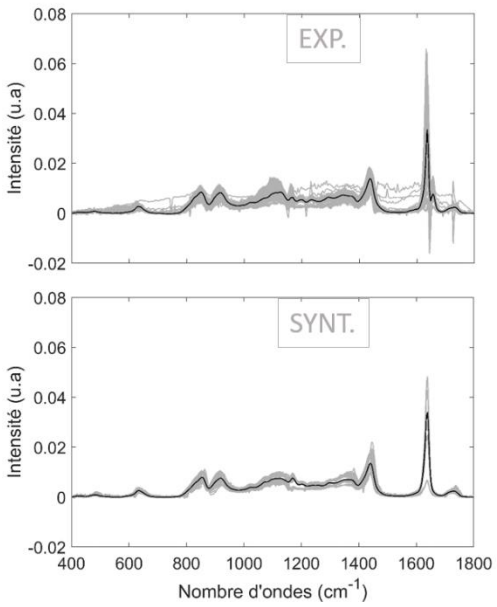


Figure V-12 – Spectres résiduels (gris) et spectre résiduel moyen (noir) des décompositions spectrales obtenues par MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le

expérimentales et synthétiques ; EXP. : expérimentales, SYNT. : synthétiques.

PNM-DL sur les données PVA-PNM-DL expérimentales (haut) et synthétiques (bas).

En définitive, bien que les films PVA-PNM-DL représentent des modèles simples de composition supposée linéaire, la méthode MCR-ALS apparait comme incapable de démélanger les spectres des deux composants dans les données spectrales expérimentales comme dans les données générées synthétiquement. Une contrainte d'égalité spectrale sur l'un des composants et plus précisément celui dont la concentration varie (c'est-à-dire le PNM-DL) est nécessaire pour obtenir une décomposition spectrale qui concorde avec la réalité. Mais malgré l'introduction de cette contrainte, l'estimation des concentrations des deux composants est erronée.

3.1.2. Mélanges PVA-GS-PNM-DL expérimentaux et synthétiques

La capacité de la méthode MCR-ALS à retrouver les spectres de composants purs contenus dans un mélange est ensuite challengée sur un modèle supposé linéaire un peu plus complexe, à savoir des films PVA-GS-PNM-DL qui sont formulés à partir de $K = 4$ composants.

3.1.2.1. Sans contrainte d'égalité spectrale

Les décompositions spectrales obtenues en appliquant la MCR-ALS sans contrainte d'égalité spectrale avec $K = 4$ composants (PVA, PNM-DL, GLY, SEPI) sur les données expérimentales et synthétiques des films PVA-GS-PNM-DL sont présentées dans la Figure V-13.

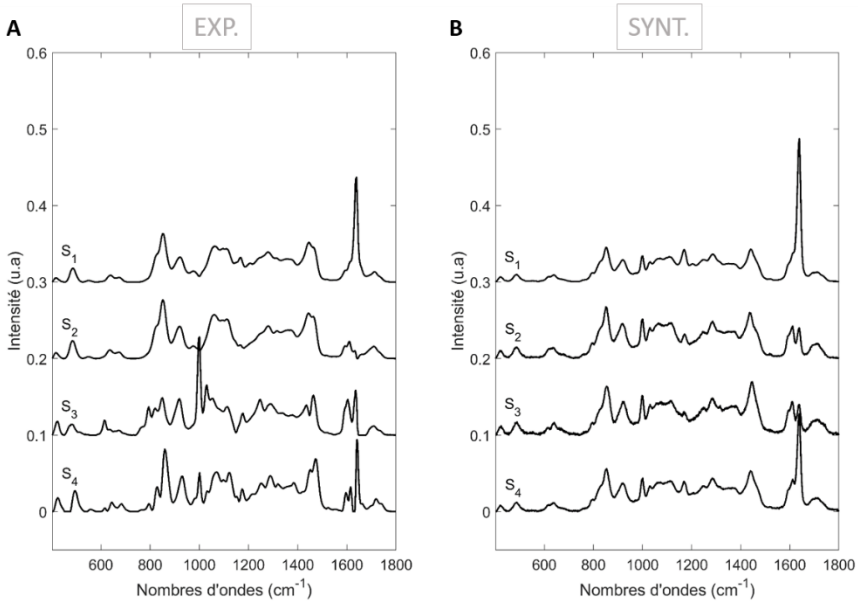


*Figure V-13 – Décompositions spectrales par la méthode MCR-ALS appliquée avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $K = 4$ composants sur les images Raman des films PVA-GS-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques.*

- Les spectres de composants purs estimés à partir des données expérimentales ne correspondent pas aux spectres de référence des composants utilisés dans la formulation des films (Figure V-13, **A**). Bien qu'il exprime la forte bande caractéristique du PNM-DL située à 1640 cm^{-1} (corrélation de 0.6526, Tableau V-5), le spectre S_1 est également composé de bandes Raman associées au PVA, au GLY et au SEPI comme en attestent ses corrélations respectives significatives: 0.5939, 0.6268 et 0.2294 (Tableau V-5). De plus, les spectres S_3 et S_4 présentent aussi la bande très

caractéristique du PNM-DL mais à plus faibles intensités : leur corrélation avec la référence spectrale du PNM-DL s'élève respectivement à 0.1789 et 0.2477 (Tableau V-5). Ces spectres sont cependant également le fruit d'un mélange des différents composants utilisés dans la formulation. Leur corrélation avec le PVA est de 0.5458 et 0.6440, celle avec le GLY est de 0.5695 et 0.7594 et celle avec le SEPI est de 0.6420 et 0.2914 (Tableau V-5). Enfin, le spectre S_2 est le seul spectre estimé dépourvu d'information concernant le PNM-DL : sa corrélation avec ce composant n'est que de -0.0356 (Tableau V-5). Toutefois, ce spectre reste lui aussi un mélange des trois autres composants : sa corrélation avec le PVA est de 0.8507, celle avec le GLY est de 0.8872 et celle avec le SEPI est de 0.1470 (Tableau V-5). La décomposition spectrale obtenue à partir des données expérimentales dans des conditions non supervisées ne fournit donc aucun des $K = 4$ spectres de références et est par conséquent loin de la réalité concernant la formulation des films PVA-GS-PNM-DL.

- Les spectres de composants purs estimés à partir des données synthétiques diffèrent de ceux obtenus à partir des données expérimentales et ne correspondent pas non plus aux spectres de référence des composants utilisés dans la formulation et la génération synthétique des mélanges PVA-GS-PNM-DL (Figure V-13, **B**). Chacun des $K = 4$ spectres de composants purs estimés présente des caractéristiques spectrales associées aux différents composants utilisés. Le spectre S_1 est à la fois corrélé au PVA, au PNM-DL, au GLY et au SEPI : les coefficients de corrélation s'élèvent dans l'ordre à 0.3695, 0.8667, 0.3109 et 0.3159 (Tableau V-5). De même, les spectres S_2, S_3 et S_4 sont tous les trois corrélés au PVA (0.8357, 0.8309 et 0.5989, Tableau V-5), au PNM-DL (0.1711, 0.7122 et 0.6627, Tableau V-5), au GLY (0.7151, 0.7122 et 0.5029, Tableau V-5) et au SEPI (0.4978, 0.4883 et 0.4248, Tableau V-5). Les spectres de composants purs estimés ne sont autre que des mélanges des composants de référence utilisés. Au final, devant la vérité de terrain avérée des données synthétiques des films PVA-GS-PNM-DL, l'application de la MCR-ALS dans des conditions d'application non supervisée sur les spectres de composants purs à estimer n'est une fois de plus pas performante pour obtenir la véritable décomposition spectrale des données.

Tableau V-5 – Coefficients de corrélation entre les spectres estimés $S_{i=1,\dots,4}$ obtenus par MCR-ALS appliquée avec une contrainte d'égalité spectrale sur le PNM-DL sur les données PVA-GS-PNM-DL expérimentales et synthétiques avec les références spectrales du PVA, du PNM-DL, du GLY et du SEPI ; EXP. : expérimentales, SYNT. : synthétiques.

	Sans contrainte d'égalité spectrale				**Avec contrainte d'égalité spectrale sur PNM-DL**			
	EXP.				**EXP.**			
Références	S_1	S_2	S_3	S_4	S_1 **PNM-DL**	S_2	S_3	S_4
PVA	0.5939	0.8507	0.5458	0.6440	-0.0708	0.8463	0.6410	0.7570
PNM-DL	0.6526	-0.0356	0.1789	0.2477	1.0000	-0.0260	0.2535	0.0411
GLY	0.6268	0.8872	0.5695	0.7594	-0.0679	0.8907	0.5691	0.8216
SEPI	0.2294	0.2470	0.6420	0.2914	0.0834	0.2745	0.5359	0.0751
	SYNT.				**SYNT.**			
PVA	0.3695	0.8357	0.8309	0.5989	-0.0708	0.8416	0.8375	0.5208
PNM-DL	0.8667	0.1711	0.1737	0.6627	1.0000	0.1528	0.1683	0.7466
GLY	0.3109	0.7151	0.7122	0.5029	-0.0679	0.7201	0.7184	0.4355
SEPI	0.3159	0.4978	0.4883	0.4248	0.0834	0.4974	0.4948	0.3891

En définitive, de même que pour les modèles estimés à partir des données Raman des mélanges avec seulement deux composants PVA-PNM-DL, l'application de la MCR-ALS sans contrainte d'égalité spectrale sur les données Raman expérimentales et synthétiques des mélanges PVA-GS-PNM-DL ne permet pas d'extraire distinctement les spectres de référence des $K = 4$ composants

utilisés. L'efficacité de la MCR-ALS sans supervision spectrale dans la décomposition spectrale et le suivi d'un principe actif à travers la peau est donc à nouveau remise en question.

3.1.2.2. Avec contrainte d'égalité spectrale sur le PNM-DL

À nouveau, pour guider son algorithme dans le démélange des spectres de composants purs contenus dans les films PVA-GS-PNM-DL à partir des données Raman expérimentales et synthétiques, une contrainte d'égalité spectrale sur le PNM-DL a été introduite dans la méthode MCR-ALS. Les décompositions spectrales obtenues en appliquant la MCR-ALS avec cette contrainte d'égalité spectrale sur le PNM-DL et $K = 4$ composants (PVA, PNM-DL, GLY, SEPI) sur les données expérimentales et synthétiques des films PVA-GS-PNM-DL sont présentées dans la Figure V-14.

- Les spectres de composants purs estimés à partir des données expérimentales ne correspondent pas aux spectres de référence des composants utilisés dans la formulation des films, excepté le spectre S_1 qui est contraint à être la référence spectrale du PNM-DL (Figure V-14, **A**). Malgré la contrainte spectrale imposée sur le spectre S_1, le spectre S_3 présente aussi la bande caractéristique du PNM-DL à 1640 cm^{-1} avec une intensité Raman plus faible. Sa corrélation avec le PNM-DL s'élève à 0.2535 (Tableau V-5). De plus, le spectre S_3 présente également des bandes caractéristiques des trois autres composants comme en attestent ses corrélations avec le PVA de 0.6410, avec le GLY de 0.5691 et avec le SEPI de 0.5359 (Tableau V-5). Les deux autres spectres de composants purs estimés, S_2 et S_4, apparaissent chacun comme un mélange différent de bandes caractéristiques associées au PVA et au GLY. Leur corrélation avec le PVA est de 0.8463 et 0.7570 et celle avec le GLY est de 0.8907 et 0.8216 (Tableau V-5). En revanche, le spectre S_2 présente en supplément des bandes Raman associées au SEPI : sa corrélation avec le SEPI est de 0.2745 (Tableau V-5). La décomposition spectrale obtenue à partir des données expérimentales à l'aide d'une supervision spectrale sur le spectre du PNM-DL présente donc trois autres spectres de composants purs estimés correspondant toujours à des mélanges de bandes caractéristiques des différents composants utilisés dans la formulation des films PVA-GS-PNM-DL.

- Les spectres de composants purs estimés à partir des données synthétiques diffèrent de ceux obtenus à partir des données expérimentales, à l'exception du spectre estimé S_1 qui est également contraint à être la référence spectrale du PNM-DL (Figure V-14, **B**). À nouveau, les trois autres spectres de composants purs estimés ne correspondent pas aux spectres de référence des composants utilisés dans la formulation et la génération synthétique des mélanges PVA-GS-PNM-DL. Chacun d'eux présente des bandes spécifiques associées à la fois au PVA, au PNM-DL, au GLY et au SEPI : leur corrélation respective avec le PVA est de 0.8416, 0.8375 et 0.5208, celle avec le PNM-DL est de 0.1528, 0.1683 et 0.7466, celle avec le GLY est de 0.7201, 0.7184 et 0.4355 et celle avec le SEPI est de 0.4974, 0.4948 et 0.3891 (Tableau V-5). Les spectres de composants purs estimés ne sont donc autres que des mélanges des composants de référence utilisés. Au final, devant la vérité de terrain avérée des données synthétiques des films PVA-GS-PNM-DL, l'application de la MCR-ALS à l'aide d'une supervision spectrale sur le spectre du PNM-DL n'est une fois de plus pas performante pour obtenir la véritable décomposition spectrale des données.

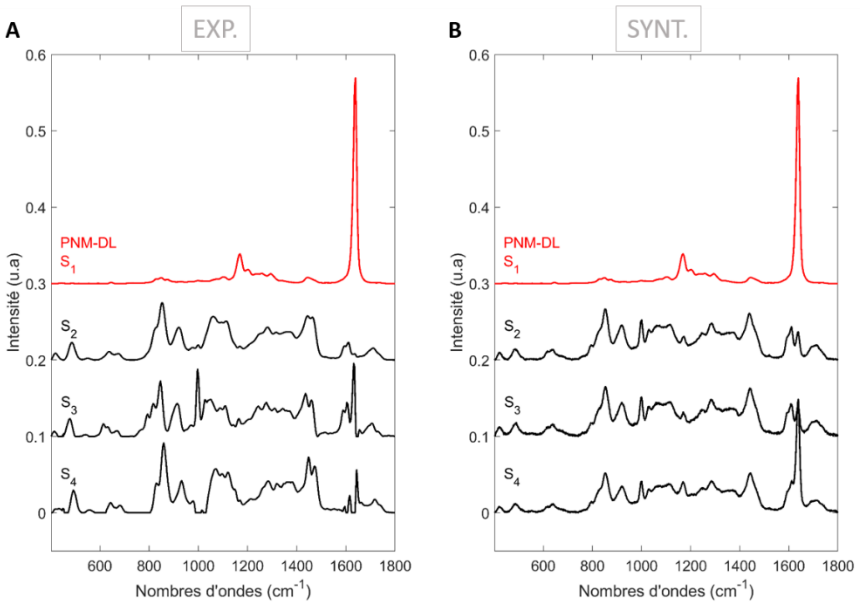


*Figure V-14 – Décompositions spectrales par la méthode MCR-ALS appliquée avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $K = 4$ composants et une contrainte d'égalité spectrale sur le PNM-DL sur les images Raman des films PVA-GS-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques.*

Qu'importe les données analysées, on remarque que la corrélation calculée entre les spectres estimés par la MCR-ALS et la référence spectrale du PVA est constamment très proche de la corrélation calculée entre ces spectres et le spectre de référence du GLY. En effet ces deux composants, PVA et GLY, ont des signatures Raman très proches. La corrélation entre leurs références spectrales est par conséquent significativement grande : 0.5850 (Tableau V-6). Cette proximité semble avoir des répercussions dans la décomposition spectrale des données :

- Par exemple, dans le spectre S_4 estimé par la MCR-ALS dont l'application a été réalisée sur les données expérimentales avec une contrainte d'égalité supplémentaire sur le PNM-DL, seules des bandes spécifiques au PVA et au GLY sont observables. Il ne contient pas d'informations concernant le PNM-DL ni concernant le SEPI. Sa corrélation avec le PNM-DL n'est que de 0.0411 (Tableau V-5). Celle-ci n'est pas excessive face aux coefficients de corrélation calculés entre le PVA et le PNM-DL et entre le GLY et le PNM-DL qui sont respectivement de -0.0708 et -0.0679 (Tableau V-6). Puis, la corrélation de ce spectre avec le SEPI de 0.0751 (Tableau V-5) est comparable à celle calculée entre le PVA et le SEPI (0.0734) et entre le GLY et le SEPI (0.0908) (Tableau V-6). L'interprétation spectrale de ce spectre permet donc de constater que la MCR-ALS est capable de dissocier le SEPI et le PNM-DL du PVA et du GLY.

- De la même façon, l'interprétation spectrale du spectre S_2 obtenu dans ce même modèle estimé montre que la MCR-ALS est aussi capable de dissocier les informations associées au PNM-DL à celles associées au SEPI : sa corrélation avec le PNM-DL n'est que de -0.0260 alors que sa corrélation avec le SEPI est de 0.2745 (Tableau V-5).

Au final, d'après l'observation de l'ensemble des coefficients de corrélation calculés entre les spectres estimés par MCR-ALS avec ou sans contrainte d'égalité spectrale sur le PNM-DL et les références spectrales du PVA et du GLY, la MCR-ALS ne dissocie pas correctement les informations concernant ces deux derniers composants de référence. Par conséquent, cette étude permet

d'entrevoir une limitation de la MCR-ALS, à savoir l'incapacité à dissocier les signatures spectrales de deux composants spectralement proches dont les variations sont corrélées.

En définitive, à l'instar des résultats obtenus pour les mélanges PVA-PNM-DL, la méthode MCR-ALS est incapable d'extraire les spectres des $K = 4$ composants purs contenus dans les mélanges PVA-GS-PNM-DL que ce soit à partir des données spectrales expérimentales ou des données spectrales générées synthétiquement. De plus, malgré l'ajout d'une contrainte d'égalité spectrale sur l'un des composants (le PNM-DL), l'information spectrale le concernant est exprimée dans plusieurs des autres spectres de composants purs qu'elle estime et les spectres des trois autres composants purs présents dans les mélanges (PVA, GLY et SEPI) ne sont pas extraits de manière distincte. Les performances de la MCR-ALS avec ou sans supervision pour étudier la composition cutanée et les changements moléculaires dans la peau sont donc remises en question.

Tableau V-6 – Coefficients de corrélation entre les références spectrales du PVA, du PNM-DL, du GLY et du SEPI.

	PVA	PNM-DL	GLY	SEPI
PVA	**1.0000**	-0.0708	**0.5850**	0.0734
PNM-DL		**1.0000**	-0.0679	0.0834
GLY			**1.0000**	0.0908
SEPI				**1.0000**

3.2. ÉVALUATION DE LA MVC-NMF

Au même titre que la MCR-ALS, la capacité de la MVC-NMF à extraire les signatures spectrales des composants purs présents dans un mélange et à estimer leurs concentrations relatives respectives est évaluée en analysant les résultats de son application : d'une part sur les images Raman acquises à partir des formulations des films PVA-PNM-DL et PVA-GS-PNM-DL, et d'autre part sur les images Raman générées synthétiquement qui reproduisent les mélanges supposés linéaires des formulations des films PVA-PNM-DL et PVA-GS-PNM-DL. Une comparaison entre les spectres de composants purs estimés par la MVC-NMF et les spectres de référence des composants utilisés dans les formulations et les générations de mélanges PVA-PNM-DL et PVA-GS-PNM-DL ainsi qu'une analyse des résidus des modèles estimés permettent d'examiner ses performances.

3.2.1. Mélanges PVA-PNM-DL expérimentaux et synthétiques

3.2.1.1. Sans contrainte d'égalité spectrale

Les décompositions spectrales obtenues en appliquant la MVC-NMF sans contrainte d'égalité spectrale avec $\tau = 0.1$ et $K = 2$ composants (PVA, PNM-DL) sur les données expérimentales et synthétiques des films PVA-PNM-DL sont présentées dans la

Figure V-15.

- Les spectres de composants purs estimés à partir des données expérimentales ne correspondent pas aux spectres de référence des composants utilisés dans la formulation des films (

Figure V-15, **A**). Bien qu'il présente en majorité les caractéristiques spectrales du PNM-DL (corrélation de 0.8882, Tableau V-7, **B**), le spectre S_1 exprime aussi des bandes caractéristiques du PVA (corrélation de 0.3894, Tableau V-7, **B**). Le spectre S_2 présente majoritairement des bandes Raman associées au PVA (corrélation de 0.9860, Tableau V-7, **B**) mais quelques bandes spécifiques au PNM-DL sont également dissociables. Sa faible corrélation avec le PNM-DL de 0.0948 (Tableau V-7, **B**) est tout de même significativement supérieure au coefficient de corrélation entre les références spectrales du PVA et du PNM-DL (-0.0708, Tableau V-6). La

décomposition spectrale obtenue à partir des données expérimentales dans des conditions non supervisées est donc loin de la réalité concernant la formulation des films PVA-PNM-DL.

- Les spectres de composants purs estimés à partir des données synthétiques sont similaires à ceux obtenus à partir des données expérimentales (

Figure V-15, **B**): la corrélation entre les premiers spectres de composants estimés S_1 s'élève à 0.9825 et celle entre les seconds spectres estimés S_2 est de 0.9948 (Tableau V-7, **A**). Par conséquent, leur proximité avec la vérité de terrain est tout aussi faible. En effet, à la fois S_1 et S_2 correspondent à des mélanges des composants de référence utilisés dans la formulation des donnés synthétiques des films PVA-PNM-DL: les corrélations de S_1 avec les spectres de référence du PNM-DL et du PVA sont respectivement de 0.9555 et 0.2266, et celles de S_2 avec ces mêmes références spectrales s'élèvent dans l'ordre à 0.9648 et 0.1949 (Tableau V-7, **B**). Au final, devant la vérité de terrain avérée des données synthétiques des films PVA-PNM-DL, l'application de la MVC-NMF dans des conditions d'application non supervisée sur les spectres de composants purs à estimer n'est elle aussi pas assez performante pour obtenir la véritable décomposition spectrale des données.

