



**HAL**  
open science

**Extraction des connaissances en imagerie  
microspectrométrique par analyse chimométrique :  
application à la caractérisation des constituants d'un  
calcul urinaire**

Waïss Elmi Rayaleh

► **To cite this version:**

Waïss Elmi Rayaleh. Extraction des connaissances en imagerie microspectrométrique par analyse chimométrique : application à la caractérisation des constituants d'un calcul urinaire. Autre. Université des Sciences et Technologie de Lille - Lille I, 2006. Français. NNT : . tel-00270116

**HAL Id: tel-00270116**

**<https://theses.hal.science/tel-00270116>**

Submitted on 3 Apr 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



N° d'ordre : 3813

**Université des sciences et technologies de Lille**  
**École Doctorale : Sciences pour l'ingénieur (SPI)**

**Thèse**

**pour obtenir le grade de**

**Docteur de l'Université des Sciences et Technologies de Lille**

**Discipline : Instrumentation et analyses avancées**

Présentée par

**Waïss ELMI RAYALEH**

**Extraction des connaissances en imagerie microspectrométrique  
par analyse chimiométrique. Application à la caractérisation des  
constituants d'un calcul urinaire.**

**Directeur de la thèse : M. Jean-Pierre HUVENNE**

**Co-encadrant : M. Ludovic DUPONCHEL.**

Soutenue publiquement le 27 juin 2006 devant le jury.

**Membres du jury :**

**Mme Mireille BAYART**, Professeur, Polytech'Lille, Université Lille 1, Villeneuve D'Ascq

**M. Ahmed ADDOU**, Professeur, Université de Mostaganem, Mostaganem, Algérie

**M. Philippe GIAMARCHI**, Professeur, Université de Bretagne Occidentale, Brest

**M. Ludovic DUPONCHEL**, Maître de conférences, Université Lille 1, Villeneuve D'Ascq

**M. Bernard SOMBRET**, Ingénieur de Recherche, LASIR -USTL, Villeneuve D'Ascq

**M. Jean-Pierre HUVENNE**, Professeur, Université Lille 1, Villeneuve D'Ascq

*A mon père*

*A ma mère*

*A mes frères et sœurs*

*A Fatouma et Mouhsin.*

## Remerciements.

---

*Ce travail, financé par une allocation MENRT, a été réalisé au Laboratoire de Spectrochimie Infrarouge et Raman (Lasir – CNRS UMR 8516). J'en suis entièrement reconnaissant.*

*Je tiens en premier lieu à adresser mes vifs remerciements et ma gratitude au Professeur Jean-Pierre Huvenne qui m'a accordé toute sa confiance pour ce travail. Je le remercie chaleureusement pour les très nombreux conseils prodigués tout au long de l'accomplissement de ce travail. Par delà, j'adresse mes reconnaissances aux responsables de l'école doctorale Science pour l'Ingénieur (SPI), et dans une plus large mesure aux autorités de Lille1.*

*Je tiens à remercier chaleureusement, mon encadrant, Ludovic Duponchel, pour ses constants et précieux conseils tout au long de la durée de ce travail. Qu'il en soit remercié pour m'avoir poussé à produire le meilleur de moi-même dans une pleine autonomie.*

*Mes vifs remerciements vont aussi au professeur Philippe GIAMARCHI de l'université de Bretagne occidentale (UBO) et au professeur Ahmed ADDOU de l'université de Mostaganem d'avoir accepté de juger ce travail et d'en être les rapporteurs. Je remercie chaleureusement madame Mireille BAYART d'accepter d'être membre du jury pour juger ce travail.*

*Je remercie Jacky Laureyns pour ses conseils et pour m'avoir initié à la microspectrométrie Raman, et Bernard Sombret pour ses très nombreuses aides techniques ; je le remercie aussi de m'honorer en acceptant de faire parti du jury. Je remercie aussi les autres membres du laboratoire qui m'ont apporté leur concours de quelle que manière que ce soit, et plus particulièrement Cyril Ruckebusch pour ses conseils très instructifs.*

*J'adresse mes sincères remerciements à Anna De Juan pour ses précieux éclairages et au professeur A. Addou de l'université de Mostaganem pour toutes les infos utiles qu'il nous a fournies.*

*Je ne saurais oublier les autorités Djiboutiennes du ministère de l'éducation nationale qui m'ont accordé leur confiance et le temps nécessaire pour accomplir sereinement ce travail.*

*Mes amis, à Djibouti, en France et ailleurs, qui m'ont encouragé, motivé et soutenu de quelle que manière que ce soit, méritent mes chaleureux remerciements.*

*Je n'oublie pas les compagnons de route pendant ces longues années, Bouziane et Maazouz, dont les échanges étaient fructueux. Je n'oublie pas non plus Ahmed, Yves, Lionel, Béatrice, Alexandra, Séverine et Olivier.*

# Sommaire

---

<i>INTRODUCTION GENERALE</i> .....	11
------------------------------------	----

<b>Chapitre 1 : Chimométrie pour la résolution des images</b> .....	<b>13</b>
---	-----------

<b>I. Images spectroscopiques et leurs problèmes</b> .....	<b>13</b>
--	-----------

<b>1. Images spectroscopiques : généralités</b> .....	<b>13</b>
---	-----------

<b>2. Problématique de reconstitution d'images</b> .....	<b>15</b>
--	-----------

<b>3. Les limitations de la production d'images spectroscopiques</b> .....	<b>17</b>
--	-----------

3.1. Limites théoriques .....	17
-------------------------------	----

3.2. Limitation instrumentale.....	17
------------------------------------	----

<b>4. Quelques outils pour améliorer la résolution ?</b> .....	<b>18</b>
--	-----------

<b>II. Méthodologies d'estimation du nombre des constituants</b> .....	<b>19</b>
--	-----------

<b>1. Définition et Bases mathématiques de l'analyse factorielle (Factor Analysis : FA)</b> .....	<b>19</b>
---	-----------

1.1. Définition.....	19
----------------------	----

1.2. Les fondements mathématiques.....	19
--	----

1.2.1. Décomposition en valeurs propres et vecteurs propres.....	21
--	----

1.2.2. Décomposition en valeurs singulières (Singular Value Decomposition : SVD).....	22
---	----

<b>2. Recherche des facteurs significatifs à l'aide des valeurs singulières</b> .....	<b>23</b>
---	-----------

<b>3. Les autres méthodes d'estimation de facteurs significatifs pour la décomposition</b> .....	<b>25</b>
--	-----------

3.1. Les méthodes empiriques .....	25
------------------------------------	----

3.2. Les méthodes statistiques .....	27
--------------------------------------	----

3.3. Les méthodes basées sur l'estimation de l'erreur expérimentale.....	28
--	----

3.4. Quelques méthodes du « self-modeling ».....	31
--	----

<b>III. Premières estimations : méthodes simples d'extraction de spectres et de concentrations</b> ..	<b>33</b>
---	-----------

<b>1. SVD et PCA</b> .....	<b>33</b>
----------------------------	-----------

<b>2. EFA et ses dérivées</b> .....	<b>35</b>
-------------------------------------	-----------

2.1. La naissance de EFA .....	35
--------------------------------	----

2.2. L'adaptation de EFA à une fenêtre fixe en déplacement : l'apparition de FSW-EFA (ou FSMW-EFA)...	37
---	----

2.3. L'utilisation de EFA pour l'imagerie 2-D, 3-D ou N-D : la création de FSIW-EFA .....	39
---	----

<b>3. SIMPLISMA (SIMPLE-to-use Interactive Self-Modeling Mixture Analysis) et OPA (Orthogonal Projection Approach)</b> .....	<b>40</b>
--	-----------

3.1. SIMPLISMA .....	40
----------------------	----

3.2. OPA .....	43
4. D'autres méthodes.....	46
<i>IV. L'optimisation et l'introduction des contraintes dans les méthodologies d'extraction pour une meilleure résolution .....</i>	<i>47</i>
1. Hypothèses sous-jacentes aux méthodologies d'extraction simples .....	47
2. Les nouvelles exigences et les contraintes associées .....	48
3. Quelques méthodologies d'extraction plus robustes .....	50
3.1. La procédure des moindres carrés Alternés (ALS : Alternating Least Squares) .....	50
3.2. Amélioration dans ALS et création de MCR-ALS (Multivariate Curve Resolution - Alternating Least Squares).....	52
3.3. D'autres méthodes comme PMF (Positive Matrix Factorization) .....	54
4. Contraintes sur la résolution des images .....	54
<b>Chapitre 2 : Imagerie microspectrométrique (infrarouge et Raman).....</b>	<b>56</b>
<i>Spectroscopie vibrationnelle .....</i>	<i>56</i>
<i>I. Spectroscopie infrarouge.....</i>	<i>56</i>
1. Un peu d'histoire .....	56
2. Domaine infrarouge et interaction .....	57
3. Vibration et modélisation .....	59
4. Les sources et les transducteurs en infrarouge.....	61
4.1. Les sources .....	61
4.2. Les transducteurs : état de l'art des détecteurs infrarouges .....	62
5. La Transformée de Fourier.....	65
5.1. Série de Fourier et Transformée de Fourier (TF).....	65
5.2. La transformée de Fourier et le traitement numérique du signal .....	66
6. Les spectromètres infrarouges .....	67
<i>II. Spectroscopie Raman .....</i>	<i>69</i>
1. Brève historique .....	69
2. Effet Raman.....	69
3. Instrumentation Raman .....	71
3.1. Les sources .....	72

3.2. Les spectromètres .....	73
<b><i>Imagerie microspectrométrique infrarouge et Raman .....</i></b>	<b>75</b>
<b>1. Aperçu général .....</b>	<b>75</b>
<b>2. Moyens de production d'images en infrarouge et Raman.....</b>	<b>76</b>
<b>3. Production d'images spectroscopiques .....</b>	<b>78</b>
3.1. Description des dispositifs.....	78
3.2. Principe d'acquisition des spectres.....	81
3.3. Production d'images spectroscopiques : la manière classique et ses limitations.....	81
<hr/>	
<b><i>Chapitre 3 : Application des méthodes de résolution sélectionnées à une matrice de données synthétiques bien caractérisées : Test d'efficacité .....</i></b>	<b>86</b>
<hr/>	
<b><i>I. Introduction .....</i></b>	<b>86</b>
<b><i>II. Les méthodes chimiométriques de l'étude.....</i></b>	<b>88</b>
<b><i>III. Le choix de données synthétiques et la simulation des conditions expérimentales .....</i></b>	<b>89</b>
<b>1. Constitution de la matrice de données synthétiques .....</b>	<b>89</b>
<b>2. Manipulation de ces données synthétiques .....</b>	<b>91</b>
<b>3. La Simulation des conditions expérimentales.....</b>	<b>92</b>
<b><i>IV. La méthodologie générale de résolution .....</i></b>	<b>95</b>
<b>1. Critère de comparaison et d'évaluation.....</b>	<b>95</b>
<b>2. Comparaison globale .....</b>	<b>95</b>
<b><i>V. Application des méthodes de résolution sélectionnées aux données synthétiques.....</i></b>	<b>97</b>
<b>1. Etude du critère particulier de SIMPLISMA : l'influence de l'offset .....</b>	<b>97</b>
<b>2. Etude de l'influence du rapport signal sur bruit sur la capacité d'extraction des méthodes de résolution avec la seule contrainte de non-négativité sur les spectres et les concentrations .....</b>	<b>99</b>
2.1. Etude comparative des capacités d'extraction des méthodes de résolution « simples », MS .....	99
2.1.1. Etude de la résolution en fonction du bruit sans décalage en nombres d'onde .....	99
2.1.2. Etude de la résolution en fonction des décalages en nombres d'onde pour un niveau de bruit donné .....	101
2.1.3. Quelques explications sur la mauvaise qualité d'extraction de certaines méthodes de résolution....	102
2.1.3.1. OPA(spec).....	102
2.1.3.2. PCA.....	104

2.2. Etude comparative des capacités d'extraction des méthodes de résolution « simples » raffinées par MCR-ALS .....	105
2.2.1. Etude de la résolution en fonction du bruit sans décalage en nombres d'onde.....	105
2.2.2. Etude en fonction des décalages .....	107
<b>3. Etude des méthodes de résolution raffinées par MCR-ALS avec les contraintes du rang local en plus de la non-négativité.....</b>	<b>109</b>
3.1. Explication de la méthodologie de la contrainte du rang local .....	109
3.1.1. Estimation du nombre de composés présents sur un pixel.....	110
3.1.2. Production des cartes du rang local.....	111
3.1.3. Production de la carte des absences .....	113
3.1.4. Utilisation des contraintes du rang local .....	114
3.2. Etude des résolutions en fonction du bruit sans décalage en nombres d'onde .....	115
3.3. Observations après décalages en nombres d'onde.....	117
<b>4. Les méthodes de résolution les plus robustes à retenir pour l'application au traitement des données de l'imagerie microspectrométrique.....</b>	<b>118</b>
 <i>VI. Conclusion du Chapitre 3.....</i>	 <i>126</i>
<b><i>Chapitre 4 : Application de notre méthodologie de résolution à un cas concret : analyse d'un calcul urinaire par traitement chimiométrique des données issues de la microspectrométrie Raman.....</i></b>	
 <i>I. Introduction .....</i>	 <i>127</i>
 <i>II. Généralité sur les calculs urinaires.....</i>	 <i>128</i>
1. Les principaux constituants des calculs urinaires .....	128
2. Le calcul urinaire étudié.....	129
 <i>III. Acquisition expérimentale.....</i>	 <i>130</i>
1. Description du spectromètre utilisé.....	130
2. Acquisition expérimentale des données.....	131
 <i>IV. Etude des données acquises.....</i>	 <i>133</i>
1. Exploration globale, estimation du nombre des constituants.....	133
1.1. Exploration globale et estimation du nombre des constituants.....	133
1.2. Les prétraitements nécessaires .....	136
1.2.1. La normalisation .....	136
1.2.2. La correction de la ligne de base.....	137

1.2.3. Le lissage .....	139
1.3. La méthodologie de résolution proposée pour l'étude des calculs urinaires.....	139
<b>2. Extraction par quelques « méthodes simples » avec la seule contrainte de non-négativité.....</b>	<b>141</b>
2.1. Extraction par OPA(var).....	141
2.2. Confirmation des résultats de OPA(var) par l'extraction par SIMPLISMA.....	147
<b>3. Couplage de OPA(var) avec MCR-ALS : résultats et analyses .....</b>	<b>149</b>
<b>4. Application des contraintes du rang local aux résultats du couplage de OPA(var) avec MCR-ALS : résultats et analyses.....</b>	<b>150</b>
<b>5. Résolution par le procédé « augmentation de matrices » en utilisant les résultats fournis par l'application des contraintes du rang local aux résultats du couplage de OPA(var) avec MCR-ALS.....</b>	<b>155</b>
<b><i>V. Bilan du chapitre 4 .....</i></b>	<b><i>159</i></b>
<b><i>CONCLUSION GENERALE .....</i></b>	<b><i>161</i></b>
<b><i>Annexe 1 : Algorithmiques des méthodes étudiées et quelques remarques sur leurs qualités et leurs défauts.....</i></b>	<b><i>164</i></b>
<b><i>Annexe 2 : Quelques fonctions Matlab utilisées .....</i></b>	<b><i>173</i></b>
<b><i>Annexe 3 : Tableaux des dissimilarités associés aux graphes présentés au chapitre 3 .....</i></b>	<b><i>176</i></b>
<b><i>Tables des illustrations.....</i></b>	<b><i>179</i></b>
<b><i>Bibliographie .....</i></b>	<b><i>185</i></b>

## Liste des abréviations

---

ALS	Alternating Least Square
CCD	Charge-Coupled Devices
Coef corr ou coeff corr	coefficient de corrélation
corrcoef max	coefficient de corrélation maximal
cov	covariance
D <sub>c</sub>	données spectrales sous forme tridimensionnelle
D <sub>d</sub>	données spectrales sous forme dépliée
DTGS	Deuterated TriGlycine Sulfate
dw	Durbin-Watson coefficient
FNNLS	Fast Non Negative Least square
FPA	Focal Plane Array
FSIW- EFA	Fixed Size Image Window- Evolving Factor analysis
FSW-EFA ou FSMW-EFA	Fixed Size (moving) Window- Evolving Factor analysis
FTIR	Fourier Transform Infrared
HPLC-DAD	High Performance Liquid Chromatography - Diode Array Detection
IE	Imbedded Error (function)
IKSFA	Iterative Key Set Factor Analysis
IND	INDicator (function)
InGaAs	Indium gallium Arsenic
InSb	Antimonure (Sb) d'Indium (In)
IR	Infrared
IRFPA	Infrared Focal Plane Array
ITTFA et TTFA	(Iterative) Target Transformation Factor Analysis
KSFA	Key Set Factor Analysis
MBE	Molecular Beam Epitaxy
MCR-ALS	Multivariate Curve Resolution – Alternating least Square
MCT (ou HgCdTe)	Mercure-cadmium-Tellure
NA	Numerical aperture (ouverture numérique)
OLE	Ouverture Lumineuse sur l'Echantillon
OPA(spec)	Orthogonal Projection Approach par sélection des spectres
OPA(var)	Orthogonal Projection Approach par sélection des variables

OPUS	OPTics User Service
pab	pas de balayage
PARAFAC	PARAllel FACtor Analysis
PbS	Sulfure de Plomb
PCA	Principal Component Analysis (analyse en composante principale)
PCR	Principal Component regression
PLS	Partial Least Square
PMF	Positive matrix factorization
PRESS	PREdiction Sum of Squares
QWIP	Quntum Well Infrared Photodetector
RE	Real error
RMN	Résonance Magnétique Nucléaire
RMS ou RMSE	Root-Mean-Square Error
RSD	Residual Standard Deviation
SCR	Somme des Carrés de Résidus
SIMPLISMA	SIMPLe-to-use Interactive Self-Modeling Mixture Analysis
SMCR	Self-Modeling Curve resolution
SNV	Standard Normal Variate
SPEXFA	Spectral Isolation Factor Analysis
SVD	Singular Value Decomposition

# INTRODUCTION GENERALE

Aujourd'hui, les images envahissent notre vie quotidienne dans tous ses aspects, de l'audio-visuel à l'imagerie médicale en passant par les images astronomiques, satellitaires, radars, ... La recherche fondamentale (académique ou privée) est le moteur du développement qui fournit les bases techniques de chacune d'elles. La spectroscopie produit de très nombreux types d'imageries. Actuellement l'imagerie microspectrométrique est en pleine expansion sous différentes formes pour de nombreuses spectroscopies.

L'un des domaines dans lequel l'imagerie est très active aussi bien au niveau industriel que militaire est l'infrarouge. L'imagerie microspectrométrique infrarouge est très utilisée pour le contrôle de qualité, la détection des défauts dans les composants électroniques, la recherche des impuretés ou des contaminants dans l'industrie pharmaceutique par exemple. Cependant la production d'images microspectroscopiques se trouve limitée par certaines contraintes telles que la limite de diffraction, la résolution des microscopes, entre autres. En outre, la méthode classique basée sur la production d'images spectroscopiques par la sélection d'une bande spectrale spécifique à un constituant donné atteint ses limites dans un mélange naturel ou industriel complexe. Dans ce cas de figure, les recouvrements spectraux des bandes d'absorption de nombreux constituants sont très importants. Le calcul de l'intensité pour produire l'image spectroscopique par intégration de l'aire sous pic est complètement faussé à cause des contributions probables des bandes d'absorption d'autres composés que celui considéré.

De ce fait depuis les années 1970, les analystes physico-chimistes ont développé et utilisent des méthodes chimiométriques pour atteindre une meilleure résolution. Ces méthodes de résolution se sont constamment améliorées avec l'apparition continue de nouvelles approches.

L'un de nos objectifs est l'étude de la capacité de résolution des méthodes chimiométriques appliquées aux données de l'imagerie microspectrométrique. Dans notre cas, l'investigation porte sur l'imagerie infrarouge et Raman et nous mettrons en évidence l'apport des méthodes chimiométriques sur la qualité de résolution. Le but d'une bonne méthodologie de résolution est d'aller au-delà des limitations inhérentes à l'imagerie spectroscopique classique. En outre, la problématique liée à la résolution spatiale est une grandeur caractéristique aux différents types d'imageries. Compte tenu de l'échelle d'étude et de l'instrumentation utilisée, d'une imagerie à une autre la résolution peut varier de quelques microns à quelques kilomètres. Par exemple, un satellite d'observation à une résolution de l'ordre du mètre ou demi-mètre alors qu'une image microspectrométrique Raman à une résolution de l'ordre du micron. Néanmoins, La

méthodologie d'étude et de résolution des images microspectrométriques infrarouges ou Raman par les méthodes chimiométriques peut être étendue aux autres types d'imageries.

Cependant l'application d'une méthode chimiométrique au traitement de données spectroscopiques est le résultat d'un choix subjectif du chercheur expérimentateur. En effet dans notre revue bibliographique, nous n'avons pas relevé une étude exhaustive portant sur l'application d'un grand nombre de ces méthodes à des données connues et caractérisées pour évaluer leur capacité de résolution respective d'une manière concrète et objective. En plus de cela, très peu de couplages entre ces méthodes de résolution ont été réalisés.

Dans notre étude nous nous proposons de faire cette évaluation pour un certain nombre de méthodes chimiométriques. Principalement certaines choisies parmi celles du « self-modeling » ont été appliquées à l'étude des données spectrales synthétiques bien caractérisées et par la suite à des données réelles. Nous proposons de mettre en valeur le couplage de certaines méthodes et de montrer la pertinence de ce couplage sur la qualité de résolution ; la résolution se faisant sans *a priori* sur les constituants du mélange à étudier.

Dans le premier chapitre nous rappellerons certaines généralités sur l'imagerie spectroscopique et les limitations existantes. Nous rappellerons quelques principes de la chimiométrie. Nous exposerons les méthodologies d'estimation du nombre de constituants et les méthodes de résolution de notre étude.

Au chapitre 2, nous présenterons quelques rappels sur la spectroscopie vibrationnelle, les bases techniques et les moyens de production d'images spectroscopiques afin de mieux mettre en relief notre apport par la suite.

Puis nous comparerons, au chapitre 3, les résultats produits par les méthodes d'intérêt sur des données spectrales synthétiques dont les paramètres sont contrôlés. Nous nous interrogerons sur le lien entre la qualité des images et des spectres extraits et la capacité de résolution de chacune des méthodes. Certaines contraintes rendant compte des propriétés au niveau du pixel, comme les contraintes du rang local, seront spécifiées et appliquées.

La méthodologie d'étude mise en place au chapitre 3 nous permettra, au chapitre 4, de traiter efficacement des données réelles issues de la microspectrométrie Raman.

Enfin nous conclurons en soulignant la pertinence de notre étude.

# Chapitre 1

## Chimimétrie pour la résolution des images

---

### I. Images spectroscopiques et leurs problèmes

#### 1. Images spectroscopiques : généralités

Une image spectroscopique est issue de la mesure d'un signal spectroscopique. Les domaines spectroscopiques sont assez nombreux et les signaux obtenus sont de natures très diverses. Quel que soit la spectroscopie, il faut disposer de l'instrumentation nécessaire pour l'acquisition du signal spectroscopique ainsi que les outils informatiques indispensables aussi bien pour le traitement des données que la production d'images spectroscopiques. Cette production est rendue possible par le gain en précision et en rapidité des instruments spectroscopiques ainsi que leur miniaturisation due aux progrès technologiques. Et cela a permis l'échantillonnage des surfaces de plus en plus petites. De ce fait, on passera du début du siècle à nos jours de la spectrométrie macroscopique à la microspectrométrie.

Les images spectroscopiques sont produites dans des nombreux domaines<sup>1</sup> tels que l'infrarouge, le Raman, la RMN, la spectroscopie UV-Visible, les rayons X (DRX, XPS, EXAFS), la spectrométrie de masse, la spectroscopie électronique de surface (SEM, STM, AFM), la fluorescence, la phosphorescence, ....etc. Cette liste n'est pas exhaustive mais donne une idée de la très grande diversité des spectroscopies et des types d'images qu'on peut produire.

La radiographie, la scintigraphie, la tomographie, le sonar, le scanner, l'échographie et l'IRM en sont les illustrations des nombreuses applications d'usage courant.

Cependant ce qui conditionne la production de l'image spectroscopique, ce sont la nature de l'échantillon étudié et la caractéristique du signal d'analyse. Pour le signal, les méthodes spectrométriques les plus courantes utilisent le rayonnement électromagnétique. Ce type d'énergie se présente sous différentes formes dont la principale est la lumière. Les rayons X, gamma et ultraviolets, les micro-ondes et les ondes radios sont des formes différentes de ce rayonnement électromagnétique.

Une image est reconstruite après un long processus d'acquisition et de traitements de données. Outre l'échantillon, un spectromètre adapté avec son détecteur sensible dans le domaine spectroscopique ciblé et un moyen de capture d'images sont nécessaires (CCD, FPA par exemple). Pour la microspectrométrie, un microscope puissant est associé au spectromètre. Après avoir obtenu l'échantillon aux propriétés voulues (tailles, épaisseurs, quantités, propriétés physiques ou chimiques, ...etc.), celui-ci est exposé au phénomène spectroscopique (radiation lumineuse, radiofréquence, flux d'électrons, ...etc.). Chaque spectroscopie a sa propre instrumentation spécifique pour l'acquisition du signal.

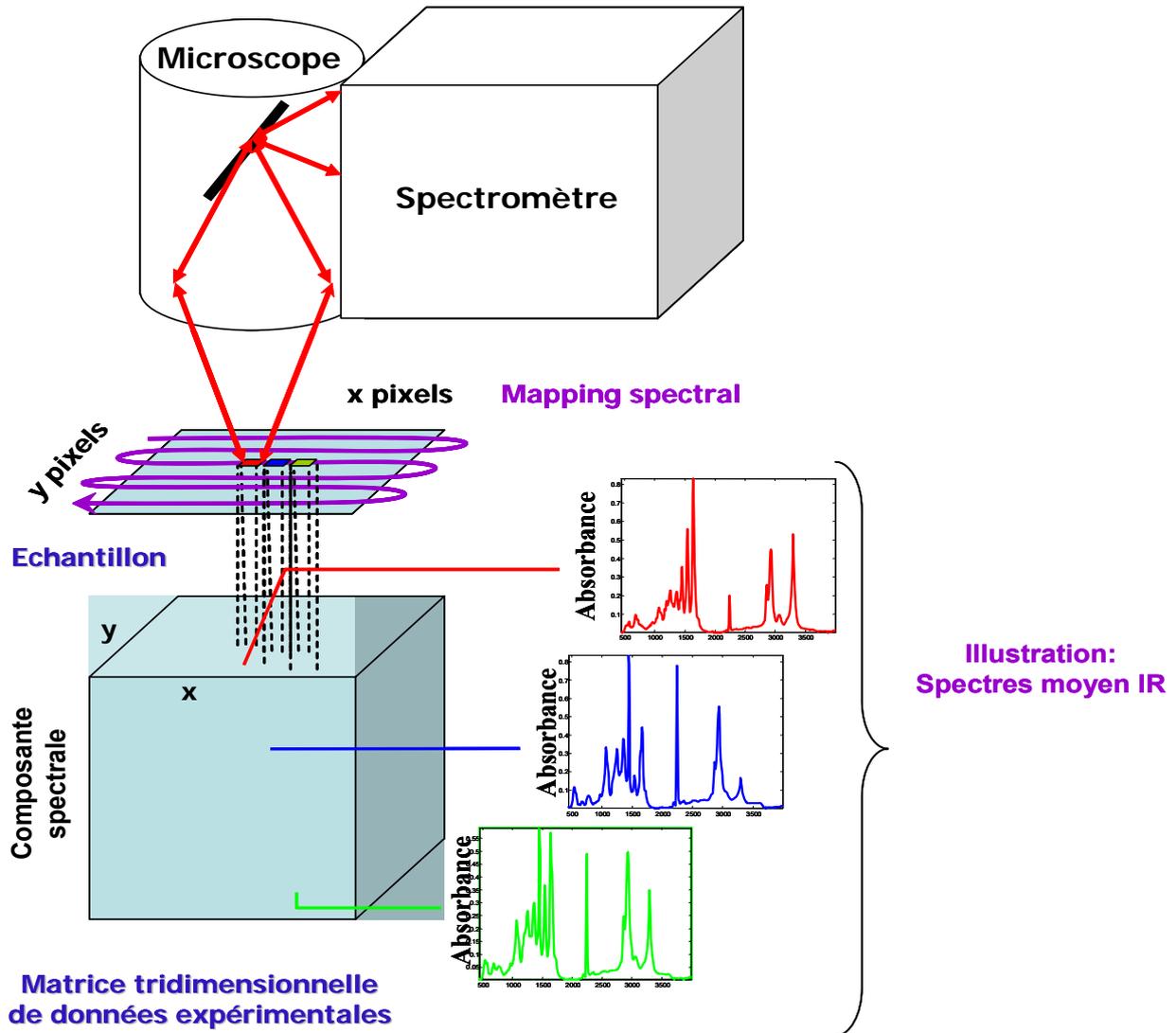
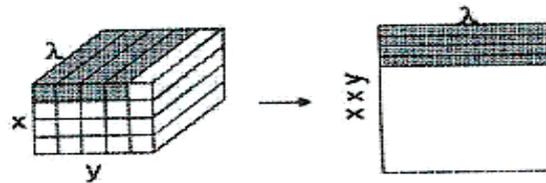


Figure 1 : Procédé d'acquisition de spectres et de constitution de données spectrales.

D'une manière générale, pour aboutir à la formation d'images, on réalise une « cartographie spectrale » sur une portion délimitée de l'échantillon. On effectue un balayage à deux dimensions (suivant x et y) avec un pas régulier. Un déplacement vertical, par contrôle piézoélectrique pour s'adapter à la non-planéité de l'échantillon, est disponible sur certains microscopes. Chaque aire de mesure est une micro zone d'analyse, ce qu'on appelle « pixel ».

Un spectre est acquis sur chacun d'eux. L'ensemble des spectres acquis forme la matrice de données à étudier. Cette matrice est tridimensionnelle. Les deux premières dimensions restituent la répartition spatiale des pixels dans le plan de la « cartographie spectrale ». La troisième dimension exprime les caractéristiques spectrales des données mesurées. Les coordonnées spatiales des pixels serviront à mieux restructurer la formation d'images. La composante spectrale est spécifique à chaque type de spectroscopie et rend compte de la mesure spectroscopique sur chaque pixel analysé. La Figure 1 : illustre le procédé d'acquisition spectrale et de constitution de la matrice de données spectrales. Dans les différentes microspectrométries, les structures des données sont semblables à celle illustrée sur cette figure. C'est seulement la composante spectrale qui change de nature.

**a) Image plane : forme des données acquises et la restructuration pour les traitements**



**b) Image multicouche**

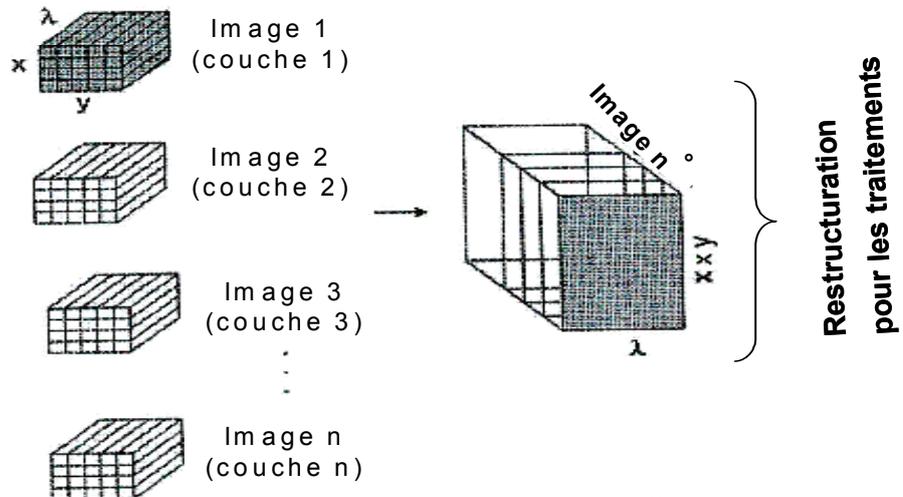


Figure 2 : Les catégories de l'image, la forme des données acquises et leur restructuration pour les traitements (A. De Juan<sup>2</sup> et al.) .

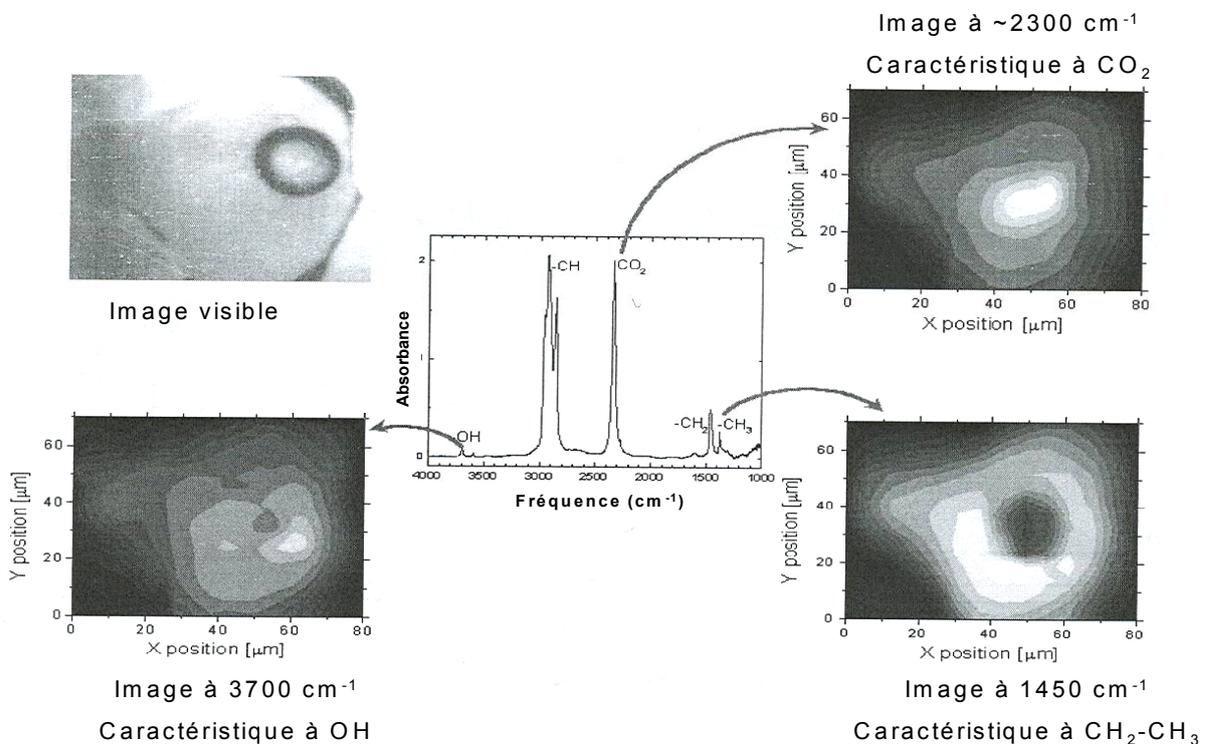
La Figure 2 montre les différentes catégories d'images et les manipulations pour restructurer les données pour les traitements.

## 2. Problématique de reconstitution d'images

Les images sont formées après traitement et analyse de la matrice tridimensionnelle de données. Chaque pixel d'analyse a fourni un spectre. Mais c'est un spectre de mélange de l'ensemble des constituants présents sur le pixel.

La problématique est alors : comment reconstituer l'image d'un composé donné de l'échantillon alors qu'on n'a que des spectres de mélange ?

Les spectroscopistes ont proposé plusieurs démarches. Une d'elles était la recherche d'une caractéristique spécifique à chacun des constituants. Par exemple en Infrarouge, on peut rechercher une longueur d'onde caractéristique, dans un intervalle donné, pour chaque composé de l'échantillon. Et on produit une image à chacune des longueurs d'onde trouvées. Depuis les années 1980, c'était la façon classique de produire les images spectroscopiques. Ceci était possible lorsqu'on avait une maîtrise sur les échantillons analysés, lorsqu'on connaissait leurs caractéristiques et le nombre de leurs constituants. Dans ce cas, l'intégrale de l'aire sous pic de la bande spectrale donnait pour chaque pixel une estimation de la concentration du composé considéré.



**Figure 3 : Illustration de la méthodologie de production d'une image par longueur d'onde caractéristique (P. DUMAS<sup>3</sup> et al.)**

La Figure 3 donne l'illustration de cette méthodologie dans le cadre d'étude d'une inclusion dans une roche. Trois images correspondant à trois fonctions moléculaires sont produites et cela a permis l'identification de la substance incluse : de l'hydrocarbure.

Mais dès que la complexité du mélange augmente, on est confronté à plusieurs difficultés. D'abord, il devient pratiquement impossible de trouver un intervalle spécifique à chaque constituant du mélange à cause des recouvrements spectraux. Ensuite même si on trouve ces intervalles, on ne pourrait exclure la contribution d'autres composés, aussi minime soit-elle. Enfin pour un échantillon complexe naturel ou industriel, dont le nombre de constituants est

inconnu, cette méthode de sélection d'une bande spectrale est inopérante. Ce qui nécessite l'invention d'autres outils de résolution.

Ces difficultés liées à la sélection ne sont pas les seules rencontrées. Il faut aussi faire face aux limitations d'ordre théorique (critère Rayleigh, ... etc.) et instrumentale (sensibilité, limite de détection, ...).

### 3. Les limitations de la production d'images spectroscopiques

#### 3.1. Limites théoriques

En microspectrométrie, on fait l'acquisition spectrale par l'intermédiaire d'un microscope. Il existe une multitude de microscopes (optique, électronique à balayage, à balayage par sonde (à effet tunnel, à force atomique)) aux propriétés et qualités différentes. Les éléments qui les constituent introduisent certaines limitations théoriques. Le pouvoir de résolution connu sous le nom « critère de Rayleigh » pour les microscopes optiques, par exemple, indique la limite théorique de séparation de deux objets proches voisins. Cette relation établie à partir de la théorie de la diffraction d'Abbe donne la distance minimale  $d_m$  entre deux objets pour qu'ils soient correctement séparables.  $d_m$  (éq.1) est exprimé en fonction de la longueur d'onde d'illumination ( $\lambda$ ), de l'indice de réfraction du milieu ( $n$ ) et le demi-angle d'ouverture de l'objectif ( $\theta$ ) :

$d_m = \frac{1,22\lambda}{2n \sin(\theta)}$  (éq.1). Après simplification et en introduisant l'ouverture numérique  $NA$

( $NA = n \sin(\theta)$ ) de l'objectif, on l'exprime souvent :  $d_m = \frac{0,611\lambda}{NA}$ .

#### 3.2. Limitation instrumentale

Lorsque la limite de résolution est atteinte, pour les systèmes optiques par exemple, on atteint la diffraction et les objets ne sont pas nettement discernables. Ce qui nécessite une ouverture plus grande de l'objectif. Mais pour échantillonner des surfaces de plus en plus petites, on est amené à réduire cette ouverture. Ce qui réduit considérablement, l'énergie reçue par l'échantillon. Alors la détection du signal qui a interagi avec l'échantillon et contenant l'information utile à extraire devient très difficile.

Donc l'imagerie microspectrométrique doit concilier le besoin d'analyser des surfaces de plus en plus petites et les différentes limitations d'ordre théorique et instrumentale.

Il apparaît évident donc que les éléments optiques de l'instrumentation, la nature du signal, les contraintes liées à l'échantillonnage introduisent des limitations, entre autres celles que nous venons de mentionner ci-dessus.

#### 4. Quelques outils pour améliorer la résolution ?

Les spectroscopistes-analystes ont introduit depuis les années 1970 l'analyse multivariée des données. Ce qui a donné un ensemble d'outils basés sur la « résolution des courbes multivariées » (Multivariate Curve Resolution : MCR). La base est l'analyse factorielle (Factor Analysis : FA) fondée sur la décomposition en valeurs propres et en vecteurs propres.

Nous avons mentionné précédemment que la mesure spectrale sur un pixel donne un spectre de mélange. L'analyse par MCR permet la résolution de la matrice de données spectrales  $D$  en produit de deux grandeurs physico-chimiquement interprétables  $C$  et  $S^t$  :

$D = C * S^t + E$  (éq.2) ;  $E$  représentant l'erreur minimale commise sur cette décomposition. Pour

l'imagerie microspectrométrique, cela revient à trouver une relation analogue à (éq.2) au niveau

de chaque pixel :  $D_p = \sum_{i=1}^q C_i S_i^t$  ;  $D_p$  le spectre mélange du pixel,  $C_i$  concentration spécifique du

composé  $i$  sur le pixel considéré,  $S_i^t$  le spectre caractéristique du composé  $i$  à extraire du mélange et  $q$  le nombre estimé des composés.

L'importance dans cette démarche est que la résolution peut se faire sans connaissance a priori sur les spectres ou les concentrations. Cependant une estimation du nombre des constituants est nécessaire pour une meilleure résolution.

De très nombreuses méthodologies sont disponibles pour faire cette estimation. Nous exposerons dans la partie suivante quelques unes d'utilisation courante.

## II. Méthodologies d'estimation du nombre des constituants

Les principales méthodes d'estimation du nombre de constituants sont basées sur l'analyse factorielle (FA). La méthodologie sous-jacente est la décomposition de la matrice de données en valeurs propres et en vecteurs propres. Le nombre de vecteurs propres les plus significatifs seront retenus comme étant le nombre de constituants à estimer. Pour une meilleure compréhension de la méthodologie de sélection du « nombre significatif de vecteurs propres », une description succincte du procédé de décomposition en « valeur propres et en vecteurs propres » est donnée dans la partie suivante.

### 1. Définition et Bases mathématiques de l'analyse factorielle (Factor Analysis : FA)<sup>4 5 6</sup>

#### 1.1. Définition

Malinowski, dans son ouvrage « Factor analysis in Chemistry » édité en 1991, donne la définition suivante :

*« L'analyse factorielle (FA) est une technique des méthodes de l'analyse multivariée pour la réduction des matrices de données à leur plus faible dimensionnalité par l'utilisation d'un espace factoriel orthogonal et des transformations qui donnent des prédictions et/ou des facteurs reconnaissables ».*

Cela suppose bien sûr la maîtrise des fondements mathématiques de l'analyse factorielle et l'interprétation de ces concepts transposés aux données physico-chimiques. FA est très utile et permet :

- ✓ l'étude des données avec une grande complexité,
- ✓ de simplifier des données énormes,
- ✓ et donne des facteurs significativement interprétables.

#### 1.2. Les fondements mathématiques

L'analyse factorielle est une technique d'étude. Le but est de décomposer une matrice des données recueillies  $D$  en produit de deux matrices  $D = C * S' + E$  (cf. éq.2),  $E$  étant la matrice des résidus exprimant la variance non expliquée par les facteurs retenus pour la décomposition.  $C$  et  $S$  les matrices de la décomposition.

Les traitements mathématiques expliquent comment effectuer cette décomposition de la manière la plus efficace possible. Il se pose deux problèmes :

- ✓ Pour faire cette décomposition, il faut estimer le nombre des facteurs significatifs.

- ✓ Ensuite il faut transformer cette décomposition mathématique pour lui donner un sens physique interprétable relatif aux données étudiées.

Le but de cette décomposition est de retrouver des « facteurs caractéristiques », des données étudiées, exempts de tout préjugé de l'expérimentateur et peu influencés par les différents types d'erreurs. La matrice de données expérimentales  $D$  est souvent un tableau rectangulaire. Les lignes représentent des individus (échantillons, observations...etc.) et les colonnes des variables à valeurs numériques continues (des mensurations, des taux, des mesures spectrales...etc.). Les lignes ont, en général, des caractères répétitifs (ou formées d'unités ayant des caractères répétitifs). Le Tableau 1 présente la forme matricielle de cette matrice de données  $D$ . Les éléments  $Ech_i$  représentent des individus, des espèces, des échantillons expérimentaux (...etc.) pour lesquels les variables  $var_i$  ont été enregistrées. L'expérimentateur cherche à savoir les comportements des  $Ech_i$  pour les  $var_i$ , la relation pouvant exister entre les  $Ech_i$ , entre les  $var_i$  ou entre les  $Ech_i$  et les  $var_i$ . On cherche surtout à ramener la grande taille des données mesurées à quelques facteurs simples facilement interprétables ayant un sens physique.

**Tableau 1 : Présentation de la matrice des données recueillies, observées, mesurées (...etc.) sous la forme d'un tableau à 2 dimensions.**

		Variables (observation, mesures...)				
		$var_1$	$var_2$	$var_3$	.... $var_j$ ....	$var_p$
Échantillon (espèce, individu...)	$Ech_1$	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	.... $d_{1,j}$ ....	$d_{1,p}$
	$Ech_2$	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	.... $d_{2,j}$ ....	$d_{2,p}$
	$Ech_3$	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$	.... $d_{3,j}$ ....	$d_{3,p}$
	.	.	.	.	.	.
	$Ech_i$	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	.... $d_{i,j}$ ....	.... $d_{i,p}$ ....
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	$Ech_m$	$d_{n,1}$	$d_{n,2}$	$d_{n,3}$	.... $d_{n,j}$ ....	$D_{n,p}$

**Matrice des données  
D**

FA décompose alors cette matrice  $D$  en un produit de deux matrices, tenant compte du nombre de facteurs choisi. Le Tableau 2 montre cette décomposition d'une manière schématique. D'où la première étape, de l'analyse factorielle, qui est l'estimation du nombre des facteurs significatifs

donnant un sens physique à cette décomposition. Plusieurs méthodes sont disponibles pour cette estimation.

**Tableau 2 : Décomposition de la matrice de données D en produit de 2 matrices R et C en fonction du nombre de facteurs significatifs choisis (n).**

Échantillon (espèce, individu...)	Variables (observation, mesures...)						=	n facteurs				*	n facteurs				n facteurs
	var <sub>1</sub>	var <sub>2</sub>	var <sub>3</sub>	....	Var <sub>j</sub>	....		var <sub>p</sub>	r <sub>11</sub>	r <sub>12</sub>	.....		r <sub>1n</sub>	c <sub>11</sub>	c <sub>21</sub>	.....	
Ech <sub>1</sub>	d <sub>1,1</sub>	d <sub>1,2</sub>	d <sub>1,3</sub>	....	d <sub>1,j</sub>	....	d <sub>1,p</sub>	r <sub>21</sub>	r <sub>22</sub>	.....	r <sub>2n</sub>	c <sub>21</sub>	c <sub>22</sub>	.....	c <sub>1p</sub>		
Ech <sub>2</sub>	d <sub>2,1</sub>	d <sub>2,2</sub>	d <sub>2,3</sub>	....	d <sub>2,j</sub>	....	d <sub>2,p</sub>	r <sub>31</sub>	r <sub>32</sub>	.....	r <sub>3n</sub>	.	.	.....	.		
Ech <sub>3</sub>	d <sub>3,1</sub>	d <sub>3,2</sub>	d <sub>3,3</sub>	....	d <sub>3,j</sub>	....	d <sub>3,p</sub>	.	.	.....	.	.	.	.....	.		
.	.	.	.	....	.	....	.	.	.	.....	.	.	.	.....	.		
Ech <sub>i</sub>	d <sub>i,1</sub>	d <sub>i,2</sub>	d <sub>i,3</sub>	....	d <sub>i,j</sub>	....	d <sub>i,p</sub>	.	.	.....	.	.	.	.....	.		
.	.	.	.	....	.	....	.	.	.	.....	.	.	.	.....	.		
Ech <sub>m</sub>	d <sub>m,1</sub>	d <sub>m,2</sub>	d <sub>m,3</sub>	....	d <sub>m,j</sub>	....	d <sub>m,p</sub>	r <sub>11</sub>	r <sub>11</sub>	.....	r <sub>mn</sub>	c <sub>n1</sub>	c <sub>n2</sub>	.....	c <sub>np</sub>		
	<b>Matrice des données D</b>							<b>Matrice ligne abstraite R</b>					<b>Matrice colonne abstraite C</b>				

Estimer le nombre des facteurs significatifs pour la décomposition d'une donnée réelle revient à estimer la part d'erreurs (erreur expérimentale, erreur systématique...etc.) et surtout la part due au bruit. Pour toute spectroscopie, la matrice de données réelles D peut être écrite :  $D = D^* + E$  (éq.2bis), E étant la matrice des résidus correspondant aux erreurs et au bruit, D\* la matrice des données expurgées des erreurs.

Une méthode très utilisée, pour l'estimation de facteurs de décomposition, est la *décomposition en valeurs singulières* ( *Singular Value Decomposition* : *SVD*) de la matrice de données physico-chimiques.

Cette décomposition en valeurs singulières est basée elle-même sur la décomposition en valeurs propres et en vecteurs propres. Nous donnons ci-dessous une explication de cette décomposition.

### 1.2.1. Décomposition en valeurs propres et vecteurs propres

Soit  $B(n \times n)$  une matrice carrée. Pour décomposer B en vecteurs propres et en valeurs propres, il faut que B soit non *singulière*, c'est-à-dire son déterminant doit être différent de zéro. Dans ce cas, B est donc inversible.

Une fois cette condition vérifiée :

- ✓ On cherche d'abord les valeurs propres  $\lambda_i$  en résolvant les équations données par le polynôme caractéristique  $\det(B - \lambda I) = 0$  (éq.3) (noté  $|B - \lambda I| = 0$ ), I étant la matrice identité. Chaque équation donne un certain nombre de valeurs propres.
- ✓ Ensuite on recherche le vecteur propre  $v_i^p$  associé à chaque valeur propre  $\lambda_i$  en résolvant l'équation  $(B - \lambda_i I) * v_i^p = 0$  (éq.4).
- ✓ Comme B est une matrice carrée, le nombre des valeurs propres et de vecteurs propres est égal à la dimension de B, soit n.

En réalité, les matrices étudiées sont rectangulaires et la démarche décrite ci-dessus n'est pas directement applicable. Soit  $D(r \times c)$  une matrice rectangulaire. Cette matrice n'est pas directement inversible et on ne peut pas calculer son déterminant. On travaille alors avec les matrices carrées  $D' * D$  et  $D * D'$ . Ces dernières ont le même rang que  $D$  et se décomposent dans le même espace vectoriel. Les vecteurs propres de  $D$  sont les mêmes que ceux de  $D' * D$ .

En posant  $X = D' * D$ , on se ramène au cas des matrices carrées. La recherche des vecteurs propres et des valeurs propres se fait alors de la même façon.

La décomposition en « vecteurs propres » et en « valeurs propres » est la base de FA mais c'est la forme « valeurs singulières » qui est souvent explicitée et utilisée. Dans la partie suivante nous rappelons cette décomposition.

### 1.2.2. Décomposition en valeurs singulières (Singular Value Decomposition : SVD)

Dans le cas d'étude des données réelles physico-chimiques, il faut décomposer en valeurs singulières la matrice  $D(r \times c)$ ,  $r$  lignes et  $c$  colonnes (souvent  $c < r$ ). Comme  $D$  est très souvent rectangulaire, la décomposition se fait sur  $X = D' * D$  ou  $X' = D * D'$  selon la forme de  $D$ , ou simultanément sur les deux.  $X$  est appelé la matrice des variances-covariances ou la matrice des corrélations. Décomposer en valeurs singulières une matrice de données  $D$  revient à trouver 2

matrices  $U$  et  $V$  orthogonales telle que  $D = U W V^t = \sum_{k=1}^N \lambda_k u_k v_k^t$  (éq.5) pour  $N$  facteurs de décomposition.  $W$  étant une matrice diagonale de dimension  $(N \times N)$  avec les éléments de la diagonale positifs,  $U(r \times N)$  la matrice des vecteurs propres de  $X' = D * D'$  (décomposition à droite) et  $V(N \times c)$  celle des vecteurs propres de  $X = D' * D$  (décomposition à gauche), le  $t$  en exposant signifiant transposé et  $N$  le rang de la matrice  $D$ .  $U$  et  $V$  sont dites orthogonales car  $U^t * U = U * U^t = I$  et  $V * V^t = V^t * V = I$ .  $I$  est la matrice identité. Ce qui veut dire que  $U^{-1} = U^t$  et  $V^{-1} = V^t$ . Pour une interprétation physique des données,  $U$  et  $V$  forment une base orthogonale respectivement pour les colonnes et les lignes de  $D$  de dimension respectivement  $r$  et  $c$ .

L'objectif de cette décomposition est d'obtenir la matrice  $W$  diagonale contenant les valeurs singulières. Pour cela, il faut diagonaliser  $X'$  et  $X$ . Alors diagonaliser  $X'$  et  $X$  revient à chercher leurs valeurs propres et leurs vecteurs propres. La démarche à suivre est celle décrite dans le paragraphe précédent.

Finalement puisque  $U$  et  $V$  sont orthogonales et contiennent des vecteurs propres orthogonaux aussi, la matrice diagonale  $W$  rattachée à  $D$  est calculée de la manière suivante

(éq.6)  $W = U^t D V$ . Dans  $W$  les valeurs singulières, qui sont en réalité les racines carrées des valeurs propres de  $X = D^t * D$ , sont classées dans l'ordre décroissant  $w_1 \geq w_2 \geq \dots \geq w_N \geq 0$ .

Tout cela découle du théorème mathématique qui s'énonce comme suit :

*Pour toute matrice  $A(m \times n)$ , il existe deux matrices orthogonales  $U(m \times m)$  et  $V(n \times n)$ , telles que :  $U^t A V = W(m \times n)$  où  $W = \text{diag}[w_1, \dots, w_p]$ , avec  $p = \min(m, n)$ .*

Ce théorème se démontre d'une manière simple. Cette démonstration, la décomposition en valeurs propres et en vecteurs propres ainsi que la décomposition en valeurs singulières sont assez abondantes dans les ouvrages à caractère pédagogique. Nous citons pour l'exemple celui de Rotella<sup>7</sup> et al.

## 2. Recherche des facteurs significatifs à l'aide des valeurs singulières

L'obtention des valeurs singulières, issues de la décomposition de la matrice de données, ne termine pas notre résolution pour la recherche des facteurs significatifs. Pour une matrice de dimension  $(r \times c)$  comme notre matrice de données  $D$  et supposant que  $c < r$ , on aura au maximum  $c$  valeurs singulières. Mais toutes les colonnes de la matrice  $D$  ne sont pas, toujours, linéairement indépendantes. Alors la liste des valeurs singulières contient des valeurs nulles, autant de fois qu'il y a de relations de colinéarité. Ce qui revient à dire que le rang de  $D$  n'est pas égal à  $c$  mais à un nombre  $p$  inférieur à  $c$ . Pour l'estimation des facteurs significatifs à l'aide des valeurs singulières, on procède comme suit :

- ✓ **première étape de résolution** : on retient dans  $W$  que les valeurs singulières non nulles. Ce qui réduit la dimension de l'espace de décomposition et le nombre de vecteurs propres non nuls à  $p$ .
- ✓ **deuxième étape de résolution** : pour une matrice de données  $D$  contenant un nombre élevé de lignes et de colonnes, on obtient un nombre de valeurs singulières non nulles assez élevé aussi. Un très grand nombre de ces valeurs singulières ont une valeur très proche de zéro. On ne retient alors que les grandes valeurs singulières qui donnent une bonne reconstitution de la matrice de données  $D$ . Pour cela on s'aide souvent du diagramme du pourcentage individuel et pourcentage cumulé des valeurs propres en s'autorisant dans la plupart des cas une marge d'erreur de 5%. Les  $q$  plus grandes valeurs propres servent à reconstituer la matrice de données de départ supposée exempte d'erreurs (bruit et tout type d'erreur). Ces  $q$  valeurs propres sont associées aux  $q$  vecteurs propres représentant les  $q$  facteurs significatifs recherchés.

Le Tableau 3 montre cette procédure de restriction du nombre de valeurs propres pour trouver le nombre de facteurs significatifs.

Donc la matrice  $W^*(r \times q)$ , ne contenant que les plus grandes valeurs singulières correspondant aux facteurs significatifs, va être retenue. Le Tableau 4 illustre cette démarche. Les valeurs propres les plus significatives, dans  $W^*(r \times q)$ , sont appelées les « valeurs singulières primaires » et les autres les « valeurs singulières secondaires ».

La matrice  $D^*$  reconstruite par cette décomposition est obtenue alors par la relation suivante qui est (éq.5) modifiée légèrement :

$$D^* = \bar{U}W^*\bar{V}^t = \sum_{k=1}^q \lambda_k u_k v_k^t \quad (\text{éq.7}).$$

$\bar{U}$  et  $\bar{V}^t$  étant les matrices

orthogonales dont les bases sont les  $q$  vecteurs propres associés aux  $q$  valeurs propres significatives retenues,  $W^*$  la matrice définie ci-dessus (Cf. Tableau 4). Les résidus de la décomposition correspondent alors à  $D - D^* = E$  (éq.8). Connaissant les valeurs dans  $D$  et celles dans  $D^*$ , dans la plupart des procédures itératives, il s'agit de minimiser cette différence.

Une autre façon d'écrire (éq8) est la suivante en explicitant les termes :

$$\sum_{i=1}^r \sum_{k=1}^c (d_{ik}^2 - d_{ik}^{*2}) = \sum_{j=q+1}^c \lambda_j = \sum_{i=1}^r \sum_{j=q+1}^c (\sigma_{ij})^2 \quad (\text{éq.8bis}),$$

$\sigma_{ij}$  étant un facteur ligne rattaché à l'erreur

résiduelle. La première partie de l'égalité se démontre facilement par le fait que la trace de  $D$  est égale à celle de  $D^*$  et chacune est égale à la somme des  $\lambda_j$  ( $j=1$  à  $c$ ). La deuxième égalité nous indique que l'erreur résiduelle est égale à la somme des valeurs propres secondaires.

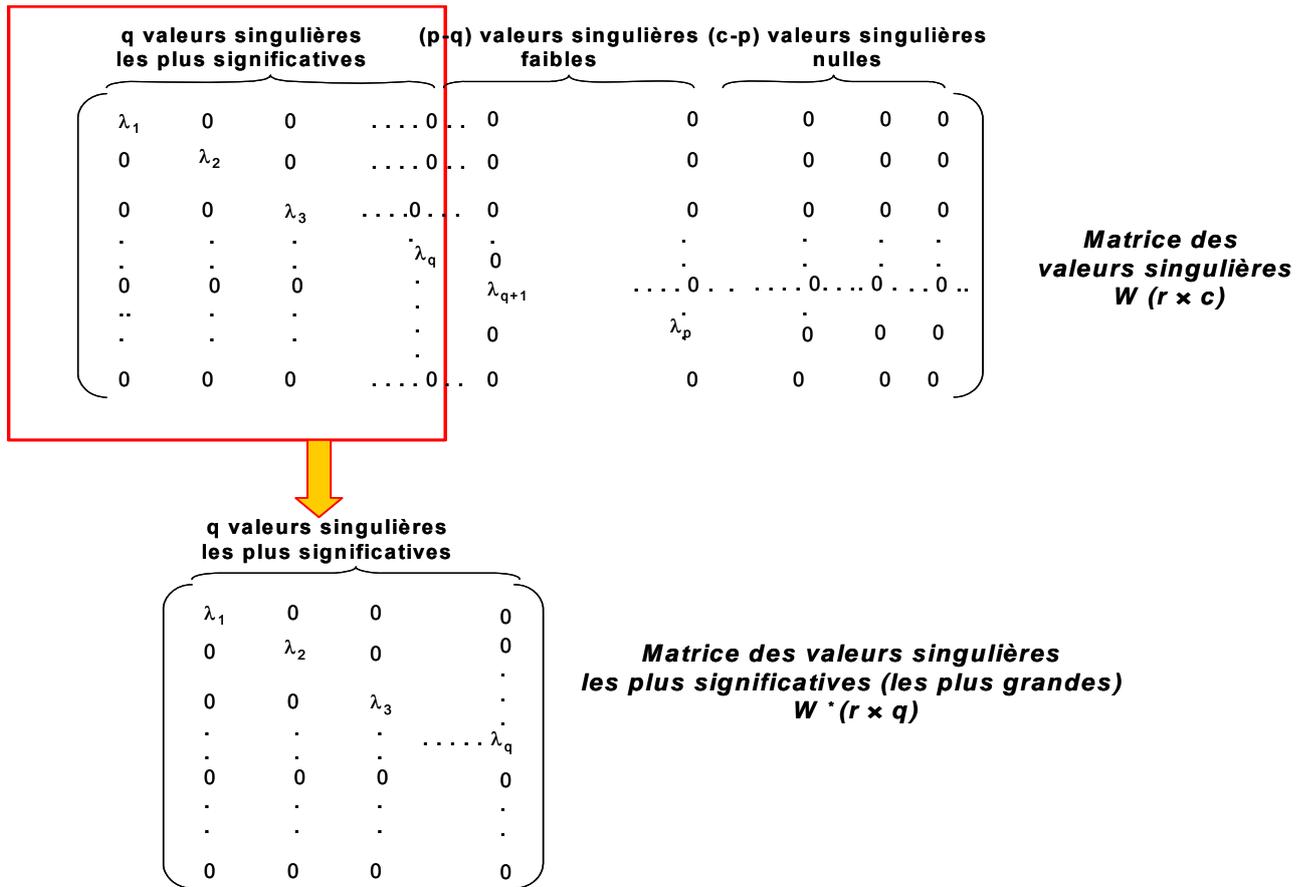
**Tableau 3 : Classement des valeurs singulières dans  $W$  diagonale ( $\lambda_1 > \lambda_2 > \dots > \lambda_p$ ) pour ne retenir dans la décomposition en facteurs significatifs que les «  $q$  » valeurs singulières significatives.**

q valeurs singulières les plus significatives			(p-q) valeurs singulières faibles			(c-p) valeurs singulières nulles		
$\lambda_1$	0	0	....0..	0	0	0	0	0
0	$\lambda_2$	0	....0..	0	0	0	0	0
0	0	$\lambda_3$	....0..	0	0	0	0	0
.	.	.	.	.	.	.	.	.
.	.	.	$\lambda_q$	.	.	.	.	.
0	0	0	.	0	....0..	....0..	....0..	....0..
.	.	.	.	$\lambda_{q+1}$	.	.	.	.
.	.	.	.	.	$\lambda_p$	.	.	.
.	.	.	.	0	.	0	0	0
.	.	.	.	.	.	.	.	.
0	0	0	....0..	0	0	0	0	0

**Matrice des Valeurs singulières  $W$  ( $r \times c$ )**

Cette dernière démonstration n'est pas évidente et demande des transformations. Plus de détails peuvent être trouvés dans l'ouvrage de Malinowski<sup>8</sup>.

**Tableau 4 : Sélection de matrice  $W^*$  ( $r \times q$ ) des valeurs propres significatives associées aux principaux vecteurs propres donnant la dimension de l'espace de décomposition, c'est-à-dire les facteurs significatifs.**



Comme nous pouvons voir, minimiser cette différence (éq.8bis) revient à minimiser la somme des carrés des résidus.

### 3. Les autres méthodes<sup>9 10</sup> d'estimation de facteurs significatifs pour la décomposition

Nous citons ci-dessous, en donnant succinctement quelques formules et définitions, les principales méthodes utilisées pour l'estimation du nombre de facteurs significatifs pour une matrice de données  $D(r \times c)$ . La classification des méthodes d'estimation que nous avons adoptée, dans ce qui suit, est celle de Malinowski.

#### 3.1. Les méthodes empiriques

Ces méthodes empiriques sont apparues au fur et à mesure dans l'analyse de données. Elles sont dites « empiriques » car elles sont souvent les résultats de l'intuition des chercheurs expérimentateurs. Les principales méthodes sont les suivantes.

- **Fonction d'erreur incluse ou intégrée (Imbedded Error function) IE.**

C'est une fonction définie à partir des « valeurs singulières secondaires  $\lambda^s$  » et qui s'exprime de

la façon suivante : 
$$IE = \left( \frac{q \sum_{j=q+1}^c \lambda_j^s}{rc(c-q)} \right)^{1/2}$$
 (éq.9). Le nombre  $q$  qui minimise cette fonction donne une

estimation du nombre de facteurs cherchés.

- **Fonction d'indicateur de facteur (factor indicator function) IND.**

Cette fonction est proposée en 1977 par Malinowski. Elle est plus précise que celle précédente

(IE). Cette fonction se définit avec les mêmes variables que IE,  $IND = \frac{RE}{(c-q)}$  (éq.10), RE étant

l'erreur réelle qui est la déviation standard résiduelle (RSD) que nous verrons dans la partie suivante. Lorsqu'on a une estimation de l'erreur réelle RE, cette fonction IND atteint un minimum lorsque le nombre de facteurs significatifs  $q$  est employé.

- **Variance du pourcentage cumulé.**

C'est une méthode simple très souvent utilisée et qui permet de choisir un seuil pour ne retenir que les plus grandes valeurs propres donnant le nombre des composés significatifs présents dans les données étudiées ; autrement dit le nombre de facteurs nécessaires pour la décomposition.

Cette variance du pourcentage cumulé se calcule par la relation :

$$Variance \text{ \% cumulé} = \left( \frac{\sum_{i=1}^r \sum_{k=1}^c d_{ik}^{*2}}{\sum_{i=1}^r \sum_{k=1}^c d_{ik}^2} \right) \text{ (éq.11).}$$

- **Fonction d'Exner  $\psi$ .**

Cette fonction, dont l'utilisation dans le traitement des données physico-chimiques est suggérée par Weiner<sup>11</sup> en 1974, est préconisée dans le cas où une bonne mesure de l'erreur expérimentale n'est pas possible. Cette fonction s'exprime de la manière suivante:

$$\Psi(n) = \left( \frac{\sum_{i=1}^r \sum_{k=1}^c (d_{ik} - d_{ik}^*)^2}{\sum_{i=1}^r \sum_{k=1}^c (d_{ik} - \bar{d})^2} \times \frac{rc}{rc-n} \right) \text{ (éq.12) où } \bar{d} \text{ est la moyenne totale de la matrice}$$

expérimentale D. Le résultat fourni par cette fonction d'Exner<sup>12</sup> doit être soumis à un raisonnement critique car il y a une limite de validité pour la signification physique (limite=1)

comme on peut la voir dans le Tableau 5. Elle n'est jugée crédible d'une manière certaine que lorsque la valeur obtenue est de l'ordre de 0.1.

**Tableau 5 : Signification et interprétation des résultats fournis par la fonction d'Exner.**

<b>Fonction d'Exner <math>\psi = (0 \rightarrow \infty)</math></b>	<b>1</b>	<b>0.5</b>	<b>0.3</b>	<b>0.2</b>	<b>0.1</b>
<b>interprétation</b>	Limite supérieure pour la signification physique	Plus grande valeur à accepter selon Exner	Corrélation lointaine	Bonne corrélation	Excellente corrélation

Il existe beaucoup d'autres méthodes empiriques que nous ne mentionnerons pas.

### 3.2. Les méthodes statistiques<sup>9</sup>

Ces méthodes statistiques sont assez nombreuses et il est impossible de les citer toutes. Nous n'en citerons que deux.

- **La « Cross Validation ».**

Cette méthode est assez utilisée dans le cadre de la calibration mais ici il s'agit d'estimer le nombre de facteurs significatifs pour une matrice de données D. L'implémentation de la Cross-Validation connue sous le nom « leave-one-out » (laisser-un-dehors) est appliquée pour extraire le nombre de facteurs significatifs cherchés. Dans cette méthodologie, la matrice D est divisée en r matrices de dimension (r-1, c) en excluant chaque fois une ligne de D. Chacune des sous-matrices est analysée par FA en faisant varier le nombre de facteurs abstraits jusqu'à atteindre un résultat similaire à celui obtenu par « target testing » sur la ligne manquante. La différence entre le point cible et le point prédit par l'utilisation des facteurs abstraits est calculée. L'évaluation de la méthode porte sur la « prédiction de la somme des carrées des erreurs » PRESS(j) calculée pour chaque niveau des facteurs j. PRESS(j) se calcule par la relation :

$$PRESS(j) = \sum_{i=1}^r \sum_{k=1}^c (d_{ik} - \hat{d}_{ik}(j))^2 \text{ (éq.13) avec } \hat{d}_{ik}(j) \text{ les éléments prédits des lignes supprimées}$$

obtenues pour les facteurs j. D'une manière générale et préférentiellement, on utilise le rapport R des PRESS(j) :  $R = \frac{PRESS(j+1)}{PRESS(j)}$  (éq.14). Le calcul de ce rapport nous montre si le facteur

(j+1) apporte ou pas une amélioration. Si  $R > 1$ , il n'y a pas d'amélioration. Par contre si  $R < 1$ , alors il y a une amélioration. D'autres transformations sont possibles pour ramener cette expression sous forme d'un F-test et définir d'autres critères sur ce test.

- **Test de Durbin-Watson<sup>13</sup>** .

C'est un test de la statistique mathématique dont le résultat (la distribution de Durbin-Watson) est souvent lié à la nature de la matrice de données D. Dans chaque type d'analyse<sup>14,15,16</sup> une adaptation est nécessaire. Le critère de Durbin-Watson  $dw$  est utilisé en général dans les régressions pour étudier l'aspect aléatoire de leurs résidus et détecter la corrélation existant entre les variables considérées dans D. De ce fait, ce critère est défini à partir des résidus de deux variables consécutives (lignes, colonnes, dissimilarité, pureté ...etc.). L'une des hypothèses sous-jacentes est que l'on suppose, dans ce test, que les observations et les résidus ont un ordre naturel. La définition de ce critère  $dw$  dépend des données étudiées et de ce l'on veut obtenir. On

peut le définir d'une manière assez générale comme suit :  $dw = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$  (éq.15),  $dw$

étant le coefficient de Durbin-Watson et  $e_i$  le résidu rattaché à la valeur  $i$ ,  $n$  le nombre total des valeurs dans le calcul. Cela aboutit à un intervalle de confiance et une « bande d'indétermination ». De très nombreux travaux ont été entrepris depuis les années 1950 pour améliorer les résultats et lever l'indétermination. Notre but n'est pas de faire l'historique ni de présenter un détail exhaustif sur cette méthode mais de rappeler le principe et la méthodologie d'estimation du nombre de facteurs significatifs.

- ❖ **Méthodologie** : s'il existe une forte corrélation entre les points successifs (mesures, variables, observations...etc.), les valeurs  $dw$  convergent vers zéro (0). Dans le cas d'une très faible corrélation, la convergence est vers 2 ; ce qui est le signe de données dont les variables sont non corrélées. Le nombre  $q$  ( $q < n$ ) pour lequel on atteint la convergence est retenu comme le nombre de facteurs significatifs cherchés.

Le *F-test* de Fisher-Snedecor, le *t-test* de Student sont d'autres tests statistiques.

### 3.3. Les méthodes basées sur l'estimation de l'erreur expérimentale

Ces méthodes utilisent des éléments simples comme la variance, l'écart-type, la déviation standard, les éléments  $(i, j)$  de la matrice de données étudiées, les valeurs propres. On se sert de ces techniques lorsqu'une estimation de l'erreur expérimentale est disponible.

- **La « déviation standard résiduelle » RSD (Residual Standard Deviation).**

Cette « déviation standard résiduelle » est l'erreur réelle RE. RSD (éq16) se définit à partir des « valeurs singulières secondaire » qui, elles-mêmes, sont rattachées aux erreurs résiduelles :

$$RSD = \left( \frac{\sum_{j=q+1}^c \lambda_j^s}{r(c-q)} \right)^{1/2} \quad (\text{éq.16}), \quad \lambda^s \text{ représentant les « valeurs singulières secondaires » pour une}$$

matrice  $D(r \times c)$ ,  $q$  étant le nombre des « valeurs singulières primaires ». A l'aide des transformations et en introduisant les éléments  $d_{ik}$  de la matrice des données  $D$ , cette même

$$\text{relation peut s'écrire } RSD = \left( \frac{\sum_{i=1}^r \sum_{k=1}^c d_{ik}^2 \sum_{j=q+1}^c \lambda_j^s}{rc(c-q)} \right)^{1/2} \quad (\text{éq.17}). \text{ Cette formule inclus les éléments } d_{ik}^2$$

de la matrice de données  $D$  et successivement chacune des « valeurs propres secondaires »  $\lambda_j^s$ .

❖ *Méthodologie :*

- 1°) Il faut avoir une estimation de l'erreur expérimentale (EE).
- 2°) On calcule la RDS sur la base d'un seul facteur pour toute la matrice  $D$ .
- 3°) La valeur propre la plus grande est considérée comme la seule constituante de l'ensemble « valeurs propres primaires », les autres valeurs propres étant toutes considérées comme les « valeurs propres secondaires ».
- 4°) On compare la RSD calculée par la formule de (éq.16) et l'erreur expérimentale estimée (EE). Si  $RSD \cong EE$ , le nombre de facteurs retenus est bon (1 dans le premier cas). Par contre si  $RSD > EE$ , le nombre de facteurs choisis n'est pas suffisant. Alors on intègre la deuxième valeur propre dans le calcul de la RSD dans la formule de (éq.16).
- 5°) Le processus de calcul de la RSD et de comparaison se poursuit jusqu'à ce que  $RSD \cong EE$ . A ce moment là, on est arrivé à séparer les « valeurs propres primaires » et les « valeurs propres secondaires » et à repérer le nombre de facteurs significatifs (nombre de constituants) pour une bonne décomposition de  $D$ .

• ***L'erreur quadratique moyenne (Root-Mean-Square Error) RMS ou RMSE.***

Cette méthode est utilisée aussi dans des nombreux cas. RMS (éq.18) mesure la différence entre la donnée brute  $D(r \times c)$  et celle reconstruite avec  $q$  facteurs significatifs à choisir  $D^*(r \times c)$ :

$$RMS = \left( \frac{\sum_{i=1}^r \sum_{k=1}^c (d_{ik} - d_{ik}^*)^2}{rc} \right)^{1/2} \quad (\text{éq.18}). \text{ Lorsque RMS est minimale, le nombre de facteurs}$$

cherchés est atteint.

- **Chi-carré ( $\chi^2$ ).**

Barlett<sup>17</sup> a proposé d'utiliser ce critère en 1950 pour les données où la déviation standard varie d'un point à un autre. C'est une méthode qui prend en compte la variabilité de l'erreur d'un point à un autre mais qui demande qu'on ait une estimation de l'erreur pour chaque point. Lorsque ces conditions sont remplies, le  $\chi^2$  est donné par la relation :  $\chi_q^2 = \sum_{i=1}^r \sum_{k=1}^c \frac{(d_{ik} - d_{ik}^*)^2}{\sigma_{ik}^2}$  (éq.19), avec  $\sigma_{ik}$  la déviation standard associée à la valeur mesurable  $d_{ik}$  et  $d_{ik}^*$  la valeur correspondant au point reconstruit par l'utilisation des  $q$  plus grandes valeurs propres.

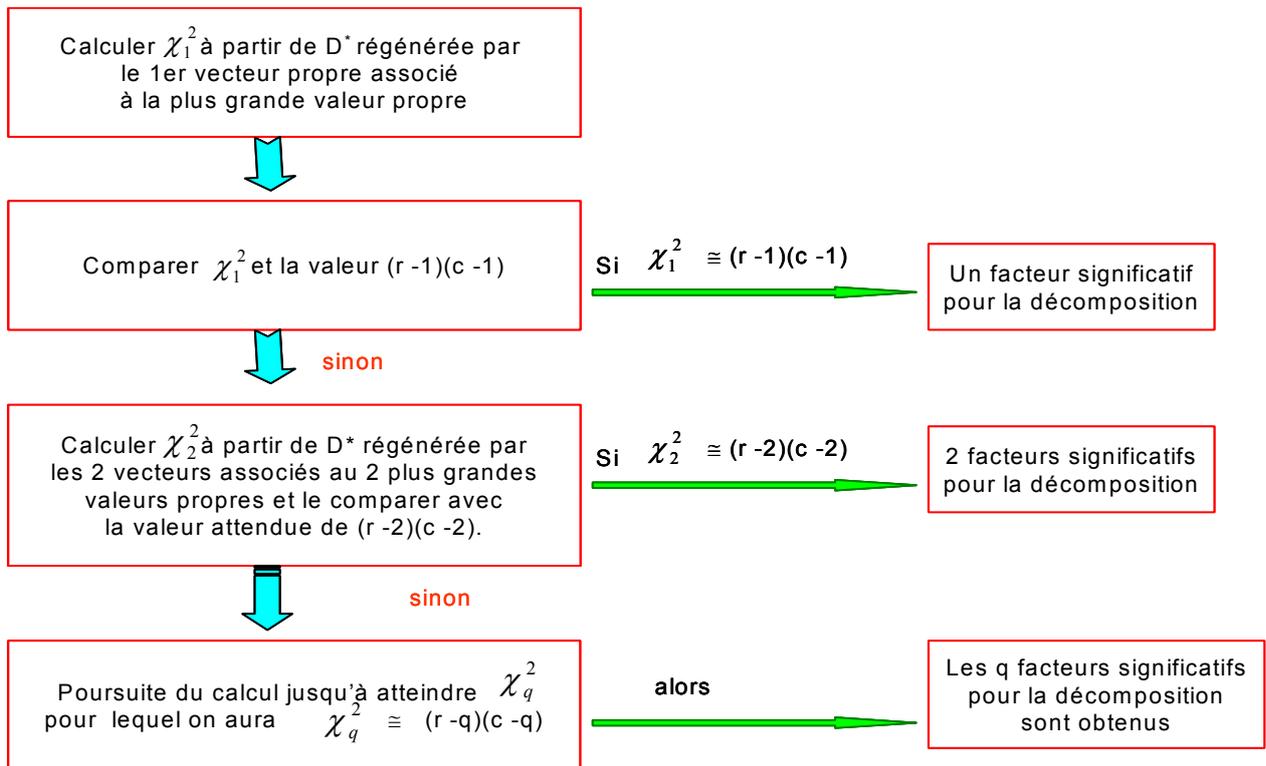


Figure 4 : Diagramme montrant le calcul itératif pour obtenir les  $q$  facteurs significatifs fournis par le  $\chi^2$  pour la décomposition de la matrice de données D.

❖ **Méthodologie** : le diagramme de la Figure 4 résume le calcul itératif jusqu'à ce que le  $\chi_q^2$  calculé avec la relation (éq.19) pour  $q$  facteurs soit pratiquement égal au  $\chi_q^2(\text{attendu}) = (r - q)(c - q)$ . A ce moment-là les  $q$  facteurs qui réalisent cette condition sont retenus pour la décomposition de la matrice de données et sont les facteurs significatifs recherchés.

### 3.4. Quelques méthodes<sup>18</sup> du « self-modeling »

Parmi les très nombreuses méthodes du « self-modeling », nous ne citerons que quatre qui peuvent être utilisées pour estimer le nombre de constituants (facteurs significatifs) d'une matrice de données d'un mélange.

- ***Evolving Factor Analysis (EFA).***

Cette méthode est utilisée au départ dans la chromatographie liquide. Ensuite elle a été généralisée. Elle comprend deux méthodologies : EFA progressive (EFA-forward) et EFA régressive (EFA-backward). La technique est basée sur la détection du rang de la matrice de données D par la décomposition en valeurs singulières des sous-matrices de D. Le rang de D estimé donne le nombre de facteurs cherchés ou du moins une approximation s'il y a une déficience. Dans la partie suivante consacrée aux premières estimations de concentrations et de spectres des constituants d'un mélange, nous reviendrons plus en détail sur EFA et ses dérivées.

