

50376  
1992  
25

61026

50376  
1992  
25

N° d'ordre: 851

# THESE

présentée à

L'UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE

pour obtenir le titre de

DOCTEUR

Mention: SPECTROCHIMIE

par

Frédéric DE LENE MIROUZE

Ingénieur E.N.S.C.L

ANALYSE QUANTITATIVE EN SPECTROMETRIE MOYEN INFRAROUGE  
PAR TRANSFORMEE DE FOURIER: PERSPECTIVES D'APPLICATION  
DANS LE DOMAINE AGRO-ALIMENTAIRE



Soutenue le 24 janvier 1992 devant la Commission d'Examen

Président :	M. F. WALLART
Rapporteurs :	M. J.P. HUVENNE M. M. MEURENS
Examineurs :	M. J.C. BOULOU M. S. COMINI M. P. LEGRAND

030 35

*Ce travail a été effectué au laboratoire de Spectrochimie Infrarouge et Raman du CNRS (LP 2641) dirigé par Monsieur J. Corset, Directeur de Recherche.*

*J'exprime ma profonde gratitude à Monsieur le Professeur F. Wallart, Délégué Régional à la Recherche et à la Technologie, pour l'honneur qu'il me fait en acceptant la présidence du jury.*

*Je voudrais remercier le Professeur P. Legrand, Responsable de l'équipe IRTF du LASIR, pour la confiance qu'il m'a témoignée en m'accueillant dans son groupe et en me proposant ce sujet de recherche.*

*Monsieur J.P. Huvenne, Professeur à l'Université de Lille II, a bien voulu suivre ce travail. Qu'il soit sincèrement remercié pour m'avoir fait profiter de son expérience de la recherche et de ses conseils sans lesquels cette thèse n'aurait pu être menée à bien.*

*J'adresse mes remerciements à Monsieur M. Meurens, Chef de travaux à l'Université Catholique de Louvain-La-Neuve en Belgique, pour les nombreuses et intéressantes discussions que nous avons eues durant l'année qu'il a passée au laboratoire.*

*Je tiens à remercier Monsieur J-C. Boulou, Ingénieur d'Application à la société SADIS BRUKER, pour son soutien exemplaire et le suivi constant qu'il a manifesté en animant régulièrement des réunions de travail au laboratoire. Je remercie également l'ensemble du personnel du département infrarouge de la société pour leur accueil toujours sympathique et leur compétence technique dont ils ont su me faire profiter durant mes différents séjours à Wissembourg.*

*J'exprime ma reconnaissance à Mr S. Comini, Ingénieur de Recherche de la société Roquette Frères, pour les séries d'échantillons qu'il nous a fournies. Sa présence dans ce jury témoigne de l'intérêt qu'il porte à ce travail.*

*La liste des remerciements serait incomplète si j'omettais de citer Monsieur B. Sombret qui m'a fait profiter de sa grande compétence en spectrométrie infrarouge.*

*Je remercie également le personnel technique du LASIR et plus particulièrement Messieurs B. Papillon, R. Demol et A. Fugère pour leur patience et le dévouement dont ils ont fait preuve notamment lors de la conception du matériel indispensable à une grande partie de ce travail.*

*Je n'oublie pas non plus l'ensemble du personnel du LASIR, chercheurs ou techniciens, pour l'accueil et l'ambiance sympathiques qu'ils ont contribués à créer.*

*Ce travail a bénéficié du soutien financier de la société SADIS BRUKER notamment par une convention CIFRE de l'ANRT.*

*"Hâtez vous lentement, et sans perdre courage  
Vingt fois sur le métier remettez votre ouvrage  
Polissez-le sans cesse et le repolissez."*

Boileau, L'art poétique.