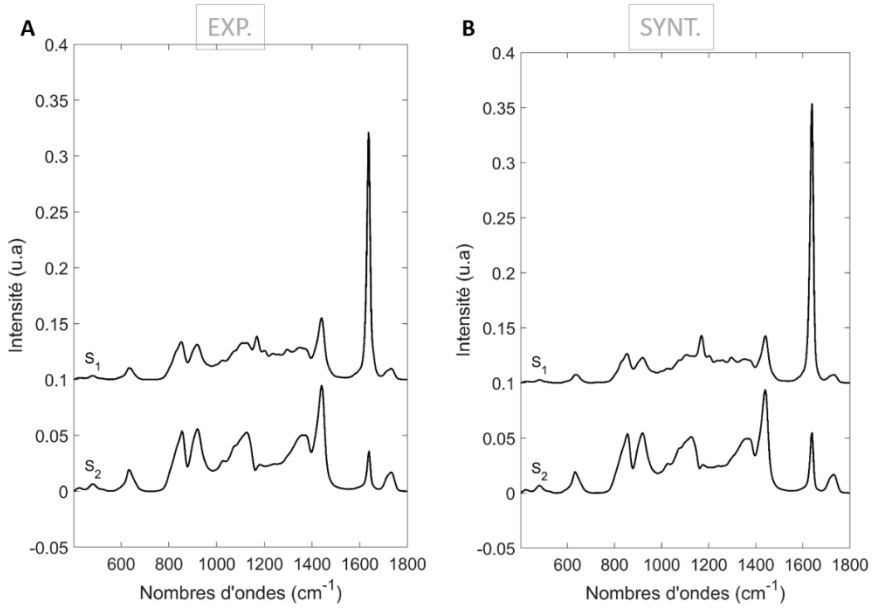


*Figure V-15 – Décompositions spectrales par la méthode MVC-NMF appliquée avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $\tau = 0.1$ et $K = 2$ composants sur les images Raman des films PVA-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques.*

*Tableau V-7 – Coefficients de corrélation entre (**A**) les spectres estimés $S_{i=1,2}$ obtenus par MVC-NMF appliquée sans contrainte d'égalité spectrale sur les données PVA-PNM-DL expérimentales et synthétiques et (**B**) les spectres estimés $S_{i=1,2}$ et les références spectrales du PVA et du PNM-DL ($\tau = 0.1$) ; EXP. : expérimentales, SYNT. : synthétiques.*

A	**EXP.**	
SYNT.	S_1	S_2
S_1	0.9825	0.3843
S_2	0.6172	0.9948

B	**EXP.**		**SYNT.**	
Références	S_1	S_2	S_1	S_2
PVA	0.3894	0.9860	0.2266	0.9648
PNM-DL	0.8882	0.0948	0.9555	0.1940

En définitive, de même que la MCR-ALS, la MVC-NMF appliquée sans aucune contrainte d'égalité spectrale est incapable de retrouver distinctement les spectres du PVA et du PNM-DL quelles que

soient les images Raman analysées. Alors que le mélange est constitué seulement de deux composants, les performances de la MVC-NMF sont, elles aussi, incertaines. Par conséquent, son utilisation pour le suivi d'un AC à travers l'épiderme humain sans aucune supervision spectrale est discutable.

3.2.1.2. Avec contrainte d'égalité spectrale sur le PNM-DL

De la même manière que pour l'évaluation de la MCR-ALS, une contrainte d'égalité spectrale sur le PNM-DL a été intégrée dans l'algorithme de la méthode MVC-NMF afin de la guider dans le démélange des spectres de composants purs contenus dans les films PVA-PNM-DL à partir des données Raman expérimentales et synthétiques. Les décompositions spectrales obtenues en appliquant la MVC-NMF avec une contrainte d'égalité spectrale sur le PNM-DL, $\tau = 0.1$ et $K = 2$ composants (PVA, PNM-DL) sur les données expérimentales et synthétiques des films PVA-PNM-DL sont présentées dans la Figure V-16.

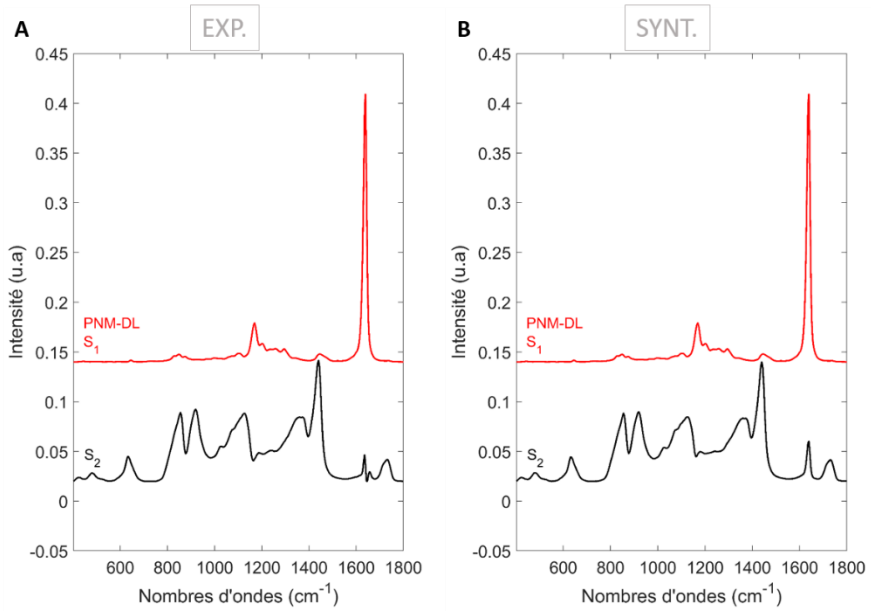


*Figure V-16 – Décompositions spectrales par la méthode MVC-NMF appliquée avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $\tau = 0.1$ et $K = 2$ composants et une contrainte d'égalité spectrale sur le PNM-DL sur les images Raman des films PVA-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques.*

*Tableau V-8 – Coefficients de corrélation entre (**A**) les spectres estimés $S_{i=1,2}$ obtenus par MVC-NMF appliquée avec une contrainte d'égalité spectrale sur le PNM-DL sur les données PVA-PNM-DL expérimentales et synthétiques et (**B**) les spectres estimés $S_{i=1,2}$ et les références spectrales du PVA et du PNM-DL ($\tau = 0.1$); EXP. : expérimentales, SYNT. : synthétiques.*

A	EXP.	
SYNT.	S_1 PNM-DL	S_2
S_1 PNM-DL	1.0000	-0.0083
S_2	0.0789	0.9949

B	EXP.		SYNT.	
Références	S_1 PNM-DL	S_2	S_1 PNM-DL	S_2
PVA	-0.0708	0.9968	-0.0708	0.9888
PNM-DL	1.0000	-0.0083	1.0000	0.0789

Ces décompositions spectrales diffèrent quelque peu l'une de l'autre (Tableau V-8, **A**). Alors que le spectre S_1 est contraint à être la référence spectrale du PNM-DL dans les deux modèles estimés, le spectre S_2 contient plus ou moins d'informations spécifiques au PNM-DL selon les données à partir desquelles il a été estimé. Tandis que le spectre S_2 estimé à partir des données synthétiques présente des bandes caractéristiques du PNM-DL, celui estimé à partir des données

expérimentales n'en exprime aucune de manière significative. Leur corrélation avec le PNM-DL confirme cette observation: la corrélation avec le spectre estimé à partir des données expérimentales est de -0.0083 et celle avec le spectre estimé à partir des donnés synthétiques est de 0.0789 (Tableau V-8, **B**). La première est plus proche de la corrélation entre les références spectrales du PVA et du PNM-DL que la seconde. Au final, les spectres de composants purs estimés à partir des données expérimentales sont plus proches des spectres de références des composants utilisés dans la formulation des films PVA-PNM-DL que les spectres de composants purs estimés à partir des données synthétiques.

Comme les spectres de composants purs estimés par la MVC-NMF à l'aide d'une supervision spectrale sur le spectre du PNM-DL à partir des données expérimentales sont proches des spectres de référence des composants utilisés dans la formulation des films PVA-PNM-DL, les concentrations relatives du PNM-DL estimés dans ce modèle peuvent être étudiées. Les moyennes de ces concentrations relatives estimées augmentent progressivement avec le pourcentage de PNM-DL introduit dans leurs formulations (Figure V-17). Cependant, l'observation de ces moyennes permet une nouvelle fois de constater une perte de linéarité dans l'estimation des concentrations relatives du PNM-DL en comparaison aux quantités réellement utilisées dans la formulation des films PVA-PNM-DL. Cette perte de linéarité s'explique à nouveau par la présence de la signature spectrale du PNM-DL dans les résidus de modèle estimé par la MVC-NMF (Figure V-18). Cette présence signifie qu'une partie de l'information concernant ce composant n'a pas été reproduite par ce modèle estimé et par conséquent que les estimations de concentrations relatives associées sont erronées.

D'autre part, bien que la décomposition spectrale obtenue par la MVC-NMF dans les mêmes conditions à partir des données synthétiques ne soit pas idéale, les moyennes des concentrations relatives associées au PNM-DL mènent à la même observation : elles augmentent en même temps que la quantité de PNM-DL utilisée dans la génération des données. Mais de façon inattendue, les moyennes des concentrations relatives du PNM-DL estimées à partir des données synthétiques se montrent un peu plus élevées que celles estimées à partir des données expérimentales (Figure V-17). Le caractère inattendu de cette observation s'explique par le fait qu'une partie de l'information concernant le PNM-DL dans les données synthétiques est contenue à la fois dans le second spectre de composants purs estimés (Figure V-16, **B**) et dans les spectres résiduels du modèle estimé (Figure V-18) et n'est donc pas prise en compte dans l'estimation de ses concentrations relatives associées. Ces précisions confirment donc bel et bien que les concentrations relatives estimées dans ce modèle sont elles aussi erronées.

En définitive, bien que les films PVA-PNM-DL représentent des modèles simples de composition supposée linéaire, la méthode MVC-NMF apparait elle aussi comme incapable d'extraire les spectres des deux composants dans les données spectrales expérimentales comme dans les données générées synthétiquement. De plus, l'introduction d'une contrainte d'égalité spectrale sur l'un des composants et plus précisément celui dont la concentration varie (c'est-à-dire le PNM-DL) ne permet pas non plus d'assurer une estimation de ses concentrations relatives sans biais. Les performances de la MVC-NMF pour étudier la composition cutanée et les changements moléculaires dans la peau sont donc remises en question

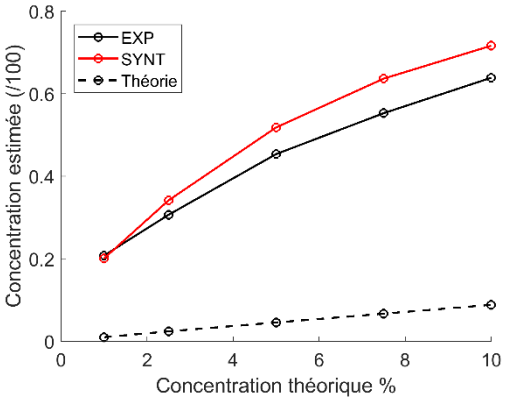


Figure V-17 – Comparaison des concentrations relatives moyennes du PNM-DL estimées par MVC-NMF sous contrainte d'égalité spectrale sur le PNM-DL avec $\tau = 0.1$ à partir des données PVA-PNM-DL expérimentales et synthétiques ; EXP. : expérimentales, SYNT. : synthétiques.

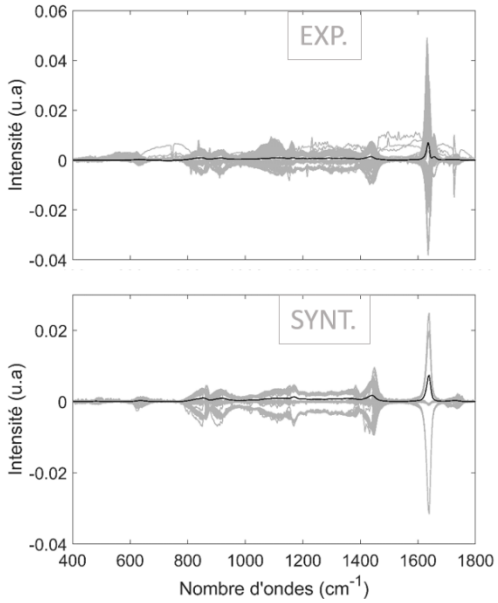


Figure V-18 – Spectres résiduels (gris) et spectre résiduel moyen (noir) des décompositions spectrales obtenues par MVC-NMF appliquée avec une contrainte d'égalité spectrale sur le PNM-DL sur les données PVA-PNM-DL expérimentales (haut) et synthétiques (bas) pour $\tau = 0.1$.

3.2.2. Mélanges PVA-GS-PNM-DL expérimentaux et synthétiques

Au même titre que la MCR-ALS, la capacité de la méthode MVC-NMF à retrouver les spectres de composants purs contenus dans un mélange est ensuite défiée sur un modèle supposé linéaire un peu plus complexe, à savoir des films PVA-GS-PNM-DL qui sont formulés à partir de $K = 4$ composants.

3.2.2.1. Sans contrainte d'égalité spectrale

Les décompositions spectrales obtenues en appliquant la MVC-NMF sans contrainte d'égalité spectrale avec $\tau = 0.1$ et $K = 4$ composants (PVA, PNM-DL, GLY, SEPI) sur les données expérimentales et synthétiques des films PVA-GS-PNM-DL sont présentées dans la Figure V-13.

- Les spectres de composants purs estimés à partir des données expérimentales ne correspondent pas aux spectres de référence des composants utilisés dans la formulation des films PVA-GS-PNM-DL (Figure V-19, **A**). Bien qu'il comporte en majorité les bandes caractéristiques du PNM-DL (corrélation de 0.8362, Tableau V-9), le spectre S_1 exprime aussi les bandes Raman associées aux trois autres composants utilisés dans la formulation des films. Sa corrélation avec le PVA, le GLY et le SEPI s'élève dans l'ordre à 0.4004, 0.4248 et 0.2386 (Tableau V-9). Les spectres S_2, S_3 et S_4 sont tous les trois principalement corrélés au PVA (0.8380, 0.8424 et 0.8298, Tableau V-9), au GLY (0.8857, 0.8787 et 0.8897, Tableau V-9) et au SEPI (0.2327, 0.3225 et 0.1914, Tableau V-9). La décomposition spectrale obtenue à partir des données expérimentales dans des conditions non supervisées ne fournit donc aucun des $K = 4$ spectres de référence et est par conséquent à son tour loin de la réalité concernant la formulation des films PVA-GS-PNM-DL.

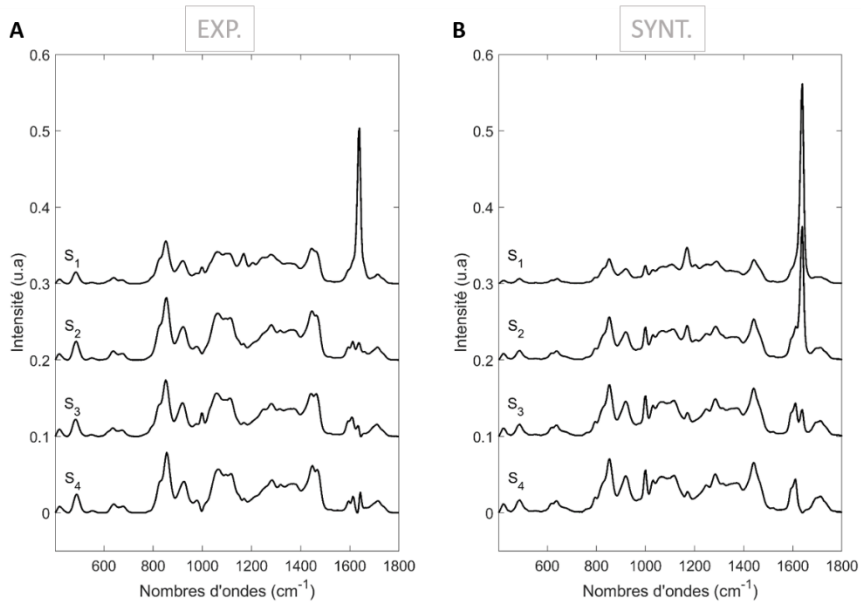


*Figure V-19 – Décompositions spectrales par la méthode MVC-NMF appliquée avec uniquement des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $\tau = 0.1$ et $K = 4$ composants sur les images Raman des films PVA-GS-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques.*

- Les spectres de composants purs estimés à partir des données synthétiques diffèrent de ceux obtenus à partir des données expérimentales et ne correspondent pas non plus aux spectres de référence des composants utilisés dans la formulation et la génération synthétique des mélanges PVA-GS-PNM-DL (Figure V-19, **B**). La bande spécifique du PNM-DL située à 1640 cm^{-1} est observable à la fois dans les spectres S_1 et S_2 avec une forte intensité et dans le spectres S_3 avec une plus faible intensité comme en attestent leurs corrélations avec ce composant (0.9693, 0.7803 et 0.1384, Tableau V-9). Le spectre S_4 ne présente quant à lui pas cette bande. Sa faible corrélation avec le PNM-DL de -0.0623 certifie qu'il est dépourvu d'informations associées au PNM-DL puisqu'elle est d'une part comparable à la corrélation observée entre les spectres de référence du PVA et du PNM-DL et entre ceux du GLY et du PNM-DL et d'autre part inférieure à la corrélation observée entre les références spectrales du SEPI et du PNM-DL. En revanche, chacun des 4 spectres estimés présente des caractéristiques spectrales relatives aux PVA, au GLY et au SEPI : leurs coefficients de corrélation avec le PVA, le GLY et le SEPI s'élèvent dans l'ordre à 0.1431, 0.1163 et 0.2013 pour S_1, 0.4862, 0.4101 et 0.3689 pour S_2, 0.8451, 0.7237 et 0.4970 pour S_3 puis 0.8656, 0.7426 et 0.4848 pour S_4 (Tableau V-9). Les spectres de composants purs estimés ne sont finalement autres que des mélanges des composants de référence utilisés. Devant la vérité de terrain avérée des données synthétiques des films PVA-GS-PNM-DL, l'application de la MVC-NMF dans des conditions d'application non supervisées sur les spectres de composants purs à estimer n'est une fois de plus pas performante pour obtenir la véritable décomposition spectrale des données et estimer les concentrations relatives de chacun des composants utilisés sans biais.

Tableau V-9 – Coefficients de corrélation entre les spectres estimés $S_{i=1,\dots,4}$ obtenus par MVC-NMF appliquée avec une contrainte d'égalité spectrale sur le PNM-DL sur les données PVA-GS-PNM-DL expérimentales et synthétiques avec les références spectrales du PVA, du PNM-DL, du GLY et du SEPI ; EXP. : expérimentales, SYNT. : synthétiques.

	Sans contrainte d'égalité spectrale				**Avec contrainte d'égalité spectrale sur PNM-DL**			
	EXP.				**EXP.**			
Références	S_1	S_2	S_3	S_4	S_1 **PNM-DL**	S_2	S_3	S_4
PVA	0.4004	0.8380	0.8424	0.8298	-0.0708	0.8228	0.8386	0.8444
PNM-DL	0.8362	0.0589	-0.0202	0.0164	1.0000	0.1450	-0.0283	-0.0457
GLY	0.4248	0.8857	0.8787	0.8897	-0.0679	0.8661	0.8935	0.8810
SEPI	0.2386	0.2327	0.3225	0.1914	0.0834	0.2944	0.2013	0.3149
	SYNT.				**SYNT.**			
PVA	0.1431	0.4862	0.8451	0.8656	-0.0708	0.6341	0.7665	0.8658
PNM-DL	0.9693	0.7803	0.1384	-0.0623	1.0000	0.6198	0.3925	-0.0683
GLY	0.1163	0.4101	0.7237	0.7426	-0.0679	0.5390	0.6545	0.7426
SEPI	0.2013	0.3689	0.4970	0.4848	0.0834	0.4349	0.4826	0.4842

En définitive, de même que pour la MCR-ALS et les modèles estimés par la MVC-NMF à partir des données Raman des mélanges avec seulement deux composants PVA-PNM-DL, l'application de la MVC-NMF sans contrainte d'égalité spectrale sur les données Raman expérimentales et synthétiques des mélanges PVA-GS-PNM-DL ne permet pas d'extraire distinctement les spectres de références des $K = 4$ composants utilisés. L'efficacité de la MVC-NMF sans supervision spectrale dans la décomposition spectrale et le suivi d'un principe actif à travers la peau est donc également remise en question.

3.2.2.2. Avec contrainte d'égalité spectrale sur le PNM-DL

À nouveau, pour guider son algorithme dans le démélange des spectres de composants purs contenus dans les films PVA-GS-PNM-DL à partir des données Raman expérimentales et synthétiques, une contrainte d'égalité spectrale sur le PNM-DL a été introduite dans la méthode MVC-NMF. Les décompositions spectrales obtenues en appliquant la MVC-NMF avec une contrainte d'égalité spectrale sur le PNM-DL, $\tau = 0.1$ et $K = 4$ composants (PVA, PNM-DL, GLY, SEPI) sur les données expérimentales et synthétiques des films PVA-GS-PNM-DL sont présentées dans la Figure V-20.

- Les spectres de composants purs estimés à partir des données expérimentales ne correspondent pas aux spectres de référence des composants utilisés dans la formulation des films, excepté le spectre S_1 qui est contraint à être la référence spectrale du PNM-DL (Figure V-20, **A**). Malgré la contrainte spectrale imposée sur le spectre S_1, le spectre S_2 présente aussi la bande caractéristique du PNM-DL à 1640 cm^{-1} avec une intensité Raman plus faible. Sa corrélation avec le PNM-DL s'élève à 0.1450 (Tableau V-9). En revanche, les deux autres spectres de composants purs estimés, S_3 et S_4, n'expriment pas de bandes spécifiques au PNM-DL. Plus précisément, leur corrélation avec le PNM-DL est comparable à celle du PVA et du GLY avec le PNM-DL (-0.0283 et -0.0457, Tableau V-9) ce qui atteste de l'absence d'informations concernant ce composant dans ces spectres. Par ailleurs, les 3 spectres non contraints estimés S_2, S_3 et S_4 sont également le fruit d'un mélange des trois autres composants utilisés dans la formulation. Leur corrélation avec le PVA est de 0.8228, 0.8661 et 0.2944, celle avec le GLY est de 0.8386, 0.8935 et 0.2013 et celle avec le SEPI est de 0.8444, 0.8810 et 0.3149 (Tableau V-9).

- Les spectres de composants purs estimés à partir des données synthétiques diffèrent de ceux obtenus à partir des données expérimentales, à l'exception du spectre estimé S_1 qui est

également contraint à être la référence spectrale du PNM-DL (Figure V-20, **B**). Malgré la contrainte spectrale imposée sur le spectre S_1, les spectres S_2 et S_3 expriment aussi la forte bande très caractéristique du PNM-DL : leur corrélation avec le PNM-DL s'élève à 0.6198 et 0.3925 (Tableau V-9). Le spectre S_4 ne contient quant à lui aucune information sur le PNM-DL : aucune de ses bandes caractéristiques n'est exprimée puis sa corrélation avec ce composant est très faible (-0.0683, Tableau V-9) et, qui plus est, est comparable à la corrélation du PVA et du GLY avec le PNM-DL. Par ailleurs, chacun des 3 spectres non contraints estimés présentent des caractéristiques spectrales relatives aux PVA, au GLY et au SEPI : leurs coefficients de corrélation avec le PVA, le GLY et le SEPI s'élèvent dans l'ordre à 0.6341, 0.5390 et 0.4349 pour S_2, 0.7665, 0.6545 et 0.4846 pour S_3, puis 0.8658, 0.7426 et 0.4842 pour S_4 (Tableau V-9). Au final, devant la vérité de terrain avérée des données synthétiques des films PVA-GS-PNM-DL, l'application de la MVC-NMF à l'aide d'une supervision spectrale sur le spectre du PNM-DL n'est une fois de plus pas performante pour obtenir la véritable décomposition spectrale des données.