- ***L'Analyse en Composantes Principales (PCA : Principal Component Analysis).***

PCA est l'une des méthodes de l'analyse factorielle la plus ancienne. PCA est basée sur la décomposition en « valeurs singulières », la « SVD ». Par conséquent détecter le nombre de facteurs significatifs revient, comme on l'a déjà vu, à chercher le nombre des « valeurs propres significatives ». Nous reviendrons pour plus de détails sur PCA dans la partie suivante.

- ***SIMPLISMA (SIMPLLe-to-use Interactive Self-Modeling Mixture Analysis).***

SIMPLISMA est l'une des méthodes du « self-modeling » qui n'est pas basée sur l'analyse factorielle. Cette méthode permet à la fois la détection de nombre de facteurs significatifs et l'extraction de spectres associés aux différents constituants. Par une procédure itérative simple basée sur un « critère de pureté », on arrive à obtenir le nombre de facteurs significatifs ou constituants d'un mélange sous analyse. On procède par extraction successive des variables les plus pures jusqu'à ce que le « spectre de la pureté » ne représente que le bruit de fond. Plus de détails sur la procédure seront donnés au paragraphe III.3.1 suivant.

- ***OPA (Orthogonal Projection Approach).***

OPA est une méthode basée sur l'orthogonalisation de Gram-Schmidt. Par une transformation de la matrice de données, on cherche à extraire les spectres les plus dissimilaires les uns aux autres (les plus orthogonaux). Par une procédure itérative similaire à celle de SIMPLISMA, mais en s'intéressant cette fois à la dissimilarité de chaque variable, on arrive à déterminer le nombre de facteurs significatifs. Lorsque le spectre de dissimilarité, par exemple, ne manifeste que le bruit de fond, on atteint une bonne estimation du nombre facteurs cherchés. OPA sera traité plus en détail dans le paragraphe III.3.2 ci-dessous.

Après avoir estimé correctement le nombre des constituants de données, il reste à extraire le spectre et la concentration de chacun d'eux. Dans la partie suivante, nous exposerons cette méthodologie d'extraction.

### III. Premières estimations : méthodes simples d'extraction de spectres et de concentrations

Dans cette démarche, la décomposition de la matrice de données par une méthode d'extraction doit aboutir à l'extraction des spectres des composés du mélange et de leurs concentrations associées. Le traitement chimiométrique est appliqué à la matrice de données spectrales. En effet, comme écrit S. Wold<sup>19</sup>, père de la régression PLS, toute donnée acquise expérimentalement peut être écrite d'une manière générale :

**Donnée spectrale (D) = Modèle chimique (M) + Bruit (E)** (éq.20) du point de vue statistique. Mais chaque expérimentation a ses propres contraintes et ses conditions particulières qui doivent être définies convenablement.

Le but, dans tous les cas, est de retrouver les principales caractéristiques des constituants de l'échantillon sous analyse. Il s'agit d'extraire le spectre caractéristique de chaque constituant et la concentration de cette même espèce. Ce qui nous a déjà permis d'écrire l'équation générale de la résolution sous la forme  $D = C * S^t + E$  (cf. éq.3), C la matrice des concentrations des espèces,  $S^t$  celle des spectres de ces mêmes espèces et E la matrice des résidus (erreur sur la décomposition ou variance non expliquée par les facteurs retenus). Nous décrirons ci-dessous quelques méthodes pour cette réalisation.

#### 1. SVD et PCA

Bien que leur dénomination soit différente, les deux méthodologies SVD et PCA sont similaires et procèdent de la même façon pour l'extraction des profils spectraux caractéristiques de différents constituants.

PCA est l'une des méthodes factorielles les plus anciennement utilisées dans l'analyse des données surtout non physico-chimiques telles que les données de la psychométrie<sup>20</sup>, de la psychanalyse, de sciences de comportement social ...etc. PCA est conçu pour la première fois par K. Pearson<sup>21</sup> en 1901, puis intégré à la statistique mathématique par H. Hotelling en 1933. H. H. Harmon<sup>22</sup> nous rapporte, dans son ouvrage de référence, l'historique de « l'analyse factorielle ». Les travaux de C. Spearman sont assez conséquents dans ce sens, notamment la « théorie de deux-facteurs » qui a trouvé une bonne application dans la recherche psychologique. Du fait de sa très grande contribution au développement de « l'analyse factorielle », Spearman est considéré comme le véritable père fondateur. Dès lors l'aspect statistique de l'analyse factorielle, dont PCA, va être utilisé dans les traitements et l'analyse des données de nature très diverse à l'exception de celles de physico-chimie. En effet, PCA est une technique mathématique

qui ne demande de spécifier ni le modèle statistique qui sous-tend ni beaucoup d'hypothèses sur la distribution des variables sous étude. Des détails mathématiques plus approfondis pour la compréhension de PCA sont assez abondants dans la littérature comme par exemple le livre de H. Martens<sup>23</sup> et *al.*

Quant à l'analyse des données physico-chimiques, ce sont W.H. Lawton et E.A. Sylvestre<sup>24</sup> qui ont introduit PCA. Ils l'ont appliqué à un ensemble de spectres issus de la spectrophotométrie. Dans un mélange spectrophotométrique où on ne considère que la présence de deux espèces différentes, ils ont réussi à interpréter et à transformer les facteurs abstraits mathématiques extraits en composantes physico-chimiquement appréciables.

Avant de procéder au traitement, on réalise en général quelques prétraitements (centrage et réduction par exemple) pour homogénéiser les données étudiées et supprimer la dispersion relative aux variables (valeurs trop différentes, valeurs non exprimées dans la même unité, ...etc.).

PCA<sup>25,26</sup> aide à l'étude et à l'interprétation des données récoltées en les décomposant en peu de facteurs simples interprétables. Cette décomposition obéit aux propriétés mathématiques de l'analyse factorielle. La recherche des axes factoriels et la projection des données sur ces axes permettent leurs analyses et l'extraction d'informations utiles. L'idée, dans PCA, est de ramener un nombre important de variables des données expérimentales à un nombre réduit de variables synthétiques. Ce nombre de variables synthétiques est sensé bien représenter les principales caractéristiques des variables de données étudiées. Dans l'étude des données physico-chimiques, l'utilisation de PCA n'est devenue efficace que depuis les trente dernières années grâce aux puissants moyens de calcul.

- **SVD**

Dans le paragraphe II.1.2.2 précédent, nous avons exposé la méthodologie de la SVD. L'application de la SVD à une matrice de données  $D$  peut être résumée par la relation suivante :  $SVD(D) = [U, W, V]$  ;  $W$  la matrice diagonale contenant les valeurs singulières,  $U$  et  $V$  les matrices de vecteurs propres de la décomposition gauche et droite de  $D$ . Les vecteurs propres contenus dans  $V$  donnent une estimation des spectres des constituants. L'estimation des concentrations  $C$  est donnée par le produit  $U^*W$ . Cela aide à donner une première interprétation des constituants du mélange analysé.

- **PCA**

Comme la SVD, PCA est basée sur la décomposition en valeurs propres et en vecteurs propres. Ces vecteurs propres constituent les axes factoriels recherchés. Ensuite, le but est de retrouver les spectres caractéristiques des composés étudiés et des concentrations associées. La projection de l'ensemble des variables sur les axes factoriels (*les composantes principales*) donne la matrice  $S$

des *Loadings* représentant les spectres extraits. De même la projection de l'ensemble des individus-spectres sur les composantes principales donne la matrice *C* des *Scores* représentant les concentrations associées aux spectres extraits. Les *loadings* indiquent le poids de la contribution de chaque variable à chacune des composantes principales extraites. Les *scores* sont les coordonnées de chaque individu-spectre sur ces composantes principales.

Cette décomposition donne en général une première idée des spectres des constituants. L'interprétation de ces *loadings* aide à l'orientation de l'analyse et à la recherche de meilleures solutions.

## 2. EFA et ses dérivées

Nous avons donné une description synthétique de EFA au §II.3.4 précédent. C'est une méthode qui était, au départ, proposée pour détecter l'apparition de différents produits en élution dans la chromatographie liquide à haute performance par détection par la barrette de diode (HPLC-DAD). C'est une technique basée sur la décomposition en valeurs singulières d'une sous matrice de la matrice de données étudiées *D*.

### 2.1. La naissance de EFA

En chromatographie liquide, les composés apparaissent dans un certain ordre. Dès qu'il y a apparition d'un composé nouveau, le rang de la matrice de données  $D(r \times c)$  croît d'une unité. L'idée dans EFA mis au point par Maeder<sup>27</sup> est, en utilisant des sous-matrices  $M_i$  dont le nombre de lignes augmente régulièrement de 1 à  $n$ , de détecter la variation du rang de la matrice qui indique l'apparition du nouveau composé. Pour confirmer ce résultat, on effectue aussi la même démarche en sens inverse à l'aide des sous-matrices  $Q_i$  en commençant par le dernier spectre et en remontant jusqu'au premier spectre. Dans ce dernier cas, on détecte la disparition du composé. La Figure 5 montre, d'une manière schématique, les comportements de ces deux procédures de EFA. Pour les deux façons de procéder de EFA, on effectue une décomposition en valeurs singulières de la sous-matrice  $M_i$  ou  $Q_i$ . Le nombre de valeurs singulières « non nulles » donne le rang de la sous-matrice. Ce rang de la sous-matrice est relié au nombre de composés présents dans le mélange chromatographique. Le graphe du logarithme des valeurs singulières, en fonction du paramètre d'étude, permet de visualiser l'apparition (EFA forward) ou la disparition (EFA backward) d'un composé du fait de la brusque variation des valeurs singulières par rapport au niveau de bruit défini. En général, les logarithmes des valeurs singulières sont représentés en fonction du temps de rétention, du pH, du numéro de mesure ou des coordonnées spatiales.

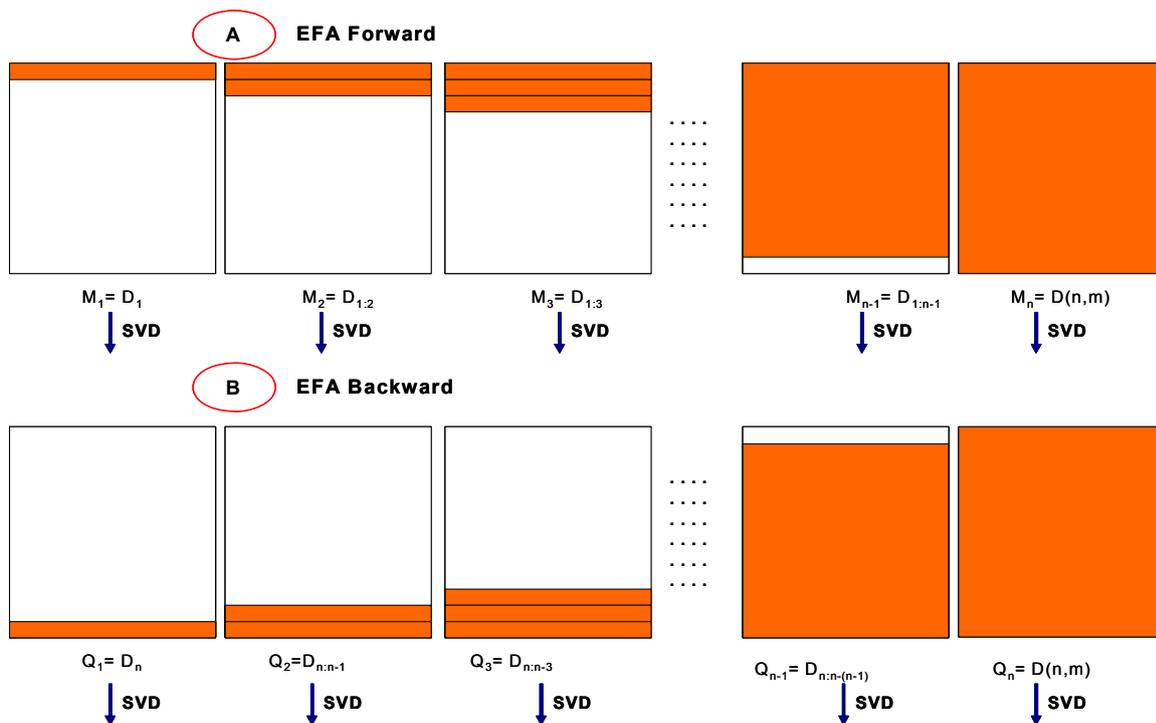


Figure 5 : Explication de la méthodologie de la procédure EFA. A) EFA forward par addition successive des lignes en commençant à la 1<sup>ère</sup> ligne ; B) EFA backward par addition successive des lignes en commençant par la dernière ligne.

C'est pour cette raison que EFA est souvent utilisé pour obtenir une estimation du nombre des facteurs significatifs pour la décomposition de la matrice de données  $D$ . Le nombre de facteurs significatifs recherchés est le rang de la matrice.

La Figure 6 montre un graphe de *EFA*, celle obtenue par Maeder pour l'étude d'un mélange chromatographique contenant 4 composés en élution. On observe facilement sur ce graphe l'apparition du premier composé au temps de rétention 6, le 2<sup>ème</sup> au temps 12, le 3<sup>ème</sup> au temps 19 et le 4<sup>ème</sup> au temps 26 pour EFA forward. Pour EFA backward les composés apparaissent en ordre inverse. Les traits pleins en dessous indiquent la fenêtre de concentration montrant le temps de présence de chaque composé dans le mélange.

EFA a donné des bons résultats lorsqu'on étudie des mélanges chromatographiques ou des systèmes réactifs pour lesquels les constituants ont des concentrations similaires d'une part, et d'autre part pour lesquels les spectres ou chromatogrammes ne sont pas très corrélés.

Dans les mélanges où il y a des composés minoritaires, par cette méthode classique de EFA, on risque de ne pas détecter ces composés minoritaires. Lorsque les spectres sont très corrélés on risque d'aboutir à une déficience du rang. Ce qui pose un problème.

Une adaptation devenait nécessaire pour surmonter ces difficultés. Cela aboutit à la naissance de FSW-EFA ou FSMW-EFA.

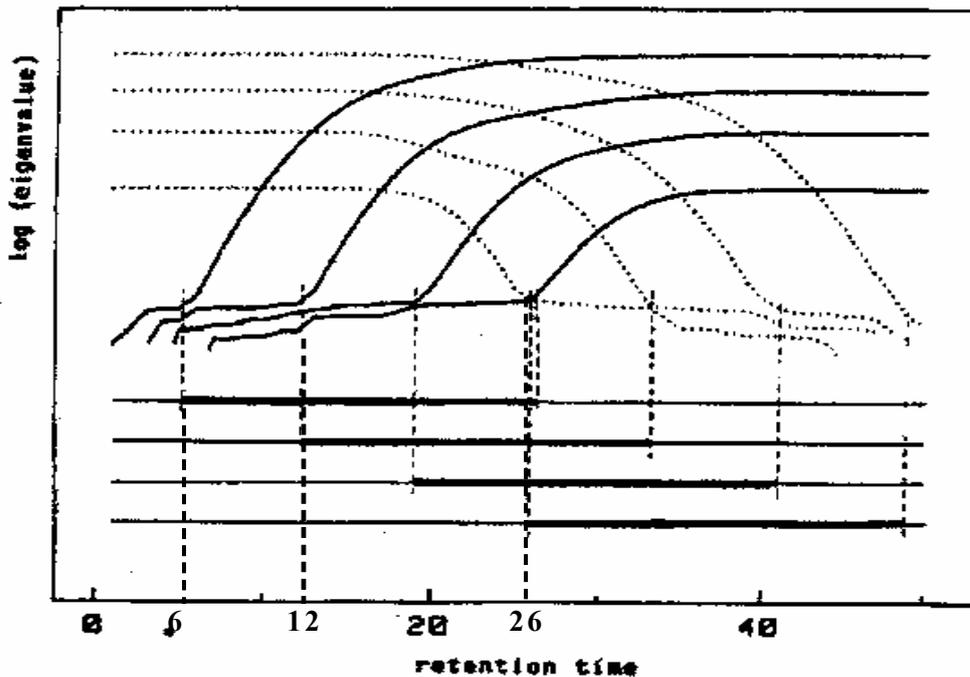


Figure 6 : Graphe EFA, représentation du logarithme des valeurs singulières en fonction du temps de rétention. Les courbes en tiret plein sont fournies par EFA Forward et celles en pointillé (.....) par EFA Backward. Les traits pleins en dessous indiquent la fenêtre de concentration pour chaque composé (M. Maeder<sup>27</sup>).

## 2.2. L'adaptation de EFA à une fenêtre fixe en déplacement : l'apparition de FSW-EFA (ou FSMW-EFA)

FSMW-EFA (Fixed Size Moving Window – Evolving Factor Analysis) est mis au point par Keller<sup>28</sup> et al. en 1992. Pour détecter les composés minoritaires qui ne sont pas décelables par EFA, l'idée est de choisir une fenêtre de taille fixe et de balayer toute la matrice de données avec cette fenêtre. A chaque position, cette fenêtre englobe un certain nombre de spectres. La dimension de cette fenêtre doit être égale au moins au nombre de composés estimés présents dans le mélange plus 1 ( $q+1$ ). Lors du balayage, pour chaque position de la fenêtre, une SVD est réalisée sur la matrice correspondant à l'ensemble de spectres couverts par la fenêtre. La fenêtre se déplace chaque fois d'un spectre jusqu'à atteindre le dernier. On obtiendra autant de valeurs singulières que le nombre de spectres inclus dans la fenêtre.

Pour la visualisation des résultats, le graphe *FSMW-EFA* représente le logarithme des valeurs singulières en fonction du numéro des fenêtres. On effectue une analyse similaire à celle de EFA pour la détection de l'apparition de différents composés en élution dans le mélange. Pour une donnée spectrale contenant  $r$  spectres, le nombre de fenêtres ( $nw$ ) est donné par la relation suivante :  $nw = r - ws + 1$ ,  $ws$  étant la dimension de la fenêtre c'est-à-dire le nombre de spectres contenus dans la fenêtre.

La Figure 7, tirée des travaux de F. Cuesta Sanchez<sup>29</sup> et al., illustre la représentation du logarithme des valeurs singulières en fonction du numéro des fenêtres pour une fenêtre de taille égale à 3. Comme on peut voir, pour chaque fenêtre on obtient 3 valeurs singulières. Ce qui donne en fin de compte 3 courbes qui indiquent l'évolution de la concentration de 3 composés différents dans le mélange en fonction du temps. En mettant en parallèle l'évolution de la concentration en fonction du temps (cf. Figure 7b) et le graphe du logarithme des valeurs singulières en fonction des fenêtres (cf. Figure 7a), on détecte l'ordre et le temps d'apparition de chaque composé. La Figure 8 schématise la procédure FSMW-EFA avec une fenêtre de dimension fixée à 4.

FSMW-EFA permet donc de surmonter certaines difficultés rencontrées par EFA. Son application à l'analyse de données assez diverses est mentionnée abondamment dans la littérature<sup>30, 31</sup>.

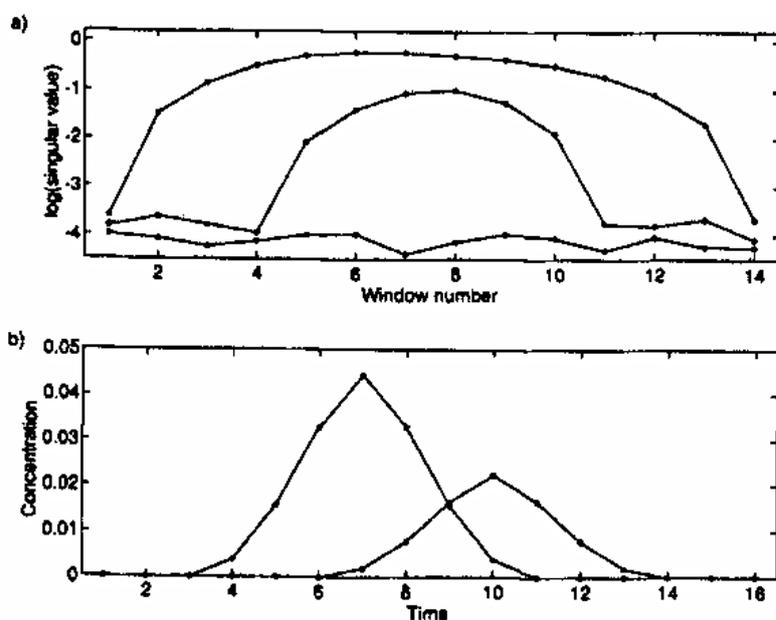
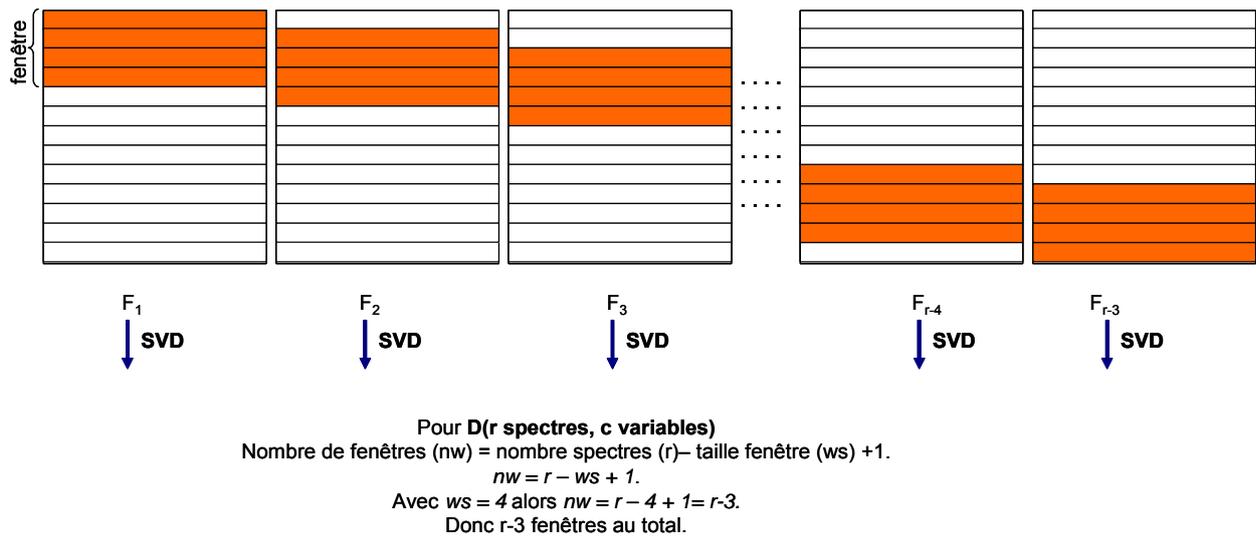


Figure 7 : Graphe *FSMW-EFA* ; a) représentation du logarithme des valeurs propres (ou singulières) en fonction du numéro des fenêtres ; b) concentration en fonction du temps de rétention. (F. Cuesta Sanchez<sup>29</sup> et al.).

EFA est bien adapté à l'analyse de données de structure simple. *FSMW-EFA* renforce cette capacité mais pour l'analyse des données multidimensionnelles EFA et *FSMW-EFA* ne sont pas bien adaptés. Par exemple, l'application de EFA ou *FSMW-EFA* au traitement des données de l'imagerie 2-D (ou plus) rencontre un certain nombre de problèmes.

Il y a une difficulté de prendre en compte les qualités « proches voisins » des pixels ou l'organisation spatiale pour les images multicouches. Une adaptation de *FSMW-EFA* pour l'imagerie multidimensionnelle est donc nécessaire.



**Figure 8 : Méthodologie de la procédure FSMW-EFA. Dans cette illustration la fenêtre de balayage a une taille de 4 lignes. A chaque étape, elle se déplace d'un spectre et donne les différentes fenêtres  $F_i$  schématisées dans le schéma. A chaque position, une SVD est réalisée.**

### 2.3. L'utilisation de EFA pour l'imagerie 2-D, 3-D ou N-D : la création de FSIW-EFA

Pour pouvoir analyser des données multidimensionnelles (N-D avec  $N \geq 2$ ) avec une fenêtre de dimension fixe et restituer correctement les propriétés relatives à l'organisation spatiale des données étudiées, FSIW-EFA qui est une adaptation de FSMW-EFA est proposé par Anna De Juan<sup>32</sup> et *al.* en 2005. L'algorithme FSIW-EFA mime le protocole suivi lors de la collecte des données de l'image spectroscopique et la procédure de scans sur tous les pixels d'analyse de l'échantillon étudié. Le choix de la dimension de la fenêtre, le balayage ainsi que la décomposition en valeurs singulières s'effectuent de la même façon que EFA. Mais cette fois-ci la fenêtre est aussi multidimensionnelle : par exemple une dimension 3D,  $(x,y,\lambda)$  pour une image 2D ;  $(x,y,z)$  ou  $(x,y,z,\lambda)$  pour une image multicouche (4D dans ce dernier cas).

La Figure 9 montre la méthodologie de cette procédure FSIW-EFA. Le graphe *FSIW-EFA* permet de choisir un seuil pour distinguer les valeurs singulières rattachées au bruit et celles caractéristiques des composés présents dans le système étudié. Ce graphe des valeurs singulières est en fonction du numéro des fenêtres ou des indices des valeurs singulières, ou plus principalement en fonction du numéro des pixels. Nous reviendrons sur cette procédure FSIW-EFA au chapitre 3 (Cf. chapitre3, §V.3.1.1) lorsque nous étudierons une matrice de données synthétiques simulant les données de l'imagerie microspectrométrique. Et cela dans le cadre de la détermination du rang local d'une image.

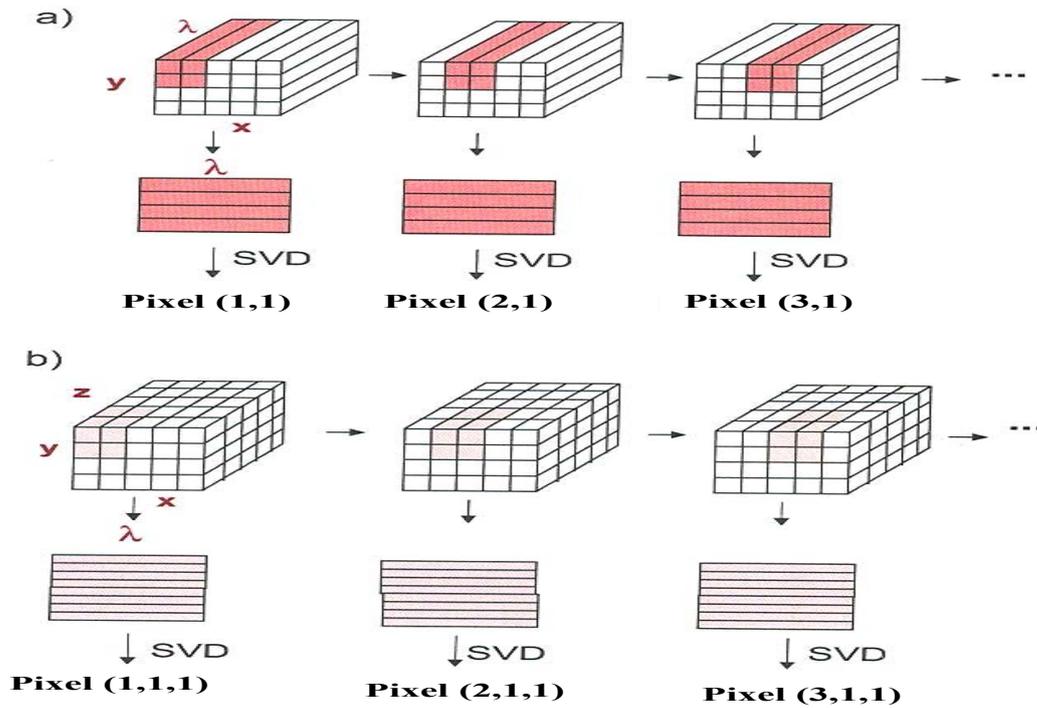


Figure 9 : Explication du déroulement de la procédure FSIW-EFA appliquée à une image 2D (a) et à une image 3D multicouche (b). (Anna De Juan<sup>32</sup> et al.).

### 3. SIMPLISMA (SIMPLE-to-use Interactive Self-Modeling Mixture Analysis) et OPA (Orthogonal Projection Approach)

#### 3.1. SIMPLISMA

C'est une méthode de décomposition proposée par W. Windig<sup>33</sup> et al. en 1991. Ils ont montré l'efficacité d'extraction de celle-ci en l'appliquant aux données déjà traitées en 1976 par G.L. Riter<sup>34</sup> et al. Ces derniers, quant à eux, ont utilisé l'approche du « self-modeling mixture analysis » (ISMA). Windig et son coauteur ont montré que les résultats sont identiques à ceux obtenus par G.L. Riter et al. Et cela, que ce soit les variables pures extraites, les spectres purs obtenus et les concentrations associées. Ils ont exposé les bases théoriques d'extraction de différentes variables pures successives. SIMPLISMA est utilisé pratiquement dans tous les domaines de la spectroscopie. Une bonne synthèse de différents travaux dans lesquels SIMPLISMA est utilisé comme méthode de résolution peut être trouvée dans l'article de Windig<sup>35</sup> et al.

SIMPLISMA peut être utilisé aussi pour obtenir une estimation du nombre de facteurs significatifs pour la décomposition d'une matrice de données D comme brièvement mentionné au paragraphe II.3.4 de ce chapitre. En effet SIMPLISMA est une méthode de résolution basée sur la recherche des « pures variables » caractéristiques des constituants du mélange étudié. En utilisant des éléments simples comme l'écart-type  $\sigma$ , la moyenne  $\mu$  et un offset d'ajustement  $\alpha$ , on évalue la pureté  $p_j$  de chaque variable colonne  $j$  de la matrice de données D.

Dans leur interprétation géométrique de SIMPLISMA, Windig<sup>33</sup> et *al.* ont démontré la relation liant la longueur  $\delta$  (norme), la moyenne  $\mu$  et la déviation standard  $\sigma$  d'une variable colonne  $j$  par

$\delta_j^2 = \mu_j^2 + \sigma_j^2$  (éq.21), avec  $\sigma_j = \sqrt{\frac{\sum_{i=1}^m (d_{i,j} - \mu_j)^2}{m}}$  et  $\mu_j = \frac{\sum_{i=1}^m d_{i,j}}{m}$ . Cette relation s'apparente à

celle de Pythagore dans un triangle rectangle. Après avoir défini la signification de chaque terme et la propriété rattachée, ils définissent la pureté d'une variable  $j$  par  $p_j^1 = \frac{\sigma_j}{\mu_j}$ ; le 1 signifiant le

calcul de la première pureté. Le problème rencontré est que des variables avec une moyenne trop faible vont avoir une pureté très grande. Le rôle de l'offset  $\alpha$  introduit au dénominateur est

d'éviter cet inconvénient. La pureté s'exprime alors  $p_j^1 = \frac{\sigma_j}{(\mu_j + \alpha)}$  (éq.22). La valeur de  $\alpha$  est

un pourcentage de la valeur moyenne maximale d'absorbance. Une valeur de 1 à 3 est préconisée dans la plupart des publications. En outre, un autre ajustement est nécessaire. Les variables n'ayant pas la même moyenne, la pureté doit tenir compte de l'importance de chaque variable. Pour cela on introduit un poids  $w_i$  dans le calcul pour tenir compte de cette importance.

L'expression générale de la pureté (cf. annexe 1.2), s'écrit donc  $p_j^k = \frac{\sigma_j}{(\mu_j + \alpha)} * w_j^k$  (éq.23) pour

le  $k^{\text{ème}}$  calcul de la pureté pour la variable  $j$ . La pureté ainsi définie nous permet d'extraire correctement les variables avec leur influence relative.

Après avoir défini les paramètres nécessaires et choisi l'offset convenable, SIMPLISMA fournit les tracés du spectre de la pureté et du spectre de la pureté corrigé par la déviation standard en fonction du numéro « des variables colonnes » de la matrice de données D.

La Figure 10 montre la première interface obtenue lorsqu'on utilise SIMPLISMA (avec un offset de 10) pour le traitement d'une matrice de données D contenant 2500 spectres dont chaque spectre est mesuré pour 460 longueurs d'onde.

Pour l'estimation du nombre de facteurs significatifs ou des constituants d'un mélange à analyser, on procède par extraction successive des variables les plus pures jusqu'à ce que le spectre de la pureté ne représente que le bruit de fond. Cette Figure 10 montre l'aspect du spectre de la pureté après calcul de la première pureté de chaque variable colonne  $j$ .

A partir d'une matrice de données  $D(2500 \times 460)$ , nous avons extrait 6 variables pures correspondant à 6 composés connus du mélange. La Figure 11 visualise les différents aspects pris par le spectre de la pureté après la première et la sixième extraction de variable pure. On observe qu'après la sélection de la 6<sup>ème</sup> variable pure, le spectre de la pureté prend un aspect

assez déformé. Lorsque cet aspect est atteint, on a alors une estimation du nombre de facteurs significatifs, ou de constituants, qui est égal au nombre de variables pures sélectionnées.

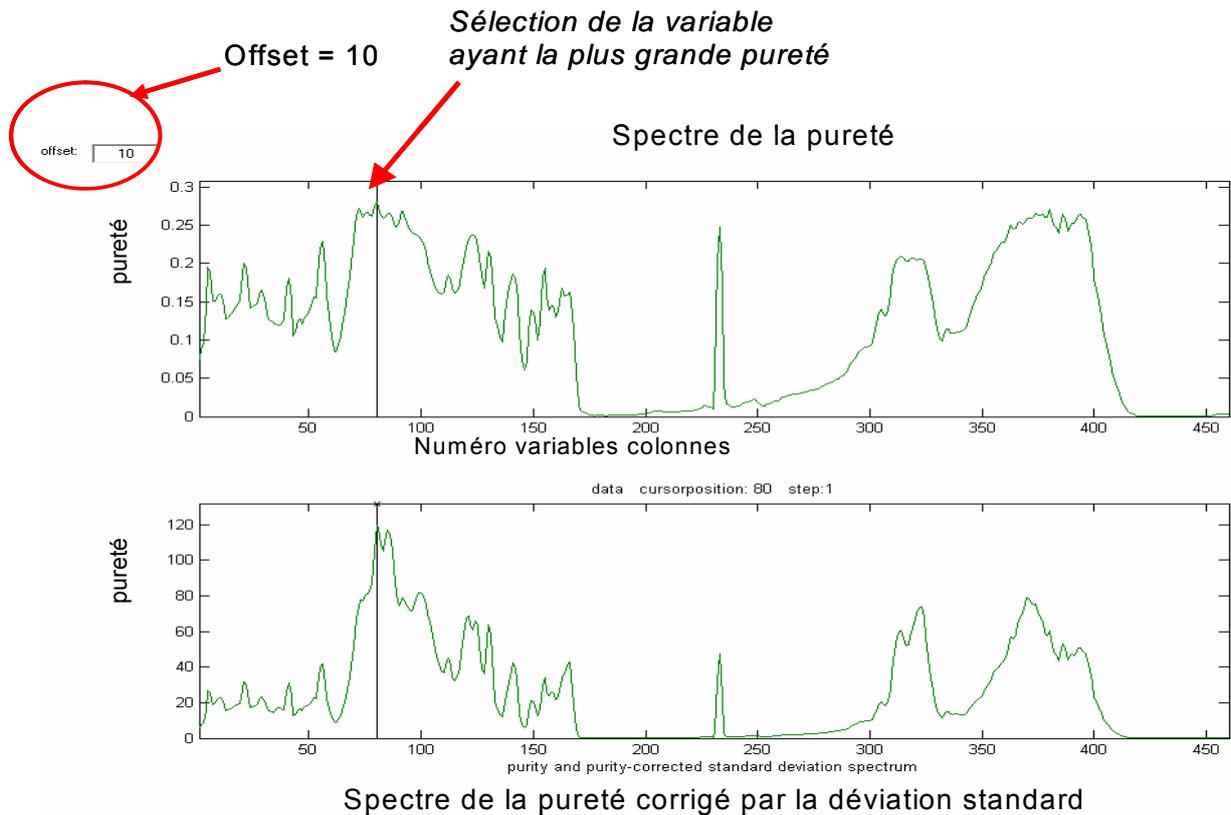


Figure 10 : Interface utilisateur produit par SIMPLISMA pour sélectionner les variables les plus pures. En haut, le spectre de la pureté produit après le calcul de la 1<sup>ère</sup> pureté pour chaque variable colonne d'une matrice de données D(2500×460). En bas le même spectre corrigé par la déviation standard. L'offset choisi dans ce cas est égal à 10. La position du curseur indique la variable, ici 80, ayant la plus grande pureté.

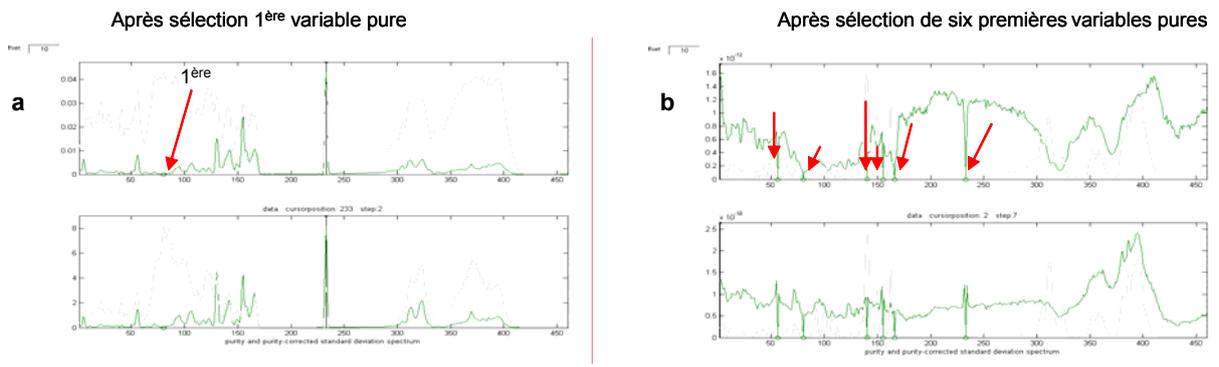


Figure 11 : Sélection des variables pures successives pour une matrice de données D(2500×460) par la procédure interactive SIMPLISMA. Après sélection de chaque variable pure, un nouveau spectre de pureté est produit. La sélection s'arrête lorsque le spectre de pureté prend un aspect assez aléatoire comme on peut le voir sur la partie b.

On note aussi sur ce graphique que l'absorbance devient de plus en plus faible après chaque extraction.

- **Estimation des spectres associés aux variables pures sélectionnées.**

Toutes les variables sélectionnées sont regroupées dans la matrice C et constituent la matrice des concentrations estimées initialement. Pour chaque variable pure sélectionnée, on récupère les concentrations correspondantes à partir de la matrice de données D.

La deuxième étape consiste à résoudre l'équation  $D = C * S^t$  pour la recherche des Spectres purs correspondant aux différents constituants du mélange étudié. La matrice des spectres purs  $S^t$  est obtenue donc par :  $S^t = (C^t * C)^{-1} * C^t * D$  (éq.24). Cette dernière formulation tient compte que les matrices C et D peuvent être non carrées et donc non inversibles directement.

### 3.2. OPA

OPA est d'abord un algorithme de sélection et de recherche du nombre de constituants d'un mélange complexe. En 1996, F. Cuesta Sanchez<sup>29</sup> et ses coauteurs présentent cet algorithme qui est l'aboutissement de plusieurs travaux<sup>36,37</sup> qu'ils ont menés auparavant. Ils ont démontré qu'on peut extraire les purs constituants des données issues de la chromatographie liquide (LC-DAD ou HPLC-DAD) par le traitement par OPA.

Les principales hypothèses sous-jacentes sont la bilinéarité de la matrice du mélange étudié et que la condition de la loi de Beer-Lambert soit remplie. Lorsque ces conditions sont vérifiées, on peut décomposer la matrice de données D en produit d'une matrice des concentrations C et d'une matrice S des coefficients d'extinction molaires (spectres).

OPA procède par résolution par étapes successives dans une approche d'orthogonalisation par comparaison de chaque spectre mesuré à un spectre pris comme référence. Cette orthogonalisation est basée sur celle de Gram-Schmidt<sup>38</sup>.

D'une manière générale on choisit un spectre de référence ( $s_{ref}$ ) qui est dans la plupart des cas le spectre moyen. Mais on peut choisir aussi, comme référence, celui qui a la plus grande norme (notamment dans le domaine de la chromatographie liquide).

Après avoir normalisé le spectre de référence  $s_{ref}$  et tous les autres spectres à 1, on projette chacun de spectres sur le spectre de référence. En s'aidant de l'orthogonalisation de Gram-Schmidt, on calcule le vecteur spectre  $d_i$  orthogonal au vecteur spectre référence correspondant à chacun des spectres d'origine  $s_i$  de la matrice de données  $D(r \times c)$ . Prenons par exemple le cas général où  $s_{ref}$  est le spectre moyen.  $d_i$  est calculé alors de la manière suivante :

$$d_i = s_i - (s_i^t * s_{ref}) * s_{ref} \quad (\text{éq.25}).$$

Ensuite la norme euclidienne  $\|d_i\|$  de chaque spectre  $d_i$  est calculée. Cette norme est ce qu'on appelle la « dissimilarité ». Ayant calculé une valeur de dissimilarité pour chaque spectre  $d_i$  de  $D$ , on trace le graphe de la dissimilarité en fonction du paramètre de l'étude (temps de l'étude, du numéro de spectres ou des variables colonnes, ...etc.). Ce graphe nous permettra de sélectionner le spectre, la variable, le temps, (...etc.) ayant la plus grande valeur de dissimilarité. Par exemple, le spectre sélectionné est le plus dissimilaire du spectre moyen et sera considéré comme le premier spectre pur. Il en sera de même pour la variable, le temps, ...

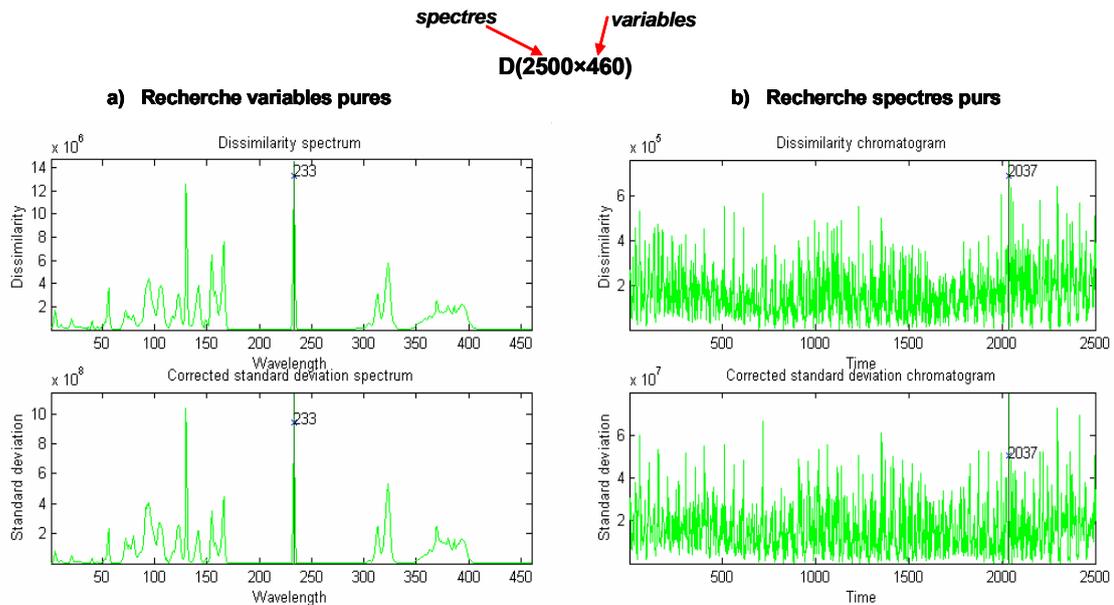


Figure 12 : Interface utilisateur fourni par OPA après application à une matrice de données  $D(2500 \times 460)$ , a) pour la sélection des variables pures, b) pour la sélection des spectres purs.

OPA propose deux méthodologies de recherche : soit une recherche des variables les plus pures suivant les colonnes de la matrice  $D$ , soit une recherche des spectres enregistrés les plus purs suivant les lignes de  $D$ . Dans les deux cas, le graphe produit va être en fonction de la méthodologie de recherche choisie. La Figure 12 montre la première interface obtenue pour chacune de deux méthodologies. Il faut noter que dans le cas de recherche des variables pures, on obtient le spectre de dissimilarité qui est en fonction des longueurs d'onde et dans l'autre cas le « chromatogramme de dissimilarité » qui est en fonction du numéro d'ordre des spectres.

Pour extraire les autres variables ou spectres les plus purs, comme pour SIMPLISMA on va utiliser une procédure itérative de calcul des dissimilarités pour les variables ou spectres restants. Puis on va extraire la contribution de chaque variable ou spectre extrait à la matrice de données par le biais de corrélation. Cela aboutira, par exemple, à un spectre de dissimilarité ayant un aspect aléatoire lorsque le nombre correct de variables est atteint. Cette procédure itérative est expliquée plus en détail dans l'algorithmique donnée en annexe 1.3.

Sur la Figure 13, on peut observer la procédure de sélection des variables les plus pures dont la démarche est semblable à celle de SIMPLISMA. La seule différence est le calcul de la dissimilarité d'une manière différente. Et sur la partie Figure 13c, on peut voir l'aspect totalement déformé du spectre de dissimilarité lorsque le nombre de composés présents dans le mélange est atteint (dans notre cas, 6 composés). OPA est donc une méthode approchée d'estimation du nombre de composés présents dans un mélange donné. Le nombre de facteurs peut être estimé à 1 ou 2 facteurs près au plus.

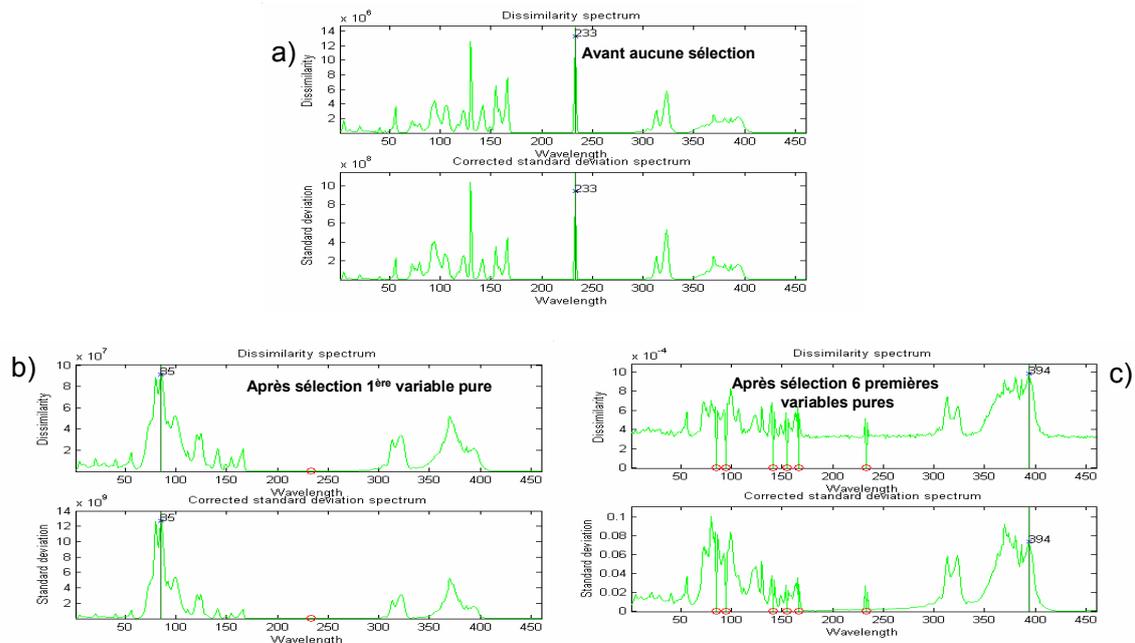


Figure 13 : Procédure de sélection successive des variables pures (concentrations) par OPA; a) 1<sup>ère</sup> interface produite par OPA pour la sélection des variables pures avant qu'aucune sélection n'est effectuée ; b) nouveau spectre de dissimilarité recalculé après sélection de la 1<sup>ère</sup> variable la plus pure ; c) aspect du spectre de dissimilarité après avoir sélectionné les 6 premières variables les plus pures pour un mélange constitué de 6 composés purs. A noter l'aspect assez déformé de ce spectre de dissimilarité après la 6<sup>ème</sup> sélection.

- **Extraction des variables pures et des spectres purs.**

L'extraction de l'ensemble des variables caractéristiques, par la méthodologie itérative d'OPA, donne l'estimation de la matrice des concentrations  $C$ . Une variable pure sélectionnée est représentative d'un constituant du mélange. L'utilisation appropriée de la procédure des moindres carrés alternés (ALS) produit les spectres de différents composés du mélange regroupés dans la matrice  $S^t$ .

En revanche, la procédure d'extraction ciblant les spectres purs identifie parmi les spectres mesurés du mélange, s'il en existe certains qui peuvent être caractéristiques d'un constituant du mélange. Après avoir identifié et extrait l'ensemble des spectres spécifiques, on les regroupe dans la matrice d'estimation  $S^t$ . Par la suite l'extraction des concentrations  $C$  est réalisée par ALS (cf. Annexe 1.3)

## 4. D'autres méthodes

Il existe beaucoup d'autres méthodes pour obtenir une première estimation des concentrations et des spectres des constituants d'un mélange. Les énumérer toutes est fastidieux et hors de notre cadre d'étude. Nous ne citerons que quelques unes que nous allons utiliser dans notre étude. Les deux méthodologies IKSFA (Iterative Key Set Factor Analysis) et SPEXFA (Spectral Isolation Factor Analysis) basées sur l'analyse factorielle permettent d'extraire aussi les spectres et les concentrations des composés d'un mélange. Ces méthodes, dont l'auteur principal est Malinowski<sup>39,40,41</sup>, ciblent des caractéristiques particulières des données recueillies pour extraire les informations utiles.

Nous verrons aussi d'autres méthodes comme PMF et MCR-ALS qui sont des méthodologies de résolution assez évoluées. Nous reviendrons sur celles-ci dans la partie suivante consacrée à l'optimisation des résolutions.

## IV. L'optimisation et l'introduction des contraintes dans les méthodologies d'extraction pour une meilleure résolution

Nous venons de passer en revue quelques méthodes qui ont servi à produire les premières estimations des spectres et des concentrations des différents composés présents dans un mélange étudié. Certaines hypothèses, plus ou moins évidentes, sous-tendent ces méthodologies.

### 1. Hypothèses sous-jacentes aux méthodologies d'extraction simples

PCA est basée sur la contrainte mathématique d'orthogonalité deux à deux des *composantes principales* extraites. Ceci est une propriété mathématique importante pour le calcul mais qu'en est-il de la signification physique vis-à-vis des spectres des constituants ou de leurs concentrations à estimer ? Chaque spectre extrait correspond-il à un constituant ? Les valeurs trouvées pour les concentrations sont-elles toutes valides pour l'appréciation des espèces ? Nous avons montré que l'extraction par PCA donnait une première idée des constituants. Mais ce qui ressort, c'est qu'une quantification de la concentration des espèces avec les *Scores* ou une interprétation approfondie avec les *loadings* extraits n'est pas satisfaisante. Au chapitre 3, nous proposons une discussion sur les résultats extraits par PCA sur un exemple concret.

Quant à SIMPLISMA, toute la méthodologie d'extraction est basée sur la pureté des variables. Mais aucune contrainte sur les spectres ni sur les concentrations à extraire n'est imposée. Ce qui peut aboutir à des concentrations ou des spectres négatifs. Cela contredit la principale propriété de ces deux grandeurs.

Pour OPA, le critère de base est l'orthogonalisation de Gram-Schmidt. Le critère de dissimilarité, stipulant que le spectre à extraire est le plus dissimilaire d'un spectre pris comme référence, ne garantit pas forcément l'orthogonalité des deux. Même si c'était le cas, cela leur confère-t-il un sens physique ? Cependant ce qu'il faut ajouter, ce que lors d'utilisation de la procédure ALS, le critère de non-négativité des spectres et des concentrations est utilisé. OPA est donc en réalité associé à ALS pour l'extraction des spectres lorsque les variables sont estimées ou vice versa. Nous traiterons ALS et ses dérivées dans le paragraphe 3.1 suivant.

IKSFA et SPEXFA utilisent des caractéristiques locales de certaines régions spectrales. SPEXFA, par exemple, isole les principales composantes spectrales caractéristiques des constituants d'un mélange lorsqu'il existe une gamme spectrale spécifique pour chaque constituant. Pour IKSFA, il faut qu'il y ait au moins un point de mesure considéré pur pour chaque spectre. La procédure itérative tente d'extraire, à partir de la matrice de données  $D$ , les lignes ou les colonnes les plus orthogonales sensées représenter des caractéristiques différentes

associées aux composés recherchés. Mais la résolution se fait sans d'autres contraintes sur les concentrations ou les spectres extraits, ce qui aboutit parfois à des aberrations et à une véritable difficulté d'interprétation !

Toutes ces remarques ont amené les physico-chimistes, et en particuliers les chimiométriciens, à injecter des contraintes à ces méthodes ou à inventer de nouvelles méthodologies de résolution intégrant plus de contraintes spécifiques. Et cela dans le but d'extraire des profils réalistes et conformes aux types d'expériences en usage.

## 2. Les nouvelles exigences et les contraintes associées

Depuis les années 1970, les données à étudier ont considérablement évolué. Leur nature a varié. Leur structure a constamment changé : de la structure plane, on est passé à des matrices multidimensionnelles.

Pour s'adapter à tous ces changements, les méthodes de résolution ont évolué aussi. La nécessité à mener à la création des nouvelles au fil du temps. Certaines existantes sont améliorées pour élargir leur champ d'application et intégrer plus des paramètres ou des contraintes. A la statistique analytique du début du siècle, se sont substituées des méthodes assez robustes d'extraction comme MCR-ALS ou de quantification comme PLS et ses dérivées. Ces méthodologies sont l'aboutissement d'une laborieuse évolution passant par la méthode de Tucker<sup>42</sup>, PARAFAC<sup>43</sup>, PCA, « target analysis » comme ITTFA<sup>44,45</sup>, les méthodes itératives comme IKSFA, les méthodologies d'extraction simple comme SIMPLISMA, OPA, ...etc. Les travaux de Riter<sup>34</sup> et *al.*, de Rasmussen<sup>46</sup> et *al.*, et d'Antoon<sup>47</sup> et *al.* sont une bonne indication de cette évolution. Des micro-ordinateurs de plus en plus puissants ont permis ce progrès rapide. L'acquisition des données de plus en plus complexes est devenue courante : d'où à la fois l'abondance des informations à extraire et certaines difficultés d'analyse.

Cependant un paramètre très important pour la physico-chimie est la non-négativité des spectres et des concentrations. Une bonne méthode d'extraction doit tenir compte de cette réalité physique. On doit avoir aussi la possibilité d'inclure des contraintes liées à la connaissance des propriétés locales sur les concentrations ou les spectres. La structure des données et le type de spectroscopie peuvent introduire des contraintes dans la méthodologie de résolution.

De ce fait, la méthodologie du « hard-modeling » s'appuie sur des contraintes tenant compte des conditions physico-chimiques en vue de respecter des modèles ou des conditions postulées. Cependant dans la plupart des cas, il est difficile de postuler une méthode concrète avec tous les paramètres bien connus et définis. Cette dernière contrainte limite l'utilisation de cette méthodologie à des cas contrôlables. A l'inverse, le « soft-modeling »<sup>48, 49</sup> nécessite peu d'hypothèses contraignantes et les principales méthodes de résolution appartiennent à ce

domaine. La combinaison<sup>50, 51</sup> de ces deux types de méthodologies peut donner une résolution robuste.

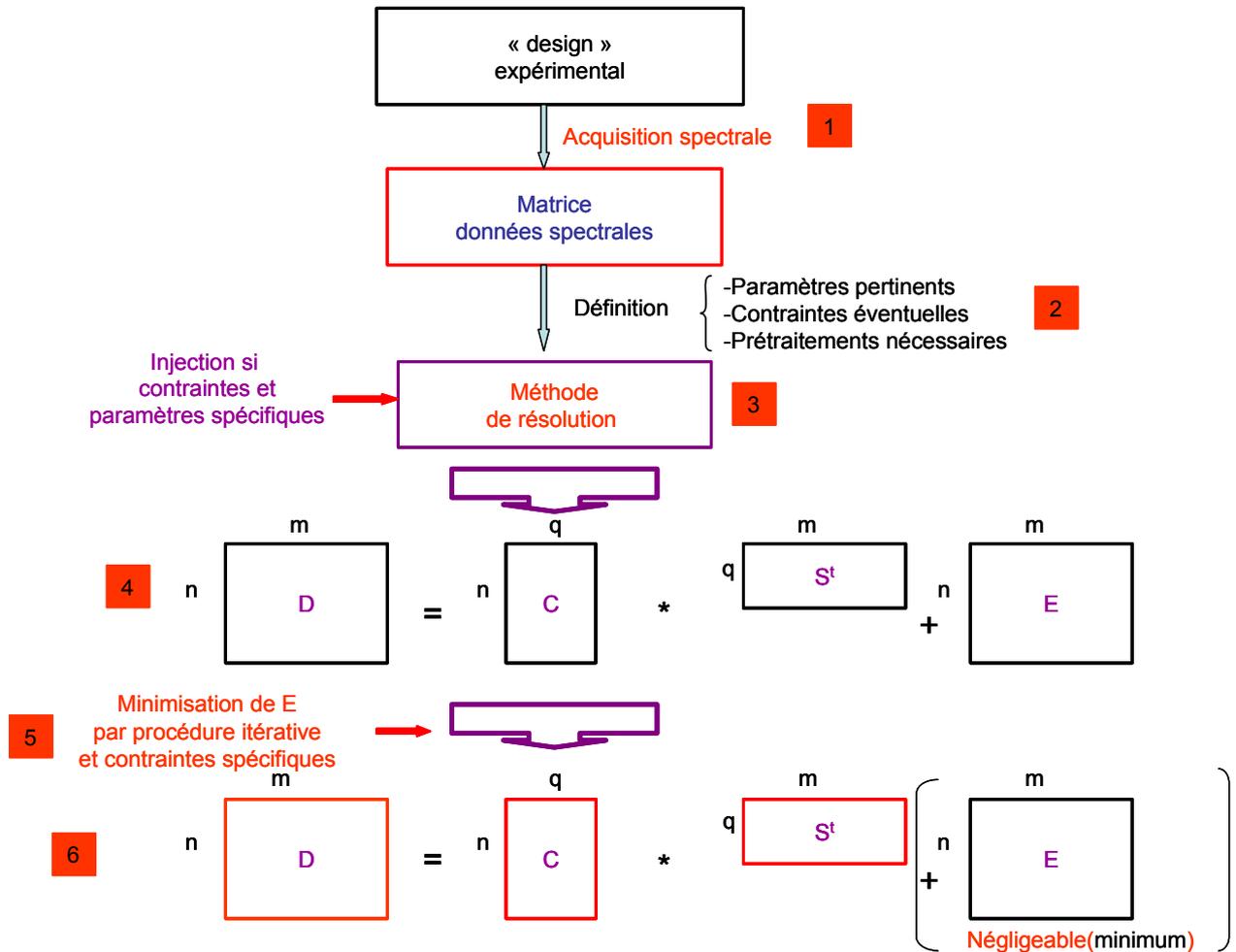


Figure 14 : Schéma général de résolution de données. L'optimisation des résultats est la minimisation des résidus par des procédures itératives avec injection des contraintes spécifiques.

Quant à l'imagerie, avec les récentes technologies (caméra CCD, FPA, IRFPA, barrette Hyperspectrale, ...), la résolution d'images nécessite des nouvelles contraintes à respecter (répartition spatiale, propriété de voisinage, problème de bord, qualité de reproduction, ...). Des nombreux algorithmes sont développés à cette fin.

En tout cas, dans presque tous les domaines, la stratégie tend vers des méthodes de résolution puissantes intégrant à chaque étape des nombreuses contraintes comme ALS ou MCR-ALS. La Figure 14 rappelle le schéma général de résolution avec l'optimisation possible par injection des contraintes adaptées.

Nous présentons certaines de ces méthodologies et leurs principes de base dans la partie suivante.

### 3. Quelques méthodologies d'extraction plus robustes

Les principales méthodes de résolutions pour optimiser l'extraction sont basées sur la procédure des moindres carrés alternés (ALS).

#### 3.1. La procédure des moindres carrés Alternés (ALS : Alternating Least Squares)

ALS est basé sur la méthode des moindres carrés. Avant d'approfondir celle-ci, quelques rappels sur la méthode des moindres carrés sont présentés.

- **Les moindres carrés**

La procédure des moindres carrés a une longue histoire. Les premiers travaux significatifs cités dans la littérature sont ceux de Gauss et de Legendre dans la détermination de trajectoires des comètes, des astéroïdes ou des planètes dans leur révolution autour du soleil.

La méthode des moindres carrés sert à faire un ajustement entre deux grandeurs (ou deux ensembles de grandeurs) liées par une relation de la forme  $y = f(x, a, b_1, \dots, b_m)$  par exemple,  $a, b_1, \dots, b_m$  étant les paramètres de la fonction. Dans le cas où il existe une relation linéaire entre chacun des couples  $(y_i, x_i)$ , il s'agit de déterminer la droite d'ajustement (droite de régression).

Cette droite est sensée représenter la vraie relation entre  $x$  et  $y$ . Cependant entre la valeur observé  $y_i$  et celle prédite par le modèle  $\hat{y}_i = ax + b$ , il existe une différence  $e_i$  qui est le résidu exprimé par la relation suivante  $e_i = y_i - \hat{y}_i = (y_i - ax_i - b)$  (éq.26). La variable  $x_i$  est appelée « variable indépendante ». L'erreur commise sur la mesure de  $x_i$  est supposée nulle ou plus petite que l'erreur expérimentale.  $y_i$  est « la variable dépendante » avec une variance  $\sigma^2$ . Pour pouvoir effectuer le calcul, un certain nombre d'hypothèses sont faites sur les résidus  $e_i$  telles que leur indépendance, une distribution aléatoire normale, une même variance  $\sigma^2$  et la propriété de l'homoscédasticité (indépendance par rapport à  $x$ ).

La procédure des moindres carrés estime les paramètres de la régression par minimisation de la somme des carrés des résidus  $SCR$  exprimée par la relation :

$$SCR = \sum_i e_i^2 = \sum_i (y_i - ax_i - b)^2 \quad (\text{éq.27}).$$

Les dérivées partielles de  $SCR$  par rapport à  $a$  et  $b$  doivent être nulles et donnent les équations suivantes qui permettent de calculer  $a$  et  $b$  :

$$\frac{\partial(SCR)}{\partial a} = \sum_i 2(y_i - ax_i - b)(-x_i) = 0 \quad (\text{éq.28}) \quad \text{et} \quad \frac{\partial(SCR)}{\partial b} = \sum_i 2(y_i - ax_i - b)(-1) = 0 \quad (\text{éq.29}).$$

Ces équations sont appelées « équations normales ». En général, on obtient autant d'équations normales que de paramètres à estimer.

Après quelques transformations et en introduisant les expressions de  $x_i$  et  $y_i$  en fonction de leur moyennes et de leurs écarts-types,  $a$  et  $b$  sont exprimés de la manière suivante :

$$a = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})} = \frac{n \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n \sum_i x_i^2 - (\sum_i x_i)^2} \quad \text{et} \quad b = \bar{y} - a\bar{x} = \frac{\sum_i y_i \sum_i x_i^2 - \sum_i x_i \sum_i x_i y_i}{n \sum_i x_i^2 - (\sum_i x_i)^2}$$

(*éq.30*) et (*éq.31*).

Ce cas exprimé ci-dessus correspond à une régression linéaire simple. On peut avoir une régression linéaire multiple (plusieurs variables explicatives indépendantes  $x_{1i}, x_{2i}, \dots, x_{ni}$ ), ... etc. Dans chaque cas, les hypothèses et le calcul des paramètres de la régression doivent être adaptés.

Le traitement est souvent effectué sous forme matricielle :  $X * B = Y$  (*éq.32*) pour la régression

linéaire simple avec  $X = \begin{bmatrix} 1 & x_1 \\ .. & .. \\ 1 & x_n \end{bmatrix}$ ,  $B = \begin{pmatrix} b \\ a \end{pmatrix}$  et  $Y = \begin{bmatrix} y_1 \\ . \\ y_n \end{bmatrix}$ . La résolution se fait aisément en

multipliant à gauche par la matrice transposée de X, ce qui donne  $X' * X * B = X' * Y$ . La matrice  $X' * X$  est une matrice carrée inversible (si  $|X' * X| \neq 0$ ) et dans la résolution on retrouve les mêmes équations normales décrites ci-dessus. La matrice des paramètres se calcule :  $B = (X' * X)^{-1} * (X' * Y)$  (*éq.33*). Pour les autres types de régression, telle la régression linéaire multiple, ce sont les dimensions des matrices X et B qui changent compte tenu du nombre des variables explicatives retenues et des paramètres de la régression à estimer.

D'une manière générale, l'équation de Beer-Lambert généralisée sous forme matricielle peut être résolue par la procédure des moindres carrés.

Pour améliorer la résolution et affiner l'extraction simultanée des deux grandeurs liées par une relation établie, on a mis au point la procédure de la régression alternée des moindres carrés (ALS).

- **Début de l'ALS et les contraintes initiales.**

Karjalainen<sup>52</sup>, après avoir construit auparavant des méthodes mathématiques pour résoudre le recouvrement spectral des composés dans la chromatographie gaz associée à la spectrométrie de masse (GC/MS), proposa en 1989 la régression alternée (Alternating Regression AR). Il l'appliqua à la décomposition des données physico-chimiques pour extraire correctement la matrice des concentrations C et celle des spectres S<sup>t</sup> des constituants avec le minimum d'erreur représenté par la matrice des résidus E (cf. *éq.2*). Karjalainen n'a pas utilisé l'analyse factorielle dans cette décomposition. Par un tâtonnement simple avec les méthodes d'estimation, un nombre q de facteurs de décomposition est proposé et la matrice S' est remplie par des nombres aléatoires au départ. Ensuite la régression et les contraintes permettront d'obtenir les solutions correctes. L'un des problèmes rencontrés est l'ambiguïté rotationnelle car théoriquement il existe une

infinité de décomposition de D en produit de deux matrices, car  $D = C * S^t = (C * T) * (T^{-1} * S^t)$  pour toute matrice T exprimant une rotation possible. Pour surmonter ces difficultés Karjalainen proposa d'imposer deux contraintes : la non-négativité sur les concentrations et les spectres, et un seul maximum local pour la concentration de chaque espèce. Cette dernière condition est valable pour la chromatographie par exemple. La régression alternée consiste en la résolution de C et  $S^t$ . Après le choix aléatoire des valeurs dans S, soit  $S_{ini}$ , il s'agit de calculer C par la relation  $C_{cal} = (D * S_{ini}) * (S_{ini}^t * S_{ini})^{-1}$  (éq.34) et d'imposer à C les contraintes nécessaires comme la non-négativité. Ce qui donne  $C_{cor}$  la matrice des concentrations corrigées. Puis on évalue la somme des carrés des résidus  $SCR = \sum_i \sum_j e_{ij}^2$  (cf. éq.27) car  $E = D - C_{cor} * S_{cor}^t$  et on vérifie si le critère de convergence est atteint. En utilisant la matrice de concentrations calculées  $C_{cor}$ , on calcule à nouveau la matrice des spectres  $S^t$  par la relation  $S_{cal}^t = (C_{cor}^t * C_{cor})^{-1} * (C_{cor}^t * D)$  (éq.35). Ensuite il faut imposer à  $S^t$  la contrainte de non-négativité, normaliser chaque vecteur-spectre à 1 et évaluer à nouveau SCR. Cette résolution alternée en C et en  $S^t$  se poursuit en imposant à chaque résolution les contraintes adaptées jusqu'à atteindre la convergence voulue qui est obtenue d'une manière relativement rapide dans la plupart des cas. La convergence est atteinte lorsque SCR devient inférieure à un nombre  $\epsilon$  prédéfini. Une autre façon est de calculer la différence des SCR entre deux itérations consécutives et si la variation est négligeable, alors la convergence est atteinte.

### 3.2. Amélioration dans ALS et création de MCR-ALS (Multivariate Curve Resolution - Alternating Least Squares)

La régression alternée basée sur la procédure des moindres carrés et n'utilisant pas « l'analyse factorielle » était une avancée considérable. Certaines difficultés, comme *l'ambiguïté rotationnelle*, pouvaient être surmontées par l'introduction des contraintes générales ou spécifiques au problème étudié. Certaines difficultés comme « *l'ambiguïté sur l'intensité* », c'est-à-dire un problème d'échelle sur la concentration, subsistaient et ne permettaient pas une quantification exacte des espèces extraites. Les contraintes injectées étaient aussi peu nombreuses et ne permettaient pas une résolution généralisée. A partir de 1991, R. Tauler<sup>53, 54</sup> et son groupe, entre autres, ont commencé à introduire plusieurs contraintes pour une résolution plus efficace. Cette résolution étant orientée à la quantification de la concentration des espèces chimiques dans des processus réactionnels ou dans la chromatographie liquide à haute performance (HPLC). Les nombreuses contraintes introduites dans ALS, en plus de la *non-négativité* et un maximum local pour chaque espèce (*unimodalité*) introduits par Karjalainen, sont la possibilité d'utiliser la connaissance locale sur la concentration (*sélectivité*), la contrainte

sur la concentration totale ou la masse (*closure*), des contraintes sur la forme des courbes (*shape constraints*), des contraintes sur la structure des données (*structure constraints*). En plus de cette incorporation, ALS est généralisé pour s'adapter aux différents types de structures (3-D et plus). Dans cette nouvelle version, l'estimation initiale des concentrations ou des spectres est renforcée et se fait par des méthodes du « self-modeling » comme EFA, SIMPLISMA, OPA, PCA ou SVD et non aléatoirement. Cette version améliorée et modifiée de ALS est souvent appelée MCR-ALS (Multivariate Curve Resolution – Alternating Least Squares). La Figure 15 rappelle le déroulement algorithmique de MCR-ALS.

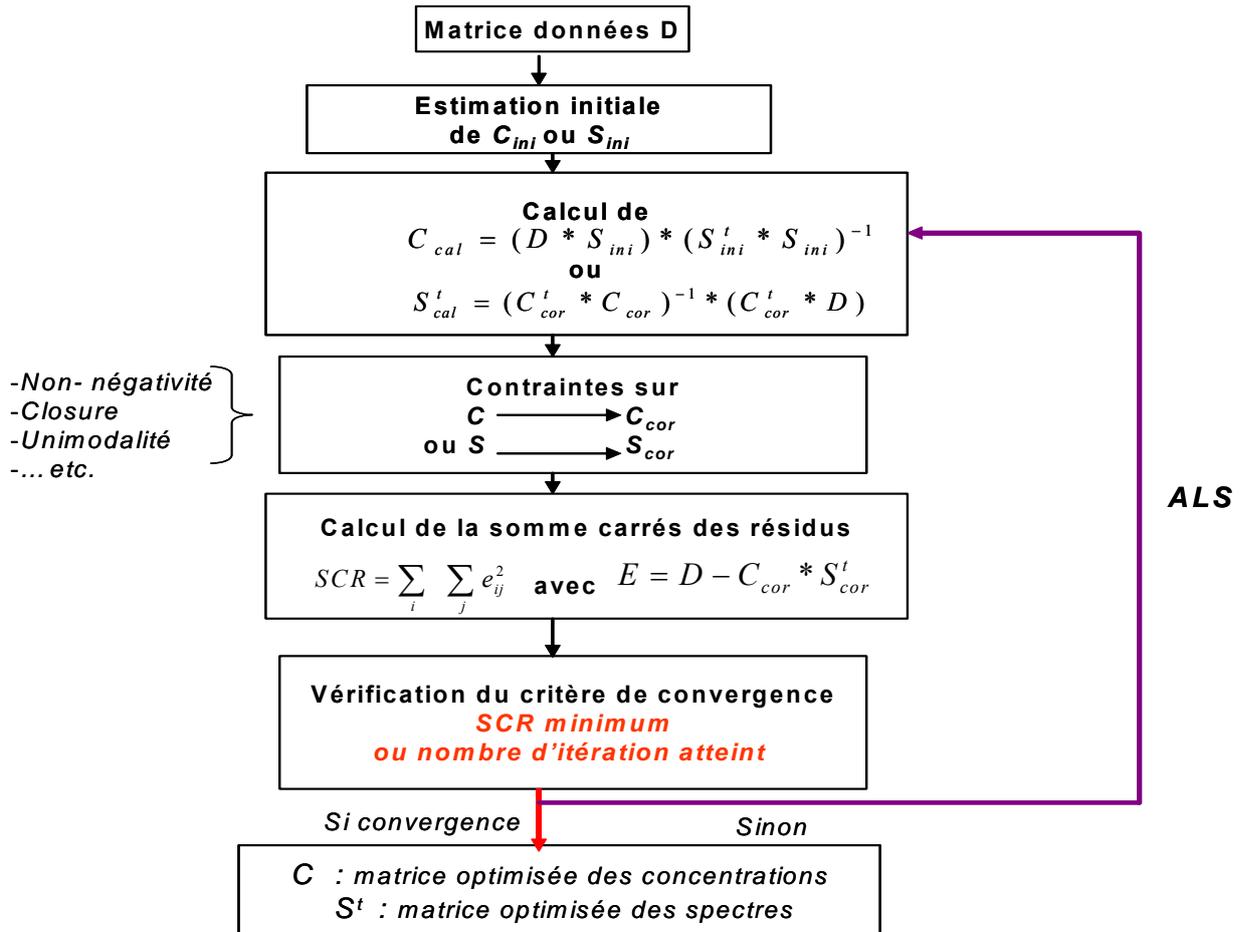


Figure 15 : Schéma du déroulement de l'algorithme MCR-ALS.

Les principales innovations introduites par MCR-ALS sont :

- ✓ la résolution alternée de deux grandeurs à optimiser après une première estimation de l'une d'elles ;
- ✓ l'injection des contraintes à chaque étape et la vérification du critère de convergence.

Lorsque la convergence est atteinte, on obtient la meilleure optimisation possible pour les deux grandeurs simultanément. MCR-ALS est la principale méthode de « résolution multivariée des courbes ». De ce fait, elle est l'une des plus utilisées. E. J. Karjalainen<sup>52</sup> a montré son intérêt dans la problématique de reconstruction des spectres purs des composés d'un mélange. R.

Tauler<sup>55,53</sup> et son groupe ont appliquée cette méthode aux traitements de données de nature diverse.

### 3.3. D'autres méthodes comme PMF (Positive Matrix Factorization)

D'autres méthodes utilisant la régression des moindres carrés alternés existent. Une que nous utiliserons dans notre étude est PMF.

PMF est un modèle bilinéaire du PARAFAC. Le modèle du PARAFAC2 proposée en 1972 par R.A. Harshman<sup>43</sup> était assez lourd à mettre en œuvre. Depuis lors, cet algorithme a subi une constante amélioration et une adaptation aux différentes données issues des domaines très variés. En 1997, P. Paatero<sup>56</sup> propose cette version bilinéaire très améliorée du PARAFAC appelée PMF. D'autres versions<sup>57</sup> étaient auparavant testées. L'auteur définit PMF comme une procédure des moindres carrés pondérés. La pondération étant basée sur la déviation standard connue des éléments de la matrice de données. C'est une variante de « l'analyse factorielle » où les facteurs sont contraints à être non négatifs. PMF permet aussi des « facteurs non-orthogonaux ». Cette méthode est assez utilisée dans le domaine environnemental, notamment dans l'étude de la pollution atmosphérique<sup>58,59</sup>.

Le modèle est défini sur la minimisation des résidus, représentés par la matrice E, par l'intermédiaire des deux facteurs G et F similaires aux scores et loadings dans PCA. On exprime

la relation comme suit :  $E_{ij} = D_{ij} - Y_{ij} = D_{ij} - \sum_{k=1}^q G_{ik} * F_{kj}$  (éq.36),  $D_{ij}$  étant les valeurs mesurées

et  $Y_{ij}$  les valeurs fournies par le modèle pour  $q$  facteurs de décomposition. Pour le calcul, on définit la «fonction objet Q», fonction des facteurs G et F, qui doit être minimisée

$Q(E) = \sum_{i=1}^r \sum_{j=1}^c (E_{ij} / \sigma_{ij})^2$  (éq.37) pour toute matrice de données  $D(r \times c)$ ,  $\sigma_{ij}$  étant la déviation

standard de la valeur mesurée  $D_{ij}$ . Tous les éléments dans G et F sont contraints à être non-négatifs. Il faut préciser cependant que les valeurs dans G et F étaient, au départ, choisies aléatoirement. Par la suite la procédure itérative ALS similaire à celle de MCR-ALS optimise les deux grandeurs.

Des procédures assez détaillées sur PMF peuvent être trouvées dans la littérature<sup>60</sup> ou dans les ressources du net<sup>61</sup>.

## 4. Contraintes sur la résolution des images

Dans le paragraphe I, nous avons abordé quelques aspects de l'imagerie spectroscopique. Pour l'extraction de la concentration de chaque espèce au niveau de chaque pixel, et pour une meilleure résolution, on a besoin d'injecter des contraintes. Certaines contraintes sont générales

comme le nombre de constituants à estimer, ou locales comme le nombre des composés présents sur un pixel donné. On peut utiliser aussi la connaissance de la concentration sur certaines parties du domaine spectral. On peut définir aussi des critères sur les propriétés «proches voisins» des pixels. Nous avons déjà décrit l'algorithme FSIW-EFA qui tient compte de cette propriété importante pour les images.

Il est possible aussi d'ajouter des contraintes tenant compte de la profondeur pour les images multicouches. La structure des données (bilinéaire, trinéaire) peut être une contrainte à intégrer lors de la résolution. M. Jiang<sup>62</sup> et al. rapportent dans leur article un bon nombre de contraintes pour la résolution des images et leur quantification. Des contraintes spécifiques sur le signal peuvent avoir une incidence très importante sur la qualité des images.

Nous verrons au chapitre 3 que les contraintes du rang local peuvent nous fournir une estimation du nombre de composés présents sur chaque pixel. On pourra intégrer ces connaissances dans MCR-ALS ou ALS pour mieux orienter la résolution et aboutir à des meilleures extractions.

Dans le chapitre 2 suivant, nous aborderons l'imagerie microspectrométrique en générale et plus particulièrement en infrarouge et en Raman après avoir exposé les bases de la spectroscopie vibrationnelle.

## Chapitre 2 :

### Imagerie microspectrométrique (infrarouge et Raman)

---

#### Spectroscopie vibrationnelle

##### I. Spectroscopie infrarouge

###### 1. Un peu d'histoire

En avril 1800, Sir W. Herschel<sup>63,64</sup> réalise des expériences de diffraction de la lumière solaire après avoir construit un spectromètre rudimentaire utilisant un thermomètre comme détecteur. Il remarque en dehors du domaine spectral visible, en dessous du rouge (infrarouge), une augmentation de la température en fonction du temps d'exposition. C'est la première constatation du rayonnement infrarouge rapportée par la littérature. Par la suite les travaux de H. Rubens, professeur à l'université de Berlin, sont assez considérables pour la caractérisation du rayonnement infrarouge et la mise au point d'instruments de détection de ce rayonnement.