## TABLE DES MATIERES

<b>1. INTRODUCTION</b>	<b>1</b>
<b>2. BASES THEORIQUES DE L'ANALYSE QUANTITATIVE</b>	<b>4</b>
2.1 Notations utilisées	4
2.2 Abréviations utilisées	7
2.3 L'analyse quantitative: un outil de haute technologie	8
2.4 Généralités	10
2.5 Les techniques de traitements de spectres	13
2.6 Fondements théoriques des algorithmes d'analyse quantitative	19
2.6.1 Estimateurs des paramètres statistiques	19
2.6.2 Les modèles théoriques	20
2.6.3 Ecriture matricielle	21
2.6.4 Le critère des moindres carrés linéaires	22
2.6.4.1 Le théorème de Gauss-Markow et le critère des moindres carrés	22
2.6.4.2 Résolution de l'ajustement des moindres carrés	24
2.7 Application à l'analyse quantitative	27
2.7.1 La méthode CLS	27
2.7.2 La méthode ILS	28
2.7.3 Méthode de la matrice Q	30
2.7.4 Les techniques de condensation des données	30
2.7.4.1 L'analyse en composantes principales et la méthode PCR	34
2.7.4.2 La méthode PLS	38
2.8 Comparaison des méthodes CLS, ILS, PCR et PLS	41
2.9 Validation d'une analyse	43
<b>3. ACQUISITION ET TRANSFERT DE DONNEES</b>	<b>50</b>
<b>4. ETUDE CRITIQUE DE DEUX METHODES D'ECHANTILLONNAGE</b>	<b>53</b>
4.1 La réflexion diffuse	53
4.1.1 Introduction théorique	53
4.1.2 Instrumentation	56
4.1.3 Obtention des spectres infrarouges par réflexion diffuse	59
4.1.3.1 Réflexion spéculaire et non linéarité	60
4.1.3.2 Influence de la dilution	62
4.1.3.3 Nature du diffuseur	66
4.1.3.4 Influence de la granulométrie	67
4.1.3.5 Compacité et répétabilité	70
4.2 Réflexion totale atténuée	80
4.2.1 Introduction théorique	80
4.2.2 Instrumentation	83
4.2.3 Obtention de spectres infrarouges par ATR	85
4.2.3.1 Etude de la répétabilité en ATR des liquides	85

<b>5. APPLICATION A L'ANALYSE QUANTITATIVE</b>	87
5.1 Fonctionnement du logiciel de calcul	88
5.2 La réflexion diffuse	93
5.3 Analyses par ATR	100
5.4 Application a des mélanges réels	106
5.4.1 Jus de fruits	106
5.4.2 Sirops de glucose	111
<b>6. CONCLUSION GENERALE</b>	120
<b>7. BIBLIOGRAPHIE</b>	123

# **CHAPITRE I**

## **I- INTRODUCTION**

## 1) INTRODUCTION

Les besoins en analyse quantitative vont sans cesse croissants, en particulier dans les laboratoires de contrôle qualité des produits industriels. Or faute d'instruments appropriés, la spectroscopie infrarouge a semblé longtemps inadaptée à toute analyse sophistiquée. Les progrès en instrumentation optoélectronique, l'apparition de nouvelles techniques d'échantillonnage et surtout le développement de l'informatique ont permis à des pionniers tels Karl Norris de valoriser l'infrarouge comme technique d'analyse adaptée aux besoins industriels, notamment agro-alimentaires. L'ouvrage devenu classique d'Osborne<sup>(1)</sup> témoigne de cette maturité. La prolifération sur le marché de nombreux appareils commerciaux en est une autre illustration.

Une des premières applications quantitatives de l'infrarouge date de 1964 avec le spectromètre IRMA (Infra Red Milk Analyser) de Goulden<sup>(2)</sup>. L'appareil effectuait le dosage dans le moyen infrarouge (MIR) du lactose, des protéines et des matières grasses du lait liquide dans des cellules de 37  $\mu\text{m}$  d'épaisseur à fenêtres de  $\text{CaF}_2$ . Ces premiers essais ont très vite mis en évidence certaines faiblesses du MIR que ne rencontre pas le proche infrarouge (NIR):

- \* un rapport signal/bruit bas en raison de sources moins lumineuses et de détecteurs moins sensibles.
- \* les coefficients d'absorption étant plus élevés en MIR, on est obligé d'utiliser des cellules de quelques dizaines de microns avec les difficultés de construction et d'utilisation que cela implique. Par comparaison les cellules peuvent atteindre une épaisseur de plusieurs millimètres dans le NIR.
- \* peu de matériaux sont transparents en MIR, chimiquement inertes et insolubles dans l'eau.
- \* les bandes de l'eau sont très intenses dans le moyen infrarouge et posent des problèmes à la fois d'interférences et de saturation.
- \* l'échantillonnage est plus difficile (homogénéisation des globules gras).