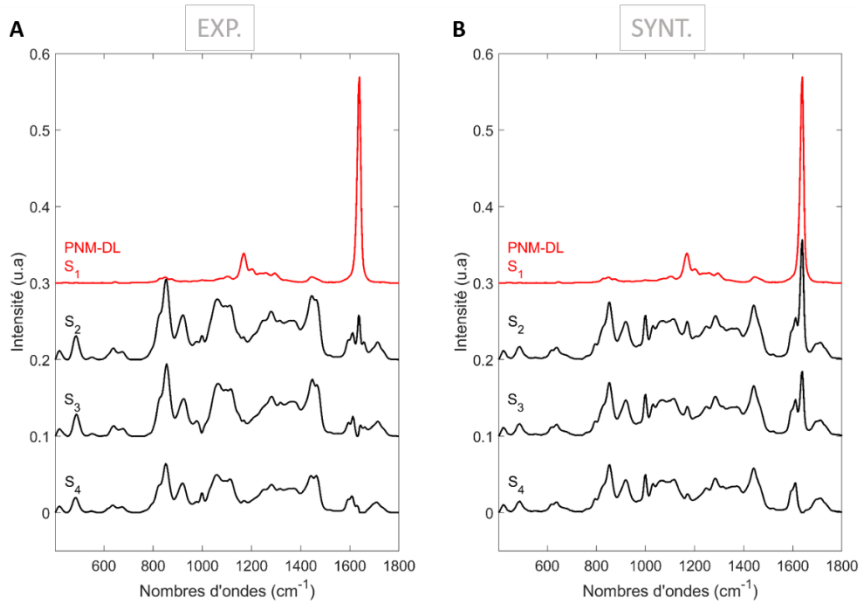


*Figure V-20 – Décompositions spectrales par la méthode MVC-NMF appliquée avec des contraintes de non-négativité et de normalisation des spectres et des concentrations relatives pour $\tau = 0.1$ et $K = 4$ composants et une contrainte d'égalité spectrale sur le PNM-DL sur les images Raman des films PVA-GS-PNM-DL obtenues de manière (**A**) expérimentale et (**B**) synthétique (les spectres estimés sont décalés en intensité pour plus de clarté) ; EXP. : expérimentales, SYNT. : synthétiques.*

De façon analogue à la MCR-ALS, la MVC-NMF rencontre le même problème concernant la dissociation de deux composants spectralement proches. En effet, les coefficients de corrélation entre les spectres estimés avec ou sans supervision sur le PNM-DL et les références spectrales du PVA et du GLY évoluent toujours conjointement. En revanche, alors que la MCR-ALS réussit à dissocier le SEPI des trois autres composants, la MVC-NMF a également des difficultés à dissocier sa signature de celles du PVA et du GLY : les coefficients de corrélation entre les spectres de composants purs estimés et le SEPI sont toujours comparables à ceux calculés entre les spectres estimés et ceux du PVA et du GLY.

En définitive, à l'instar des résultats obtenus par la méthode MCR-ALS, la méthode MVC-NMF est incapable d'extraire les spectres des $K = 4$ composants purs contenus dans les mélanges PVA-GS-PNM-DL que ce soit à partir des données spectrales expérimentales ou des données spectrales générées synthétiquement. De plus, malgré l'ajout d'une contrainte d'égalité spectrale sur l'un des composants (le PNM-DL), l'information spectrale le concernant est exprimée dans plusieurs

des autres spectres de composants purs qu'elle estime et les spectres des trois autres composants purs présents dans les mélanges (PVA, GLY et SEPI) ne sont pas extraits de manière distincte. Les performances de la MVC-NMF avec ou sans supervision pour étudier la composition cutanée et suivre les concentrations relatives d'un principe actif dans la peau sont donc également remises en question.

CONCLUSION

En résumé, les formulations de films de polymères fréquemment utilisés en cosmétique représentent un bel exemple de mélange linéaire de différents composants. La complexité de ces mélanges est moindre par rapport à celle du mélange que constitue la peau. Les données spectrales Raman enregistrées à partir de ces films de polymères se sont donc avérées être plus simples à étudier. De plus, la connaissance de leur composition qualitative et quantitative exacte nous a permis de construire une vérité de terrain pour évaluer les performances des méthodes de démélange spectral, la MCR-ALS et la MVC-NMF, utilisées dans le chapitre précédent. En effet, ces connaissances *a priori* nous ont permis de générer des données spectrales synthétiques qui soient proches de celles enregistrées expérimentalement et qui sont donc réalistes. Cette simulation de données Raman a été réalisée en s'affranchissant de tout évènement externe à la formulation des mélanges, tel que les potentielles perturbations spectrales induites par diverses interactions moléculaires, qui peut déstabiliser le processus de démélange spectral entrepris par les méthodes d'analyse multivariée. Les données produites par l'approche de simulation proposée se sont révélées proches, au sens des critères quantitatifs et qualitatifs utilisés, de celles mesurées par imagerie Raman.

Les résultats obtenus par la MCR-ALS et la MVC-NMF de manière non supervisée ou semi-supervisée à partir des données expérimentales et synthétiques ont ensuite été comparés à la vérité de terrain. Dans cette étude préliminaire, réalisée en fin de doctorat, l'analyse comparative de ces résultats n'a abouti à une qualité de résultat comparable à celle obtenue dans les autres études que nous avons menées. Les deux méthodes ont rencontré des difficultés à retrouver la composition qualitative et semi-quantitative utilisée pour la création de vérité de terrain dans certaines conditions.

Ces résultats préliminaires invitent selon nous à poursuivre le travail sur la simulation de données plutôt qu'à tirer des conclusions hâtives sur l'utilisation des méthodes MCR-ALS et MVC-NMF pour le suivi d'un composant cosmétique actif dans des films de polymères. Ils soulignent deux perspectives au travail réalisé dans ce chapitre :

- d'une part que la simulation de données réaliste est importante pour l'évaluation rigoureuse des approches multivariées et qu'elle doit en conséquence être poursuivie afin de parvenir à une simulation qui n'aboutisse pas à des résultats des méthodes d'analyse multivariée contraires à ceux obtenus sur des données réelles ;

- d'autre part que les profils purs estimés par les méthodes de démélange doivent d'une façon générale être considérés avec prudence et avec une validation avant d'être employés pour l'analyse des images Raman.

Ces perspectives, ainsi que d'autres sont présentées plus en détails dans le Chapitre VI.

Chapitre VI.
PERSPECTIVES D'AMÉLIORATION

Résumé
Ce chapitre présente des axes d'améliorations à explorer pour rendre les méthodes de démélange spectral plus adaptées aux données spectrales enregistrées à partir d'échantillons de peau humaine reconstruite. Pour cela, nous soumettons l'idée de reconsidérer en grande parte la formulation du problème de minimisation à résoudre pour obtenir une décomposition spectrale plus proche de la réalité : l'ajout de termes de pénalisation sur les matrices des spectres S et des concentrations relatives locales C, une redéfinition du modèle linéaire ainsi qu'un changement de fonction de coût à minimiser sont suggérés afin d'outrepasser les limitations des méthodes MCR-ALS et MVC-NMF. Puis nous indiquons les différentes étapes à mettre en place pour résoudre le nouveau problème de minimisation auquel nous avons pensé et évaluer les performances du nouvel algorithme potentiel.

Introduction 194
1. Amélioration des méthodes d'analyse multivariée adaptées aux problématiques cosmétiques 194
1.1. Évolution de la formulation du problème d'optimisation à résoudre 194
1.1.1. La sous/sur-décomposition spectrale 195
1.1.2. L'homogénéité du voisinage 196
1.1.3. La cohérence des spectres estimés avec les données analysées 197
1.1.4. Les perturbations spectrales 197
1.1.5. Changement de la fonction de coût dans le terme d'attache aux données 200
1.2. Choix de l'algorithme de minimisation à utiliser 202
1.3. Tests de la robustesse du nouvel algorithme 203
2. Création d'une interface Matlab 203

INTRODUCTION

Les travaux de cette thèse réalisés sur des données réelles ont permis de mettre en évidence l'intérêt des méthodes MCR-ALS et MVC-NMF pour l'analyse des spectromètres Raman. De nombreuses recommandations ont également été formulées pour une utilisation optimisée. Cependant, bien que préliminaires, les simulations menées dans le Chapitre V sous-tendent à ce que l'utilisation des méthodes MCR-ALS et MVC-NMF pour le suivi d'un composant cosmétique actif dans des films de polymères ne soit pas sans limite. En particulier, il a été souligné qu'aucune de ces deux méthodes ne prend en considération les éventuels problèmes de colinéarité spectrale présents entre les différents composants purs du mélange. De même, ni l'une ni l'autre ne prend en compte les possibles perturbations spectrales induites suite à l'interaction moléculaire entre les différents composants purs pouvant être différentes d'un spectre à un autre.

Nous proposons dans ce chapitre différents axes d'améliorations à prendre en considération pour améliorer la qualité du suivi d'un AC à travers la peau. De plus, le suivi d'un AC à travers la peau étant un sujet très plébiscité par les industriels, un projet de création d'application accessible au plus grand nombre pour pouvoir utiliser ces méthodes d'analyse de données est également présenté.

1. AMÉLIORATION DES MÉTHODES D'ANALYSE MULTIVARIÉE ADAPTÉES AUX PROBLÉMATIQUES COSMÉTIQUES

Afin de surpasser les potentielles limites de la MCR-ALS et la MVC-NMF et d'augmenter la qualité de la décomposition spectrale, plusieurs perspectives d'améliorations pour les méthodes de démélange spectral peuvent être envisagées.

- Dans un premier temps, il est essentiel de réadapter la formulation du problème d'optimisation à résoudre à notre problématique. L'ajout de contraintes supplémentaires et le choix d'une fonction de coût différente sont des solutions sur lesquelles il peut être intéressant de se pencher.

- Puis, comme les méthodes de résolution telles que la régression des moindres carrés ou la descente de gradient utilisées respectivement dans la MCR-ALS et MVC-NMF risquent de ne plus être les plus adaptées à la nouvelle formulation du problème d'optimisation, d'autres techniques de résolution de courbe multivariée plus adaptée pourraient être envisagées.

- Enfin, les performances des nouvelles versions améliorées des méthodes de démélange spectral développées devront être évaluées. Pour ce faire, il est indispensable de poursuivre le développement de la simulation de données réalistes présentées dans le Chapitre V, avec l'objectif combiné d'une justesse de simulation et d'une cohérence de résultat avec l'utilisation de données réelles.

1.1. ÉVOLUTION DE LA FORMULATION DU PROBLÈME D'OPTIMISATION À RÉSOUDRE

Nous suggérons plusieurs axes d'amélioration dans la formulation du problème de minimisation pour tenter d'augmenter la qualité de la décomposition spectrale des données Raman enregistrées à partir d'échantillons de peau exposés à un AC. Tout d'abord, nous proposons de gérer les éventuelles colinéarités présentes entre différents composants purs du mélange et les différences de contributions considérables entre chacun d'eux en intégrant des termes de

régularisation adaptés dans le problème d'optimisation. De même, nous pensons que la cohérence des spectres de composants purs estimés avec ceux des données Raman analysées ainsi que l'homogénéité des contributions spatiales peuvent être prises en considération en ajoutant aussi des termes de régularisation supplémentaires appropriés. En revanche, pour tenir compte des éventuelles perturbations spectrales engendrées par des interactions moléculaires dans les données Raman, nous pensons qu'il est nécessaire de repenser l'expression du modèle linéaire et de la considérer autrement. Enfin, nous soumettons l'idée d'utiliser une autre fonction de coût plus appropriée pour évaluer la proximité entre le modèle estimé et les données analysées. Les détails de nos différentes suggestions concernant ces multiples modifications à apporter sont présentées ci-après.

1.1.1. La sous/sur-décomposition spectrale

Dans les études réalisées dans le Chapitre IV, le problème de sous- et sur-décomposition spectrale a été régulièrement mentionné pour évaluer la qualité de la décomposition spectrale des données Raman. Lorsque le nombre de composant utilisé était trop faible, les méthodes MCR-ALS et MVC-NMF avaient tendance à sous-décomposer les données : les bandes caractéristiques de la peau et de l'actif étudié restaient mélangées dans un même spectre de composant pur estimé. Ce phénomène s'est également produit lorsqu'un des composants (par exemple la caféine) avait une très faible contribution dans le mélange le rendant plus difficilement détectable. En revanche, lorsque le nombre de composant utilisé était trop élevé, les méthodes MCR-ALS et MVC-NMF avaient tendance à sur-décomposer les données : les bandes caractéristiques de l'actif étudié (par exemple le Delipidol®) apparaissaient dans plusieurs spectres de composants purs estimés. Ces observations indiquent donc que ces deux méthodes ont des difficultés à sélectionner les spectres de composants les plus pertinents.

De la même manière, les tests réalisés dans le Chapitre V ont permis de mettre en lumière ce problème de sous- et sur-décomposition des données : les spectres de deux composants purs proches (PVA et GLY) n'ont pas été séparés distinctement et ont été considérés comme un seul et même spectre de composant pur quelle que soit la méthode utilisée (MCR-ALS ou MVC-NMF). Une fois de plus, la sélection des spectres de composants les plus pertinents s'est montrée être une limite de ces deux méthodes lorsqu'une multi-colinéarité entre spectres de composants purs est présente. À noter qu'il n'est par ailleurs pas exclu que les signatures spectrales des composants de la peau ou des actifs cosmétiques appliquées soient corrélés entre eux.

Par conséquent, afin de mieux gérer la sélection des spectres de composants purs les plus significatifs nous suggérons d'utiliser une régularisation « Elastic-Net » dans la formulation du problème de minimisation. Cette régularisation est une combinaison de deux régularisations très connues : la régularisation ℓ_1, dite régularisation *Lasso*, et la régularisation ℓ_2, dite régularisation *Ridge*.

- La régularisation ℓ_1 a tendance à privilégier qu'un seul des éléments corrélés et éliminer les autres. Pour cela, elle fournit une matrice solution des abondances parcimonieuses. Elle force les coefficients de contribution à être nuls si le composant considéré n'a que très peu d'influence dans le mélange. Toutefois, lorsque deux composants sont très corrélés, cette régularisation ne sélectionne qu'un seul des composants corrélés.

- La régularisation ℓ_2 a tendance à réduire tous les coefficients de contributions associés aux composants corrélés ensemble et pénalisent ceux trop élevés pour éviter un sur-ajustement. Pour se faire, elle force les coefficients de contribution associés aux composants proches ou peu

significatifs à être petits mais non nuls. Ainsi, elle traite les problèmes de multi-colinéarité grâce à une restriction des abondances et une conservation de tous les composants.

Au final, l'utilisation d'une régularisation *Elastic-Net* permet de gérer la situation où des composants du mélange sont corrélés entre eux. Elle permet d'éviter la sélectivité trop forte que peut proposer la régularisation ℓ_1 en conservant des composants corrélés grâce à la régularisation ℓ_2. Ainsi, des composants très corrélés tels que le PVA et le GLY pourraient être dissociés et des composants à faible contribution dans le mélange, tels que la caféine, pourraient être identifiés. L'expression de ce terme de pénalité supplémentaire est le suivant :

$$\Omega_1 = \frac{\alpha_1}{2}\|C\|_1 + \frac{1-\alpha_1'}{2}\|C\|_2^2 \tag{VI-1}$$

où $\alpha_1, \alpha_1' \in \mathbb{R}$ sont des paramètres contrôlant la force de régularisation (par défaut, $\alpha_1 = \alpha_1' = 1$). Des valeurs élevées de α_1, α_1' impliquent une forte régularisation, autrement dit elles engendrent moins de sur-ajustement, voire même un sous-ajustement. À l'inverse, de faibles valeurs de α_1, α_1' impliquent une faible régularisation, c'est-à-dire elles engendrent un sur-ajustement.

Le problème d'optimisation à résoudre deviendrait donc le suivant :

$$\min_{S\geq 0, C\geq 0, \sum_{k=1}^{K} C_k = 1} \|D - CS^T\|_2^2 + \frac{\alpha_1}{2}\|C\|_1 + \frac{\alpha_1'}{2}\|C\|_2^2 \tag{VI-2}$$

1.1.2. L'homogénéité du voisinage

La peau se compose de différentes couches et sa composition moléculaire est variable selon les couches. Mais chaque couche cutanée est constituée de manière homogène. Cette connaissance *a priori* est alors très informative sur la répartition des coefficients de contributions de chaque composant de la peau : la distribution spatiale des concentrations relatives de chaque composant doit alors respecter une certaine homogénéité.

Cependant, les méthodes MCR-ALS et MVC-NMF décrites dans le Chapitre IV s'appliquent sur les matrices de données dépliées. Elles ne prennent donc pas en compte la dimension spatiale des données spectrales enregistrées. Cet inconvénient n'a pas été évoqué jusqu'ici, mais celui-ci constitue un autre point d'amélioration à travailler pour définir une formulation du problème d'optimisation plus adaptée à notre problématique.

Afin de prendre en compte l'homogénéité du voisinage dans les données et donc d'assurer la faible variation spatiale des coefficients de concentrations relatives, nous proposons d'ajouter une pénalité spatiale de lissage sur ces coefficients dont l'expression est la suivante :

$$\Omega_2 = \frac{\alpha_2}{2}\|C\Delta\|_2^2 \tag{VI-3}$$

avec Δ la matrice calculant les différences finies spatiales au premier ordre dans les trois directions de la matrice de données et $\alpha_2 \in \mathbb{R}$ un paramètre contrôlant la force de lissage spatial des coefficients.

Le problème d'optimisation à résoudre deviendrait donc le suivant :

$$\min_{S\geq 0, C\geq 0, \sum_{k=1}^{K} C_k = 1} \|D - CS^T\|_2^2 + \frac{\alpha_1}{2}\|C\|_1 + \frac{\alpha_1'}{2}\|C\|_2^2 + \frac{\alpha_2}{2}\|C\Delta\|_2^2 \tag{VI-4}$$

1.1.3. La cohérence des spectres estimés avec les données analysées

Un des plus importants critères pour juger de la qualité d'une décomposition spectrale des données est la cohérence des spectres de composants purs estimés avec ceux exprimant des mélanges dans les données acquises. En effet, les spectres de composants purs estimés ne doivent pas être totalement sans rapport avec les spectres analysés, ils doivent être absolument liés aux informations spectrales fournies par les données. Alors qu'il fait défaut à la MCR-ALS, ce critère est l'essence même de la proposition de la MVC-NMF : la contrainte de volume minimale imposée sur le simplex formé par les spectres de composants purs permet d'assurer une certaine proximité entre les spectres exprimant les mélanges dans les données analysées et les spectres de composants purs estimés par la méthode.

Toutefois dans la littérature, il existe d'autres manières d'imposer ce critère de cohérence spectrale. Par exemple, il pourrait être envisager de changer cette contrainte de volume minimal par une autre pénalité de proximité entre les spectres initiaux et estimés dont l'expression est plus simple :

$$\Omega_3 = \frac{\alpha_3}{2} \left\| S - S^{(0)} \right\|_2^2 \tag{VI-5}$$

où $S^{(0)}$ représente la matrice formée par les spectres initiaux et $\alpha_3 \in \mathbb{R}$ contrôle le compromis de proximité. Ce remplacement de pénalité est suffisamment pertinent pour chercher une solution finale proche d'une pré-estimation grossière des spectres de composants purs.

Deux problèmes d'optimisation pourraient alors être résolus :

$$\min_{S\geq 0, C\geq 0, \sum_{k=1}^{K} C_k} \|D - CS^T\|_2^2 + \frac{\alpha_1}{2}\|C\|_1 + \frac{\alpha_1'}{2}\|C\|_2^2 + \frac{\alpha_2}{2}\|C\Delta\|_2^2 + \frac{\alpha_3}{2}\left\|S - S^{(0)}\right\|_2^2 \tag{VI-6}$$

$$\min_{S\geq 0, C\geq 0, \sum_{k=1}^{K} C_k=1} \|D - CS^T\|_2^2 + \frac{\alpha_1}{2}\|C\|_1 + \frac{\alpha_1'}{2}\|C\|_2^2 + \frac{\alpha_2}{2}\|C\Delta\|_2^2 + 2\,\alpha_3 V(S^T) \tag{VI-7}$$

où $V(S^T) \in \mathbb{R}$ est le volume du simplex formé par les spectres contenus dans la matrice S.

1.1.4. Les perturbations spectrales

Le choix d'utiliser des méthodes de démélange spectral telles que la MCR-ALS et la MVC-NMF pour notre objectif de suivi des cinétiques de pénétration d'un AC dans la peau découle des inconvénients identifiés dans l'utilisation de la CLS déjà employée à ces fins dans la littérature [44], [105]. En effet, la NCLS nécessite d'avoir accès aux signatures spectrales Raman de tous les composants de la peau et de celle de l'actif étudié pour pouvoir être appliquée et ce besoin ne laisse aucune liberté aux composants purs d'exprimer des modifications spectrales induites par des interactions moléculaires avec d'autres composants par rapport à leur expression Raman à l'état pur. L'utilisation de la MCR-ALS et de la MVC-NMF avait donc pour but de s'affranchir de ce besoin restrictif de connaître les signatures spectrales de tous les composants présents et d'obtenir plus de liberté dans l'expression de leurs signatures spectrales qui sont toutes estimées uniquement à partir des données acquises.

Pourtant, les études menées dans les Chapitres IV et V ont montré que les méthodes MCR-ALS et MVC-NMF étaient plus ou moins capables de surpasser ce problème de variabilité spectrale engendrée par certaines interactions moléculaires. Dans nos applications, selon la facilité à détecter le composant actif d'intérêt et selon si une supervision spectrale est appliquée ou non,

les méthodes ont sous- ou sur-décomposé les données car elles n'ont pas forcément tenu compte des éventuelles perturbations spectrales pouvant être différentes d'un spectre à un autre.

De manière générale, les méthodes de démélange spectral supposent que les données peuvent être reproduites par un modèle linéaire en raison de sa simplicité et sa facilité d'utilisation. Cependant, la plupart d'entre elles supposent implicitement que chaque composant est représenté par un seul et unique spectre qui n'est pas modifié lorsque le composant est mélangé avec les autres composants. Sous cette hypothèse implicite d'alignement parfait des bandes Raman entre spectres de composants purs et spectres de données, les méthodes de démélange ne tiennent alors pas compte des éventuelles perturbations spectrales qui peuvent apparaître dans les signatures spectrales de chaque composant lorsque les différentes molécules du mélange interagissent entre elles ni de leurs différentes expressions dans l'espace. De ce fait, les décompositions spectrales obtenues par ces méthodes ne sont pas toujours de bonne qualité. Un axe d'amélioration serait donc de prendre en compte les variabilités spatio-spectrales représentant les perturbations induites dans les spectres de composants purs afin d'augmenter la qualité de modélisation des données.

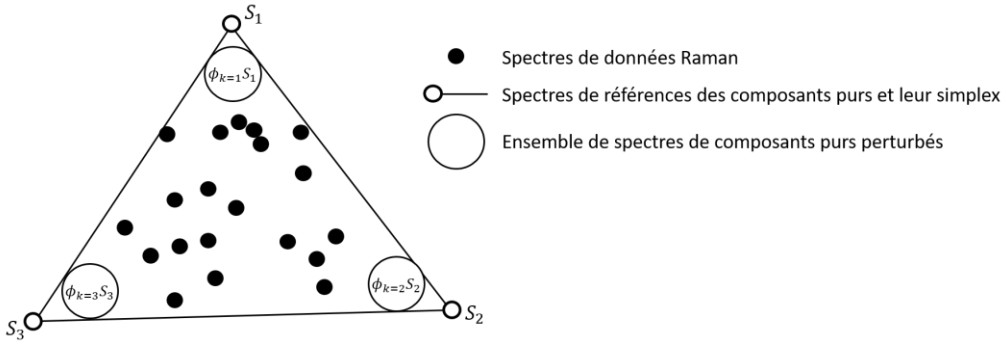


Figure VI-1 – Représentation géométrique des données et du principe de « dictionnaire » de spectres de composants purs perturbés.