L'utilisation de l'infrarouge et son développement sont intimement liés au développement instrumental<sup>65</sup>. C'est ce qui a permis de mieux détecter ce rayonnement et d'obtenir une réponse d'une manière de plus en plus rapide. Dès 1900 aux Etats-Unis, Coblentz enregistrait déjà la transmittance de très nombreuses substances chimiques et faisait des hypothèses sur les groupes fonctionnels relatifs aux structures moléculaires. Les principales innovations dans la détection et la mesure des radiations infrarouges sont apparues au vingtième siècle.

Les premiers détecteurs utilisés sont les thermopiles (Nobili<sup>66</sup>, 1829; Melloni<sup>67</sup>, 1833), les thermocouples (Seebeck<sup>68,66</sup>, 1821) et des bolomètres (Langley<sup>69</sup>, 1880). Les premières tentatives de l'étude des matériaux photoconducteurs ne datent que de 1917 (Case<sup>70</sup>, 1917). Certains détecteurs photoconducteurs (comme semi-conducteur PbS) apparaissent ensuite en 1933.

Le premier spectromètre automatisé est réalisé par E. Lehrer<sup>71</sup> en 1937.

A partir de 1946, commencent à apparaître les premiers catalogues des spectres infrarouges notamment sous l'impulsion de l'*American Petroleum Research (Programme 4)*.

En 1948, C. La Lau<sup>72</sup>, dans son article traitant de l'application analytique de la spectrophotométrie, décrit un appareil infrarouge tout à fait nouveau, *le pfund*, pour l'analyse des

gaz. Cette apparition fait suite à celle du premier spectrographe double faisceau<sup>73</sup> permettant la représentation directe de la transmittance en fonction de la longueur d'onde  $\lambda$  ( $T = f(\lambda)$ ) mis au point en 1947. Le premier domaine de l'infrarouge à trouver une application analytique dans l'analyse quantitative de certains produits est le proche infrarouge<sup>74</sup>. Dès 1955, W. Kaye<sup>75</sup> propose une synthèse des techniques et de l'instrumentation infrarouge disponibles à cette époque.

En 1959 on s'interroge déjà, pour le compte de Perkin-Elmer, sur les avancées de l'instrumentation<sup>76</sup> infrarouge. Avant 1960, les appareils disponibles sur le marché étaient essentiellement dispersifs. A cette date, apparaissent les détecteurs quantiques MCT (Mercure-Cadmium-Tellure) sachant que la technique spectroscopique utilisée était essentiellement la transmission. Les spectromètres à transformée de Fourier (cf. §I.5), suite à la mise en place de l'algorithme FFT (Fast Fourier Transform), sont mis sur le marché.

Dans les années 1970, apparaissent les détecteurs à couplage de charge (CCD) à base de silicium ou d'un alliage de platine/silicium mis au point par Boyle<sup>77</sup> et Smith.

Début des années 1980, c'est le retour en force des détecteurs MCT associés aux nouvelles technologies tels que les capteurs CCD et l'apparition des détecteurs « SPRITE » pour le système d'imagerie thermique.

Dans les années 1990<sup>78</sup>, il y a un progrès considérable dans l'imagerie infrarouge et apparition des FPA<sup>79</sup> (Focal Plane Array) très sophistiqués tels que les FPA à bolomètres ou FPA pyroélectriques.

Actuellement des versions de plus en plus sophistiquées des détecteurs<sup>80,81</sup>, des capteurs et des spectromètres continuent à apparaître.

## 2. Domaine infrarouge et interaction

L'infrarouge est un domaine de la spectroscopie qui a pris de plus en plus d'importance depuis les années 1970.

Le domaine infrarouge du spectre correspond aux radiations de nombres d'onde compris entre environ 12800 et 10  $\text{cm}^{-1}$ , ou des longueurs d'onde comprises entre 0,78 et 1000  $\mu\text{m}$ . Ce domaine infrarouge est partagé en trois sous domaines aux caractéristiques différentes : le proche (0,7 à 2,5  $\mu\text{m}$ ), moyen (2,5 à 30  $\mu\text{m}$ , domaine exploité 2,5 à 25  $\mu\text{m}$ ) et lointain (30 à 1000  $\mu\text{m}$ ) infrarouge.

Dans l'infrarouge les molécules peuvent absorber, émettre ou réémettre, réfléchir les radiations infrarouges. L'interaction d'une molécule avec le rayonnement infrarouge induit une transition entre les niveaux de vibration ou de rotation.

S'il y a une différence d'électronégativité entre les atomes d'une liaison, cela crée un moment dipolaire permanent qui peut interagir avec l'onde incidente. Les molécules diatomiques homonucléaires ( $H_2$ ,  $N_2$ ,  $O_2$ , ...) n'ayant pas un moment dipolaire permanent n'absorbent pas la radiation infrarouge.

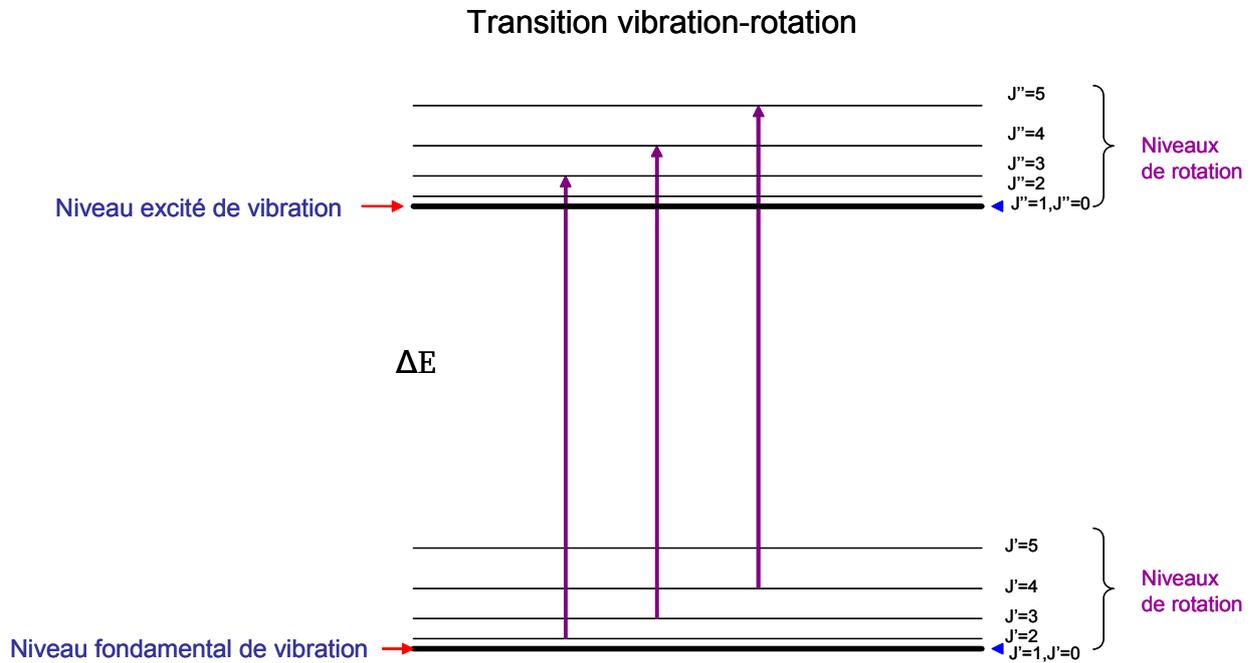


Figure 16 : Visualisation schématique des transitions entre les niveaux de vibration et rotation.

Si la fréquence naturelle de vibration de la molécule est égale à la fréquence de l'onde incidente alors il y a absorption d'énergie et une modification de l'amplitude de la vibration moléculaire sans modification de la fréquence de vibration du dipôle moléculaire. On dit alors qu'il y a *résonance*.

L'énergie entre deux niveaux de vibration est quantifiée. Chaque niveau de vibration possède plusieurs sous-niveaux de rotation. La Figure 16 montre schématiquement la transition entre les niveaux de transition et de rotation pour un gaz. Pour les liquides et les solides les sous-niveaux de rotation ne se manifestent pas, on a seulement accès aux différents niveaux de vibration définis par le nombre quantique  $V$ .

Les molécules vibrent de différentes façons. Les deux modes les plus souvent cités sont :

- ✓ Elongation : une variation de la longueur de liaison interatomique.
- ✓ Déformation angulaire : une modification d'angle entre deux liaisons. Ces déformations sont de quatre sortes : le cisaillement (*déplacement en sens inverse de deux liaisons dans le même plan*), la rotation (*déplacement dans le même sens de deux liaisons dans le même plan*), le balancement (*sortie symétrique du plan de 3 atomes dont 2 en opposition par rapport au 3<sup>ème</sup>*) et la torsion (*sortie symétrique du plan de deux atomes*).

Pour les molécules polyatomiques, un couplage de différentes vibrations est possible.

### 3. Vibration et modélisation

Le nombre de vibrations possibles pour une molécule est fonction du nombre d'atomes qui la constituent. Une molécule non linéaire constituée par N atomes possède  $3N-6$  modes de vibration alors que celle linéaire en possède  $3N-5$ .

La vibration d'une liaison simple entre deux atomes est modélisée par l'oscillateur harmonique (Cf. Figure 17). La force  $f$  de cet oscillateur est exprimée en fonction de la constante de force de rappel  $k$  :

$f = -ky$  (éq.38), l'origine étant prise à la position d'équilibre. L'énergie d'un oscillateur harmonique s'exprime  $E = (1/2) * ky^2$  (éq.39) et la fréquence de vibration pour une

seule masse  $\nu_m = \frac{1}{2\pi} \sqrt{\frac{k}{m}}$ . La fonction  $y = A \cos(2\pi\nu_m t)$  (éq.40) exprime le déplacement de

l'oscillateur par rapport à sa position d'équilibre en fonction du temps  $t$ , A étant l'amplitude de l'oscillation. Pour deux masses  $m_1$  et  $m_2$  différentes, c'est la masse réduite  $\mu$  qui intervient à la

place de  $m$  et on a  $\nu_{m_1, m_2} = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} = \frac{1}{2\pi} \sqrt{\frac{k(m_1 + m_2)}{m_1 * m_2}}$  (éq.41).

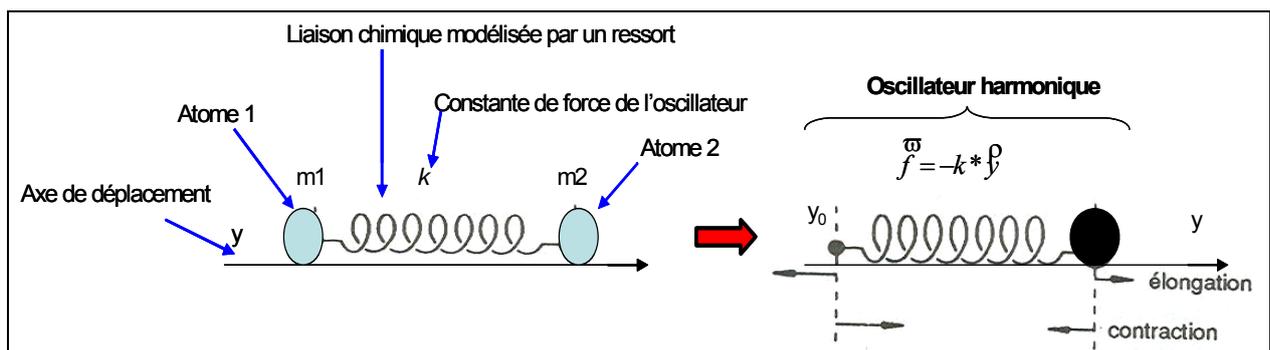


Figure 17 : Modélisation d'une liaison chimique entre deux atomes et la représentation de cette dernière par un oscillateur harmonique.

Dans le traitement quantique de l'oscillateur harmonique, l'énergie peut être exprimée en

fonction du nombre quantique de vibration  $V$  :  $E = (V + \frac{1}{2}) * \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} = (V + \frac{1}{2}) * h\nu_\mu$  (éq.42). Et

pour l'énergie de transition entre deux niveaux de vibration, on obtient :  $\Delta E = h\nu_\mu = \frac{h}{2\pi} \sqrt{\frac{k}{\mu}}$

(éq.43),  $h$  étant la constante de Planck. D'après la règle de sélection, une transition n'est possible que si et seulement si  $\Delta V = \pm 1$ .

Pour le cas polyatomique, chaque liaison entre deux atomes est représentée par un oscillateur harmonique. Soit par exemple une molécule ayant 3 vibrations fondamentales de nombre quantique respectif  $V_1$ ,  $V_2$  et  $V_3$  et de fréquence respective  $\nu_1$ ,  $\nu_2$  et  $\nu_3$ , l'énergie de vibration  $E_V$

s'exprime :  $E_V(V_1, V_2, V_3) = h\nu_1(V_1 + 1/2) + h\nu_2(V_2 + 1/2) + h\nu_3(V_3 + 1/2)$  (éq.44). Cela donne en nombre d'onde  $\bar{\nu}$  :  $G(V_1, V_2, V_3) = \bar{\nu}_1(V_1 + 1/2) + \bar{\nu}_2(V_2 + 1/2) + \bar{\nu}_3(V_3 + 1/2)$  (éq.45).

Ce modèle harmonique est une bonne approximation lorsqu'on est proche de la position d'équilibre. Lorsque les atomes sont assujettis à une vibration de plus grande amplitude qui les éloigne de cette position d'équilibre, les approximations ne sont plus valables. Le modèle qui exprime mieux, et d'une manière plus réaliste, la vibration moléculaire est l'oscillateur anharmonique. L'anharmonicité peut révéler ou prendre en compte le couplage de différents types de vibration.

Pour un oscillateur anharmonique simple, l'énergie potentielle s'exprime ainsi :  $E = (1/2)ky^2 + k'y^3 + \dots$  (éq.46) avec  $k' \ll k$ . L'énergie de vibration quantifiée  $E_V$  peut s'écrire alors en fonction de la fréquence de vibration, après application du modèle de Morse<sup>82,83</sup> dans l'équation de Schrödinger indépendante du temps,  $E_V = (V + 1/2)h\nu_m - (V + 1/2)^2 h\nu_m \chi_e$  (éq.46bis) avec  $\chi_e$  la constante d'anharmonicité. Entre deux niveaux de vibration, la règle de sélection de  $\Delta V = \pm 1$  est valable mais cette fois on peut avoir aussi les cas  $\Delta V = \pm 2, \pm 3, \dots$

Pour le cas polyatomique, l'énergie tient compte en plus des termes harmoniques les termes de couplages deux à deux de différentes vibrations. Ce qui donne en fonction du nombre d'onde pour trois vibrations<sup>84</sup> fondamentales de nombres quantiques de vibration  $V_1, V_2, V_3$  (éq.47) :

$$G(V_1, V_2, V_3) = \bar{\nu}_1(V_1 + 1/2) + \bar{\nu}_2(V_2 + 1/2) + \bar{\nu}_3(V_3 + 1/2) + \chi_{11}(V_1 + 1/2)^2 + \chi_{22}(V_2 + 1/2)^2 + \chi_{33}(V_3 + 1/2)^2 + \chi_{12}(V_1 + 1/2)(V_2 + 1/2) + \chi_{23}(V_2 + 1/2)(V_3 + 1/2) + \chi_{13}(V_1 + 1/2)(V_3 + 1/2)$$

Le terme  $\chi_{ij}$  représente la constante d'anharmonicité correspondant au couplage des vibrations  $i$  et  $j$ .

Compte tenu de la constante de couplage  $C$  et de celle de l'anharmonicité  $\chi$ , en évaluant le rapport ( $r = \frac{C}{\chi}$ ) entre ces deux constantes, on peut estimer quel est le mode de vibration prédominant. Si  $r$  est élevé, ce sont les modes normaux de vibration qui dominent. Si par contre  $r$  est faible ou inférieur à 1, alors c'est le modèle des modes locaux qui est le mieux représentatif. Ce dernier cas est utilisé dans le cadre des hautes énergies.

### Fréquence de vibration et identification moléculaire.

En infrarouge chaque liaison entre deux atomes a une fréquence de vibration spécifique. Cette fréquence caractéristique permet d'identifier chaque type de liaison dans une molécule complexe et permet *a priori* de détecter la présence de tel ou tel atome dans la molécule. Par exemple les liaisons  $C-O$ ,  $C=O$ ,  $C-N$ ,  $C=N$ ,  $C=C$ ,  $C \equiv C$ , (...etc.), ont des fréquences de vibration différentes qui leur sont spécifiques. Cependant certains déplacements des bandes

peuvent être observés selon l'environnement atomique de l'atome considéré et le groupe fonctionnel. L'identification moléculaire en infrarouge se fait dans le moyen infrarouge. Dans le proche infrarouge, on se sert de certaines bandes caractéristiques pour faire une quantification de certains types de produits<sup>85,82</sup>, (sucre, lait, matière grasse...etc.). Certains<sup>86</sup> ont démontré la potentialité de quantification du moyen infrarouge.

## 4. Les sources et les transducteurs en infrarouge<sup>87</sup>

### 4.1. Les sources

Les sources en infrarouge sont, le plus souvent, des sources de rayonnement continu utilisant l'émission du corps noir. Certaines sources discontinues (laser CO<sub>2</sub> par exemple) sont disponibles aussi. Pour les sources continues, un solide inerte est porté à une température comprise entre 1500K et 2200K par chauffage électrique. Cependant dans tout le domaine spectral, l'intensité n'est pas constante. Ce qui va nécessiter une bonne sélection du domaine à utiliser ou un ajustement instrumental.

Les principales sources utilisées en infrarouge sont les suivantes avec quelques informations supplémentaires pour certaines d'entre elles.

- **La lampe de Nernst** : constituée par un cylindre d'oxydes de terres rares de hauteur 20 mm avec un diamètre de 1 à 2 mm. Ces oxydes de terres rares sont le plus souvent mélangés avec l'oxyde de zirconium. Le contact électrique est assuré par des fils de platine. Il y a nécessité de protéger le circuit d'alimentation par un dispositif limitant le courant en vue de protéger la lampe de la surchauffe.
- **La lampe Globar** : c'est un barreau de carbure de silicium de 50 mm de long et de 5 mm de diamètre chauffé électriquement jusqu'à 1300 à 1500K. Le principal avantage est, qu'on a avec ce système, un coefficient de température positif (augmentation de la résistance du barreau lorsque la température augmente). L'inconvénient majeur est qu'il faut refroidir les contacts par circulation d'eau ou d'air.
- **Les lampes à filament incandescent (enveloppe verre ou quartz)**:
  - filament nichrome : roulé en spirale chauffé par un courant électrique jusqu'à 1100K. Le principal avantage est une durée de vie plus longue. L'inconvénient majeur est une intensité plus faible que celle des lampes Globar et Nernst.
  - Filament de tungstène : utilisée pour le proche infrarouge (NIR ou PIR), domaine 4000 à 12800cm<sup>-1</sup> (2.5 à 0.78µm.)
- **L'arc à mercure** (pour l'infrarouge lointain ( $\lambda > 50\mu m$ )) : tube en quart rempli de vapeur de mercure à une pression supérieure à 1 atm ( $\approx 1$  bar). Le passage du courant à travers la vapeur provoque la formation d'un plasma qui émet un rayonnement continu

dans l'infrarouge lointain. L'avantage est que l'intensité de cette source est plus importante que celle des autres sources citées ci-dessus.

- **Le laser à dioxyde de carbone** : c'est un laser accordable utilisé pour le contrôle des polluants atmosphériques et l'analyse des espèces absorbantes en solution aqueuse. Son domaine d'utilisation est entre 900 et 1100  $\text{cm}^{-1}$  et selon l'échantillon à analyser, on peut sélectionner une raie convenable parmi l'ensemble des raies disponibles. La puissance du rayonnement est beaucoup plus importante que celle fournie par les autres sources.

#### 4.2. Les transducteurs : état de l'art des détecteurs infrarouges<sup>78 65 80</sup>.

Disposer des sources adaptées pour obtenir un rayonnement infrarouge continu dans un domaine spectral donné est une bonne chose. Mais arriver à détecter ce rayonnement infrarouge d'une manière efficace en est une autre. Pour une détection efficace, il faut disposer des matériaux bien adaptés pouvant fournir des réponses adéquates aux stimulations infrarouges.

Ces matériaux spécifiques sont appelés « transducteurs ». En infrarouge, les transducteurs sont divisés en deux grands groupes qui se subdivisent eux-mêmes en plusieurs sous groupes :

- **Les détecteurs thermiques (transducteurs de chaleur)** qui regroupent les thermocouples (jonction Peltier, ...), les thermopiles (série de thermocouples), les bolomètres, les cellules pneumatiques (Golay cell), les pyroélectriques (TGS (Triglycine Sulfate), DTGS, des gaufrettes cristallines ...etc.). Ces détecteurs thermiques sont les premiers transducteurs qui ont été utilisés.
- **Les détecteurs quantiques (transducteurs à photo-induction)** qui regroupent de très nombreux types de détecteurs utilisés. On les classe en deux catégories : ceux à détection **intrinsèque** et ceux à détection **extrinsèque**. Les *détecteurs quantiques à détection intrinsèque* comprennent les photoconducteurs (PbS, InSb, HgCdTe, ...) et les photovoltaïques (InGaAs, InAs, HgCdTe, ...). Les *détecteurs quantiques à détection extrinsèque* sont constitués essentiellement par les semi-conducteurs en germanium (Ge) ou en silicium (Si) dopés par des éléments métalliques (cf. Tableau 6).

Le Tableau 6 tiré d'une des documentations techniques<sup>88</sup> de la firme japonaise Hamamatsu, l'un des grands fabricants des détecteurs infrarouges, résume les différents types de détecteurs infrarouges, leur fenêtre spectrale de fonctionnement, leur température spécifique de fonctionnement et leur réponse spectrale (défectivité  $D^*$ ). La Figure 18 rappelle brièvement l'historique de l'apparition des principaux détecteurs infrarouges.

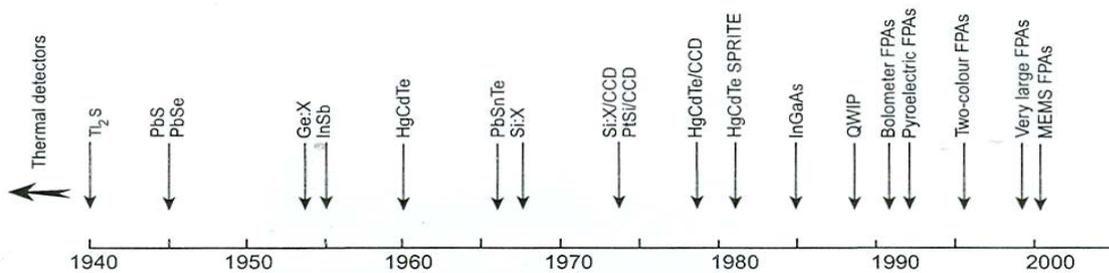
La seconde guerre mondiale a eu un rôle stimulateur pour le développement des détecteurs infrarouges en  $\text{Tl}_2\text{S}$  (sulfure de thallium) et en sels de plomb (PbS essentiellement). Juste après la guerre Golay<sup>89</sup> a construit la cellule pneumatique à expansion de gaz. Dans les années 1950

apparaissent les détecteurs aux semi-conducteurs à germanium à dopage métallique (Ge:X photoconducteur extrinsèque) et d'antimoniure d'indium (InSb photovoltaïque).

**Tableau 6 : Les principaux détecteurs infrarouges (thermiques & quantiques) et leurs principales caractéristiques (documentation technique, Hamamatsu<sup>88</sup>).**

Type		Détecteur	Réponse spectrale (µm)	Température de fonctionnement	D*(cm Hz <sup>1/2</sup> /W)	
Thermique	Thermocouple , Thermopile	Cellule de Golay, ... PZT, TGS, LiTaO	Dépend du matériau de la fenêtre	300	D*(λ, 10, 1)=6x10 <sup>8</sup>	
	Bolomètre			300	D*(λ, 10, 1)=1x10 <sup>8</sup>	
	Cellule pneumatique			300	D*(λ, 10, 1)=1x10 <sup>8</sup>	
	Détecteur pyroélectrique			300	D*(λ, 10, 1)=2x10 <sup>8</sup>	
Quantique	Photoconducteur	PbS	1 à 3,6	300	D*(500,600,1)=1x10 <sup>9</sup>	
		PbSe	1,5 à 5,8	300	D*(500,600,1)=1x10 <sup>8</sup>	
		InSb	2 à 6	213	D*(500,1200,1)=2x10 <sup>9</sup>	
		HgCdTe	2 à 16	77	D*(500,1000,1)=2x10 <sup>10</sup>	
	Intrinsèque	Photovoltaïque	Ge	0,8 à 1,8	300	D*(λp)=1x10 <sup>11</sup>
			InGaAs	0,7 à 1,7	300	D*(λp)=1x10 <sup>12</sup>
			InGaAs (exemple)	1,2 à 2,55	253	D*(λp)=2x10 <sup>11</sup>
			InAs	1 à 3,1	77	D*(500,1200,1)=1x10 <sup>10</sup>
			InSb	1 à 5,5	77	D*(500,1200,1)=2x10 <sup>10</sup>
			HgCdTe	2 à 16	77	D*(500,1000,1)=1x10 <sup>10</sup>
	Extrinsèque	Ge : Au	1 à 10	77	D*(500,900,1)=2x10 <sup>11</sup>	
		Ge : Hg	2 à 14	4,2	D*(500,900,1)=8x10 <sup>9</sup>	
		Ge : Cu	2 à 30	4,2	D*(500,900,1)=5x10 <sup>9</sup>	
Ge : Zn		2 à 40	4,2	D*(500,900,1)= 5x10 <sup>9</sup>		
Si : Ga		1 à 17	4,2	D*(500,900,1)= 5x10 <sup>9</sup>		
Si : As		1 à 23	4,2	D*(500,900,1)= 5x10 <sup>9</sup>		

En 1960, le détecteur quantique MCT (HgCdTe) fait son apparition. Le MCT révolutionne la qualité de détection au-delà de 12 µm et la capacité d'échantillonnage. En 1970, c'est la mise au point des dispositifs à couplage de charge (charge-coupled devices) CCD par Boyle<sup>77</sup> et Smith.



**Figure 18 : historique du développement des détecteurs infrarouges (d'après A. ROGALSKI<sup>78</sup>)**

L'association du détecteur à silicium dopé par des métaux (Si:X) et les CCDs ouvre des nouveaux horizons d'investigation et un gain de temps et de précision dans les années 1970. Une autre association verra le jour à la fin des années 1970, celle des détecteurs MCT avec les CCDs suivis par l'invention des détecteurs MCT SPRITE tout au début des années 1980.

Un progrès remarquable est réalisé dans les détecteurs quantiques. Les photodétecteurs quantiques (QWIP : quantum well infrared photodetector)<sup>81</sup> à large zone 2-D de détection

apparaissent à la fin des années 1980. Au début des années 1990, c'est le retour en force des détecteurs thermiques avec les détecteurs aux plans focaux FPA (Focal Plan Array), essentiellement des bolomètres et des pyroélectriques. Les matières pyroélectriques ont un temps de réponse beaucoup plus rapide que les autres types de détecteurs pour une certaine gamme spectrale.

Au milieu des années 1990, on commence à parler de la mise en œuvre de la troisième génération des détecteurs infrarouges tels que les FPAs non refroidis, les détecteurs à 2-couleurs, la barrette hyperspectrale (hyperspectral arrays) ...etc. A partir de 2000 certains sont déjà en test.

### Le détecteur MCT.

Ce détecteur MCT était révolutionnaire dès son apparition. Il peut être de nature diverse (photoconducteur ou photovoltaïque). Il est en constante évolution depuis son apparition en 1960. Selon le taux Hg/Cd, des quantités de détecteurs avec une détectivité spectrale adaptée à la gamme spectrale voulue peuvent être produits. Pour cette raison on représente ce détecteur souvent par la formule  $Hg_{1-x}Cd_xTe$ ,  $x$  représentant le taux Hg/Cd.

Actuellement dans la troisième génération, les détecteurs MCT sont formés par épitaxie par jets moléculaires, MBE (Molecular Beam Epitaxy en anglais).

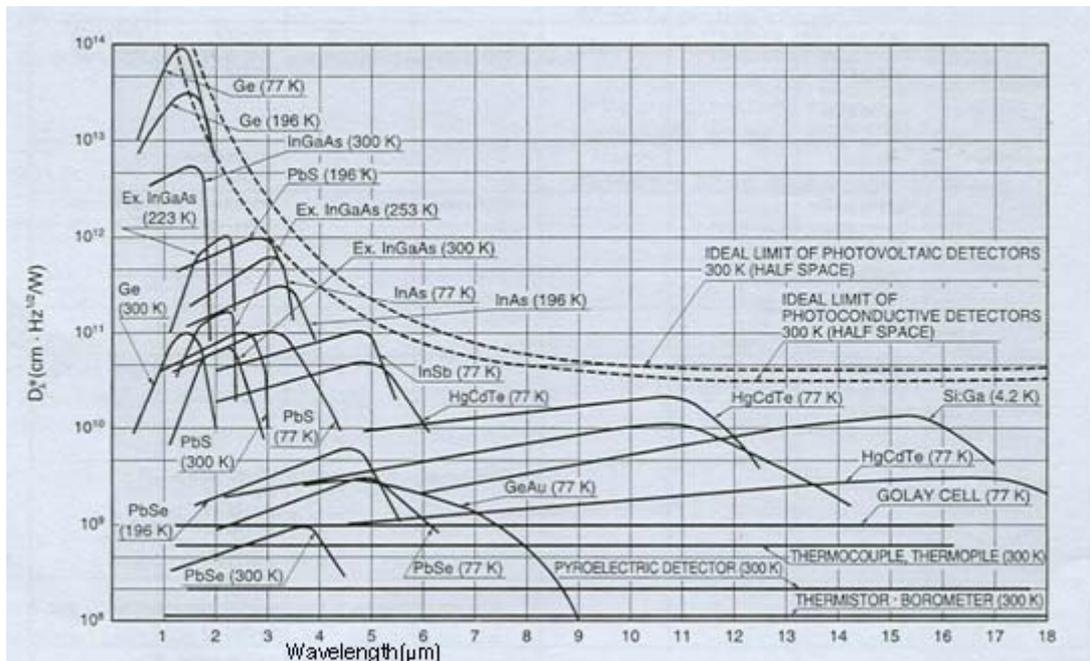


Figure 19 : Les réponses spectrales caractéristiques ( $D^*$ ) de nombreux détecteurs infrarouges pour l'intervalle [1-18  $\mu\text{m}$ ] avec leur température spécifique de fonctionnement (documentation technique, Hamamatsu<sup>88</sup>).

Une bonne réponse spectrale (détectivité  $D^*$ ) d'un détecteur dépend de la nature de ses constituants, de la température de fonctionnement et de la longueur d'onde du rayonnement (gamme spectrale d'étude).

La Figure 19 représente la détectivité spectrale  $D^*(\lambda, f, \Delta f)$  (en  $\text{cm Hz}^{1/2} \text{ W}^{-1}$ ) pour différents types de détecteurs utilisés en infrarouge pour un champ de vision FOV (fiel of view) de  $180^\circ$ .  $f$  est la fréquence de mesure et  $\Delta f$  la bande d'amplification du signal détecté (souvent  $\Delta f = 1$ ).

Nous remarquons, qu'à faibles longueurs d'onde, le semi-conducteur en germanium est le plus efficace avec la plus grande détectivité. Pour les longueurs d'onde supérieures à  $5\mu\text{m}$ , le MCT est l'un des plus efficaces. La variation du taux Hg/Cd le rend plus efficace sur un grand intervalle de longueurs d'onde. Pour toutes ces raisons, le détecteur MCT, en outre les CCDs, est l'un des plus utilisés.

Toutes ces technologies associées à la puissance des ordinateurs et à la transformée de Fourier (TF) permettent une meilleure analyse des signaux spectroscopiques.

## 5. La Transformée de Fourier

### 5.1. Série de Fourier et Transformée de Fourier (TF)

La série de Fourier transforme une fonction  $f(x)$  périodique de période  $T$ , monotone par tranches et bornée sur un intervalle de longueur  $T$ , remplissant les conditions de Dirichlet, en une somme des cosinus et des sinus. Soit  $f(x)$ , son développement en série de Fourier s'écrit

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos(nx) + b_n \sin(nx)) \quad (\text{éq.48})$$

où  $a_0$ ,  $a_n$  et  $b_n$  sont les coefficients de Fourier. Le

problème de la transformation se ramène alors au calcul de ces derniers par les relations :

$$a_0 = \frac{2}{T} \int_0^T f(x) dx \quad (\text{éq.49}), \quad a_n = \frac{2}{T} \int_0^T f(x) \cos(nx) dx \quad (\text{éq.50}), \quad \text{et} \quad b_n = \frac{2}{T} \int_0^T f(x) \sin(nx) dx \quad (\text{éq.51}).$$

Pour plus de facilité d'écriture mais aussi de résolution, la notation complexe est utilisée. Rappelons que  $e^{j\theta} = \cos \theta + j \sin \theta$  avec  $j^2 = -1$  le nombre complexe. Si une fonction  $f(x)$  est périodique de période  $T$ , sa transformée de Fourier est aisée et s'écrit :

$$f(x) = \sum_{n=-\infty}^{\infty} C_n e^{j \frac{2\pi n x}{T}} \quad (\text{éq.52}) \quad \text{où} \quad C_n = \frac{1}{T} \int_0^T f(t) \cdot e^{-j \frac{2\pi n t}{T}} \cdot dt \quad (\text{éq.53})$$

sont les coefficients de

Fourier complexes. Si la fonction n'est pas périodique ou de période infinie, elle peut être cependant transformée mais la somme dévient « l'intégrale de Fourier ». On considère un continuum de fréquences dans le traitement du signal par exemple. D'une manière générale, la

$$\text{transformée de Fourier } F(u) \text{ d'une fonction } f(x) \text{ s'écrit : } F(u) = TF(f(x)) = \int_{-\infty}^{+\infty} e^{-2j\pi u x} f(x) dx$$

(éq.54). La TF d'une fonction est un complexe et possède donc une amplitude et une phase. Dans le traitement du signal, c'est le module au carré de l'amplitude (fonction de puissance) qui est utilisé dans les représentations spectrales. L'étude des phases permet d'approfondir et de détecter des caractéristiques spécifiques aux composantes spectrales.

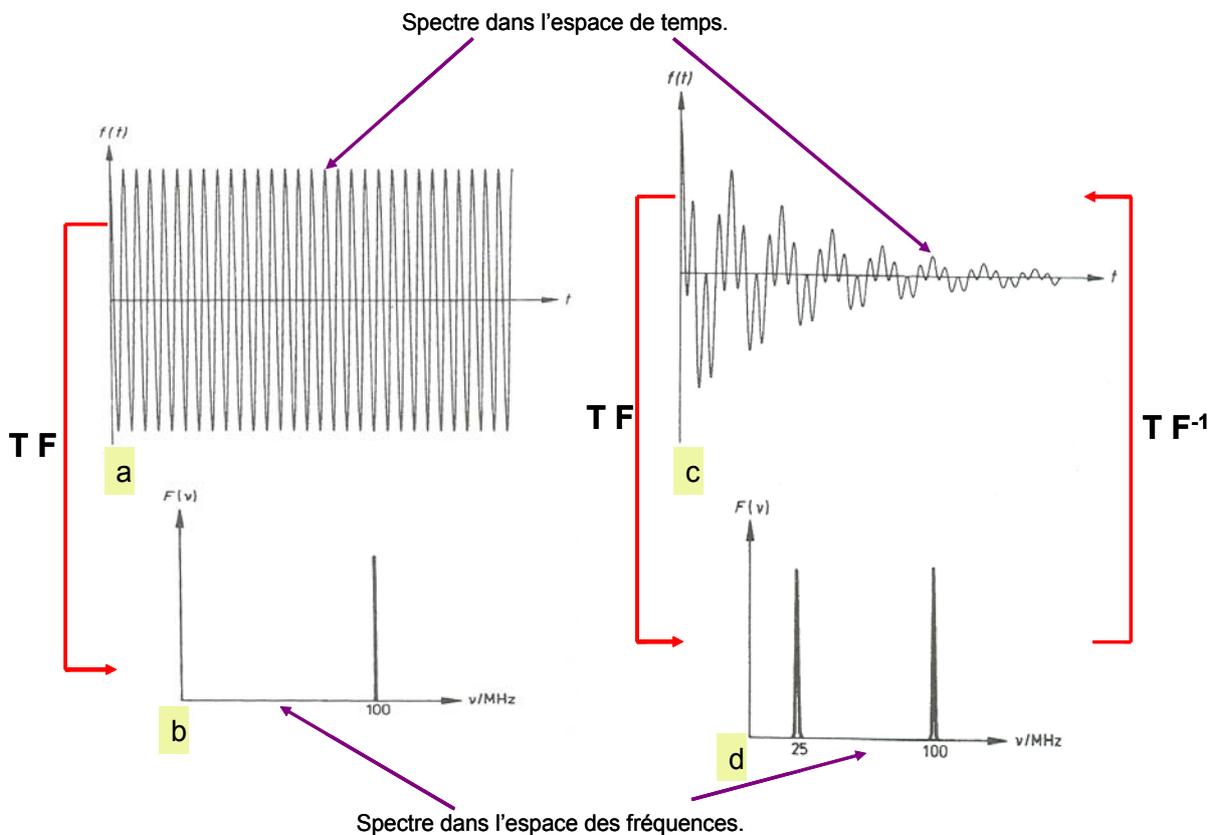
## 5.2. La transformée de Fourier et le traitement numérique du signal<sup>90 91</sup>

Dans le traitement du signal, la TF nous permet de passer d'une représentation temporelle à une représentation fréquentielle d'un signal. Soit  $f(t)$  un signal dans l'espace temporel, sa TF  $F(\nu)$  dans l'espace fréquentiel est donnée par (éq.54) où  $u$  est remplacé par  $\nu$  et  $x$  par  $t$  :

$$F(\nu) = TF(f(t)) = \int_{-\infty}^{+\infty} e^{-2j\pi\nu t} f(t) dt \quad (\text{éq.54bis}).$$

La transformée de Fourier inverse (cf. Figure 20c-d), notée  $TF^{-1}$ , permet le passage inverse :  $f(t) = TF^{-1}(F(\nu)) = \int_{-\infty}^{+\infty} F(\nu) e^{2j\pi\nu t} d\nu$  (éq.55).

La TF possède des nombreuses propriétés intéressantes qui rendent son utilisation et sa maniabilité assez pratiques : respect de la linéarité, retard (phase affectée uniquement), dérivabilité, intégrabilité, inversibilité, respect du produit de convolution, ...etc.



**Figure 20 : Illustration de la Transformée de Fourier TF. a) et c) Spectres dans l'espace de temps avec une et deux fréquences respectivement ; b) et d) les spectres par Transformée de Fourier pour a) et c), les fréquences sont bien mises en évidence (J. Michael Hollas<sup>91</sup>). La transformée de Fourier inverse  $TF^{-1}$  réalise le passage inverse. L'élargissement des raies dépend d'un certain nombre de paramètres (durée de vie de l'état excité pour la RMN par exemple).**

Dans l'espace des temps, les spectres ont une allure compliquée difficilement interprétable. Leur transformée de Fourier dans l'espace des fréquences est plus aisée pour l'analyse. La Figure 20 illustre la transformation de Fourier d'un spectre d'une source monochromatique ( $a \rightarrow b$ ) et d'une source à deux longueurs d'onde ( $c \rightarrow d$ ). Nous voyons que la TF simplifie énormément le spectre et que l'interprétation est pratique dans l'espace des fréquences.

La transformée de Fourier rapide FFT (Fast Fourier Transform en anglais) mise au point par Cooley et Tukey<sup>92</sup> est utilisée dans le traitement numérique du signal. Cet algorithme est depuis lors très amélioré avec plusieurs variantes<sup>93</sup> (algorithme de Good-Thomas, celui split-radix, ...etc.). Les coefficients de Fourier sont calculés d'une manière itérative et catégorisée (entre pairs et impairs). La convergence est atteinte plus rapidement.

- **Transformée de Fourier et traitement d'images.**

Pour l'analyse et le traitement d'images, la transformée de Fourier est très utilisée. Une image est multidimensionnelle (au moins deux dimensions). De ce fait la TF multidimensionnelle (intégrale multiple) est utilisée (dans la RMN 2D par exemple). En chaque pixel  $(i,j)$ , on a une intensité  $f[i,j]$  pour une image bidimensionnelle. Dans le traitement numérique des signaux, c'est la TF discrète<sup>93</sup> qui entre en jeu prenant en compte le mode d'échantillonnage du signal considéré. La TF est utilisée aussi comme filtre numérique par convolution linéaire (ou déconvolution).

- **Transformée de Fourier et spectromètres.**

Les spectromètres à TF ont une instrumentation simplifiée et possèdent peu d'éléments optiques. Ce qui fait que l'intensité lumineuse n'est pas atténuée et arrive directement au détecteur où elle est échantillonnée : c'est *l'avantage de Jacquinot*. Le rapport signal/bruit est significativement plus important. En outre, l'ensemble d'informations atteint le détecteur. Le multiplexage permis par la TF réalise la mesure simultanée de toutes les longueurs d'onde de la plage spectrale et permet ainsi un formidable gain de temps : c'est *l'avantage Fellgett*. La TF permet aussi d'atteindre une meilleure résolution de recouvrements spectraux.

## 6. Les spectromètres infrarouges

Le progrès dans la technologie des détecteurs infrarouges est étroitement lié au progrès réalisé dans les détecteurs semi-conducteurs infrarouges.

Les spectromètres utilisés en infrarouge sont classés en deux catégories<sup>94,87</sup> : les appareils dispersifs et ceux à transformée de Fourier.

- **Les appareils dispersifs.**

Ces appareils fonctionnent suivant un mode séquentiel simple. Un monochromateur muni d'un réseau motorisé est utilisé pour balayer la plage spectrale étudiée. Ces spectromètres se subdivisent aussi en deux catégories : les spectromètres à simple faisceau et ceux à double faisceau. Les spectromètres à simple faisceau (comme Perkin-Elmer 12) étaient les premiers à être utilisés mais l'interférence des éléments du milieu ambiant (principalement CO<sub>2</sub>, vapeur d'eau) ont gravement nuit à la qualité des spectres obtenus. Ce qui a mené à l'invention des spectromètres à double faisceau (IR 435 de Shimadzu par exemple), un faisceau traversant

l'échantillon, l'autre non. Les deux faisceaux se rejoignent au niveau du monochromateur où leur comparaison permet de déduire directement le spectre de l'échantillon analysé.

- **Les appareils à transformée de Fourier.**

Quant aux spectromètres à transformée de Fourier, ils s'apparentent à ceux « simple faisceau » où un interféromètre est mis après la source. Les principales différences sont le remplacement du monochromateur par un interféromètre de Michelson<sup>95</sup> et le trajet des faisceaux lumineux à travers un système de miroirs (un mobile et un fixe) et d'une séparatrice. Avant l'enregistrement, le signal est traité numériquement. Les spectromètres à transformée de Fourier (IRTF) effectuent une analyse simultanée de toute la bande spectrale sélectionnée à partir de mesures interférométriques<sup>96</sup>. La Figure 21 montre le schéma de principe d'un spectromètre à transformée de Fourier comprenant l'interféromètre de Michelson. Le principe de multiplexage des signaux optiques à l'aide des ordinateurs de plus en plus puissants est appliqué. La mesure est effectuée avec une grande précision des longueurs d'onde grâce à la détermination des longueurs d'onde par le laser associé au dispositif. Ce type de spectromètre est aussi très utile dans les régions infrarouges où des détecteurs plus sensibles tels que les photomultiplicateurs ne sont pas disponibles.

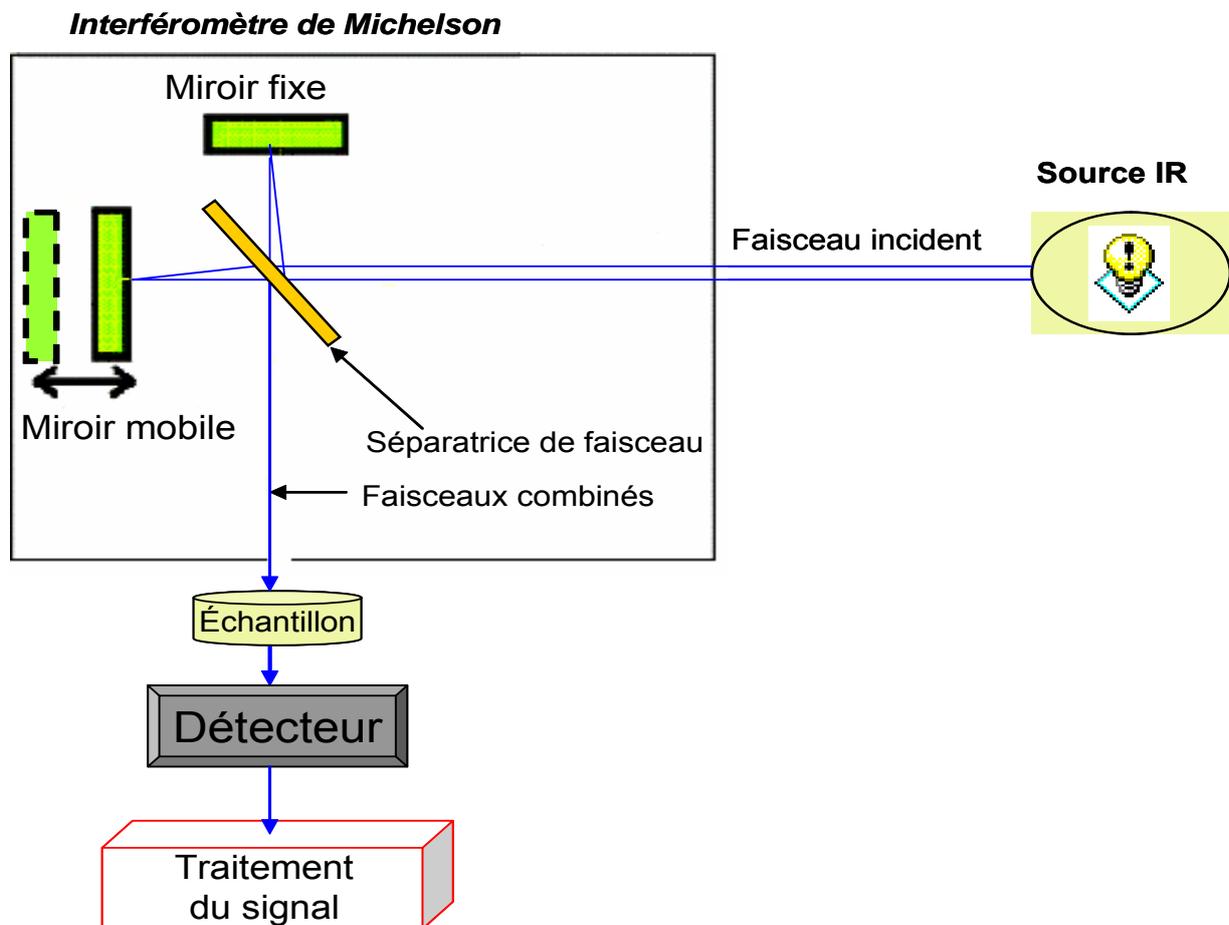


Figure 21 : Schéma de principe d'un spectromètre à Transformée de Fourier avec interféromètre de Michelson.

## II. Spectroscopie Raman

### 1. Brève historique

La spectroscopie Raman est l'un des principaux moyens pour la caractérisation moléculaire.

L'effet Raman semble être prédit théoriquement par A.G. Smekal<sup>97</sup> en 1923. Mais celui qui l'a bien mis en évidence par des observations rigoureuses est le physicien indien C.V. Raman<sup>98</sup>. En 1928, il constata que la longueur d'onde d'une petite fraction du rayonnement diffusé par certaines molécules était différente de celle du rayonnement incident. Il montra que ce déplacement chimique dépendait de la structure des molécules du milieu diffusant irradié et non de la nature de la lumière incidente. Tous ces travaux valurent à Raman le prix Nobel de Physique en 1930. Ce phénomène est par la suite qualifié « d'effet Raman ».

Cependant la littérature nous rapporte que cet « effet Raman » était indépendamment découvert en Union Soviétique, à la même époque, par les deux physiciens G.S. Landsberg et L.I. Mandelstam<sup>99</sup> dans leur étude des systèmes cristallins.

Par la suite avec le progrès instrumental, on distinguera plusieurs types de Raman : Raman conventionnel, Raman de résonance, Raman stimulé...etc.

### 2. Effet Raman

#### Mécanisme de l'effet Raman

Pour qu'une molécule manifeste la diffusion Raman, il faut qu'il y ait une variation de la polarisabilité<sup>100</sup> de la molécule en fonction de la distance interatomique. Une vibration fondamentale sera active en Raman si elle a les mêmes propriétés de symétrie que l'un des éléments actifs du tenseur de polarisabilité  $\alpha$ . L'interaction du champ électrique incident  $\vec{E}_i$  avec le dipolaire moléculaire entraîne une modification de la forme du nuage électronique créant ainsi un dipôle électrique  $\vec{P} = \alpha_0 * \vec{E}_i$ ,  $\alpha_0$  étant la polarisabilité linéaire de la molécule. La théorie des groupes est utilisée, comme en infrarouge, afin de dénombrer et de classer les vibrations moléculaires dans les différentes représentations irréductibles d'un groupe de symétrie. L'examen des « tables de caractères » des groupes ponctuels de symétrie permet de prédire si une vibration sera active ou non en Raman pour des molécules simples. Cependant, selon la symétrie de la molécule certaines vibrations peuvent être actives à la fois en Raman et en infrarouge.

En Raman, après absorption d'un photon ( $E = h\nu$ ), l'énergie de la molécule peut donc prendre toutes les valeurs comprises entre celle de l'état électronique fondamental et le premier état électronique excité (Cf. Figure 22a). Après absorption, on dit que la molécule se trouve dans un état virtuel avant de se désexciter et de revenir au niveau fondamental.

**Application à l'analyse des échantillons.**

Lorsqu'un échantillon est irradié, ses molécules diffusent dans toutes les directions une partie de la lumière incidente. La grande partie de la lumière diffusée a la même longueur d'onde que celle de la lumière incidente ( $\lambda=\lambda_0$ ), on parle alors de diffusion élastique ou « Rayleigh ». Une petite fraction a une longueur d'onde différente ( $\lambda\neq\lambda_0$ ) et c'est celle-là qu'on qualifie de « déplacement Raman ». Si  $\lambda>\lambda_0$  ( $\bar{\nu} < \bar{\nu}_0$ ), on a les raies Stokes avec une intensité Raman plus importante (cf. Figure 22). Par contre si  $\lambda<\lambda_0$  ( $\bar{\nu} > \bar{\nu}_0$ ), on a les raies Anti-Stokes avec une intensité Raman plus faible que celles Stokes. La Figure 22 illustre cette transition et montre ces différentes raies en fonction du déplacement Raman. Pour l'exploitation spectrale, et pour une raison pratique, le déplacement Raman  $\Delta\bar{\nu}$  est exprimé par la différence entre le nombre d'onde incident ( $\bar{\nu}_0$ ) et celui diffusé ( $\bar{\nu}$ ) :  $\Delta\bar{\nu} = \bar{\nu}_0 - \bar{\nu}$ . Ce déplacement chimique, comme déjà précisé, ne dépend pas de la longueur d'onde incidente mais bien des caractéristiques des molécules présentes dans l'échantillon. Cependant il faut préciser que la longueur d'onde incidente et celle diffusée sont toutes les deux dans le même domaine spectral (visible, UV, proche infrarouge).

Les raies Raman Anti-Stokes sont moins intenses que celles Raman Stokes mais sont non sujettes à la fluorescence qui perturbent ces dernières. Précisons cependant que l'intensité Raman diffusée est proportionnelle au nombre des molécules qui diffusent. Ce qui permet de calculer, en principe, la concentration de l'espèce étudiée. Mais le nombre des molécules sujettes à l'effet Raman est proportionnellement très faible (de 1 pour  $10^9$  à 1 pour  $10^6$ ).

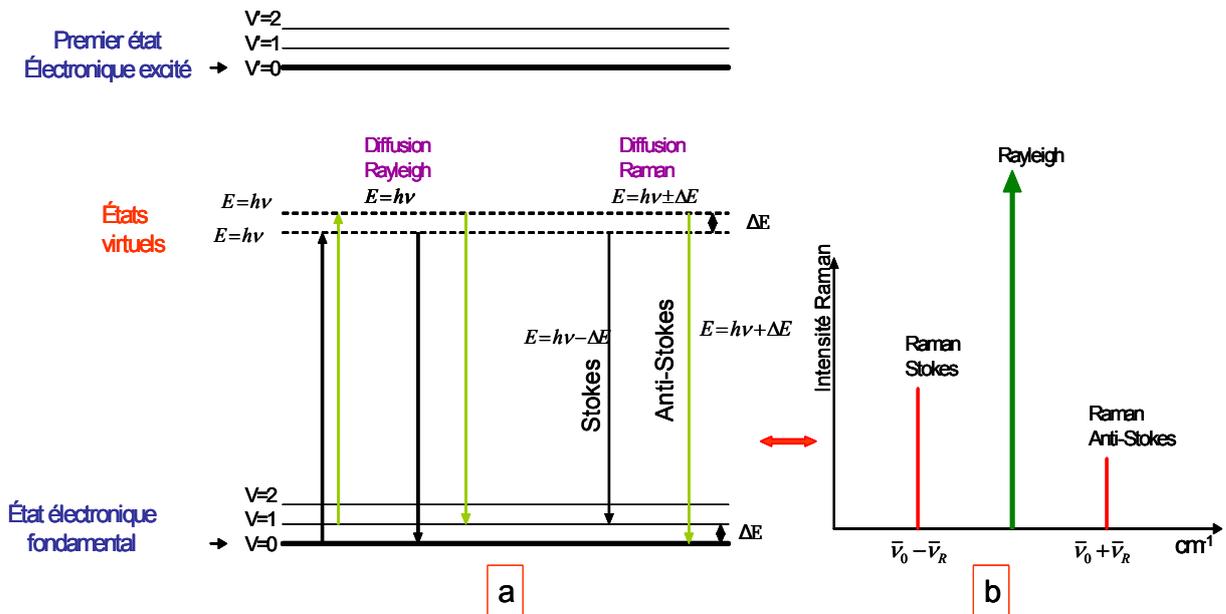


Figure 22 : a) Transition montrant la diffusion Rayleigh et Raman<sup>87</sup>, b) Déplacement Raman Stokes et Anti-Stokes en fonction de l'intensité Raman ( $\bar{\nu}_0$  fréquence de l'onde incidente,  $\bar{\nu}_R$  fréquence diffusion Raman).

### Identification moléculaire.

La spectroscopie Raman, bien qu'ayant ses spécificités, apporte un complément d'informations non négligeable à l'infrarouge, notamment dans l'identification moléculaire. On obtient des spectres semblables à ceux obtenus en infrarouge. D'une manière analogue à l'infrarouge, chaque liaison chimique donnée a un déplacement Raman qui lui est spécifique. Disposant d'un catalogue des déplacements Raman, on peut identifier les structures moléculaires. En s'intéressant d'une manière appropriée à certains types de liaisons, on peut faire une quantification<sup>101</sup> de certains produits.

### 3. Instrumentation Raman

Dans un échantillon donné, très peu de molécules diffusent en Raman. Par conséquent l'intensité diffusée est elle aussi très faible. Cette contrainte a longtemps limité l'utilisation de la spectroscopie Raman. Le Raman n'est apparu très intéressant qu'après l'apparition des puissantes sources d'énergie tels que les lasers dans les années 1960. En effet, il faut la combinaison de trois bons éléments :

- ✓ une puissante source d'énergie tels que les lasers ;
- ✓ un bon système d'irradiation de l'échantillon ;
- ✓ et un spectromètre adéquat.

La Figure 23 montre schématiquement ce dispositif.

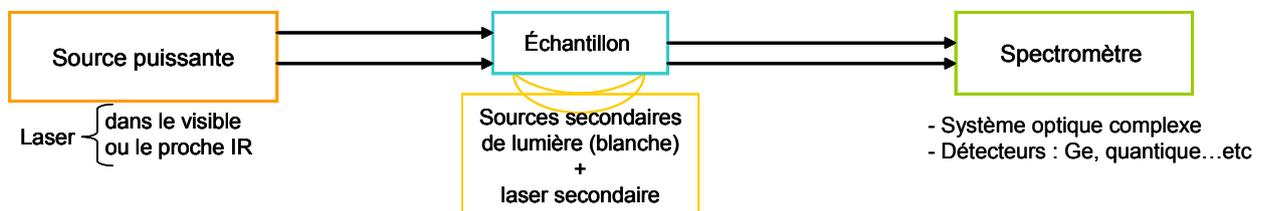


Figure 23 : Schéma du dispositif de l'instrumentation Raman.

Pour la microspectrométrie Raman, un microscope optique sert à focaliser à la fois les différentes radiations sur l'échantillon et à collecter les rayonnements diffusés par celui-ci. Actuellement, la spectroscopie Raman possède l'une des meilleures résolutions, une résolution de 1  $\mu\text{m}$  (meilleure que l'infrarouge). On l'utilise dans des nombreuses études telles que l'identification des phases, la caractérisation des matériaux, la détermination de la structure moléculaire, l'étude des systèmes amorphes et cristallins.

Comme l'infrarouge, la spectroscopie Raman<sup>102</sup> est une méthode non destructive et non intrusive. Elle a de nombreux avantages telles que l'étude des échantillons de très faible concentration ou de très petite taille, l'indépendance par rapport à la longueur d'onde excitatrice utilisée, ...etc. Quelques gênes rencontrées auparavant par la spectroscopie Raman telles que la fluorescence, des réactions photochimiques (décomposition des substances colorées...) ou

multiphotoniques, la décomposition des échantillons par échauffement sont actuellement surmontées. Dans les spectromètres à transformée de Fourier sont utilisés des excitatrices moins énergétiques (Nd:YAG par exemple), des multicanaux de mesure et la détermination des longueurs d'onde par des lasers secondaires.

### 3.1. Les sources

La source doit être à la fois puissante et monochromatique. Ce rôle est bien rempli par les différents types de lasers utilisés.

#### Caractéristique des lasers<sup>103</sup>

Les lasers se caractérisent par leur grande directivité, leur monochromaticité, leur grande cohérence mais essentiellement aussi par leur grande intensité et leur efficacité d'échantillonnage. L'émission stimulée est la base de l'effet laser. Pour exciter les atomes de l'espèce réactive, un pompage (par passage d'un courant électrique, par décharge ou par réaction chimique) est nécessaire. La stimulation de l'espèce excitée, par des photons ayant une énergie égale à la différence d'énergie entre l'état excité et l'état fondamental, produit l'émission des photons cohérents à ceux incidents.

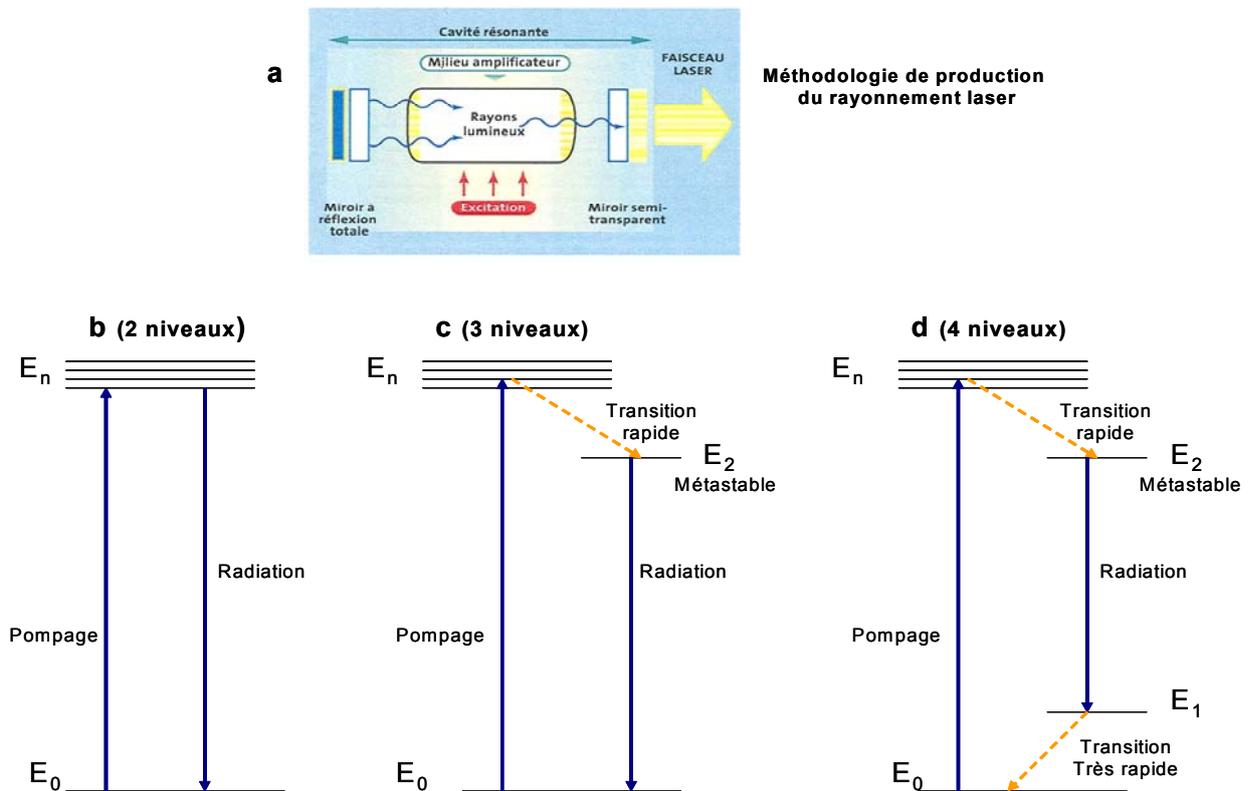


Figure 24 : a) Mode de production du rayonnement laser dans la cavité résonante par excitation et amplification (inrs<sup>103</sup>) ; b, c et d) les différents types de lasers selon le nombre de niveaux d'énergie. Pour le laser à 4 niveaux d'énergies (d), l'espèce excitée dans l'état métastable  $E_2$  a une durée de vie plus longue que celle dans l'état  $E_1$ , ce qui permet une inversion de la population et une amplification de l'effet laser.

Tableau 7 : Quelques sources lasers et leurs caractéristiques.

<b>Spectre (gamme spectrale d'émission)</b>	<b>Élément (Laser)</b>	<b>Longueur d'onde caractéristique</b>
<b>Visible</b>	Argon ionisé	488 nm (Bleu) ou 514,5 nm (vert)
	Krypton ionisé	530.9 nm (vert) ou 647.1 nm (rouge)
	Hélium/Néon (He/Ne)	632.8 nm (Rouge)
<b>Infrarouge</b>	Diode laser	782 ou 830 nm (PIR)
	Nd:YAG	1064 nm (PIR)

La caractéristique d'un laser dépend aussi du nombre de niveaux d'énergie. Les lasers à plusieurs niveaux d'énergie (comme Nd :YAG : 4 niveaux d'énergie) sont les plus pratiques car ils sont peu coûteux en énergie. La Figure 24 présente le mode de production de l'effet laser et les différents types de lasers selon le nombre de niveaux d'énergie.

Des lasers émettent dans le visible tels que ceux à Argon ionisé ou Krypton ionisé. Dans ce domaine, le plus utilisé est le laser à Hélium/Néon (He/Ne) émettant dans le rouge. D'autres types de lasers ont des caractéristiques très intéressantes et émettent dans le proche infrarouge tels que les diodes lasers et le laser Nd:YAG. Le Tableau 7 présente les caractéristiques des lasers mentionnés ci-dessus. En outre, le système d'irradiation doit être adapté à la nature de l'échantillon étudié (solide, liquide, gaz). Beaucoup de systèmes existent pour focaliser la lumière sur l'échantillon ou la collecter. Ces systèmes peuvent être des fibres optiques, des systèmes optiques ou des microscopes optiques.

### 3.2. Les spectromètres

Anciennement, les spectromètres en Raman étaient semblables aux appareils dispersifs utilisés en infrarouge ou en visible et utilisaient des photomultiplicateurs comme transducteurs pour amplifier le signal Raman par nature faible.

Actuellement, les spectromètres à Transformée de Fourier<sup>104</sup> avec des transducteurs à germanium refroidis à la température de l'azote liquide (77K) ou des spectromètres dispersifs multicanaux<sup>105</sup> aux détecteurs à couplage de charges (CCD) sont les plus utilisés. Les spectromètres comportent plusieurs autres éléments pour améliorer la qualité d'échantillonnage tels que des filtres holographiques ou monochromatiques pour éliminer, par exemple, la longueur d'onde de la source ou sélectionner certaines gammes spectrales.

L'imagerie microspectrométrique infrarouge et Raman utilisent les différentes potentialités offertes par ces deux types de spectroscopies vibrationnelles. Dans la partie suivante, les principes de base et les méthodologies de ces imageries sont exposés.

### III. Imagerie microspectrométrique infrarouge et Raman

#### 1. Aperçu général

Au 19<sup>ème</sup> siècle, les hommes ont pu produire des images sur des plaques photographiques grâce aux progrès techniques et scientifiques notamment dans les domaines de l'optique et de la chimie.

La première image produite qu'on peut citer du point de vue spectroscopique est celle obtenue par l'utilisation des rayons X dont la découverte est due à W.C. Roentgen en 1895. La radiographie et le scanner utilisent cette technique d'imagerie par les rayons X. Spectroscopiquement<sup>106</sup> parlant, c'est la diffraction des rayons X qui permettra tout un domaine d'imagerie appliquée aux différents types d'analyses (des structures cristallines, de la caractérisation des surfaces, d'identification des microparticules<sup>107</sup>, ...).

La découverte de la radioactivité naturelle par Becquerel (1896) et artificielle par P. et M. Curie (1898) va conduire au développement de la spectroscopie nucléaire. Plus tard, dans le domaine médical, le développement de l'imagerie nucléaire (comme la scintigraphie, 1960) et la tomographie par émission des positons (TEP, 1990) en sont des conséquences.

L'autre domaine dont les images sont très utilisées dans les domaines militaires et industriels est l'infrarouge<sup>108</sup> dont la première utilisation en imagerie remonterait à la seconde guerre mondiale. Dans le domaine industriel, l'imagerie microspectrométrique infrarouge est, depuis les années 1980, assez intensivement appliquée dans le processus de contrôle de la fabrication des composants électroniques, à la recherche des contaminants ou de l'impureté dans l'industrie pharmaceutique<sup>109</sup>, au contrôle des imperfections dans les couches des polymères<sup>110</sup>, dans le textile, dans l'industrie du tabac, dans l'industrie agroalimentaire, dans l'analyse et l'authentification des objets d'art<sup>111</sup> et beaucoup d'autres produits<sup>112,113,114,115</sup>.

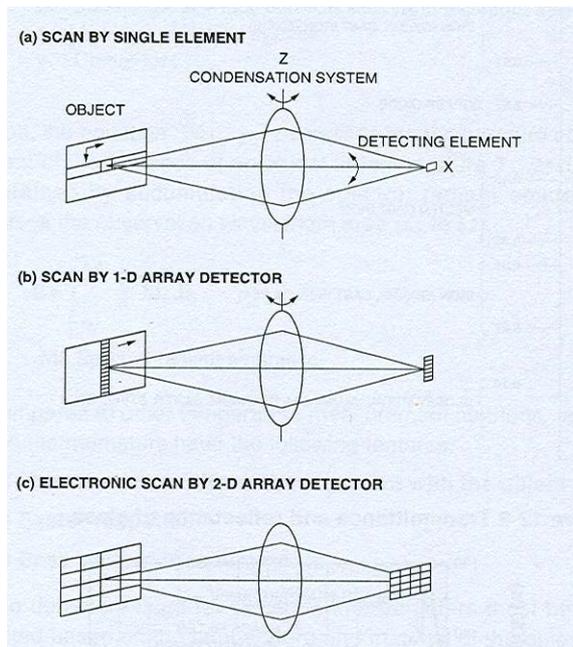
L'imagerie microspectrométrique Raman est aussi efficacement utilisée dans des nombreuses applications<sup>116,117,118</sup>. Nous citons comme exemple l'identification des particules contaminantes dans l'étude de la pollution aérienne industrielle<sup>119</sup>. Les spectroscopies infrarouge et Raman sont avantageuses grâce à leur grande sélectivité et à la disponibilité d'une instrumentation d'utilisation très pratique.

D'autres spectroscopies telles que la spectrométrie de masse<sup>120</sup> (appliquée par exemple à l'analyse de l'organisation spatiale de certains constituants de tissus cellulaires mammaires), l'UV, la RMN<sup>121</sup>, la fluorescence ou la phosphorescence produisent des images spectroscopiques de grande qualité. Ces dernières méthodes sont beaucoup plus intéressantes pour la quantification. Pourtant, leur manque d'une grande efficacité sélective limite leur utilisation intensive.

## 2. Moyens de production d'images en infrarouge et Raman

Depuis la seconde guerre mondiale, la production d'images en infrarouge s'est constamment améliorée avec le perfectionnement des moyens de détection et d'échantillonnage. A l'origine, les images produites au début étaient des images non scannées obtenues au moyen des détecteurs thermiques. Des évaporographes<sup>122</sup> étaient utilisés dans lesquels les radiations étaient focalisées sur une membrane noircie recouverte d'une fine couche d'huile. Le taux différentiel d'évaporation de l'huile était proportionnel à l'intensité de la radiation. Le film était ensuite éclairé par la lumière blanche pour produire la figure d'interférence correspondant à l'image thermique. Le deuxième dispositif de production d'images thermiques utilisé était basé sur « le convertisseur d'image de la limite d'absorption »<sup>123</sup> (absorption-edge image converter). Le fonctionnement du dispositif était basé sur la dépendance de la température avec la limite d'absorption du semi-conducteur utilisé. Cependant la performance de ces deux dispositifs était faible à cause d'une constante de temps importante et d'une faible résolution spatiale.

Par la suite des photodétecteurs refroidis, comme le PEV (pyroelectric vidicom), sont devenus très compétitifs pour la production d'images thermiques. L'invention des détecteurs quantiques MCT (HgCdTe) en 1960 a inspiré par la suite le développement de trois générations de détecteurs basés sur cette technologie. Ce développement était stimulé par la demande dans le domaine militaire, notamment aux Etats-Unis. Le développement se faisant sous contrat militaire classé secret défense, d'autres pays comme la France<sup>124</sup> ont entrepris de développer leur propre technologie sur ce type de détecteur.



**Figure 25 : Représentation schématique de la détection de 3 générations des détecteurs basés sur le MCT (documentation technique, Hamamatsu<sup>88</sup>); a) détection élément par élément, b) détection d'une rangée entière, c) détection 2D d'une surface.**

Cette technologie a permis de produire des images « par scans ».

En effet la première génération permettait de scanner élément par élément, la deuxième de scanner toute une rangée en une seule fois et la troisième un scan 2-D d'une surface. La Figure 25 montre schématiquement les différents types de scans pour les différents « arrays detectors ».

Pour les FPA, les détecteurs sont placés au plan focal du système optique de condensation.

L'association des détecteurs MCT et des caméras CCD a amélioré à la fois la qualité de détection et de scanning. Dans la troisième génération, on tente de réduire la taille du pixel du

CCD, d'améliorer l'efficacité de détection au niveau de chaque pixel (le rendement quantique) ...etc. On augmente de plus en plus la taille du détecteur en intégrant plus de pixels.

Les « uncooled IRFPA »<sup>125</sup>, détecteurs infrarouges aux plans focaux non refroidis, sont actuellement à la mode.

Au laboratoire, nous disposons une caméra CCD associée à un détecteur MCT refroidi à la température de l'azote liquide. Ce refroidissement a un rôle essentiel : pour améliorer le rapport signal sur bruit (S/N), il faut maintenir le détecteur à basse température constante pour que les mesures ne soient pas influencées par une brusque variation de la température. La Figure 26 montre un exemple de la configuration d'un dispositif d'acquisition d'images en infrarouge. Nous voyons l'emplacement du détecteur InSb ou MCT refroidi par l'azote liquide. La caméra CCD qui sert à visualiser est mise à la sortie « to display unit ».

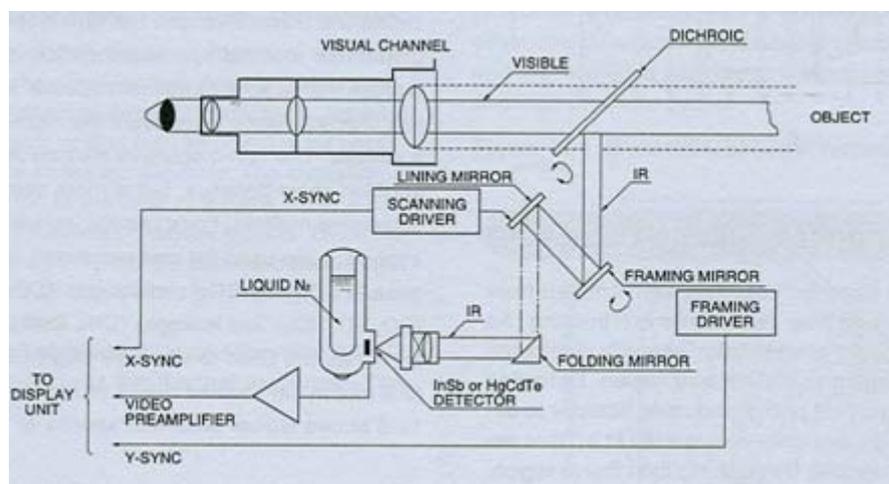


Figure 26 : Exemple de configuration d'un dispositif d'acquisition d'image en infrarouge (documentation technique, Hamamatsu<sup>88</sup>)

Les caméras CCD possèdent de nombreuses caractéristiques<sup>126</sup> telles que la résolution, la sensibilité, la qualité du rapport Signal/Bruit, ...etc.

La *résolution TV* (ou *résolution horizontale*) indique le nombre maximal de lignes noires et blanches que la caméra est capable de résoudre sur une largeur égale à la hauteur de l'image. C'est la résolution la plus élevée que peut atteindre réellement la caméra.

La *sensibilité* d'une caméra CCD traduit la réponse minimale de sortie produite par unité de surface par un nombre minimal de photons incidents. Elle dépend de plusieurs paramètres comme la surface réellement active du pixel et leur nombre, l'efficacité quantique, la perte de lumière par réflexion, la perte de charge lors de la phase de balayage, ...etc.

Le *rapport Signal/Bruit* est une caractéristique importante car au sein d'une caméra de multiples facteurs sont générateurs de bruit (bruit de grenaille, bruit de courant d'obscurité, ...). Compte tenu de la nature du capteur et de la taille de ses photoéléments (pixels), ce rapport évolue en général entre 45 et 60 dB.

Le *niveau de noir*, la *dynamique* et la *caractéristique de transfert* (éclairage, temps d'intégration) sont d'autres propriétés importantes des CCDs à tenir en compte.

La *numérisation*, c'est-à-dire la conversion du flux reçu par les pixels en un signal électrique (courant ou tension), est un processus complexe nécessitant une interpolation et un convertisseur analogique/numérique (CAN). Ce CAN comporte un nombre de bits adapté à la dynamique de la caméra.

Ces dispositifs que nous venons de décrire contribuent à la production des images spectroscopiques.

### 3. Production d'images spectroscopiques<sup>127 114</sup>

#### 3.1. Description des dispositifs

Au chapitre 1, nous avons brièvement rappelé quelques caractéristiques générales des images spectroscopiques. Nous avons indiqué que la production d'une image spectroscopique dépend du type de spectroscopie et donc de la nature du signal mis en jeu.

Dans le cadre de notre étude, nous nous intéressons aux microspectrométries infrarouge et Raman. Pour la microimagerie en IRTF, dans le livre de B. C. Smith<sup>128</sup> sont contenues des explications sur la description du système IRFT, son fonctionnement, les différentes manipulations spectrales, l'échantillonnage et le couplage de différentes techniques spectroscopiques.

En infrarouge et en Raman, pour l'acquisition de l'image, on associe au spectromètre un microscope<sup>129</sup> optique. Celui-ci sert aussi à la focalisation de la lumière en Raman et en Infrarouge et à la collecte du rayonnement réémis par l'échantillon. Ce microscope est associé à une platine (x-y) contrôlée par un moteur qui peut être lui-même commandé automatiquement par un ordinateur. Certains microscopes sont munis d'un objectif « autofocus Z » (déplacement vertical) par contrôle piézoélectrique pour mieux se positionner à la surface l'échantillon. Le système optique associé au dispositif infrarouge réduit le faisceau du rayonnement à une dimension proche de celle analysée sur l'échantillon. La Figure 27 montre l'appareillage pour la microspectrométrie infrarouge. La Figure 28 détaille les éléments optiques du microscope optique IRscope2 utilisé au laboratoire pour l'imagerie infrarouge. On observe bien l'association d'un détecteur MCT et d'une caméra CCD. Les trajets différents de deux modes de fonctionnement, transmission et réflexion, sont schématisés aussi.

Après avoir obtenu l'échantillon voulu, défini les principales caractéristiques à chercher, on sélectionne sur l'échantillon, après une investigation globale, la zone d'analyse.

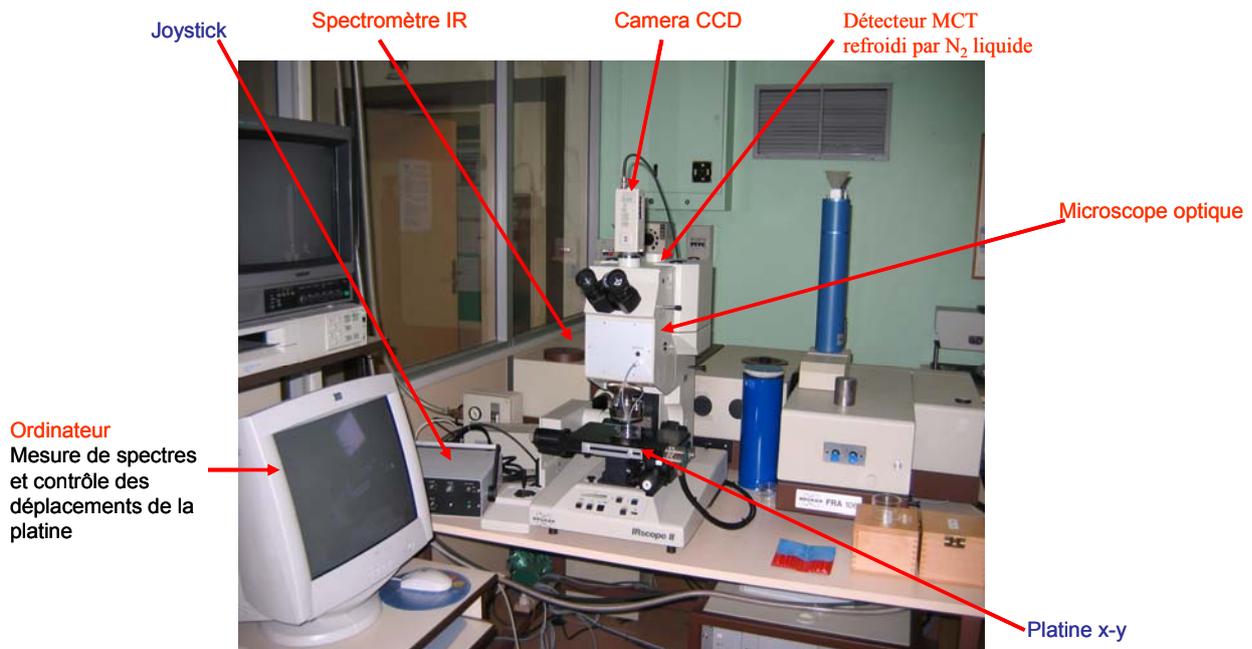


Figure 27 : Appareillage de la microspectrométrie infrarouge (IRscopeII et spectromètre IR ...etc.).