Pour toutes ces raisons l'analyse quantitative infrarouge s'est jusqu'à présent surtout développée dans le proche infrarouge, exceptées quelques réalisations comme le dosage du lait où le MIR s'est imposé depuis 25 ans<sup>(5)</sup>. Actuellement, le proche infrarouge intervient pour les produits les plus divers<sup>(4,6)</sup> et dans des conditions parfois inaccessibles au MIR: par exemple les sociétés TREBOR et TECATOR commercialisent un analyseur grain entier capable d'assurer les besoins en dosages de l'industrie céréalière sans nécessiter le broyage de l'échantillon.

De nos jours les difficultés précédemment évoquées sont partiellement levées grâce au développement de la spectrométrie infrarouge par transformée de Fourier<sup>(5,6)</sup> (IRTF) qui offre notamment un gain important du rapport signal/bruit. L'essor de la réflexion diffuse sous l'impulsion de Griffiths<sup>(5,6)</sup> permet de simplifier considérablement l'échantillonnage des poudres tandis que l'apparition récente de l'ATR horizontal rend possible l'étude des solutions aqueuses dans de bien meilleures conditions qu'auparavant. Enfin, l'apport le plus intéressant du moyen infrarouge est une meilleure spécificité. On peut par exemple doser en MIR les insaturations trans des acides gras qui n'ont pas de bandes spécifiques dans le proche infrarouge<sup>(7)</sup>. Pendant longtemps peu de travaux en analyse quantitative ont été consacrés à la spectrométrie MIR, peut-être en raison des réticences qu'éprouvent les fondamentalistes (utilisateurs principaux de ce domaine) à l'égard de l'approche statistique et souvent empirique qu'implique l'analyse quantitative.

C'est la raison pour laquelle nous nous sommes proposés, en collaboration avec la société SADIS BRUKER, d'étudier du point de vue de l'analyse quantitative le nouveau contexte offert par les développements récents des techniques d'échantillonnage. Nous nous sommes placé essentiellement dans l'optique de l'analyse de routine en milieu industriel.

En raison de la grande diversité des besoins il est difficile d'être précis dans la liste des qualités attendues pour ce type d'application. Toutefois certains critères peuvent être dégagés:

- robustesse de la méthode vis à vis de l'expérimentateur ou des fluctuations dans les échantillons.
- fiabilité et précision des résultats.
- rapidité de la mesure.
- coût peu élevé de l'échantillonnage (réactifs...).
- échantillonnage pratique.

Les bases théoriques nécessaires à la réalisation d'analyses quantitatives complexes (cas multicomposants par exemple) sont encore relativement méconnues. Il a donc paru important de détailler cette question aux chapitres suivants.

## **CHAPITRE II**

### **II- BASES THEORIQUES DE L'ANALYSE QUANTITATIVE**

1. Notations utilisées
2. Abréviations utilisées
3. L'analyse quantitative: un outil de haute technologie
4. Généralités
5. Les techniques de traitements de spectres
6. Fondements théoriques des algorithmes d'analyse quantitative
7. Application à l'analyse quantitative
8. Comparaison des méthodes CLS, ILS, PCR et PLS
9. Validation d'une analyse

## 2) BASES THEORIQUES DE L'ANALYSE QUANTITATIVE

### 2.1) NOTATIONS UTILISEES

---

A notre connaissance, aucune normalisation officielle n'existe en chimiométrie aussi bien pour les terminologies employées que pour les notations. Celles rencontrées dans la littérature sont donc aussi variées que contradictoires, bien qu'une tentative de normalisation ait été faite par Martens et Wold<sup>(\*)</sup> en 1984.

Les notations employées dans la suite de ce mémoire, à l'exception de quelques variables locales, sont détaillées ci-après.

a) Les variables scalaires sont notées par des lettres minuscules normales.

Exemple:  $i, j, k$ .

b) Les matrices seront représentées par des lettres majuscules en caractère gras, à l'exception des vecteurs. Lorsque les dimensions de la matrice doivent être explicitées, elles seront indiquées en indice.

Exemple:  $A_{mn}$  ou  $A_m$  si  $m = n$ .

c) Les vecteurs (matrices colonnes) sont représentés par des minuscules en caractère gras.

Exemple:  $x, y, z$ .

d) La matrice transposée de  $A$  est notée  $A^t$  ( $a^t$  pour un vecteur).

e) L'opérateur  $E(x)$  désignera la moyenne de la variable  $x$ .

f) L'estimateur d'un paramètre est représenté par un accent circonflexe.