En reprenant la représentation géométrique des données dans un simplex, la considération des sommets comme définissant les spectres de composants purs est trop restrictive car elle ne laisse pas de place à de possibles perturbations spectrales. L'idée serait donc de prendre en compte les variabilités spatio-spectrales des signatures spectrales de chaque composant pur en créant un « dictionnaire » de spectres qui leur sont respectivement associés comme cela a été développé dans la littérature dans d'autres domaines d'application [115]–[118]. De cette façon, un composant pur ne serait plus représenté par une seule et unique signature spectrale mais plutôt par un ensemble de spectres abritant les différentes variations identifiées dans les données. Cet ensemble de spectres associés à un seul et même composant pur dans les données est représenté sous forme de cercle accolé au sommet du simplex dans la Figure VI-1. Mathématiquement, cela revient à redéfinir l'expression de la matrice composée des spectres de composants purs S comme étant égale à:

$$S_k = f_{ik}(S_{REFS,k}) \tag{VI-8}$$

où $f_{ik}: \mathbb{R}^{N_\lambda} \rightarrow \mathbb{R}^{N_\lambda}$ représente la perturbation appliquée au k-ème spectre de composant pur de référence $S_{REFS,k}$ pour le i-ème spectre de données. Si on considère ϕ_{ik} une matrice de taille $(N_\lambda \times N_\lambda)$ contenant l'ensemble des facteurs de perturbations du k-ème spectre de composants purs identifiés dans le i-ème spectre de données, alors $f_{ik}(S_{REFS,k}) = \phi_{ik}\, S_{REFS,k}$. La matrice ϕ_{ik} peut être définie de diverses façons dont deux représentent des cas particuliers de perturbations :

- Lorsque les perturbations représentent des variations d'intensités spectrales, la matrice ϕ_{ik} est définie par

$$\phi_{ik} = \begin{pmatrix} m_1 & 0 & 0 & \cdots & 0 \\ 0 & m_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & m_{N_\lambda} \end{pmatrix}_{N_\lambda \times N_\lambda}$$

où $m_{j=1,\dots,N_\lambda} \in \mathbb{R}^+$ sont des facteurs de multiplication d'intensité associés à chaque nombre d'ondes.

- Lorsque les perturbations représentent des décalages spectraux dans les nombres d'ondes plus grands ou plus petits, la matrice ϕ_{ik} est définie par

$$\phi_{ik} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \cdots & \cdots & 0 \end{pmatrix}_{N_\lambda \times N_\lambda} \quad \text{ou} \quad \phi_{ik} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}_{N_\lambda \times N_\lambda}$$

selon si le décalage spectral est de 1 cm^{-1} vers les nombres d'ondes plus grands ou plus petits. À noter que pour un décalage de x cm^{-1} ($x \in \mathbb{N}$), la matrice ϕ_{ik} est composée de 1 sur la x-ème diagonale de la matrice et de 0 partout ailleurs.

Bien que cela ne soit pas forcément vérifié en général, il est supposé que les coefficients non nuls de la matrice ϕ_{ik} doivent être positifs. L'ajout de cette contrainte supplémentaire permet de s'assurer que le produit $\phi_{ik}\, S_{REFS,k}$ reste positif pour tout $i \in \{1, \dots, N\}$ et $k \in \{1, \dots, K\}$. Par ailleurs, pour qu'elle puisse bien représenter une matrice définissant les perturbations spectrales engendrées sur le k-ème spectre de composant pur de référence dans le i-ème spectre de données, la matrice ϕ_{ik} doit être parcimonieuse, c'est-à-dire que seuls quelques-uns de ses coefficients peuvent être non nuls.

Au final, l'expression du modèle de mélange linéaire perturbé pour un spectre de données est la suivante :

$$\begin{aligned} D_i &= \sum_{k=1}^{K} C_{ik} \left(\phi_{ik}\, S_{REFS,k}\right)^T + E_i \\ &= C_i\, \phi_i^T\, (S_{REFS})^T + E_i \end{aligned} \qquad \text{(VI-9)}$$

avec ϕ_i contenant les matrices de perturbations des K spectres de composants purs à estimer. Cela peut être écrit de manière plus compacte pour tous les spectres de données sous une forme matricielle :

$$D = (C\Phi^T)\, (S_{REFS})^T + E \qquad \text{(VI-10)}$$

où $\Phi = [\phi_i]_{i \in N}$ rassemble l'ensemble des matrices de perturbations associées aux K composants purs pour chaque spectre de données i.

En définitive, en prenant en considération les possibles perturbations spatio-spectrales dans la modélisation des données et en tenant compte de la propriété de parcimonie des matrices de perturbations ϕ_i au moyen d'une régularisation ℓ_1 ($\Omega_4 = \alpha_4 \|\Phi\|_1$), le problème d'optimisation à résoudre deviendrait le suivant :

$$\min_{S\geq 0, C\geq 0, \Phi\geq 0, \sum_{k=1}^{K} C_k=1} \quad \|D-(C\Phi^T)\,(S_{REFS})^T\|_2^2+\frac{\alpha_1}{2}\|C\|_1+\frac{\alpha'_{1_1}}{2}\|C\|_2^2 +\frac{\alpha_2}{2}\|C\Delta\|_2^2+\frac{\alpha_3}{2}\left\|S_{REFS}-S^{(0)}\right\|_2^2+\alpha_4\|\Phi\|_1 \tag{VI-11}$$

1.1.5. Changement de la fonction de coût dans le terme d'attache aux données

Dans la littérature des méthodes de Factorisation Matricielle Non-négative (NMF : *Non-negative Matrix Factorisation*), la norme euclidienne est fréquemment utilisée pour mesurer les différences de similarité entre les données expérimentales et les données reproduites par le modèle de mélange linéaire. L'utilisation de cette fonction de coût quadratique s'explique par le fait que la plupart du temps, le bruit présent dans les données analysées est supposé suivre une loi normale. Par exemple en spectroscopie Raman, il est généralement admis que le bruit présent dans les données est de nature gaussienne, autrement dit sa distribution est supposée suivre une loi normale. L'utilisation de la fonction de coût quadratique dans les méthodes de démélange spectral, telles que la MCR-ALS et la MVC-NMF, pour mesurer la proximité des données reproduites par le modèle estimé avec nos données expérimentales semble donc théoriquement appropriée. Cependant, l'observation du bruit estimé dans les différentes images Raman utilisées tout au long de ce manuscrit ne semble pas toujours corroborer cette hypothèse. En effet, les histogrammes des spectres résiduels moyens de chaque image Raman utilisée dans les études présentées précédemment ont pour certaines une distribution plus proche de celle d'une loi de Student que de celle d'une loi normale (Figure VI-2). L'utilisation d'une fonction quadratique s'avère donc peut-être pas le plus adapté pour nos données spectrales.

Pour mieux s'adapter à la distribution du bruit contenu dans leurs données, Cavalcanti *et al.* [119]–[121] ont remplacé la norme euclidienne par une β-divergence pour mesurer les erreurs d'estimation du modèle linéaire par rapport aux données expérimentales. En effet, cette mesure de dissemblance leur offre plus de flexibilité puisqu'elle prend en compte trois types de distribution de bruit à elle seule : la distribution de Poisson ($\beta = 0$), la distribution Gamma ($\beta = 1$) et la distribution gaussienne ($\beta = 2$). Dans le même esprit que Cavalcanti *et al.*, nous suggérons d'adapter la fonction de coût associée au terme d'attache aux données à la distribution de bruit observée dans nos données spectrales. En d'autres mots, nous proposons d'utiliser une fonction de coût adaptée à une distribution de Student. Dans le domaine du traitement de vidéo, Y. Zhou et S. Maskell [122] ont développé un modèle linéaire pour estimer le mouvement de fond entre deux images d'une vidéo dans le but de déduire le mouvement de la caméra en supposant que la différence entre deux images suivait une loi de Student. La fonction de coût qu'ils ont employée correspond à la log-vraisemblance de la fonction de densité de la loi de Student et interpole à la fois la distribution Gaussienne ($\nu = +\infty$ avec ν le degré de liberté de la loi de Student) et la distribution Cauchy-Lorentz ($\nu = 1$). Cette fonction de coût pourrait donc être plus en adéquation avec nos données.

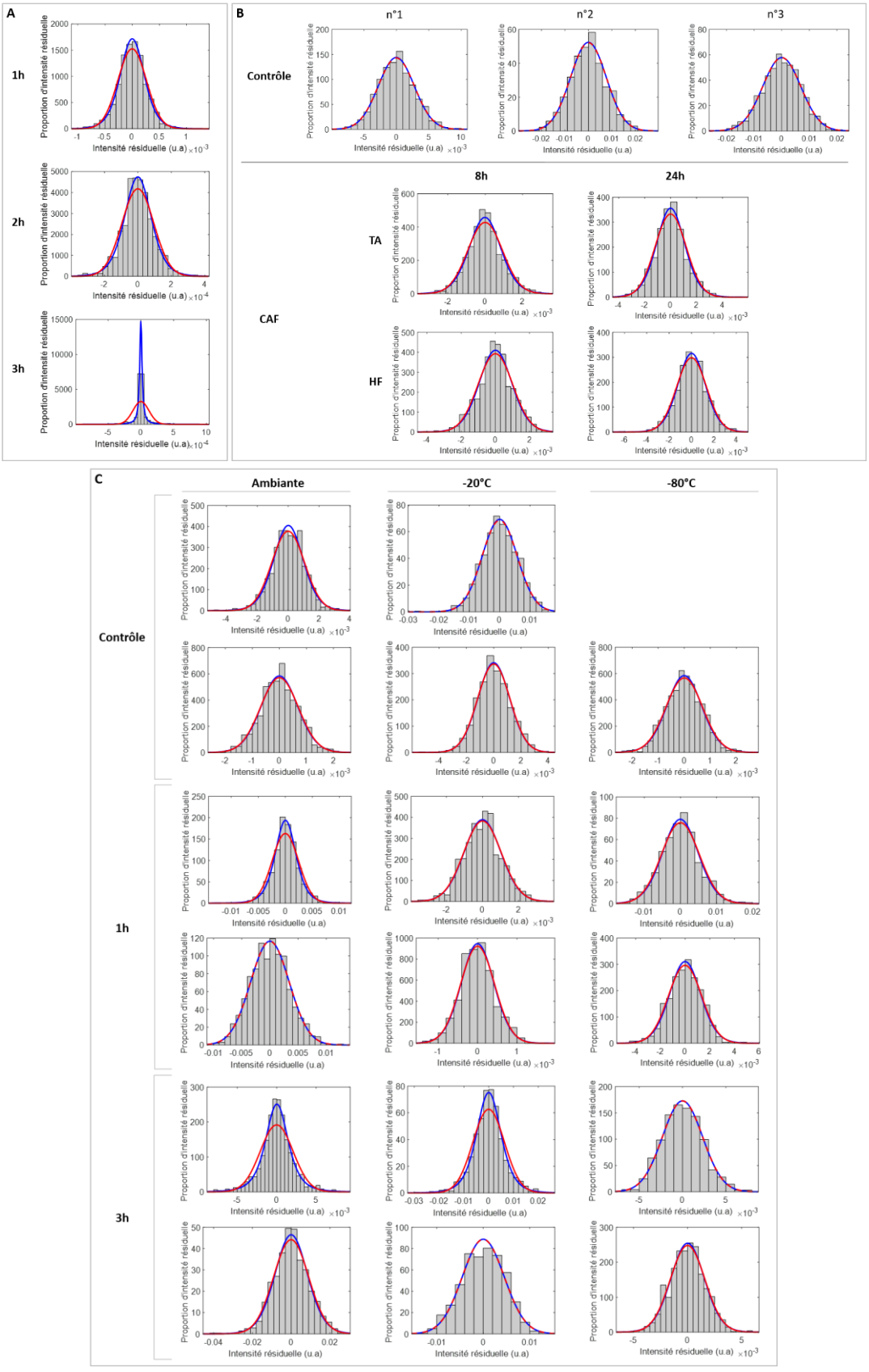


Figure VI-2 – Distribution du bruit présent dans les données Raman utilisées dans les études (**A**) *du suivi du Delipidol® et* (**B**) *de la détection et du suivi de la caféine dans un épiderme humain reconstruit, puis* (**C**) *pour l'étude des conditions de stockage des échantillons de peau reconstruit in vitro ; Une loi normale (courbe rouge) et une loi de Student (courbe bleue) sont ajustées à chaque histogramme ; Les histogrammes sont tracés à partir du spectre résiduel moyen de chaque image Raman.*

En notant $PDF(y)$ la fonction de densité de probabilité de la distribution de Student :

$$PDF(y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{y^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{VI-12}$$

pour $-\infty < y < +\infty$ avec $\Gamma(\cdot)$ la fonction Gamma et $\nu > 0$ degrés de liberté, la fonction de coût associée à la distribution de Student s'exprime par :

$$\Psi(y) = \frac{\nu+1}{2}\ln\left(1+\frac{y^2}{\nu}\right) - \ln\left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\right) \tag{VI-13}$$

En définitive, en tenant compte de toutes les améliorations suggérées dans les parties 1.1.1 à 1.1.5 de ce Chapitre VI, le problème de minimisation à résoudre serait le suivant :

$$\min_{S\geq 0, C\geq 0, \Phi\geq 0\ ,\sum_{k=1}^{K} C_k = 1} \sum_{i=1}^{N} \Psi\left(D_i - \left(C_i \phi_i^T\right)(S_{REFS})^T\right) + \sum_{n=1}^{4} \Omega_n \tag{VI-14}$$

1.2. CHOIX DE L'ALGORITHME DE MINIMISATION À UTILISER

Précédemment nous avons suggéré d'estimer un modèle de mélange linéaire perturbé d'expression $(C\Phi^T)(S_{REFS})^T$ pour reproduire au mieux nos données spectrales Raman enregistrées à partir d'échantillons de peau exposés à un AC. Plus exactement, nous avons proposé une nouvelle formulation de problème de minimisation à résoudre (VI-14) qui semble plus adaptée aux propriétés que respectent nos données Raman pour parvenir à les modéliser plus simplement. Dans ce nouveau problème d'optimisation, les matrices de concentrations relatives C, de perturbations spectrales Φ et de spectres de composants purs S_{REFS} sont toutes trois inconnues. Pour mener à bien la résolution de ce nouveau problème (VI-14), il semble judicieux de recourir à une approche de minimisation par bloc. Autrement dit, le problème d'optimisation doit être considéré sous forme de 3 sous-problèmes de minimisation à résoudre de manière alternative:

$$\begin{cases} \min\limits_{\substack{C\geq 0 \\ \sum_{k=1}^{K} C_k = 1}} & J(S_{REFS}, C, \Phi) \\ \min\limits_{S_{REFS}\geq 0} & J(S_{REFS}, C, \Phi) \\ \min\limits_{\Phi\geq 0} & J(S_{REFS}, C, \Phi) \end{cases} \tag{VI-15}$$

où $J(S, C, \Phi) = \sum_{i=1}^{N} \Psi\left(D_i - \left(C_i \phi_i^T\right)(S_{REFS})^T\right) + \sum_{n=1}^{4} \Omega_n$.

Cette façon de procéder évoque celle également utilisée dans les méthodes MCR-ALS et MVC-NMF : chacune d'elle résout deux sous-problèmes de minimisation, l'un en fonction de la matrice S et l'autre en fonction de la matrice C, de manière alternative. Cependant, ces deux méthodes ont recours au même algorithme d'optimisation pour résoudre les deux sous-problèmes de minimisation. Alors que la MCR-ALS utilise une régression linéaire des moindres carrés, la MVC-NMF opte plutôt pour une méthode de descente de gradient qui est plus adaptée en présence d'un terme supplémentaire contraignant le volume du simplex formé par les spectres de composants purs à être minimal. Cette contrainte supplémentaire concerne cependant uniquement la matrice S. L'algorithme de descente de gradient est donc assurément approprié pour résoudre le sous-problème de minimisation en fonction de la matrice S mais il n'est peut-

être pas le plus adéquat et le plus optimal pour résoudre le sous-problème de minimisation en fonction de la matrice C.

À l'instar de la MVC-NMF, des termes de régularisation différents et des contraintes différentes sont appliqués sur les matrices à estimer dans les trois sous-problèmes de minimisation $(\text{VI-}16)$ à résoudre. Chacun d'eux doit donc être pris en compte pour définir quelle méthode de minimisation il est préférable d'adopter. Toutefois bien que cela ne soit pas pratiqué dans les méthodes MCR-ALS et MVC-NMF, l'utilisation d'une stratégie de minimisation par bloc offre la possibilité d'utiliser un algorithme différent et adapté à chaque sous-problème. C'est même en réalité tout l'intérêt de cette approche. Par exemple, contrairement à la minimisation selon la matrice S_{REFS}, les minimisations selon les matrices C ou Φ ne pourront être réalisées par les méthodes de descente de gradient ou de Newton en raison de la non-différentiabilité des régularisations ℓ_1 appliquées sur ces deux dernières matrices. En revanche, des minimisations à l'aide d'un algorithme dit de Majoration-Minimisation (MM) pourraient être éventuellement plus adaptées. Une étude approfondie devra par conséquent être menée pour définir quel algorithme est le plus adapté pour chaque estimation de matrice.

1.3. TESTS DE LA ROBUSTESSE DU NOUVEL ALGORITHME

Une fois algorithme d'optimisation du problème $(\text{VI-}14)$ mis en place, sa robustesse devra être testée.

- Dans un premier temps, l'efficacité de cet algorithme pourra être évaluée sur les données générées synthétiquement dans le Chapitre V. À noter qu'à cette étape les données synthétiques étant générées sans aucune perturbation spectrale, la matrice des perturbations Φ devrait être estimée comme égale à la matrice identité. Ces premiers tests permettraient de vérifier si l'algorithme développé est capable de surpasser les problèmes liés à la multi-colinéarité entre spectres de composants purs et donc d'évaluer la pertinence d'ajouter des termes de régularisations sur des matrices dans la formulation du problème.

- Dans un second temps, l'algorithme développé pourrait être testé sur des données générées synthétiquement un peu plus complexes. Ces dernières pourraient cette fois-ci intégrer des perturbations de bandes (décalage, déformation, changement d'intensité, etc…) afin de vérifier la pertinence d'utiliser un modèle de mélange linéaire perturbé pour modéliser les données et/ou simuler des mélanges intégrant des diffusions progressives de composants purs afin de justifier l'intérêt de cet algorithme pour nos applications cosmétiques.

Les résultats obtenus par ces tests devront être évalués en termes de statistiques de modélisation globale et de qualité de décomposition spectrale à travers une analyse des résidus et d'une interprétation des spectres de composants purs estimés. Puis, seulement après validation des résultats de ces tests sur ces différents types de données synthétiques, l'algorithme pourra être appliqué sur des données réelles provenant d'échantillons de peau exposés à un AC.

2. CRÉATION D'UNE INTERFACE MATLAB

À ce jour, les industriels et les chercheurs du domaine cosmétique sont en quête d'outils de recherche et de diagnostic à appliquer sur leurs nombreuses données qui soient rapides et faciles d'utilisation afin d'obtenir sans trop d'effort des informations immédiatement compréhensibles pour la mise au point et le suivi de la stabilité d'une formulation ou encore pour qualifier et/ou contrôler la qualité d'un produit en observant ses composants, leur répartition et leur homogénéité. Pour répondre à ces besoins, une routine de prétraitement de données ainsi que

deux méthodes d'analyse de données ont été présentées dans ce manuscrit pour étudier la pénétration et les perturbations moléculaires induites suite à l'application d'un AC à la surface de la peau pour répondre à ces besoins.

Cependant, les acteurs du cosmétique ne sont pas forcément experts en traitement de données. Pour leur simplifier l'utilisation de ces méthodes, nous proposons de créer une interface utilisateur Matlab intégrant ces différentes méthodes. À noter qu'une fois validée, la nouvelle méthode suggérée dans la partie 1. de ce chapitre pourrait être également intégrée à l'interface utilisateur Matlab. L'utilisateur n'aurait donc pas besoin de connaissances particulières en analyse de données ou en programmation informatique pour accéder à ces méthodes puisqu'il n'aurait qu'à cliquer sur des boutons pour rentrer ses données dans l'application, réaliser les calculs sur les données et afficher les résultats. Il ne lui resterait plus qu'à lire et interpréter les résultats obtenus.

CONCLUSION GÉNÉRALE

Au cours de ce travail de recherche, quatre axes de recherche majeurs ont été abordés : i) la création d'un protocole de prétraitement de données spectrales adapté aux images Raman enregistrées à partir d'échantillons de peau humaine *in vitro* ou d'autres supports cosmétiques, ii) l'application de méthodes de démélange spectral, utilisées dans d'autres domaines d'application ou sur d'autres types de données, sur des images Raman réelles dans le cadre d'études cosmétiques ciblées, iii) la création d'une vérité de terrain permettant d'évaluer la pertinence d'utilisation des méthodes de démélange spectral pour répondre à nos problématiques cosmétiques, et iv) la suggestion de modifications à apporter dans la formulation du problème à résoudre afin d'améliorer et de mieux adapter une méthode de démélange spectral aux contraintes qu'imposent les images Raman enregistrées dans le cadre d'études cosmétiques.

Dans le cadre du premier axe, une analyse comparative de différentes méthodes de prétraitement spectral régulièrement employées en spectroscopie Raman a permis d'identifier celles étant les plus adaptées à nos données. Des étapes de prétraitement spectral supplémentaires ont été proposées afin de rendre le prétraitement des données complet : une méthode de linéarisation spectrale a été utilisée dans un soucis de réalignement spectral des données et une méthode de correction de raies cosmiques hybride a été proposée. Ce travail a finalement abouti à la mise en place d'un protocole de prétraitement de données Raman qui est jusqu'à ce jour inexistant.

Dans le cadre du deuxième axe, l'utilisation de deux méthodes de démélange spectral a été proposée pour analyser des images Raman enregistrées à partir d'échantillons exposés à un AC et estimer ses cinétiques de pénétration. Trois études portant sur la recherche cosmétique ont permis de démontrer l'applicabilité, le potentiel et l'intérêt d'utilisation de ces deux méthodes pour ce type d'analyse. Ces trois études ont même mené à des conclusions importantes pour le secteur cosmétique : les cinétiques de pénétration de deux actifs cosmétiques (le Delipidol® et la caféine) ont pu être définies et les dégradations engendrées par le stockage des échantillons d'épiderme humain reconstruit à -20°C et -80°C sur leur perméabilité et leur fonction barrière ont pu être mises en évidence. D'autre part, une attention particulière a été portée tout au long de la présentation des différentes étapes d'analyse afin de considérer ce travail de thèse comme un guide pratique pour toute personne souhaitant utiliser ces méthodes pour ce type d'analyse sur leurs données sans faire appel à un expert. Les expérimentations réalisées ont poussé les méthodes dans leur retranchement et il a été mis en évidence des limitations de performance lorsque les profils de pénétration cutanée des AC sont difficilement détectables dans les données Raman sans connaissance *a priori* sur leur référence spectrale. Il s'agit ici des conditions d'applications les plus difficiles qui restent à ce jour encore un défi en terme d'analyse multivariée. Un autre challenge est la capacité à prendre en compte les différentes perturbations spectrales locales engendrées suite aux interactions moléculaires entre les AC et les composants de la peau.