Avant la mesure des spectres, certains aspects techniques<sup>130,127,131</sup> doivent être bien définis.

### En infrarouge.

Selon la nature de l'échantillon et de ce que l'on veut obtenir, on doit choisir le domaine spectral : proche, moyen ou lointain infrarouge.

En moyen infrarouge par exemple, on doit définir l'ouverture du faisceau lumineux sur l'échantillon (**OLE**), « aperture » en anglais, les pas de balayage dans l'espace (x-y), les dimensions (x,y) de la zone à analyser. En effet, la netteté de l'image obtenue est en rapport avec la relation<sup>127</sup> entre cette ouverture lumineuse sur l'échantillon (**OLE**) et le pas de balayage (**pab**).

Trois cas sont possibles :

- ✓ Si le pas de balayage est égal à l'ouverture (**pab = OLE**), il y a un bon compromis et la résolution spatiale est acceptable.
- ✓ Si le pas de balayage est supérieur à l'ouverture (**pab > OLE**), il y a beaucoup de zones mortes, c'est-à-dire non mesurées, et donc la possibilité de ne pas prendre en compte certaines caractéristiques importantes de l'échantillon. Ce n'est pas recommandé pour la microimagerie car dans celle-ci on tente de prendre en compte toutes les caractéristiques de la zone analysée sur l'échantillon.
- ✓ Par contre si le pas de balayage est inférieur à « l'ouverture » (**pab < OLE**), on peut obtenir une bonne résolution spatiale. Ce cas est celui qui est le plus intéressant et le plus souvent utilisé.

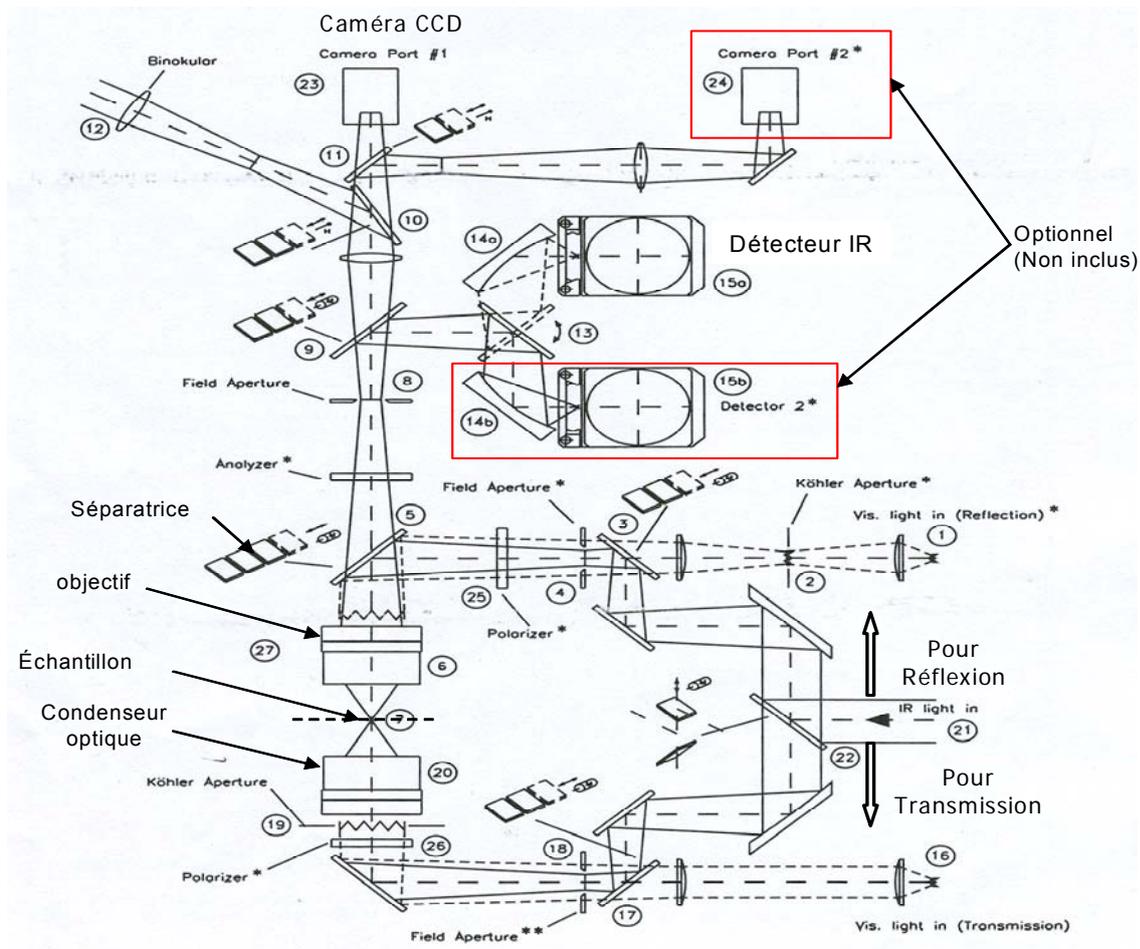


Figure 28 : Configuration optique du microscope optique IRscope2 utilisé au laboratoire dans l'imagerie microspectrométrique infrarouge pour les deux modes d'utilisation (transmission et réflexion) (documentation technique Bruker).

### En Raman.

Après le choix de l'échantillon, c'est le choix de la source d'irradiation qui détermine la longueur d'onde d'irradiation (dans le visible ou le proche infrarouge). Si par exemple, l'échantillon est assez fluorescent, une source dans le proche infrarouge tel que le laser Nd:YAG ( $\lambda = 1064 \text{ nm}$ ) est particulièrement efficace pour fortement diminuer voire totalement éliminer cette fluorescence. Cependant ces types de sources sont moins énergétiques et on risque d'avoir une diminution de l'intensité du signal.

En Raman<sup>132</sup>, le faisceau lumineux a une étendue d'environ  $1 \mu\text{m}$ . Ce qu'il faut choisir, c'est le pas de balayage, le temps d'irradiation avant chaque mesure, le temps d'acquisition et le nombre d'accumulations pour chaque position de mesure.

### Problème de diffraction en infrarouge et en Raman

Cependant que ce soit en infrarouge ou en Raman, il faut être vigilant aux problèmes liés à la diffraction<sup>133</sup>. En infrarouge, connaissant la longueur d'onde du rayonnement, c'est l'ouverture du faisceau lumineux sur l'échantillon qui définit la limite de la diffraction. En moyen

infrarouge, le domaine spectral utilisé étant compris entre 2,5 et 25  $\mu\text{m}$  ( $4000 - 400 \text{ cm}^{-1}$ ), la résolution spatiale<sup>134</sup> limitée par la diffraction est de l'ordre de 10  $\mu\text{m}$  (cf. éq.1). Il se pose un problème similaire en Raman lié à l'ouverture de la fente et à la longueur d'onde d'illumination. Un traitement méthodique et analytique de la diffraction peut être trouvé dans les articles de F. E. Lytle<sup>135,136</sup>. Nous ne donnerons pas de détails supplémentaires à ce propos.

### 3.2. Principe d'acquisition des spectres<sup>112</sup>

Avec le pas de balayage défini, on balaye dans le sens spécifié (horizontal puis vertical ou vertical puis horizontal) la zone à analyser de l'échantillon. A chaque position, c'est-à-dire dans chaque pixel, un spectre est mesuré et correspond à la surface couverte par le faisceau lumineux. De ce fait, l'intensité mesurée en transmission en infrarouge est proportionnelle<sup>137,138</sup> au coefficient d'absorptivité molaire, de la concentration de l'espèce présente dans la zone d'analyse et des pas de balayage (suivant x et y).

En Raman<sup>87</sup>, l'intensité mesurée dépend, entre autres, de la polarisabilité moléculaire de l'espèce chimique, de l'intensité de la source émettrice et de la concentration de l'espèce chimique étudiée.

Après avoir mesuré tous les spectres souhaités, on les regroupe, selon l'ordre de mesure, dans une matrice de données spectrales. Cette matrice ( $x \times y \times \lambda$ ) est tridimensionnelle comme nous avons déjà expliqué au chapitre 1 (§I.1).  $n = x * y$  est le nombre total des spectres mesurés.  $\lambda$  est de dimension ( $1 \times m$ ),  $m$  étant le nombre de variables acquises lors de la mesure de chaque spectre. La Figure 1 : (cf. chapitre 1) donne une illustration de la « cartographie spectrale » microspectrométrique. L'enregistrement et le traitement des spectres sont réalisés numériquement par des moyens informatiques.

### 3.3. Production d'images spectroscopiques<sup>139 141</sup>: la manière classique et ses limitations

Pour rappel, chaque domaine spectroscopique a ses instruments et sa façon de produire les images. En ce qui nous concerne, notre domaine d'étude est l'infrarouge et le Raman. Nous nous limiterons dans les exemples et les explicitations aux procédés très spécifiques à ces deux techniques mais, d'une manière générale, la démarche peut être étendue aux autres types de spectroscopies.

Disposant de la matrice de données spectrales, après avoir acquis le nombre de spectres voulu, dans la démarche classique il faut connaître tous les constituants de l'échantillon étudié et disposer de leurs spectres de référence.

Pour produire l'image spectroscopique donnant la répartition cartographique sur la zone d'analyse de l'espèce chimique considérée, il faut pouvoir sélectionner sur son spectre une bande spectrale. Dans celle-ci, il faut qu'il n'y ait aucune contribution des autres composés présents dans l'échantillon. Les bandes de plus grande intensité sont souvent utilisées pour caractériser les constituants du produit étudié. Sur chaque spectre mesuré sur chacun des pixels, on calcule, à l'aide du logiciel informatique, l'aire sous le pic de cette bande<sup>140</sup>. Chaque aire calculée pour chaque pixel est proportionnelle à l'intensité mesurée sur ce pixel. Cette intensité est elle-même proportionnelle à la concentration de l'espèce présente sur ce même pixel.

En reportant cette valeur de l'intégrale (aire calculée) sur chaque pixel et en la codant à l'aide d'une échelle de couleurs, on visualise la répartition de la concentration du constituant ciblé sur l'ensemble des pixels d'analyse. En procédant de la même manière pour tous les constituants de l'échantillon, on obtient la cartographie globale de chacun d'eux. La Figure 29 illustre la démarche pour deux composés (A et B) fictifs sur une zone d'analyse de quatre pixels. On constate, après production de la cartographie, que le composé A est absent sur le pixel P<sub>2,1</sub> (totalement noir) et qu'il est le seul constituant du pixel P<sub>1,1</sub> (totalement blanc).

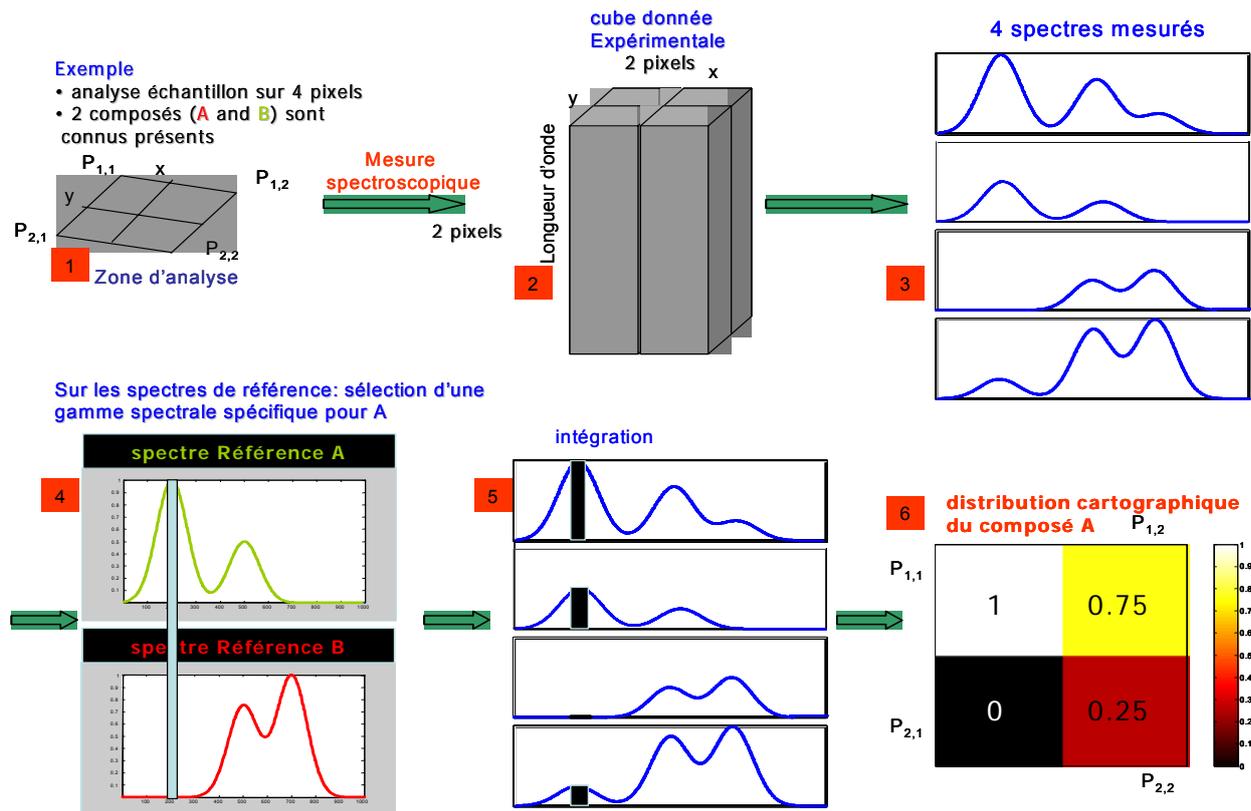


Figure 29 : Illustration de la production d'images spectroscopiques à partir des bandes spectrales par intégration.

Cette technique a donné des très bons résultats lorsqu'on connaît les constituants et qu'on recherche, par exemple, certaines imperfections telle que dans la fabrique des composants électroniques. Mais dans d'autres domaines, où d'une part le nombre de constituants est très

important (comme dans la pollution atmosphérique, la pollution des sol, des processus chimiques en évolution), et d'autre part les données acquises sont de plus en plus importantes, il est impossible de s'affranchir des recouvrements spectraux inextricables. Alors ces types de conditions réduisent l'intérêt de cette façon classique de produire les images.

Cette méthodologie classique de production d'images spectroscopiques soulève donc quelques critiques. D'abord il y a une nécessité de connaître tous les constituants du mélange et l'obligation d'attribuer une bande spectrale strictement spécifique à chacun d'eux. Cela nous amène à nous poser plusieurs questions.

Premièrement : pour une zone spectrale donnée, peut-on exclure la contribution de plusieurs autres constituants ? La réponse n'est pas aussi simple et cela risque d'avoir des incidences considérables sur la mesure de l'intensité du constituant considéré, sur sa concentration et enfin sur son image produite.

Deuxièmement : pouvons-nous facilement trouver des zones caractéristiques lors d'analyse d'échantillons complexes ? Les limites semblent donc importantes. D'autres questions restent en suspens.

Dans cette méthodologie classique, il apparaît clairement que la formation d'image est directement reliée à l'intensité mesurée, celle-ci étant elle-même proportionnelle à l'aire sous le pic considéré pour le composé spécifié. Lors de la mesure, la prise en compte de la contribution d'autres constituants n'est pas à exclure et donc l'intensité mesurée ne sera pas seulement celle du composé spécifié. Par conséquent la concentration obtenue est faussée ainsi que l'image formée grâce à celle-ci. Et toute éventuelle interprétation ou utilisation sera entachée par cette erreur

Cette incertitude est d'autant plus importante lorsqu'on analyse des échantillons industriels ou naturels complexes pour lesquels il est presque impossible de trouver des zones caractéristiques à cause des recouvrements spectraux toujours omniprésents.

La Figure 30 présente les spectres du cellophane pur (C), du polyéther-uréthane pur (D) ainsi qu'un mélange de ces deux produits. Ces deux composés sont tirés de l'exemple concret que nous allons étudier dans le chapitre suivant. Ainsi par exemple, afin d'extraire une information quantitative pour le produit D (polyéther-uréthane), il est nécessaire de trouver une zone spectrale caractéristique où il n'y a pas de contribution de C. La seule zone spectrale susceptible d'être sélectionnée sur le spectre de mélange pour le produit D se situe aux alentours de  $1650\text{ cm}^{-1}$  (zone grisée). On note néanmoins sur les spectres des produits purs qu'il existe une petite contribution de C au domaine spectral sélectionné. Cela induit donc irrémédiablement une surestimation de la concentration en D. Cette erreur se répercutera directement au niveau de l'image (distribution des concentrations) du composé D sur la zone d'analyse. De plus la

recherche d'une zone spectrale caractéristique pour C dans le spectre du mélange est encore plus délicate. On peut penser que les erreurs seront encore plus importantes. Cet exemple simple permet d'apercevoir les limites de l'imagerie infrarouge classique. De plus, lorsque les constituants du produit analysé sont inconnus, il est encore plus difficile de sélectionner une zone spectrale caractéristique pour un composé donné.

Toutes ces observations amènent les chimistes, et en particulier les chimiométriciens, à utiliser de plus en plus des outils de « curve modeling » pour mieux résoudre les problèmes rencontrés et d'extraire ainsi les spectres et les concentrations des produits purs dans les mélanges<sup>141,142</sup> sans *a priori*.

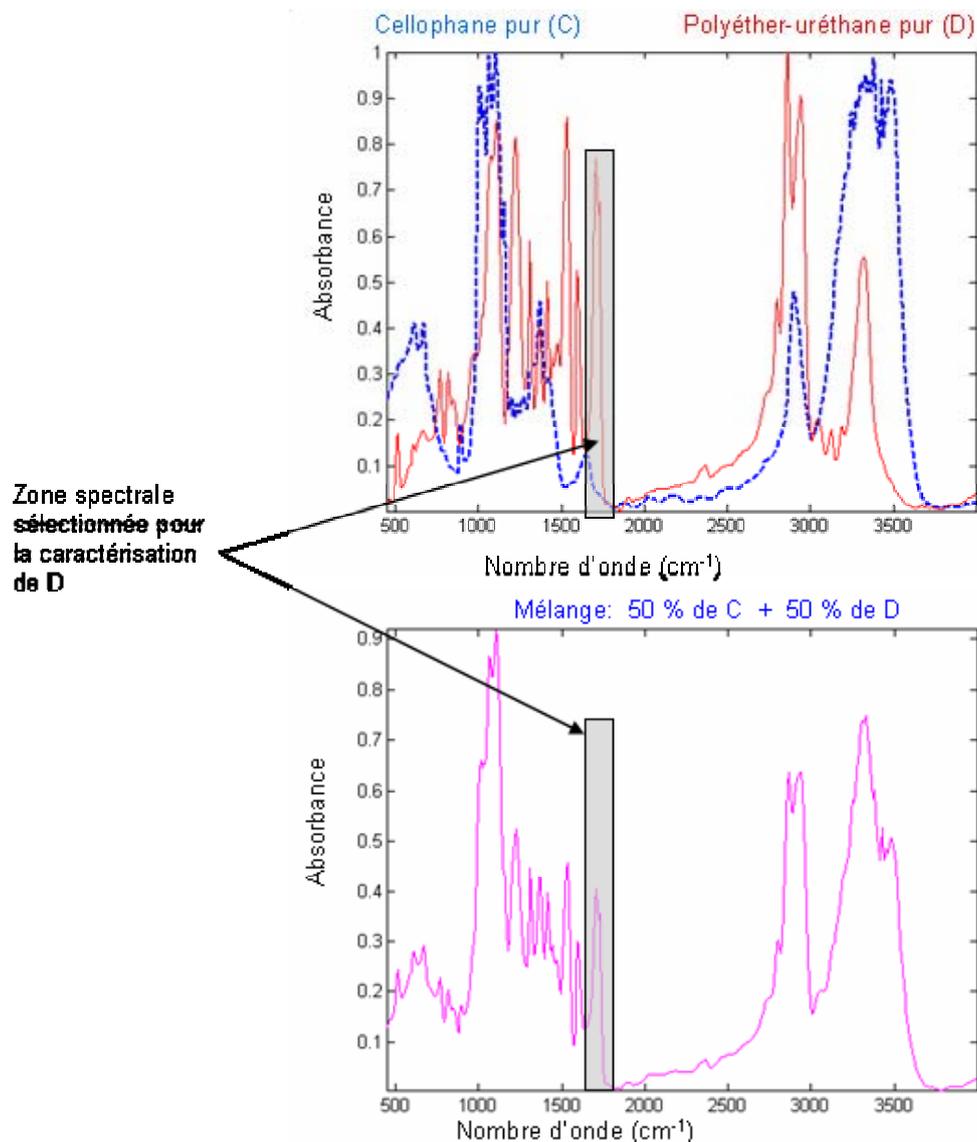


Figure 30 : Visualisation du mode de sélection d'une zone spectrale caractéristique pour le composé pur D.

Nous avons exposé les bonnes bases de compréhension à la fois sur la chimiométrie et l'imagerie microspectrométrique respectivement aux chapitres 1 et 2. Après avoir sélectionné nos méthodes chimiométriques d'étude, nous allons comparer l'efficacité d'extraction, c'est-à-dire la capacité de résolution, de ces méthodes sur une matrice de données de référence connues. En effet, nous allons aborder l'étude de ces méthodes chimiométriques sur une matrice de données synthétique que nous avons fabriquée.

Le chapitre 3 suivant présente les explications et les résultats de l'étude.

## **Chapitre 3 :**

# **Application des méthodes de résolution sélectionnées à une matrice de données synthétiques bien caractérisées : Test d'efficacité**

---

## **I. Introduction**

Au chapitre 1, nous avons exposé quelques problématiques sur les images et quelques méthodes utilisées pour leur résolution. Nous allons ici présenter celles qui sont l'objet de notre étude. Puis nous soumettrons ces méthodes de résolution à la même matrice de données spectrales. Le but de cette étude est d'optimiser ces méthodes de résolution en détectant les points de faiblesse et l'exactitude de la résolution de chacune d'entre elles.

En effet nous allons aborder avec un esprit critique l'application de ces méthodes de résolution aux données de l'imagerie microspectrométrique et nous allons chercher à montrer leurs comportements différenciés faces aux différentes conditions expérimentales. Les méthodes de résolutions que nous avons retenues sont toutes des méthodes linéaires appartenant au groupe « self-modeling ». La raison de ce choix est d'abord leur usage courant chez les chimométriciens, leur facilité d'utilisation (interactivité simple) et le nombre de contraintes à injecter souvent assez réduit.

En outre, dans la littérature, une synthèse ou une comparaison<sup>143,166,170</sup> des méthodes chimométriques ne porte souvent que sur un nombre assez réduit de méthodes. Cependant ces méthodes de résolution sont appliquées aux données de tous les domaines spectroscopiques. Le choix de la méthode de résolution à appliquer est souvent le résultat d'une démarche assez subjective. Au contraire, nous allons appliquer ces méthodes de résolution à une matrice de données synthétiques bien caractérisées que nous allons construire. Puis nous simulerons les perturbations expérimentales en ajoutant à cette matrice de données synthétiques différents niveaux de bruit blanc et en réalisant des décalages en nombres d'onde.

Après avoir défini un critère objectif d'évaluation pour caractériser la différence de comportement entre les méthodes, nous exposerons l'analyse assez détaillée dans trois cas de traitements différents qui sont :

- L'analyse des résultats de 6 méthodes : SIMPLISMA, OPA(var), OPA(spec), PCA, IKSFA et SPEXFA, en ne tenant compte que la contrainte de non-négativité, en fonction du bruit et des décalages en longueur d'onde. Elles sont qualifiées de l'adjectif « simples » parce qu'utilisées seules.
- Ces 6 méthodes couplées à MCR-ALS, c'est-à-dire en utilisant comme estimation initiale les résultats fournis par chacune de ces méthodes dans leur version « simple ». Nous verrons l'effet de ce couplage des méthodes avec MCR-ALS toujours en fonction du bruit et des décalages en longueurs d'onde. Nous introduirons aussi PMF dans la comparaison.
- L'application des contraintes du rang local aux couplages des méthodes avec MCR-ALS. Nous exposerons, dans ce cas, la méthodologie d'obtention de ces contraintes du rang local. Et de la même manière que les cas précédents, nous ferons l'analyse des résultats en fonction du bruit et des décalages en longueurs d'onde.

Enfin, dans un dernier paragraphe, nous ferons la synthèse de ces différents cas étudiés pour dégager la méthode la plus robuste parmi celles étudiées. Nous illustrerons notre argumentation par les images et les spectres extraits d'un des 6 composés de référence.

Après cette analyse, nous pourrions dire laquelle des méthodes est la plus apte à traiter les données de l'imagerie microspectrométrique puisque sa stabilité et son efficacité d'extraction auront été démontrées.

## II. Les méthodes chimiométriques de l'étude

Après une étude bibliographique, nous avons retenu pour notre étude les méthodes de résolution suivantes : PCA, SIMPLISMA, OPA, IKSFA, SPEXFA, MCR-ALS et PMF. Le Tableau 8 présente l'ensemble de ces méthodes.

Tableau 8 : Présentation des méthodes chimiométriques de notre étude.

Méthode	Qualité de la méthodologie	Itération basée sur
<b>SIMPLISMA</b>	Méthodologie d'extraction simple (MS)	Extraction des puretés successives
<b>PCA</b>	Méthodologie d'extraction simple (MS)	Extraction des composantes principales
<b>OPA(var)</b>	Mixte MS+MR	Extraction dissimilarités successives puis ALS
<b>OPA(spec)</b>	Mixte MS+MR	Extraction dissimilarités successives puis ALS
<b>IKSFA</b>	Méthodologie d'extraction simple (MS)	Extraction composantes spectrales (orthogonales)
<b>SPEXFA</b>	Méthodologie d'extraction simple (MS)	Extraction composantes spectrales (orthogonales)
<b>MCR-ALS</b>	Méthodologie d'extraction robuste (MR)	ALS
<b>PMF</b>	Méthodologie d'extraction robuste (MR)	ALS

L'essentiel de chacune de ces méthodes a été présenté au chapitre 1 (cf. §III). On trouvera dans l'annexe 1, le déroulement algorithmique et quelques indications sur leurs qualités et leurs défauts dans la résolution de données. OPA étant utilisé de deux façons différentes (OPA(var) et OPA(spec)), ces deux procédés sont considérés comme deux méthodologies d'extraction différentes.

La capacité d'extraction des images et des profils spectraux connus par chaque méthode est étudiée et une comparaison globale est réalisée.

Pour effectuer ces traitements d'une manière rigoureuse, il nous faut disposer des données de qualité. La partie suivante présente la méthodologie que nous avons utilisée pour obtenir une matrice de données synthétiques.

### III. Le choix de données synthétiques et la simulation des conditions expérimentales

Le choix des données synthétiques obéit à des critères objectifs. Nous devons disposer des données dont les paramètres sont bien contrôlables pour l'optimisation de l'ensemble de ces méthodes de résolution. Ces « données synthétiques » sont regroupées dans une matrice.

#### 1. Constitution de la matrice de données synthétiques

Notre démarche, dans ce cas de figure, est de construire une matrice de données synthétiques équivalentes aux données microspectrométriques acquises expérimentalement (cf. chapitre 4) et que cette matrice construite ne soit pas entachée d'erreur.

Pour cela, nous nous sommes référés aux bibliothèques spectrales disponibles. Dans la bibliothèque Hummel, bibliothèque spectrale des polymères, nous avons sélectionné 6 polymères qui sont le cellophane, le poly-acrylonitrile, le polyamide6, le polyesterterephthalate, le polyéther-uréthane, le fluorure de polyvinylidène. Chaque spectre de cette bibliothèque est enregistré pour un nombre de variables de 460 nombres d'onde dans l'intervalle  $[455, 4000 \text{ cm}^{-1}]$  avec une résolution de  $8 \text{ cm}^{-1}$ .

Ce choix est guidé par le recouvrement spectral très dense des spectres de ces 6 composés, sachant que la démarche est d'imposer cette complexité à l'ensemble des méthodes de résolution pour en juger leur qualité d'extraction.

Ensuite, nous avons associé à chacun des 6 composés une image (distribution de concentration) connue, déterminée et facilement reconnaissable. La Figure 31 montre les spectres de ces 6 composés et les images qui leur sont associées. Puis, nous avons constitué mathématiquement la matrice de données spectrales. Chaque image de référence a une dimension de 50 pixels sur 50. Chaque pixel d'une image présente une valeur de la concentration codée par une échelle de couleur comprise entre 0 et 255. Ces valeurs vont avoir un rôle important par la suite pour la caractérisation des images extraites.

Pour former cette matrice tridimensionnelle ( $50 \times 50 \times 460$ ) de données expérimentales, nous avons réalisé une combinaison linéaire de la matrice des concentrations représentant les 6 images et la matrice formée par les 6 spectres de 6 composés de référence.

La Figure 32 explique la démarche pour l'obtention de la matrice de données synthétiques. Chaque image de départ ayant une dimension carrée de 50 pixels sur 50, la matrice mélange des concentrations aura la même dimension en pixels ( $50 \times 50$ ).

Le spectre correspondant à chaque pixel de la matrice obtenue est une combinaison linéaire de 6 concentrations correspondant à ce pixel et de 6 spectres associés.

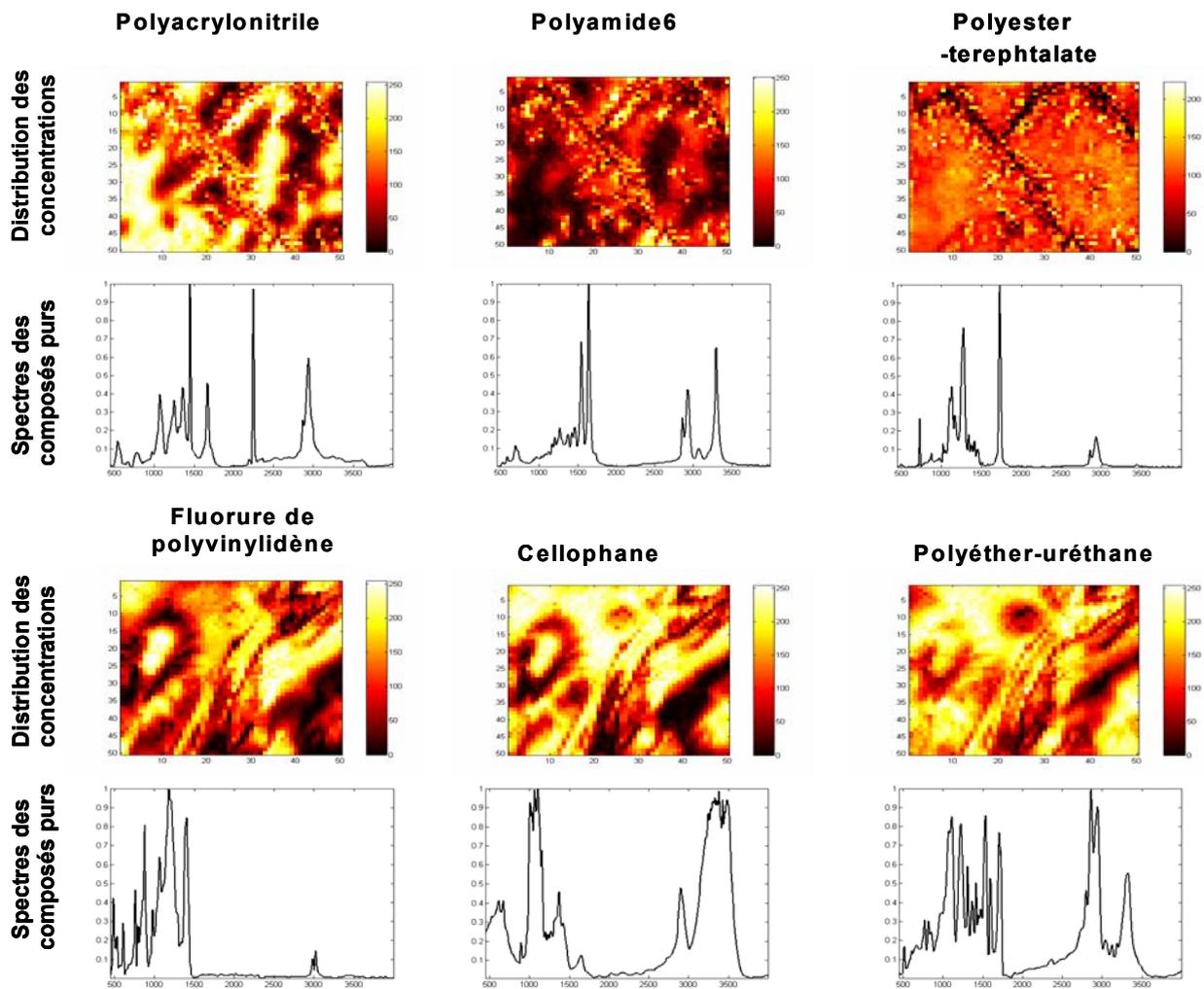


Figure 31 : Les cartes de distribution (images) et les spectres de 6 composés de référence.

Par exemple sur le 1<sup>er</sup> pixel de D, indexé  $D(1,1,\lambda)$ , on a la combinaison linéaire suivante :

$$D(1,1,\lambda) = I_1(1,1) * S_1(\lambda) + I_2(1,1) * S_2(\lambda) + I_3(1,1) * S_3(\lambda) + I_4(1,1) * S_4(\lambda) + I_5(1,1) * S_5(\lambda) + I_6(1,1) * S_6(\lambda).$$

Et pour un pixel (i,j) de D, i et j allant de 1 à 50, nous devons avoir :

$$D(i,j,\lambda) = I_1(i,j) * S_1(\lambda) + I_2(i,j) * S_2(\lambda) + I_3(i,j) * S_3(\lambda) + I_4(i,j) * S_4(\lambda) + I_5(i,j) * S_5(\lambda) + I_6(i,j) * S_6(\lambda).$$

I indique, comme on peut le voir sur la Figure 32, l'image considérée et S le spectre pur du composé associé à cette image.

Cela nous donnera en tridimension la matrice  $D_c(50 \times 50 \times 460)$ .

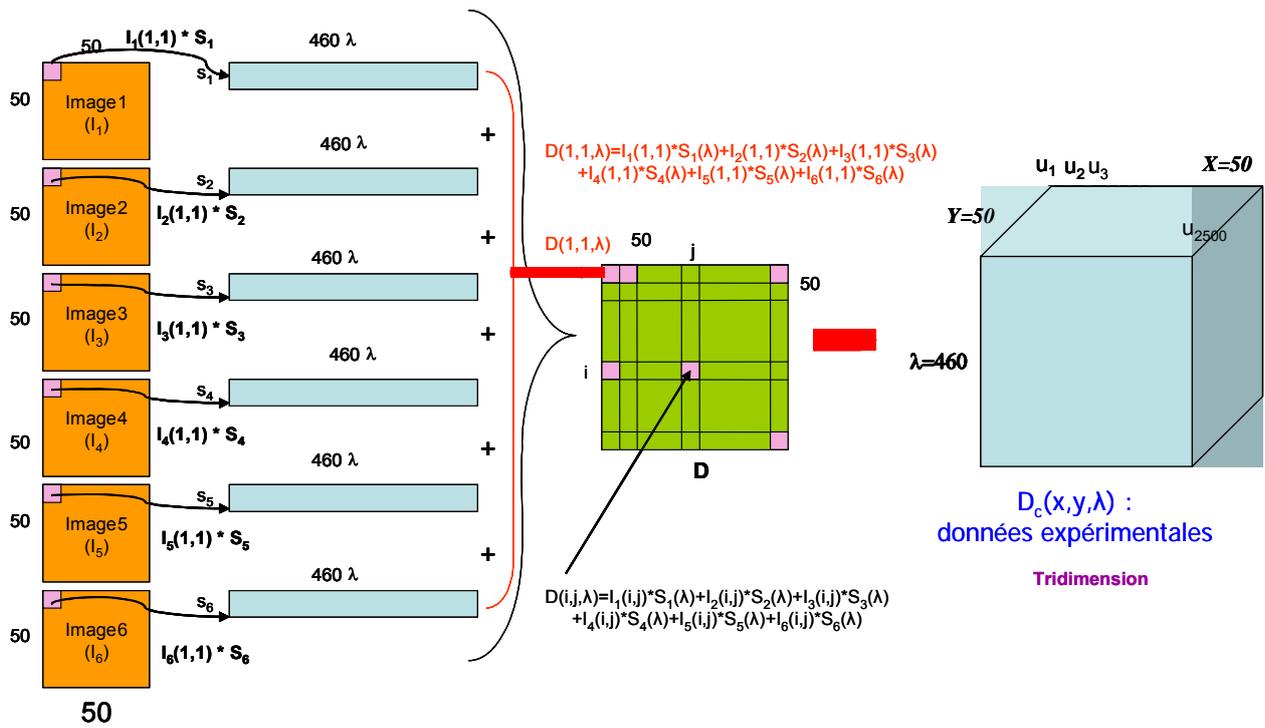


Figure 32 : La constitution de la matrice de données expérimentales,  $S_i$  ( $i=1:6$ ) spectre du composé pur,  $u_i$  ( $i=1:2500$ ) spectre de mélange formé par la combinaison linéaire des concentrations et des spectres purs de 6 composés de référence.  $I_i$  ( $i=1:6$ ) image associée au spectre  $S_i$ .

## 2. Manipulation de ces données synthétiques

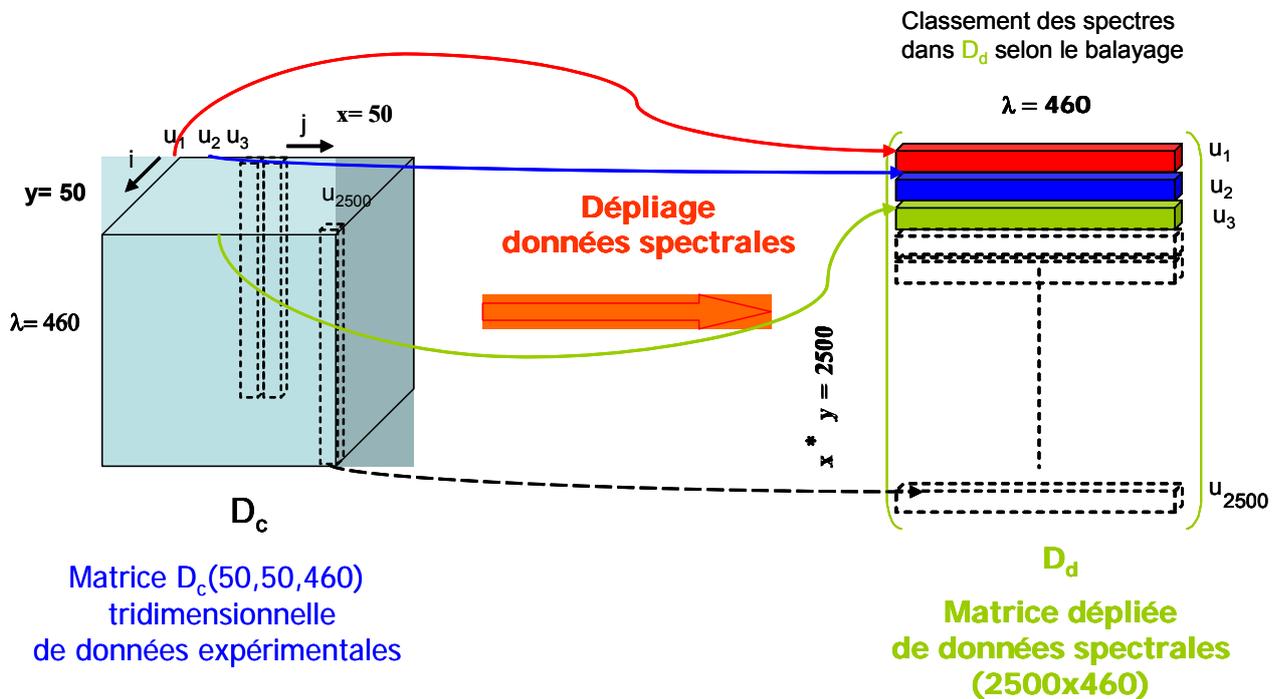


Figure 33 : La procédure de conversion de la matrice tridimensionnelle  $D_c(x,y,\lambda)$  en matrice dépliée bidimensionnelle  $D_d(x*y,\lambda)$  dans laquelle les spectres sont placés selon l'ordre de balayage.

Lors de l'utilisation de cette matrice de données spectrales tridimensionnelle dans les méthodes de résolution, un dépliage est opéré pour obtenir une matrice bidimensionnelle de  $50 \times 50$  spectres, soient 2500 spectres. Dans cette matrice bidimensionnelle, l'ordre des spectres est

respecté selon le sens du parcours de la zone d'analyse. La Figure 33 explique la démarche. Par exemple, le 1<sup>er</sup> spectre est celui correspondant au 1<sup>er</sup> pixel de la 1<sup>ère</sup> ligne de  $D_c$ , le 2<sup>ème</sup> spectre est celui correspondant au 2<sup>ème</sup> pixel de la même ligne et ainsi de suite. Le spectre du 1<sup>er</sup> pixel de la 2<sup>ème</sup> ligne est à la 51<sup>ème</sup> position dans la matrice dépliée que nous appellerons  $D_d$  et le 2500<sup>ème</sup> spectre, le dernier, est le spectre correspondant au dernier pixel de la 50<sup>ème</sup> ligne de  $D_c$ . Soit d'une manière générale, le spectre du pixel  $(i, j)$  dans  $D_c$  va donner le  $(50 * (i - 1) + j)$ <sup>ème</sup> spectre ligne dans  $D_d$ . Cela veut dire que lorsque les pixels de la première ligne ont tous été enregistrés, le microspectromètre se positionne sur le premier pixel de la deuxième ligne, et ainsi de suite.

Des détails techniques sur la manipulation et le traitement des données 3-D spectroscopiques sont relatés dans la littérature<sup>144,145</sup>.

### 3. La simulation des conditions expérimentales

La matrice bidimensionnelle  $D_d(2500 \times 460)$  est donc prête pour l'emploi. Mais pour être encore plus proche des conditions expérimentales réelles, nous avons décidé de réaliser certaines perturbations dans nos données expérimentales. Ces perturbations sont celles qu'on peut rencontrer naturellement lors des mesures spectroscopiques. Une des perturbations toujours omniprésente est le bruit<sup>146</sup>. Ce bruit peut avoir différentes origines<sup>87</sup> : une origine chimique (variation de la pression, variation de la température, variation du taux d'humidité de l'air...etc.) ou instrumentale (bruit thermique, bruit de grenaille, bruit de scintillation, bruit de l'environnement).

De ce fait, nous avons ajouté à la matrice de données spectrales  $D_d$  un bruit blanc ayant une distribution normale d'intensité de 1%, 3% et 10% par rapport au maximum pour simuler des conditions relativement simples et des conditions assez critiques au niveau de bruit. En effet, en imagerie spectroscopique<sup>147</sup>, le bruit est un phénomène très perturbant. Il faut concilier le besoin d'acquérir un nombre très élevé de spectres et la qualité spectrale de chaque spectre. Pour une bonne qualité spectrale, il faut un temps d'acquisition important. Cependant pour avoir un nombre élevé de spectres, la durée d'enregistrement d'un spectre ne doit être longue. Ce qui fait que le rapport Signal sur Bruit (S/B) est un critère important dont il faut tenir compte. Ce rapport est souvent l'un des principaux facteurs qui limitent la précision de la mesure. La Figure 34 montre un spectre de mélange avec les différents niveaux de bruit. Nous pouvons observer la nette différence entre le spectre de mélange bruité à 1% et celui bruité à 10%.

Pour se rapprocher davantage de la réalité (plus complexe), nous avons réalisé dans ces données synthétiques, en plus du bruit, un décalage en nombres d'onde effectué d'une manière aléatoire.

En effet, ce phénomène de décalage n'est plus observé en spectrométrie moyen infrarouge depuis l'utilisation de l'interférométrie mais reste néanmoins possible en spectrométrie proche infrarouge utilisant des instrumentations dispersives alors que l'imagerie de spectrométrie proche infrarouge est actuellement en pleine expansion pour la caractérisation de nombreux produits complexes<sup>148,149,150,145</sup>.

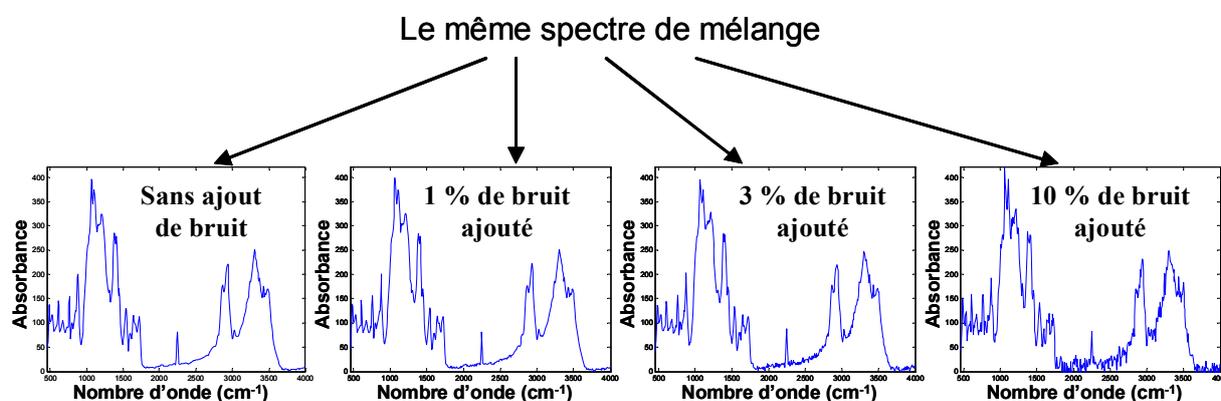


Figure 34 : Aspect d'un spectre de mélange en fonction de différents niveaux de bruits ajoutés.

Pour cette expérimentation, nous appliquerons des décalages aléatoires en nombres d'onde de 1 ou de 2 incréments spectraux vers les nombres d'onde croissants ou décroissants sur les spectres de la matrice de données spectrales  $D_d$ .

Cinq cas sont possibles :

- soit le spectre n'est pas décalé et donc reste identique à lui-même : 1<sup>er</sup> cas.
- Soit le spectre est décalé d'un incrément de nombres d'onde (1 fois la résolution spectrale) à gauche (2<sup>ème</sup> cas) ou à droite (3<sup>ème</sup> cas).
- Soit le spectre est décalé de 2 incréments de nombres d'onde (2 fois la résolution spectrale) à gauche (4<sup>ème</sup> cas) ou à droite (5<sup>ème</sup> cas).

La Figure 35 montre un spectre de mélange de données bruitées à 1% qui a subi un décalage d'un et de deux incréments de nombres d'onde vers les nombres d'onde croissants. Un incrément est égal à la résolution spectrale de  $8 \text{ cm}^{-1}$ . Nous avons zoomé une partie du spectre pour mettre en évidence l'effet de ces décalages sur les données à traiter.

En imposant ces perturbations assez importantes aux données expérimentales, nous voulons avoir un jugement d'ensemble qui nous permettra de tirer les conclusions nécessaires quant à la robustesse de ces méthodes face à de telles perturbations.

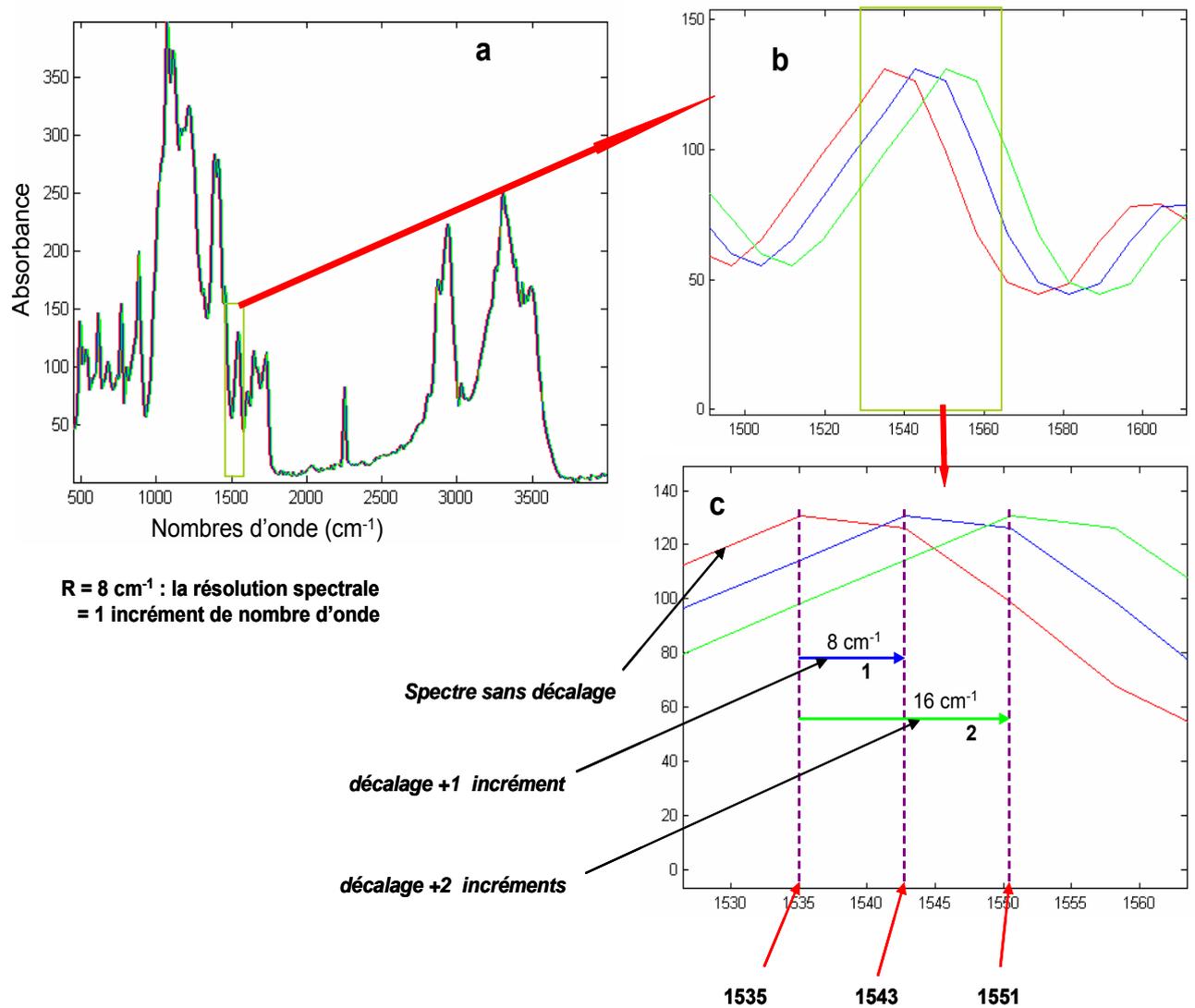


Figure 35 : Illustration de l'effet des décalages sur le premier spectre de  $D_a$ . a) les trois spectres (spectre sans décalage (rouge) et spectres avec des décalages de 1 (bleu) et 2 (vert) incréments de nombres d'onde) ; b) et c) zoom pour mieux visualiser ces décalages.

## IV. La méthodologie générale de résolution

La démarche de résolution d'une manière générale suit le schéma du principe général de résolution défini dans le paragraphe IV du chapitre 1 (cf. Figure 14).

### 1. Critère de comparaison et d'évaluation

Pour un jugement objectif, il nous faut un critère efficace de comparaison qui nous permettra d'apprécier les images extraites entre elles et les spectres extraits entre eux. Ce critère de comparaison que nous allons utiliser est l'indice de dissimilarité. Nous définirons ce critère et il nous permettra de comparer l'élément extrait et sa référence. Cet indice de dissimilarité est défini à partir du coefficient de corrélation et s'exprime ainsi :  $dis(X, Y) = \sqrt{1 - cor^2(X, Y)}$  (éq48). Ce coefficient de corrélation ( $cor$ ) est lui-même défini à partir de la covariance et des écarts-types des éléments X et Y, vecteurs à N dimensions. X représente l'élément de référence (spectre ou image) et Y l'élément extrait (spectre ou image) par la méthode de résolution.

Soit donc :  $cor(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$  (éq49) avec  $cov$  la covariance des variables X et Y,  $\sigma_X$  et  $\sigma_Y$

les écarts-types des variables X et Y ; avec  $\sigma_X = \sqrt{\sum_{i \in N} \frac{(X_i - \bar{X})^2}{N-1}}$  et  $\sigma_Y = \sqrt{\sum_{i \in N} \frac{(Y_i - \bar{Y})^2}{N-1}}$ .  $\bar{X}$  et

$\bar{Y}$  sont les valeurs moyennes des variables X et Y. La covariance  $cov$  des variables X et Y se calcule de la manière suivante :  $cov = \sum_{i \in N} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{N-1}$  (éq50). Les variables ont le même

poids.

Comme on peut s'attendre, un coefficient de corrélation élevé (proche de 1) donne un indice de dissimilarité faible (proche de 0) et indique une forte similitude de l'élément extrait avec la référence correspondante. Inversement, un indice de dissimilarité élevé signifie une forte dissemblance des deux éléments.

### 2. Comparaison globale

L'indice de dissimilarité que nous venons de définir nous permettra de faire une comparaison générale. Chaque image ou spectre extrait va être comparé à sa référence.

Mais pour pouvoir comparer les éléments extraits par une méthode à ceux extraits par les autres méthodes, nous sommes amenés à faire une moyenne des dissimilarités sur les 6 composés pour chaque méthode. Cette valeur de dissimilarité moyenne pour chaque méthode permettra de faire un jugement sur la capacité d'extraction de ces méthodes de résolution entre elles.

La Figure 36 visualise la procédure générale à suivre dans cette étape de résolution et de comparaison.

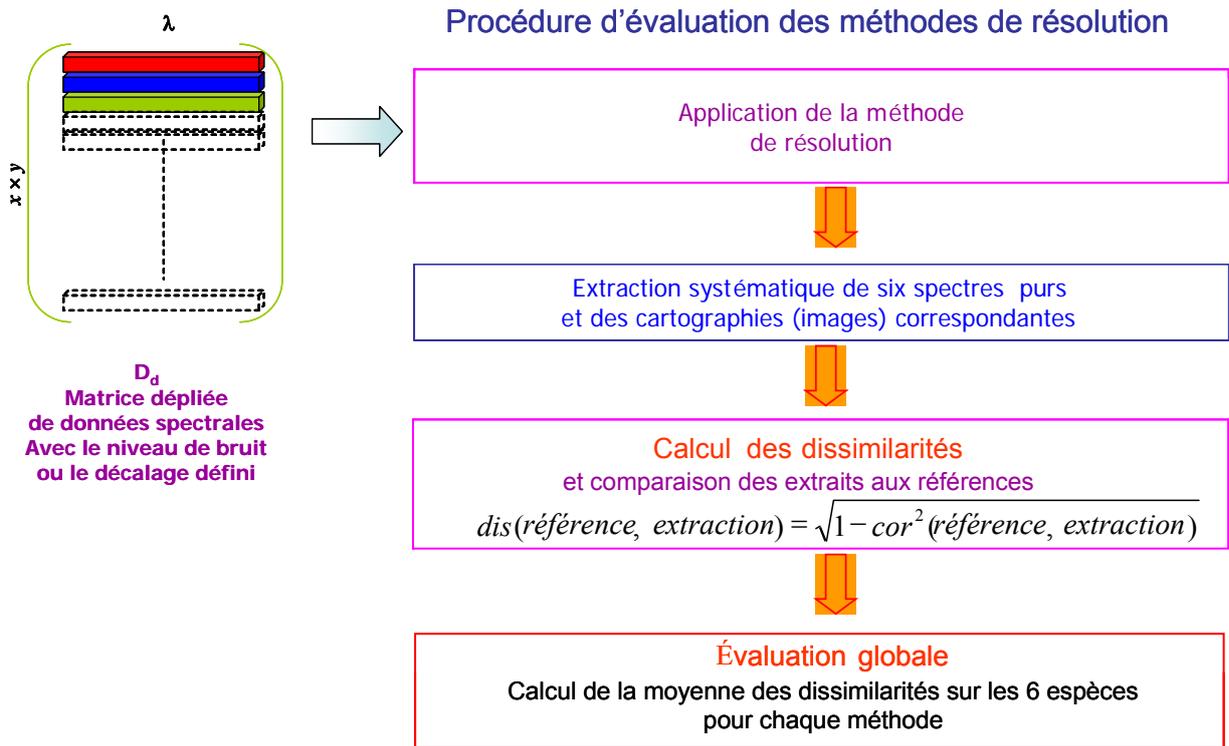


Figure 36 : Critère de comparaison entre les éléments extraits et leurs références. Evaluation des méthodes de résolution entre elles.

## V. Application des méthodes de résolution sélectionnées aux données synthétiques

Avant présenter les résultats de cette étude, nous présentons ci-dessous l'étude réalisée pour SIMPLISMA en ce qui concerne son offset.

### 1. Etude du critère particulier de SIMPLISMA : l'influence de l'offset

La pureté calculée par SIMPLISMA est fonction de la déviation standard, de la moyenne et de l'offset (cf. chapitre 1 §III.3.1). L'offset agit sur la moyenne. Mais quelle est la meilleure valeur de l'offset à choisir pour un type de données ? Pour apporter une réponse à cette question dans notre cas, nous avons entrepris une étude. Le Tableau 9 montre les 6 variables pures extraites dans l'ordre par SIMPLISMA à partir de la matrice de données  $D_d$  bruitée à 10%.

**Tableau 9 : Extraction par SIMPLISMA de 6 variables les plus pures pour chaque offset.**

offset		variable 1	variable 2	variable 3	variable 4	variable 5	variable 6
1		434	431	423	432	433	424
3		394	434	431	423	424	432
10		80	233	155	423	434	424
30		80	233	155	166	56	394
50		80	233	155	166	56	394
80		80	233	155	166	95	394
100		80	233	155	166	95	394
150		80	233	155	166	95	313
200		80	233	155	166	95	313

Nous remarquons, en passant de l'offset 1 à l'offset 3, que l'ordre des variables pures extraites change. A partir de 10, il y a une stabilité de l'ordre d'extraction pour un bon nombre de variables. L'analyse comparative est réalisée pour la première variable pure extraite car celle-ci est calculée par une relation simple (cf. éq22).

Le Tableau 10 présente les résultats de l'étude de cette première variable pure en fonction de l'offset. La moyenne de chaque variable « nombre d'onde » extraite et sa déviation standard sont indiquées. Nous remarquons lorsqu'une variable extraite possède une moyenne faible pour un offset petit (cas offset=1), on obtient une grande valeur de pureté et cette variable est considéré comme étant la plus pure. Cela fausse le processus d'extraction de SIMPLISMA. L'erreur commise sur la première variable se répercute sur la suite d'extraction. Cette remarque reste valable pour l'offset de 3. Dans ce Tableau 10, la première variable pure extraite à partir de l'offset 10 est la variable numéro 80.

Tableau 10 : Etude de l'évolution du rapport offset/moyenne pour la première variable pure extraite par SIMPLISMA.

offset		1	3	10	30	50	80	100	150	200
Première pure variable extraite	N°variable	434	394	80	80	80	80	80	80	80
	moyenne	2.6272	152.2664	387.6488	387.6488	387.6488	387.6488	387.6488	387.6488	387.6488
	std	10.0531	75.7141	143.1628	143.1628	143.1628	143.1628	143.1628	143.1628	143.1628
	offset/moyenne	0.3806	0.0197	0.0258	0.0774	0.1290	0.2064	0.2580	0.3869	0.5159

On peut donc considérer que l'offset choisi est bon lorsqu'on obtient une stabilité de l'ordre d'extraction des variables pures pour un jeu d'offset. Dans notre étude, il ressort qu'un offset compris entre 10 et 50 est suffisant. En effet, la moyenne de cette variable 80 est égale à 387,6488 et un offset de 10 à 50 n'a qu'une influence limitée sur le calcul de la pureté. Par conséquent, pour ces offsets, on arrive à extraire correctement la variable la plus pure.

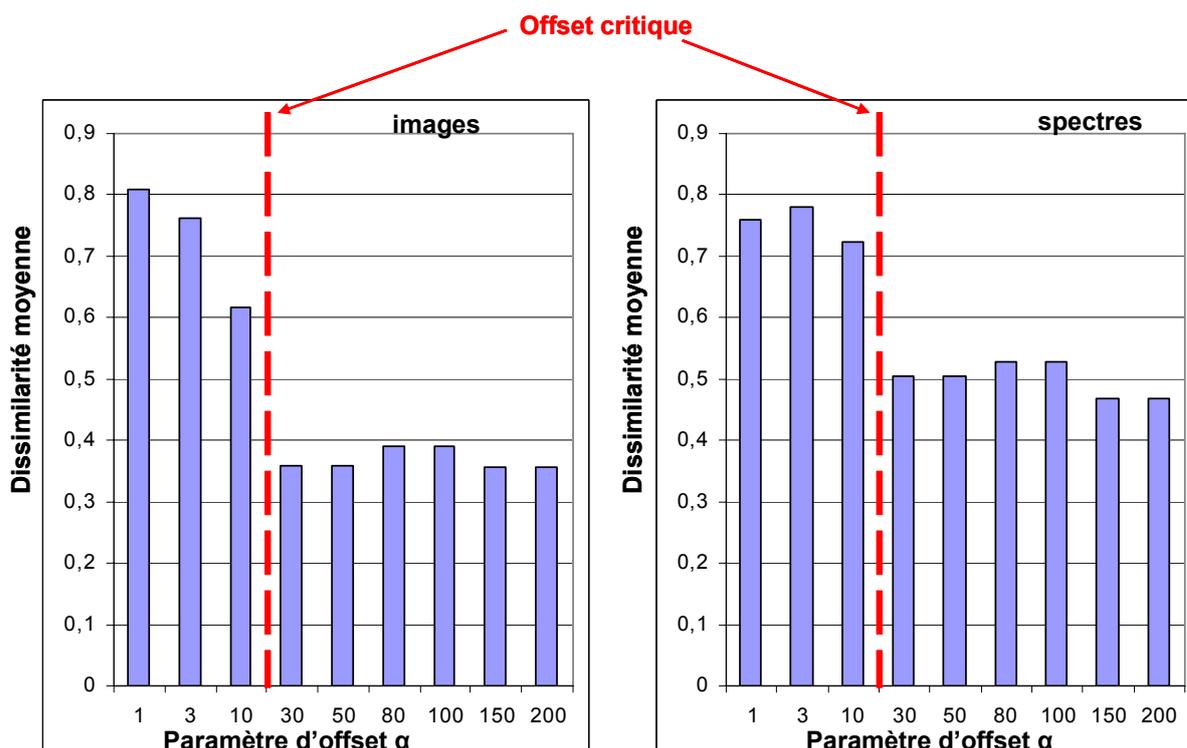


Figure 37 : Influence du paramètre d'offset  $\alpha$  sur la capacité de résolution de l'algorithme SIMPLISMA. Le meilleur offset dans notre cas est compris entre 10 et 50.

En outre, l'étude faite sur les 6 composés de référence confirme ce résultat. Nous avons réalisé l'extraction des spectres et des images pour les 6 composés de référence pour les offsets de l'étude. La moyenne des dissimilarités est calculée pour chaque offset. Le résultat est présenté à

la Figure 37 sur laquelle la dissimilarité moyenne est fonction de l'offset choisi. Nous notons que cette étude sur la moyenne globale exclut même l'offset de 10 du bon intervalle à retenir. Nous observons que l'offset de 3 préconisé en général dans de nombreuses publications<sup>151,152</sup> ne donne pas, et de loin, les meilleurs résultats.

N'existant pas de procédure rigoureuse pour le choix de ce paramètre, cette nouvelle limitation s'ajoute à certaines faiblesses de SIMPLISMA comme le manque des contraintes de non-négativité.

## 2. Etude de l'influence du rapport signal sur bruit sur la capacité d'extraction des méthodes de résolution avec la seule contrainte de non-négativité sur les spectres et les concentrations

Pour être toujours plus prêt des conditions expérimentales, les seules contraintes que nous allons imposer aux méthodes de résolution sont la non-négativité sur les concentrations et les spectres.

En microspectroscopie, il faut que le rapport (Signal/Bruit) soit le meilleur<sup>153</sup> possible (le plus grand possible) pour avoir un bon signal exploitable par la suite pour la production d'images par exemple, ou pour la caractérisation des produits chimiques sous analyse. Nous allons étudier l'influence du bruit sur la qualité d'extraction des méthodes chimiométriques dans les différents cas expérimentaux (1%, 3% et 10%) que nous avons simulés. Une illustration d'un bruit blanc avec une distribution normale peut être trouvée dans l'article de P. Barone<sup>154</sup>.

Après avoir appliqué la procédure d'extraction et d'évaluation, l'étude comparative des résultats obtenus pour l'ensemble des méthodes de l'étude est présentée dans les parties suivantes.

### 2.1. Etude comparative des capacités d'extraction des méthodes de résolution « simples », MS

#### 2.1.1. Etude de la résolution en fonction du bruit sans décalage en nombres d'onde

Les résultats du traitement de cette analyse sont présentés à la Figure 38. Comme nous observons sur cette Figure 38, nous obtenons un comportement différent pour les images et les spectres.

Pour les images, deux méthodes se distinguent par leur mauvaise qualité d'extraction mise en évidence par une valeur de dissimilarité moyenne élevée. Ces deux méthodes sont OPA(spec) et PCA avec des valeurs de dissimilarité supérieures ou égales à 0,8. Les quatre autres méthodes SIMPLISMA, OPA(var), IKSFA et SPEXFA manifestent des valeurs de dissimilarité relativement faibles montrant une meilleure qualité d'extraction des images.

Cependant l'extraction des images semble ne pas être perturbée par le niveau de bruit. A l'exception de SIMPLISMA et de SPEXFA, toutes les autres méthodes donnent pratiquement les mêmes valeurs de dissimilarité pour les trois niveaux de bruit étudiés.

Au faible niveau de bruit, SIMPLISMA reste l'une des meilleures mais montre une sensibilité plus grande face au bruit. Car la variation de la dissimilarité moyenne est significative en passant de 0,2 à 1% à environ 0,4 pour le niveau de bruit de 10%.

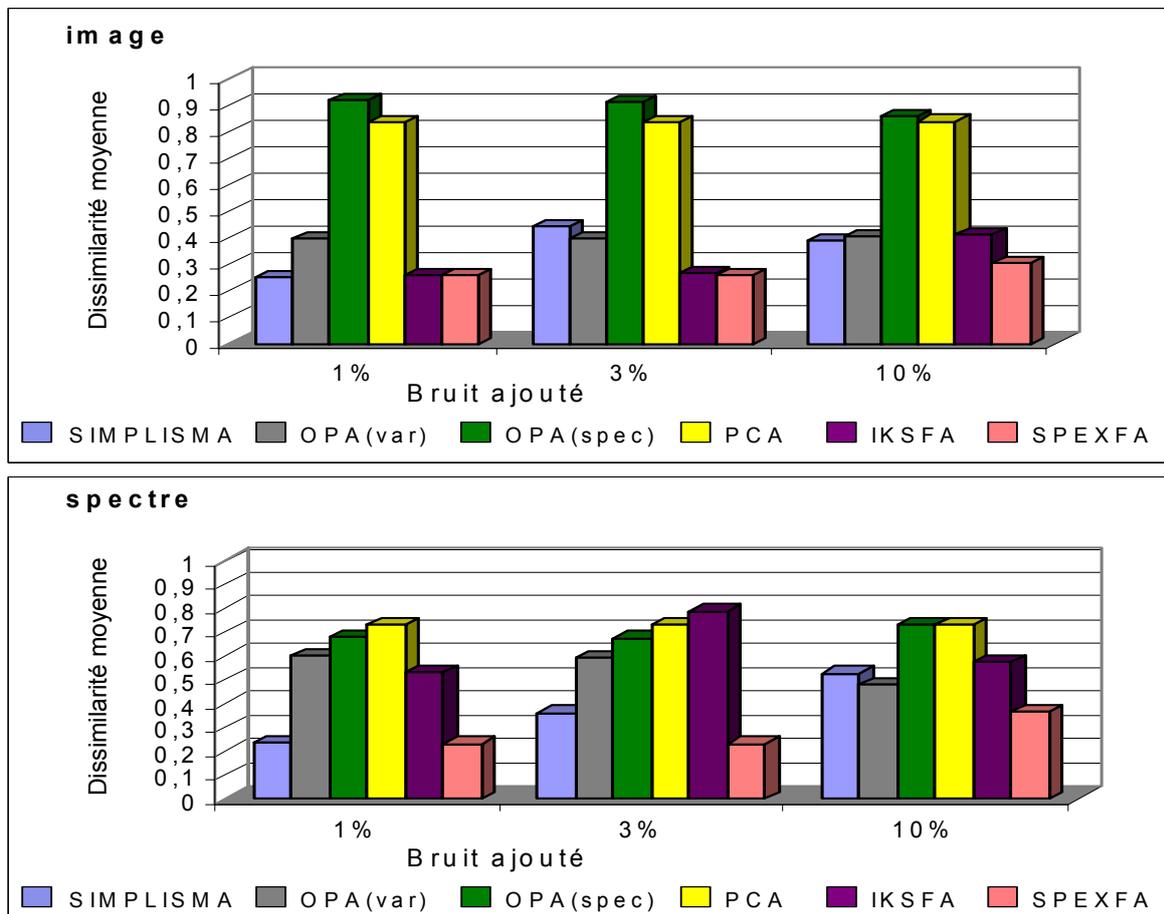


Figure 38 : Comparaison globale de 6 méthodes « simples » pour les images (cartes de concentrations) et les spectres en fonction de la dissimilarité moyenne globale.

Quant aux spectres, la plupart des méthodes donnent une valeur de dissimilarité comprise entre 0,4 et 0,8. Ce qui indique une plus grande perturbation pour l'extraction des spectres.

SIMPLISMA et SPEXFA produisent les plus faibles valeurs de dissimilarité moyenne pour les faibles niveaux de bruit (1% et 3%). Cependant pour SIMPLISMA, plus le niveau de bruit augmente plus la perturbation est grande. Comme pour les images, les deux méthodes OPA(spec) et PCA sont les plus perturbées du fait des valeurs de dissimilarité moyenne toujours supérieures 0,7.

### 2.1.2. Etude de la résolution en fonction des décalages en nombres d'onde pour un niveau de bruit donné

Nous étudions les décalages en nombres d'onde pour un niveau de bruit donné. Comme dans le paragraphe précédent, nous obtenons aussi dans ce cas présent les deux comportements caractéristiques : celui des images et celui des spectres. Les résultats sont présentés à la Figure 39. Ces illustrations concernent les deux niveaux de bruit de 1% et de 10%.

**Pour les images**, quel que soit le niveau de bruit, le comportement des méthodes d'extraction est le même. Les deux méthodes OPA(spec) et PCA possèdent toujours les plus grandes valeurs de dissimilarité moyenne et donc possèdent les plus mauvaises qualités d'extraction.

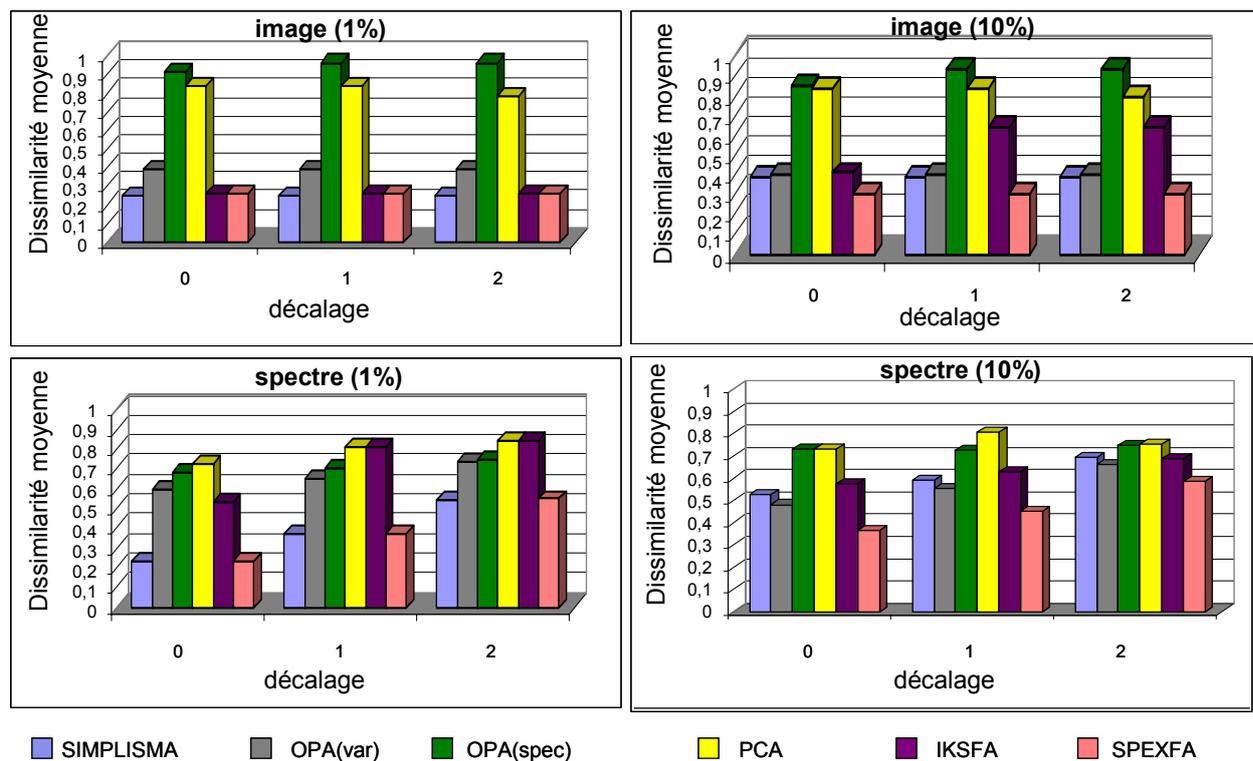


Figure 39 : Etude de la capacité d'extraction des méthodes « simples » en fonction des décalages pour un niveau de bruit donné (à gauche 1%, à droite 10%).

Pour le niveau de bruit de 1%, quel que soit le décalage, toutes les autres méthodes ont une valeur de dissimilarité moyenne comprise entre 0,2 et 0,4. Pour le niveau de bruit de 10%, nous observons une augmentation sensible de la dissimilarité moyenne pour toutes les méthodes puisque cette fois-ci elle est plutôt proche de 0,4.

Pour un niveau de bruit donné, les images ne sont pas sensibles aux décalages, et on obtient pratiquement les mêmes résultats quel que soit le décalage. Ce comportement est bien visible sur la Figure 39. Ce phénomène assez remarquable reste pour l'instant inexpliqué.

**Quant aux spectres**, pour les deux niveaux de bruit considérés, lorsque le décalage augmente les valeurs de dissimilarité deviennent plus importantes. Ce qui est synonyme de grande

perturbation sur la qualité d'extraction de ces méthodes. Cependant OPA(spec) et PCA restent les plus perturbées. Pour un faible niveau de bruit (1%), SIMPLISMA et SPEXFA ont les meilleurs qualité d'extraction quel que soit le décalage mais manifestent une sensibilité plus grande face aux décalages. Pour le niveau de bruit de 10%, aucune méthode ne se démarque d'une manière assez significative.

Pour faire la synthèse, nous pouvons dire :

- les images ne sont pas sensibles aux décalages en nombres d'onde pour un niveau de bruit donné. Mais comparativement, pour un niveau de bruit élevé, on observe une sensibilité relative.
- Quant aux spectres, pour un faible niveau (1%), le comportement est similaire et toutes les méthodes restent légèrement sensibles aux décalages. Pour un bruit assez important, la perturbation due aux décalages sur la qualité d'extraction est assez notable d'une manière générale. Mais celle-ci se manifeste plus sur les méthodes ayant une meilleure qualité d'extraction au faible niveau de bruit comme SIMPLISMA ou SPEXFA.

### 2.1.3. Quelques explications sur la mauvaise qualité d'extraction de certaines méthodes de résolution

Une remarque importante s'impose au stade actuel de notre étude. Il ressort sans équivoque que OPA(spec) et PCA sont les deux méthodes qui possèdent la plus mauvaise qualité d'extraction. Et cela, quel que soit le niveau de bruit imposé et le décalage réalisé. Cela nécessite une analyse pour expliquer ces comportements observés.

#### 2.1.3.1. OPA(spec)

Pour mieux apprécier les défauts d'extraction de OPA(spec), nous allons comparer ses résultats à ceux extraits par OPA(var) et cela pour le même niveau de bruit de 1% et sans réaliser des décalages en nombres d'onde. Mais d'abord analysons le fonctionnement de OPA(spec).

Comme on peut le voir dans le Tableau 11A, OPA(spec) cherche à extraire, à partir de la matrice de données de mélange, dépliée (bidimensionnelle), les spectres purs correspondant aux 6 composés de référence. Rappelons que les spectres de mélange sont obtenus par une combinaison linéaire de 6 spectres et des cartes de concentrations associées (images) de 6 composés de référence. Ce qui veut dire en clair dans ce cas présent qu'il est pratiquement impossible de trouver un spectre totalement pur correspondant à un composé de référence.

Dans le Tableau 11A, nous avons les numéros de 6 spectres extraits par OPA(spec) et considérés les plus purs. Ensuite, nous avons récupéré, à partir de la matrice référence des concentrations, les valeurs de concentrations correspondantes. C'est ce qu'indiquent les nombres dans le Tableau 11A qui sont compris entre 0 et 255 car chaque valeur de la concentration est codée par une

coloration (entre 0 et 255). Ces valeurs dans le Tableau 11A sont en unité arbitraire. Pour que chaque spectre extrait corresponde à un composé de référence, il faut qu'une valeur de concentration domine d'une manière significative les 5 autres valeurs.