Dans le cadre du troisième axe, les données Raman enregistrées à partir de formulations de films de polymère, représentants des matrices d'application très utilisés en cosmétique, ont été présentées comme délivrant un fort potentiel pour créer une vérité de terrain. Des données synthétiques ont pu être générées de façon à être proches des données enregistrées à partir des formulations expérimentales et ont pu être considérées comme définissant une vérité de terrain avérée qui soit réaliste pour évaluer les performances des méthodes. Une analyse comparative des résultats obtenus sur ces données synthétiques et les données expérimentales avec la vérité

de terrain n'a pas permis de retrouver la qualité des résultats obtenus sur des données réelles. Ces résultats invitent à poursuivre le développement de la simulation de données ainsi qu'à étudier de manière plus approfondie la qualité des spectres estimés par les méthodes de démélange proposées lorsque les composants du mélange étudié ont des signatures spectrales proches ou lorsque des perturbations spectrales sont différentes localement. Ces deux limitations étant rencontrées lorsque des échantillons de peau humaine *in vitro* sont étudiés (existence de nombreux composants dans la peau dont les signatures Raman sont proches et présence de perturbations locales différentes selon les propriétés chimiques de l'AC appliqué en surface), ce travail a permis d'exposer les limites d'utilisation de ces deux méthodes pour répondre à notre problématique.

Dans le cadre du dernier axe, un plan d'action à mettre en place pour améliorer et adapter la méthode de démélange spectral à notre problématique et aux limitations mises en évidence dans la partie précédente a été proposé. La réflexion menée a nécessité l'utilisation de connaissances mathématiques. Les suggestions données ne sont que théoriques. Elles concernent la reformulation partielle du problème d'optimisation à résoudre pour mener à bien le démélange spectral des données Raman : il a été suggéré d'ajouter des termes de régularisation, de reconsidérer la formulation du modèle linéaire pour reproduire les données et de reconsidérer la fonction de coût à minimiser.

De façon plus générale, les études menées dans le cadre d'application cosmétique dans ces travaux de thèse ont démontré le potentiel du couplage de l'imagerie confocale Raman avec les méthodes d'analyse multivariée. Ces travaux mènent à reconnaitre cette combinaison de techniques comme outil d'analyse puissant aussi bien pour le dosage de molécules d'intérêt dans des formulations complexes ou l'objectivation de molécules actives que pour le contrôle qualité des produits cosmétiques. Ils ont permis de souligner l'applicabilité des méthodes d'analyse multivariée présentées et de dégager un référentiel pour leur mise en application pratique tant pour les non experts du traitement de données souhaitant réaliser ce type d'analyse que pour les spécialistes désireux de les utiliser ou d'aller plus loin dans le domaine. Enfin, ces travaux de thèse ont également mis à jour les potentielles failles des méthodes disponibles et des perspectives ont été proposées pour leur amélioration.

LISTES DES PUBLICATIONS

CONFÉRENCES NATIONALES

(1) Aline Stella, Franck Bonnier, Lynda Miloudi, Ali. Tfayli, Florent. Yvergnaux, Emilie Munnier, Clovis Tauber. “Estimation of penetration profiles with unsupervised unmixing approaches in Raman spectroscopic imaging,”. *Chimiométrie 2019*. Montpellier, France, Janv. 2019. [Oral]

(2) Aline Stella, Franck Bonnier, Lynda. Miloudi, Ali. Tfayli, Florent Yvergnaux, Emilie Munnier, Clovis. Tauber. “Estimation of prenetration profils based on unsupervised unmixing approaches applied to Raman imaging”. *Colloque Cosmétosciences*.Orléans, France, 15 Feb. 2019.[Poster]

(3) Aline Stella, Franck Bonnier, Lynda Miloudi, Ali. Tfayli, Florent. Yvergnaux, Emilie Munnier, Clovis Tauber : “Estimation de profils de pénétration à l’aide d’approches de démélange non-supervisées en imagerie spectroscopique Raman,”. *Recherche en Imagerie et recherché pour la Santé 2019 (RITS 2019)*. Tours, France, Mai 2019. [Poster]

(4) Aline Stella, Franck Bonnier, Ali Tfayli, Florent. Yvergnaux, Emilie Munnier, Clovis Tauber (2020): “Supervised vs unsupervised MCR-ALS to estimate spectral components and penetration profile in Raman imaging: application to active cosmetic ingredient tracking into the skin,”. *Chimiométrie 2020*. Liège, Belgique, Janv. 2020. [Poster]

CONFÉRENCES INTERNATIONALES

(5) Aline Stella, Franck Bonnier, Lynda Miloudi, Ali. Tfayli, Florent. Yvergnaux, Emilie Munnier, Clovis Tauber (2019): “Minimum Volume Constrained Non-negative Matrix Factorization Applied to the Monitoring of Active Cosmetic Ingredient into the Skin in Raman Imaging,”. *CPRS 2019: 10th International Conference on Pattern Recognition Systems.* Tours, France, Juil. 2019 [Oral et poster].

(6) Aline Stella, Franck Bonnier, Ali Tfayli, Florent. Yvergnaux, Emilie Munnier, Clovis Tauber: “Simultaneous estimation of spectra and penetration profile in Raman imaging by non-supervised MCR-ALS: application to active cosmetic ingredient tracking into the skin,”. *The Studium Conference: Skin Models in Cosmetic Science: Bridging Established Methods and Novel Technologies.* Tours, France, Déc. 2019. [Oral]

JOURNAUX INTERNATIONAUX

(7) Aline Stella, Franck Bonnier, Ali. Tfayli, Florent. Yvergnaux, Hugh J.Byrne, Igor Chourpa, Emilie Munnier, Clovis Tauber., “Raman mapping coupled to self-modelling MCR-ALS analysis to estimate active cosmetic ingredient penetration profile in skin,” *J. Biophotonics*, vol. 13, no. 11, pp. 1–13, 2020. DOI: 10.1002/jbio.202000136

DOI: 10.1002/jbio.202000136

Received: 9 April 2020 | Revised: 30 June 2020 | Accepted: 1 July 2020

DOI: 10.1002/jbio.202000136

Check for updates

JOURNAL OF BIOPHOTONICS

FULL ARTICLE

Raman mapping coupled to self-modelling MCR-ALS analysis to estimate active cosmetic ingredient penetration profile in skin

Aline Stella[1] | **Franck Bonnier**[2] | **Ali Tfayli**[3] | **Florent Yvergnaux**[4] | **Hugh J Byrne**[5] | **Igor Chourpa**[2] | **Emilie Munnier**[2] | **Clovis Tauber**[1*]

[1]UMR U1253, iBrain, Université de Tours, Inserm, Tours, France

[2]EA6295 Nanomédicament et Nanosondes, Université de Tours, Tours, France

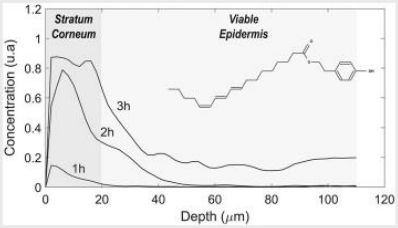
[3]U-Psud, Univ. Paris-Saclay, Chatenay-Malabry, France

[4]Bioeurope Solabia Group, R&D Cosmetic and nutrition, Anet, France

[5]FOCAS Research Institute, TU Dublin, Dublin, Ireland

***Correspondence**
Clovis Tauber, UMR U1253 iBrain, Université de Tours, Inserm, 10 Boulevard Tonnellé, 37032 Tours, France.
Email: clovis.tauber@univ-tours.fr

Abstract



Confocal Raman mapping (CRM) is a powerful, label free, non-destructive tool, enabling molecular characterization of human skin with applications in the dermo-cosmetic field. Coupling CRM to multivariate analysis can be used to monitor the penetration and permeation of active cosmetic ingredients (ACI) after topical application. It is presently illustrated how multivariate curve resolution alternating least squares (MCR-ALS) can be applied to detect and semi-quantitatively describe the diffusion profile of Delipidol, a commercially available slimming ACI, from Raman spectral maps. Although the analysis outcome can be critically dependent on the a priori selection of the number of regression components, it is demonstrated that profiling of the kinetics of diffusion into the skin can be established with or without additionnal spectral equality constraints in the multivariate analysis, with similar results. Ultimately, MCR-ALS, applied without spectral equality contraints, specifically identifies the ACI as one of main spectral components enabling to investigate its distribution and penetration into the stratum corneum and underlying epidermis layers.

KEYWORDS

confocal Raman mapping, Delipidol, human skin, multivariate curve resolution alternating least squares, penetration profiles

1 | INTRODUCTION

The dense corneocyte layers and lipid rich composition of the stratum corneum (SC) of the skin ensure a particularly effective defence against exogenous agents. It is well accepted that the resulting barrier function can strongly compromise the efficiency of active cosmetic ingredients (ACI) by limiting their penetration. However,

 © 2020 WILEY-VCH GmbH

DOI: 10.1002/jbio.202000136

many commercialized ACI in skin products have cellular targets located in deeper layers of the skin, below the SC. As a result, evaluation of cosmetic products, that is, biological testing, to validate the diffusion of the ACI into the skin cannot be circumvented. While chromatographic techniques can be considered as gold standards [1], enabling determination of the kinetics of penetration with high accuracy and sensitivity, large amounts of solvents are used for such analysis, making the method non eco-friendly (an increasing concern in the cosmetic industry). Moreover, tape stripping protocols employed for chromatographic analysis lead to lack of accuracy at a sub-micrometric level, due the nature of the mechanical separation of cell layers with adhesive films. Over the last 30 years, Raman spectroscopy has gained popularity in this specific field of application [2]. This non-destructive, label free and reagent (solvent) free technique provides molecular fingerprints based on the intrinsic, inelastic scattering of light, when the sample is illuminated with a laser source without adding any tracer. Moreover, coupling the molecular specificity of the analysis with the micrometric spatial resolution offers a relevant platform for evaluation of local diffusion and distribution of active molecules [3–9].

However, bio-analytical applications of Raman spectroscopy, along with near infrared (NIR) and Fourier transform infrared (FT-IR) spectroscopy, are hampered by the complexity of the spectral signatures originating from combined contributions from all the chemical compounds in the analysed area. Therefore, advanced data mining protocols are usually necessary to extract and interpret the spectral information inside the large data sets generated by hyperspectral mapping, and preprocessing can also play a key role in correcting or diminishing spectral variability originating from several optical and instrumental effects [10]. Recently, the development of multivariate analysis (ie, chemometrics) has greatly contributed to increasing the potential of spectroscopy mapping for tissue, cell and surface analysis of material but also in pharmaceutical and cosmetic fields [3, 7, 11–20]. Principal components analysis (PCA) and K-means clustering methods are examples of descriptive unsupervised methods, widely used to better understand the hyperspectral data [21]. The former, PCA, is often applied to reduce the number of variables in a multidimensional data set and the analysis of the identified principal components can provide some information about the origin of the variability of the data. The second approach, K-means clustering, is a method commonly employed in spatial mapping applications, by which regions of similar spectral profiles are associated. A number of examples of applications of PCA and/or K-means clustering to understand how skin components evolve under specific conditions [11, 12] and to detect the presence of exogenous molecules [7, 13] have been reported. Such methods are particularly interesting for interpretation of large hyperspectral maps containing hundreds to thousands of data points. However, they are unable to provide the spectral signatures of specific biochemical components and their respective abundance from data. Thus, to access the distribution of specific chemicals, it is necessary to use more elaborate unmixing methods. Decomposing each pixel of an image in a combination of estimated constituent spectra, according to their respective abundance, enables both qualitative and quantitative characterization. Therefore, biomedical profiles of cells and tissues can be analysed to access subtle changes linked to biological mechanisms or pathological development [22, 23]. Among the available unmixing approaches, classic least squares (CLS) and independent components analysis (ICA) are methods commonly employed in a wide range of industrial applications, including monitoring the penetration of molecules such as penetration enhancers, vitamins or other active ingredients [3, 13–15], control of pharmaceutical products or detection of counterfeit medicines [16, 17]. While CLS uses a collection of spectral references of all components in the sample to estimate their respective abundance, ICA only requires an estimation of the number of components present to decompose them. Singular value decomposition (SVD) is widely used to estimate this parameter, while Durbin-Watson Criterion (DWC) and Kaiser Meyer Olkin Index (KMOI) are also regularly used [24, 25]. However, since ICA does not impose a non-negativity constraint on the constituent spectra, the biochemical interpretation of spectral decomposition is more complicated.

A recent study has investigated the penetration of Delipidol through the skin by means of non-negative constrained least squares (NCLS) applied to skin cross sections [26]. Reconstructed distribution maps and penetration profiles of this molecule were estimated using a measured pure spectrum of ACI. However, this method presents some limitations as its performance relies on the availability of a large number of spectra from control samples but also on the feasibility of collecting highly reliable reference spectra from pure compounds. The use of such a reference spectral data base excludes the possibility of, for example, band shifts due to molecular interactions.

In contrast, multivariate curve resolution alternating least squares (MCR-ALS) regression analysis is an unmixing method which can provide an accurate molecular decomposition of the spectroscopic information

DOI: 10.1002/jbio.202000136

contained in the data set, while also dealing with such band shifts [27]. MCR-ALS has been applied to detect modification in ingredients and their dosing in falsified medicine [16, 18–20, 28]. The method has also been applied on FT-IR and Raman data to study effects of pathology, anatomy, environmental or genetic factors [27] and on Raman data collected from plant cell walls to study the mechanical stability of cells by changing their form, their thickness or their composition [29]. The advantages of MCR-ALS [30] over other multivariate approaches lies in the possibility to add some constraints to the model, such as non-negativity or equality constraints. The equality constraints can be a known spectrum or/and concentration, while retaining a reduced number of components. Additionally, the method does not require large control data sets. Therefore, MCR-ALS is an attractive alternative to most common methods proposed to evaluate the efficiency and innocuity of cosmetic products, starting with establishing their distribution profiles of the ACI to study their kinetics of penetration in skin layers.

This work aims to evaluate the MCR-ALS algorithm for the tracking of an ACI through the skin with Raman spectroscopy. The investigation addresses whether an a priori knowledge of the tracked molecule is necessary, and thus whether forcing one spectral signature to be that of the ACI signature changes the performance of the MCR-ALS approach in its data decomposition and fitting.

2 | EXPERIMENTAL DATA SET

The present work proposes to revisit and optimise the data analysis of previously published material [26]. The initial work was performed using NCLS and full details are provided in reference. Briefly, Raman spectra were acquired using a LabRam (Horiba Jobin-Yvon, Villeneuve-d'Ascq, France). Samples of *Punica granatum* seed oil hydroxyphenethyl esters, commercialized as Delipidol, were kindly provided by BioEurope Solabia, France Solabia (Pantin, France). It is a patented formulation used as an anti-cellulite ACI. Ethanolic solution of 5% Delipidol has been topically applied for 1, 2 and 3 hours onto human reconstructed skin samples (Episkin, Lyon, France). Raman maps were collected from 20 μm thick cross-sections prepared using a cryomicrotome and placed on Raman grade CaF_2 substrates.

As mentioned in [26], ethanol is a solvent which can act as penetration enhancer leading to superior diffusion of Delipidol. However, the aim of the experiment was to construct a set of Raman images with noticeable variations in the ACI concentrations to illustrate the application of multivariate analysis, that is, MCR-ALS.

3 | DATA PROCESSING AND ANALYSIS

3.1 | Data pre-processing

All Raman spectra were subject to pre-processing, including spectral denoising, baseline correction, followed by normalization. All pre-processing steps were performed in MATLAB (R2017b).

3.1.1 | Spectral denoising

The spectra were first filtered with a PCA to remove the noise [31]. PCA is a multivariate data analysis technique that aims to summarize and simplify the information available with as little information loss as possible. This simplification consists of transforming the variables to ones uncorrelated with each other. Once the uncorrelated principal components have been selected, all data are projected onto them, to go back to the original space. Hence, the noise is reduced. Note that, in the spectroscopy field, it is common to reduce the dimensionality of the spectral data set while preserving its main features to make it possible to remove the noise [32, 33]. In this work, 1% of the number of spectral data was used as the number of loadings for the data projection (more details in Figure S1).

3.1.2 | Baseline correction

During spectral acquisition, instrumental phenomena, scattering or intrinsic fluorescence by certain molecules under the excitation laser can introduce a background to spectra. This baseline has to be corrected prior to data analysis.

The Rubberband method is a well-established and previously employed background correction algorithm [10, 34–37] that was preferred in this study.

The Rubberband method estimates a piecewise polynomial baseline. First, a set of supports points is determined, such that the region below these points form a convex hull. Then, a polynomial curve is estimated between each support point. The polynomial order was set to 1 to avoid overcorrection of spectra. Finally, the set of these estimated polynomials compose the baseline and they are subtracted from the spectra (Figure S2).

3.1.3 | Normalization

The last step of the pre-processing corresponds to the data normalization. In this work, it was achieved by

DOI: 10.1002/jbio.202000136

calculating the ratio of spectra to their respective Euclidian norms.

3.2 | Data analysis

3.2.1 | Multivariate curve resolution alternating least squares

Unmixing methods assume that the data matrix D, of size $N \times N_\lambda$ where N and N_λ are respectively the number of pixels/spectra and Raman shifts, can be expressed as a linear combination of pure constituent spectra, weighted by their respective abundance (Figure 1). Mathematically, the data D are considered to be the product of a pure spectra matrix S^T, composed of K components, multiplied by the matrix of relative concentration C of these K components in each pixel, as following:

$$D = CS^T + E, \tag{1}$$

where E is the residual matrix of the linear model (1).

Furthermore, the MCR-ALS algorithm is a self-modelling method. It resolves an optimization problem through the minimization of the adjustment errors E of the linear model. The minimization problem to be solved is expressed as:

$$\min \left\| D - CS^T \right\|. \tag{2}$$

Some constraints have been added to this optimization problem to respect some logical and chemical properties, such as positivity of the estimated Raman spectra and their relative concentrations. The non-negativity constraint on spectra and relative concentration is introduced to the minimization problem, such that Equation (1) becomes:

$$\min_{S \geq 0, C \geq 0} \left\| D - CS^T \right\|^2. \tag{3}$$

The minimization problem (3) can be solved by different algorithms. The MCR-ALS method is one of them and the algorithm is designed to better fit the model to the data by alternatively minimizing residuals by least squares fitting, according to the spectral matrix S and to the associated abundance matrix C. Hence, the ALS aspect of the implementation is expressed as:

$$C = DS\left(S^T S\right)^{-1}, \tag{4}$$

$$S^T = \left(C^T C\right)^{-1} C^T D. \tag{5}$$

Ultimately, MCR-ALS relaxes the hypothesis of the presence of pure components within the data and it assumes that each pixel is a mixed signal (Figure S3).

Although the MCR-ALS algorithm aims at solving a global minimization problem, its initialization can influence the quality of the results. Therefore, it is preferable to use a method such as SIMPLISMA (SIMPLe-to-use-interactive-self-modelling-mixture-analysis) [25] to provide initial estimations of pure spectra, somehow close to the solution, as a starting point, rather than a random approach that could lead to unrealistic, outlier local minima [38].

SIMPLISMA estimates components of a mixture based on the purity of spectra, which is defined as a modified coefficient of variation:

$$P_i = \frac{\sigma_i}{\mu_i + \beta}, \tag{6}$$

where P_i is the purity coefficient for a given spectrum S_i, σ_i the SD, μ_i the mean spectrum and β is a coefficient offset between 1% and 5% of the mean. The higher the coefficient, the more significant is the difference to the mean, and therefore the component spectrum is considered purer. Ultimately, the purest spectra are selected based on the purity coefficient and also on their orthogonality to each other.

Alternatively, if constituent spectra or/and relative concentrations are known, the MCR-ALS method allows the addition of constraints to impose this information onto the constructed model. It consists of including the known spectra into the spectral initialization and setting the constraints on the spectra and relative abundances in

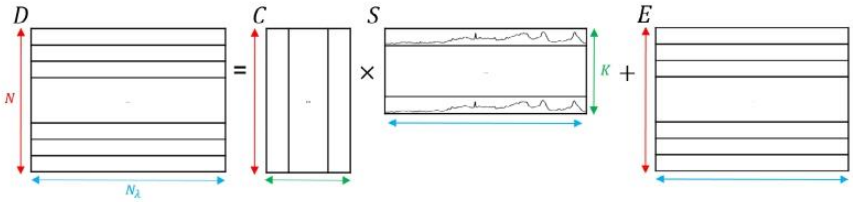


FIGURE 1 Linear mixing model

DOI: 10.1002/jbio.202000136

the respective matrices S and C at each iteration. It should be noted, that, in the absence of calibration, concentrations C of constituents are relative and their manifestation in the complex mixture is dependent on the relative strength of the Raman scattering cross sections, giving rise to the spectra S.

The MCR-ALS algorithm converges when the difference in the model adjustment between two iterations is smaller than a set threshold. MCR-ALS was applied using a multiset approach, that is, all images corresponding to 1, 2 and 3 hours have been decomposed simultaneously to consider the kinetics of Delipidol penetration and increase the coherency of the component extracted to more specifically reconstruct semi-quantitative distribution maps.

It should be noted that the application of added constraints such as non-negativity of spectra/abundances or correspondence of one component with a reference spectrum and the multiset approach promote singularity of the solution and reduces its rotational ambiguity [39]. The MCR-Bands algorithm described in [39] makes it possible to estimate the rotational ambiguity of a solution. It consists of calculating the relative signal contribution of every component of the maximum and minimum band boundaries of the feasible solutions and computes their differences. In the case where there is no rotational ambiguity, the estimated maximum and minimum band boundaries are close to the solution and the computed difference is close to 0. In this work, MCR-ALS and MCR-bands algorithms have been applied using the scripts from MCR-ALS GUI 2.0 software in Matlab [40].

3.2.2 | Number of components

Components are estimated Raman signatures used to decompose each pixel of the image. Therefore, it is necessary to set the number of components K to be estimated in the MCR-ALS computation. The parameter K plays a critical role in the decomposition of spectral maps, and the difficulty with biological samples lies in the fact that this number is unknown and difficult to estimate. Therefore, in order to reduce the subjectivity, different mathematical approaches are available to provide an estimation and guide the analysis. The most commonly used is SVD. Two other criteria, the Durbin-Watson criterion [25] and the Kaiser-Meyer-Olkin Index (KMO) [24] have also been recently proposed.

All these criteria have been applied to the evaluation of Raman maps of reconstructed skin samples exposed to Delipidol. While SVD is determined directly from the data maps, the other two decision criteria are based on SIMPLISMA outputs using a number of constituents ranging from 1 to 20. Further information can be found in supplementary data.

3.2.3 | Analysis tools

Some statistic tools are needed to analyse the data decomposition resulting from MCR-ALS. Correlation coefficients are used to quantitatively compare spectral components and highlight any possible matches to reference spectra, while histograms of experimental residuals are used to qualitatively evaluate the model fitting. They are described below:

Correlation coefficient – The correlation coefficient measures the linear link between several variables. Two components are positively and negatively correlated if their correlation coefficient is respectively close to 1 and − 1. Note that a null value does not imply independence. Moreover, the squared correlation coefficient, known as coefficient of determination, corresponds to the percentage of variations in the spectral values of a component that can explain the variations of the spectral values of another one. In other words, this squared value corresponds to the recovery rate of spectral bands between two components. In general, a correlation coefficient superior to 0.9 in absolute value is statistically considered as very high correlation, and therefore such a correlation coefficient is considered to be satisfactory, since more than 80% of variations of spectral values will be interrelated [41].

Histogram of residuals – A histogram is a graphical representation to visualize the data set repartition. All ranges of values are binned in classes, categorized according to the number of values included in them. In other words, for this work, all Raman intensities of mean model residuals are organized in a series of intervals and the number of Raman shifts having a Raman intensity included in each of them is associated to them. The histogram shape provides some information about the value distribution; in particular, a Gaussian shape means that the residual noise is Gaussian.

4 | RESULTS AND DISCUSSION

4.1 | Spectral characterization

Delipidol has a characteristic Raman signature, with a combination of intense specific peaks easily discernible and weaker spectral bands partially overlapping with skin features (Figure 2E). As such, it is an interesting candidate to evaluate unmixing approaches such as MCR-ALS. Spectral characterization of Delipidol is

DOI: 10.1002/jbio.202000136

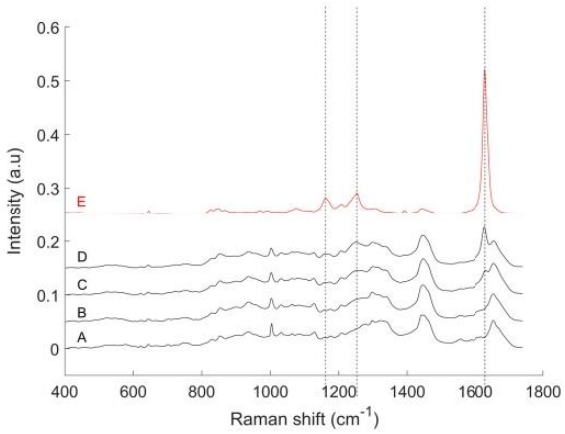


FIGURE 2 Comparison of data mean spectrum at each exposure time to the ACI with the reference spectrum of the ACI. **A**: control; **B**: 1 hour; **C**: 2 hours; **D**: 3 hours and **E**: pure spectrum of Delipidol. Data are pre-processed with the Rubberband baseline. ACI, active cosmetic ingredients

presented in Figure 2E, showing dominant spectral features at 1630 cm^{-1} (C═C stretching of the aromatic ring), 1254 cm^{-1} (C─C─C) and 1162 cm^{-1}. Other weaker bands at 1441, 1393, 1307, 1251, 1207, 1076, 991, 969, 865, 843 and 827 cm^{-1} can also be observed.

Mean spectra of Raman maps of control, 1, 2 and 3 hours exposure time are also illustrated in Figure 2 as an overall initial comparison of spectra. For illustration purposes, mean spectra presented have been obtained using the pre-processing strategy described above. In addition to those of Delipidol, Raman spectra collected from skin samples exhibit a complex signature originating from combined spectral features of physiological constituents. The amide I band located at 1655 cm^{-1} indicates the skin is predominantly composed of proteins such as keratin and other cellular proteins [11]. Content may vary between the upper layers corresponding to SC and other layers containing higher proliferating cell density (eg, basal layer). The position and shape of the amide I band reflects changes in protein composition but also in their conformation (secondary structure) [42], and therefore can substantially vary across the skin tissue section. Similarly, spectra contain a wealth of information about lipids, contributing peaks identified at 1064, 1129 and 1297 cm^{-1} but also bands of DNA and nucleic acids observed at 783 cm^{-1} (O─P─O), 813 cm^{-1} (O─P─O), 1086 cm^{-1} (O─P─O), 1252 cm^{-1} (guanine, cytosine), 1324 cm^{-1} (DNA) and 1340 cm^{-1} (DNA) [42]. However,

Delipidol exhibits a characteristic Raman signature, with a combination of intense specific peaks and weaker spectral bands (Figure 2E), which are easily discernible superimposed on the spectrum of skin (Figure 2A-D) even in the mean spectra of the skin cross-section, in which its localized contribution is naturally diluted. As a result, each pixel of the image can be considered a complex spectral mixture corresponding to the local biochemical composition tempered with contributions from the ACI. MCR-ALS applied to Raman maps is expected to deliver more specific information about its distribution in the different layers of the epidermis and consequently enable construction of penetration profiles.

4.2 | MCR-ALS applied to ACI penetration study using $K = 4$ components with non-negativity constraint only

This section presents results of MCR-ALS applied only with non-negativity constraints on spectra and relative concentrations and no ACI reference to guide the algorithm, using $K = 4$ components. The four estimated spectral components are displayed in Figure 3A, with corresponding reconstructed distribution map for Delipidol. Results are illustrated with data subjected to Rubberband baseline correction, derived from a multiset analysis, including maps from 1, 2 and 3 hours.

Spectral components are identified as S_i for $i = 1, ..., 4 \in \mathbb{N}$ (Figure 3A1) and their display order has been organized so that they are more easily readable and comparable. Compared to the reference spectrum of Delipidol, spectral component S_1 exhibits strong similarities. Although a shoulder at 1661 cm^{-1} suggests that a small residual feature from the amide I band remains, the MCR-ALS applied with only non-negativity constraints to spectra and abundances was able to identify a spectral component with strong similarities to the pure spectrum of the ACI studied (correlation coefficient of .9746), highlighting that the reference spectrum is not necessary to construct accurate distribution maps. Distribution maps for components S_2, S_3 and S_4 obtained with $K = 4$ are displayed in Figure 4. These components exhibit features related to biochemical constituents of the skin, such as lipids, proteins and nucleic acids [11], consistent with those observed in the mean spectra of Figure 2. Although it is not the subject of the current study, the combination of Raman/MCR-ALS can be further employed to study the skin composition, and any changes to it as a result of permeation of the ACI.

Moreover, the value of the differences of the relative contributions of the four components of maximum and minimum band boundaries estimated by application of the MCR-bands algorithm is respectively 1.16×10^{-3}, 8.00×10^{-3}, -5.81×10^{-3} and -25.61×10^{-3} for S_1, S_2, S_3 and S_4, indicating that the rotational ambiguities are negligible.

DOI: 10.1002/jbio.202000136

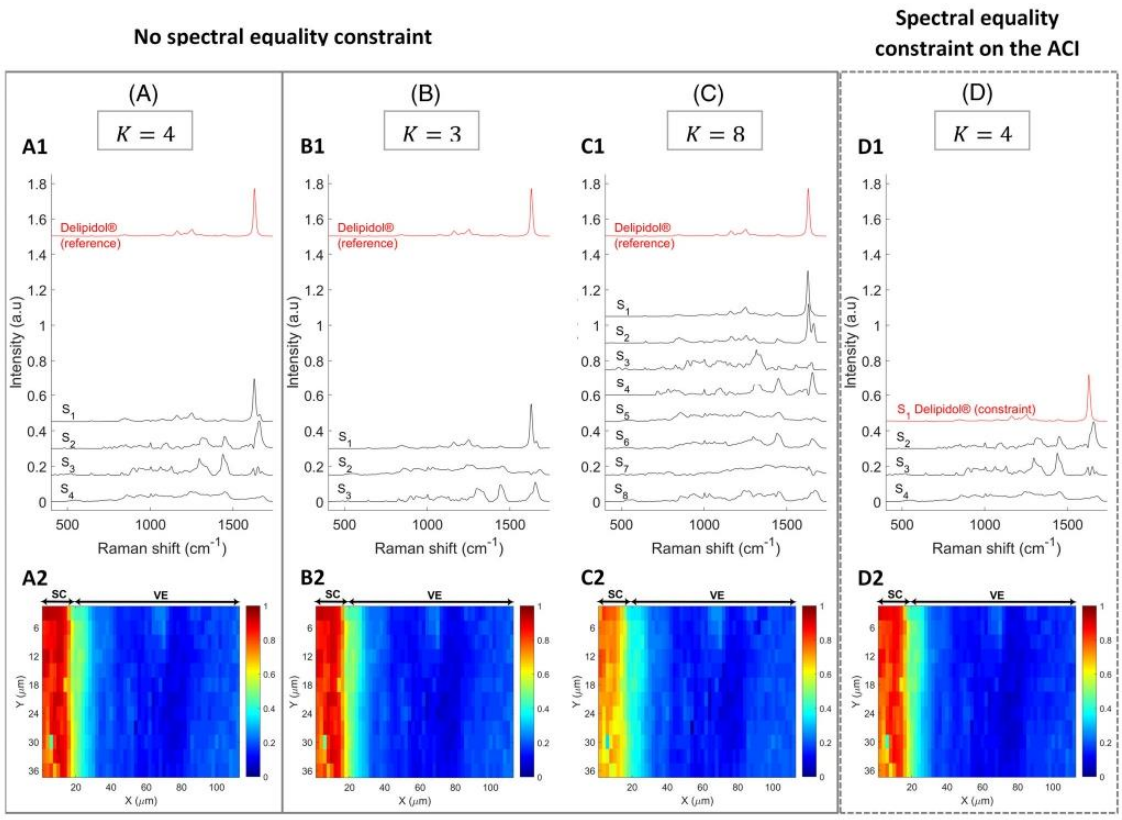


FIGURE 3 MCR-ALS applied with only non-negativity constraints for $K = 4$ **A**, and $K = \{3, 8\}$ B, C, and with an additional spectral equality constraint on the ACI for $K = 4$ D, on pre-processed data; A1-B1-C1-D1: Spectra estimated with K components initialized by SIMPLISMA (offset for clarity); A2-B2-C2-D2: Reconstructed maps of Delipidol in skin samples at 3 hours exposure time. Maps obtained using estimating spectrum of ACI S_1 (A2-B2-C2) or referential spectrum S_1(D2). ACI, active cosmetic ingredients; MCR-ALS, multivariate curve resolution alternating least squares; SIMPLe-to-use-interactive-self-modelling-mixture-analysis

For each exposure time, penetration profiles of the ACI were averaged over all samples (black curves in Figure 5) displaying the permeation of Delipidol into the SC for the 1 hour experiment, followed by progressive diffusion into deeper layers of the epidermis after 2 and 3 hours exposure with a concomitant increase in the outer layer (0–20 μm). It was observed that average penetration profiles calculated using the MCR-ALS applied with only non-negativity constraints were close to the profiles obtained with a more constrained approach such as NCLS [26] (Figure 5).

4.3 | Comparison of numbers of K parameters

The choice of $K = 4$ in Section 4.2 was rather subjective and based on the similarity of S_1 with the pure spectrum of Delipidol. However, the number of parameters chosen can have an impact on the model performance.

It was anticipated that selecting a number of components for the MCR-ALS applied to biological samples would be a difficult task. Despite the existence of criteria to support the analyst in this decision, the observation of SVD, DWC and KMOI all suggested that a high number of components would be necessary to encompass the entire biochemical spectral variability (Figure S4). However, the purpose of using advanced unmixing methods is to provide simplified information about the distribution of Delipidol which is easily interpretable. Moreover, the criteria described took into account the global information contained in the spectral maps and they did not necessarily directly reflect the optimal number of components to accurately and specifically detect and localize the ACI onto chemical maps.

DOI: 10.1002/jbio.202000136

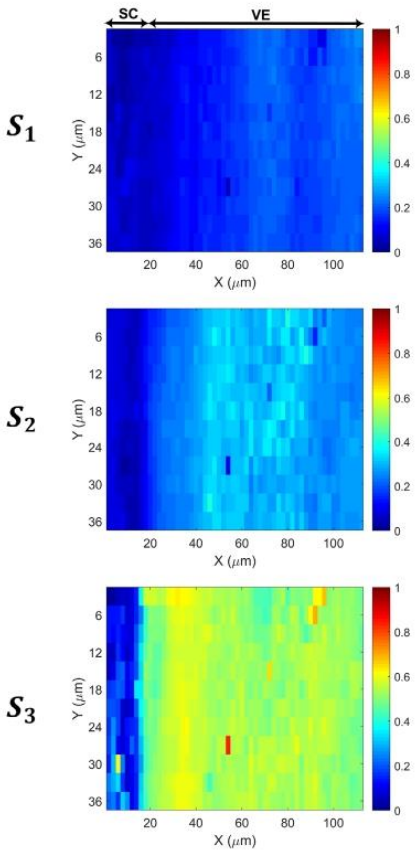


FIGURE 4 Reconstructed concentration distribution maps of S_1, S_2 and S_3 in skin samples at 3 hours exposure time. They resulted from MCR-ALS applied only with non-negativity constraints applied with $K = 4$. MCR-ALS, multivariate curve resolution alternating least squares

Plots obtained using SVD, DWC and KMOI are provided in Figure S4. There is no unambiguous decision about the optimal number of components to use for the spectra decomposition. While SVD suggests 11 components would be necessary to reach 95% of explained variance, DWC and KMOI tends to indicate that numbers between 4 and 8 could capture the relevant spectral information contained in Raman maps. Here are presented results from different sizes of the spectral base ($K = \{3, 8\}$) to better understand the impact of the parameter K on the spectral unmixing and the model fitting to the data. Hence, in order to evaluate the reliability of MCR-ALS, the correlation coefficients have been computed to estimate similarities between spectral components and the reference spectrum of Delipidol, and the interpretation of experimental residuals distribution by means of histograms has been explored.

Spectral components S_i for $i = 1, ..., K \in \mathbb{N}$ with $K = \{3, 8\}$ used for the MCR-ALS applied with only non-negativity constraints are displayed in Figure 3B1and Figure 3C1, with corresponding reconstructed distribution maps for Delipidol in Figure 3B2-C2. While distribution maps for 3 hours exposure time appeared similar for $K = 3$ and $K = 4$, the outcome with $K = 8$ diverged for the first two models with clearly an significantly lower abundance of Delipidol in the 20 μm, corresponding to the SC. Observations of spectral components highlighted that, with $K = 8$, features of Delipidol at 1630, 1442, 1254 and 1162 cm^{-1} have been also found in S_2 along with other skin features. It is an example of over decomposition of the spectral component of interest. While increasing the number of K parameters should deliver overall a better fitting of the model, presently the Delipidol is no longer considered as single pure component thus no interpretation of its distribution can be made.

In comparison S_1 components respectively calculated with $K = 3$ and $K = 4$ appears quite comparable. To estimate the degree of similarity correlation coefficients of estimated spectral components using $K = \{3, 4\}$ are presented in Table 1 (table including $K = 8$ is provided in Table S1). To be noted that S_1 Delipidol refers to the pure spectrum of the ACI. It is observed that the correlation coefficients between the spectral component of Delipidol estimated from the algorithm using only the non-negativity constraints and the reference pure spectrum used (S_1 Delipidol) are really high, with respective values of .9753, .9746 for $K = 3$ and $K = 4$. Moreover, the estimated spectral component S_2 and S_3 for $K = 3$ and $K = 4$ do not match with the reference ACI spectrum (correlation coefficients are respectively .0054, .2664 and .2518, .0681), suggesting the component S_1can be considered close to spectroscopically pure. In contrast, S_2 for $K = 8$ has a relatively large correlation coefficient of .8801 with the signature of Delipidol (Table S1). As mentioned above, this is a result of over-decomposition of the spectral signatures resulting with presence of Delipidol features in multiple spectral components.

The study investigates the applicability of MCR-ALS to construct distribution maps of the ACI as function of time. It is expected that a gradient will be observed from the surface of the skin (SC) toward deeper layers of the viable epidermis (VE). Therefore, the comparison criteria between models relies on the abundance of Delipidol according to depth.

Here, numerical values of the relative mean concentrations for different depth ranges 0-20 μm, 20-40 μm and 40-110 μm for $K = \{3, 4\}$ have been calculated and are presented in Table 2 (top part for analysis using only non-negativity constraints). These values are also available for $K = 8$ in Table S2. For $K = 3$ and $K = 4$ sensibly

DOI: 10.1002/jbio.202000136

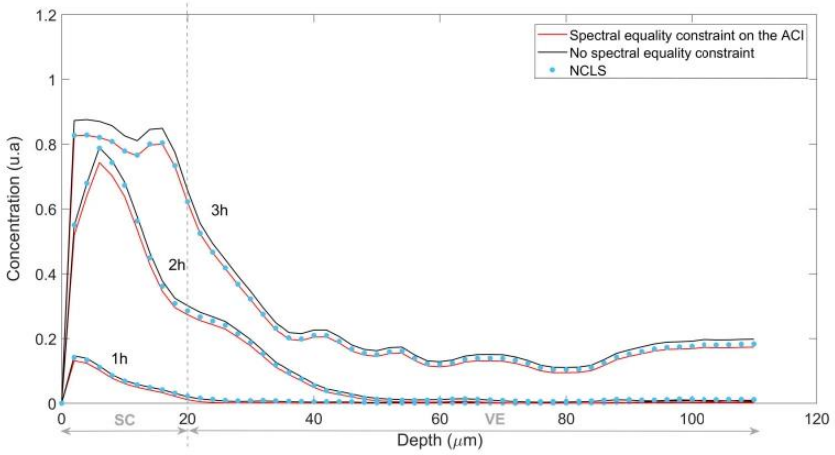


FIGURE 5 Average penetration profiles of the Delipidol for 1, 2 and 3 hours of exposure time obtained by MCR-ALS applied with (red) vs without (black) spectral equality constraints on the ACI and using $K = 4$. The average concentrations obtained by NCLS are illustrated by blue points. ACI, active cosmetic ingredients; MCR-ALS, multivariate curve resolution alternating least squares; NCLS, non-negative constrained least squares

TABLE 1 Correlation coefficients between estimated spectra $S_{l=1,\ldots,K}$ obtained by MCR-ALS applied with and without spectral equality constraints on the ACI ($K = \{3, 4\}$) from data corrected with Rubberband method

	$K = 3$			$K = 4$			
	Spectral equality constraint						
No spectral on the ACI equality constraint	S_1 **Delipidol**	S_2	S_3	S_1 **Delipidol**	S_2	S_3	S_4
S_1	0.9753	0.0512	0.3452	0.9746	0.3218	0.1393	0.1821
S_2	0.0054	0.9994	0.4070	0.2518	0.9983	0.4214	0.3911
S_3	0.2664	0.4425	0.9971	0.0681	0.4535	0.9979	0.7216
S_4				0.1110	0.4206	0.6790	0.9999

Abbreviations: ACI, active cosmetic ingredients; MCR-ALS, multivariate curve resolution alternating least squares.

TABLE 2 Mean concentration of estimated ACI after 1, 2 and 3 hours exposure time obtained by MCR-ALS applied with and without spectral equality constraints on the ACI for $K = \{3, 4\}$ from data corrected with Rubberband method

		$K = 3$			$K = 4$		
		1 h	2 h	3 h	1 h	2 h	3 h
No spectral equality constraint	0 − 20μm	0.0635	0.5439	**0.8050**	0.0691	0.5364	**0.8098**
	20 − 40μm	0.0012	0.1737	0.3477	0.0067	0.1753	0.3498
	40 − 110μm	0.0058	0.0129	0.1532	0.0036	0.0115	0.1623
Spectral equality constraint on the ACI	0 − 20μm	0.0534	0.5072	**0.7643**	0.0586	0.4995	**0.7638**
	20 − 40μm	0.0001	0.1553	0.3246	0.0013	0.1567	0.3247
	40 − 110μm	0.0016	0.0043	0.1343	0.0009	0.0049	0.1426

Abbreviations: ACI, active cosmetic ingredients; MCR-ALS, multivariate curve resolution alternating least squares.

identical results are obtained for 1, 2 and 3 hours. For instance, after 3 hours of exposure, relative mean respective concentrations are 0.8050 and 0.8098 in the 0 to 20 μm range. Delipidol reached inferior layers of the epidermis progressively (40-110 μm) and remained very concentrated on the upper skin layer (0-20 μm) after 3 hour.

Figure 6C,D presents the distributions of model residuals obtained by MCR-ALS using only non-negativity constraints for $K = \{3, 4\}$. The analysis of residuals can bring additional insights about the impact of the number of components K used. While for $K = 4$ residuals display a nearly normal Gaussian distribution, which can

DOI: 10.1002/jbio.202000136

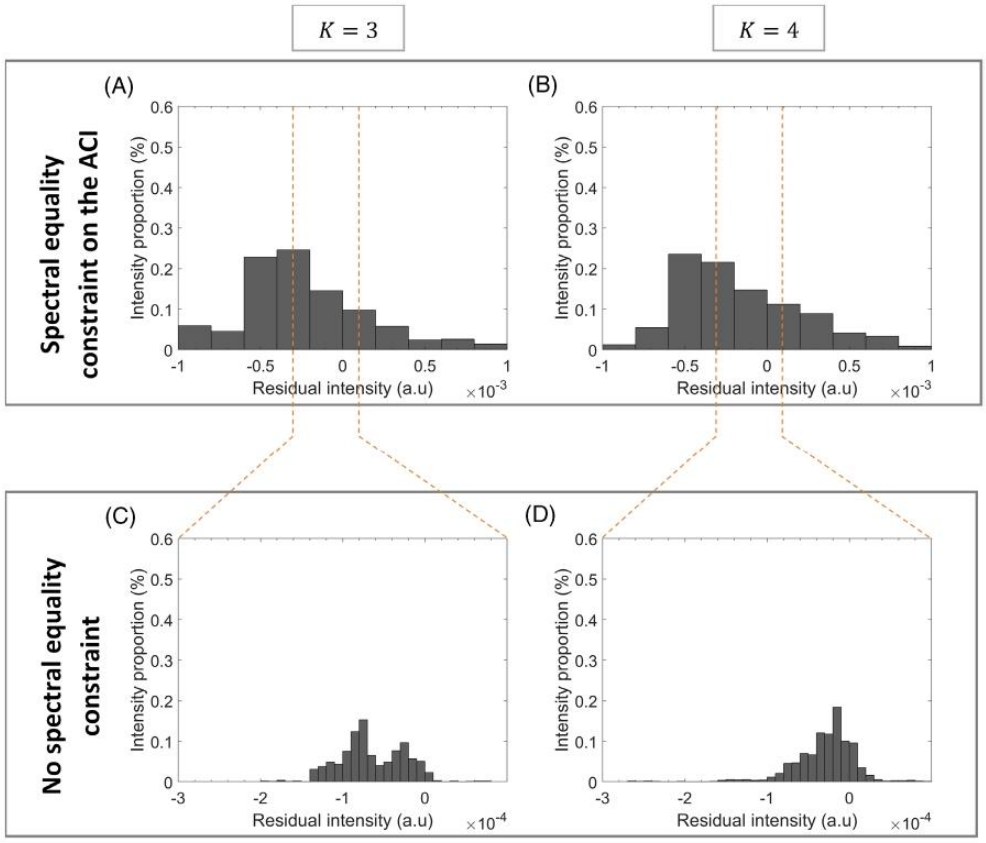


FIGURE 6 Histograms of mean residuals obtained by MCR-ALS applied with A-B, and without C-D, spectral equality constraints on the ACI for $K = \{3, 4\}$ from data preprocessed with Rubberband baseline correction. ACI, active cosmetic ingredients; MCR-ALS, multivariate curve resolution alternating least squares

correspond with the remaining noise, the shape of the distribution of residuals for $K = 3$ is composed of two Gaussian, which tends to indicate that not all spectral components have been extracted from the data. The right-hand Gaussian part of the distribution can be interpreted as residual noise, whereas the left Gaussian indicates that some data variability is mathematically not modelled. Increasing to $K = 4$ restores a unimodal Gaussian distribution, suggesting that no other significant spectral components remain in the residuals. Moreover, it is noted that the Gaussian distribution of the residuals shifts to more centered values when K is increased, with a central tendency shifting from approximately -0.85×10^{-4} to -0.25×10^{-4} and finally 0.05×10^{-4}, for respectively$K = 3$, $K = 4$ and $K = 8$. It suggests there is a mathematical overexpression of some data variability in the estimated model when K is too small.