**Tableau 11 : Comparaison de la qualité d'extraction de OPA(spec) (tableau A) et de OPA(var) (tableau B). Comparaison des composés extraits par chacune des méthodes à ceux de la référence.**

		concentrations références (unité arbitraire)					
		Poly(acrylo-Nitrile)	Polyamide 6	Polyester terephthalate	Fluorure de Polyvinylidène	Cellophane	Polyéther-uréthane
<b>A</b>	Spectre le plus pur sélectionné						
	#2037	255	136	196	23	6	86
	#1663	116	94	81	212	255	255
	#450	255	192	127	69	226	113
	#2198	33	129	115	19	73	233
	#1978	127	201	200	196	151	66
	#2340	118	32	133	84	233	229

		Corrélations entre les variables sélectionnées et les concentrations références					
		Poly(acrylo-Nitrile)	Polyamide 6	Polyester terephthalate	Fluorure de Polyvinylidène	Cellophane	Polyéther-uréthane
<b>B</b>	Variable la plus Pure sélectionnée						
	#233	0.99	0.01	0.08	0.00	0.01	0.02
	#85	0.04	0.01	0.01	0.48	0.90	0.58
	#155	0.01	0.86	0.10	0.06	0.14	0.05
	#166	0.12	0.03	0.44	0.02	0.21	0.57
	#95	0.02	0.02	0.01	0.86	0.76	0.22
	#141	0.01	0.16	0.03	0.06	0.44	0.78

En analysant le Tableau 11A, nous pouvons dire que le spectre n°2198 correspond au polyether-uréthane car la valeur de la concentration relative de 233 domine toutes les autres. Il en est de même pour le spectre n°2037 qui correspond au polyacrylonitrile, mais on peut remarquer qu'il possède une valeur de concentration de 196 sur le polyesterterephthalate non négligeable devant 255. L'ambiguïté est encore plus évidente pour les 4 autres composés. On obtient pour chacun d'eux au moins 2 valeurs de concentrations assez proches, d'où une ambiguïté de choix entre 2 composés candidats. Cela met en évidence la mauvaise qualité d'extraction et de sélection de OPA(spec) dans ce cas d'étude.

Quant à OPA(var), la recherche est faite suivant les variables « nombres d'onde » dont les valeurs sont suivant les colonnes. OPA(var) recherche dans la matrice de données dépliée s'il existe une variable pour laquelle on a uniquement un composé. Et puis OPA(var) sélectionne cette variable comme étant pure et caractéristique de ce composé du mélange. Une fois ces variables caractéristiques sélectionnées, nous avons récupéré dans la matrice dépliée  $D_d$  bruité à 1%, les valeurs d'absorbance correspondantes. Puis nous avons calculé le coefficient de

détermination ( $R^2 = r^2$ ) entre chaque variable extraite et la concentration référence de chacun des 6 composés de référence. Les résultats sont regroupés dans le Tableau 11B.

Tableau 12 : Attribution des variables extraites par OPA(var) et leur caractérisation en nombre d'onde.

Nombre d'onde (cm <sup>-1</sup> )	Variable (numéro)	composé	Coefficient de corrélation (variable/composé)
<b>2245</b>	<b>233</b>	<b>Polyacrylonitrile</b>	<b>0.99</b>
<b>1103</b>	<b>85</b>	<b>Cellophane</b>	<b>0.90</b>
<b>1643</b>	<b>155</b>	<b>Polyamide 6</b>	<b>0.86</b>
<b>1728</b>	<b>166</b>	<b>Polyesterterephthalate</b>	<b>0.44</b>
<b>1180</b>	<b>95</b>	<b>Fluorure de Polyvinylidène</b>	<b>0.86</b>
<b>1535</b>	<b>141</b>	<b>Polyéther-uréthane</b>	<b>0.78</b>

Dans ce Tableau 11B, nous observons que chaque variable extraite possède un fort coefficient de détermination sur un composé, les 5 autres coefficients étant souvent assez négligeables devant celui-ci. Le coefficient de détermination sur le composé correspondant est assez élevé ( $> 0,78$ ) à l'exception pour le polyesterterephthalate pour lequel on a 0,44. Mais cette valeur de 0,44 est nettement supérieure aux autres coefficients qui sont négligeables devant celle-ci (suivant la colonne dans le Tableau 11B). Le Tableau 12 donne la correspondance entre les variables extraites par OPA(var) et les composés correspondants. Les coefficients de détermination relatifs et la variable « nombre d'onde » caractéristique associée à chaque variable extraite sont indiqués aussi. Nous observons, à l'exception du polyesterterephthalate, un coefficient de détermination élevé entre la variable extraite et le composé de référence correspondant.

#### 2.1.3.2. PCA

On rappelle que l'analyse en composante principale (PCA), lors de la décomposition de la matrice de données, recherche les principaux axes factoriels sur lesquels la variance des variables étudiées est maximale. Les spectres extraits sont les projections des variables sur ces axes principaux et obéissent à la propriété mathématique d'orthogonalité inhérente à PCA (l'angle entre deux spectres extraits est égal à 90°).

Pour appuyer l'argumentation, nous avons calculé les angles entre les spectres purs de référence, chaque spectre étant un vecteur à 460 composantes. Le Tableau 13 présente les résultats. Ce tableau est bien entendu symétrique avec une diagonale nulle. Nous observons dans ce tableau que les angles entre les spectres purs de référence sont loin d'être orthogonaux, car ils sont tous compris entre 47° et 74°. PCA ne respecta pas donc les angles entre les spectres de référence.

Cela induit forcément un écart entre les spectres de référence et ceux extraits. C'est ce qui expliquerait en grande partie la mauvaise qualité de ces derniers et probablement une perte de certaines propriétés spectrales dans le processus de transformation.

Tableau 13 : Angles (exprimés en degré) entre les spectres purs de 6 composés de référence.

	Polyacrylonitrile	Polyamide 6	Polyesterterephthalate	Fluorure de Polyvinylidène	Cellophane	Polyéther-uréthane
Polyacrylonitrile	0	57	63	63	63	47
Polyamide 6		0	72	74	61	48
Polyesterterephthalate			0	59	71	56
Fluorure de Polyvinylidène				0	63	57
Cellophane					0	49
Polyéther-uréthane						0

Et de ce fait PCA, dans notre étude et compte tenu de la matrice de données utilisée, est l'une des méthodes la plus inapte à extraire les spectres corrects des composés de référence. Dans la plupart des études, PCA est utilisé pour estimer le nombre de constituants d'un mélange<sup>155</sup> au même titre que la SVD<sup>156,157\*</sup> (singular value decomposition).

## 2.2. Etude comparative des capacités d'extraction des méthodes de résolution « simples » raffinées par MCR-ALS

### 2.2.1. Etude de la résolution en fonction du bruit sans décalage en nombres d'onde

Une fois étudiée les comportements des 6 méthodes qualifiées « simples », dans cette partie, nous analyserons le couplage de ces méthodes avec MCR-ALS. Nous mettrons en évidence la pertinence de cette association dont nous sommes l'un des pionniers<sup>158</sup>. MCR-ALS est une procédure itérative assez évoluée. Comme nous l'avons expliqué au chapitre 1 (Cf. §IV.3), MCR-ALS a besoin d'une estimation initiale. Cette estimation peut concerner soit les spectres, soit les concentrations. Chacune de nos méthodes « simples » a fourni la matrice des concentrations et celle des spectres. Nous fournirons à MCR-ALS comme estimation initiale l'une des 2 matrices estimées. Pour l'écriture, par exemple pour OPA(var), nous noterons ce couplage OPA(var)/MCR-ALS. Cette démarche est visualisée par le diagramme de la Figure 40. En plus de cela, nous allons comparer les résultats fournis par ces 6 couplages avec celui fourni par PMF qui utilise une procédure itérative analogue à MCR-ALS.

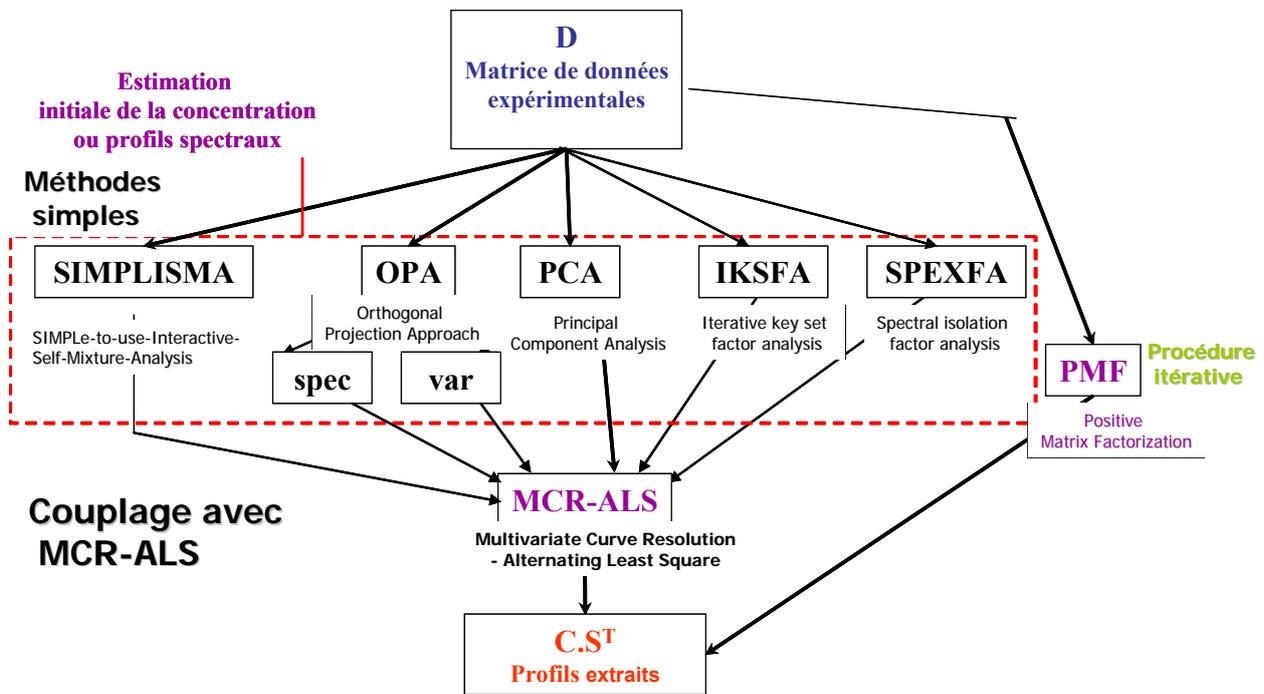


Figure 40 : Diagramme détaillant le couplage des 6 méthodes « simples » avec MCR-ALS.

Nous utilisons la procédure d'évaluation globale déjà définie. Les résultats sont présentés sur les graphes de la Figure 41. Pour mieux apprécier l'effet de MCR-ALS, nous avons repris dans la partie gauche du graphique les résultats des méthodes « simples ».

En analysant cette figure 41, nous tirons les observations suivantes :

- ✓ MCR-ALS affine toutes les méthodes car nous observons une forte diminution de la dissimilarité moyenne pour toutes les méthodes. Une diminution de moitié pour certaines et davantage pour d'autres (diminution d'environ  $\frac{3}{4}$ ) par rapport au cas des méthodes « simples ».
- ✓ Lorsque le niveau de bruit est élevé, la qualité d'extraction de OPA(var)/MCR-ALS est de loin la meilleure avec une faible dépendance de la dissimilarité moyenne avec le bruit alors que toutes les autres méthodes sont sensibles à l'augmentation du niveau de bruit.
- ✓ A l'exception de PCA/MCR-ALS et de OPA(spec)/MCR-ALS, toutes les autres méthodes ont une dissimilarité moyenne inférieure à 0,3 aussi bien pour les images que pour les spectres. Pour un faible niveau de bruit, presque toutes ces méthodes possèdent une dissimilarité moyenne d'environ 0,1 pour les images et pour les spectres. SIMPLISMA/MCR-ALS est l'une des méthodes le plus efficaces vis-à-vis d'un faible niveau de bruit.
- ✓ Les méthodes OPA(spec)/MCR-ALS et PCA/MCR-ALS sont les moins précises et possèdent donc les plus mauvaises qualités d'extraction quel que soit le niveau de bruit considéré. Cela reste vrai à la fois pour les images et pour les spectres.

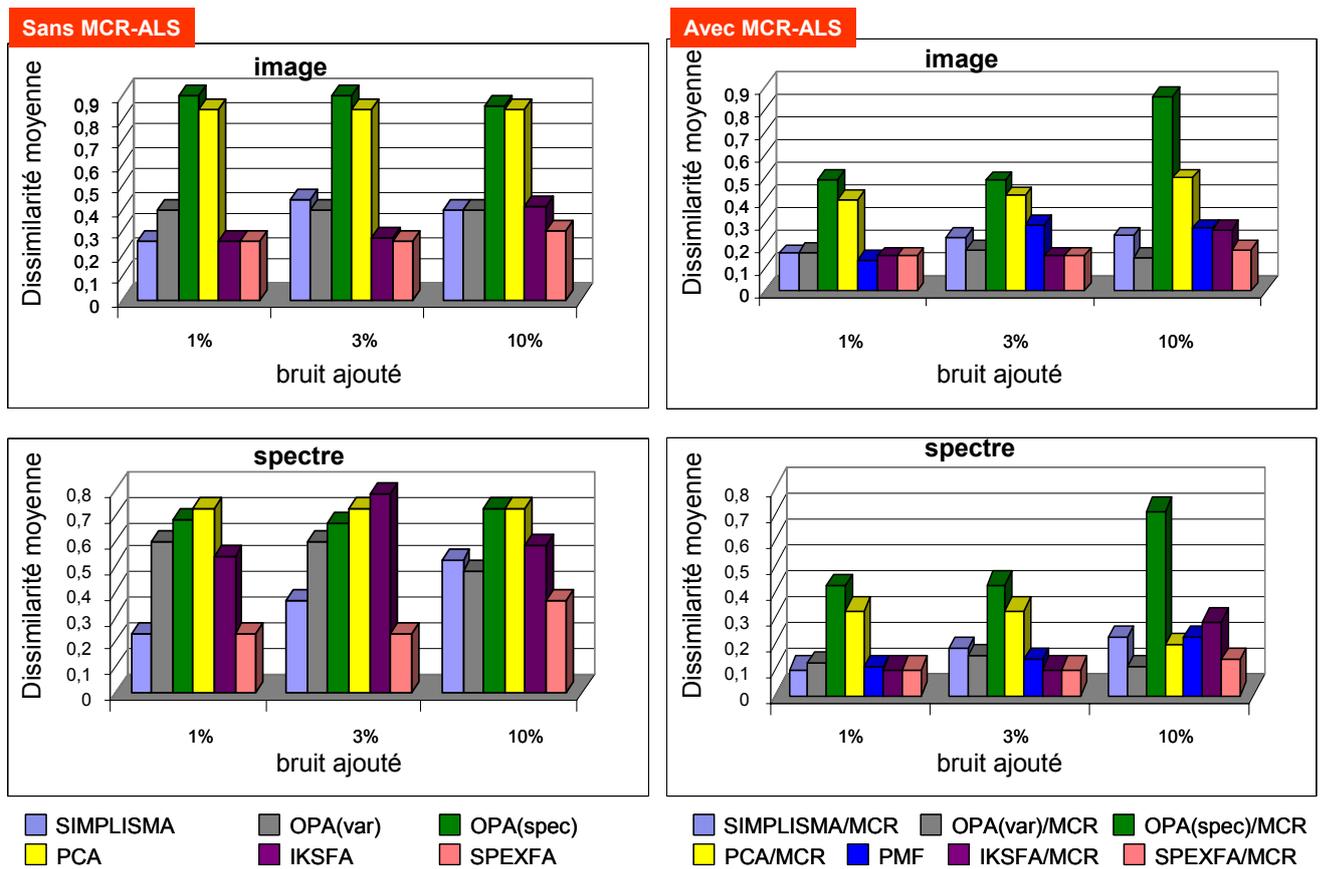


Figure 41 : Comparaison des résultats du couplage des méthodes « simples » avec MCR-ALS en fonction du bruit.

De cette étude, nous devons retenir deux conclusions importantes :

- L'utilisation de MCR-ALS à la suite des méthodes « simples » améliore d'une manière considérable la qualité d'extraction de toutes les méthodes. D'où la pertinence du couplage que nous proposons. Dans les faibles bruits, PMF donne un résultat similaire à OPA(var)/MCR-ALS.
- Une méthode « simple » ayant déjà une mauvaise qualité d'extraction aura aussi avec le couplage avec MCR-ALS une mauvaise qualité d'extraction. C'est ce qui est le cas pour PCA et OPA(spec).

### 2.2.2. Etude en fonction des décalages

Nous venons de voir les comportements des méthodes simples couplées à MCR-ALS en fonction du bruit. Nous allons analyser maintenant l'effet de ce couplage vu sous l'angle des décalages en nombres d'onde.

Pour mieux observer à la fois en fonction des décalages en nombres d'onde mais aussi en fonction des niveaux de bruit, nous faisons l'interprétation pour deux niveaux de bruit (1% et 3%). Les résultats sont représentés sur le graphique de la Figure 42.

Nous retrouvons les deux comportements : celui propre aux images et celui caractéristique des spectres.

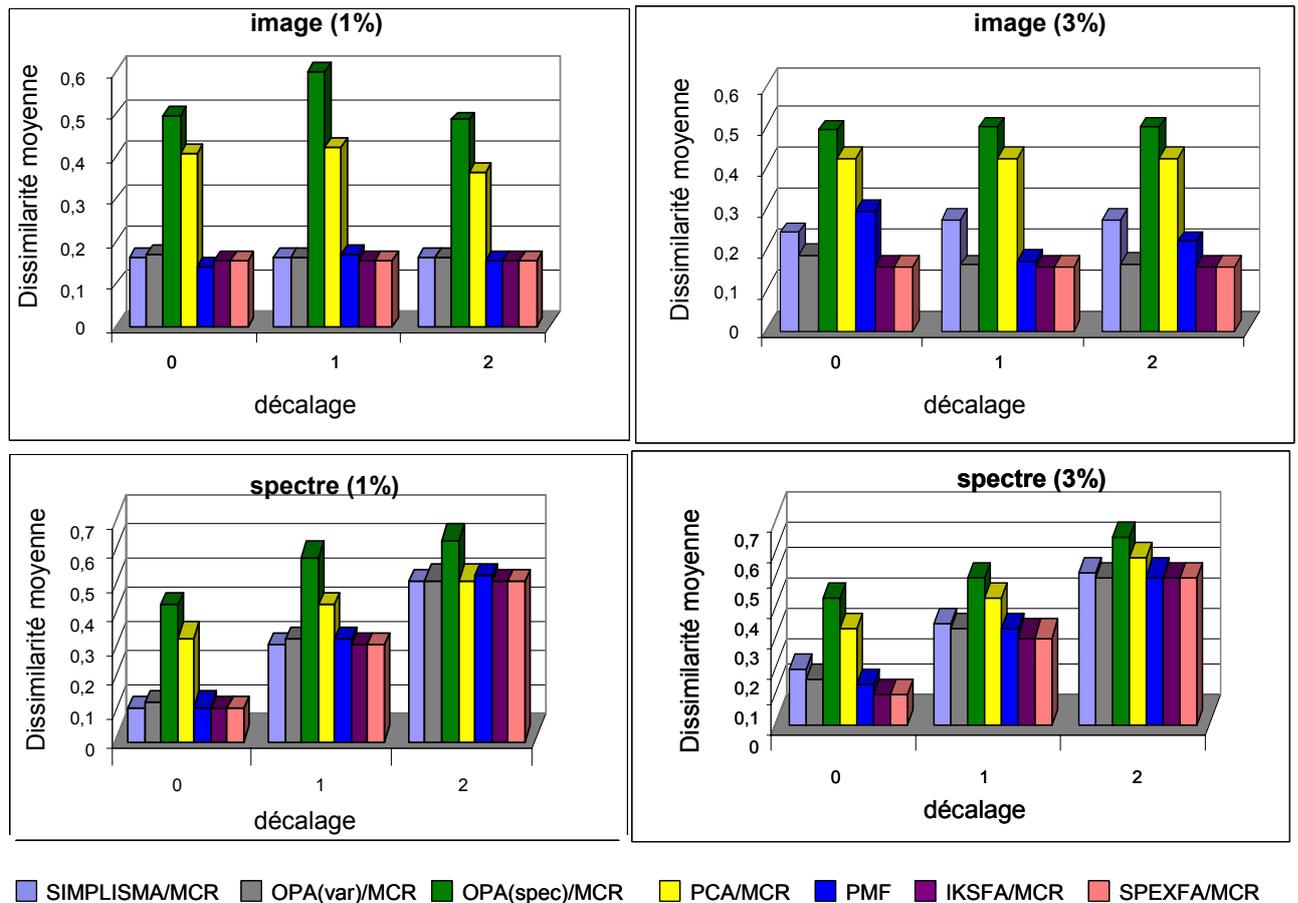


Figure 42 : Etude des résultats du couplage des méthodes « simples » avec MCR-ALS en fonction des décalages à l'aide des dissimilarités moyennes globales pour les niveaux de bruit de 1% et 3%.

Pour les images, nous notons les observations suivantes :

- ✓ Pour un niveau de bruit donné, nous obtenons la même valeur de dissimilarité moyenne pour chaque méthode quel que soit le décalage. Ce qui traduit une grande stabilité des images face aux décalages.
- ✓ Entre deux niveaux de bruit, on note une légère sensibilité de SIMPLISMA/MCR-ALS pour qui la dissimilarité moyenne passe de 0,15 à 1% à environ 0,25 à 3%.
- ✓ Comme déjà noté, OPA(spec)/MCR-ALS et PCA/MCR-ALS restent les plus perturbées avec les plus mauvaises qualités d'extraction.

Quant aux spectres, le comportement observé est totalement différent.

- ✓ Quel que soit le niveau de bruit lorsque le décalage croît la perturbation augmente.
- ✓ Pour un faible niveau de bruit (1%), SIMPLISMA/MCR-ALS est la meilleure mais pour un niveau de bruit plus important (3%) celle-ci perd son avantage et les autres méthodes telles que OPA(var)/MCR-ALS, IKSFA/MCR-ALS ou SPEXFAM/MCR-ALS produisent une meilleure extraction.

Cette étude nous amène à retenir les conclusions suivantes :

- L'extraction des images n'est pas significativement perturbée par les décalages pour un niveau de bruit donné. Même entre deux niveaux de bruit différents, les résultats sont presque identiques.
- L'extraction des spectres est fortement perturbée par les décalages quel que soit le niveau de bruit. On observe une croissance linéaire en fonction des décalages. A un niveau de bruit important, quelques méthodes comme OPA(var)/MCR-ALS donnent la meilleure extraction.

### 3. Etude des méthodes de résolution raffinées par MCR-ALS avec les contraintes du rang local en plus de la non-négativité

Dans le paragraphe précédent, nous avons passé en revue les différentes conditions expérimentales avec la seule contrainte de non-négativité. Dans cette partie nous allons utiliser la même démarche en gardant la contrainte de non-négativité mais cette fois nous introduirons dans le processus de résolution les contraintes du rang local afin d'en dégager l'intérêt. Avant d'exposer les résultats, nous devons d'abord expliquer la procédure de calcul des contraintes du rang local.

#### 3.1. Explication de la méthodologie de la contrainte du rang local

Ayant constitué notre matrice de données expérimentales par une combinaison linéaire aléatoire, le nombre de composés présents sur un pixel donné n'est pas toujours égal à 6. C'est un nombre compris entre 0 et 6 inclus.

Pour aboutir aux contraintes du rang local, on procède par plusieurs étapes. Avant de détailler cette partie importante de la procédure nous allons l'exposer d'une manière globale.

La première étape est d'estimer le nombre des composés présents sur un pixel donné. Puis à l'aide des spectres des composés de référence et par les biais des corrélations, on identifie les composés absents sur chaque pixel ; ce qui nous donne la carte des absences. Ensuite en utilisant ces connaissances, on déduit le rang local de chaque pixel. Le diagramme de la Figure 43 résume la démarche globale.

La procédure détaillée concernant l'estimation et le calcul du rang local d'une image bidimensionnelle ou tridimensionnelle, les techniques et les routines utilisées peuvent être trouvés dans l'un des travaux de Anna De Juan<sup>32</sup> et *al.*

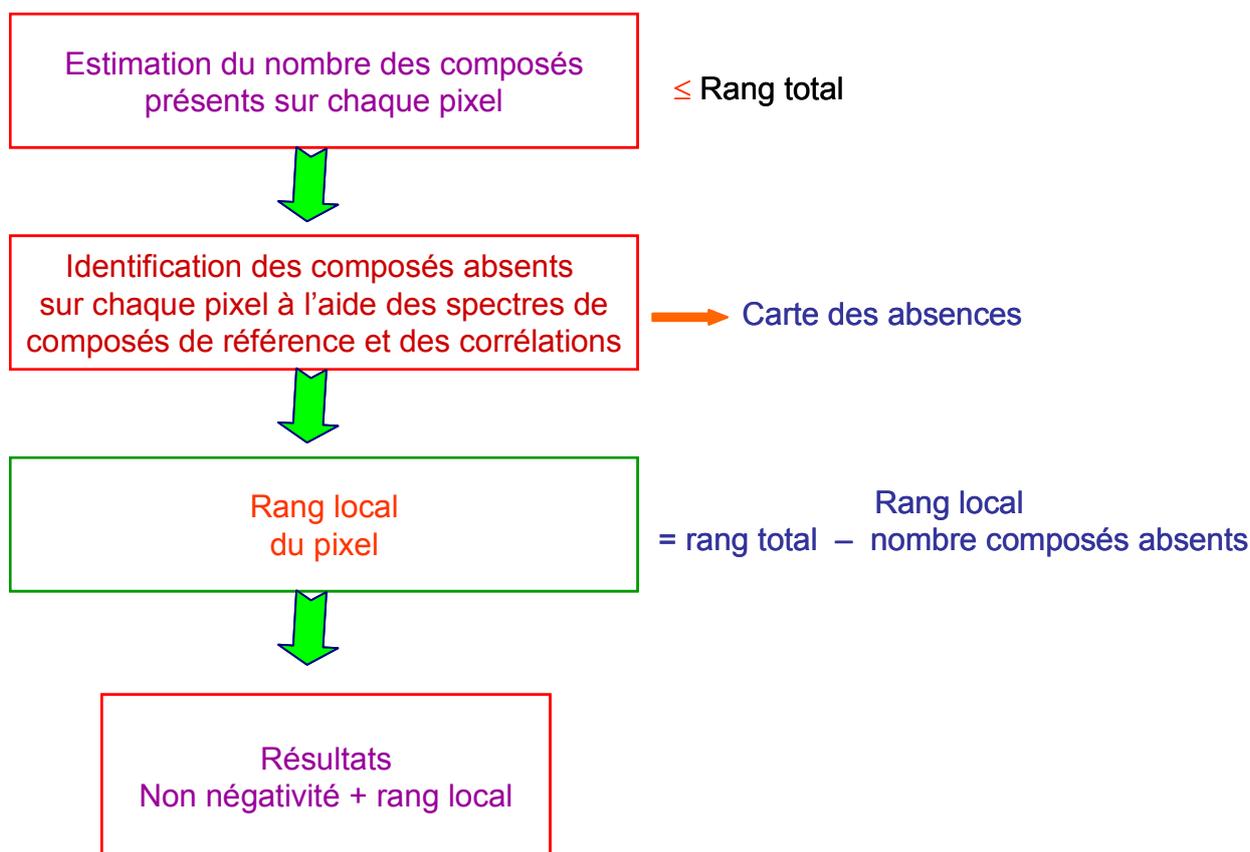


Figure 43 : La procédure globale pour le calcul des contraintes du rang local.

### 3.1.1. Estimation du nombre de composés présents sur un pixel

Cette estimation nécessite d'utiliser la matrice de données sous forme tridimensionnelle ( $x \times y \times \lambda$ ). Et cela pour plusieurs raisons. D'abord il s'agit de garder cette matrice comme l'acquisition expérimentale dans l'espace  $(x,y)$ , la troisième dimension  $\lambda$  représentant les variables « nombres d'onde ». Ensuite il ne faut pas perdre les propriétés « proches voisins » des pixels de l'image étudiée. Le but, c'est d'estimer le nombre de composés présents sur chaque pixel sans faire subir des déformations à l'image. Pour cela, la définition d'une fenêtre en évolution est nécessaire. A une dimension, l'utilisation de FSMW-EFA<sup>159,160</sup> (cf. chapitre 1 §III.2) donne des résultats satisfaisants mais cela n'est pas applicable à une image 2D ou 3D. Pour cette raison, nous avons utilisé l'algorithme modifié FSIW-EFA proposé par Anna De Juan<sup>32</sup> et *al.* (cf. chapitre 1 §III.2.3). Ce qui est critique, c'est le choix la taille de la fenêtre. Cette fenêtre carrée doit avoir une taille égale au moins au nombre de constituants. Par exemple si on recherche 4 constituants, une fenêtre de  $(2 \times 2)$  suffit ; pour 6 on prendra une fenêtre de  $(3 \times 3)$ . Dans notre cas, on aura donc à faire le calcul pour 9 constituants. Parmi ceux-là nous retrouverons les 6 composés recherchés. Ensuite, on effectue un balayage dans l'espace  $(x,y)$  en se déplaçant d'un pixel chaque fois comme schématisé à la Figure 9a.

A chaque position, on effectue la SVD de l'ensemble des spectres des pixels couverts par la fenêtre. Le nombre des valeurs singulières significatives de cette SVD donne l'estimation du nombre de composés présents sur le pixel considéré.

### 3.1.2. Production des cartes du rang local

La deuxième étape est la production des cartes du rang local en utilisant ce nombre estimé sur chaque pixel. En effet l'information issue de l'analyse SVD de différentes fenêtres est restructurée pour tenir compte de la structure 2D ou 3D de l'image analysée.

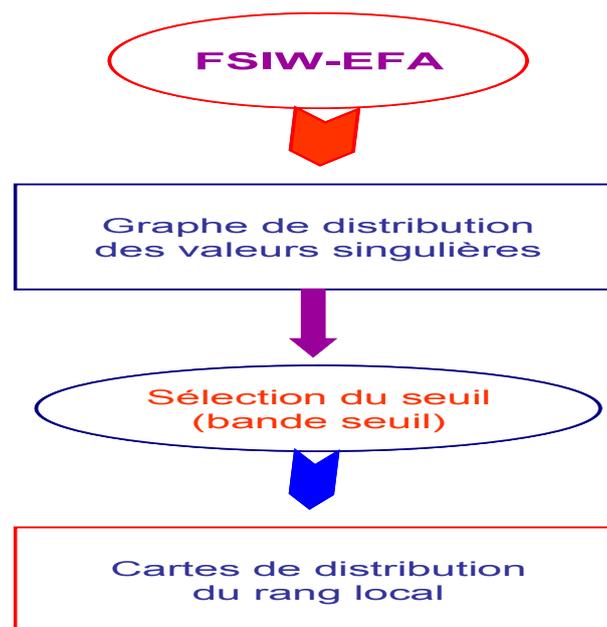


Figure 44 : Schéma explicatif de la production des cartes du rang local par l'utilisation de FSIW-EFA

La démarche se déroule comme suit :

- 1°. On produit alors le graphe des valeurs singulières en fonction du numéro des pixels ou de l'index des valeurs singulières.
- 2°. Sur le graphe des valeurs singulières en fonction du numéro des pixels, on sélectionne le seuil, une bande « seuil » en fait (cf. Figure 45 et Figure 71), pour distinguer les valeurs singulières rattachées au bruit de celles caractéristiques des constituants.
- 3°. On produit la carte du rang local complet correspondant à la position médiane entre les deux limites de la « bande seuil » et les cartes du rang local correspondant aux deux limites de cette même bande (cf. Figure 46 et Figure 70). Le rang local sur un pixel est exprimé par le nombre estimé de composés présents sur ce pixel. Ce qui est le plus important et que nous allons utiliser par la suite, c'est la carte du rang local partiel. Cette carte du rang local partiel ne comporte que les pixels pour lesquels la valeur du rang local ne varie pas entre les deux limites de la « bande seuil » (cf. Figure 46). Ceux pour lesquels le rang local varie, on ne se prononce. Ceux-ci sont en

bleu-marine sur la carte du rang local partiel sur la Figure 46. Le schéma de la Figure 44 synthétise les différentes étapes dans la production des cartes du rang local.

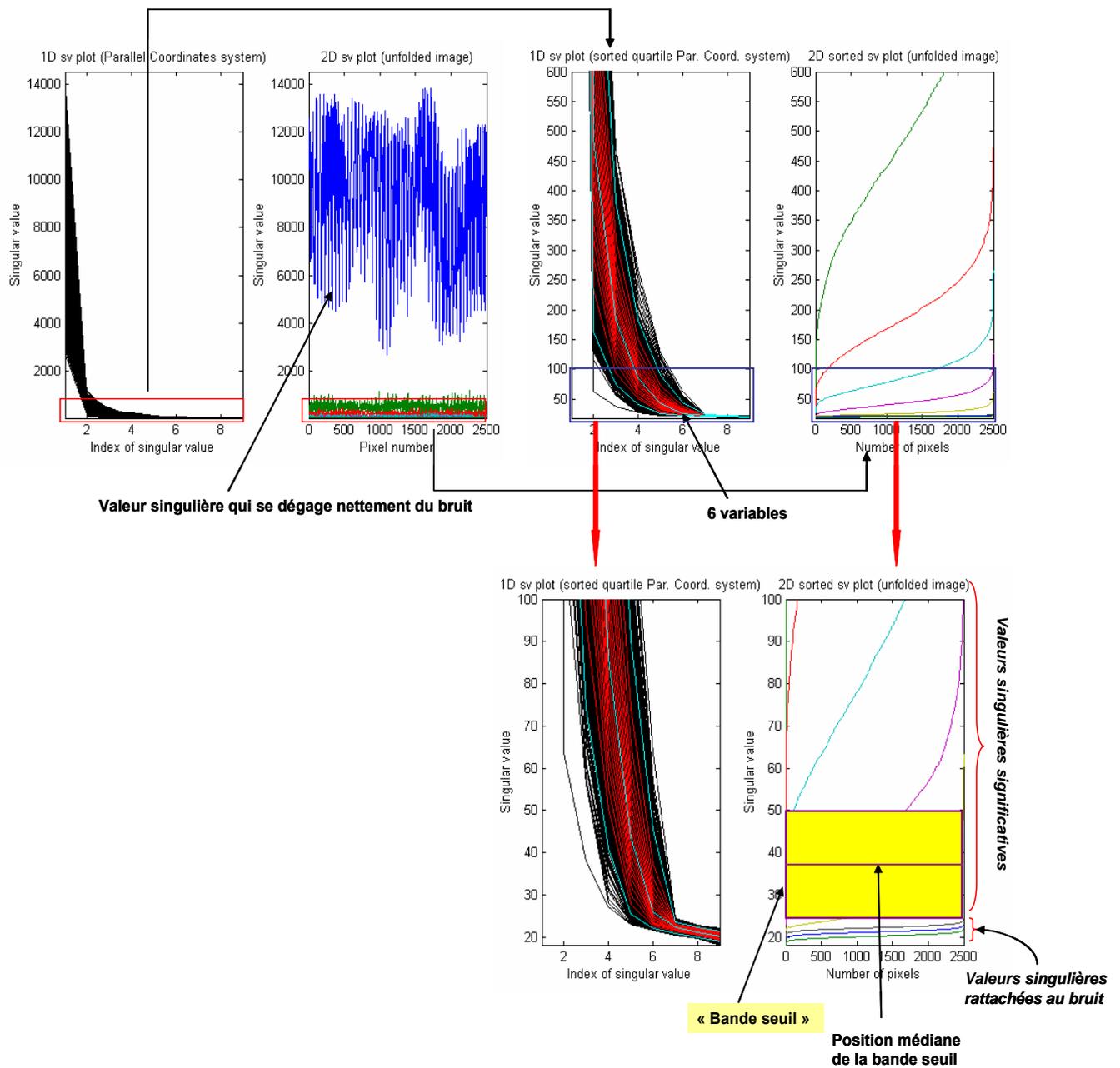


Figure 45 : Réorganisation des valeurs singulières pour la sélection de la bande « seuil » servant à la production des cartes du rang local. Dans le graphe le plus à droite, les valeurs singulières rattachées au bruit sont entassées en bas.

Sur la Figure 45 (deuxième graphe à partir de la gauche), nous voyons qu'il y a une valeur singulière très élevée pour chaque pixel et celle-ci se distingue très nettement du bruit. Le zoom réalisé, sur le graphe le plus à droite, montre aussi 4 valeurs singulières (4 courbes) assez distinctes du bruit, la cinquième en jaune est proche du niveau de bruit. Cela fait au total 6 valeurs singulières distinctes de celles rattachées au bruit entassées en bas du graphique. Les cartes du rang local sur la Figure 46 sont produites alors pour un nombre de constituants égal à 6.

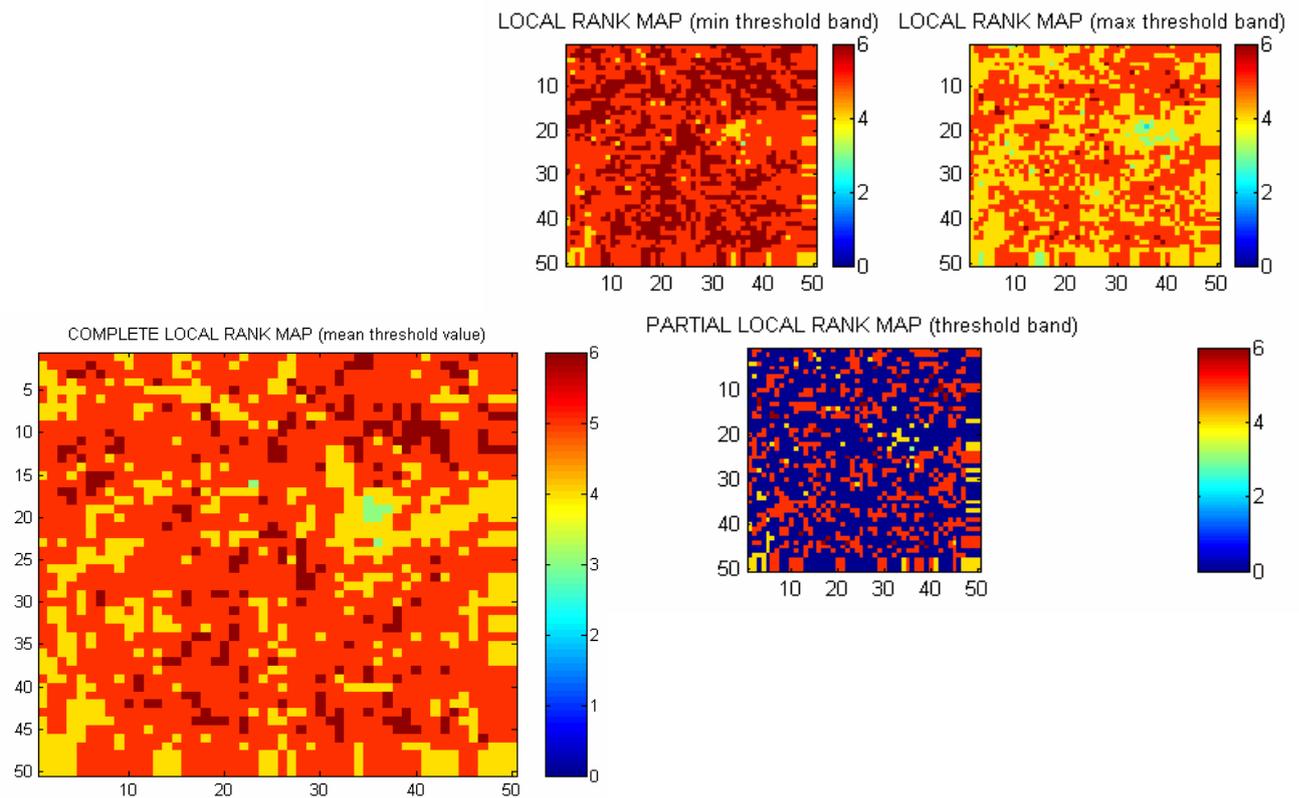


Figure 46 : Cartes du rang local produites pour les extrémités de la bande « seuil », pour la position médiane de la bande (rang local complet) et la carte du rang local partiel pour toute la bande.

### 3.1.3. Production de la carte des absences

Disposant des cartes du rang local, nous allons utiliser ces connaissances pour faire une analyse plus approfondie. Nous tenterons d'identifier les composés absents sur chaque pixel. Nous nous servirons de la carte du rang local partiel puisque dans ce cas le rang local reste invariant sur une large zone. En plus de cela, nous devons avoir à notre disposition les spectres des 6 composés de référence et l'analyse se fait sur la matrice tridimensionnelle ( $x \times y \times \lambda$ ). Chaque composé est pris individuellement et l'analyse est faite sur tous les pixels. On aura alors une carte des absences pour chaque composé sur l'ensemble des pixels. Puis l'ensemble de 6 cartes des absences formera la matrice exprimant les contraintes du rang local puisqu'elle indiquera aux algorithmes qu'un tel composé est absent sur tel pixel de l'image.

Pratiquement pour dire qu'un composé donné est absent sur un pixel quelconque, il nous faut un critère de sélection. Ce choix est fait en utilisant les coefficients de corrélation entre les spectres purs fournis. On calcule les coefficients de corrélation de l'un ( $S_1$  par exemple) par rapport au 5 autres. Parmi les 5 coefficients de corrélation calculés, on retient le plus grand comme représentatif du spectre considéré ( $S_1$ ).

Lors de l'analyse, on calcule le coefficient de corrélation entre le spectre du pixel  $i$  et chaque spectre pur  $S_j$  ( $j = 1 : 6$ ) fourni. Si ce coefficient est inférieur au coefficient maximal sélectionné

pour le spectre pur  $S_j$ , alors on considère que le composé correspondant à ce spectre pur  $S_j$  est absent du pixel  $i$ .

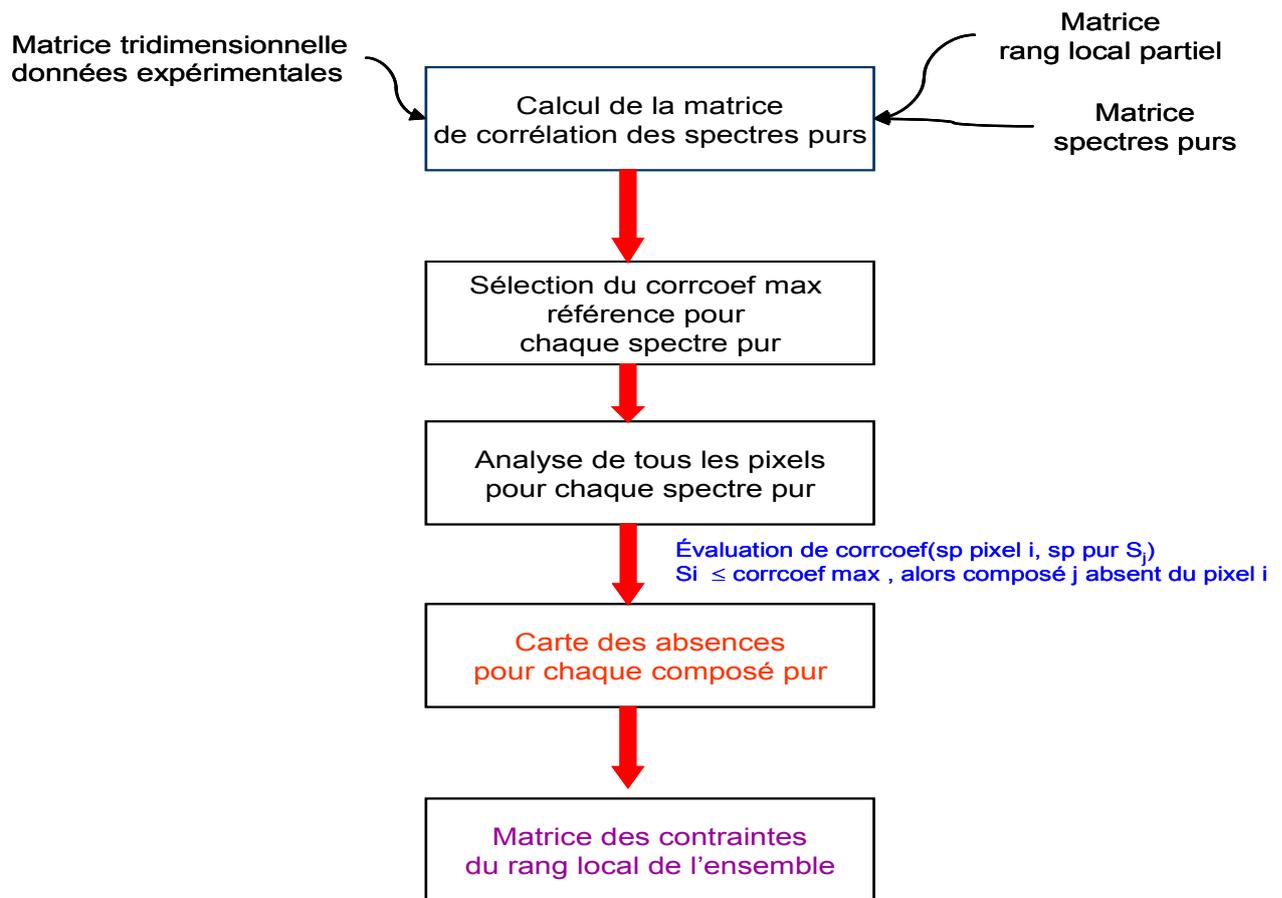


Figure 47 : Procédure de production des cartes des absences des composés de référence et de la matrice des contraintes du rang local.

Cette démarche est résumée d'une manière schématique sur la Figure 47. La production de la matrice des contraintes du rang local nous apporte une connaissance supplémentaire de la répartition des composés sur les pixels d'analyse.

#### 3.1.4. Utilisation des contraintes du rang local

La matrice des contraintes du rang local, obtenue à partir de la carte du rang local partiel, va être utilisée dans la procédure itérative ALS pour obtenir une optimisation à la fois des concentrations et des spectres. Comme on peut voir sur la Figure 48, les autres paramètres à injecter dans ALS doivent être bien définis.

Il est à préciser que nous utilisons dans cette résolution la matrice de données expérimentales sous forme dépliée et les spectres initiaux résultats du couplage avec MCR-ALS.

Un nombre d'itération assez élevé est utilisé pour que le processus converge automatiquement.

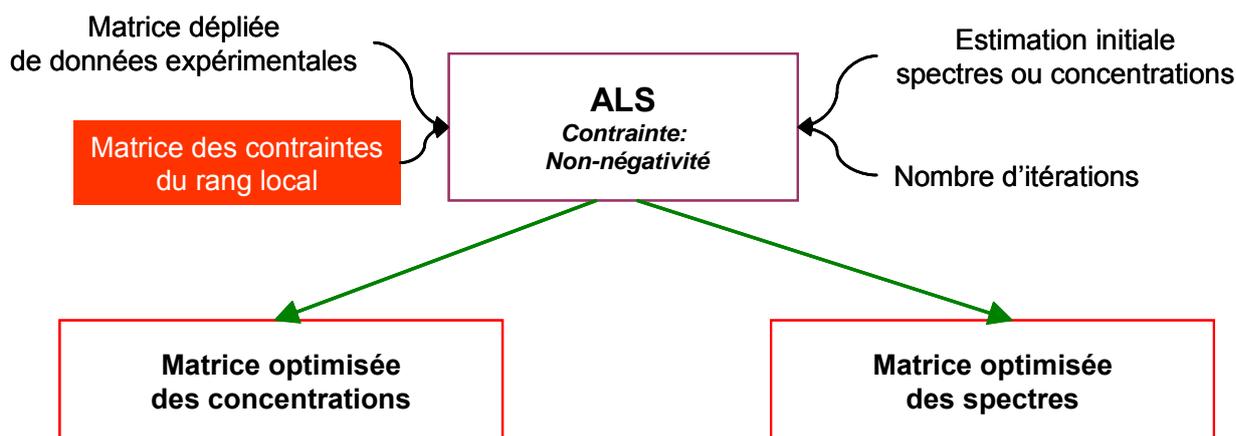


Figure 48 : Visualisation des productions des résultats par ALS en utilisant les contraintes du rang local, les données expérimentales et les autres paramètres nécessaires à la résolution.

Après convergence, nous obtenons les matrices optimisées des spectres et des concentrations.

C'est à partir de ces résultats que nous calculerons les indices des dissimilarités et analyserons l'effet apporté à l'extraction par les contraintes du rang local.

### 3.2. Etude des résolutions en fonction du bruit sans décalage en nombres d'onde

Dans le paragraphe précédent §2.2, nous avons obtenu les résultats du couplage des méthodes « simples » avec MCR-ALS. Les spectres fournis par ce couplage sont utilisés comme estimation initiale dans la procédure itérative ALS lors de l'application des contraintes du rang local. Nous appliquons la procédure d'évaluation globale déjà explicitée. Pour mieux ressortir, l'effet du rang local, sur la Figure 49, nous présentons le même cas mais sans rang local. Pour faciliter aussi l'appréciation visuelle, les graphes sont à la même échelle aussi bien pour les images que pour les spectres.

En analysant ces résultats, nous notons les observations suivantes :

- Le rapport d'amélioration de qualité pour les méthodes (OPA(spec)/MCR-ALS et PCA/MCR-ALS) qui avaient une mauvaise qualité d'extraction est considérable. La qualité d'extraction de ces deux méthodes devient proche de celle des autres méthodes.
- Les contraintes du rang local apportent une amélioration sur toutes les méthodes de résolution.
- Pour les images :
  - Entre les deux cas présentés sur cette Figure 49, en comparant les niveaux de bruit entre eux et les méthodes entre elles, on note que l'amélioration reste faible.
  - Comme déjà constaté, on observe une différence sensible entre le niveau de bruit de 1% et celui de 10%.

- Seule la méthode OPA(var)/MCR-ALS possède la plus faible valeur de la dissimilarité moyenne (environ 0,1) quel que soit le niveau de bruit. Par conséquent, c'est la méthode qui possède la meilleure qualité d'extraction et cela est renforcé par l'application des contraintes du rang local.

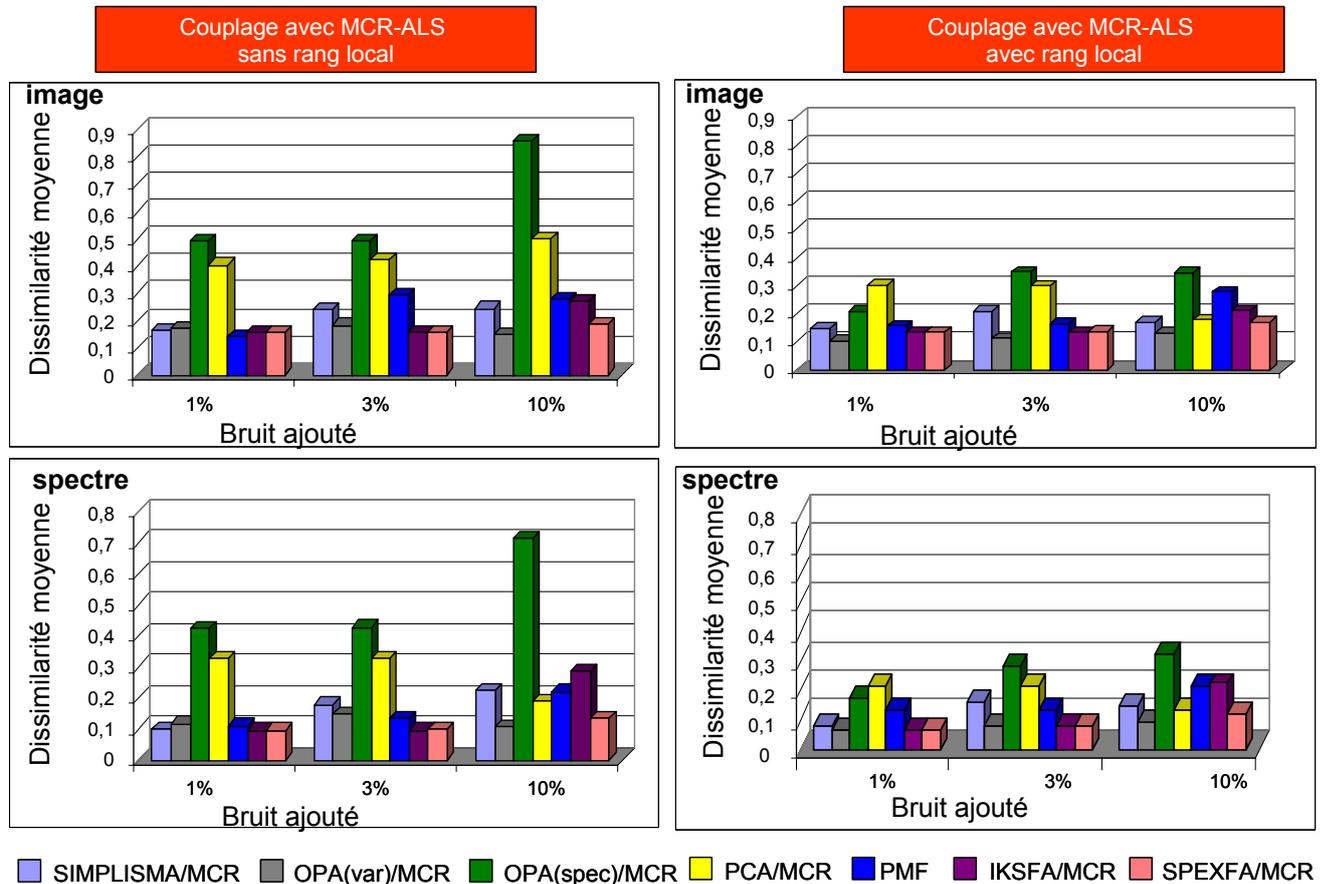


Figure 49 : Comparaison globale des résultats avec le couplage avec MCR-ALS, à gauche sans contraintes du rang local et à droite avec les contraintes du rang local.

- Pour les spectres :
  - L'amélioration de la qualité d'extraction est plus importante. Même les méthodes qui avaient une bonne qualité d'extraction comme SIMPLISMA/MCR-ALS, OPA(var)/MCR-ALS, IKSFA/MCR-ALS ou SPEXFA/MCR-ALS produisent une dissimilarité moyenne diminuée au moins de moitié par rapport au cas sans contraintes de rang local. Même dans ce cas, PMF reste influencé par le bruit.
  - Comme pour les images, OPA(var)/MCR-ALS est la plus stable et est de loin celle qui possède la meilleure qualité d'extraction dans un contexte de perturbation plus importante (bruit de 10% par exemple).

### 3.3. Observations après décalages en nombres d'onde

Les principales remarques, lorsque nous réalisons des décalages en nombres d'onde et appliquons les contraintes du rang local, sont exposées dans cette partie. La Figure 50 montre les résultats pour les niveaux de bruit de 1% et de 3%.

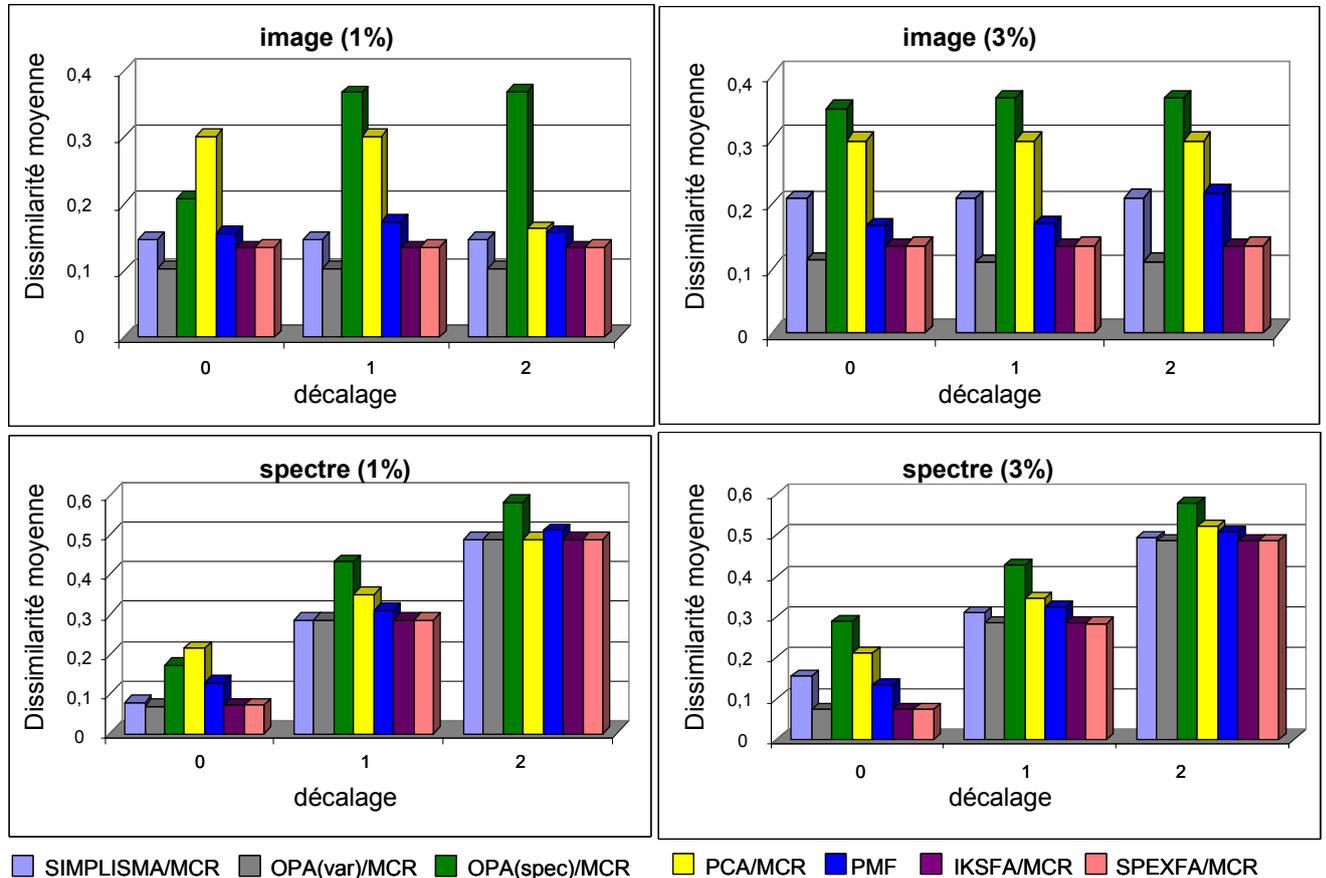


Figure 50 : Comparaison des résultats pour les niveaux de bruit de 1% et 3% lors de l'application des contraintes du rang local après réalisation des décalages en nombres d'onde.

Nous retiendrons les observations suivantes.

- Comme pour le cas sans les contraintes du rang local précédemment expliqué, nous notons pour les images :
  - Les images ne sont pas sensibles aux décalages puisque nous obtenons les mêmes valeurs de dissimilarité moyenne quel que soit le décalage pour chaque méthode de résolution (cf. 3% par exemple).
  - Les images ne sont pas non plus perturbées par le niveau de bruit. Seule SIMPLISMA/MCR-ALS montre une très légère sensibilité puisque la dissimilarité moyenne est légèrement plus importante pour le niveau de bruit de 3% (de 0,15 à 0,20). En effet, dans les faibles dissimilarités moyennes, une légère différence entre deux méthodes traduit une perturbation relativement importante.

- Les spectres sont fortement perturbés par les décalages et la dégradation de la qualité d'extraction est linéairement croissante avec les décalages. A un faible niveau de bruit, OPA(var)/MCR-ALS est l'une des meilleures méthodes et garde cette qualité même pour des conditions expérimentales plus difficiles. On n'observe pas un comportement significativement différent entre les deux niveaux de bruit.

#### 4. Les méthodes de résolution les plus robustes à retenir pour l'application au traitement des données de l'imagerie microspectrométrique

Dans les paragraphes précédents, nous avons passé en revue les trois cas d'étude : les méthodes « simples », ces mêmes méthodes « simples » couplées avec MCR-ALS et enfin les contraintes du rang local injectées à ces couplages. Dans cette partie, nous allons dégager les méthodes les plus robustes sur un exemple donné en proposant les images et les spectres extraits. Parmi les différents cas étudiés dans les paragraphes précédents, nous avons sélectionné les 4 méthodes ayant les meilleures qualités d'extraction.

Le schéma de la Figure 51 résume la démarche et présente cette sélection. Le numéro entre parenthèses indique le classement de la méthode de résolution en fonction de la qualité d'extraction par rapport à l'ensemble des méthodes étudiées ; et cela en fonction de la dissimilarité moyenne sur les 6 composés.

En analysant les méthodes sélectionnées ayant les meilleures qualités d'extraction pour chaque cas étudié, nous pouvons dire :

- Pour les méthodes « simples » :
  - SPEXFA et IKSFA donnent un résultat similaire à celui de SIMPLISMA et de OPA(var). Ces deux dernières n'étant pas toujours les mieux classées.
  - En observant ces classements pour les spectres et les images, on peut appréhender un comportement différent possible entre les images et les spectres pour chaque méthode de résolution.
- Pour le couplage avec MCR-ALS et l'application des contraintes du rang local :
  - Dans ces deux cas, l'ordre de classement des méthodes pour la qualité d'extraction est toujours la même.
  - Comme nous avons indiqué par le souligné, c'est OPA(var)/MCR-ALS qui possède la meilleure qualité d'extraction et qui devance d'une manière significative SPEXFA/MCR-ALS et SIMPLISMA/MCR-ALS.
  - L'application des contraintes du rang local renforce cet avantage de OPA(var)/MCR-ALS sur les autres.

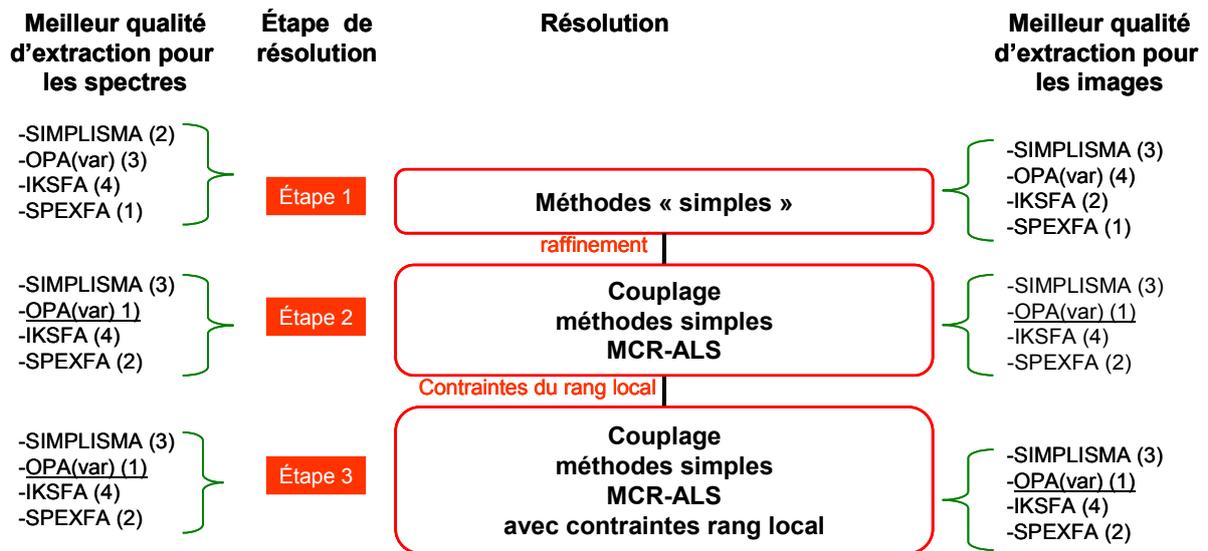


Figure 51 : Qualité d'extraction des 4 meilleures méthodes appliquées dans différents conditions de résolution. Le nombre entre parenthèses indique le classement de la méthode selon le critère de la dissimilarité moyenne.

Pour mieux apprécier la différence de comportement de ces méthodes de résolution dans les différents cas étudiés, les images et les spectres extraits par chacune pour un composé, sont présentés. Mais cette comparaison et l'interprétation qui en découle peuvent être faites pour chaque composé de référence. Nous présentons cette visualisation pour le composé « polyesterterephthalate ». Le spectre et l'image de référence de ce composé sont aisément identifiables. Cependant les méthodes de résolution ont une certaine difficulté à l'extraire correctement. Les figures (Figure 52, Figure 53, Figure 54) montrent ces visualisations comparatives.

Dans chaque cas, nous avons décidé de présenter cette visualisation comparative pour deux niveaux de bruit de 1% et 10%.

Dans chaque cas, en face des images et des spectres, nous présentons l'image et le spectre de référence du « polyesterterephthalate ». C'est ce qui nous permettra de juger la qualité d'extraction. Sur chaque image extraite, nous avons affiché la dissimilarité pour apprécier la différence des images extraites par rapport à celle de la référence. Il en est de même pour les spectres.

Pour mieux montrer la différence de comportement de ces quatre méthodes, nous avons représenté sur les images des ellipses bleues montrant une certaine surestimation de la concentration. Et sur les spectres, des ellipses et des flèches rouges qui indiquent l'apparition des bandes supplémentaires ou des nombreux artefacts.

Nous allons donner une analyse synthétique dans les trois cas d'étude en insistant un peu plus sur les caractéristiques à mettre en évidence.

- Cas des méthodes « simples » : la Figure 52 nous présente la visualisation comparative des méthodes « simples ».

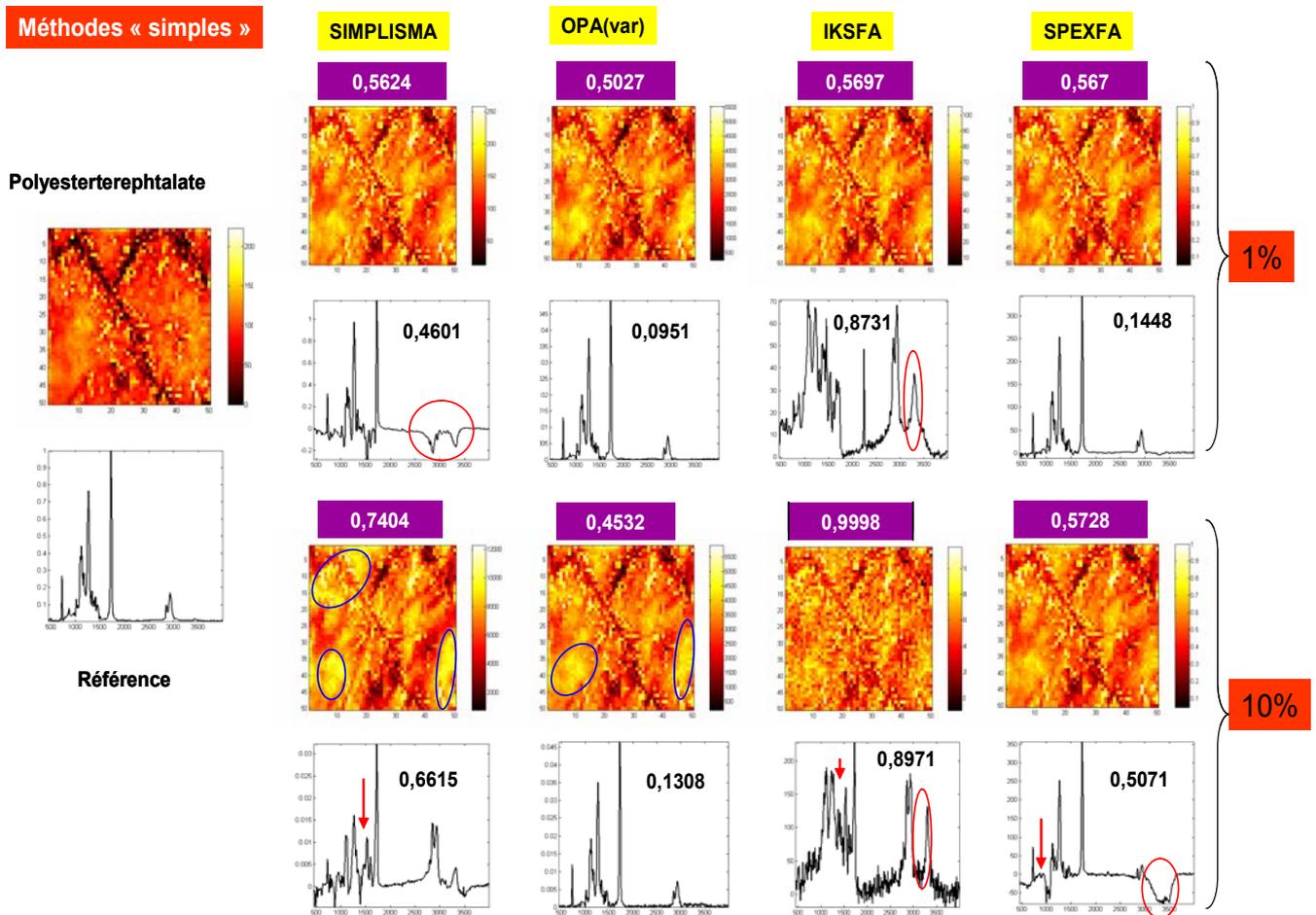


Figure 52 : Visualisation des images et des spectres extraits pour les 4 meilleurs algorithmes dans le cas des méthodes « simples ». Les ellipses et les flèches rouges indiquent l'apparition des bandes supplémentaires (ou artefacts). Les ellipses bleues sur les images indiquent une surestimation de la concentration.

En observant cette figure, nous pouvons retenir les observations suivantes :

- ✓ Pour les images, d'une manière générale, nous pouvons dire qu'il y a une surestimation. Cette surestimation est mise en évidence d'une part par le coefficient de corrélation assez élevé souvent supérieure à 0,50 ; et d'autre part par une coloration plus claire en se référant à l'échelle des concentrations.
- ✓ Pour le cas de 1%, malgré la différence des dissimilarités, la différence sur les images est très peu remarquable. Quant aux spectres, SIMPLISMA et IKSFA présentent des nombreux artefacts soulignés par les ellipses rouges et cela indique une qualité d'extraction dégradée. OPA(var) et SPEXFA extraient un spectre de qualité avec un net avantage pour OPA(var).

- ✓ Pour le cas de 10%, nous observons d'une manière générale, une dégradation de la qualité des images et des spectres extraits comparativement au cas de 1%. Cela est caractérisé d'une part par l'augmentation d'une manière significative des dissimilarités, et d'autre part par le renforcement des colorations sur les images ou le multiplication des artefacts sur les spectres. Il est à noter la très mauvaise qualité d'extraction de IKSFA pour lequel ni l'image ni le spectre ne sont reconnaissables avec des dissimilarités (respectivement de 0,9998 et 0,8971) très élevées; ce qui est à la limite extrême de la dissemblance. Mais aussi la dégradation assez significative pour SIMPLISMA et SPEXFA aussi bien pour les images que les spectres. Seul OPA(var) manifeste la plus faible variation des dissimilarités et subit de ce fait la plus faible perturbation due au bruit. Ce qui traduit une stabilité relative.
- Cas des méthodes « simples » couplées à MCR-ALS. La Figure 53 présente les résultats d'extraction. Nous indiquerons dans ce cas les principales différences notables et significatives, et par rapport au cas précédent.

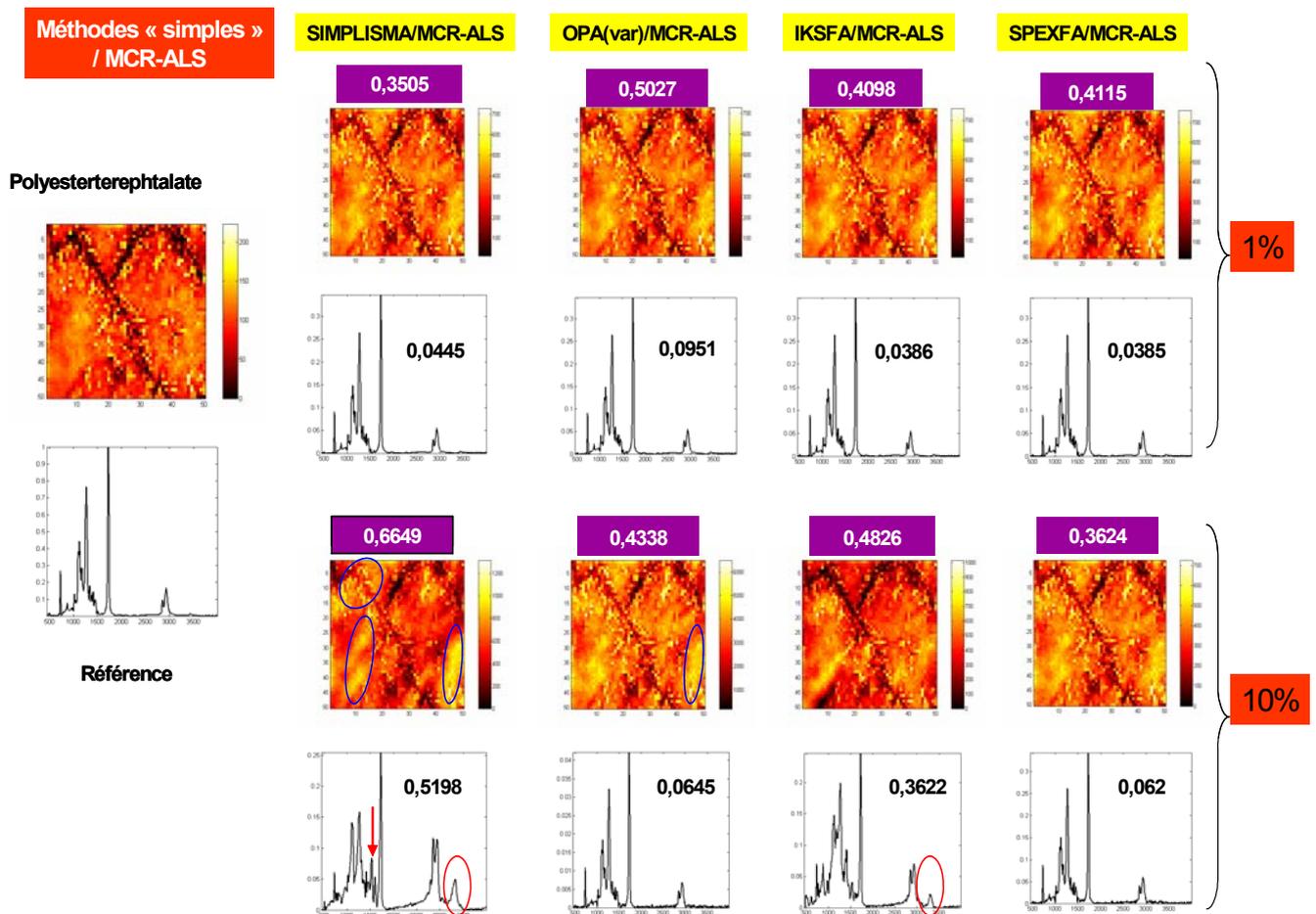


Figure 53 : Visualisation des images et des spectres extraits pour les 4 meilleurs algorithmes avec le couplage avec MCR-ALS. Les ellipses et les flèches rouges indiquent l'apparition des bandes supplémentaires (ou artefacts). Les ellipses bleues sur les images indiquent une surestimation de la concentration.

- ✓ Ce qui se remarque tout de suite quel que soit le niveau de bruit considéré, c'est la meilleure qualité des images et des spectres extraits mis en évidence par la diminution d'une manière significative des dissimilarités par rapport au cas précédent
- ✓ Pour le 1%, on n'observe aucun artefact sur les spectres ni une grosse surestimation des concentrations.
- ✓ Pour 10%, on note la présence des artefacts sur les spectres seulement pour SIMPLISMA et ISKFA avec les plus grandes valeurs de dissimilarité. Mais aussi l'étendue des zones claires caractéristiques de surestimation des concentrations est relativement réduite.
- ✓ Entre 1% et 10%, la différence due à la perturbation du bruit est moindre que le cas des méthodes « simples ». Nous voyons clairement que dans les conditions expérimentales difficiles (cas 10% par exemple), OPA(var)/MCR-ALS et SPEXFA/MCR-ALS sont les meilleures méthodes avec des très faibles valeurs de dissimilarité surtout pour les spectres.
- Cas des méthodes « simples » couplées à MCR-ALS avec les contraintes du rang local. La Figure 54 nous présente le résultat visuel.
  - ✓ D'une manière générale, nous observons aussi une amélioration par rapport au cas du couplage des méthodes « simples » avec MCR-ALS aussi bien pour les images que les spectres.
  - ✓ Pour les images, nous notons une ressemblance plus forte des extraits avec la référence même si les dissimilarités ne sont pas assez faibles.
  - ✓ Pour le 1%, les spectres sont de très bonne qualité et les images ont la meilleure qualité de tous les cas étudiés.
  - ✓ Pour 10%, nous notons une certaine persistance des surestimations de concentrations, pour SIMPLISMA/MCR-ALS par exemple, et la présence de quelques artefacts sur les spectres extraits par SIMPLISMA/MCR-ALS et IKSFA/MCR-ALS. Cependant pour les images, leur qualité est proche de celle obtenue pour le cas de 1%.

Après avoir décrit d'une manière synthétique les trois cas présentés pour le composé « polyesterephthalate » aussi bien pour les images que les spectres, il nous reste à faire l'observation suivant la troisième dimension. Cela nous permettra d'avoir une vision globale de trois cas étudiés.

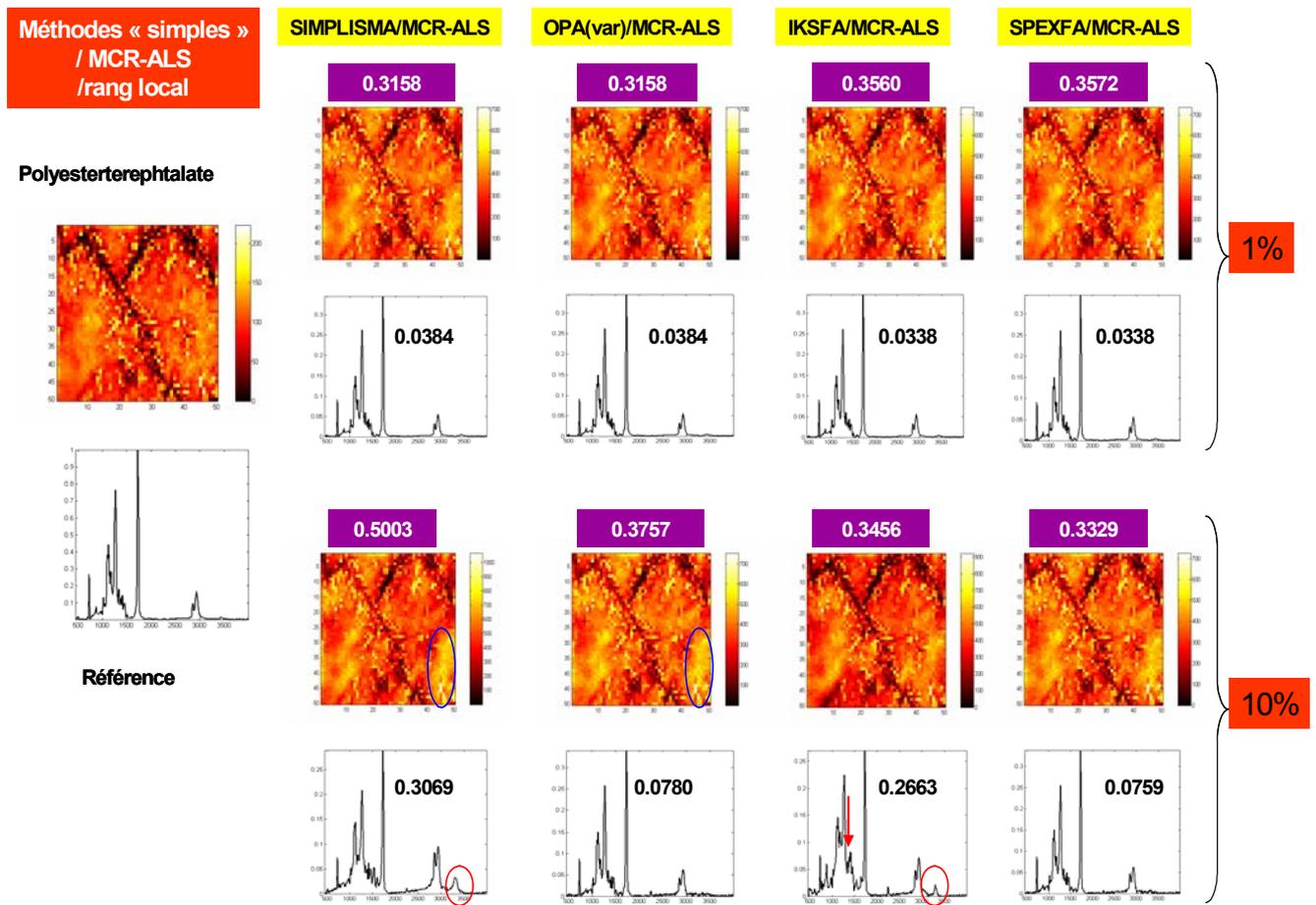


Figure 54: Visualisation des images et des spectres extraits pour les 4 meilleurs algorithmes avec le couplage de MCR-ALS et l'application des contraintes du rang local. Les ellipses et les flèches rouges indiquent l'apparition des bandes supplémentaires (ou artefacts). Les ellipses bleues sur les images indiquent une surestimation de la concentration.