MCR-ALS enables extraction of the spectrum of ACI from the data set collected from the skin samples. Selecting too few components leads to an unsatisfactory model in terms of fitting errors. While the correlation coefficients (Table 1) and relative mean concentrations (Table 2) indicate that $K = 3$ and $K = 4$ deliver sensibly similar results, the experimental residuals suggest optimal fitting is obtained with 4 components. Ultimately, $K = 4$ is a good trade-off between an acceptable fitting error and a coherent spectral decomposition.

4.4 | Comparison of MCR-ALS applied with and without spectral equality constraints on the ACI

Applying spectroscopic techniques for the analysis of penetration and/or permeation profiles of ACIs in the cosmetics field generally means the products themselves are available, and that pure, or reference, spectra can be collected before application onto the skin. In terms of

DOI: 10.1002/jbio.202000136

application of multivariate analysis protocols, these reference spectra can be either employed to guide the analysis, by using them as spectral equality constraints, or to validate the analysis. In the following, the reference spectrum of Delipidol is used as a spectral equality constraint on the ACI to help MCR-ALS to estimate its relative concentrations.

As mentioned in [27], when comparing methods with and without spectral equality constraints, it should be kept in mind that inserting more constraints in the MCR-ALS algorithm will naturally generally lead to a higher lack of fit due to a set of conditions to be respected that diminishes the flexibility of the model to find the optimal solution. Indeed, as MCR-ALS adjusts the fitting of spectral components through an iterative algorithm, using a fixed reference spectrum of a component leads to fewer degrees of freedom than an approach using only the non-negativity constraints. This should be considered when analysing the residuals. Adding some spectral equality constraints on the ACI can, however, in some situations, reduce overfitting to spectra that MCR-ALS would estimate as pure but that could actually have a mixed character. For the current application, MCR-ALS was applied with $K = 4$ components and with a spectral equality constraint on the first component of the model, S_1, fixed as a measured pure spectrum of Delipidol, while components S_2, S_3 and S_4 were calculated by the algorithm. As expected, adding a spectral equality constraint on S_1 leads to higher residuals (Figure 6). The analysis using a spectral equality constraint on the ACI indicates distributions closer to normal, but the fitting error is tenfold higher (10^{-3}) compared to that not including it (10^{-4}). The lack of fit [27, 40] was also calculated. The estimated model obtained by MCR-ALS applied without any spectral equality constraint had a lack of fit of 5.21%. As expected, adding a spectral equality constraint on the ACI in the analysis increased this lack of fit by 2.17%.

Nevertheless, the estimated spectral components and the reconstructed distribution maps for Delipidol displayed in Figure 3D do not exhibit any noticeable difference compared to the ones obtained by MCR-ALS applied without spectral equality constraint (Figure 3A). Indeed, correlation coefficients for S_2, S_3 and S_4 between approaches using spectral equality constraint on the ACI or not were close to 1, with respective values of 0.9983, 0.9979 and 0.9999 (Table 1). The spectral components estimated by applying a spectral equality constraint or not in the MCR-ALS analysis were found to be similar, for this application. This is understandable, as the component S_1 from the MCR-ALS applied without spectral equality constraint on ACI is really close to the reference spectrum of Delipidol, and thus the mathematical decomposition for other components leads to similar outcomes. This is confirmed by the trends in penetration profiles which are comparable for both approaches (Figure 5). Both average penetration profiles present the same higher accumulation of Delipidol in the SC and progressive decrease in deeper layers (Table 2). Note also that the kinetics of penetration described in this paper are close to the ones previously obtained with an NCLS approach [26] (Figure 5). However, comparison of relative concentrations highlights a slightly lower estimated abundance for the analysis using an additional spectral equality constraint on the ACI. For example, in the SC (0–20 μm), when only non-negativity is applied to spectra and relative concentrations, estimated abundances were 0.0691 (1 hour), 0.5364 (2 hours) and 0.8098 (3 hours), while when a spectral equality constraint on the ACI is added to the analysis, values of 0.0586, 0.4995 and 0.7638 were found. Despite this slight difference of relative abundances, the similarity of kinetics of penetration for both approaches support the possibility of using an MCR-ALS without spectral equality constraint on the ACI for our application.

In the field of cosmetic research and development, Raman spectroscopy can be a powerful tool to confirm a molecule is able to diffuse trough the main barrier function of the *SC* and reach deeper cellular targets, without the requirement to access absolute concentrations. Images are a highly relevant means of representation of results with high impact in the industrial sector, especially in marketing, enabling comparative studies between different conditions, which is essential to support for example optimization of formulations for skin care products. Presently, for the case of Delipidol, both approaches applied with and without spectral equality constraint on ACI enable to reach converging results as reconstructed images convey the same information (Figure 2A-D) and it may be considered that the method addresses current needs for improved capabilities of Raman spectroscopy for skin analysis.

Previously NCLS was explored to probe the distribution of the ACI, rather than the specific skin components [26]. The approach remain to model as closely as possible the underlying skin contribution in spectra to more specifically extract Delipidol relative concentrations. MCR-ALS, both approaches applying spectral equality constraint or not on the ACI, appears easier to employ, due to the low requirement for reference spectra to perform the analysis. In comparison, methods such as NCLS are strongly limited by their dependence on the degree of resemblance between the pure spectra and the constituents found in the skin, making the analysis less adaptive and thus prone to fitting errors.

MCR-ALS appears as potentially powerful alternative for unmixing of spectral component and investigation of

DOI: 10.1002/jbio.202000136

the distribution of an ACI within skin cross sections. The approach using a spectral equality constraint on the ACI is quite relevant when a priori knowledge of the compound of interest is available. Adding a constraint of fixing one of the components with its pure spectrum enables to profile its distribution within Raman maps and reduces the rotational ambiguity of the results [39]. For the case of Delipidol, it is demonstrated that, despite a slight loss in the fitting quality, the advantage lies in the absence of ambiguity in the results.

However, the use of a spectral equality constraint in MCR-ALS could also suffer from limitations in the case where the ACI spectrum is changed because of, for example, the influence of the local environment as it progresses through the layers of the skin, crystallization, accumulation in reservoirs, or metabolization, or indeed when the target of the active is the molecular constituents of the skin itself. In such cases, the outcome of analysis constrained or not in spectral equality can be expected to diverge, and optimised protocols for the more flexible MCR-ALS without supplementary constraint could be more suited.

5 | CONCLUSION

Multivariate analysis provides promising unmixing methods to understand and track the kinetics of diffusion of ACI in skin samples from Raman data. In this work, results confirm that Raman microspectroscopy coupled with MCR-ALS is a powerful tool to detect and monitor an ACI penetration in the upper layers of the skin. It is observed that for the case of Delipidol an a priori knowledge on the spectrum of ACI was not necessary for the analysis when using the MCR-ALS approach. The method applied without spectral equality constraint on the ACI was able to estimate a characteristic spectrum very close to that of ACI.

Ultimately, MCR-ALS decomposition provides similar penetration profile estimations to the previously tested NCLS, with the great advantage that it is a self-modelling approach that does not require complex and sometimes unreliable large reference database of spectra from pure components or an extended number of control data set collected from control maps. In other words, the MCR-ALS coupled to Raman mapping appears to be an efficient tool to analyse and evaluate the effectiveness of ACI penetration and could support further transfer of the technique into the cosmetics industry.

ACKNOWLEDGMENTS

This work is part of the MISTIC project (2017 – 00117681). We thank Conseil Régional Centre Val de Loire, ARD-2020 Cosmétosciences program and FEDER for financial support.

ORCID

Aline Stella https://orcid.org/0000-0001-5062-5410

REFERENCES

[1] European Medicines Agency, *Comm. Med. Prod. Hum. Use* **2018**, *44*(October), 1.
[2] L. Franzen, M. Windbergs, *Adv. Drug Deliv. Rev.* **2015**, *89*, 91.
[3] M. Essendoubi, C. Gobinet, R. Reynaud, J. F. Angiboust, M. Manfait, O. Piot, *Skin Res. Technol.* **2016**, *22*, 55.
[4] A. Tfayli, O. Piot, F. Pitre, M. Manfait, *Eur. Biophys. J.* **2007**, *36*(8), 1049.
[5] L. Binder, E. M. Kulovits, R. Petz, J. Ruthofer, D. Baurecht, V. Klang, C. Valenta, *Eur. J. Pharm. Biopharm.* **2018**, *130*(July), 214.
[6] D. Lunter, R. Daniels, *J. Biomed. Opt.* **2014**, *19*(12), 126015.
[7] S. Mujica Ascencio, C. S. Choe, M. C. Meinke, R. H. Müller, G. V. Maksimov, W. Wigger-Alberti, J. Lademann, M. E. Darvin, *Eur. J. Pharm. Biopharm.* **2016**, *104*, 51.
[8] G. J. Puppels, C. Nico, T. C. Bakker-Schut, J. de Sterke, P. J. Caspers, *Proc. SPIE* **2020**, *11236*, 1123609.
[9] P. J. Caspers et al., *Transl. Biophoton.* **2019**, *1*(1–2), 1.
[10] H. J. Byrne, P. Knief, M. E. Keating, F. Bonnier, *Chem. Soc. Rev.* **2016**, *45*(7), 1865.
[11] S. M. Ali, F. Bonnier, K. Ptasinski, H. Lambkin, K. Flynn, F. M. Lyng, H. J. Byrne, *Analyst* **2013**, *138*(14), 3946.
[12] F. J. González, C. Castillo-Martínez, M. Martínez-Escanamé, M. G. Ramírez-Elías, F. I. Gaitan-Gaona, C. Oros-Ovalle, B. Moncada, *Skin Res. Technol.* **2012**, *18*(4), 442.
[13] V. K. Tippavajhala, T. de Oliveira Mendes, A. A. Martin, *AAPS PharmSciTech* **2017**, *19*(2), 753.
[14] L. dos Santos et al., *Vib. Spectrosc.* **2019**, *100*, 123.
[15] F. D. Fleischli, S. Mathes, C. Adlhart, *Vib. Spectrosc.* **2013**, *68*, 29.
[16] P.-A. Bonnet, C. Ghyselinck, C. Brenier, M. Martin, H. Rebiere, *J. Pharm. Biomed. Anal.* **2017**, *148*, 316.
[17] M. Boiret, D. N. Rutledge, N. Gorretta, Y. M. Ginot, J. M. Roger, *J. Pharm. Biomed. Anal.* **2014**, *90*, 78.
[18] L. Coic et al., *Talanta* **2019**, *198*, 457.
[19] B. Vajna, G. Patyi, Z. Nagy, A. Bódis, A. Farkas, G. Marosi, *J. Raman Spectrosc.* **2011**, *42*(11), 1977.
[20] S. J. Mazivila, A. C. Olivieri, *Trends Anal. Chem.* **2018**, *108*, 74.
[21] C. Krafft, G. Steiner, C. Beleites, R. Salzer, *J. Biophotonics* **2009**, *2*(1–2), 13.
[22] K. W. Short, S. Carpenter, J. P. Freyer, J. R. Mourant, *Biophys. J.* **2005**, *88*(6), 4274.
[23] I. R. Ramos, A. D. Meade, O. Ibrahim, H. J. Byrne, M. McMenamin, M. McKenna, A. Malkin, F. M. Lyng, *Faraday Discuss.* **2016**, *187*, 187.
[24] A. Kassouf, D. Jouan-Rimbaud Bouveresse, D. N. Rutledge, *Talanta* **2018**, *179*, 538.
[25] S. Gourvénec, D. L. Massart, D. N. Rutledge, *Chemom. Intel. Lab. Syst.* **2002**, *61*(1–2), 51.
[26] L. Miloudi, F. Bonnier, A. Tfayli, F. Yvergnaux, H. J. Byrne, I. Chourpa, E. Munnier, *J. Biophotonics* **2018**, *11*(4), e201700221.

DOI: 10.1002/jbio.202000136

[27] J. Felten, H. Hall, J. Jaumot, R. Tauler, A. De Juan, A. Gorzsás, *Nat. Protoc.* **2015**, *10*, 217.
[28] D. N. Rutledge, E. de Paula, R. J. Poppi, H. Mitsutake, S. R. Castro, M. C. Breitkreitz, *Int. J. Pharm.* **2018**, *552*(1–2), 119.
[29] B. Prats-Mateu, M. Felhofer, A. de Juan, N. Gierlinger, *Plant Methods* **2018**, *14*(1), 1.
[30] A. De Juan, J. Jaumot, R. Tauler, *Anal. Methods* **2014**, *6*(14), 4964.
[31] P. Koziol et al., *Sci. Rep.* **2018**, *8*(1), 1.
[32] J. M. Shaver, "Chemometrics for Raman Spectroscopy," in *Handbook of Raman spectroscopy. From the research laboratory to the process line*., I. R. Lewis and H. G. M. Edwards, Eds. New York, NY: Marcel Dekker, Inc., **2001**, pp. 292–298.
[33] H. Martens, T. Naes, *Multivariate Calibration*, John Wiley and Sons, Chichester, UK **1989**.
[34] M. Pirzer and J. Sawatzki, Method and Device for Correcting a Spectrum, US patent US20060212275A1; **2006**.
[35] S. Wartewig, C. Schorn, P. Bigler, *IR and Raman Spectroscopy*, Darmstadt, Germany: Wiley-VCH Verlag GmbH & Co., KGaA, **2003**, pp. 77.
[36] D. R. Parachalil, B. Brankin, J. McIntyre, H. J. Byrne, *Analyst* **2018**, *143*(24), 5987.
[37] J. T. Bulmer, D. E. Irish, F. W. Grossman, G. Herriot, M. Tseng, A. J. Weerheim, *Appl. Spectrosc.* **1975**, *29*(6), 506.
[38] W. Windig, J. Guilment, W. Winding, J. Guilment, *Anal. Chem.* **1991**, *63*(14), 1425.
[39] J. Jaumot, R. Tauler, *Chemom. Intel. Lab. Syst.* **2010**, *103*(2), 96.
[40] J. Jaumot, A. de Juan, R. Tauler, *Chemom. Intel. Lab. Syst.* **2015**, *140*, 1.
[41] R. Taylor, *J. Diagnos. Med. Sonogr.* **1990**, *6*(1), 35.
[42] S. M. Ali, F. Bonnier, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T. C. Lee, F. M. Lyng, H. J. Byrne, *Anal. Methods* **2013**, *5*(9), 2281.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Stella A, Bonnier F, Tfayli A, et al. Raman mapping coupled to self-modelling MCR-ALS analysis to estimate active cosmetic ingredient penetration profile in skin. *J. Biophotonics*. 2020;e202000136. https://doi.org/10.1002/jbio.202000136

RÉFÉRENCES

[1] J. Fenner and R. A. F. Clark, "Anatomy, Physiology, Histology, and Immunohistochemistry of Human Skin," in *Skin Tissue Engineering and Regenerative Medicine*, Elsevier, 2016, pp. 1–17.

[2] A. Mélissopoulos and C. Levacher, *La peau Structure et physiologie*, 2e edition. Lavoisier, 2012.

[3] P. A. J. Kolarsick, M. A. Kolarsick, and C. Goodwin, "Anatomy and Physiology of the Skin," *J. Dermatol. Nurses. Assoc.*, vol. 3, no. 4, pp. 203–213, Jul. 2011.

[4] C. Twarog, "Particules Janus pour des applications industrielles : stabilités et incorporation de principes actifs," Sciences Pharmaceutiques, hal-01932273, 2017.

[5] E. Abd *et al.*, "Skin models for the testing of transdermal drugs," *Clin. Pharmacol. Adv. Appl.*, vol. 8, pp. 163–176, 2016.

[6] F. Netzlaff, C. M. Lehr, P. W. Wertz, and U. F. Schaefer, "The human epidermis models EpiSkin®, SkinEthic® and EpiDerm®: An evaluation of morphology and their suitability for testing phototoxicity, irritancy, corrosivity, and substance transport," *Eur. J. Pharm. Biopharm.*, vol. 60, no. 2, pp. 167–178, 2005.

[7] M. Van Gele, B. Geusens, L. Brochez, R. Speeckaert, and J. Lambert, "Three-dimensional skin models as tools for transdermal drug delivery: Challenges and limitations," *Expert Opin. Drug Deliv.*, vol. 8, no. 6, pp. 705–720, 2011.

[8] U. F. Schaefer, S. Hansen, M. Schneider, J. L. Contreras, and C.-M. Lehr, "Models for Skin Absorption and Skin Toxicity Testing," *Drug Absorpt. Stud.*, pp. 3–33, 2007.

[9] V. Planz, C. M. Lehr, and M. Windbergs, "In vitro models for evaluating safety and efficacy of novel technologies for skin drug delivery," *J. Control. Release*, vol. 242, pp. 89–104, 2016.

[10] "Les conditions de mise sur le marché d'un produit cosmétique." [Online]. Available: https://solidarites-sante.gouv.fr/soins-et-maladies/autres-produits-de-sante/article/les-conditions-de-mise-sur-le-marche-d-un-produit-cosmetique.

[11] J. B. Nielsen, I. Plasencia, J. A. Sørensen, and L. A. Bagatolli, "Storage conditions of skin affect tissue structure and subsequent in vitro percutaneous penetration," *Skin Pharmacol. Physiol.*, vol. 24, no. 2, pp. 93–102, 2011.

[12] A. M. Barbero and H. F. Frasch, "Effect of Frozen Human Epidermis Storage Duration and Cryoprotectant on Barrier Function Using Two Model Compounds," *Skin Pharmacol. Physiol.*, vol. 29, no. 1, pp. 31–40, 2016.

[13] European Parliament, "DIRECTIVE 2004/23/CE DU PARLEMENT EUROPÉEN ET DU CONSEIL du 31 mars 2004 relative à l'établissement de normes de qualité et de sécurité pour le don, l'obtention, le contrôle, la transformation, la conservation, le stockage et la distribution des tissus et," in *Official Journal of the European Union*, no. L102, 2004, pp. 48–58.

[14] Conseil de l'Europe, "Convention pour la protection des droits de l'homme et de la dignité de l'être humain à l'égard des applications de la biologie et de la medicine: Convention sur les droits de l'homme et de la biomédecine," Oviedo, Espagne, 1997.

[15] M. Ponec, “Skin constructs for replacement of skin tissues for in vitro testing,” *Adv. Drug Deliv. Rev.*, vol. 54, no. SUPPL., pp. 19–30, 2002.

[16] E. Boelsma, S. Gibbs, C. Faller, and M. Ponec, “Characterization and comparison of reconstructed skin models: Morphological and immunohistochemical evaluation,” *Acta Derm. Venereol.*, vol. 80, no. 2, pp. 82–88, 2000.

[17] M. Ponec, E. Boelsma, A. Weerheim, A. Mulder, J. Bouwstra, and M. Mommaas, “Lipid and ultrastructural characterization of reconstructed skin models,” *Int. J. Pharm.*, vol. 203, no. 1–2, pp. 211–225, 2000.

[18] F. P. Schmook, J. G. Meingassner, and A. Billich, “Comparison of human skin or epidermis models with human and animal skin in in-vitro percutaneous absorption,” *Int. J. Pharm.*, vol. 215, no. 1–2, pp. 51–56, 2001.

[19] P. Portes, M. J. Pygmalion, E. Popovic, M. Cottin, and M. Mariani, “Use of human reconstituted epidermis Episkin for assessment of weak phototoxic potential of chemical compounds,” pp. 96–102, 2002.

[20] S. H. Mathes, H. Ruffner, and U. Graf-Hausner, “The use of skin models in drug development,” *Adv. Drug Deliv. Rev.*, vol. 69–70, pp. 81–102, 2014.

[21] A. Tfayli, O. Piot, F. Pitre, and M. Manfait, “Follow-up of drug permeation through excised human skin with confocal Raman microspectroscopy,” *Eur. Biophys. J.*, vol. 36, no. 8, pp. 1049–1058, 2007.

[22] G. Zhang, C. R. Flach, and R. Mendelsohn, “Tracking the dephosphorylation of resveratrol triphosphate in skin by confocal Raman microscopy,” *J. Control. Release*, vol. 123, no. 2, pp. 141–147, 2007.

[23] L. Franzen and M. Windbergs, “Applications of Raman spectroscopy in skin research - From skin physiology and diagnosis up to risk assessment and dermal drug delivery,” *Adv. Drug Deliv. Rev.*, vol. 89, pp. 91–104, 2015.

[24] Y. M. Ginot *et al.*, “L’imagerie chimique: un apport essentiel pour apprécier la qualité pharmaceutique,” *STP Pharma Prat.*, vol. n°04, no. juillet/août, pp. p04-21, 2019.

[25] C. Krafft, G. Steiner, C. Beleites, and R. Salzer, “Disease recognition by infrared and Raman spectroscopy,” *Journal of Biophotonics*, vol. 2, no. 1–2. pp. 13–28, 2009.

[26] G. Simon, *Spectroscopies vibrationnelles*. France: Editions des archives contemporaines, 2020.

[27] Unversité de Lyon 1, “Interactions rayonnement matière.” [Online]. Available: eplanets.univ-lyon1.fr.

[28] L. Miloudi, “Application des techniques spectroscopiques vibrationnelles couplées aux analyses statistiques multivariées pour la caractérisation et l ’ objectivation des produits de soins cosmétiques,” Université de Tours, 2018.

[29] Horiba, “Raman Spectroscopy for Analysis and Monitoring.” [Online]. Available: https://static.horiba.com/fileadmin/Horiba/Technology/Measurement_Techniques/Molecular_Spectroscopy/Raman_Spectroscopy/Raman_Academy/Raman_Tutorial/Raman_bands.pdf.

[30] F. Bonnier *et al.*, “Analysis of human skin tissue by Raman microspectroscopy: Dealing with the background,” *Vib. Spectrosc.*, vol. 61, pp. 124–132, 2012.

[31] A. Tfayli, “Caractérisation structurale et moléculaire de la peau par microspectroscopies optiques vibrationnelles. Applications au diagnostic précoce des tumeurs cutanées et à l’étude de la diffusion de principes actifs à visée thérapeutique,” Université de Reims Champagne-Ardenne, 2007.

[32] Z. Movasaghi, S. Rehman, and I. U. Rehman, “Raman Spectroscopy of Biological Tissues,” *Appl. Spectrosc. Rev.*, vol. 42, no. 5, pp. 493–541, Sep. 2007.