Sur le graphique de la Figure 55, nous avons représenté l'histogramme donnant le comportement, de chacune de quatre méthodes pour ce même composé, vu sous l'angle de trois cas en fonction de la dissimilarité.

Ce graphique résume, à lui seul et d'une manière condensée, toutes les observations mentionnées pour les trois cas précédents. Nous avons représenté les deux niveaux de bruit côte à côte pour mieux comparer le comportement de chaque méthode entre ces deux niveaux de perturbation.

Nous pouvons tirer sans ambiguïté les conclusions suivantes.

- Pour les images :
  - ✓ Quelle que soit la méthode de résolution et quel que soit le niveau de bruit considéré, nous obtenons une décroissance des dissimilarités qui témoigne du gain de la qualité d'extraction avec le couplage avec MCR-ALS et l'injection des contraintes du rang local. OPA(var)/MCR-ALS et SPEXFA/MCR-ALS possèdent la meilleure qualité d'extraction quel que soit le cas. Il faut remarquer ici la sensibilité de SIMPLISMA/MCR-ALS avec l'accroissement du bruit (en vert pour 10%).

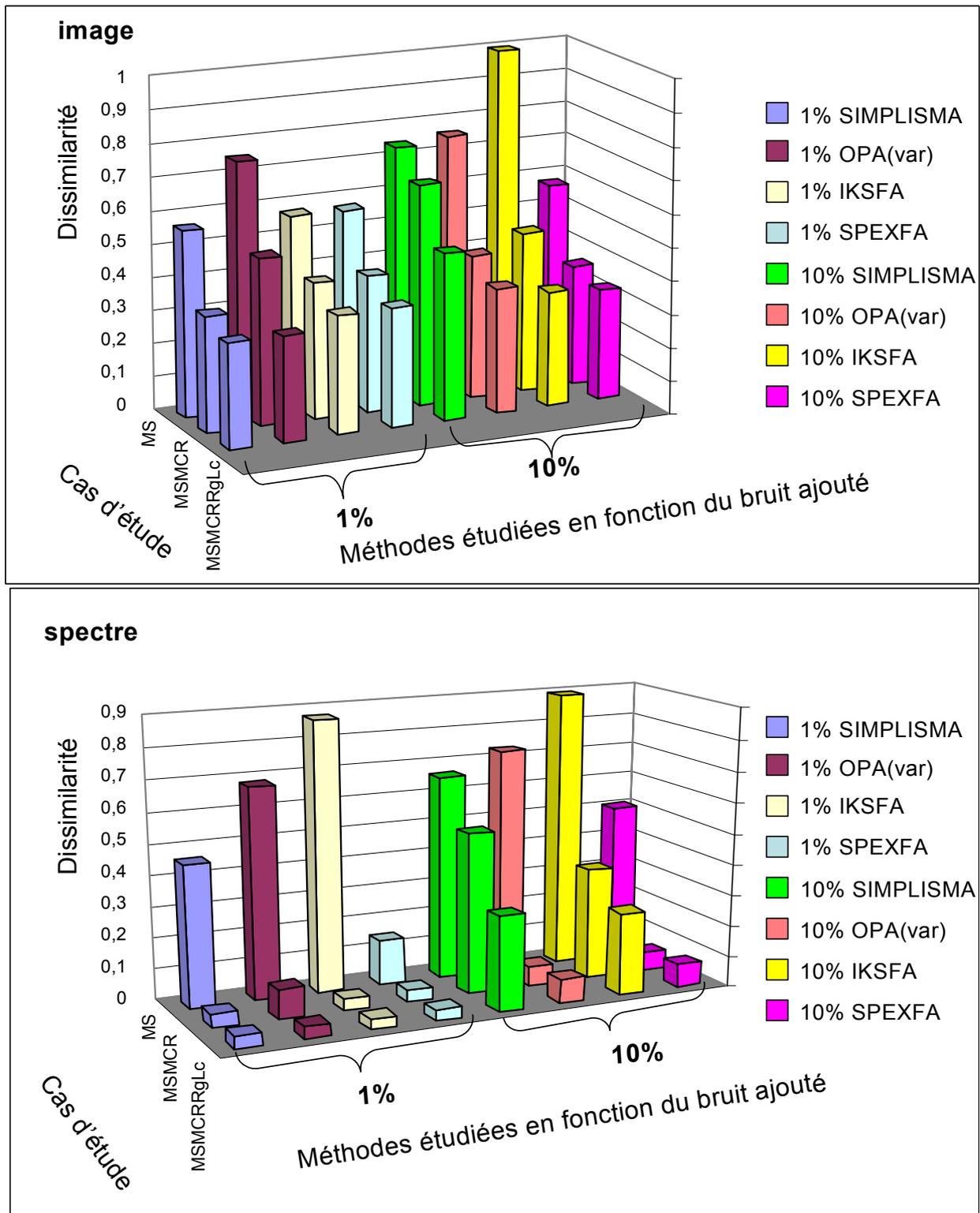


Figure 55 : Comparaison générale des comportements pour les 3 cas étudiés et pour les 2 niveaux de bruit pour les images et les spectres. Signification des initiales : MS - Méthode Simples ; MSMCR : Méthodes Simples /MCR-ALS ; MSMCRRgLc : Méthodes Simples/MCR-ALS/Rang Local.

- Pour les spectres :
  - ✓ La diminution très considérable des dissimilarités pour toutes les méthodes pour le niveau de bruit de 1% est remarquable ! Des résultats similaires sont obtenus pour toutes les méthodes après le couplage avec MCR-ALS et l'application des contraintes du rang local

par la suite. Pour le niveau de bruit de 10%, cette remarque reste valable pour OPA(var) et SPEXFA.

- ✓ Même si le couplage avec MCR-ALS et l'application des contraintes du rang local apportent une amélioration notable, SIMPLISMA et IKSFA restent les méthodes les plus perturbées par le bruit.

Compte tenu de toutes les analyses que nous venons de faire, **quelle méthode de résolution pouvons-nous raisonnablement retenir pour l'appliquer au traitement de données issues de l'imagerie microspectrométrique ?**

Il est à noter d'abord sans ambiguïté que SIMPLISMA, même couplée à MCR-ALS et avec les contraintes du rang local, reste sensible au bruit. Pour des données réelles assez complexes SIMPLISMA avec ou sans couplage ne garantit pas de fournir le meilleur résultat. Cependant du fait de son interactivité simple, SIMPLISMA<sup>161,162,163,164</sup> est utilisé dans le traitement des données dans des nombreux domaines analytiques.

Même pour ces données synthétiques avec des spectres bien traités, IKSFA a de très grandes difficultés à extraire des spectres assez corrects. IKSFA n'est pas à retenir pour un échantillon complexe. Cela n'exclut pas bien sûr que IKSFA est utilisé à bon escient dans la HPLC-DAD où certains ont obtenu des bons résultats<sup>165,166</sup>.

Quant à SPEXFA, nous avons obtenu des résultats satisfaisants comparables à ceux de OPA(var). Nous ne savons pas si cela est du à la nature de données utilisées ou aux prétraitements des spectres de la bibliothèques. En tout cas, la littérature ne plaide pas en sa faveur car il n'y a pratiquement pas (ou très peu) de publications<sup>167</sup> qui mentionnent son utilisation et l'obtention des résultats analytiques concrets.

Il nous reste, parmi les quatre méthodes présélectionnées, OPA(var). Nous avons noté dans l'analyse précédente de trois cas étudiés et dans les autres paragraphes du chapitre que OPA(var) est régulièrement améliorée par le couplage avec MCR-ALS et l'application des contraintes du rang local. Mais aussi que OPA(var) reste l'une des moins perturbées par le niveau de bruit.

Nous pouvons conclure en tenant compte à la fois de notre étude et de la littérature abondante<sup>29,168,169,170,171</sup> que OPA(var) couplé à MCR-ALS, avec les contraintes du rang local, est la plus robuste des méthodes qu'on peut appliquer sans grande difficulté aux échantillons assez complexes, notamment aux données de l'imagerie microspectrométrique.

## VI. Conclusion du Chapitre 3

Dans ce chapitre, notre but était de tester la capacité d'extraction d'un ensemble des méthodes de résolution appliqué à une matrice de données synthétiques en contrôlant quelques paramètres expérimentaux.

Nous avons réalisé cette étude dans trois cas de traitement différent que nous avons explicités.

Dans chaque cas, nous avons mis en évidence la méthode ayant la meilleure qualité d'extraction.

Puis, nous nous sommes posé la question essentielle de notre étude : ***quelle méthode de résolution est la plus robuste parmi cet ensemble des méthodes étudiées ?***

Dans le paragraphe 3, nous avons répondu à cette question en étudiant plus particulièrement les quatre meilleures méthodes de résolution. Nous avons démontré la robustesse inébranlable de OPA(var) dans ces différents cas de traitement. Nous pouvons donc appliquer cette méthode de résolution couplée à MCR-ALS (OPA(var)/MCR-ALS) au traitement de données de l'imagerie microspectrométrique. Nous pouvons nous aider aussi des contraintes du rang local pour améliorer significativement la qualité de résolution.

Dans le chapitre 4 suivant, nous appliquerons cette méthodologie d'étude à une matrice de données expérimentales issues de la microspectrométrie Raman. Nous montrerons la pertinence de cette méthodologie sur un cas réel avec toutes les complications que les données réelles peuvent impliquer.

## Chapitre 4 :

# Application de notre méthodologie de résolution à un cas concret : analyse d'un calcul urinaire par traitement chimiométrique des données issues de la microspectrométrie Raman

---

### I. Introduction

Dans ce chapitre, nous allons aborder les traitements chimiométriques d'une matrice de données spectrales réelles.

Il s'agit de poursuivre la méthodologie appliquée aux données synthétiques étudiées au chapitre 3. Les données auxquelles nous allons soumettre cette méthodologie sont issues d'une « cartographie spectrale » sur un calcul urinaire.

Nous exposerons dans une première partie quelques aspects généraux sur les calculs urinaires. Après une brève description de l'appareillage spectroscopique et de la procédure d'échantillonnage, nous aborderons l'étude propre dite de ces données spectrales. Nous chercherons d'abord à estimer le nombre de constituants significatifs. Avant les traitements, nous réaliserons les prétraitements nécessaires pour améliorer la qualité spectrale des données à traiter. Ensuite, nous appliquerons les différentes procédures de résolution explicitées au chapitre 3 qui sont : l'extraction par les « méthodes simples », le couplage des « méthodes simples » avec MCR-ALS, l'application des contraintes du rang local aux résultats du couplage précédent.

En dernier lieu, nous ajouterons un autre type de résolution : le procédé « *d'augmentation de matrices* » appliqué aux résultats fournis par l'application des contraintes du rang local. Cela nous permettra d'apprécier l'effet sur la résolution apporté par plus de contraintes.

Dans chaque cas de traitement, nous présenterons les résultats globaux et les éléments identifiés ainsi que les différents procédés d'identification utilisés.

Cela nous permettra de tirer les conclusions pertinentes dans le cadre de notre étude.

## II. Généralité sur les calculs urinaires<sup>191</sup>

Les calculs sont très divers et très variés que ce soit en ce qui concerne leur nature, leur forme, leurs constituants, le lieu de leur formation et leur origine. Il y a les calculs du rein, de la vessie et des voies urinaires, les calculs de la vésicule biliaire.

Ils sont repérables par leur morphologie, par leur coloration, par leur taille et leur structure cristalline. En effet leur forme, leur coloration et leur cristallinité dépendent de leurs constituants. Ces constituants sont très souvent liés au régime alimentaire<sup>172,173</sup> des patients, en plus des facteurs liés aux environnements écologiques. Ces calculs urinaires, qui résultent eux-mêmes de certaines pathologies<sup>174</sup>, peuvent causer de nombreuses infections douloureuses.

### 1. Les principaux constituants des calculs urinaires<sup>175 176 191</sup>

Parmi ces calculs, nous nous sommes intéressés aux calculs urinaires. D'une manière visuelle, la coloration, la taille et la forme donnent une première idée à l'œil expert. Mais pour le commun d'entre nous, une analyse rigoureuse est nécessaire. Il existe, au moins, une centaine de constituants dans les calculs urinaires. Dans chaque type de calcul urinaire, il y a des composés majoritaires. Les grands groupes de classification sont les suivants :

- ✓ les oxalates de calcium<sup>177</sup> (whewellite, weddellite),
- ✓ les phosphates de calcium (brushite, carbapatite, ...etc.),
- ✓ les phosphates de magnésium (struvite, ...etc.),
- ✓ les purines<sup>178</sup> (acide urique anhydre, acide urique dihydraté, xanthine, ...etc.),
- ✓ les protéines (albumine, fibrine, mucine, ...etc.)
- ✓ les lipides,
- ✓ les acides aminés<sup>179</sup> (cystine, tyrosine, leucine, ...etc.),
- ✓ les impuretés (drogues, ...etc.).

Parmi cet ensemble, ce sont les composés du calcium tels que les oxalates, les composés du magnésium (les phosphates), les purines et quelques protéines qui sont souvent les plus observés dans l'analyse de nombreux calculs urinaires. Actuellement différentes techniques<sup>175,180,181,182</sup> d'analyses<sup>183,184</sup> sont disponibles.

Nous avons étudié quelques calculs urinaires. Nous présenterons par la suite les résultats obtenus pour l'un d'entre eux et nous exposerons ces principaux constituants extraits par les différents traitements chimiométriques appliqués. Du point de vue de l'imagerie, c'est la répartition de chacun des constituants dans la zone d'analyse qui est ciblée. Cela permet d'avoir une idée sur le processus de formation du calcul pour diagnostiquer les causes et proposer un remède adéquat au patient.

## 2. Le calcul urinaire étudié

Les calculs urinaires étudiés nous ont été aimablement fournis par le professeur A. Addou du laboratoire STEVA de l'université de Mostaganem en Algérie.

Le calcul urinaire étudié a une coloration blanchâtre externe et foncée tendant vers le bleu-marron à l'intérieur. Sa forme est allongée. Les principales informations le concernant sont regroupées dans la carte d'identité ci-dessous. Le donneur est un patient victime de plusieurs récurrences. Cette extraction était la neuvième.

<b>Carte d'identité du calcul urinaire étudié :</b>	
<b>Donneur :</b>	sexe : homme age : 62 ans
<b>Dimensions :</b>	15 x 10 (en mm).
<b>Localisation anatomique :</b>	rein gauche
<b>Mode d'élimination :</b>	opération chirurgicale.
<b>Lieu de provenance :</b>	Mostaganem, Algérie.
<b>Régime alimentaire dominant du patient :</b>	non spécifié
<b>Principaux constituants :</b>	Etude préliminaire par IRTF par l'équipe du professeur Addou : des oxalates, acide urique, des protéines.



### III. Acquisition expérimentale

Quelques détails sur l'appareillage du spectromètre d'acquisition sont donnés. La procédure d'échantillonnage microspectrométrique en Raman et en infrarouge est expliquée au chapitre 2. Dans ce qui suit, nous donnerons les principales informations concernant l'échantillon étudié.

#### 1. Description du spectromètre utilisé

Les calculs urinaires, par la mesure microspectrométrique infrarouge, nous ont posé des difficultés. On atteint une saturation totale dans presque toutes les bandes que ce soit en transmission ou en réflexion même pour des coupes fines et on observe aussi des nombreuses réflexions spéculaires. Ce qui pose un véritable problème pour l'interprétation spectrale.

Quant à la microspectrométrie Raman, une forte fluorescence est observée. Celle-ci perturbe fortement la qualité des spectres acquis.

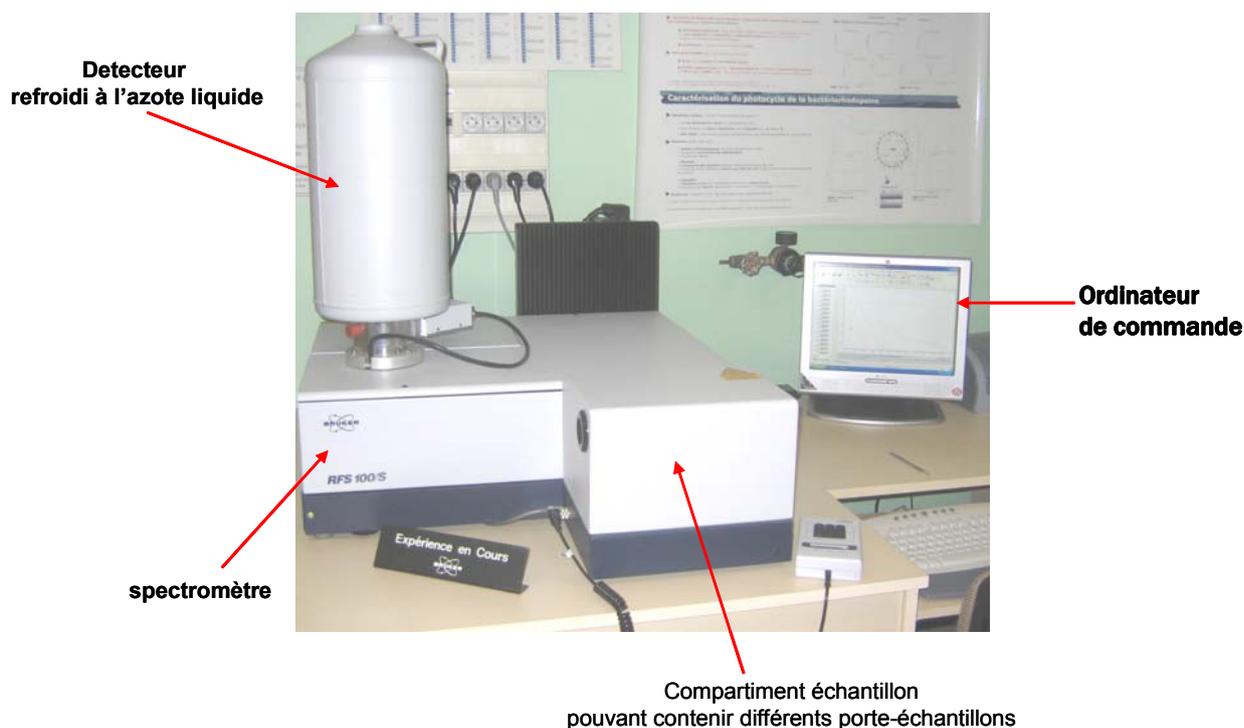


Figure 56 : Spectromètre<sup>185</sup> FT-Raman RFS 100/S utilisé pour l'acquisition spectrale sur le calcul urinaire.

Pour se rapprocher à la fois aux conditions microspectrométriques et diminuer considérablement la fluorescence, nous avons effectué les mesures spectrales en utilisant un spectromètre FT-Raman; le spectromètre RFS 100/S de Bruker récemment acquis par le laboratoire (Cf. Figure 56). Le source Raman utilisé est le laser Nd :YAG dont la longueur d'onde d'émission est de 1064 nm dans le proche infrarouge. Celui-ci, refroidi à air, est totalement contrôlé par le logiciel OPUS<sup>®</sup> via un micro-ordinateur. Dans sa configuration standard, les mesures spectrales peuvent

être effectuées dans les gammes  $3600-70\text{ cm}^{-1}$  pour les déplacements Stokes et  $-100$  à  $-2000\text{cm}^{-1}$  pour le domaine Anti-Stokes avec une résolution d'environ  $2\text{ cm}^{-1}$ .

Grâce à la grande variabilité d'angles de mesure, de porte-échantillons et des éléments optiques, ce spectromètre permet la mesure adéquate de très nombreux types d'échantillons. Le détecteur en germanium refroidi à l'azote liquide permet d'atteindre des bonnes qualités spectrales.

## 2. Acquisition expérimentale des données

Pour la mesure spectrale, il nous a fallu découper le calcul urinaire, le polir pour uniformiser la surface. Après avoir fixé une coupe fine sur un support, nous l'avons polie par un papier de carbure de silicium.

Ensuite nous avons présenté la surface polie à la radiation lumineuse. Sur la surface polie, nous avons délimité une partie aux caractéristiques intéressantes après une exploration globale.

Le but est de réaliser une cartographie (mapping) après avoir délimité l'aire de balayage, défini le pas de balayage et le nombre de spectres à mesurer, réalisé le réglage nécessaire de l'optique. L'aire de mesure spectrale délimitée est d'environ  $9\text{ mm}^2$ . C'est un rectangle de  $3.15\text{ mm}$  de longueur (suivant l'axe des x horizontale) sur  $3\text{ mm}$  de large (suivant l'axe des y vertical). Le pas de balayage est de  $50\text{ }\mu\text{m}$ , c'est-à-dire la distance entre deux points de mesure. Ce qui donne 63 points de mesure suivant x et 60 suivant y. Cela donne un nombre total de spectres mesurés égal à 3780. La gamme spectrale d'enregistrement des spectres est l'intervalle  $[3500 - 156\text{ cm}^{-1}]$  mais celle d'étude est l'intervalle  $[3500 - 400\text{cm}^{-1}]$ .

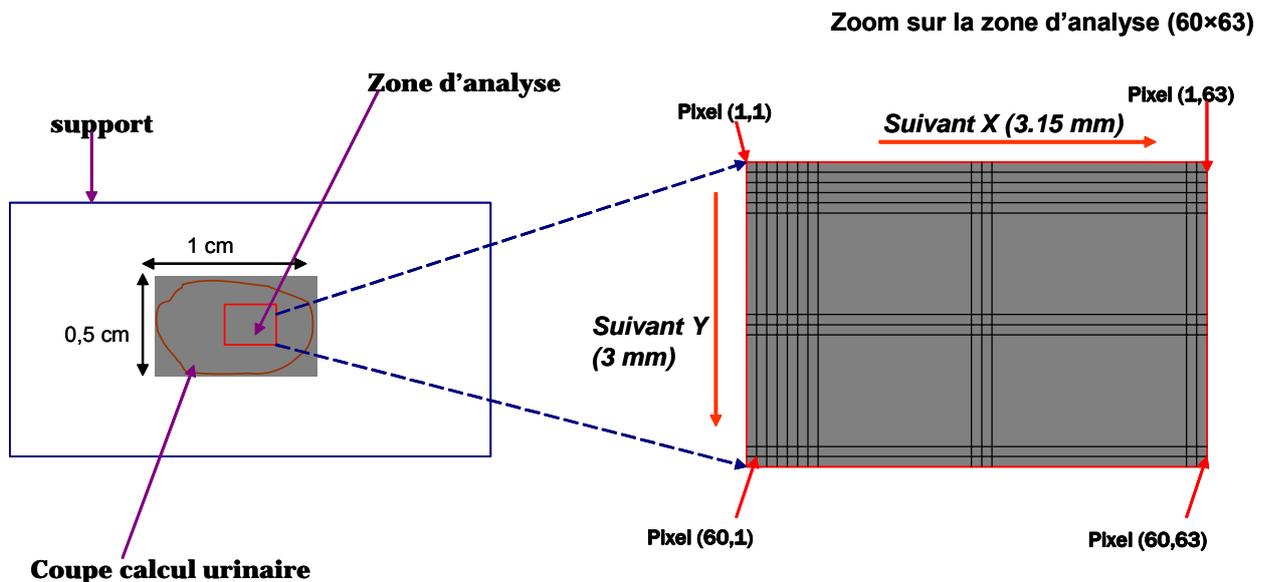


Figure 57 : Schématisation de la zone de « cartographie spectrale » sur la coupe du calcul urinaire.

La procédure de « cartographie spectrale » est celle expliquée au chapitre 2 consacré à l'imagerie microspectrométrique. La zone d'analyse (zone de mapping) du calcul urinaire est schématiquement représentée à la Figure 57.

Chaque spectre mesuré comporte 1735 points de mesure correspondant aux 1735 variables « nombres d'onde » différentes ; ceci est le spectre brut acquis avec la résolution du spectromètre d'environ  $2 \text{ cm}^{-1}$ .

Ces spectres bruts vont être analysés, prétraités et traités par différentes méthodes pour tirer l'information essentielle concernant les principaux éléments constituant ce calcul urinaire. C'est ce que nous aborderons dans la partie suivante.

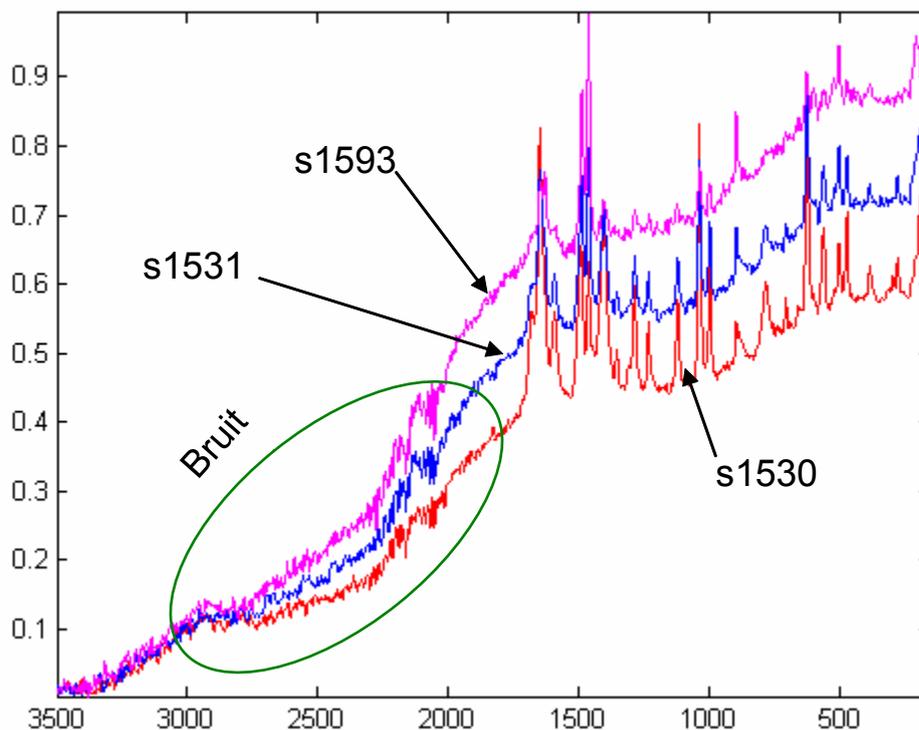
## IV. Etude des données acquises

### 1. Exploration globale, estimation du nombre des constituants et les prétraitements

#### 1.1. Exploration globale et estimation du nombre des constituants

Les données spectrales acquises vont être soumises à une exploration globale afin de tirer les informations pertinentes.

Malgré le laser Nd :YAG utilisé et l'exploration de la surface de l'échantillon du calcul urinaire avant l'acquisition spectrale, les spectres mesurés ne sont pas exempts des défauts. Sur certains spectres, on observe une forte fluorescence causant une déformation de ces spectres avec une grande dérive de la ligne de base. En plus de cela une fluctuation due à la persistance d'un fond de bruit est visible. La Figure 58 présente quelques spectres expérimentaux où la dérive de la ligne de base et la persistance du bruit de fond sont manifestes. Pour tout cela, un prétraitement de ces spectres est nécessaire.

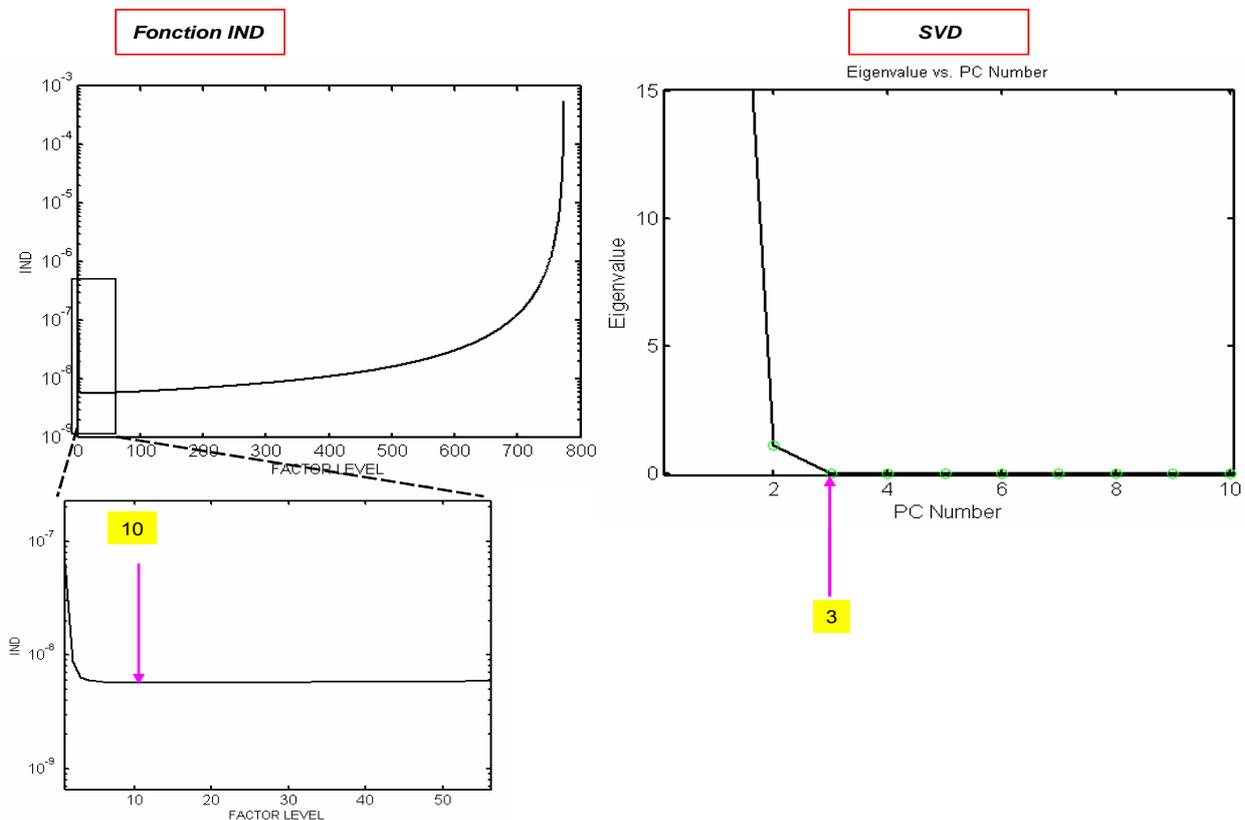


**Figure 58 : Quelques spectres expérimentaux. La dérive de la ligne de base est importante et un bruit de fond persistant est visible.**

Mais avant d'aborder ces différents types de prétraitements, nous allons examiner un aspect très important pour la suite de nos résolutions. En effet, nous avons acquis expérimentalement nos données spectrales, mais nous ne connaissons pas *a priori* le nombre exact de constituants. Nous

devrons estimer ce nombre en utilisant quelques unes des méthodes d'estimation que nous avons définies dans le paragraphe II du chapitre 1.

Les méthodes d'estimation que nous allons employer sont la fonction IND, la SVD, SIMPLISMA, OPA et FSIW-EFA. Pour ne perdre aucune information, nous avons réalisé ces estimations sur la matrice de données brutes interpolée. Le bruit résiduel, ou la dérive de la ligne, est traité comme un des facteurs.



**Figure 59 : Estimation du nombre de constituants par la fonction IND et par la SVD. IND fournit 10 facteurs alors que la SVD n'indique que 3.**

Sur les figures (Figure 59 et Figure 60), nous avons représenté les résultats graphiques pour mieux apprécier la qualité d'estimation de chaque méthode. La fonction IND fournit sans ambiguïté un nombre de facteurs égal à 10 alors que la SVD nous indique qu'il n'y a que 3 facteurs dominants. Les données expérimentales étant bruitées, la fonction IND inclut dans « les facteurs significatifs » un nombre de facteurs supplémentaires. Il y a donc forcément une surestimation de facteurs. Pour avoir plus d'idées, nous avons utilisé la méthodologie d'extraction par SIMPLISMA, OPA et FSIW-EFA. Les estimations par SIMPLISMA et OPA sont similaires. Nous avons représenté l'aspect du spectre de pureté pour SIMPLISMA et du spectre de dissimilarité pour OPA avant sélection et après avoir sélectionné 5 variables. Nous nous sommes arrêtés à la 5<sup>ème</sup> sélection ; car à ce stade de sélection le spectre de pureté ou le spectre de dissimilarité prend un aspect ne manifestant que le bruit de fond comme nous avons déjà expliqué au paragraphe III.3.1 du chapitre 1. Quant à FSIW-EFA (cf. Figure 70), nous

avons extrait jusqu'à 9 variables. Nous voyons bien que parmi ces 9, il n'y en a que 3 qui sont bien distinctes. Ce sont celles indiquées par les nombres 1, 2 et 3. Du 4<sup>ème</sup> au 9<sup>ème</sup>, nous avons pratiquement la même image (même distribution). Cela nous indique que 3 constituants peuvent être extraits sans ambiguïté.

Remarquant cette grande disparité du nombre de facteurs estimé par ces méthodes, nous nous sommes posés la question de savoir quelle serait la meilleure valeur à retenir. En effectuant une exploration plus approfondie en s'appuyant sur les scores et les loadings de PCA, les spectres et les images extraits par SIMPLISMA et OPA, les distributions des valeurs singulières fournies par FSIW-EFA, le nombre 5 est celui qui représenterait le mieux le nombre de facteurs significatifs. SVD et FSIW-EFA ne retiennent que les facteurs les plus dominants. C'est une attitude proche de la réalité. Pour cela, nous avons décidé à extraire 5 facteurs et parmi ceux-ci, nous retrouverons les principaux composés présents dans l'échantillon étudié.

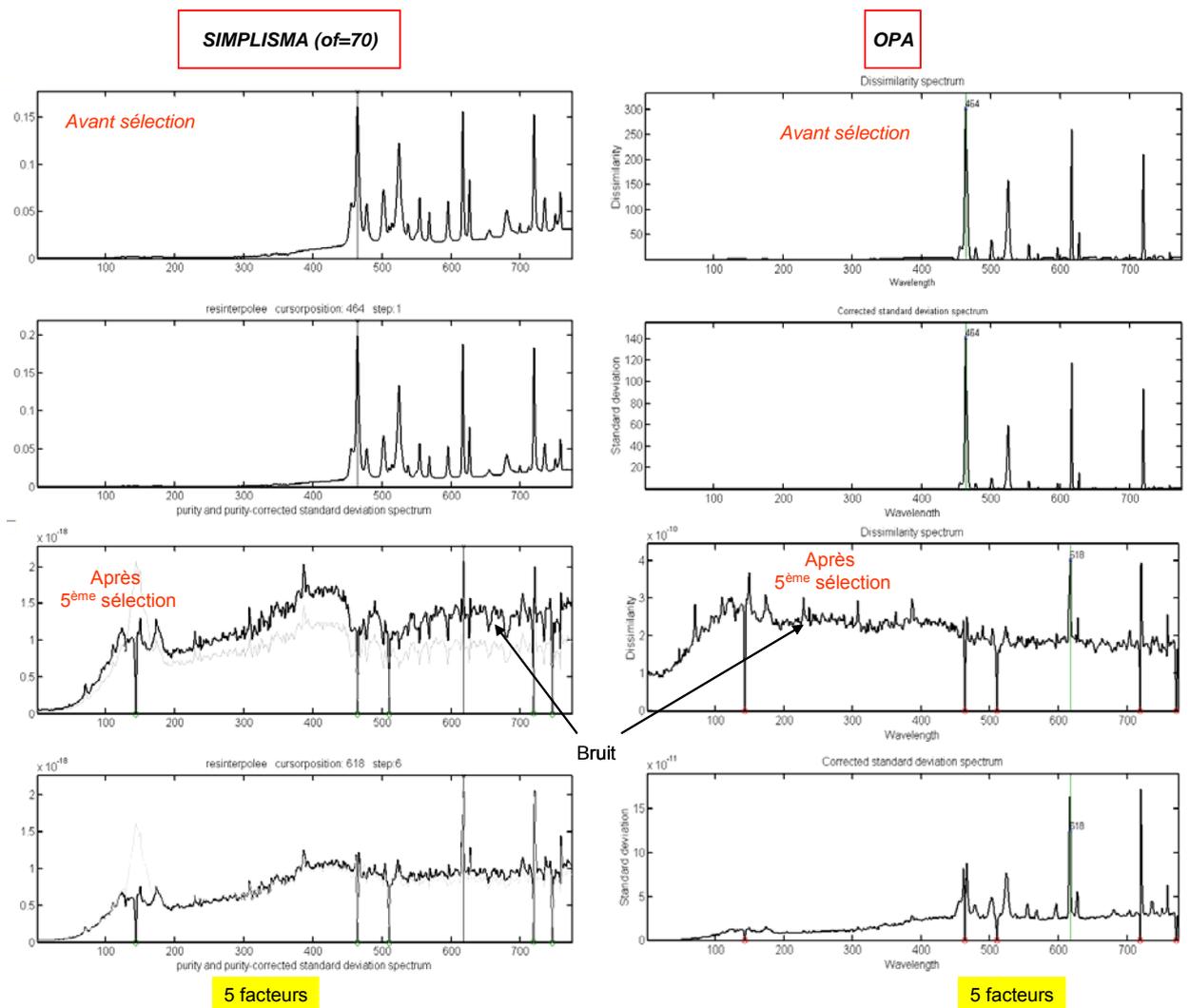


Figure 60 : Estimation du nombre de constituants par SIMPLISMA et par OPA. Nous obtenons une estimation maximale de 5 facteurs.

Ne retenir que 3 facteurs s'apparente à sous-estimer le nombre de facteurs. Et un nombre supérieur à 5 nous est apparu comme une surestimation qui nous mènera à une mauvaise qualité

d'extraction. En considérant 5 facteurs, on peut espérer extraire correctement 3 constituants, les deux autres correspondant à des composés minoritaires, des artefacts ou du bruit. Le Tableau 14 suivant résume les différentes méthodes d'estimation, le nombre estimé par chacune d'elle et une interprétation dans chaque cas.

Ayant fixé le nombre de constituants à extraire, nous allons donner brièvement la description de quelques prétraitements nécessaires pour l'amélioration de la qualité spectrale des données à traiter.

**Tableau 14 : Résultats d'estimation fournis par les méthodes.**

<b>IND function</b>	<b>10</b>	Un bon minimum pour la fonction mais indique un nombre élevé de facteurs
<b>SVD</b>	<b>3</b>	Bonne estimation
<b>SIMPLISMA (offset=70)</b>	<b>5</b>	Légère surestimation ou existence composés trop minoritaires
<b>OPA(var) OPA(sp)</b>	<b>5 5</b>	Légère surestimation ou composés trop minoritaires
<b>FSIW-EFA</b>	<b>2 bien clair Le 3<sup>ème</sup> minoritaire</b>	Légère sous-estimation

## 1.2. Les prétraitements nécessaires<sup>186 187 188</sup>

Il existe de très nombreuses façons de faire un prétraitement de données spectrales. Cela dépend du type de données à traiter, donc du domaine spectral d'étude, et de ce qu'on veut obtenir. Un prétraitement est une ou plusieurs transformations des collections spectrales dans le but de les rendre plus facilement exploitables par la suite. Dans notre cas, il s'agit d'aboutir à une meilleure interprétation des données spectrales. Nous appliquerons à nos collections spectrales les prétraitements nécessaires pour « améliorer le signal ». Il s'agira en effet de corriger adéquatement la ligne de base, de réduire le bruit de fond et de normaliser les spectres.

### 1.2.1. La normalisation

La normalisation consiste d'une manière générale à supprimer ou à diminuer l'effet lié à la variation incontrôlée de l'intensité des spectres. Pour cela, on peut s'y prendre de plusieurs façons selon le type de données à traiter. Par la normalisation, on peut imposer que les spectres aient la même hauteur ou la même longueur (dimension) par exemple. Ce qui revient

globalement à diviser chaque valeur  $x_i$  d'un spectre  $x$  par un nombre  $\alpha$  qui peut être la valeur maximale d'absorbance, la somme de tous les  $x_i$  ou la moyenne du vecteur  $x_i$ .

D'autres types de normalisations existent comme la soustraction de la moyenne, par exemple, ou l'application de la SNV (Standard Normal Variate).

Dans notre étude, nous avons veillé à ce que les spectres aient la même longueur dans tous les cas et la même hauteur lorsque cela s'avérait nécessaire.

Dans la procédure de comparaison des spectres que nous verrons par la suite, chaque spectre est normalisé par la division par la valeur maximale d'intensité.

### 1.2.2. La correction de la ligne de base

Sans déviation, la ligne de base d'un spectre est proche de l'axe des abscisses. Dans la plupart des spectres acquis expérimentalement une dérive est présente. On cherche à éliminer cette dérive par des traitements simples car le signal est supposé être composé d'un certain « bruit de fond » et le signal contenant « l'information utile ». Ce qu'on cherche à soustraire, c'est la contribution de ce « fond de bruit » soit par une correction linéaire soit par une correction polynomiale.

Nous avons utilisé la correction linéaire qui est la plus simple et qui consiste à tracer une ligne brisée là où il n'y a pas de bandes d'absorption (points au plus bas creux de la courbe).

A partir de ces points, on estime le spectre de fond. Ensuite l'opération consiste à soustraire cette ligne de base (spectre de fond) du spectre d'origine.

La Figure 61 présente les prétraitements que nous avons réalisés. Nous voyons que les données spectrales présentent une forte fluorescence qui entraîne une forte déviation de la ligne de base pour bon nombre de spectres. Le spectre 1531 illustre bien cela.

Cependant une autre difficulté se présente. Malgré tous les types de correction linéaire essayés, il apparaît une partie négative sur le spectre corrigé et cela est bien mis en évidence sur la Figure 61b (Cf. rectangle rouge). Pour remédier à cela, nous avons essayé de supprimer un intervalle plus ou moins grand correspondant à cette zone pour obtenir le meilleur résultat possible.

Après plusieurs essais d'intervalles divers et résolutions par les méthodes sélectionnées pour juger les résultats, l'intervalle retenu donnant les meilleurs (du point de vue de la reproductibilité, de la répétabilité et de stabilité face aux différentes méthodes de traitement) est la gamme spectrale  $[3500-2600] \cup [1800-400\text{cm}^{-1}]$ . La partie supprimée  $[2600-1800\text{cm}^{-1}]$  est une partie du rectangle rouge sur la Figure 61b.

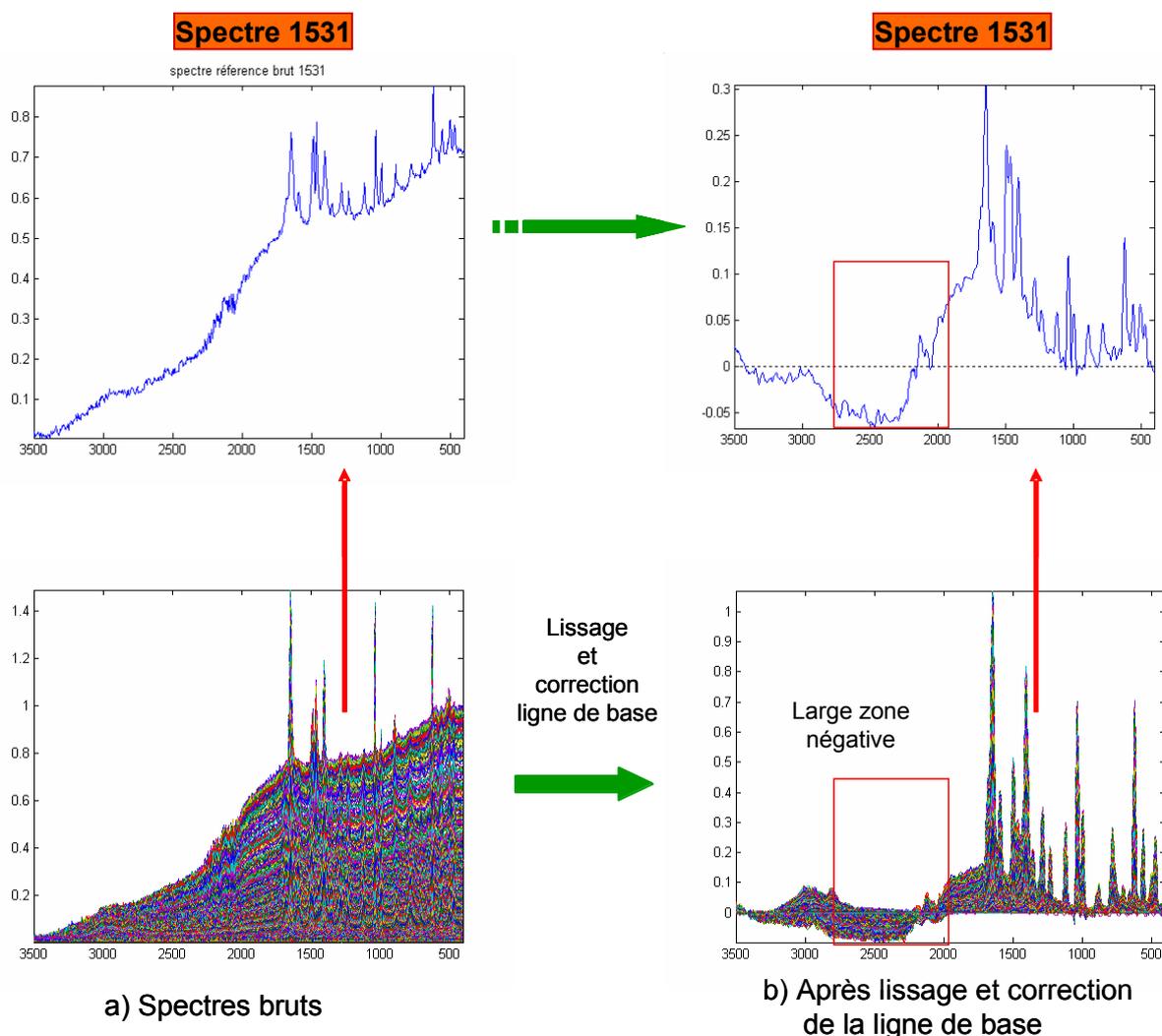


Figure 61 : Prétraitements de nos collections spectrales ; a) données spectrales brutes avant prétraitement, b) données prétraitées avec correction de la ligne de base et lissage. Le spectre 1531 donne une illustration de ces deux cas.

D'autre part, les spectres expérimentaux et ceux de la référence n'ont pas la même longueur (dimension) ni la même résolution. Les spectres de la référence sont enregistrés pour 775 points, c'est-à-dire pour 775 variables nombres d'onde avec une résolution de  $4 \text{ cm}^{-1}$ . Pour comparer deux spectres qui sont pris comme deux vecteurs, il faut qu'ils aient la même longueur (c'est-à-dire mesurés pour le même nombre de longueurs d'onde) et la même résolution. Pour réaliser cette condition une interpolation linéaire des spectres de la matrice de données est effectuée. Cette interpolation de la matrice des données réduit la longueur des spectres expérimentaux de 1735 à 775 points. La suppression de l'intervalle  $[2600-1800 \text{ cm}^{-1}]$  réduit encore la dimension à 576 points. Dans tous les traitements qui vont suivre, la matrice des données à traiter a une dimension de  $3780 \times 576$ . La Figure 62 montre globalement les différentes étapes de notre étude, en passant par l'étape d'analyse exploratoire et de prétraitements. Quant aux traitements chimiométriques, nous les verrons dans les parties suivantes.

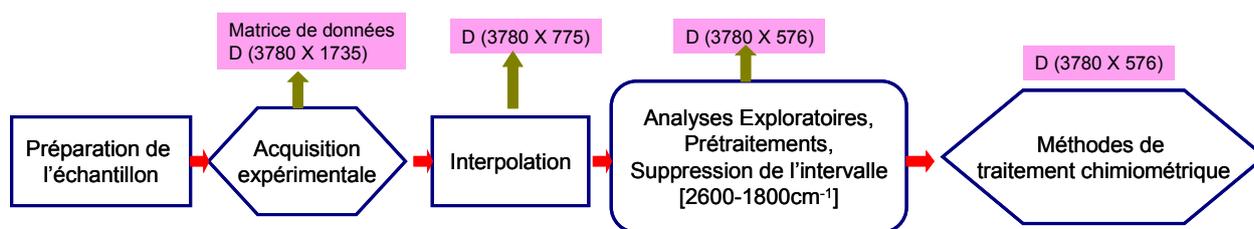


Figure 62 : Les différentes étapes de notre étude, de la préparation de l'échantillon aux traitements chimiométriques.

### 1.2.3. Le lissage

Bien que la correction de la ligne de base ait fortement amélioré l'aspect des spectres expérimentaux, il persiste cependant une bonne fluctuation due à la part du bruit résiduel. Pour améliorer davantage la qualité de ces spectres, un lissage est donc nécessaire. On peut faire ce lissage de plusieurs façons. Le lissage couramment utilisé par les spectroscopistes est celui mis au point par Savitzky et Golay<sup>189,190</sup> en 1964 et constamment perfectionné. Certaines précautions sont à prendre telles que la largeur de la fenêtre de lissage et le degré du polynôme qui doivent être adaptés à la nature du signal à étudier. Il est à noter, plus la fenêtre de lissage est large plus le rapport S/N (signal/bruit) est meilleur mais l'atténuation et l'élargissement des pics sont plus grandes. Sur la Figure 61b, le spectre 1531 illustratif présente cet aspect du lissage et on observe que la fluctuation due au bruit résiduel est considérablement diminuée. Il faut préciser cependant que dans certains logiciels, comme *Unscrambler* (6.5) par exemple, le spectre lissé est réduit de deux extrémités. Cette réduction à chaque extrémité étant égale au nombre de points dans la fenêtre de lissage. La fonction *Savgol*, qui réalise le lissage de Savitzky-Golay dans Matlab<sup>®</sup>, n'effectue pas cette réduction et le spectre lissé a la même longueur que celui d'origine. Dans notre cas, nous avons utilisé cette fonction avec un polynôme de degré 2 et une fenêtre de lissage de 15 points.

D'autres types de corrections peuvent être appliquées comme la dérivation (première ou seconde) et la « *multiplicative scatter correction* (MSC) ».

### 1.3. La méthodologie de résolution proposée pour l'étude des calculs urinaires

Comme nous venons de voir, avant d'aborder les traitements par les méthodes chimiométriques, il nous a fallu une exploration globale des données expérimentales acquises. Cela nous a permis à la fois d'évaluer le nombre de facteurs significatifs, c'est-à-dire les composés « extractibles », et de repérer grâce à une bibliothèque les composés de référence potentiellement candidats dans les données à étudier. L'organigramme de la Figure 63 présente schématiquement les différentes étapes. La première étape, l'étape d'exploration, est l'une des plus importante qui nous a pris beaucoup de temps pour avoir une idée assez claire des composés présents dans les données à

étudier ainsi que leur nombre. Dans les parties précédentes, nous avons présenté les résultats de ces explorations. La partie « repérage des candidats potentiels » nous a aidé à cadrer nos choix mais la méthodologie étant assez lourde nous ne la présentons pas. Après toutes les manipulations et les prétraitements effectués, ayant obtenu des données spectrales de bonne qualité, nous allons d'abord les soumettre aux étapes de résolution 2 à 5 comme explicité à la Figure 63. Dans chaque cas, nous analyserons la qualité des images et des spectres extraits.

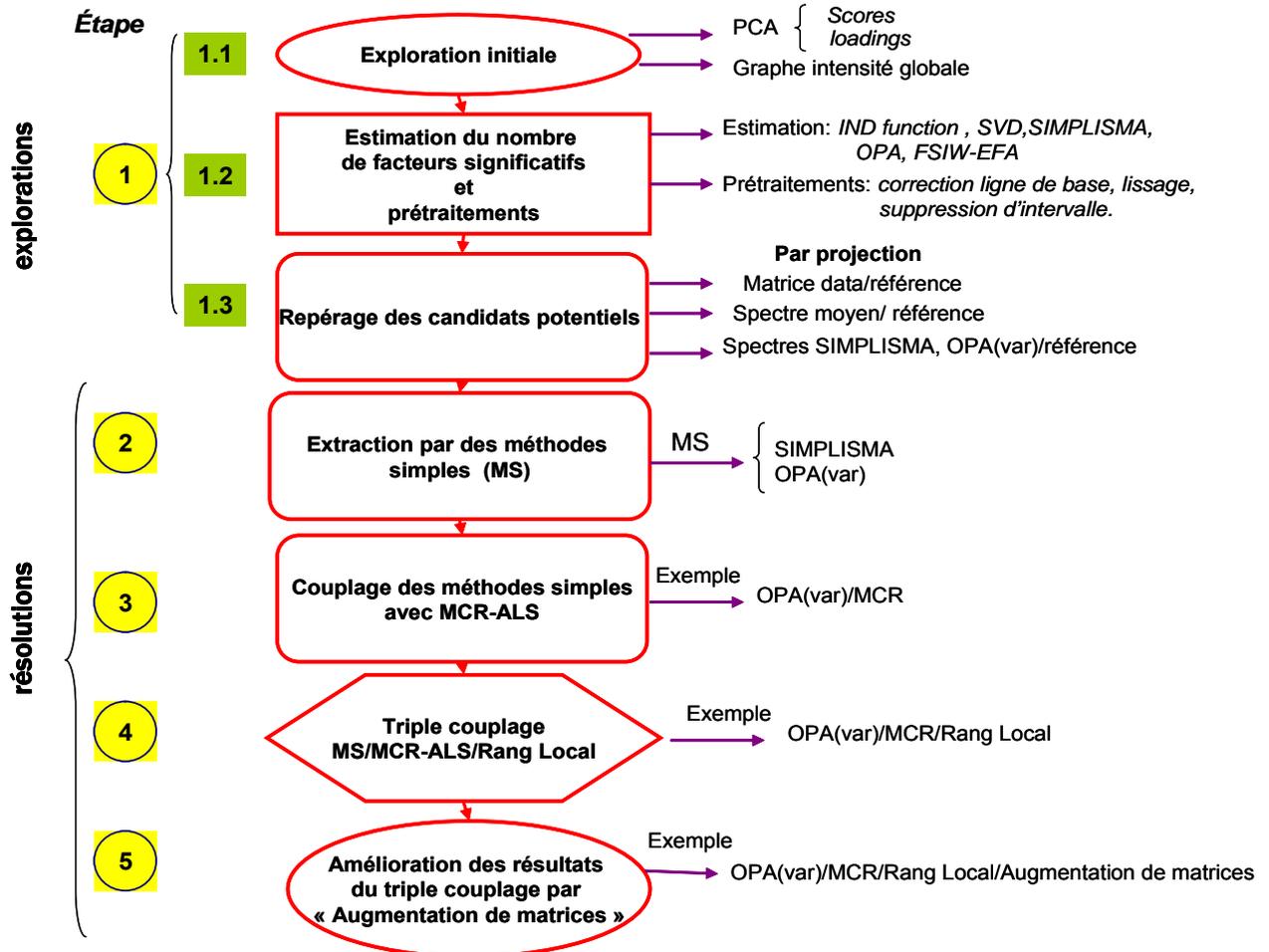


Figure 63 : Méthodologie générale de résolution adoptée pour l'analyse du calcul urinaire.

- **Données de référence.**

Pour faire notre analyse, nous allons nous appuyer sur une bibliothèque spectrale regroupant les spectres purs Raman de 105 composés qu'on trouve essentiellement dans les calculs urinaires. Nous disposons ces spectres sous forme numérique dans le CD accompagnant le très documenté livre de Nguyen Quy Dao<sup>191</sup> et Michel Daudon. Ce livre nous a été un précieux recours pour identifier, distinguer, départager, juger l'importance relative des éléments et apprécier la qualité des éléments extraits. Parmi ces 105 composés seulement 6 sont fournis sans spectre Raman et dans la matrice de référence, ils ont des valeurs nulles.

Lorsque nous allons extraire des spectres à partir de la matrice de données expérimentales par les méthodes de traitement chimiométrique, nous comparerons ces spectres extraits à ceux de la

bibliothèque de référence regroupés dans une matrice Y appelée matrice de référence. Pour cela nous avons calculé le coefficient de corrélation de chaque spectre extrait par rapport à chacun de 105 spectres de la référence. Ceux ayant les plus grands coefficients de corrélation sont sélectionnés et comparés visuellement au spectre extrait correspondant en ce qui concerne la forme et la position des bandes. Dans certains cas, beaucoup de spectres de référence sont candidats au même spectre extrait. Dans d'autres cas, le nombre de candidats est assez réduit et l'attribution est facilitée. Dans les parties suivantes, nous approfondirons la discussion dans chaque cas pour aboutir à la meilleure sélection possible. Nous avons construit les routines Matlab<sup>®192</sup> nécessaires aux comparaisons et à la discrimination des candidats (cf. Annexe 2).

## 2. Extraction par quelques « méthodes simples » avec la seule contrainte de non-négativité

Dans cette partie, l'objectif est d'extraire les principaux composés du mélange en appliquant les méthodes chimiométriques que nous avons qualifiées de « méthodes simples ». La décomposition est faite pour 5 facteurs. Dans le chapitre 3, nous avons démontré que la méthode d'extraction OPA(var) possédait la meilleure qualité d'extraction. Nous appliquerons cette méthodologie aux données expérimentales. Pour renforcer notre argumentation, nous réaliserons une comparaison parallèle des résultats extraits par SIMPLISMA dans cette première partie.

### 2.1. Extraction par OPA(var)

La résolution est appliquée à la matrice de données prétraitée. La Figure 64 présente les résultats. Sur la partie A, nous avons représenté l'ensemble des images et des spectres extraits. Par les puces rouges, nous avons désigné les 3 composés identifiables. Les composés extraits 2 et 5 ne sont pas identifiables par leur spectres. L'image associée au 2<sup>ème</sup> est nette et ressemble à celle du 4<sup>ème</sup> mais le spectre correspondant est d'une allure aléatoire semblable à celle du bruit de fond. Quant au 5<sup>ème</sup> élément, son image extraite indique du bruit alors que son spectre ne correspond à aucun spectre de notre référence.

Sur la partie B, nous avons représenté les images et les spectres de 3 composés identifiables. L'ordre correspond à celui de leur importance relative dans l'échantillon. Le nombre au dessus de chaque spectre indique le coefficient de corrélation (r) du spectre extrait avec celui de la référence correspondante.

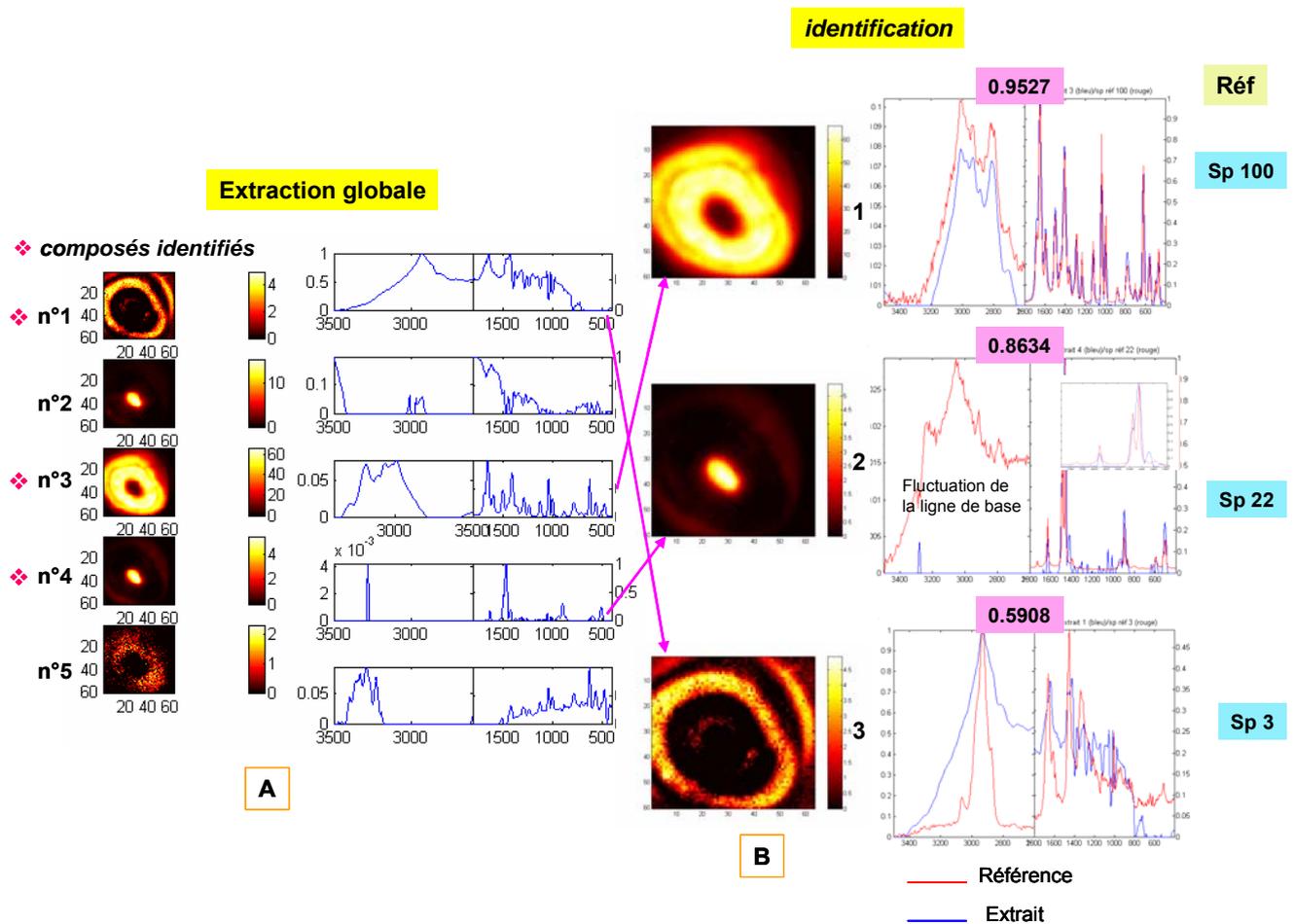


Figure 64 : Images et spectres extraits par la « méthode simple » OPA(var). A) visualisation globale de l'ensemble des images et spectres extraits ; B) représentation des composés extraits identifiés et comparaison de leur spectre avec celui de la référence correspondante.

### Identification du composé extrait 1.

Le composé 1 (cf. Figure 64B) est celui qui est le mieux identifié avec un coefficient de corrélation supérieur à 0,95 avec le composé de référence numéro 100 (Sp 100) qui est l'acide urique anhydre.

Cependant l'autre forme, l'acide urique dihydraté, a avec celui-ci un coefficient de corrélation non négligeable de l'ordre de 0,9. Ce qui voudrait dire probablement qu'elle est présente, ne serait-ce qu'à l'état de traces, dans l'échantillon. La Figure 65 compare le spectre du composé extrait avec ceux de deux composés de référence. Dans l'intervalle  $[3500-2600 \text{ cm}^{-1}]$ , nous observons le bon accord de l'extrait avec l'acide urique anhydre.

En regardant plus en détail, la position des bandes et la forme, ce composé 1 correspond bien à l'acide urique anhydre. Le numéro de chaque composé de référence identifié par chaque spectre extrait est indiqué à droite des représentations dans la colonne Réf sur la Figure 64. Le spectre du composé de référence est en rouge et celui extrait en bleu (cf. Figure 64B). L'image du composé indique la répartition de la concentration de celui-ci dans la zone d'analyse. L'échelle associée indique l'ordre de grandeur de celle-ci sur chaque pixel.

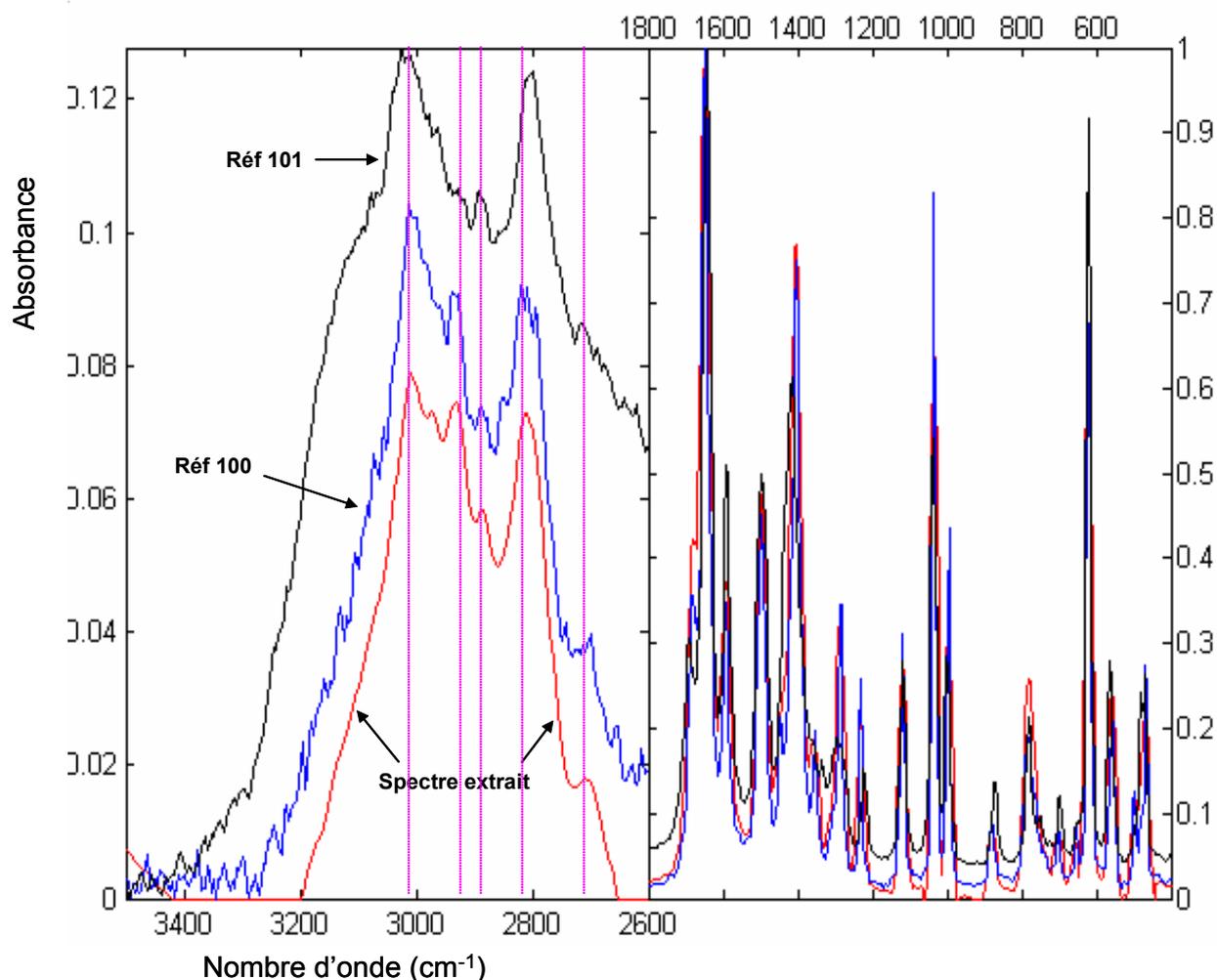


Figure 65 : Spectre du composé extrait 1 (rouge) avec les spectres de l'acide urique anhydre (bleu) et l'acide urique dihydraté (noir). L'accord est meilleur avec l'acide urique anhydre.

### Identification du composé extrait 2.

Le composé 2 (cf. Figure 64B) est extrait aussi avec un fort coefficient de corrélation de l'ordre de 0,86 avec le composé de référence 22 (oxalate de calcium monohydraté : whewellite). Cependant un autre candidat potentiel est le 21 (oxalate de calcium dihydraté : weddellite). Ce qui nous a permis de discriminer ces deux composés est à la fois la position des bandes et le nombre de pics entre 1500 et 1440  $\text{cm}^{-1}$ .

Dans cet intervalle le 22 possède 2 pics alors que le 21 n'en possède qu'un. La Figure 66b visualise bien cette différence avec un zoom dans cette partie. Pour mieux montrer cet aspect important sur le spectre extrait, nous avons fait un zoom dans cette partie sur la Figure 64B comparant les éléments extraits à ceux de la référence. Ce composé 2 extrait est bien en accord avec le composé de référence 22 dans l'intervalle 1800-400  $\text{cm}^{-1}$ . Quant à l'intervalle 3500-2600  $\text{cm}^{-1}$ , la seule différence qu'on observe est la fluctuation de la ligne de base car la valeur la plus grande de l'absorbance est de l'ordre de 0,03. En effet dans chaque partie, l'échelle s'adapte à l'écart entre les deux valeurs extrêmes, la minimale et la maximale, des absorbances. Quant à

l'image associée à ce composé, elle nous indique que ce composé est présent uniquement au centre de la zone d'analyse

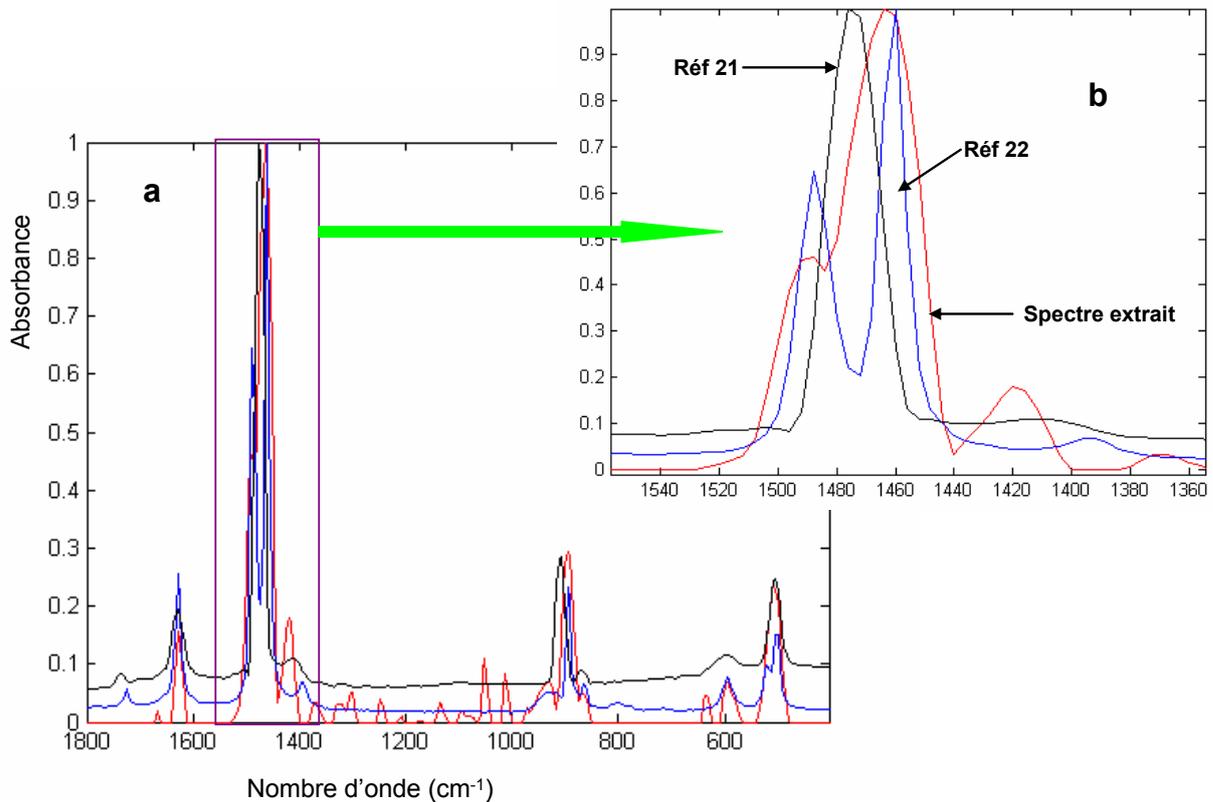


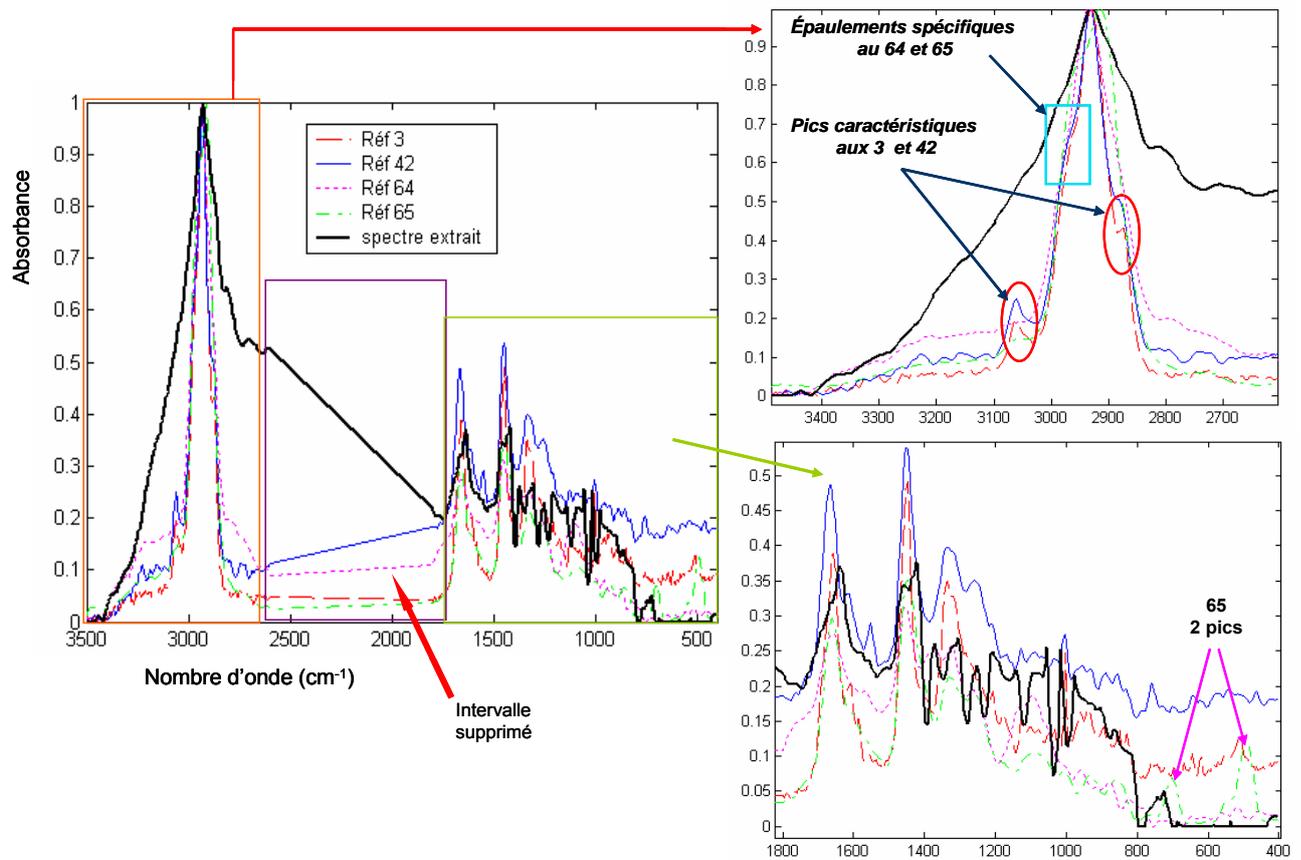
Figure 66 : Discrimination entre les composés de référence 21 et 22. a) comparaison des spectres de 2 composés de référence 21 et 22 avec le composé extrait 2 (rouge) dans l'intervalle 1800-400  $\text{cm}^{-1}$  ; b) un zoom dans la partie 1530-1360  $\text{cm}^{-1}$  pour mieux voir l'accord du 22 avec l'extrait.

### Identification du composé extrait 3.

Pour le 3<sup>ème</sup> composé extrait identifiable (cf. Figure 64B), de nombreux candidats se présentent. Ceux-ci sont assez semblables du point de vue de la forme. Une analyse approfondie en ce qui concerne la position des bandes est nécessaire. Les 4 composés de référence les plus candidats pour ce 3<sup>ème</sup> composé extrait sont : le 64 (*mucine*) qui a le plus grand coefficient de corrélation avec celui-ci, le 65 (*mucopolysaccharides*) qui le suit, le 42 (*fibrine*) et le 3 (*albumine*). Ces deux derniers sont les plus ressemblants du point de vue de la forme à ce 3<sup>ème</sup> composé extrait. Les spectres de ces 4 composés, représentés sur la Figure 67 avec la même échelle, sont fort ressemblants comme on peut les constater. Quelques détails nous permettront de les distinguer cependant (Cf. Tableau 15).

Pour les discriminer, nous avons entrepris de relever les principaux pics caractéristiques pour chacun des spectres de référence (Cf. Tableau 15). Ensuite en pointant les pics du spectre extrait et en les comparant à ceux de ces 4 spectres de référence, nous avons essayé de savoir lequel s'accordait le mieux, en ce qui concerne la forme et la position des bandes, au spectre de ce 3<sup>ème</sup>

composé. Les pics, qui sont les plus caractéristiques à chacun d'eux (Cf. Tableau 15), permettent de juger leur différence.



**Figure 67 : Comparaison du spectre du composé extrait 3 avec les spectres de 4 composés de référence candidats. Pointages de quelques pics caractéristiques à ces spectres de référence.**

Ce sont les nombres en gros caractère gras, italique et souligné dans ce Tableau 15. Dans ce tableau, nous avons reporté aussi quelques bandes du composé extrait à identifier. Les rectangles rouges indiquent une absence des bandes. Cela peut être une bonne indication pour départager les composés candidats.

Nous avons fait ces vérifications dans ce cas présent pour OPA(var). Nous en ferons de même dans les autres cas que nous verrons par la suite. Pour OPA(var), nous avons remarqué que le spectre extrait s'accordait bien en forme et en position de bandes avec le 3 et 42. Ce qui nécessite encore la recherche des détails pour départager ces deux derniers. En outre, ce qui discrimine un peu plus le 64 et le 65 par rapport au 3 et 42, c'est leur occurrence très faible. Le 64 est un composé très minoritaire dans la matrice des protéines dans les calculs urinaires. Le 65 est aussi associé aux protéines ou aux triglycérides avec une occurrence inférieure à 0,5%.

Tableau 15 : Principales bandes caractéristiques (en  $\text{cm}^{-1}$ ) de chacun de 4 composés. Les nombres en gros caractère gras, italique et souligné indiquent les pics spécifiques qui nous permettent de différencier les 4 composés. Les zones colorées indiquent l'absence des pics qui peut être intéressante pour la distinction. La dernière ligne correspond au composé extrait à identifier.