[33] A. Tfayli, O. Piot, F. Draux, F. Pitre, and M. Manfait, “Molecular characterization of reconstructed skin model by Raman microspectroscopy: Comparison with excised human skin,” *Biopolymers*, vol. 87, no. 4, pp. 261–274, Nov. 2007.

[34] S. Tfaili, C. Gobinet, G. Josse, J. F. Angiboust, M. Manfait, and O. Piot, “Confocal Raman microspectroscopy for skin characterization: A comparative study between human skin and pig skin,” *Analyst*, vol. 137, no. 16, pp. 3673–3682, 2012.

[35] S. M. Ali *et al.*, “A comparison of Raman, FTIR and ATR-FTIR micro spectroscopy for imaging human skin tissue sections,” *Anal. Methods*, vol. 5, no. 9, pp. 2281–2291, 2013.

[36] S. Tfaili, G. Josse, J. F. Angiboust, M. Manfait, and O. Piot, “Monitoring caffeine and resveratrol cutaneous permeation by confocal Raman microspectroscopy,” *Journal of Biophotonics*, vol. 7, no. 9. pp. 676–681, 2014.

[37] M. Janssens, J. Van Smeden, G. J. Puppels, A. P. M. Lavrijsen, P. J. Caspers, and J. A. Bouwstra, “Lipid to protein ratio plays an important role in the skin barrier function in patients with atopic eczema,” *Br. J. Dermatol.*, vol. 170, no. 6, pp. 1248–1255, 2014.

[38] C. Alonso, V. Carrer, C. Barba, and L. Coderch, “Caffeine delivery in porcine skin: a confocal Raman study,” *Arch. Dermatol. Res.*, vol. 310, no. 8, pp. 657–664, 2018.

[39] P. J. Caspers, G. W. Lucassen, E. A. Carter, H. A. Bruining, and G. J. Puppels, “In vivo confocal raman microspectroscopy of the skin: Noninvasive determination of molecular concentration profiles,” *J. Invest. Dermatol.*, vol. 116, no. 3, pp. 434–442, 2001.

[40] P. J. Caspers *et al.*, “Monitoring the penetration enhancer dimethyl sulfoxide in human stratum corneum in vivo by confocal Raman spectroscopy,” *Pharm. Res.*, vol. 19, no. 10, pp. 1577–1580, 2002.

[41] R. Touti, M. Noun, F. Guimberteau, S. Lecomte, and C. Faure, “What is the fate of multi-lamellar liposomes of controlled size, charge and elasticity in artificial and animal skin?,” *Eur. J. Pharm. Biopharm.*, vol. 151, no. March, pp. 18–31, 2020.

[42] H. J. Byrne, P. Knief, M. E. Keating, and F. Bonnier, “Spectral pre and post processing for infrared and Raman spectroscopy of biological tissues and cells,” *Chem. Soc. Rev.*, vol. 45, no. 7, pp. 1865–1878, 2016.

[43] C. Krafft, T. Knetschke, R. H. W. Funk, and R. Salzer, “Studies on stress-induced changes at the subcellular level by Raman microspectroscopic mapping,” *Anal. Chem.*, vol. 78, no. 13, pp. 4424–4429, 2006.

[44] L. Miloudi *et al.*, “Confocal Raman spectroscopic imaging for in vitro monitoring of active ingredient penetration and distribution in reconstructed human epidermis model,” *J. Biophotonics*, vol. 11, no. 4, p. e201700221, Apr. 2018.

[45] S. M. Ali *et al.*, “Raman spectroscopic mapping for the analysis of solar radiation induced skin damage,” *Analyst*, vol. 138, no. 14, pp. 3946–3956, 2013.

[46] P. Y. Sacré *et al.*, “Detection of counterfeit Viagra ® by Raman microspectroscopy imaging

and multivariate analysis," *J. Pharm. Biomed. Anal.*, vol. 56, no. 2, pp. 454–461, 2011.

[47] F. Bonnier and H. J. Byrne, "Understanding the molecular information contained in principal component analysis of vibrational spectra of biological systems," *Analyst*, vol. 137, no. 2, pp. 322–332, 2012.

[48] J. Dorney, F. Bonnier, A. Garcia, A. Casey, G. Chambers, and H. J. Byrne, "Identifying and localizing intracellular nanoparticles using Raman spectroscopy," *Analyst*, vol. 137, no. 5, pp. 1111–1119, 2012.

[49] D. Lunter and R. Daniels, "Confocal Raman microscopic investigation of the effectiveness of penetration enhancers for procaine delivery to the skin," *J. Biomed. Opt.*, vol. 19, no. 12, p. 126015, 2014.

[50] M. de Veij *et al.*, "Fast detection and identification of counterfeit antimalarial tablets by Raman spectroscopy," *J. Raman Spectrosc.*, vol. 38, no. 2, pp. 181–187, Feb. 2007.

[51] M. de Veij, A. Deneckere, P. Vandenabeele, D. de Kaste, and L. Moens, "Detection of counterfeit Viagra® with Raman spectroscopy," *J. Pharm. Biomed. Anal.*, vol. 46, no. 2, pp. 303–309, 2008.

[52] O. Piot *et al.*, *A portable Raman probe for in vivo pathological tissues characterisation*, vol. 6, no. PART 1. IFAC, 2006.

[53] J. Felten, H. Hall, J. Jaumot, R. Tauler, A. De Juan, and A. Gorzsás, "Vibrational spectroscopic image analysis of biological material using multivariate curve resolution-alternating least squares (MCR-ALS)," *Nat. Protoc.*, 2015.

[54] V. K. Tippavajhala, T. de Oliveira Mendes, and A. A. Martin, "In Vivo Human Skin Penetration Study of Sunscreens by Confocal Raman Spectroscopy," *AAPS PharmSciTech*, vol. 19, no. 2, pp. 753–760, 2017.

[55] P.-A. Bonnet, C. Ghyselinck, C. Brenier, M. Martin, and H. Rebiere, "Raman chemical imaging for spectroscopic screening and direct quantification of falsified drugs," *J. Pharm. Biomed. Anal.*, vol. 148, pp. 316–323, 2017.

[56] Z. Farhane, F. Bonnier, A. Casey, and H. J. Byrne, "Raman micro spectroscopy for in vitro drug screening: Subcellular localisation and interactions of doxorubicin," *Analyst*, vol. 140, no. 12, pp. 4212–4223, 2015.

[57] Z. Farhane, F. Bonnier, and H. J. Byrne, "Monitoring doxorubicin cellular uptake and trafficking using in vitro Raman microspectroscopy: short and long time exposure effects on lung cancer cell lines," *Anal. Bioanal. Chem.*, vol. 409, no. 5, pp. 1333–1346, 2017.

[58] F. Draux, C. Gobinet, J. Sulé-Suso, M. Manfait, P. Jeannesson, and G. D. Sockalingum, "Raman imaging of single living cells: Probing effects of non-cytotoxic doses of an anti-cancer drug," *Analyst*, vol. 136, no. 13, pp. 2718–2725, 2011.

[59] F. J. González *et al.*, "Noninvasive estimation of chronological and photoinduced skin damage using Raman spectroscopy and principal component analysis," *Ski. Res. Technol.*, vol. 18, no. 4, pp. 442–446, 2012.

[60] S. Guo *et al.*, "Confocal Raman imaging of skin sections containing hair follicles using classical least squares regression and multivariate curve resolution–alternating least squares," *Quantum Electron.*, vol. 49, no. 1, pp. 6–12, 2019.

[61] M. Essendoubi, C. Gobinet, R. Reynaud, J. F. Angiboust, M. Manfait, and O. Piot, "Human skin penetration of hyaluronic acid of different molecular weights as probed by Raman

spectroscopy," pp. 55–62, 2016.

[62] B. Vajna, G. Patyi, Z. Nagy, A. Bódis, A. Farkas, and G. Marosi, "Comparison of chemometric methods in the analysis of pharmaceuticals with hyperspectral Raman imaging," *J. Raman Spectrosc.*, vol. 42, no. 11, pp. 1977–1986, 2011.

[63] F. D. Fleischli, S. Mathes, and C. Adlhart, "Vibrational Spectroscopy Label free non-invasive imaging of topically applied actives in reconstructed human epidermis by confocal Raman spectroscopy," *Vib. Spectrosc.*, vol. 68, pp. 29–33, 2013.

[64] M. E. Darvin *et al.*, "Confocal Raman microscopy combined with optical clearing for identification of inks in multicolored tattooed skin: In vivo," *Analyst*, vol. 143, no. 20, pp. 4990–4999, 2018.

[65] L. dos Santos *et al.*, "Evaluation of penetration process into young and elderly skin using confocal Raman spectroscopy," *Vib. Spectrosc.*, vol. 100, no. November 2018, pp. 123–130, 2019.

[66] Food and Drug Administration, "Development and Submission of Near Infrared Analytical Procedures Guidance for Industry," 2021.

[67] European Medicine Agency, "Guideline on the use of Near Infrared Spectroscopy (NIRS) by the pharmaceutical industry and the data requirements for new submissions and variations," 2014.

[68] P. Lasch, "Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging," *Chemom. Intell. Lab. Syst.*, vol. 117, pp. 100–114, 2012.

[69] T. Bocklitz, A. Walter, K. Hartmann, P. Rösch, and J. Popp, "How to pre-process Raman spectra for reliable and stable models?," *Anal. Chim. Acta*, vol. 704, no. 1–2, pp. 47–56, 2011.

[70] A. Savitzky and M. J. Golay, "Smoothing and Differentiation," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.

[71] M. Pirzer and J. Sawatzki, "Method and Device for correcting a spectrum," 2006.

[72] D. R. Parachalil, B. Brankin, J. McIntyre, and H. J. Byrne, "Raman spectroscopic analysis of high molecular weight proteins in solution – considerations for sample analysis and data pre-processing," *Analyst*, vol. 143, no. 24, pp. 5987–5998, 2018.

[73] A. Mahadevan-Jansen and C. A. Lieber, "Automated Method for Subtraction of Fluorescence from Biological Raman Spectra," *Appl. Spectrosc.*, vol. 57, no. 11, pp. 1363–1367, 2003.

[74] P. Knief, *Interactions of Carbon Nanotubes With human lung epithelial cells in vitro. Assessment by Raman Spectroscopy*, PhD. Technological University Dublin, 2010.

[75] V. Mazet and D. Brie, "USING HALF-QUADRATIC MINIMIZATION," pp. 2–5, 2004.

[76] V. Mazet, C. Carteret, D. Brie, J. Idier, and B. Humbert, "Background removal from spectra by designing and minimising a non-quadratic cost function," *Chemom. Intell. Lab. Syst.*, vol. 76, no. 2, pp. 121–133, Apr. 2005.

[77] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1001–1009, 2001.

[78] C. J. Behrend, C. P. Tarnowski, and M. D. Morris, "Identification of outliers in hyperspectral

Raman image data by nearest neighbor comparison," *Appl. Spectrosc.*, vol. 56, no. 11, pp. 1458–1461, 2002.

[79] O. Ryabchykov *et al.*, "Automatization of spike correction in Raman spectra of biological samples," *Chemom. Intell. Lab. Syst.*, vol. 155, pp. 1–6, 2016.

[80] L. M. L. Miao and H. Q. H. Qi, "Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 1–12, 2007.

[81] J. Plaza, E. M. T. Hendrix, I. García, G. Martín, and A. Plaza, "On endmember identification in hyperspectral images without pure pixels: A comparison of algorithms," *J. Math. Imaging Vis.*, vol. 42, no. 2–3, pp. 163–175, 2012.

[82] O. Gutierrez-Navarro, D. U. Campos-Delgado, E. R. Arce-Santana, M. O. Mendez, and J. A. Jo, "Blind end-member and abundance extraction for multispectral fluorescence lifetime imaging microscopy data," *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 2, pp. 606–617, 2014.

[83] Y. Gut *et al.*, "Application of chemometric algorithms to MALDI mass spectrometry imaging of pharmaceutical tablets," *J. Pharm. Biomed. Anal.*, vol. 105, pp. 91–100, 2015.

[84] S. J. Mazivila and A. C. Olivieri, "Chemometrics coupled to vibrational spectroscopy and spectroscopic imaging for the analysis of solid-phase pharmaceutical products: A brief review on non-destructive analytical methods," *TrAC - Trends Anal. Chem.*, vol. 108, pp. 74–87, 2018.

[85] L. Coic *et al.*, "Comparison of hyperspectral imaging techniques for the elucidation of falsified medicines composition," *Talanta*, vol. 198, no. February 2019, pp. 457–463, 2019.

[86] M. Boiret, A. de Juan, N. Gorretta, Y. M. Ginot, and J. M. Roger, "Distribution of a low dose compound within pharmaceutical tablet by using multivariate curve resolution on Raman hyperspectral images," *J. Pharm. Biomed. Anal.*, vol. 103, pp. 35–43, 2015.

[87] L. Valderrama, R. P. Gonçalves, P. H. Março, D. N. Rutledge, and P. Valderrama, "Independent components analysis as a means to have initial estimates for multivariate curve resolution-alternating least squares," *J. Adv. Res.*, vol. 7, no. 5, pp. 795–802, 2016.

[88] D. N. Rutledge, E. de Paula, R. J. Poppi, H. Mitsutake, S. R. Castro, and M. C. Breitkreitz, "Comparison of different chemometric methods to extract chemical and physical information from Raman images of homogeneous and heterogeneous semi-solid pharmaceutical formulations," *Int. J. Pharm.*, vol. 552, no. 1–2, pp. 119–129, 2018.

[89] A. Gaubert *et al.*, "Characterization of surfactant complex mixtures using Raman spectroscopy and signal extraction methods: Application to laundry detergent deformulation," *Anal. Chim. Acta*, vol. 915, pp. 36–48, Apr. 2016.

[90] B. Prats-Mateu, M. Felhofer, A. de Juan, and N. Gierlinger, "Multivariate unmixing approaches on Raman images of plant cell walls: New insights or overinterpretation of results?," *Plant Methods*, vol. 14, no. 1, pp. 1–20, 2018.

[91] V. Olmos *et al.*, "Relevant aspects of unmixing/resolution analysis for the interpretation of biological vibrational hyperspectral images," *TrAC - Trends Anal. Chem.*, vol. 94, pp. 130–140, 2017.

[92] V. Olmos *et al.*, "Combining hyperspectral imaging and chemometrics to assess and interpret the effects of environmental stressors on zebrafish eye images at tissue level," *J.*

Biophotonics, vol. 11, no. 3, pp. 1–12, 2018.

[93] S. Fornasaro, A. Vicario, L. De Leo, A. Bonifacio, T. Not, and V. Sergo, "Potential use of MCR-ALS for the identification of coeliac-related biochemical changes in hyperspectral Raman maps from pediatric intestinal biopsies," *Integr. Biol. (United Kingdom)*, vol. 10, no. 6, pp. 356–363, 2018.

[94] A. De Juan, J. Jaumot, and R. Tauler, "Multivariate Curve Resolution (MCR). Solving the mixture analysis problem," *Anal. Methods*, vol. 6, no. 14, pp. 4964–4976, 2014.

[95] J. Jaumot, A. de Juan, and R. Tauler, "MCR-ALS GUI 2.0: New features and applications," *Chemom. Intell. Lab. Syst.*, vol. 140, pp. 1–12, 2015.

[96] S. Gourvénec, D. L. Massart, and D. N. Rutledge, "Determination of the number of components during mixture analysis using the Durbin-Watson criterion in the Orthogonal Projection Approach and in the SIMPLe-to-use Interactive Self-modelling Mixture Analysis approach," *Chemom. Intell. Lab. Syst.*, vol. 61, no. 1–2, pp. 51–61, 2002.

[97] A. Kassouf, D. Jouan-Rimbaud Bouveresse, and D. N. Rutledge, "Determination of the optimal number of components in independent components analysis," *Talanta*, vol. 179, no. November 2017, pp. 538–545, 2018.

[98] H. F. Kaiser, "An index of factorial simplicity," *Psychometrika*, vol. 39, no. 1, pp. 31–36, 1974.

[99] W. Windig, J. Guilment, W. Winding, and J. Guilment, "Interactive self-modeling mixture analysis," *Anal. Chem.*, vol. 63, no. 14, pp. 1425–1432, 1991.

[100] R. Taylor, "Interpretation of the Correlation Coefficient: A Basic Review," *J. Diagnostic Med. Sonogr.*, vol. 6, no. 1, pp. 35–39, 1990.

[101] J. Jaumot and R. Tauler, "MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution," *Chemom. Intell. Lab. Syst.*, vol. 103, no. 2, pp. 96–107, 2010.

[102] R. Tauler, "Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution," *J. Chemom.*, vol. 15, no. 8, pp. 627–646, 2001.

[103] J. R. Munkres, "Analysis on Manifolds," Allan M. Wylde, 1991, p. 366.

[104] M. Arngren, M. N. Schmidt, and J. Larsen, "{MATLAB} Toolbox: Nonnegative Matrix Factorization with Volume Constraints." Department of Informatics and Mathematical Modelling, 2010.

[105] A. Stella *et al.*, "Raman mapping coupled to self-modelling MCR-ALS analysis to estimate active cosmetic ingredient penetration profile in skin," *J. Biophotonics*, vol. 13, no. 11, pp. 1–13, 2020.

[106] S. Mujica Ascencio *et al.*, "Confocal Raman microscopy and multivariate statistical analysis for determination of different penetration abilities of caffeine and propylene glycol applied simultaneously in a mixture on porcine skin ex vivo," *Eur. J. Pharm. Biopharm.*, vol. 104, pp. 51–58, 2016.

[107] C. R. Afonso *et al.*, "Biodegradable antioxidant chitosan films useful as an anti-aging skin mask," *Int. J. Biol. Macromol.*, vol. 132, pp. 1262–1273, 2019.

[108] J. S. Kwon, D. Y. Kim, H. W. Seo, S. H. Jeong, J. H. Kim, and M. S. Kim, "Preparation of

erythromycin-loaded poly(vinylalcohol) film and investigation of its feasibility as a transdermal delivery carrier," *Tissue Eng. Regen. Med.*, vol. 11, no. 3, pp. 211–216, 2014.

[109] K. Van Bocxlaer *et al.*, "Film-forming systems for the delivery of DNDI-0690 to treat cutaneous leishmaniasis," *Pharmaceutics*, vol. 13, no. 4, pp. 1–14, 2021.

[110] M. Contardi *et al.*, "Transparent ciprofloxacin-povidone antibiotic films and nanofiber mats as potential skin and wound care dressings," *Eur. J. Pharm. Sci.*, vol. 104, no. January, pp. 133–144, 2017.

[111] S. Patel, S. Srivastava, M. R. Singh, and D. Singh, "Preparation and optimization of chitosan-gelatin films for sustained delivery of lupeol for wound healing," *Int. J. Biol. Macromol.*, vol. 107, pp. 1888–1897, 2018.

[112] L. Lefeuvre and D. Josse, "Composition Dermatologique et/ou cosmétique protectrice de la peau et des muqueuses," FR 2 929 847, 2009.

[113] T. T. D. Tran and P. H. L. Tran, "Controlled release film forming systems in drug delivery: The potential for efficient drug delivery," *Pharmaceutics*, vol. 11, no. 6, pp. 1–16, 2019.

[114] L. Van Gheluwe, E. Munnier, H. Kichou, and K. Kemel, "Confocal Raman Spectroscopic Imaging for Evaluation of Distribution of Nano-Formulated Hydrophobic Active Cosmetic Ingredients in Hydrophilic Films," 2021.

[115] C. Ann Bateson, G. P. Asner, and C. A. Wessman, "Endmember bundles: a new approach to incorporating endmember variability into spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2 II, pp. 1083–1094, 2000.

[116] L. Drumetz, M. A. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, "Blind Hyperspectral Unmixing Using an Extended Linear Mixing Model to Address Spectral Variability," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3890–3905, 2016.

[117] P. A. Thouvenin, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, 2016.

[118] Y. C. Cavalcanti *et al.*, "Démélange d ' images TEP dynamiques HAL Id : hal-01912800," in *26eme Colloque GRETSI sur le Traitement du Signal et des Images (GRETSI 2017),* 2017, p. pp.1-4.

[119] Y. C. Cavalcanti *et al.*, "Factor Analysis of Dynamic PET Images: Beyond Gaussian Noise," *IEEE Trans. Med. Imaging*, vol. 38, no. 9, pp. 2231–2241, 2019.

[120] Y. C. Cavalcanti, T. Oberlin, N. Dobigeo, C. Fevotte, S. Stute, and C. Taube, "Unmixing Dynamic Pet Images: Combining Spatial Heterogeneity and Non-gaussian Noise," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2019-May, no. 1, pp. 1373–1377, 2019.

[121] Y. C. Cavalcanti, T. Oberlin, V. Ferraris, N. Dobigeon, M. Ribeiro, and C. Tauber, "Compartment model-based nonlinear unmixing for kinetic analysis of dynamic PET images," *arXiv*, 2020.

[122] Y. Zhou and S. Maskell, "Robust and Accurate Global Motion Estimation Using the Student-t Distribution," pp. 1–5, 2019.

Aline STELLA

Analyse multivariée des images hyperspectrales Raman : application à l'étude des dynamiques moléculaires de la barrière cutanée

Résumé

L'imagerie confocale Raman permet une étude non destructive des cellules et tissus. Les images collectées associent à chaque pixel un spectre de valeurs qui constitue une empreinte moléculaire. Elles permettent de mettre en évidence et de suivre des modifications dans la composition chimique de la peau. Cependant, les images hyperspectrales sont des ensembles de données complexes dont la pleine exploitation nécessite l'application de méthodes statistiques robustes afin d'extraire une information pertinente. Dans cette thèse, des méthodes d'analyse multivariée sont présentées et évaluées pour exploiter les images vectorielles Raman. Le travail s'inscrit dans une démarche d'applicabilité avec i/ une étude comparative des prétraitements des données et ii/ l'application des méthodes présentées à des données réelles sur des problématiques importantes pour l'industrie du cosmétique. Les approches proposées sont en particulier appliquées pour établir les cinétiques de pénétration d'actifs cosmétiques dans la peau et pour étudier l'impact des conditions de stockage sur les modèles de peau. En parallèle, la création de données simulées réalistes vise à estimer la pertinence d'application de ces méthodes.

Mots clés : Imagerie hyperspectrale, Fouille de données, Traitement d'image, Analyses statistiques multivariées, Microspectroscopie confocale Raman.

Résumé en anglais

Confocal Raman imaging allows non-destructive study of cells and tissues. The images collected associate with each pixel a spectrum of values that constitutes a molecular fingerprint. They allow to highlight and follow changes in the chemical composition of the skin. However, hyperpectral images are complex data sets whose full exploitation requires the application of robust statistical methods to extract relevant information. In this thesis, multivariate analysis methods are presented and evaluated to exploit Raman vector images. With the concern of the applicability of the proposed approaches we present i/ a comparative study of data preprocessing and ii/ the application of presented methods to real data on important issues for the cosmetic industry. In particular, the proposed approaches are used to establish the kinetics of penetration of active ingredients through the skin and to study the impact of storage conditions on skin models. A last part of the thesis is dedicated to the simulation of realistic Raman data to estimate the relevancy of the multivariate methods.

Keywords : Hyperspectral imaging, Data mining, Image processing, Multivariate statistical analysis, , confocal Raman microspectroscopy.