Réf 3	<b><u>3063</u></b>	2980		<b><u>~2880</u></b>	1668	1608		1449	1337			1003	<b><u>938</u></b>		505
Réf 42	<b><u>3061</u></b>	2982		<b><u>2880</u></b>	1666	1615	1552	1450	<b><u>1335</u></b>	<b><u>1255</u></b>		1005			<b><u>759</u></b>
Réf 64		2986		2882	1668			1454	1323		<b><u>1096</u></b>				
Réf 65		<b><u>2962</u></b>	<b><u>2912</u></b>		1660			1448	1335		<b><u>1091</u></b>		<b><u>927</u></b>	<b><u>701</u></b>	<b><u>492</u></b>
Sp extrait	2985	2929		2812	1638			1456	1370	1250	1140	1016	903	724	535

La discrimination du 3 et du 42 est assez délicate. La distinction ne peut se faire que dans l'intervalle  $3500\text{-}2600\text{ cm}^{-1}$ . La Figure 67 montre l'absence des différences dans l'intervalle  $[1800\text{-}400\text{ cm}^{-1}]$  car les bandes spécifiques à ces deux composés présentent une forte similitude. Les quelques bandes bien spécifiques pour 3 ( $3063, \sim 2875, 938$  et  $505\text{ cm}^{-1}$ ) et pour 42 ( $3061, 2880, 1552, 759\text{ cm}^{-1}$ ) sont déterminantes pour l'attribution du composé correspondant au spectre extrait. Après cette comparaison, l'analyse penche pour le 3, l'albumine. Il faut préciser cependant que le coefficient de corrélation avec ce dernier est de l'ordre de 0,60. C'est ce qui explique, en partie, la difficulté d'attribution que nous avons rencontrée.

Tableau 16 : Récapitulatif de l'analyse des éléments extraits par OPA(var) et de l'identification des éléments présents dans l'échantillon. S1, S3, S4 spectres des composés extraits dans l'ordre d'extraction.

Spectre Extrait (n° d'ordre)	Référence	Coefficient de corrélation
S 3	100 : acide urique anhydre	0.9527
S4	22 : oxalate de calcium monohydraté	0.8634
S1	3 : protéine (albumine?)	0.5908

Dans cette partie, nous avons extrait les principaux constituants par OPA par sélection des variables en ne retenant que 5 facteurs significatifs pour la décomposition.

Parmi les 5 éléments extraits, nous n'avons que 3 qui sont identifiables. Le Tableau 16 résume les résultats de l'analyse d'identification et d'attribution des éléments extraits.

## 2.2. Confirmation des résultats de OPA(var) par l'extraction par SIMPLISMA

Pour confirmer notre analyse pour le cas précédent et donner un champ plus large, dans cette partie nous exposons les résultats fournis par SIMPLISMA.

La méthodologie d'analyse et d'identification est la même que celle utilisée pour OPA(var). La Figure 68 présente les résultats des extractions, extraction globale (partie A) et identification (partie B).

Pour l'extraction globale, il se manifeste la qualité d'extraction moindre de SIMPLISMA puisque nous voyons bien que les images correspondant aux différents éléments extraits sont assez semblables. Les spectres sont cependant significativement différents entre eux avec quelques déformations dues à l'absence des contraintes.

Quant à la partie identification, nous obtenons bien les 3 composés extraits identifiables parmi les 5. Ces 3 composés extraits sont bien identifiés par leurs spectres qui sont de qualité suffisante. La représentation graphique comparative montre un résultat semblable à celui d'OPA(var) avec toutefois des coefficients de corrélation un peu plus faibles. Quant aux images, il en est autrement comme nous venons de préciser pour l'aspect global. Les images des composés 1 (Sp 100) et du composé 3 (Sp 3) sont très peu différentes. Ce qui montre bien la moindre qualité d'extraction de SIMPLISMA comme déjà expliqué au chapitre 3. La répartition des deux composés dans l'échantillon semble identique, ce qui *a priori* n'est pas correct. Seul le troisième composé extrait semble avoir un meilleur accord établi par un coefficient de corrélation supérieur.

Le Tableau 17 résume les résultats d'extraction et d'identification. Du point de vue d'extraction spectrale, les résultats de SIMPLISMA confirment ceux de OPA(var). Et on identifie bien les 3 mêmes composés dans l'échantillon qui sont l'acide urique anhydre, l'oxalate de calcium monohydraté et une protéine (l'albumine est la plus probable).

Par la suite, dans notre analyse, nous nous limiterons à l'interprétation des résultats extraits par OPA(var) et son association avec les autres types de traitement. En effet, nous avons démontré d'une manière claire et évidente au chapitre 3 sa robustesse face à des conditions expérimentales difficiles, ce qui se trouve confirmé ici dans un cas d'extraction réelle.

Dans la partie suivante, nous couplerons cette méthode OPA(var) retenue à MCR-ALS.

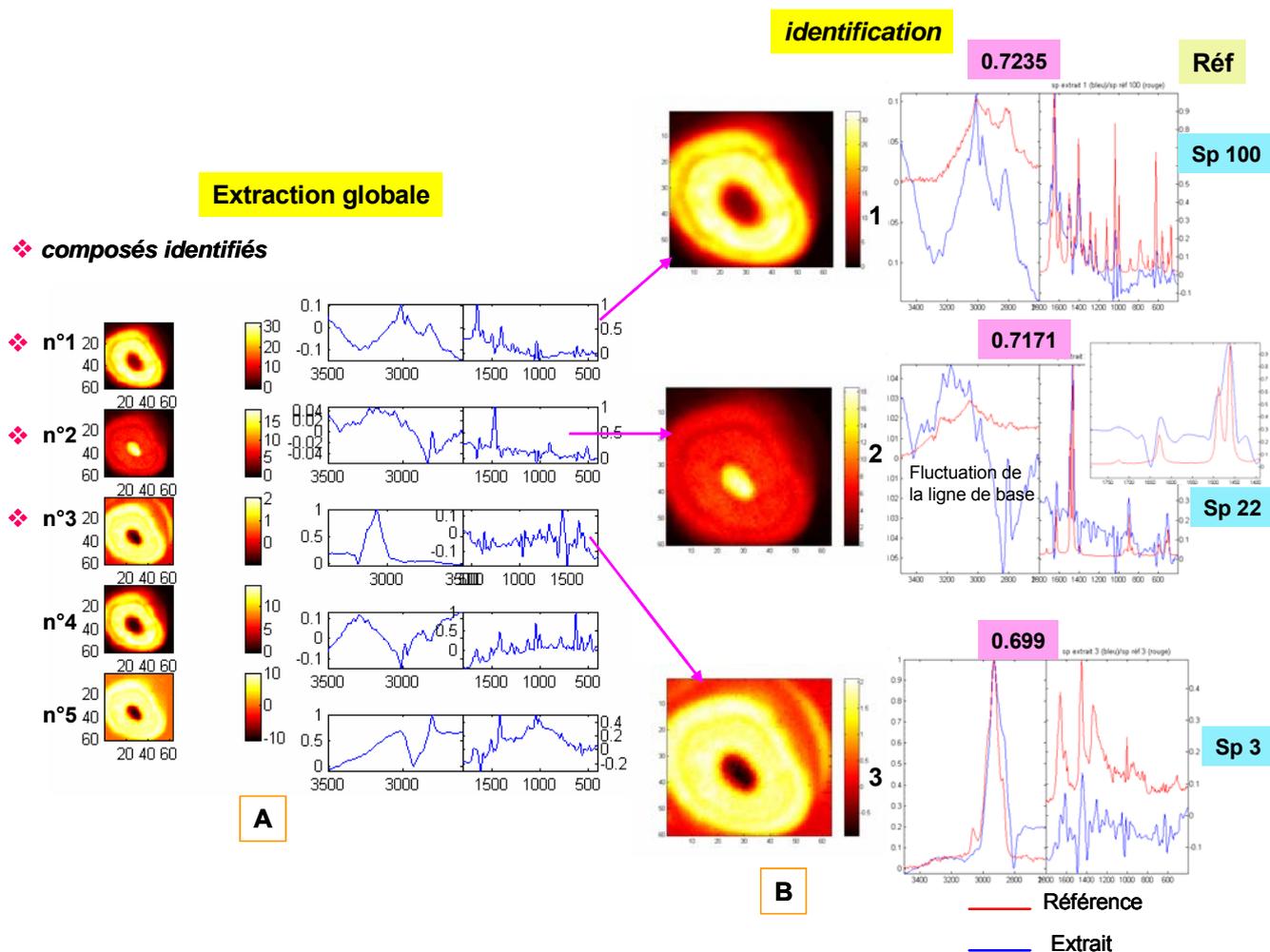


Figure 68 : Images et spectres extraits par la « méthode simple » SIMPLISMA. A) visualisation globale de l'ensemble des images et spectres extraits ; B) représentation des composés extraits identifiés et comparaison de leur spectre avec celui de la référence correspondante.

Tableau 17 : Récapitulatif de l'analyse des éléments extraits par SIMPLISMA et de l'identification des éléments présents dans l'échantillon. S1, S2, S3 spectres des composés extraits dans l'ordre d'extraction.

Spectre Extrait (n° d'ordre)	Référence	Coefficient de corrélation
S1	100 : acide urique anhydre	0.7235
S2	22 : oxalate de calcium monohydraté	0.7171
S3	3 : protéine (albumine?)	0.699

### 3. Couplage de OPA(var) avec MCR-ALS : résultats et analyses

Les résultats de OPA(var), la matrice des concentrations et celle de spectres, sont les estimations initiales utilisées dans MCR-ALS avec les contraintes de non négativité.

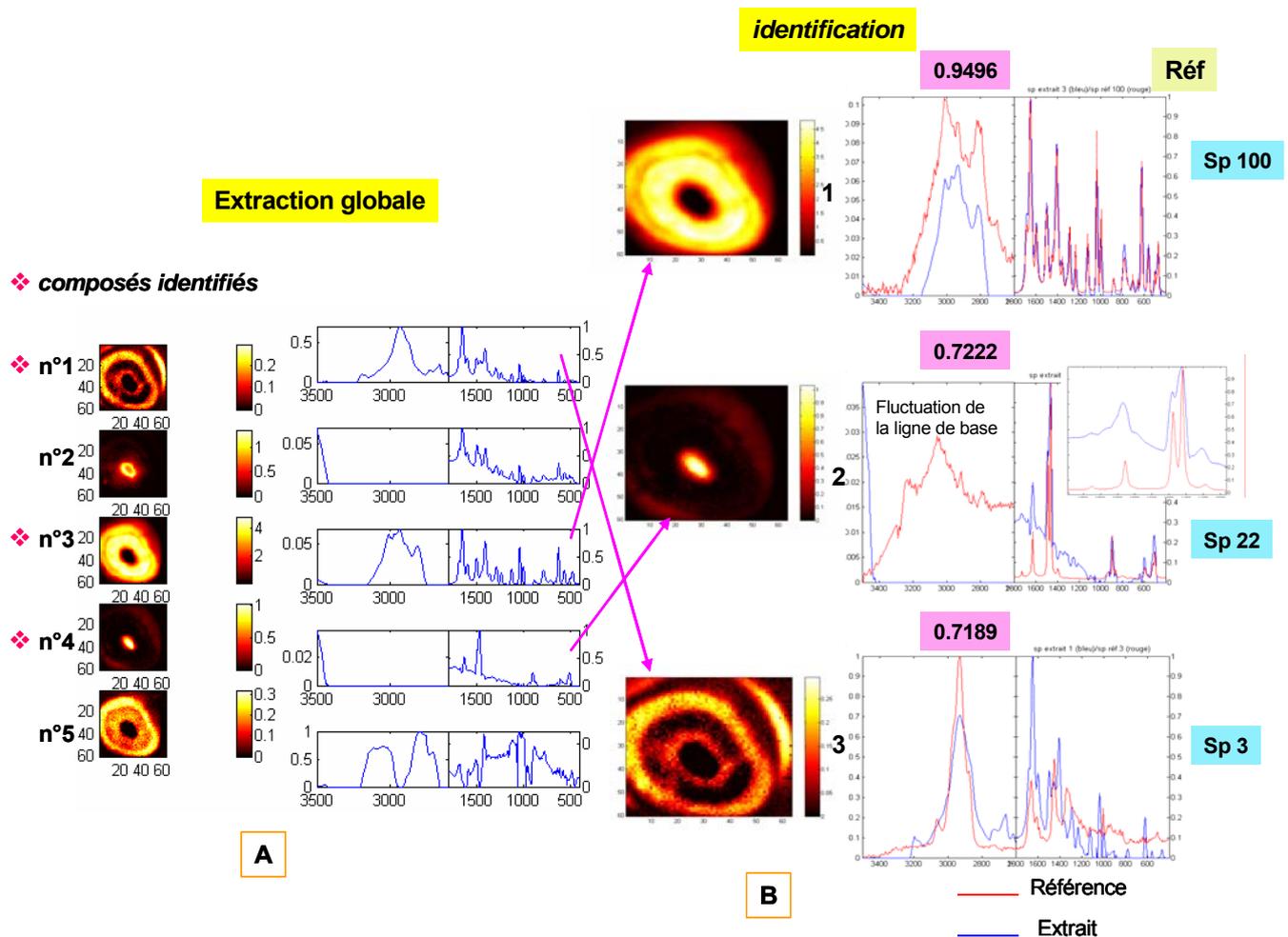


Figure 69 : Images et spectres extraits par le couplage OPA(var)/MCR-ALS. A) visualisation globale de l'ensemble des images et spectres extraits ; B) représentation des composés extraits identifiés et comparaison de leur spectre avec celui de la référence correspondante.

La Figure 69 présente les résultats de ce couplage. Dans la partie A, l'extraction globale présente les images et les spectres de 5 éléments extraits. Nous observons d'une façon claire que les 5 images extraites sont bien différentes les unes des autres. Les spectres le sont aussi sans toutefois que les spectres 2 et 5 de l'extraction globale soient bien identifiables. Ce couplage apporte donc une nette amélioration de la qualité d'extraction des images et des spectres par rapport à OPA(var) seul.

Les composés extraits 1, 3 et 4 sont bien identifiés (Cf. Figure 69A). La partie B de cette Figure 69 montre cette identification par comparaison des spectres extraits avec ceux de la référence correspondante. Nous observons le très bon accord du composé extrait 1 avec la référence 100, le 2 avec la 22 et le 3 avec la 3. Pour le composé 3, son image extraite est bien plus contrastée que celle obtenue pour OPA(var) seul. Quant à son spectre, son accord avec la référence 3 est

meilleur car le coefficient de corrélation obtenu est plus important. Cela nous a conforté dans l'analyse que nous avons effectuée pour départager les 4 composés de référence candidats pour ce 3<sup>ème</sup> composé extrait. De ce fait, ce 3<sup>ème</sup> composé extrait semble mieux correspondre à l'albumine. Le Tableau 18 suivant présente le bilan de cette extraction. L'acide urique est toujours très bien identifié avec un fort coefficient de corrélation. La protéine (l'albumine) est mieux extraite avec un coefficient de corrélation qui s'améliore par rapport au cas précédent. L'oxalate de calcium monohydraté est aussi bien identifié avec cependant un coefficient de corrélation un peu plus faible que le cas de OPA(var) seul.

**Tableau 18 : Résultats de l'extraction par le couplage OPA(var)/MCR-ALS et l'identification des éléments présents dans l'échantillon. S1, S3, S4 spectres des composés extraits dans l'ordre d'extraction.**

<b>Spectre Extrait (n° d'ordre)</b>	<b>Référence</b>	<b>Coefficient de corrélation</b>
<b>S3</b>	<b>100 : acide urique anhydre</b>	<b>0.9496</b>
<b>S4</b>	<b>22 : oxalate de calcium monohydraté</b>	<b>0.7222</b>
<b>S 1</b>	<b>3 : protéine (albumine?)</b>	<b>0.7189</b>

#### 4. Application des contraintes du rang local aux résultats du couplage de OPA(var) avec MCR-ALS : résultats et analyses

Ayant explicité l'effet du couplage entre OPA(var) et MCR-ALS, un résultat très intéressant démontré au chapitre 3 est l'application des contraintes du rang local. Dans cette partie nous allons appliquer les contraintes du rang local aux résultats du couplage OPA(var)/MCR-ALS. Toute la méthodologie de calcul et d'application des contraintes du rang local est expliquée au §V.3 du chapitre 3.

L'exploration globale et l'estimation du nombre de facteurs significatifs nous ont fourni le nombre le plus adéquat à retenir pour l'extraction (au maximum 5). Compte tenue de cela, la dimension de la fenêtre pour l'algorithme FSIW-EFA va être égale à 3, c'est-à-dire un carré de 3×3. Le nombre de valeurs singulières à calculer à chaque étape, c'est-à-dire à chaque position de la fenêtre, est alors égal à 9. Sur la Figure 70A, nous avons représenté la distribution relative à chaque valeur singulière.

Cette représentation nous permet d'avoir une idée du nombre des composés présents dans l'échantillon ainsi que leur répartition dans la zone d'analyse. Nous observons que les deux premières valeurs singulières correspondent à deux distributions nettes et distinctes, donc à deux composés différents « extractibles ». La troisième valeur singulière montre encore une distribution différente bien que moins nette. Cela indique la présence d'un troisième composé.

Quant à la quatrième, sa distribution n'est pas assez différente de la troisième ; ce qui n'aboutit pas à un avis concluant sur la présence d'un quatrième composé. Nous avons symbolisé cela par le point d'interrogation.

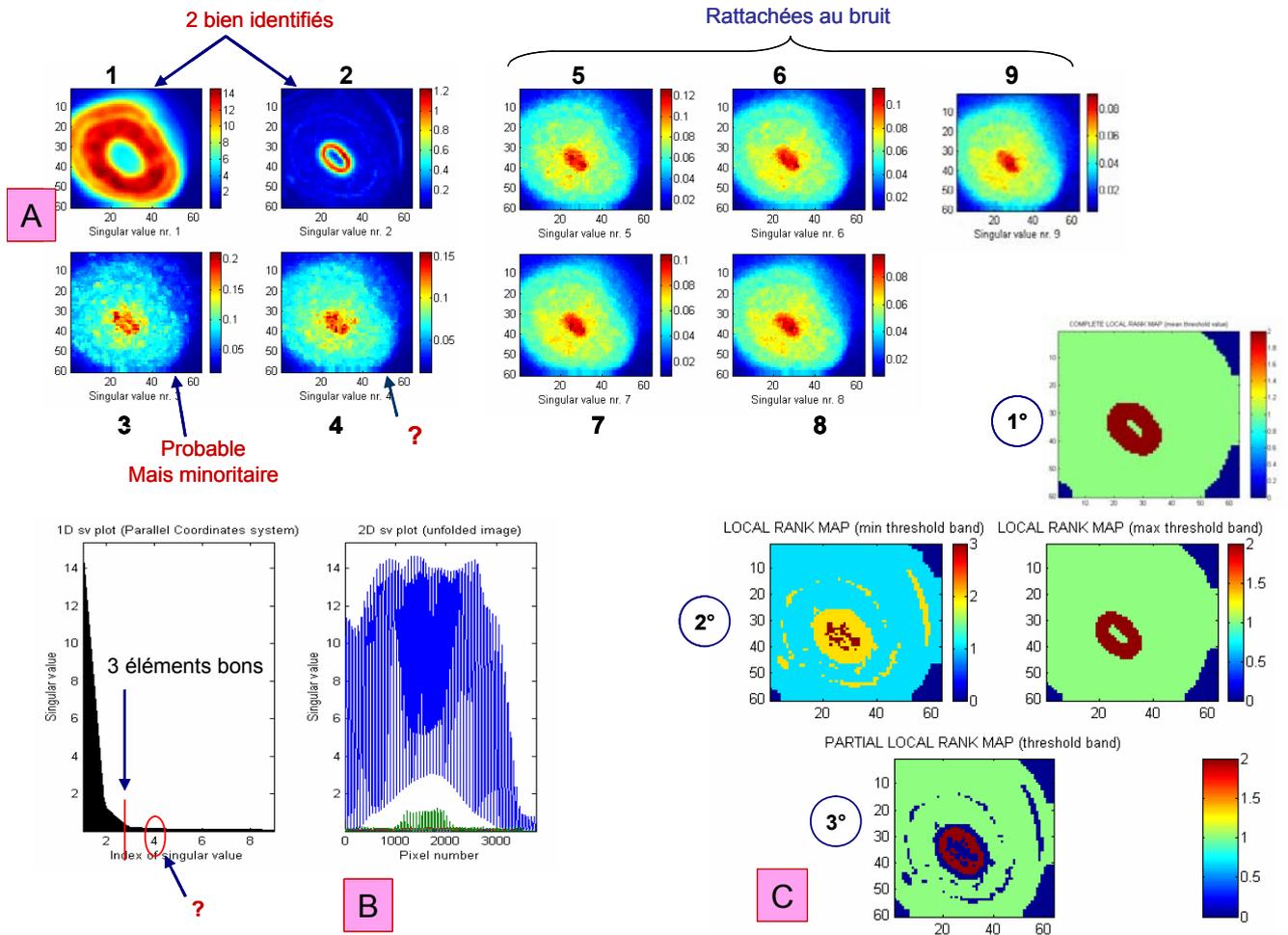


Figure 70 : Différentes étapes de calcul du rang local. A) distribution de chaque valeur propre extraite donnant une idée du nombre de composés différents dans l'échantillon étudié ; B) graphe de l'évolution des valeurs singulières en fonction de leur indice ou du numéro des pixels pour pouvoir sélectionner le seuil limite de distinction entre le bruit et les valeurs significatives ; C) représentation des cartes du rang local : 1° Carte rang local complet, 2° carte du rang local au minimum et au maximum de la « bande seuil », 3° carte du rang local partiel.

Nous voyons d'une manière évidente que la distribution de la 5<sup>ème</sup> à la 9<sup>ème</sup> est pratiquement la même. Ce qui veut certainement dire que ces valeurs singulières ne correspondent pas à un composé qu'on puisse extraire mais plutôt rattachées au bruit. Après réorganisation de ces valeurs singulières pour tenir compte de la structure 3D, les graphes de la partie B permettent de sélectionner le nombre des composés significatifs à extraire. Il apparaît bien clair que 3 composés sont, au moins, présents dans le mélange, que le pronostic est réservé pour le quatrième.

La distinction entre les valeurs singulières significatives et celles rattachées au bruit aide à la production des bonnes cartes du rang local.

La production des cartes du rang local est subordonnée à la sélection d'une « bande seuil » qui minimise l'erreur sur la sélection de la limite entre les valeurs singulières significatives et celles rattachées au bruit. La sélection de cette « bande seuil » et cette distinction des valeurs singulières sont explicitées sur la Figure 71. Les valeurs singulières rattachées au bruit s'entassent vers le bas car leur valeur est faible. Dans la partie b de la Figure 71, nous voyons que 6 valeurs singulières s'entassent vers le bas. Il n'y a que deux valeurs singulières qui se dégagent nettement du bruit, la troisième n'est pas loin du niveau de bruit. La limite inférieure de la bande doit être juste au dessus de ces valeurs singulières rattachées au bruit. Quant à la limite supérieure, la bande doit contenir au moins une valeur singulière qui se dégage du niveau de bruit. La position médiane entre les deux limites extrêmes de cette « bande seuil » donne la carte du rang local complet (cf. Figure 70C.1°). La production des cartes du rang local correspondant aux deux extrémités (cf. Figure 70C.2°) permet de produire la carte du rang local partiel (cf. Figure 70C.3°). Cette dernière montre les points pour lesquels le rang local reste invariant entre les deux extrémités de la « bande seuil ».

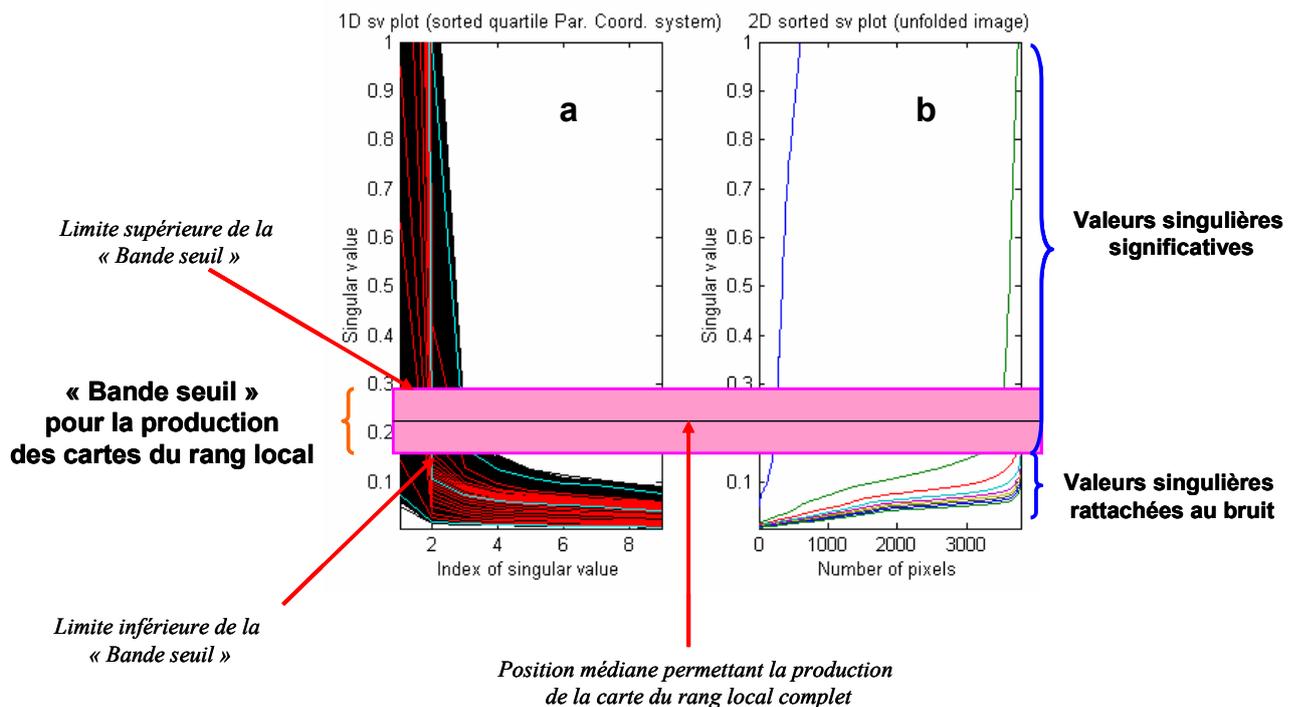


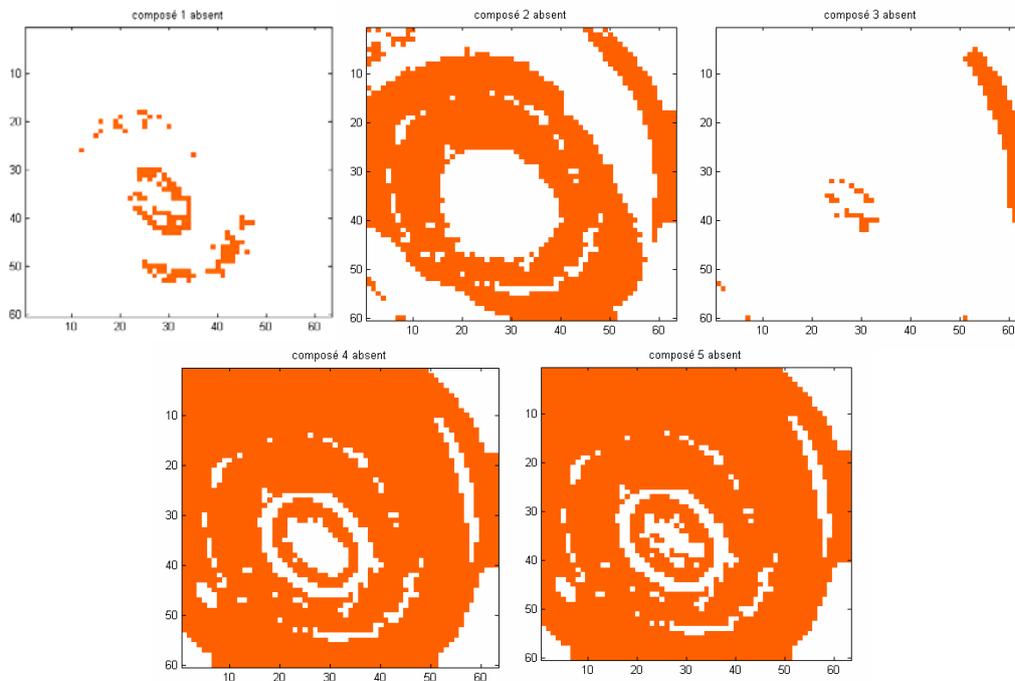
Figure 71 : Sélection de la « bande seuil » pour la distinction des valeurs singulières rattachées au bruit et celles significatives correspondant à des composés extractibles.

Cette propriété « d'invariance de rang local » est très importante car elle permet de relativiser l'incertitude de la sélection des valeurs singulières significatives. Cette carte de rang local partiel est celle qui va être utilisée pour produire les contraintes qui sont liées au rang local.

Elle va nous permettre, entre autres, de produire la carte des absences de chaque composé dans les pixels d'analyse. Ces connaissances qui traduisent les contraintes du rang local vont être

utilisées dans la procédure itérative ALS (algorithme ALS99) pour extraire les spectres et les concentrations des éléments présents dans l'échantillon analysé.

La Figure 72 visualise les exemples des cartes des absences que nous avons obtenues pour le couplage OPA(var)/MCR-ALS. Chaque point rouge indique l'absence du composé considéré sur le pixel où il est placé. Les résultats, l'ensemble des cartes des absences, sont sous forme matricielle, matrice *cse1* (3780×5) injectée dans l'algorithme.



**Figure 72 : Production des cartes des absences de 5 « composés » extraits de la matrice de données. Les points rouges indiquent les pixels sur lesquels le composé considéré est absent. A noter les zones d'absence similaires des « composés » 4 et 5.**

Le schéma de la Figure 48 montre les principaux éléments à injecter dans ALS pour cette résolution. Nous rappelons pour mémoire que les éléments à soumettre à ALS sont la matrice de données spectrales dépliée, l'estimation des spectres ou des concentrations, le nombre d'itérations nécessaires à la convergence et la matrice *cse1* traduisant les contraintes du rang local.

La matrice des concentrations optimisées nous permet de construire la carte de distributions des concentrations, c'est-à-dire les images. Et les spectres optimisés extraits nous servent à identifier les composés présents dans l'échantillon. Les résultats d'extraction et d'identification sont présentés à la Figure 73.

Nous observons d'emblée que la qualité d'extraction avec les contraintes du rang local est proche de celle du couplage OPA(var)/MCR-ALS aussi bien pour l'extraction globale que l'identification individuelle. Cela est bien mis en évidence par les coefficients de corrélation qui sont pratiquement égaux à ceux de OPA(var)/MCR-ALS au centième près.

Les 3 composés identifiés ont des images nettes bien distinctes et leurs spectres sont mis en évidence aussi bien pour la forme que la position des bandes. Cependant dans ce cas, la position des bandes semble mieux correspondre à celles des spectres de référence. On observe aussi plus de contraste sur les images.

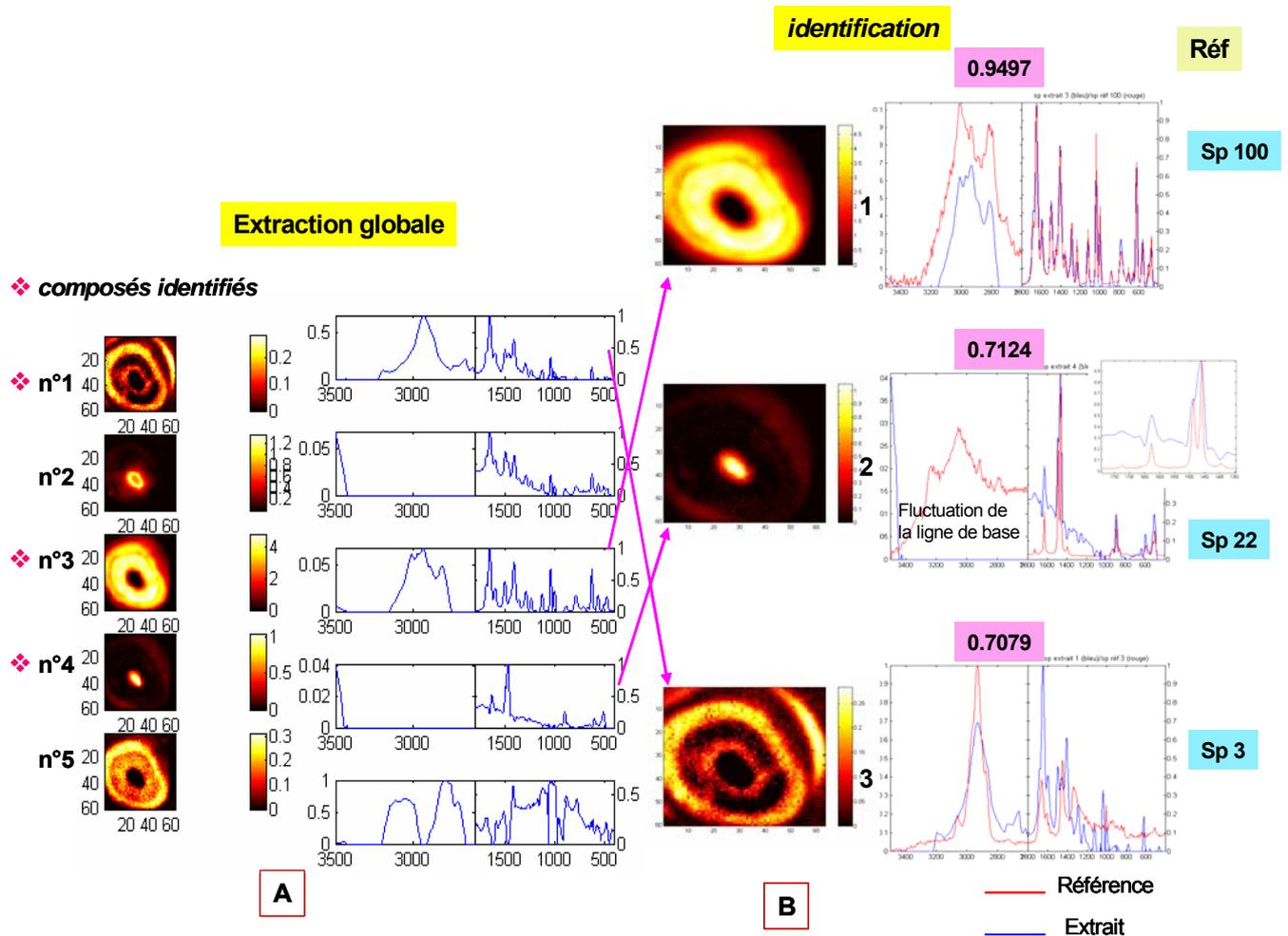


Figure 73 : Images et spectres extraits après application des contraintes du rang local aux résultats du couplage OPA(var)/MCR-ALS. A) visualisation globale de l'ensemble des images et spectres extraits ; B) représentation des composés extraits identifiés et comparaison de leurs spectres avec ceux de leurs références correspondantes.

Mais l'amélioration de la qualité de la forme des spectres extraits ne se traduit pas par une forte augmentation des coefficients de corrélation qui pourrait signifier un meilleur accord des extraits avec leurs références (Cf. Tableau 19). Un quatrième spectre bien net identifiant un quatrième composé du mélange n'est pas extrait.

Dans la partie suivante, nous emploierons le procédé « *d'augmentation de matrices* » pour essayer de mieux résoudre encore.

Tableau 19 : Résultats de l'extraction après application des contraintes du rang local au couplage OPA(var)/MCR-ALS et l'identification des éléments présents dans l'échantillon. S1, S3, S4 spectres des composés extraits dans l'ordre d'extraction.

Spectre Extrait (n° d'ordre)	Référence	Coefficient de corrélation
S3	100 : acide urique anhydre	0.9497
S4	22 : oxalate de calcium monohydraté	0.7124
S 1	3 : protéine (albumine?)	0.7079

5. Résolution par le procédé « augmentation de matrices »<sup>193 194</sup> en utilisant les résultats fournis par l'application des contraintes du rang local aux résultats du couplage de OPA(var) avec MCR-ALS

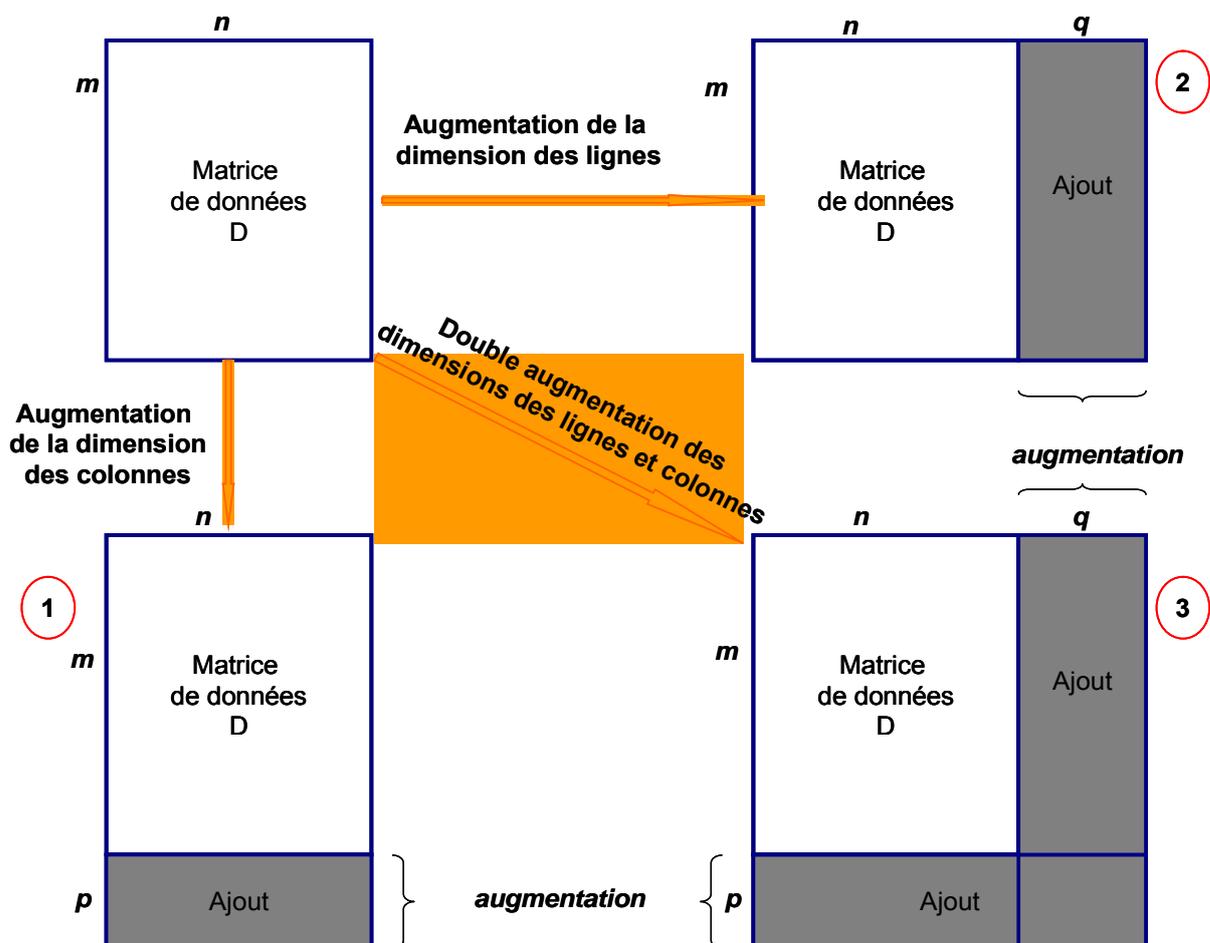


Figure 74 : Les différents procédés « d'augmentation de matrices ». 1) augmentation des dimensions des colonnes sans modification des dimensions des lignes ; 2) procédé inverse du 1) en n'augmentant que les dimensions des lignes ; 3) augmentation à la fois des dimensions des lignes et colonnes.

Ce procédé « augmentation de matrices »<sup>195</sup> est de plus en plus utilisé ces derniers temps. Il permet de résoudre des données complexes dont on connaît quelques caractéristiques (des

constituants, leurs concentrations, des caractéristiques locales d'une région spectrale, ...etc.). On adjoint à la matrice de données spectrales expérimentales  $D$  à étudier une autre matrice  $S$  contenant les informations spectrales sur les éléments connus dans le mélange à étudier. On obtient donc une matrice dont les dimensions sont plus grandes que celles de  $D$ . L'addition peut se faire suivant les lignes seulement, c'est une augmentation de la dimension des colonnes ; ou suivant les colonnes, c'est une augmentation de la dimension des lignes. Dans certains cas, l'addition peut se faire suivant les deux dimensions et la matrice voit ses deux dimensions augmenter. Les schémas de la Figure 74 résument ces différents procédés «d'augmentation de matrices ».

Dans notre cas, nous avons, d'une manière certaine, identifié deux constituants qui sont l'acide urique anhydre et l'oxalate de calcium monohydraté. Nous utiliserons le procédé «d'augmentation de la dimension des colonnes ». Nous ajouterons la matrice de 2 lignes ( $2 \times 576$ ) contenant le deux spectres purs, de deux composés identifiés, à la suite de la matrice de données  $D$ . Ces spectres ajoutés sont normalisés aussi par rapport à la valeur maximale d'intensité. Ces deux spectres purs sont tirés de la référence. Ce procédé ajoute des contraintes supplémentaires à la résolution puisqu'on précise certaines caractéristiques importantes, comme ici les spectres de deux composés. Ces contraintes sont exprimées dans une matrice, matrice  $ISP(2 \times 5)$ , qui indique la présence ou l'absence de chacun des éléments dans chacune de deux sous-matrices.

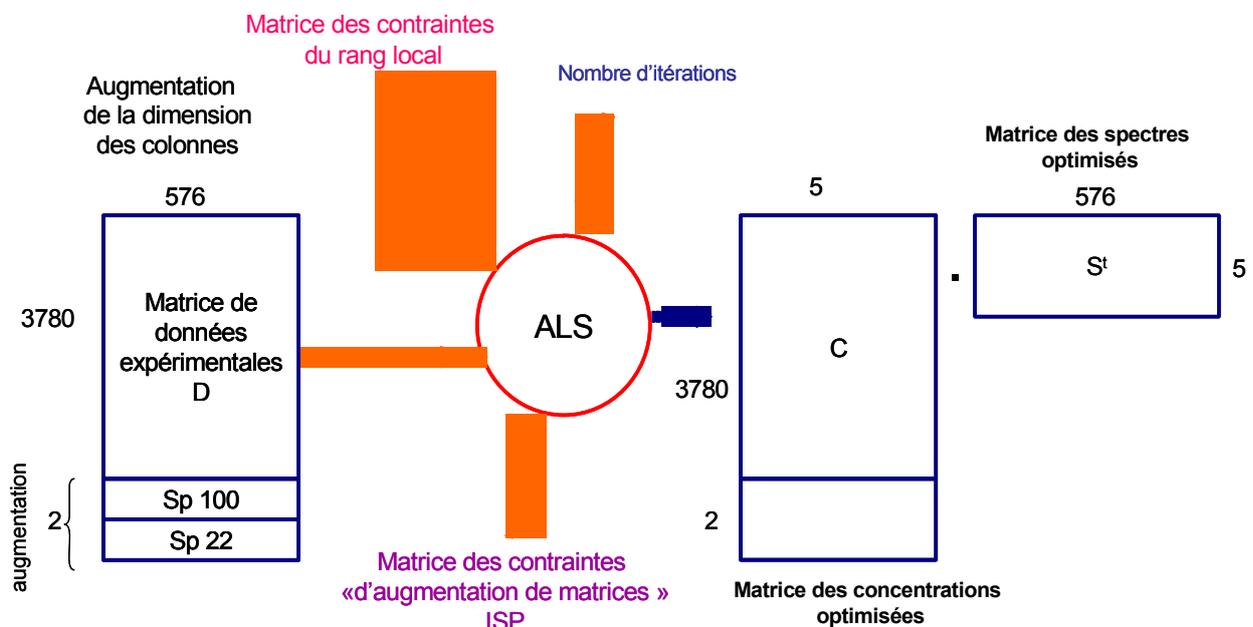


Figure 75 : Résolution de la matrice augmentée et les différentes contraintes à injecter dans ALS pour optimiser les spectres et les concentrations.

La Figure 75 résume les différents éléments à soumettre à ALS pour obtenir la meilleure optimisation à la fois des concentrations et des spectres. Les spectres sont normalisés. Dans ALS

(comme dans MCR-ALS), la contrainte de non négativité fait appel à l'algorithme FNNLS pour une meilleure souplesse de convergence.

Le but de notre résolution est bien sûr de mieux caractériser le 3<sup>ème</sup> élément extrait mais aussi d'avoir si possible une meilleure résolution sur un des deux éléments extraits non encore identifiés. Ce qui constituerait un 4<sup>ème</sup> élément à considérer dans l'échantillon sous analyse.

Les résultats de l'extraction sont représentés sur la Figure 76. L'extraction globale donne un résultat meilleur que celui obtenu pour l'application des contraintes du rang local. Le composé extrait n°2, bien que son image semble être distincte des autres, est associé à un spectre incomplet. Par conséquent il est non identifiable. Il en est de même pour le n°5 dont le spectre associé est déformé avec une image différente des autres.

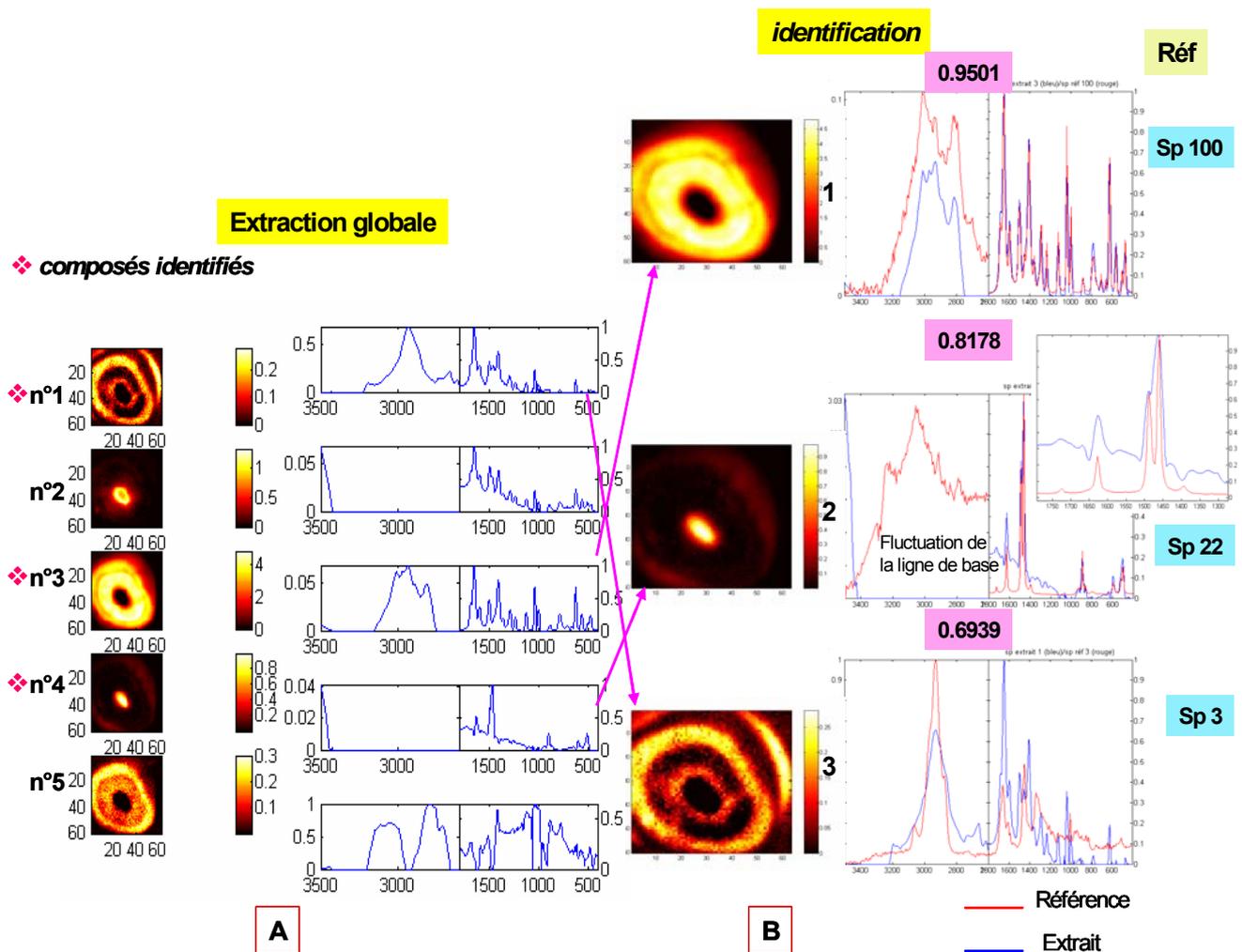


Figure 76 : Résultats du procédé « augmentation de matrices ». Images et spectres extraits de la matrice augmentée après application des contraintes du rang local et des contraintes liées à « l'augmentation de matrices » aux résultats du couplage OPA(var)/MCR-ALS. A) visualisation globale de l'ensemble des images et spectres extraits ; B) représentation des composés extraits identifiés et comparaison de leurs spectres avec leurs références correspondantes.

Quant à l'identification, les deux composés utilisés pour le procédé « augmentation de matrices », l'acide urique et l'oxalate de calcium monohydraté, sont très bien retrouvés avec des coefficients de corrélation élevés. L'accord spectral est parfait même si pour le 22, on n'observe

que la fluctuation de la ligne de base dans l'intervalle [3500-2600 cm<sup>-1</sup>]. Le troisième composé extrait identifié, le 3, présente un coefficient de corrélation presque identique à celui obtenu pour le cas précédent d'application des contraintes du rang local. Parmi les deux éléments restants, aucun n'est vraiment bien identifiable malgré l'effort de calcul et d'identification entrepris.

Le Tableau 20 présente le bilan de l'identification et d'attribution des éléments extraits par le procédé « augmentation de matrices ». L'extraction du quatrième composé espéré n'est pas atteinte. Mais dans ce dernier cas, la qualité des images et de spectres extraits est indiscutablement meilleure.

**Tableau 20 : Bilan d'identification après résolution par le procédé « augmentation de matrices » en utilisant les résultats de l'application des contraintes du rang local au couplage OPA(var)/MCR-ALS. S1, S3, S4 spectres des composés extraits dans l'ordre d'extraction.**

<b>Spectre Extrait (n° d'ordre)</b>	<b>Référence</b>	<b>Coefficient de corrélation</b>
<b>S3</b>	<b>100 : acide urique anhydre</b>	<b>0.9501</b>
<b>S4</b>	<b>22 : oxalate de calcium monohydraté</b>	<b>0.8178</b>
<b>S 1</b>	<b>3 : protéine (albumine?)</b>	<b>0.6939</b>

## V. Bilan du chapitre 4

Dans ce chapitre, nous avons étudié des données réelles acquises sur une petite surface d'un calcul urinaire par cartographie spectrale microspectrométrique.

Après estimation du nombre de facteurs pour la décomposition et prétraitement de données, nous avons réalisé l'extraction des spectres et des images par les différentes méthodes en allant du plus simple au plus compliqué. La méthode OPA(var) a extrait d'une manière nette 2 composés. Ces deux composés extraits et identifiés ont un coefficient de corrélation élevé avec les deux composés de référence correspondants. SIMPLISMA a confirmé ces résultats. L'interprétation du 3<sup>ème</sup> composé extrait était délicate. Nous avons expliqué que les composés candidats pour celui-ci sont tous des protéines dont les spectres sont assez semblables.

Le couplage de OPA(var) avec MCR-ALS a confirmé pleinement les résultats de OPA(var) avec cependant une qualité visuelle meilleure sur les spectres extraits et une différenciation plus nette sur les images. Le coefficient de corrélation évolue différemment pour les composés extraits. On a observé une relative stabilité pour le premier, une légère diminution pour le second et une augmentation pour le troisième.

Les contraintes du rang local appliquées au couplage OPA(var)/MCR-ALS améliorent légèrement les résultats de ce couplage.

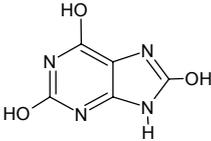
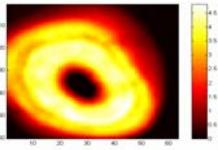
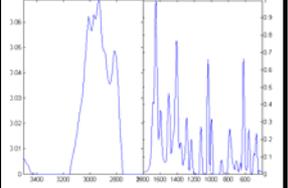
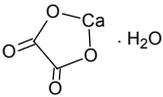
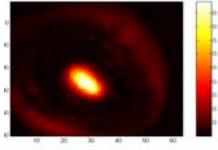
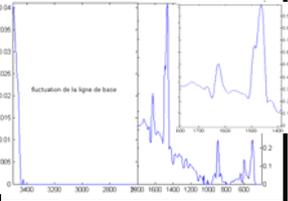
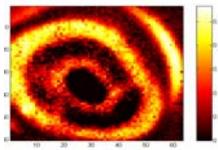
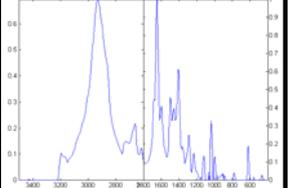
Pour renforcer notre argumentation, nous avons aussi utilisé la technique « *augmentation de matrices* » qui impose des contraintes supplémentaires. Pour cela, nous avons utilisé les spectres de deux composés clairement identifiés. L'extraction nous a fourni « ces deux composés tests » avec d'excellents coefficients de corrélation. La qualité d'extraction sur le 3<sup>ème</sup> composé est meilleure. Cependant, nous n'avons pas identifié un quatrième composé d'une manière catégorique.

En somme, nous pouvons dire que nos différents traitements nous ont permis d'extraire et d'identifier 3 composés constituant la partie analysée de l'échantillon. Le composé majoritaire toujours extrait sans ambiguïté est l'acide urique anhydre, uricite, identifié avec un très grand coefficient de corrélation de l'ordre de 0,95. Le deuxième composé avec un coefficient de corrélation qui varie entre 0,70 et 0,87 est l'oxalate de calcium monohydraté (whewellite). Le troisième, qui a nécessité une analyse plus poussée, est extrait avec un coefficient de corrélation de l'ordre de 0,70 et correspondrait à l'albumine.

La distribution des concentrations de chaque composé est indiquée par l'image associée extraite. Le centre de la zone d'analyse est occupé par l'oxalate de calcium, puis vient tout autour l'acide urique anhydre. La protéine est mélangée aux dernières couches de l'acide urique anhydre qui occupe une zone assez étendue.

L'identification correcte des constituants du calcul urinaire est indispensable pour bien diagnostiquer la pathologie. Identifier l'emplacement de chacun des constituants dans l'édifice calculaire est primordial aussi. En effet, l'élément central de l'édifice est à l'origine de la formation du calcul urinaire. Dans notre cas, l'élément central est le whewellite (oxalate de calcium monohydraté), ce qui traduit une hyperoxalurie. La forte concentration de l'acide urique témoigne d'une urine acide (par hyperuricémie ou par acidification des urines par apport diététique). Cela donne une indication sur le régime du patient. Dans l'analyse des calculs urinaires, on retrouve souvent des traces de protéines. Elles jouent le rôle essentiel du ciment de l'édifice calculaire. Il y a un danger lorsque la concentration de certaines d'entre elles (comme l'albumine par exemple) devient importante.

Tableau 21 : Récapitulatif des éléments extraits et identifiés avec nom, formule, image et spectre dans le cas du procédé « augmentation de matrices ».

Nom	Formule brute	Formule développée	Image (distribution de la concentration)	spectre
<b>Acide urique anhydre (uricite)</b>	$C_5H_4N_4O_3$			
<b>Oxalate de calcium monohydraté (whewellite)</b>	$CaC_2O_4 \cdot H_2O$			
<b>Protéine : (Albumine?)</b>	Macromolécule	Plusieurs formes et Plusieurs conformations		

Le Tableau 21 récapitule dans le cas du procédé « augmentation de matrices », les trois composés identifiés avec leurs images (la distribution des concentrations) et leurs spectres dans les deux intervalles d'étude  $[3500-2600cm^{-1}]$  et  $[1800-400cm^{-1}]$ .

## CONCLUSION GENERALE

L'un des buts de notre étude était de mettre en place, au niveau du laboratoire, une méthodologie d'étude des images microspectroscopiques par les moyens chimiométriques. L'originalité de notre étude porte, en effet, sur les méthodes de résolution appliquées à la résolution des données de l'imagerie microspectrométrique, en particulier l'imagerie microspectrométrique infrarouge et Raman. L'une des problématiques était d'améliorer la capacité de résolution de certaines méthodes de traitement chimiométrique pour mieux extraire les informations essentielles de ces données. La palette des méthodes chimiométriques disponibles étant très large, nous avons été amenés à faire un choix. Une démarche préliminaire nous a conduit à retenir pour notre étude les méthodes « du self-modeling » au nombre de sept pour leur usage courant et leur maniabilité. Mais nous avons étudié, dans notre investigation préliminaire, beaucoup d'autres non présentées.

Deux autres aspects importants de notre étude étaient le fait de tester un nombre assez important des méthodes chimiométriques sur une même matrice de données spectrales d'une part et d'autre part de réaliser d'une façon originale des couplages entre ces méthodes de résolution. L'étude systématique de cet ensemble de méthodes appliqué à une matrice de données spectrales synthétiques parfaitement maîtrisées nous a montré la capacité d'extraction qualitative de chacune d'elles. La comparaison approfondie de cet ensemble a mis en évidence la bonne qualité d'extraction de la méthode OPA(var) (Orthogonal Projection Approach par sélection des variables).

Le procédé de couplage des méthodes entre elles, consistant à associer d'une manière intelligente les méthodes de résolution « simple », comme SIMPLISMA, à celles robustes, comme MCR-ALS, est appliqué. Cette démarche s'est révélée assez efficace. En effet, les dissimilarités calculées, montrant le désaccord avec les références, ont fortement diminué. Cela traduit alors une amélioration assez significative de la qualité d'extraction par ce procédé de couplage. Le plus efficace des couplages est celui réalisé entre OPA(var) et MCR-ALS.

Etant donné que nous nous intéressons à l'imagerie microspectrométrique, pour améliorer davantage la qualité de résolution, nous avons introduit les contraintes liées au rang local. Cela nous a amené à expliciter la procédure sous-jacente et à calculer ces contraintes. Les contraintes du rang local donnent des informations locales au niveau de chaque pixel. Cela aide le processus de résolution à cibler le meilleur choix. Dans l'étude de la matrice de données synthétiques, les contraintes liées au rang local sont appliquées aux résultats du couplage de OPA(var) avec MCR-ALS. Le résultat obtenu est encore meilleur à la fois sur la qualité des spectres et des

concentrations extraites. Bien entendu, une meilleure extraction des concentrations aboutit à une meilleure production d'images spectroscopiques.

Puis nous avons appliqué ce schéma méthodologique, mis en place dans l'étude de données synthétiques, à l'analyse microspectrométrique d'un calcul urinaire. Ces données réelles microspectrométriques Raman sont acquises par le moyen d'un spectromètre FT-Raman du laboratoire. Après l'étape délicate de l'estimation des « facteurs significatifs », nous avons extrait, dans les différents cas étudiés, 3 composés identifiés par leurs spectres respectifs. Une bibliothèque de référence contenant les spectres Raman de 105 composés généralement présents dans les calculs urinaires nous a servi comme base d'identification des différents composés extraits par chacune des méthodes chimiométriques. Les spectres de 3 composés identifiés sont de bonne qualité. L'image spectroscopique, c'est-à-dire la distribution des concentrations dans la zone d'analyse, est reconstruite à partir des concentrations extraites par codage selon une échelle de couleurs (de 0 à 255). L'intensité de la coloration sur un pixel donné est proportionnelle à la concentration calculée pour ce pixel. Chaque image extraite indique la répartition du constituant considéré dans la portion de l'échantillon analysé.

Les 3 composés extraits sont l'acide urique, très majoritaire, l'oxalate de calcium monohydraté (whewellite), tous 2 identifiés sans ambiguïté, et une protéine dont l'identification moléculaire est délicate (l'albumine étant la forme la plus probable d'après notre investigation). Bien que nous en soupçonnions la présence, un quatrième composé du mélange n'a pu être identifié formellement, probablement du fait d'une trop faible concentration ; le cinquième extrait correspondant au bruit de fond. Quant à la répartition des composés, nous avons montré que l'oxalate de calcium monohydraté occupe le centre de la portion analysée correspondant à la partie centrale du calcul urinaire. Puis vient tout autour l'acide urique qui occupe une surface assez étendue, la protéine étant majoritairement mélangée avec les dernières couches de l'acide urique. Médicalement parlant, cette identification de l'élément central de l'édifice calculaire permet de remonter aux causes de la pathologie et d'apporter un traitement efficace.

Enfin dans une dernière étape, nous avons utilisé le procédé « d'augmentation de matrices ». Nous l'avons appliquée aux résultats du couplage OPA(var)/MCR-ALS avec les contraintes du rang local. Nous avons réalisé cela pour apporter plus de contraintes et affiner davantage la capacité de résolution. De ce fait, nous avons tiré profit des spectres de référence des composés précédemment identifiés pour réaliser ce procédé. Nous avons observé que la qualité d'extraction obtenue est relativement proche de celle obtenue avec les contraintes du rang local pour les spectres extraits. La qualité des images obtenues est cependant meilleure.

Notre étude a montré qu'un processus de résolutions successives (couplages des méthodes avec contraintes adaptées) judicieusement choisi permet d'atteindre une meilleure qualité d'extraction. Par exemple, dans un processus d'identification par l'image des impuretés, des inclusions, des défauts de fabrication, (...etc.), la qualité d'extraction de la méthodologie de résolution est l'un des facteurs le plus important.

En utilisant toutes ces connaissances mises en place, on peut aller par la suite dans le sens d'une résolution plus poussée pour s'affranchir davantage des contraintes instrumentales. L'optimisation ciblée de certains paramètres des algorithmes peut nous y aider.

## Annexe 1 : Algorithmiques des méthodes étudiées et quelques remarques sur leurs qualités et leurs défauts

### ❖ Annexe 1.1. PCA

#### 1 . Algorithmique de PCA

- **1. Prétraitement** : le centrage et la réduction des colonnes de la matrice de données en utilisant la moyenne et la déviation standard pour réduire la disparité des variables à étudier.

- **2. Décomposition mathématique** : sous la forme  $D = C * S^t$  (cf. éq.2).

1°). Évaluation des dimensions de la matrice  $D(r \times c)$  : si  $r < c$  alors la matrice de covariances  $\text{cov} = X = D^t * D$  sinon  $\text{cov} = X^t = D * D^t$ .

2°). Décomposition en valeurs singulières de la matrice des covariances :  $[U, S, V] = SVD(\text{cov})$  (Cf. §II.2.2.2).

3°). La matrice des concentrations est égale alors  $C = U * S$  et celle des spectres purs  $S = V^t$  (éq23) ou bien  $C = \bar{U} * \bar{S}$  et  $S = \bar{V}^t$  (cf. éq.7) avec  $\bar{U}$ ,  $\bar{V}$  les matrices relatives aux valeurs singulières les plus significatives retenues dans la matrice diagonale  $\bar{S}$ .

- **3. Calculs statistiques** : des variances, des erreurs standards, du  $T^2$  limite, des résidus à 5% près de telle sorte que la différence  $E = \sum_i \sum_j e_{ij}^2 = \sum_i \sum_j (d_{ij} - \sum_k c_{ik} s_{kj})^2$  (cf. éq.8bis et éq.27) soit minimale.

#### 2° . Quelques avantages et inconvénients de PCA.

##### Méthode PCA)

##### Avantages :

- ❖ L'une des meilleures méthodes pour l'étude exploratoire des données,
- ❖ Permet de réduire significativement la taille des données à traiter.
- ❖ L'une des meilleures méthodes pour l'estimation du nombre de facteurs significatifs (composantes spectrales d'un mélange).

##### Inconvénients :

- ❖ Absence des contraintes qui entraînent souvent une mauvaise qualité d'extraction spectrale,
- ❖ Impossibilité d'utiliser les profils extraits par PCA pour une quantification (calcul des concentrations) des espèces présentes dans un mélange

❖ **Annexe 1.2. SIMPLISMA****1. L'algorithmique de SIMPLISMA.**

La résolution se fait en 2 étapes : l'extraction des variables pures caractéristiques des constituants du mélange et puis la résolution d'une équation matricielle appropriée.

- **1<sup>ère</sup> étape** : L'extraction de variables pures de la matrice des données D, ces variables pures étant considérées caractéristiques d'un constituant du mélange étudié.

- Tout d'abord, évaluation de la pureté  $p_j$  de chaque colonne  $j$  de  $D(r \times c)$  par la

relation  $p_j^{(1)} = \frac{\sigma_j}{(\mu_j + \alpha)}$  (cf. éq.22) avec  $\sigma_j = \sqrt{\frac{\sum_{i=1}^r (d_{i,j} - \mu_j)^2}{r}}$  l'écart-type et

$$\mu_j = \frac{\sum_{i=1}^r d_{i,j}}{r}$$

la moyenne de la colonne  $j$ ,  $\alpha$  étant l'offset ajouté pour éviter que des

variables avec une valeur moyenne faible obtiennent une haute valeur de pureté.

- Sélection de la variable  $j$  ayant la plus grande valeur de pureté  $p_j$  ;
- Extraction de la contribution de cette variable à D par les biais des corrélations ;
- Calcul à nouveau de la pureté de chacune des variables restantes par la relation

$$p_j^{(k)} = w_j^{(k)} \frac{\sigma_j}{(\mu_j + \alpha)}$$

(cf. éq.23) avec  $w_j^{(k)}$  le poids affecté à la variable  $j$  ; et cela

permet de réduire l'influence des variables corrélées avec les variables précédemment extraites. Ce poids  $w_j$  est défini comme le déterminant de la matrice de dispersion  $Y_i$  qui contient le spectre sélectionné normalisé à 1 associé à chacun des spectres  $i$  de D pris individuellement ;  $w_i = \det(Y_i^t * Y_i)$ .  $k$  en exposant signifie le  $k^{\text{ième}}$  calcul de la pureté.

- Répétition des étapes 2, 3 et 4 jusqu' à sélectionner toutes les variables pures souhaitées (ou atteindre le nombre préfixé des variables). Dans le cas d'estimation du nombre de facteurs significatifs, le spectre de pureté ne doit représenter que le bruit de fond.
- Toutes les variables sélectionnées sont regroupées dans la matrice C et constituent la matrice des concentrations estimées initialement. Pour chaque variable pure sélectionnée, on récupère les concentrations correspondantes à partir de la matrice de données D.

- **2<sup>ème</sup> étape** : Résolution de l'équation  $D = C * S^t$  pour la recherche des Spectres purs. La relation  $S^t = (C^t * C)^{-1} * C^t * D$  (cf. éq.35) permet d'obtenir la matrice des spectres purs.

## 2°. Quelques avantages et inconvénients de SIMPLISMA.

### Méthode SIMPLISMA

#### Avantages :

- ❖ Une interactivité simple pour la sélection et le contrôle,
- ❖ Ne nécessite pas des connaissances a priori sur les données traitées,
- ❖ N'utilise pas la décomposition fastidieuse du « factor analysis » mais des éléments simples comme l'écart-type, la moyenne...etc.

#### Inconvénients :

- ❖ Une absence des contraintes d'où le risque d'obtenir des résultats aberrants,
- ❖ Des problèmes liés à la ligne de base,
- ❖ La qualité d'extraction dépend de l'offset choisi,
- ❖ Une résolution limitée dans le cas d'un recouvrement inextricable des composantes spectrales.

## ❖ Annexe 1.3. OPA

### 1. L'algorithmique de OPA.

L'algorithme OPA propose deux façons de faire une résolution : soit en recherchant des spectres purs correspondant aux différents composés supposés être présents dans le mélange, soit en recherchant des variables pures caractéristiques d'un composé du mélange.

1°. **1<sup>ère</sup> méthodologie : Recherche des spectres purs caractéristiques des constituants du mélange.**

Dans ce cas, on travaille suivant les lignes de la matrice D, puisque les différents spectres mesurés de l'échantillon sont placés suivant ces lignes.

- 1. Calcul de la dissimilarité  $d_i$  de chaque spectre  $S_i$  de D par le déterminant  $d_i$  exprimé par la relation suivante :  $d_i = \det(Y_i \cdot Y_i^t) = [\|S_{\text{mean}}\| \cdot \|S_i\| \cdot \sin(\alpha_i)]^2$  (cf. éq.25) avec  $\|S_{\text{mean}}\|$  la norme du vecteur ligne moyen et  $\|S_i\|$  la norme du vecteur ligne  $i$ ,  $\alpha$  étant l'angle entre les vecteurs  $S_{\text{mean}}$  et  $S_i$  dans l'espace géométrique considéré.  $Y_i$  est

appelée la matrice de dispersion et ne comporte au départ que le spectre moyen associé à chacun des spectres individuels  $S_i$ ,  $Y_i = \begin{bmatrix} S_{mean} \\ S_i \end{bmatrix}$ .

- 2. Sélection du spectre ayant le plus grand  $d_i$  qui va être le premier spectre pur sélectionné et normalisé à 1. Appelons-le  $S_{1g+}$ . Ce spectre est à la fois celui qui est le plus dissimilaire (le moins corrélé) du spectre moyen  $S_{mean}$  et ayant l'absorbance moyenne la plus grande.
- 3. Le spectre  $S_{1g+}$  sélectionné remplace, dans  $Y_i$ , le spectre  $S_{mean}$  et va être la référence pour le calcul de la 2<sup>ème</sup> dissimilarité la plus importante, d'où  $Y_i = \begin{bmatrix} S_{1g+} \\ S_i \end{bmatrix}$ .
- 4. Répétition des étapes 1 et 2 sachant que  $S_{1g+}$  est à la place de  $S_{mean}$ .
- 5. Le spectre qui vient d'être sélectionné pour la 2<sup>ème</sup> dissimilarité la plus importante,  $S_{2g+}$  normalisé aussi à 1, est ajouté à la matrice de dispersion  $Y_i$  qui devient  $Y_i = \begin{bmatrix} S_{1g+} \\ S_{2g+} \\ S_i \end{bmatrix}$ .
- 6. Répétition des étapes 1 à 5 en ajoutant à chaque fois le nouveau spectre sélectionné, normalisé à 1, à la matrice  $Y_i$  dont la dimension augmente chaque fois d'une unité.
- 7. Le processus s'arrête lorsque le nombre de spectres cherchés est atteint ou lorsqu'un critère de convergence préfixé est atteint. Ou bien, dans le cas d'estimation de facteurs significatifs pour la décomposition d'une matrice de données  $D$ , lorsque le « spectre de dissimilarité » ou le « chromatogramme de dissimilarité » ne représente que le bruit de fond.
- 8. L'ensemble des spectres sélectionnés constitue l'estimation initiale de la matrice des spectres  $S$ . Ces spectres estimés sont utilisés par la suite dans ALS pour obtenir les concentrations correspondant aux différents constituants.

2°. 2<sup>ème</sup> méthodologie : ***Recherche des variables pures caractéristiques des différents constituants du mélange.***

- 1. La même démarche que pour la recherche des spectres purs est effectuée. Les étapes 1 à 7 sont identiques, la seule différente est que cette fois les vecteurs considérés sont les colonnes de la matrice de données  $D$ . Et la matrice de dispersion  $Y_i$  voit ses colonnes augmenter au fur et à mesure,  $Y_i = [S_{1g+} \quad S_{2g+} \quad \dots \quad S_i]$ .

- 2. Les concentrations associées à ces variables (colonnes) constituent l'estimation initiale de la matrice des concentrations. Comme pour les spectres purs sélectionnés, cette matrice est utilisée dans ALS pour obtenir les spectres purs des composés associés à ces « variables concentrations » sélectionnées.

## 2. Quelques avantages et inconvénients de OPA.

### Méthode OPA.

#### Avantages :

- ❖ Une interactivité simple avec deux procédures de recherches différentes,
- ❖ Ne nécessite pas des connaissances a priori,
- ❖ L'introduction des contraintes permet une meilleure qualité d'extraction,
- ❖ L'optimisation par ALS permet de retrouver les bons profils de concentrations et de spectres.

#### Inconvénients :

- ❖ Influence de la dérive de la ligne de base sur la sélection du premier spectre le plus dissimilaire du spectre moyen.
- ❖ Pour des données de grande taille la convergence peut être lente.

## ❖ Annexe 1.4. MCR-ALS

### 1. L'algorithmique de MCRA-ALS.

1°. Estimer par des méthodes simples (EFA, SIMPLISMA...etc.)  $C$  ou  $S$ . Cette estimation est nécessaire pour l'optimisation de la résolution. Ce qui nous donne  $C$  estimé ( $C_{ini}$ ) ou  $S$  estimé ( $S_{ini}$ ).

2°. Calculer par moindres carrés la matrice restante, non estimée, à l'aide de  $D$  et la matrice estimée. Si par exemple on a estimé  $S$ , soit  $S_{ini}$ , alors  $C$  ( $C$  calculé  $C_{cal}$ ) va être calculée de la façon suivante :

$$C_{cal} = (D * S_{ini}) * (S_{ini}^t * S_{ini})^{-1} \text{ (cf. éq34).}$$

3°. Injecter les contraintes nécessaires (non négativité, unimodalité, closure, sélectivité...etc.) à  $C_{cal}$  qui subit de ce fait les corrections souhaitées. Appelons-la  $C_{cor}$ .

4°. Avec  $C_{cor}$  obtenue et  $D$ , calculer la matrice des concentrations  $S_{cal}$  associée à  $C_{cor}$  par la relation :  $S_{cal}^t = (C_{cor}^t * C_{cor})^{-1} * (C_{cor}^t * D)$  (cf. éq.35).

5°. Injecter les contraintes nécessaires (non négativité, sélectivité... etc.) à  $S_{cal}$  qui subit de ce fait les corrections souhaitées et  $S_{cal}$  est normalisé à 1. Appelons-la  $S_{cor}$ .

6°. Evaluer la différence  $E = D - C_{cor} * S_{cor}^t$  et  $SCR = \sum_i \sum_j e_{ij}^2$  (éq24) la somme des carrés des résidus. Et vérifier si le critère de convergence est atteint.

7°. Utilisation de **ALS**, calculer **C** et **S** alternativement : étapes 2 à 6.

8°. Arrêt du processus itératif lorsque SCR atteint une valeur fixée (matrice des erreurs E est minimale) ou le nombre d'itération préfixé est atteint et production de  $C_{opt}$  (matrice optimisée des concentrations) et de  $S_{opt}^t$  (matrice optimisée des spectres).

La Figure 15 résume synthétiquement l'algorithmique de MCR-ALS.

## 2. Quelques avantages et inconvénients de MCR-ALS.

### Méthode MCR-ALS

#### Avantages :

- ❖ Aucune connaissance a priori sur le modèle chimique ni mathématique n'est nécessaire,
- ❖ Une méthode adaptée à traiter tout type de données,
- ❖ Grâce aux nombreuses contraintes incorporées au cours de la résolution dans la procédure itérative ALS, on obtient des solutions réalistes (résultats corrects).

#### Inconvénients :

- ❖ Besoin d'une estimation initiale des spectres ou des concentrations,
- ❖ La rapidité de la convergence peut dépendre de la taille des données ou risque de convergence vers un minimum incorrect (dans certains cas).

## ❖ Annexe 1.5. PMF (Positive Matrix Factorization)

### 1. L'algorithmique de PMF.

1°. Choix du nombre de facteurs et la contrainte (non négativité) sur C et S.

2°. Initialisation des matrices C et S (par des valeurs aléatoires).

3°. Utilisation de **ALS** (calcul alterné de C et S) avec évaluation, à chaque étape, de la

fonction objet Q relative aux résidus :  $F = Q(\mathbf{E}) = \sum_{i,j} \frac{(\mathbf{D} - \mathbf{C} \cdot \mathbf{S}^t)_{i,j}^2}{\sigma_{ij}^2} = \sum_i \sum_j \left( \frac{e_{ij}}{\sigma_{ij}} \right)^2$  (cf.

éq.37),  $\sigma_{ij}$  étant la déviation standard représentant l'incertitude dans la mesure  $D_{ij}$  et E la matrice des résidus.

4°. Arrêt de la résolution lorsque cette fonction **Q** est minimale, c'est-à-dire lorsque l'une des conditions suivantes est remplie :

- ✓ **F** atteint une valeur limite fixée ;
- ✓ la variation entre 2 itérations consécutives (change) est faible.

## 2. Quelques avantages et inconvénients de PMF.

### Méthode PMF

#### *Avantages :*

- ❖ Une procédure de calcul qui permet la non-orthogonalité des composantes extraites,
- ❖ La contrainte de non-négativité imposée sur les deux profils permet d'obtenir des résultats corrects,
- ❖ L'optimisation utilisant ALS donne des bons résultats,
- ❖ C'est une procédure qui peut être étendue à des dimensions supérieures à 2 (3-D ...etc.)

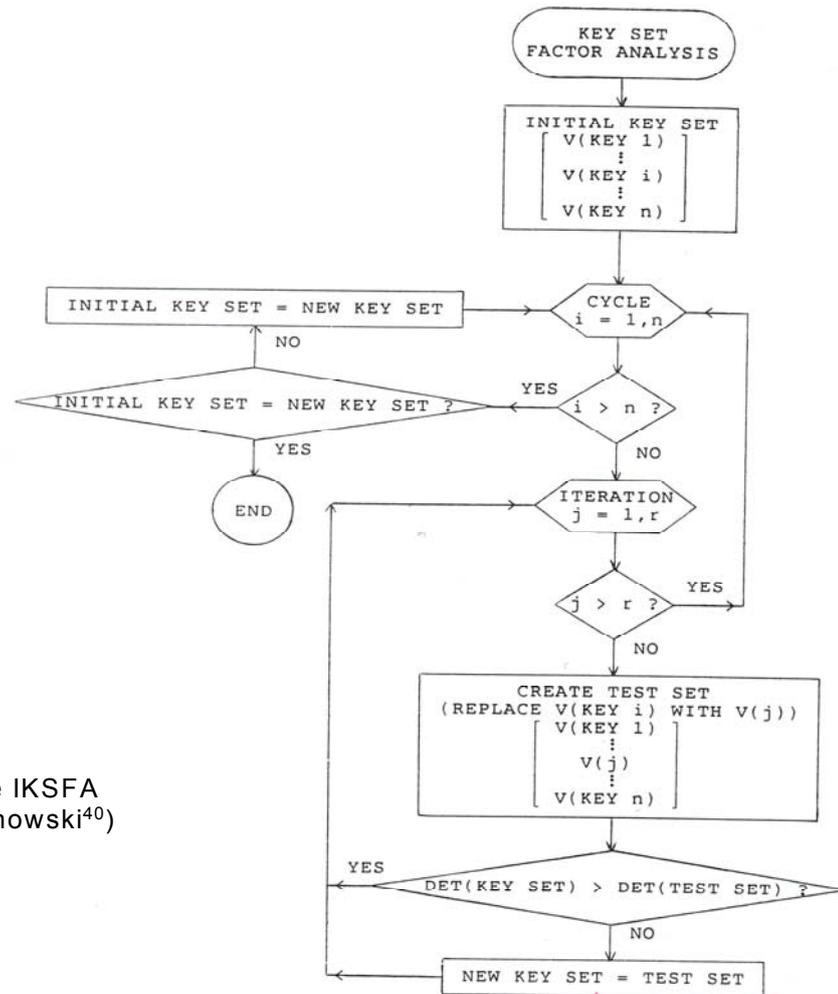
#### *Inconvénients :*

- ❖ Une initialisation aléatoire qui a un effet sur la convergence,
- ❖ Pour des données de grande taille, une lourdeur de calcul et une convergence lente.

## ❖ Annexe 1.6. IKSFA (Iterative Key Set Factor Analysis)

### 1. L'algorithmique de IKSFA.

IKSFA est une procédure itérative d'optimisation d'un ensemble de  $n$  facteurs caractéristiques sélectionnés par KSFA pour une matrice  $D(r \times c)$ . Cet ensemble constitue le « *key set* » initial sélectionné par un critère d'orthogonalité parmi les vecteurs propres de la décomposition de D. Le diagramme<sup>40</sup> suivant explique cette procédure d'itération à la fois sur les  $n$  vecteurs du « *key set* » et les  $r$  lignes de D.



Procédure itérative IKSFA  
(d'après E.R. Malinowski<sup>40</sup>)

## 2. Quelques avantages et inconvénients de IKSFA.

### Méthode IKSFA

#### Avantages :

- ❖ Méthode basée sur l'analyse factorielle, étude liée à la caractéristique des données.
- ❖ La procédure itérative utilisée permet d'obtenir des profils réalistes dans des nombreux cas.

#### Inconvénients :

- ❖ Initialisation basée sur KSFA qui ne sélectionne pas forcément les composantes spectrales les plus orthogonales,
- ❖ Malgré l'optimisation par itération, il n'est pas certain que toutes les lignes ou colonnes, de la matrice de données, sélectionnées correspondent à des composantes spectrales présentes dans le mélange (c'est-à-dire la vraie solution).
- ❖ Absence des contraintes sur les profils qui risquent de donner des résultats aberrants.

❖ **Annexe 1.7. SPEXFA (Spectral Isolation Factor Analysis)**

Il s'agit d'isoler les principales composantes spectrales caractéristiques des constituants d'un mélange lorsqu'il existe une gamme spectrale spécifique pour chaque constituant. Malinowski nous précise que les incertitudes sur les spectres isolés et leurs concentrations sont calculées par la *méthode de Clifford*<sup>196, 197</sup>.

Les procédures et les démarches sont similaires à celle de IKSFA. La seule grande différence étant que les spectres expérimentaux sont mis dans les colonnes de D pour faciliter la décomposition et la recherche en valeurs propres et en vecteurs propres orthogonaux. Les mêmes avantages et inconvénients que IKSFA sont constatés.

## Annexe 2 : Quelques fonctions Matlab utilisées

1. Fonction pour calculer la dissimilarité entre un spectre (ou image) extrait et sa référence.

```
function dis=disimil(x,y)
% x spectre extrait (par exemple)
% y spectre de référence (par exemple)
    temp=corrcoef(x,y);
    dis=sqrt(1-temp(1,2)^2);
end;
```

2. Fonction pour calculer le coefficient de corrélation des spectres extraits par rapport à l'ensemble des spectres de référence.

```
function [coeffcorr]=calcorreff(sp,ref,val);

% sp matrice des spectres des éléments extraits,[
% ref matrice des spectres de référence (105 spectres purs Raman)
% val valeur "seuil" minimale choisie à laquelle le coeffcorr calculé doit être
% supérieur.

[m,n]=size(sp);
[p,q]=size(ref);
if (n ~= q)
    disp('les spectres à comparer n"ont pas les memes dimensions');
    return;
else
    for i=1:m
        for j=1:p
            coeffcorr(i,j)=corr(sp(i,:),ref(j,:));
        end;
    end;
    for i=1:m
        tab=find(coeffcorr(i,*)>val);
        if tab==[]
            disp(['pour le spectre extrait 'num2str(i) ' aucun spectre de référence ne
vérifie ce seuil 'num2str(val) ' choisi pour le coeff corr']);
        else
            tab2=coeffcorr(i,tab);
            [x,y]=size(tab);
            table=zeros(2,y);
            table(1,:)=tab;
            table(2,:)=tab2;
            disp(['spectre 'num2str(i) ' corrélé avec les spectres de référence suivants
respectant la valeur "seuil" 'num2str(val) ' choisie:']);
            disp(table);
        end
    end
end;
```

3. Fonction pour représenter les 2 spectres (extrait et référence) dans les deux intervalles d'étude [3500-2600  $\text{cm}^{-1}$ ] et [1800-400  $\text{cm}^{-1}$ ] pour la comparaison.

**function plotspcomp3(lam1,lam2,sp,ref,ind,ind2);**

*% plotspcomp(lam1,lam2,sp(ind,:),ref(ind2,:)) fonction pour comparer un spectre extrait et celui de la référence correspondante pour mettre cote à cote les deux morceaux du spectre en faisant abstraction de la coupure intermédiaire; les spectres étant des vecteurs-lignes.*

*% lam1 vecteur longueur d'onde associé à la deuxième moitié de la figure;*

*% lam2 vecteur longueur d'onde associé à la première moitié de la figure pour un ordre décroissant de longueur d'onde.*

*% sp spectre extrait à comparer avec celui de la référence ref.*

*% ind numéro d'ordre du spectre extrait, ind2 numéro d'ordre du spectre de référence.*

[m,n]=size(lam2);

[p,q]=size(lam1);

if p~=1

    lam1=lam1';

    for i=1:p

        c2(1,i)=i;

    end

    k=c2(1,p);

else

    for i=1:q

        c2(1,i)=i;

    end

    k=c2(1,q);

end

if m~=1

    lam2=lam2';

    for i=1:m

        c1(1,i)=k+i;

    end

else

    for i=1:n

        c1(1,i)=k+i;

    end

end

figure;subplot('position',[0.05 0.05 0.45 0.9]);plot(lam1,sp(ind,c2),'blue'); hold on;

plot(lam1,ref(ind2,c2),'red');axis tight; subplot('position',[0.5 0.05 0.45

0.9]);plot(lam2,sp(ind,c1),'blue'); hold on; plot(lam2,ref(ind2,c1),'red'); axis tight;

title(['sp extrait 'num2str(ind) ' (bleu)', '/sp réf 'num2str(ind2) ' (rouge)']);

4. Fonction pour la représentation globale des images et des spectres extraits dans la même figure.

**function afimagsp3(x,y,ncomp,C,A,lam1,lam2);**

*% afimagsp(x,y,ncomp,C,A,lam1,lam2) fonction qui permet d'afficher en même temps les % images et les spectres des composés extraits*

```

% dans les 2 intervalles.
% x et y dimensions de la zone d'analyse; ncomp nombre de composantes à extraire;
% C matrice des concentrations et A matrice des spectres extraits.
% lam1 et lam2: les vecteurs contenant les variables longueurs d'onde sous forme
% (1xn) pour les deux intervalles; dans l'ordre décroissant
% des longueurs d'onde.
es=reshape(C,x,y,ncomp);
i=1;
j=1;
while (i<=((2*ncomp)-1))
    subplot(ncomp,2,i); imagesc(es(:,j)); axis image; colormap('hot');colorbar;
    i=i+2;
    j=j+1;
end;
[m,n]=size(lam2);
[p,q]=size(lam1);
if p~=1
    lam1=lam1';
    for i=1:p
        c2(1,i)=i;
    end
    k=c2(1,p);
else
    for i=1:q
        c2(1,i)=i;
    end
    k=c2(1,q);
end
if m~=1
    lam2=lam2';
    for i=1:m
        c1(1,i)=k+i;
    end
else
    for i=1:n
        c1(1,i)=k+i;
    end
end
i=2;
j=1;
while (i<=(2*ncomp))
    subplot(ncomp,2,i); subplot('position',[0.55 (1-j*(1/ncomp)+0.05) 0.2
0.11]);plot(lam1,A(j,c2));axis tight;subplot('position',[0.75 (1-j*(1/ncomp)+0.05)
0.2 0.11]);plot(lam2,A(j,c1)); axis tight;
    i=i+2;
    j=j+1;
end;

```

## Annexe 3 : Tableaux des dissimilarités associés aux graphes présentés au chapitre 3

Les nombres dans ces tableaux sont arrondis au quatrième chiffre après la virgule.

- Tableaux A correspondent à la Figure 38 : méthodes « simples ».

images

Bruit	SIMPLISMA	OPA(var)	OPA(spec)	PCA	IKSFA	SPEXFA
1%	0,2537	0,3940	0,9170	0,8364	0,2572	0,2560
3%	0,4429	0,3968	0,9129	0,8367	0,2675	0,2569
10%	0,3904	0,4018	0,8560	0,8383	0,4091	0,3045

spectres

Bruit	SIMPLISMA	OPA(var)	OPA(spec)	PCA	IKSFA	SPEXFA
1%	0,2357	0,5986	0,6795	0,7308	0,5336	0,2300
3%	0,3600	0,5919	0,6734	0,7307	0,7867	0,2296
10%	0,5270	0,4812	0,7291	0,7316	0,5765	0,3649

- Tableaux B1 et B2 correspondent à la Figure 39 : méthodes « simples » avec décalages.

B1 : bruit 1%

images

décalage	SIMPLISMA	OPA(var)	OPA(spec)	PCA	IKSFA	SPEXFA
0	0,2537	0,3940	0,9170	0,8364	0,2572	0,2560
1	0,2540	0,3940	0,9627	0,8365	0,2572	0,2560
2	0,2537	0,3940	0,9627	0,7868	0,2572	0,2560

spectres

décalage	SIMPLISMA	OPA(var)	OPA(spec)	PCA	IKSFA	SPEXFA
0	0,2357	0,5986	0,6795	0,7308	0,5336	0,2300
1	0,3749	0,6503	0,7032	0,8066	0,8097	0,3754
2	0,5482	0,7330	0,7454	0,8468	0,8409	0,5483

B2 : bruit 10%

images

	SIMPLISMA	OPA(var)	OPA(spec)	PCA	IKSFA	SPEXFA
0	0,3904	0,4018	0,8560	0,8383	0,4091	0,3045
1	0,3904	0,4018	0,9330	0,8383	0,6376	0,3045
2	0,3904	0,4018	0,9398	0,7996	0,6376	0,3045

spectres

	SIMPLISMA	OPA(var)	OPA(spec)	PCA	IKSFA	SPEXFA
0	0,5270	0,4812	0,7291	0,73155	0,57653333	0,36491667
1	0,5909	0,5530	0,7244	0,80723333	0,62853333	0,45013333
2	0,6937	0,6637	0,7476	0,7511	0,68736667	0,58556667

- Tableaux C : correspondent à la Figure 41 (couplage des méthodes « simples » avec MCR-ALS en fonction du bruit).

images

Bruit	SIMPLISMA/MCR	OPA(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
1%	0,1654	0,1714	0,4944	0,4047	0,1391	0,1558	0,1561
3%	0,2391	0,1836	0,4909	0,4210	0,2949	0,1545	0,1576
10%	0,2444	0,1484	0,8603	0,5000	0,2780	0,2697	0,1861

## spectres

Bruit	SIMPLISMA/MCR	OPA(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
1%	0,0974	0,1172	0,4234	0,3264	0,1098	0,0964	0,0965
3%	0,1786	0,1487	0,4282	0,3259	0,1329	0,0964	0,0974
10%	0,2227	0,1045	0,7149	0,18782	0,2211	0,2847	0,1349

- Tableaux D1 et D2 : correspondent à la Figure 42 (couplage des méthodes « simples » avec MCR-ALS en fonction des décalages).

D1 : bruit 1%

## images

décalage	SIMPLISMA/MCR	OPA(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
0	0,1654	0,1714	0,4944	0,4047	0,1391	0,1558	0,1561
1	0,1656	0,1646	0,5978	0,4215	0,1733	0,1558	0,1561
2	0,1654	0,1647	0,4834	0,3611	0,1572	0,1559	0,1561

## spectres

décalage	SIMPLISMA/MCR	OP(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
0	0,0974	0,1172	0,4234	0,3264	0,1098	0,0964	0,0965
1	0,2954	0,3102	0,5740	0,4247	0,3098	0,2969	0,2969
2	0,4945	0,5034	0,6312	0,5022	0,5094	0,4956	0,4956

D2 : bruit 3%

## images

décalage	SIMPLISMA/MCR	OPA(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
0	0,2391	0,1836	0,4909	0,4210	0,2949	0,1545	0,1576
1	0,2705	0,1613	0,4964	0,4208	0,1698	0,1545	0,1576
2	0,2701	0,1622	0,4966	0,4208	0,2182	0,1545	0,1576

## spectres

décalage	SIMPLISMA/MCR	OPA(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
0	0,1786	0,1487	0,4282	0,3259	0,1329	0,0964	0,0974
1	0,3428	0,3309	0,5046	0,4241	0,3259	0,2967	0,2971
2	0,5091	0,5022	0,6332	0,5665	0,5066	0,4957	0,4957

- Tableaux E : correspondent à la Figure 49 (couplage des méthodes « simples » avec MCR-ALS et application des contraintes du rang local) en fonction de différents niveaux de bruit.

## images

Bruit	SIMPLISMA/MCR	OPA(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
1%	0,1475	0,1039	0,2081	0,3022	0,1571	0,1346	0,1348
3%	0,2097	0,1132	0,3487	0,2993	0,1658	0,1340	0,1360
10%	0,1690	0,1333	0,3471	0,1781	0,2777	0,2110	0,1697

## spectres

Bruit	SIMPLISMA/MCR	OPA(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
1%	0,0764	0,0667	0,173	0,214	0,1274	0,0712	0,0712
3%	0,1553	0,0733	0,289	0,2109	0,1328	0,0726	0,073
10%	0,1459	0,0866	0,3255	0,1279	0,2209	0,2223	0,1158

- Tableaux F1 et F2 : correspondent à la Figure 50 (couplage des méthodes « simples » avec MCR-ALS et application des contraintes du rang local) en fonction des décalages.

Tableaux F1 : Bruit 1%

images

décalage	SIMPLISMA/MCR	OPA(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
0	0,1475	0,1039	0,2081	0,3022	0,1571	0,1346	0,1348
1	0,1480	0,1027	0,3684	0,3022	0,1733	0,1346	0,1348
2	0,1475	0,1027	0,3688	0,1634	0,1572	0,1346	0,1348

spectres

décalage	SIMPLISMA/MCR	OPA(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
0	0,0764	0,0667	0,173	0,2140	0,1274	0,0712	0,0712
1	0,2854	0,2843	0,4308	0,3490	0,3098	0,2838	0,2838
2	0,4873	0,4863	0,5814	0,4871	0,5094	0,4861	0,4861

Tableaux F2 : Bruit 3%

images

décalage	SIMPLISMA/MCR	OPA(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
0	0,2097	0,1132	0,3487	0,2993	0,1658	0,1340	0,1360
1	0,2096	0,1108	0,36345	0,2992	0,1697	0,1340	0,1360
2	0,21	0,1111	0,3633	0,2992	0,2181	0,1339	0,1359

spectres

décalage	SIMPLISMA/MCR	OPA(var)/MCR	OPA(spec)/MCR	PCA/MCR	PMF	IKSFA/MCR	SPEXFA/MCR
0	0,1553	0,0733	0,2890	0,2109	0,1328	0,0726	0,0730
1	0,3114	0,2864	0,4264	0,3472	0,3258	0,2842	0,2841
2	0,4950	0,4879	0,5779	0,5210	0,5066	0,4864	0,4862

- Tableaux G : correspondent à la Figure 55 (comparaison générale des comportements des méthodes étudiées dans les différents cas d'étude ; MS : Méthodes Simples, MSMCR : couplage des méthodes simples avec MCR-ALS, MSMCRRgLC : application des contraintes du rang local au couplage MSMCR).

images

Cas étudié	1%				10%			
	SIMPLISMA	OPA(var)	IKSFA	SPEXFA	SIMPLISMA	OPA(var)	IKSFA	SPEXFA
MS	0,5624	0,7495	0,5697	0,5670	0,7404	0,7575	0,9998	0,5728
MSMCR	0,3505	0,5027	0,4098	0,4115	0,6649	0,4338	0,4826	0,3624
MSMCRRgLc	0,3158	0,3158	0,356	0,3572	0,5003	0,3757	0,3456	0,3329

spectres

Cas étudié	1%				10%			
	SIMPLISMA	OPA(var)	IKSFA	SPEXFA	SIMPLISMA	OPA(var)	IKSFA	SPEXFA
MS	0,4601	0,6823	0,8731	0,1448	0,6615	0,7294	0,8971	0,5071
MSMCR	0,0445	0,0951	0,0386	0,0385	0,5198	0,0645	0,3622	0,0620
MSMCRRgLc	0,0384	0,0384	0,0338	0,0338	0,3069	0,0780	0,2663	0,0759

## Tables des illustrations

---

### Figures

Figure 1 : Procédé d'acquisition de spectres et de constitution de données spectrales. ....	14
Figure 2 : Les catégories de l'image, la forme des données acquises et leur restructuration pour les traitements (A. De Juan et al.) .....	15
Figure 3 : Illustration de la méthodologie de production d'une image par longueur d'onde caractéristique (P. DUMAS et al.) .....	16
Figure 4 : Diagramme montrant le calcul itératif pour obtenir les q facteurs significatifs fournis par le $\chi^2$ pour la décomposition de la matrice de données D. ....	30
Figure 5 : Explication de la méthodologie de la procédure EFA. A) EFA forward par addition successive des lignes en commençant à la 1 <sup>ère</sup> ligne ; B) EFA backward par addition successive des lignes en commençant par la dernière ligne. ....	36
Figure 6 : Graphe EFA, représentation du logarithme des valeurs singulières en fonction du temps de rétention. Les courbes en tiret plein sont fournies par EFA Forward et celles en pointillé (.....) par EFA Backward. Les traits pleins en dessous indiquent la fenêtre de concentration pour chaque composé (M. Maeder). ....	37
Figure 7 : Graphe <i>FMSW-EFA</i> ; a) représentation du logarithme des valeurs propres (ou singulières) en fonction du numéro des fenêtres ; b) concentration en fonction du temps de rétention. (F. Cuesta Sanchez et al.). ....	38
Figure 8 : Méthodologie de la procédure FSMW-EFA. Dans cette illustration la fenêtre de balayage a une taille de 4 lignes. A chaque étape, elle se déplace d'un spectre et donne les différentes fenêtres $F_i$ schématisées dans le schéma. À chaque position, une SVD est réalisée. ....	39
Figure 9 : Explication du déroulement de la procédure FSIW-EFA appliquée à une image 2D (a) et à une image 3D multicouche (b). (Anna De Juan et al.). ....	40
Figure 10 : Interface utilisateur produit par SIMPLISMA pour sélectionner les variables les plus pures. En haut, le spectre de la pureté produit après le calcul de la 1 <sup>ère</sup> pureté pour chaque variable colonne d'une matrice de données D(2500×460). En bas le même spectre corrigé par la déviation standard. L'offset choisi dans ce cas est égal à 10. La position du curseur indique la variable, ici 80, ayant la plus grande pureté. ....	42
Figure 11 : Sélection des variables pures successives pour une matrice de données D(2500×460) par la procédure interactive SIMPLISMA. Après sélection de chaque variable pure, un nouveau spectre de pureté est produit. La sélection s'arrête lorsque le spectre de pureté prend un aspect assez aléatoire comme on peut le voir sur la partie b. ....	42
Figure 12 : Interface utilisateur fourni par OPA après application à une matrice de données D(2500×460), a) pour la sélection des variables pures, b) pour la sélection des spectres purs. ....	44
Figure 13 : Procédure de sélection successive des variables pures (concentrations) par OPA; a) 1 <sup>ère</sup> interface produite par OPA pour la sélection des variables pures avant qu'aucune sélection n'est effectuée ; b) nouveau spectre de dissimilarité recalculé après sélection de la 1 <sup>ère</sup> variable la plus pure ; c) aspect du spectre de dissimilarité après avoir sélectionné les 6 premières variables les plus pures pour un mélange constitué de 6 composés purs. A noter l'aspect assez déformé de ce spectre de dissimilarité après la 6 <sup>ème</sup> sélection. ....	45

Figure 14 : Schéma général de résolution de données. L'optimisation des résultats est la minimisation des résidus par des procédures itératives avec injection des contraintes spécifiques.....	49
Figure 15 : Schéma du déroulement de l'algorithme MCR-ALS.....	53
Figure 16 : Visualisation schématique des transitions entre les niveaux de vibration et rotation.....	58
Figure 17 : Modélisation d'une liaison chimique entre deux atomes et la représentation de cette dernière par un oscillateur harmonique.....	59
Figure 18 : historique du développement des détecteurs infrarouges (d'après A. ROGALSKI).....	63
Figure 19 : Les réponses spectrales caractéristiques ( $D^*$ ) de nombreux détecteurs infrarouges pour l'intervalle [1-18 $\mu\text{m}$ ] avec leur température spécifique de fonctionnement (documentation technique, Hamamatsu).....	64
Figure 20 : Illustration de la Transformée de Fourier TF. a) et c) Spectres dans l'espace de temps avec une et deux fréquences respectivement ; b) et d) les spectres par Transformée de Fourier pour a) et c), les fréquences sont bien mises en évidence (J. Michael Hollas). La transformée de Fourier inverse $TF^{-1}$ réalise le passage inverse. L'élargissement des raies dépend d'un certain nombre de paramètres (durée de vie de l'état excité pour la RMN par exemple).....	66
Figure 21 : Schéma de principe d'un spectromètre à Transformée de Fourier avec interféromètre de Michelson.....	68
Figure 22 : a) Transition montrant la diffusion Rayleigh et Raman, b) Déplacement Raman Stokes et Anti-Stokes en fonction de l'intensité Raman ( $\bar{\nu}_0$ fréquence de l'onde incidente, $\bar{\nu}_R$ fréquence diffusion Raman).....	70
Figure 23 : Schéma du dispositif de l'instrumentation Raman.....	71
Figure 24 : a) Mode de production du rayonnement laser dans la cavité résonante par excitation et amplification (inrs) ; b, c et d) les différents types de lasers selon le nombre de niveaux d'énergie. Pour le laser à 4 niveaux d'énergies (d), l'espèce excitée dans l'état métastable $E_2$ a une durée de vie plus longue que celle dans l'état $E_1$ , ce qui permet une inversion de la population et une amplification de l'effet laser.....	72
Figure 25 : Représentation schématique de la détection de 3 générations des détecteurs basés sur le MCT (documentation technique, Hamamatsu) ; a) détection élément par élément, b) détection d'une rangée entière, c) détection 2D d'une surface.....	76
Figure 26 : Exemple de configuration d'un dispositif d'acquisition d'image en infrarouge (documentation technique, Hamamatsu).....	77
Figure 27 : Appareillage de la microspectrométrie infrarouge (IRscopeII et spectromètre IR ... etc.).....	79
Figure 28 : Configuration optique du microscope optique IRscope2 utilisé au laboratoire dans l'imagerie microspectrométrique infrarouge pour les deux modes d'utilisation (transmission et réflexion) (documentation technique Bruker).....	80
Figure 29 : Illustration de la production d'images spectroscopiques à partir des bandes spectrales par intégration... 82	82
Figure 30 : Visualisation du mode de sélection d'une zone spectrale caractéristique pour le composé pur D.....	84
Figure 31 : Les cartes de distribution (images) et les spectres de 6 composés de référence.....	90
Figure 32 : La constitution de la matrice de données expérimentales, $S_i$ ( $i=1 :6$ ) spectre du composé pur, $u_i$ ( $i=1 :2500$ ) spectre de mélange formé par la combinaison linéaire des concentrations et des spectres purs de 6 composés de référence. $I_i$ ( $i=1 :6$ ) image associée au spectre $S_i$ .....	91
Figure 33 : La procédure de conversion de la matrice tridimensionnelle $D_c(x,y,\lambda)$ en matrice dépliée bidimensionnelle $D_d(x*y,\lambda)$ dans laquelle les spectres sont placés selon l'ordre de balayage.....	91
Figure 34 : Aspect d'un spectre de mélange en fonction de différents niveaux de bruits ajoutés.....	93

Figure 35 : Illustration de l'effet des décalages sur le premier spectre de $D_a$ . a) les trois spectres (spectre sans décalage (rouge) et spectres avec des décalages de 1 (bleu) et 2 (vert) incréments de nombres d'onde) ; b) et c) zoom pour mieux visualiser ces décalages. ....	94
Figure 36 : Critère de comparaison entre les éléments extraits et leurs références. Evaluation des méthodes de résolution entre elles. ....	96
Figure 37 : Influence du paramètre d'offset $\alpha$ sur la capacité de résolution de l'algorithme SIMPLISMA. Le meilleur offset dans notre cas est compris entre 10 et 50. ....	98
Figure 38 : Comparaison globale de 6 méthodes « simples » pour les images (cartes de concentrations) et les spectres en fonction de la dissimilarité moyenne globale. ....	100
Figure 39 : Etude de la capacité d'extraction des méthodes « simples » en fonction des décalages pour un niveau de bruit donné (à gauche 1%, à droite 10%). ....	101
Figure 40 : Diagramme détaillant le couplage des 6 méthodes « simples » avec MCR-ALS. ....	106
Figure 41 : Comparaison des résultats du couplage des méthodes « simples » avec MCR-ALS en fonction du bruit. ....	107
Figure 42 : Etude des résultats du couplage des méthodes « simples » avec MCR-ALS en fonction des décalages à l'aide des dissimilarités moyennes globales pour les niveaux de bruit de 1% et 3%. ....	108
Figure 43 : La procédure globale pour le calcul des contraintes du rang local. ....	110
Figure 44 : Schéma explicatif de la production des cartes du rang local par l'utilisation de FSIW-EFA. ....	111
Figure 45 : Réorganisation des valeurs singulières pour la sélection de la bande « seuil » servant à la production des cartes du rang local. Dans le graphe le plus à droite, les valeurs singulières rattachées au bruit sont entassées en bas. ....	112
Figure 46 : Cartes du rang local produites pour les extrémités de la bande « seuil », pour la position médiane de la bande (rang local complet) et la carte du rang local partiel pour toute la bande. ....	113
Figure 47 : Procédure de production des cartes des absences des composés de référence et de la matrice des contraintes du rang local. ....	114
Figure 48 : Visualisation des productions des résultats par ALS en utilisant les contraintes du rang local, les données expérimentales et les autres paramètres nécessaires à la résolution. ....	115
Figure 49 : Comparaison globale des résultats avec le couplage avec MCR-ALS, à gauche sans contraintes du rang local et à droite avec les contraintes du rang local. ....	116
Figure 50 : Comparaison des résultats pour les niveaux de bruit de 1% et 3% lors de l'application des contraintes du rang local après réalisation des décalages en nombres d'onde. ....	117
Figure 51 : Qualité d'extraction des 4 meilleures méthodes appliquées dans différentes conditions de résolution. Le nombre entre parenthèses indique le classement de la méthode selon le critère de la dissimilarité moyenne. ....	119
Figure 52 : Visualisation des images et des spectres extraits pour les 4 meilleurs algorithmes dans le cas des méthodes « simples ». Les ellipses et les flèches rouges indiquent l'apparition des bandes supplémentaires (ou artefacts). Les ellipses bleues sur les images indiquent une surestimation de la concentration. ....	120
Figure 53 : Visualisation des images et des spectres extraits pour les 4 meilleurs algorithmes avec le couplage avec MCR-ALS. Les ellipses et les flèches rouges indiquent l'apparition des bandes supplémentaires (ou artefacts). Les ellipses bleues sur les images indiquent une surestimation de la concentration. ....	121
Figure 54 : Visualisation des images et des spectres extraits pour les 4 meilleurs algorithmes avec le couplage de MCR-ALS et l'application des contraintes du rang local. Les ellipses et les flèches rouges indiquent	

l'apparition des bandes supplémentaires (ou artefacts). Les ellipses bleues sur les images indiquent une surestimation de la concentration. ....	123
Figure 55 : Comparaison générale des comportements pour les 3 cas étudiés et pour les 2 niveaux de bruit pour les images et les spectres. Signification des initiales : MS - Méthode Simples ; MSMCR : Méthodes Simples /MCR-ALS ; MSMCRRgLc : Méthodes Simples/MCR-ALS/Rang Local. ....	124
Figure 56 : Spectromètre FT-Raman RFS 100/S utilisé pour l'acquisition spectrale sur le calcul urinaire. ....	130
Figure 57 : Schématisation de la zone de « cartographie spectrale » sur la coupe du calcul urinaire. ....	131
Figure 58 : Quelques spectres expérimentaux. La dérive de la ligne de base est importante et un bruit de fond persistant est visible. ....	133
Figure 59 : Estimation du nombre de constituants par la fonction IND et par la SVD. IND fournit 10 facteurs alors que la SVD n'indique que 3. ....	134
Figure 60 : Estimation du nombre de constituants par SIMPLISMA et par OPA. Nous obtenons une estimation maximale de 5 facteurs. ....	135
Figure 61 : Prétraitements de nos collections spectrales ; a) données spectrales brutes avant prétraitement, b) données prétraitées avec correction de la ligne de base et lissage. Le spectre 1531 donne une illustration de ces deux cas. ....	138
Figure 62 : Les différentes étapes de notre étude, de la préparation de l'échantillon aux traitements chimiométriques. ....	139
Figure 63 : Méthodologie générale de résolution adoptée pour l'analyse du calcul urinaire. ....	140
Figure 64 : Images et spectres extraits par la « méthode simple » OPA(var). A) visualisation globale de l'ensemble des images et spectres extraits ; B) représentation des composés extraits identifiés et comparaison de leur spectre avec celui de la référence correspondante. ....	142
Figure 65 : Spectre du composé extrait 1 (rouge) avec les spectres de l'acide urique anhydre (bleu) et l'acide urique dihydraté (noir). L'accord est meilleur avec l'acide urique anhydre. ....	143
Figure 66 : Discrimination entre les composés de référence 21 et 22. a) comparaison des spectres de 2 composés de référence 21 et 22 avec le composé extrait 2 (rouge) dans l'intervalle 1800-400 $\text{cm}^{-1}$ ; b) un zoom dans la partie 1530-1360 $\text{cm}^{-1}$ pour mieux voir l'accord du 22 avec l'extrait. ....	144
Figure 67 : Comparaison du spectre du composé extrait 3 avec les spectres de 4 composés de référence candidats. Pointages de quelques pics caractéristiques à ces spectres de référence. ....	145
Figure 68 : Images et spectres extraits par la « méthode simple » SIMPLISMA. A) visualisation globale de l'ensemble des images et spectres extraits ; B) représentation des composés extraits identifiés et comparaison de leur spectre avec celui de la référence correspondante. ....	148
Figure 69 : Images et spectres extraits par le couplage OPA(var)/MCR-ALS. A) visualisation globale de l'ensemble des images et spectres extraits ; B) représentation des composés extraits identifiés et comparaison de leur spectre avec celui de la référence correspondante. ....	149
Figure 70 : Différentes étapes de calcul du rang local. A) distribution de chaque valeur propre extraite donnant une idée du nombre de composés différents dans l'échantillon étudié ; B) graphe de l'évolution des valeurs singulières en fonction de leur indice ou du numéro des pixels pour pouvoir sélectionner le seuil limite de distinction entre le bruit et les valeurs significatives ; C) représentation des cartes du rang local : 1° Carte rang local complet, 2° carte du rang local au minimum et au maximum de la « bande seuil », 3° carte du rang local partiel. ....	151

Figure 71 : Sélection de la « bande seuil » pour la distinction des valeurs singulières rattachées au bruit et celles significatives correspondant à des composés extractibles. .... 152

Figure 72 : Production des cartes des absences de 5 « composés » extraits de la matrice de données. Les points rouges indiquent les pixels sur lesquels le composé considéré est absent. A noter les zones d'absence similaires des « composés » 4 et 5..... 153

Figure 73 : Images et spectres extraits après application des contraintes du rang local aux résultats du couplage OPA(var)/MCR-ALS. A) visualisation globale de l'ensemble des images et spectres extraits ; B) représentation des composés extraits identifiés et comparaison de leurs spectres avec ceux de leurs références correspondantes. .... 154

Figure 74 : Les différents procédés « d'augmentation de matrices ». 1) augmentation des dimensions des colonnes sans modification des dimensions des lignes ; 2) procédé inverse du 1) en n'augmentant que les dimensions des lignes ; 3) augmentation à la fois des dimensions des lignes et colonnes..... 155

Figure 75 : Résolution de la matrice augmentée et les différentes contraintes à injecter dans ALS pour optimiser les spectres et les concentrations..... 156

Figure 76 : Résultats du procédé « augmentation de matrices ». Images et spectres extraits de la matrice augmentée après application des contraintes du rang local et des contraintes liées à « l'augmentation de matrices » aux résultats du couplage OPA(var)/MCR-ALS. A) visualisation globale de l'ensemble des images et spectres extraits ; B) représentation des composés extraits identifiés et comparaison de leurs spectres avec leurs références correspondantes..... 157

## Tableaux

Tableau 1 : Présentation de la matrice des données recueillies, observées, mesurées (...etc.) sous la forme d'un tableau à 2 dimensions..... 20

Tableau 2 : Décomposition de la matrice de données D en produit de 2 matrices R et C en fonction du nombre de facteurs significatifs choisis (n). .... 21

Tableau 3 : Classement des valeurs singulières dans W diagonale ( $\lambda_1 > \lambda_2 > \dots > \lambda_p$ ) pour ne retenir dans la décomposition en facteurs significatifs que les « q » valeurs singulières significatives..... 24

Tableau 4 : Sélection de matrice  $W^*(r \times q)$  des valeurs propres significatives associées aux principaux vecteurs propres donnant la dimension de l'espace de décomposition, c'est-à-dire les facteurs significatifs. .... 25

Tableau 5 : Signification et interprétation des résultats fournis par la fonction d'Exner. .... 27

Tableau 6 : Les principaux détecteurs infrarouges (thermiques & quantiques) et leurs principales caractéristiques (documentation technique, Hamamatsu). .... 63

Tableau 7 : Quelques sources lasers et leurs caractéristiques. .... 73

Tableau 8 : Présentation des méthodes chimométriques de notre étude. .... 88

Tableau 9 : Extraction par SIMPLISMA de 6 variables les plus pures pour chaque offset. .... 97

Tableau 10 : Etude de l'évolution du rapport offset/moyenne pour la première variable pure extraite par SIMPLISMA. .... 98

Tableau 11 : Comparaison de la qualité d'extraction de OPA(spec) (tableau A) et de OPA(var) (tableau B). Comparaison des composés extraits par chacune des méthodes à ceux de la référence. .... 103

Tableau 12 : Attribution des variables extraites par OPA(var) et leur caractérisation en nombre d'onde. .... 104

Tableau 13 : Angles (exprimés en degré) entre les spectres purs de 6 composés de référence. .... 105

Tableau 14 : Résultats d'estimation fournis par les méthodes. .... 136

Tableau 15 : Principales bandes caractéristiques (en $\text{cm}^{-1}$ ) de chacun de 4 composés. Les nombres en gros caractère gras, italique et souligné indiquent les pics spécifiques qui nous permettent de différencier les 4 composés. Les zones colorées indiquent l'absence des pics qui peut être intéressante pour la distinction. La dernière ligne correspond au composé extrait à identifier. ....	146
Tableau 16 : Récapitulatif de l'analyse des éléments extraits par OPA(var) et de l'identification des éléments présents dans l'échantillon. S1, S3, S4 spectres des composés extraits dans l'ordre d'extraction. ....	146
Tableau 17 : Récapitulatif de l'analyse des éléments extraits par SIMPLISMA et de l'identification des éléments présents dans l'échantillon. S1, S2, S3 spectres des composés extraits dans l'ordre d'extraction. ....	148
Tableau 18 : Résultats de l'extraction par le couplage OPA(var)/MCR-ALS et l'identification des éléments présents dans l'échantillon. S1, S3, S4 spectres des composés extraits dans l'ordre d'extraction. ....	150
Tableau 19 : Résultats de l'extraction après application des contraintes du rang local au couplage OPA(var)/MCR-ALS et l'identification des éléments présents dans l'échantillon. S1, S3, S4 spectres des composés extraits dans l'ordre d'extraction. ....	155
Tableau 20 : Bilan d'identification après résolution par le procédé « augmentation de matrices » en utilisant les résultats de l'application des contraintes du rang local au couplage OPA(var)/MCR-ALS. S1, S3, S4 spectres des composés extraits dans l'ordre d'extraction. ....	158
Tableau 21 : Récapitulatif des éléments extraits et identifiés avec nom, formule, image et spectre dans le cas du procédé « augmentation de matrices ». ....	160

## Bibliographie

---

- <sup>1</sup> D. E. BUGAY. Characterization of solid-state: spectroscopic techniques. *Advanced Drug Delivery Reviews*, 48 (2001), pp43-65.
- <sup>2</sup> A. De JUAN, R. TAULER, R. DYSON, C. MARCOLLI, M. RAULT, M. MAEDER. Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis. *Trends in Analytical Chemistry*, 23 (1) (2004), pp70-79.
- <sup>3</sup> P. DUMAS, G. L. CARR, G. P. WILLIAMS. Enhancing the lateral resolution in infrared microspectrometry by using synchrotron radiation: application and perspectives. *Analisis*, 28(1) (2000), pp68-74.
- <sup>4</sup> J. TORRENS-IBERN. Modèles et méthodes de l'analyse factorielle. *Dunod*, Paris, (1972), 202p.
- <sup>5</sup> C. CHATFIELD, A. J. COLLINS. Introduction to multivariate analysis. *Chapman & Hall*, London, 2<sup>nd</sup> edit, 1986, 246p.
- <sup>6</sup> H. H. HARMAN. Modern factor analysis, 3<sup>ème</sup> edit., *University of Chicago Press*, Chicago (1976), 487p.
- <sup>7</sup> F. ROTELLA, P. BORNE. Théorie et pratique du calcul matriciel. *Collection Méthodes et pratiques de l'ingénieur*, éditions *Technip* (1995), pp92-98.
- <sup>8</sup> E. R. MALINOWSKI. Factor analysis in Chemistry (2nd édit). *John Wiley & Sons, Inc*, New York (1991), 350p.
- <sup>9</sup> M. WASIM, R. G. BRERETON. Determination of significant components in liquid chromatography nuclear magnetic resonance spectroscopy. *Chemom. Intell. Lab. Syst.*, 72 (2004), pp133-151.
- <sup>10</sup> M. MELOUN, J. ČAPEK, P. MIKŠIK, R. G. BRERETON. Critical comparison of methods predicting the number of components in spectroscopic data. *Analytica Chimica Acta*, 423 (2000), pp51-68.
- <sup>11</sup> J. H. KINDSVATER, P. H. WEINER, T. J. KLINGEN. *Anal. Chem.* 46 (1974), p982.
- <sup>12</sup> O. EXNER. *Collect. Czech. Chem. Commun.*, 31 (1966), p3222.
- <sup>13</sup> J. DURBIN, G. S. WATSON. *Biometrika*, 37 (1950), pp409-428.
- <sup>14</sup> J. JEONG, S. CHUNG. Bootstrap tests for autocorrelation. *Computational Statistics & Data Analysis*, 38(1) (2001) pp49-69.
- <sup>15</sup> D. N. RUTLEDGE, A. S. BARROS. Durbin–Watson statistic as a morphological estimator of information content. *Analytica Chimica Acta*, 454(2) (2002), pp277-295.
- <sup>16</sup> J.-L. FÉMÉNIAS. Fitting models to correlated data (large samples). *Journal of Molecular Spectroscopy*, 224(1) (2004), pp73-98.

- 
- <sup>17</sup> M. S. BARTLETT. *Br. J. Psychol. Stat.*, 3 (1950), p77.
- <sup>18</sup> K. ARTYUSHKO, J. E. FULGHUM. Identification of chemical components in XPS spectra and images using multivariate statistical analysis methods. *Journal of Electron Spectroscopy and Related Phenomena*, 121(1-3) (2001), pp33-55.
- <sup>19</sup> S. WOLD. Chemometrics. What do we mean with it, and what do we want from it? *Chemom. Intell. Lab. Syst.*, 30 (1995), pp109-115.
- <sup>20</sup> J.-J. DROESBEKE. Histoire de la Statistique. *Presses universitaires de France*, 1<sup>ère</sup> édit. (1990), pp89-100.
- <sup>21</sup> K. PEARSON. *Philos. Mag*, series 6, 2 (1901), p559.
- <sup>22</sup> H. H. HARMAN. Modern Factor Analysis, 2<sup>nd</sup> edit. *University of Chicago Press*, Chicago (1967).
- <sup>23</sup> H. & M. MARTENS. Multivariate Analysis of quality: an introduction. *J. Wiley*, Chichester (2001), pp93-110 & pp376-381.
- <sup>24</sup> W. H. LAWTON, E. A. SYLVESTRE. Self modeling curve resolution. *Technometrics*, 13 (1971), pp617-633.
- <sup>25</sup> L. LEBART, M. PIRON, A. MORINEAU. Statistique exploratoire multidimensionnelle, *Dunod*, 3<sup>ème</sup> édit, Paris (2000), pp32-66.
- <sup>26</sup> T. FOUCART. L'analyse des données : mode d'emploi. *Presses Universitaires de Rennes*, (1997), pp15-52.
- <sup>27</sup> M. MAEDER. Evolving factors analysis for the resolution of overlapping chromatographic peaks, *Anal. Chem.*, 59 (1987), pp527-530.
- <sup>28</sup> H. R. KELLER, D. L. MASSART. Evolving factor analysis. *Chemom. Intell. Lab. Syst.*, 12(3) (1992), pp209-224.
- <sup>29</sup> F. C. SANCHEZ, B. Van Den BOGAERT, S. C. RUTAN, D. L. MASSART. Multivariate peak purity approaches (Tutorial), *Chemom. Intell. Lab. Syst.*, 34 (1996), pp139-171.
- <sup>30</sup> K. De BRAEKELEER, A. De JUAN, D. L. MASSART. Purity assessment and resolution of tetracycline hydrochloride samples analysed using high-performance liquid chromatography with diode array detection. *Journal of Chromatography A*, 832(1-2) (1999), pp67-86.
- <sup>31</sup> F. C. SANCHEZ, J. TOFT, O. M. KVALHEIM, D. L. MASSART. Eigenstructure tracking analysis for assessment of peak purity in high-performance liquid chromatography with diode array detection. *Analytica Chimica Acta*, 314( 3) (1995), pp131-139.
- <sup>32</sup> A. De JUAN, M. MAEDER, T. HANCEWICZ, R. TAULER. Local rank analysis for exploratory spectroscopic image analysis. Fixed size image Window-Evolving Factor Analysis. *Chemom. Intell. Lab. Syst.*, 77 (2005), pp64-74.
- <sup>33</sup> W. WINDIG, J. GUILMENT. Interactive self-modeling mixture analysis. *Analytical Chemistry*. 63 (1991), pp1425-1432.

- 
- <sup>34</sup> G. L. RITER, S. R. LOWRY, T. L. ISENHOUR, C. L. WILKINS. Factor analysis of the mass spectra of mixtures. *Analytical Chemistry*, 48 (1976), pp591-595.
- <sup>35</sup> W. WINDIG, B. ANTALEK, J. L. LIPPERT, Y. BATONNEAU, C. BRÉMAR. Combined use of conventional and second-derivative Data in the SIMPLISMA Self-Modeling Mixture Analysis Approach. *Analytical Chemistry*, 74(6) (2002), pp.1371-1379.
- <sup>36</sup> F. C. SANCHEZ, M. S. KHOTs, D. L. MASSART, J. O. De BEER. Algorithm for the assessment of peak purity in liquid chromatography with photodiode-array detection. *Analytica Chimica Acta*, 285(1-2) (1994), pp181-192.
- <sup>37</sup> F. C. SANCHEZ, J. TOFT, B. Van Den BOGAERT, D. L. MASSART. Orthogonal Projection Approach Applied to Peak Purity Assessment. *Analytical Chemistry*, 68 (1996), pp79-85.
- <sup>38</sup> R. JAMAI, N. DAMIL. Influence of iterated Gram-Schmidt orthonormalization in the asymptotic numerical method. *Comptes Rendus Mécanique*, 331 (2003), pp351-356.
- <sup>39</sup> E. R. MALINOWSKI. Obtaining the key set of typical vectors by factor analysis and subsequent isolation of component spectra. *Analytica Chimica Acta*. 134 (1982), pp129-137.
- <sup>40</sup> K. J. SCHOSTACK, E.R MALINOWSKI. Preferred set selection by iterative key set factor analysis. *Chemom. Intell. Lab. Syst.* 6 (1989), pp21-29.
- <sup>41</sup> E. R. MALINOWSKI. Factor analysis in Chemistry (3<sup>ème</sup> édit). *John Wiley & Sons, Inc*, New York (2002), pp264-268.
- <sup>42</sup> L. R. TUCKER. *Psychometrika*, 31(3) (1966), pp279-311.
- <sup>43</sup> R. A. HARSHMAN. PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics*, 22 (1972), pp30-47.
- <sup>44</sup> P. J. GEMPERLINE. A priori estimates of the elution profiles of the pure components in overlapped liquid chromatography peaks using target factor analysis. *J. Chem. Inf. Comput. Sci.* 24 (1984), pp206-212.
- <sup>45</sup> B. G. M. VANDEGINSTE, W. DERKS, G. KATEMAN. Multicomponent self-modelling curve resolution in high-performance liquid chromatography by iterative target transformation analysis. *Analytica Chimica Acta*, 173 (1985), pp253-264.
- <sup>46</sup> G. T. RASMUSSEN, T. L. ISENHOUR, S. R LOWRY, G. L. RITER. Principal component analysis of the infrared spectra of mixtures. *Analytica Chimica Acta*, 103 (1978), p213
- <sup>47</sup> M. K. ANTOON, L. D'ESPOSITO, J. L. KOENIG. Factor analysis applied to Fourier Transform Infrared spectra. *Applied Spectroscopy*, 33 (1979), pp.351-357.
- <sup>48</sup> A. De JUAN, E. CASASSAS, R. TAULER. Soft modeling of analytical data. In *encyclopedia of analytical chemistry*. (R. A. Meyers, Eds), *J. Wiley*, Chichester (2000), pp9800-9837.
- <sup>49</sup> J.-H. JIANG, Y. LIANG, Y. OZAKI. Principles and methodologies in self-modeling curve resolution. *Chem. Intell. Lab. Syst.*, 71 (2004), pp1-12.

- 
- <sup>50</sup> A. De JUAN, M. MAEDER, M. MARTINEZ, R. TAULER. Application of a novel resolution approach combining soft- and hard-modelling features to investigate temperature-dependent kinetic processes. *Analytica Chimica Acta*, 442(2) (2001), pp337-350.
- <sup>51</sup> A. De JUAN, M. MAEDER, M. MARTINE, R. TAULER. Combining hard- and soft-modelling to solve kinetic problems. *Chemom. Intell. Lab. Syst.*, 54(2) (2000), pp123-141.
- <sup>52</sup> E. J. KARJALAINEN. The spectrum reconstruction problem: Use of alternating regression for unexpected spectral components in two-dimensional spectroscopies. *Chemom. Intell. Lab. Syst.*, 7 (1989), pp31-38.
- <sup>53</sup> R. TAULER, D. BARCELÓ. Multivariate curve resolution and calibration applied to liquid chromatography diode array detection. *Trends in Analytical Chemistry*, 12 (1993), pp319-327.
- <sup>54</sup> R. TAULER. Multivariate curve resolution applied to second order data. *Chemom. Intell. Lab. Syst.*, 30 (1995), pp133-146.
- <sup>55</sup> R. TAULER, B.R. KOWALSKI, S. FLEMMING. Multivariate Curve Resolution applied to spectral data from multiple runs of an industrial process. *Analytical Chemistry*, 65 (1993), pp2040-2047.
- <sup>56</sup> P. PAATERO. Least squares formulation of robust non-negative factor analysis. *Chemom. Intell. Lab. Syst.*, 37 (1997), pp23-35.
- <sup>57</sup> P. ANTTILA, P. PAATERO, U. TAPPER, O. JÄRVINEN. Source identification of bulk wet deposition in Finland by positive matrix factorization. *Atmospheric Environment*, 29(14) (1995), pp1705-1718.
- <sup>58</sup> Y. SONG, Y. ZHANG, S. XIE, L. ZENG, M. ZHENG, L. G. SALMON, M. SHAO, S. SLANINA. Source apportionment of PM<sub>2.5</sub> in Beijing by positive matrix factorization. *Atmospheric Environment*, 40(8) (2006), pp1526-1537.
- <sup>59</sup> K. ITO, N. XUE, G. THURSTON. Spatial variation of PM<sub>2.5</sub> chemical species and source-apportioned mass concentrations in New York City. *Atmospheric Environment*, 38(31) (2004), pp5269-5282.
- <sup>60</sup> M. JUVELA, K. LEHTINEN. The use of Positive Matrix Factorization in the analysis of molecular line spectra from the Thumbprint nebula. *Clouds, Cores, and Low Mass Stars, ASP conference series*, (D. P. Clemens, R. Barvainis, eds), 65 (1994), pp176-180.
- <sup>61</sup> [On-line : [www.epa.gov/ttnamti1/files/ambient/pm25/workshop/laymen.pdf](http://www.epa.gov/ttnamti1/files/ambient/pm25/workshop/laymen.pdf) ]: A Guide to Positive Matrix Factorization by PHILIP K. HOPKE. (15p)-(consultation: 30 novembre 2004)
- <sup>62</sup> M. JIANG, G. WANG, X.-M. MA. A general axiomatic system for image resolution quantification. *Journal of Mathematical Analysis and Applications*, 315(2) (2006), pp462-473.
- <sup>63</sup> W. HERSCHEL, *Philos. Trans.* (Royal society of London), 90 (1800) p284.
- <sup>64</sup> S. F. JOHNTON. Fourier transform infrared: a constantly evolving technology. *Ellis Horwood*, Chichester (England) (1991), p51.
- <sup>65</sup> A. ROGALSKI. Infrared detectors: an overview. *Infrared Physics & Technology*, 43 (2002), pp187-210.

- 
- <sup>66</sup> E. S. BARR. Historical survey of the early development of the infrared spectral region. *American Journal of Physics*, 28 (1960), pp42-54.
- <sup>67</sup> E. S. BARR. The infrared pioneers-II. Macedonio Melloni. *Infrared Physics*, 2 (1962), pp67-73.
- <sup>68</sup> R. A SMITH, F. E. JONES, R. P. CHASMAR. The detection and measurement of infrared radiation, Clarendon, Oxford (1958).
- <sup>69</sup> E. S. BARR. The infrared pioneers-III. Samuel Pierpont Langley. *Infrared Physics*, 3 (1963), pp195-206.
- <sup>70</sup> T. W. CASE. Notes on the change of resistance of certain substrates in light. *Physics Reviews*, 9 (1917), pp305-310.
- <sup>71</sup> E. LEHRER. *Z. Techn. Phys.* 23 (1942), pp169.
- <sup>72</sup> C. LA LAU. The analytical application of infra-red spectrophotometry. *Analytica Chimica Acta*, 2 (1948), pp734-743.
- <sup>73</sup> N. WRIGHT, L.W. HERSCHER. *Journal of Optical Society of America*, 37 (1947) p211.
- <sup>74</sup> J. R. HART, K. H. NORRIS, C. GOLUBIC. Determination of the moisture content of seeds by Near-Infrared Spectrophotometry of their methanol extracts. *Cereal. Chem.*, 39 (1962), pp94-99.
- <sup>75</sup> W. KAYE. Near-infrared. II. Instrumentation and technique (a review). *Spectrochimica Acta*, 7 (1955-1956), pp181-204.
- <sup>76</sup> S. F. D. ORR. Recent advances in infra-red instrumentation. *Spectrochimica Acta*, 14 (1959), pp86-96.
- <sup>77</sup> W. S. BOYLE, G. E. SMITH. Charge-coupled semiconductors devices, *Bell Syst. Tech., J.* 49 (1970), pp587-593.
- <sup>78</sup> A. ROGALSKI. Infrared detectors: status and trends (review). *Progress in Quantum Electronic*, 27 (2003), pp59-210.
- <sup>79</sup> J. ZIEGLER, M. BRUDER, M. FINCK, R. KRÜGER, P. MENGER, Th. SIMON, R. WOLLRAB. Advanced sensor technologies for high performance infrared detectors. *Infrared Physics & Technology*, 43 (2002), pp239-243.
- <sup>80</sup> A. ROGALSKI. Toward third generation HgCdTe infrared detectors. *J. of Alloys and Compounds*, 371 (2004), pp53-57.
- <sup>81</sup> B. DAS, P. SINGARAJU. Novel quantum wire infrared photodetectors. *Infrared Physics & Technology*, 46 (2005), pp209-218.
- <sup>82</sup> L. DUPONCHEL. Validation par les méthodes chimiométriques d'un analyseur spectrométrique de lait travaillant dans le proche infrarouge. Transfert de calibration en vue de développements industriels. *Th.* 1997, USTL Lille, 210p.
- <sup>83</sup> K. K. IRIKURA. Mass scaling for vibrational frequencies from ab initio calculations. *Chemical Physics Letters*, 403(4-6) (2005), pp275-279.

- 
- <sup>84</sup> Z.-G. WANG, H.-R. XIA. Molecular and laser spectroscopy. Springer series in chemical physics. Springer-Verlag, Berlin Heidelberg (1991), p14.
- <sup>85</sup> Y. ROGGO. Détermination de la qualité de la betterave sucrière par spectroscopie proche infrarouge et chimiométrie. *Th.* 2003, USTL Lille 1, 190p.
- <sup>86</sup> F. DE LENE MIROUZE. Analyse quantitative en spectrométrie moyen infrarouge par transformée de Fourier : perspective d'application dans le domaine agro-alimentaire. *Th.* 1992, USTL Lille1, 131p.
- <sup>87</sup> D. A. SKOOG, F. J. HOLLER. T. A. NIEMAN. Principes d'analyse instrumentale. *De Boeck*, Paris (2003), 960p.
- <sup>88</sup>: [on-line:  
[http://sales.hamamatsu.com/assets/applications/SSD/Characteristics\\_and\\_use\\_of\\_infrared\\_detectors.pdf](http://sales.hamamatsu.com/assets/applications/SSD/Characteristics_and_use_of_infrared_detectors.pdf): Technical information sd-12: Characteristics and use of infrared detectors].  
Visite:15/12/2004.
- <sup>89</sup> M. J. E. GOLAY. A pneumatic infrared detector, *Review of Scientific Instruments*, 18 (1947), pp357-362.
- <sup>90</sup> K. KPALMA, V. HAESE-COAT. Traitement numérique du signal: théorie et applications. *Ellipses, Collection Informatique industrielle*, Paris (2003), 242p.
- <sup>91</sup> J. MICHAEL HOLLAS. Spectroscopie : cours et exercices. *Dunod*, Paris (1998), 390p.
- <sup>92</sup> J. W. COOLEY, J. W. TUKEY. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.*, 19 (1965), pp297-301.
- <sup>93</sup> G. PEYRÉ. L'algèbre discrète de la transformée de Fourier. *Ellipses*, Paris (2004), 336p.
- <sup>94</sup> F. ROUESSAC. Analyse chimique : méthodes et techniques instrumentales modernes. *Masson*, Paris (1992), pp127-148.
- <sup>95</sup> [on-line : <http://www.sciences.univ-nantes.fr/physique/enseignement/tp/michelson/michp.html> : interféromètre de Michelson], visité: le 26/09/2005.
- <sup>96</sup> J. CHAMBERLAIN. The principles of interferometric spectroscopy. *J. Wiley*, Chichester (England) (1979), 347p.
- <sup>97</sup> A. G. SMEKAL. *Naturwiss*, 11 (1923), p873.
- <sup>98</sup> C. V. RAMAN, K. S. KRISHNAN. *Nature* 121 (3048), 501, (1928).
- <sup>99</sup> R. SINGH, F. RIESS. The 1930 Nobel Prize for physics: a close decision? *Notes and Records of Royal Society London* 55 (2001), pp267-283, *The Royal Society*.
- <sup>100</sup> G. D. MAHAN, K. R. SUBBASWAMY. Local density theory of Polarizability. *Plenum Press*, New York (1990), pp1-37 & pp77-129.
- <sup>101</sup> P. TRAMINI. Analyse par microspectrométrie Raman des tissus dentaires minéralisés et leur relation avec l'âge : étude statistique multivariée. *Th.* 1999, Montpellier 1, 256p.

- 
- <sup>102</sup> B. SCHRADER, H. SCHULZ, G. N. ANDREEV, H. H. KLUMP, J. SAWATZKI. Non-destructive NIR-FT-Raman spectroscopy of plant and animal tissues, of food and works of art. *Talanta*, 53(1) (2000), pp35-45.
- <sup>103</sup>[On-line : [http://www.inrs.fr/INRS-PUB/inrs01.nsf/IntranetObject-accesParReference/ED+5009/\\$File/ed5009.pdf](http://www.inrs.fr/INRS-PUB/inrs01.nsf/IntranetObject-accesParReference/ED+5009/$File/ed5009.pdf) ]: visité le 06/02/2006.
- <sup>104</sup> B. CHASE. Fourier transform Raman Spectroscopy. *Analytical Chemistry*, 59(14) (1987), pp881A-889A.
- <sup>105</sup> C. D. NEWMAN, G. G. BRET, R. L. MCCREERY. Fiber-Optic sampling combined with an imaging spectrograph for routine. Raman spectroscopy. *Applied Spectroscopy*, 46(2) (1992), pp262-265.
- <sup>106</sup> S. B. TÖRÖK, R. E. Van GRIEKEN. X-ray spectrometry. *Analytical Chemistry*, 66 (1994), pp186R-206R
- <sup>107</sup> M.-P. ISAURE, A. LABOUDIGUE, A. MANCEAU, G. SARRET, C. TIFFREAU, P. TROCELLIER, G. LAMBLE, J.-L. HAZEMANN, D. CHATEIGNIER. Quantitative Zn speciation in contaminated dredged sediment by  $\mu$ -PIXE,  $\mu$ -SXRF, EXAFS spectroscopy and principal component analysis. *Geochimica et Cosmochimica Acta*, 66(9) (2002), pp1549-1567.
- <sup>108</sup> J. E. KATON. Infrared microspectroscopy. A review of fundamentals and applications. *Micron*, 27(5) (1996), pp303-314.
- <sup>109</sup> D. S. ALDRICH, M. A. SMITH. Pharmaceutical applications of infrared microspectroscopy. In *Practical guide to infrared microspectroscopy*. Practical spectroscopy series, vol. 19 (Howard J. Humecki. Eds), *Marcel Dekker*, New York (1995), pp323-375.
- <sup>110</sup> C. KRAJEDJEL, L. MCDERMOTT. NIR analysis of polymers. *Handbook of near-infrared analysis*. Practical spectroscopy series, vol. 13 (Donald. A. Burns & Emil W. Ciurczak. Eds), *Marcel Dekker*, New York (1992), pp565-608.
- <sup>111</sup> M. R. DERRICK. Infrared microspectroscopy in the analysis of cultural artefacts. In *Practical guide to infrared microspectroscopy*. Practical spectroscopy series, vol. 19 (Howard J. Humecki. Eds), *Marcel Dekker*, New York (1995), pp287-322.
- <sup>112</sup> P. A. MARTOGLIO SMITH. Infrared microspectroscopy mapping studies of packaging materials: experiment design and data profiling considerations. *Vibrational Spectroscopy*, 24 (2000), pp47-62.
- <sup>113</sup> K. S. KALASINSKY. Drug distribution in human hair by infrared microscopy. *Cellular and Molecular Biology*. 44(1) (1998), pp81-87.
- <sup>114</sup> C. PALUSZKIEWIECZ, A. STOCH. FT-IR microscopic imaging of metal-ceramic joints. *Vibrational Spectroscopy*, 35(1-2) (2004), pp183-187.
- <sup>115</sup> N. A. NGO-THI, C. KIRSCHNER, D. NAUMANN. Characterization and identification of microorganisms by FT-IR microspectrometry. *Journal of Molecular Structure*, 661-662 (2003), pp371-380.
- <sup>116</sup> S. SAHU, D. L. EXLINE, M. P. NELSON. Identification of thaumasite in concrete by Raman chemical imaging. *Cement & Concrete Composites*, 24 (2002), pp347-350.

- 
- <sup>117</sup> A. NEMETZ, H. KNOBLOCH, W. KNOLL. Raman imaging with plasmon surface polaritons. *Thin Solid Films*, 226(1) (1993), pp48-52.
- <sup>118</sup> L. ZHANG, M. J. HENSON, S. S. SEKULI. Multivariate data analysis for Raman imaging of a model pharmaceutical tablet. *Analytica Chimica Acta*, 545(2) (2005), pp262-278.
- <sup>119</sup> N. DUPUY, Y. BATONNEAU. Reliability of the contribution profiles obtained through the SIMPLISMA approach and used as reference in calibration process: application to Raman micro-analysis of dust particles. *Analytica Chimica Acta*, 495 (2003), pp205-215.
- <sup>120</sup> P. CHAURAND, S. A. SCHWARTZ, R. M. CAPRIOLI. Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections (review article). *Current Opinion in Chemical Biology*, 6( 5) (2002), pp676-681.
- <sup>121</sup> J. DUBOIS, R. BAYDACK, E. MCKENZIE, T. BOOTH, M. JACKSON. Scrapie infection investigated by magnetic resonance imaging and Fourier transform infrared microscopy. *Vibrational Spectroscopy*, 32 (2003), pp95-105.
- <sup>122</sup> G. W. MCDANIEL, D. Z. ROBINSON. Thermal imaging by means of evaporographs. *Applied Optics*, 1 (1962), pp311-324.
- <sup>123</sup> C. HILSUM, W. R. HARDING. The theory of thermal imaging, and its application to the absorption-edge image tube, *Infrared Physics*, 1 (1961), pp67-93.
- <sup>124</sup> Ph. TRIBOLET. HgCdTe technology in France. *C. R. Physique*, 4 (2003), pp1121-1131.
- <sup>125</sup> J.-L. TISSOT. La détection infrarouge avec les plans focaux non refroidis: état de l'art. *C. R. Physique*, 4 (2003), pp1083-1088.
- <sup>126</sup> A. TREMEAU, C. FRENANDEZ-MALOIGNE, P. BONTON. Image numérique couleur : de l'acquisition au traitement. *Dunod*, Paris (2004), 460p.
- <sup>127</sup> K. KRISHNAN, J. R. POWELL, S. L. HILL. Infrared microimaging. In *Practical guide to infrared microspectroscopy*. Practical spectroscopy series, vol. 19 (Howard J. Humecki. Eds), *Marcel Dekker*, New York (1995), pp85-110.
- <sup>128</sup> B. C. SMITH. Fundamentals of Fourier Transform Infrared Spectroscopy. *CRC Press LLC*, Boca Raton (1996), 202p.
- <sup>129</sup> P. WILHELM. Applications of FT-IR microscopy with materials analyses. *Micron*, 27 (5) (1996), pp341-344.
- <sup>130</sup> R. G. MESSERSCHMIDT. Minimizing optical nonlinearities in infrared microspectroscopy. In *Practical guide to infrared microspectroscopy*. Practical spectroscopy series, vol. 19 (Howard J. Humecki. Eds), *Marcel Dekker*, New York (1995), pp3-39.
- <sup>131</sup> B. O. BUDESVKA. Minimization of optical non-linearities in Fourier transform-infrared microspectroscopic imaging. *Vibrational Spectroscopy*, 24 (2000), pp37-45.
- <sup>132</sup> I. BLAKEY, G. A. GEORGE. Raman spectral mapping of photo-oxidised polypropylene. *Polymer Degradation and Stability*, 70(2) (2000), pp269-275.

- 
- <sup>133</sup> A. J. SOMMER, J. E. KATON. Diffraction-Induced stray light in infrared microspectroscopy and its effect on spatial resolution. *Applied Spectroscopy*. 45(10) (1991), pp1633-1640.
- <sup>134</sup> J. A. REFFNER. Instrumental factors in infrared microspectroscopy. *Cellular and Molecular Biology*. 44(1) (1998), pp1-7.
- <sup>135</sup> F. E. LYTLE. An introduction to diffraction. Part I: The Near Field. *Applied Spectroscopy*. 53(6) (1999), pp212A-226A.
- <sup>136</sup> F. E. LYTLE. An introduction to diffraction. Part II: The Far Field. *Applied Spectroscopy*. 53(7) (1999), pp262A-276A.
- <sup>137</sup> M. A. HARTHCOCK, S. C. ATKIN. Imaging with functional group maps using infrared microspectroscopy. *Applied Spectroscopy*, 42 (3) (1988), pp449-455.
- <sup>138</sup> J. E. KATON, G. E. PACEY, J. F. O'KEEFE. Vibrational molecular microspectroscopy. *Analytical Chemistry*. 58 (1986), pp465A-481A.
- <sup>139</sup> M. A. HARTHCOCK, S. C. ATKIN. Infrared microspectroscopy: development and applications of imaging capabilities. In *Infrared microspectroscopy: Theory and Applications*, Practical spectroscopy series, vol.6 (R. G. Messerschmidt & M. A. Harthcock, Eds), *Marcel Dekker*, New York (1988), pp21-39.
- <sup>140</sup> N. P. CAMACHO, P. WEST, M. H. GRIFFITH, R. F. WARREN, C. HIDAKA. FTIR imaging spectroscopy of genetically modified bovine chondrocytes. *Materials Science and Engineering : C*, 17 (2001), pp3-9.
- <sup>141</sup> J. J. ANDREW, M. A. BROWNE, I. E. CLARK. Raman Imaging of emulsion systems. *Applied Spectroscopy*, 52(6) (1998), pp790-796.
- <sup>142</sup> J. J. ANDREW, T. M. HANCEWICZ. Rapid analysis of Raman image data using two-way Multivariate Curve Resolution. *Applied Spectroscopy*, 52(6) (1998), pp797-807.
- <sup>143</sup> S. ŠAŠIĆ, Y. OZAKI, A. OLINGA, H. W. SIESLE. Comparison of various chemometric evaluation approaches for on-line FT-NIR transmission and FT-MIR/ATR spectroscopic data of methyl methacrylate solution polymerization. *Analytica Chimica Acta*, 452(2) (2002), pp265-276.
- <sup>144</sup> P. COURCOUX, M.-F. DEVAUX, B. BOUCHET. Simultaneous decomposition of multivariate images using three-way data analysis : application to the comparison of cereal grains by confocal laser scanning microscopy. *Chemom. Intell. Lab. Syst.*, 62 (2002), pp103-113.
- <sup>145</sup> F. CLARKE. Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy. *Vibrational Spectroscopy*, 34(1) (2004), pp25-35.
- <sup>146</sup> Michel KERSZBERG. Noise, delays, robustness, canalization and all that. *Current Opinion in Genetics & Development*, 14(4) (2004), pp440-445.
- <sup>147</sup> X. P. HU, M. PATEL, K. UGURBIL. A New Strategy for Spectroscopic Imaging. *Journal of Magnetic Resonance*, B103(1) (1994), pp30-38.
- <sup>148</sup> C. D. TRAN. Visualising chemical composition and reaction kinetics by the near infrared multispectral imaging technique. *Journal of Near Infrared Spectroscopy*, 8 (2000), pp87-99.

- 
- <sup>149</sup> C. RIDGWAY, J. CHAMBERS. Detection of insects inside wheat kernels by NIR imaging. *Journal of Near Infrared Spectroscopy*, 6 (1998), pp115-119.
- <sup>150</sup> Y. K. KWON, R. K. CHO. Identification of rice variety using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 6 (1998), pp67-73.
- <sup>151</sup> W. WINDIG, B. ANTALEK, J. L. LIPPERT, Y. BATONNEAU, C. BRÉMARD. Combined use of conventional and second-derivative Data in the SIMPLISMA Self-Modeling Mixture Analysis Approach. *Analytical Chemistry*, 74(6) (2002), pp.1371-1379.
- <sup>152</sup> W. WINDIG. Spectral data files for self-modeling curve resolution with examples using the Simplisma approach. *Chemometrics and intelligent Laboratory Systems*, 36 (1997), pp3-16.
- <sup>153</sup> V. RAGHAVENDRA REDDY, A. GUPTA, T. GOVERDHAN REDDY, P. YADAGIRI REDDY, K. RAMA REDDY. Correlation techniques for the improvement of signal-to-noise ratio in measurements with stochastic processes. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 501(1-2) (2003), pp559-575.
- <sup>154</sup> P. BARONE. On the distribution of poles of Padé approximants to the Z-transform of complex Gaussian white noise. *Journal of Approximation Theory*, 132(2) (2005), pp224-240.
- <sup>155</sup> M. MELOUN, J. ČAPEK, P. MIKŠIK, R. G. BRERETON. Critical comparison of methods predicting the number of components in spectroscopic data. *Analytica Chimica Acta*, 423 (2000), pp51-68.
- <sup>156</sup> S. I. GASS, T. RAPCSÁK. Singular value decomposition in HAP. *European journal of Operational Research*, 154 (2004), pp573-584.
- <sup>157</sup> T. POLITI, A. PUGLIESE. Numerical methods for computing SVD in the  $D$ -orthogonal group. *Future Generation Computer Systems*, 22(4) (2006), pp423-429.
- <sup>158</sup> L. DUPONCHEL, W. ELMI-RAYALEH, C. RUCKEBUSH, J. P. HUVENNE. Multivariate Curve Resolution Methods in Imaging spectroscopy: Influence of Extraction Methods and Instrumental Perturbations. *Journal of Chemical Information and Computer Science*, 43 (2003), pp2057-2067.
- <sup>159</sup> Y. DU, J. JIANG, Y. LIANG, T. AMARI, Y. OZAKI. Chemometric resolution of ATR-IR spectra for polycondensation reaction of bis(hydroxyethylterephthalate) with a combination of self-modeling curve resolution (SMCR) and local rank analysis. *Analyst*, 128 (2003), pp1320-1325.
- <sup>160</sup> H. R. KELLER, D. L. MASSART. Peak purity control in liquid chromatography with photodiode-array detection by a fixed size moving window evolving factor analysis. *Analytica Chimica Acta*, 246(2) (1991), pp379-390.
- <sup>161</sup> G. CHEN, P. de B. HARRINGTON. SIMPLISMA applied to two-dimensional wavelet compressed ion mobility spectrometry data. *Analytica Chimica Acta*, 484(1) (2003), pp75-91.
- <sup>162</sup> F. C. SÁNCHEZ, D. L. MASSART. Application of SIMPLISMA for the assessment of peak purity in liquid chromatography with diode array detection. *Analytica Chimica Acta*, 298(3) (1994), pp331-339.

- <sup>163</sup> P. J. RAUCH, P. de B. HARRINTON, D. M. DAVIS. Near real-time self-modeling mixture analysis. *Chemom. Intell. Lab. Syst.*, 39(2) (1997), pp175-185.
- <sup>164</sup> W. WINDIG, C. E. HECKLER, F. A. AGBLEVOR, R. J. EVAN. Self-modeling mixture analysis of categorized pyrolysis mass spectral data with the SIMPLISMA approach. *Chemometrics and Intelligent Laboratory Systems*, 14(1-3) (1992), pp195-207.
- <sup>165</sup> A. BOGOMOLOV, M. MCBRIEN. Mutual peak matching in a series of HPLC-DAD mixture analyses. *Analytica Chimica Acta*, 490 (1-2) (2003), pp41-58.
- <sup>166</sup> P. V. Van ZOMEREN, H. DARWINKEL, P. M. J. COENEGRACHT, G. J. de JONG. Comparison of several curve resolution methods for drug impurity profiling using high-performance liquid chromatography with diode array detection. *Analytica Chimica Acta*, 487(2) (2003), pp155-170.
- <sup>167</sup> K. J. SCHOSTACK, E. R. MALINOWSKI. Theory of evolutionary factor analysis for resolution of quaternary-plus mixtures. *Chemom. Intell. Lab. Syst.*, 10(3) (1991), pp303-324.
- <sup>168</sup> K. De BRAEKELEER, D. L. MASSART. Evaluation of the orthogonal projection approach (OPA) and the SIMPLISMA approach on the Windig standard spectral data sets. *Chemom. Intell. Lab. Syst.*, 39(2) (1997), pp127-141.
- <sup>169</sup> S. GOURVÉNEC, D. L. MASSART, D.N. RUTLEDGE. Determination of number of components during mixture analysis using the Durban-Watson criterion in the Orthogonal Projection Approach and in the SIMPLE-to-use Interactive Self-modelling Mixture Analysis approach. *Chemom. Intell. Lab. Syst.*, 61 (2002), pp51-61.
- <sup>170</sup> A. G. FRENICH, D. P. ZAMORA, J. L. MARTINEZ VIDAL, M. MARTINEZ GALERA. Resolution (and quantitation) of mixtures with overlapped spectra by orthogonal projection approach and alternating least square. *Analytica Chimica Acta*, 449 (2001), pp143-155.
- <sup>171</sup> S. ŠAŠIĆ, Y. OZAKI, M. KLEIMANN, H.W. SIESLER. On the ambiguity of self-modeling curve resolution: orthogonal projection approach analysis of the on-line Fourier transform-Raman spectra of styrene/1,3-butadiene block-copolymerization. *Analytica Chimica Acta*, 460 (2002), pp73-83.
- <sup>172</sup> A. C. PIZZATO, E. J. G. BARROS. Dietary calcium intake among patients with urinary calculi. *Nutrition Research*, 23 (2003), pp1651-1660.
- <sup>173</sup> J. SHUTER, B. FINLAYSON, R. L. SCHEAFFER, R. SIERAKOWSKI, J. ZOLTEK, S. DZEGEDE. Primary liquid intake and urinary stone disease. *Journal of Chronic Diseases*, 38 (11) (1985), pp907-914.
- <sup>174</sup> K. -H. BICHLER, E. EIPPER, K. NABER, V. BRAUN, R. ZIMMERMANN, S. LAHME. Urinary infection stones. *International Journal of Antimicrobial Agents*, 19 (6) (2002), pp488-498.
- <sup>175</sup> L. LARSSON, B. SÖRBO, H.-G. TISELIUS, S. ÖHMAN. A method for quantitative wet chemical analysis of urinary calculi. *Clinica Chimica Acta*, 140(1) (1984), pp9-20.
- <sup>176</sup> J.-F. SABOT, C.-E. BORNET, S. FAVRE, S. SABOT-GUERIAUX. The analysis of peculiar urinary (and other) calculi : an endless source of challenge. *Clinica Chimica Acta*, 283 (1999), pp151-158.

- <sup>177</sup> J. KALOUSTIAN, T. F. EL-MOSELHY, H. PORTUGAL. Determination of calcium oxalate (mono- and dehydrate) in mixtures with magnesium ammonium phosphate or uric acid: the use of simultaneous thermal analysis in urinary calculi. *Clinica Chimica Acta*, 334 (2003), pp117-129.
- <sup>178</sup> K. SAFRANOW, Z. MACHOY, K. CIECHANOWSKI. Analysis of Purines in Urinary calculi by High-Performance Liquid Chromatography. *Analytical biochemistry*, 286(2) (2000), pp224-230.
- <sup>179</sup> R. S. PUROHIT, M. L. STOLLER. Stone clustering of patients with cystine urinary stone formation. *Urology*, 63(4) (2004), pp630-634.
- <sup>180</sup> S. Y. NAKADA, D. G. HOFF, S. ATTAI, D. HEISEY, D. BLANKENBAKER, M. POZNIAK. Determination of stone composition by noncontrast spiral computed tomography in the clinical setting. *Urology*, 55(6) (2000), pp816-819.
- <sup>181</sup> G. MOTLEY, N. DALRYMPLE, C. KEESLING, J. FISCHER, W. HARMON. Hounsfield unit density in the determination of urinary stone composition. *Urology* 58(2) (2001), pp170-173.
- <sup>182</sup> [on-line: <http://www.ftir.cz/kidlan2.pdf> ; F. KESNER, I. DOMINAK. Kidney stone analysis by Nicolet FTIR spectrometer] visité: le 15/06/2005.
- <sup>183</sup> I. KUZMANOVSKI, M. TRPKOYSKA, B. SOPTRAJANOV, V. STEFOV. Determination of the composition of human urinary calculi composed of whewellite, weddellite and carbonate apatite using artificial neural networks. *Analytica Chimica Acta*, 491(2) (2003), pp211-218.
- <sup>184</sup> O. SVEN, S. STEFFEN, H. OLIVIER, M. ANDREAS. Evaluation of chemical composition of urinary calculi by normal X-ray-scanning. *European Urology Supplements*, 1(1) (2002), p147.
- <sup>185</sup> [On-line : <http://www.bruckeroptics.com/raman/rfs100.html> ] visité le: 13/07/2005.
- <sup>186</sup> A. FAMILI, W.-M. SHEN, R. WEBER, E. SIMOUDIS. Data preprocessing and intelligent data analysis. *Intelligent Data Analysis*, 1(1-4) (1997), pp3-23.
- <sup>187</sup> D.-S. JOO, D.-J. CHOI, H. PARK. The effects of data preprocessing in the determination of coagulant dosing rate. *Water Research*, 34(13) (2000), pp3295-3302.
- <sup>188</sup> D. BERTRAND, E. DUFOUR. La spectroscopie infrarouge et ses applications analytiques-Prétraitements des données. *Editions Tec & doc*, Londres (2000), pp351-370.
- <sup>189</sup> A. SAVITZKY, M. J. E. GOLAY. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36 (1964), pp1627-1639.
- <sup>190</sup> [on-line : [http://bbm1.ucm.es/torralba/funder/theory/savitzky\\_golay\\_filters.shtml](http://bbm1.ucm.es/torralba/funder/theory/savitzky_golay_filters.shtml) ; Savitzky-Golay filters] visité le : 30/10/2004.
- <sup>191</sup> N. QUY DAO, M. DAUDON. Infrared and Raman spectra of calculi. *Editions scientifiques et médicales Elsevier*, Paris (1997), 350p.
- <sup>192</sup> “The language of the technical computing”, produit de « The Mathworks », site web: <http://www.mathworks.com> .

- <sup>193</sup> A. K. CONLIN, E. B. MARTIN, A. J. MORRIS. Data augmentation : an alternative approach to the analysis of spectroscopic data. *Chemom. Intell. Lab. Syst.*, 44 (1998), pp161-173.
- <sup>194</sup> M. J. SÁIZ-ABAJO, B.-H. MEVIK, V.H. SEGTNAN, T. NÆS. Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data. *Analytica Chimica Acta*, 533 (2005), pp147-159.
- <sup>195</sup> C. RUCKEBUSCH, A. De JUAN, L. DUPONCHEL, J. P. HUVENNE. Matrix augmentation for breaking rank-deficiency: A case study. *Chemom. Intell. Lab. Syst.*, 80(2) (2006), pp209-214.
- <sup>196</sup> C. A. NOLDER. Conjugate harmonic functions and Clifford algebras. *Journal of Mathematical Analysis and Applications*, 302(1) ( 2005), pp137-142.
- <sup>197</sup> M. A. ABUL-EZ, D. CONSTALES. On the order of basic series representing Clifford valued functions. *Applied Mathematics and Computation*, 142(2-3) (2003), pp575-584.