Exemple:  $s \rightarrow \hat{s}$ .

g) L'écriture des matrices (ou des vecteurs) peut être suivie par divers indices pour préciser qu'elles se réfèrent à des mélanges inconnus (indice  $u$ ), des standards de calibration (indice  $c$ ) ou de prédiction (indice  $p$ ).

Exemple:  $Y_u, E_c, E_p$

Les variables  $x$ ,  $y$ ,  $e$ ,  $b$  et  $x$ ,  $y$ ,  $e$ ,  $b$  peuvent être suivies d'un indice pour indiquer qu'elles sont extraites des matrices  $X$ ,  $Y$ ,  $E$ ,  $B$  respectivement.

#### Variables scalaires

$b$	coefficient de régression
$e$	résidu
$h$	nombre de loadings utilisés
$i$	$1 \leq i \leq n$ (indice)
$j$	$1 \leq j \leq m$ (indice)
$k$	$1 \leq k \leq p$ (indice)
$K$	coefficient d'absorption
$m$	nombre de constituants du mélange
$n$	nombre de spectres (standards, validation, inconnus)
$p$	nombre de réponses spectrales
$S$	coefficient de diffusion
$x$	concentration
$y$	réponse du système analytique (absorbance...)

#### Variables vectorielles (dimension entre parenthèses)

$b$	vecteur des coefficients de régression
$e$	vecteur des résidus
$t$	vecteur des scores (méthode PCR et PLS)
$x$	vecteur des concentrations ( $nx1$ )
$y$	vecteur des réponses ( $nx1$ )

**Variables matricielles (dimension entre parenthèses)**

<b>0</b>	matrice ou vecteur nul
<b>B</b>	matrice des coefficients de régression
<b>E</b>	matrice des résidus
<b>I</b>	matrice identité
<b>P<sup>t</sup></b>	matrice des loadings (hxp)
<b>T</b>	matrice des scores (nxh)
<b>X</b>	matrice des concentrations (nxm)
<b>Y</b>	matrice des données (npx)

## 2.2) ABBREVIATIONS UTILISEES

---

ATR	Attenuated Total Reflectance (→ <i>réflexion totale atténuée</i> )
CLS	Classical Least Squares (→ <i>méthode des moindres carrés ordinaires</i> )
DO	Densité Optique
ILS	Inverse Least Square (→ <i>méthode des moindres carrés inverses</i> )
IRE	Internal Reflexion Element (→ <i>milieu de réflexion interne</i> ).
IRS	Internal Reflexion Spectroscopy (→ <i>spectroscopie de réflexion interne</i> )
IRTF	InfraRouge à Transformée de Fourier (→ <i>FTIR</i> )
KM	Kubelka-Munk (unité de )
MIR	Moyen InfraRouge (→ <i>Mid InfraRed, MIR</i> )
MLR ou MLS	Multiple Linear Regression (→ <i>régression multilinéaire</i> )
MSC	Multiple Scatter Correction (→ <i>correction de l'écart à la régression</i> )
NIPALS	Nonlinear Iterative Partial Least Squares
NIR	Near InfraRed (→ <i>Proche InfraRouge, PIR</i> )
PCA	Principal Component Analysis (→ <i>analyse en composantes principales</i> )
PCR	Principal Component Regression (→ <i>régression en composantes principales</i> )
PLS	Partial Least Squares (→ <i>moindres carrés partiels</i> )
SEC	Standard Error of Calibration (→ <i>erreur standard de calibration</i> )
SEE	Standard Error of Estimate (→ <i>erreur standard d'estimation</i> )
SEP	Standard Error of Prediction (→ <i>erreur standard de prédiction</i> )
SMLR	Stepwise Multiple Linear Regression (→ <i>régression multilinéaire pas à pas</i> )

### 2.3) L'ANALYSE QUANTITATIVE: UN OUTIL DE HAUTE TECHNOLOGIE

---

Les progrès de l'analyse quantitative par les techniques spectroscopiques ont en grande partie accompagnés ceux des techniques de traitement des données.

La solution, apportée par les premiers travaux au problème de l'analyse quantitative en infrarouge, fut de repérer sur le spectre une longueur d'onde supposée caractéristique du constituant à analyser. Le tracé des variations d'absorbance du mélange à cette longueur d'onde en fonction de la concentration donne la courbe d'étalonnage du mélange. La loi de Beer-Lambert postule que cette courbe présente l'allure d'une droite.

Le développement de l'électronique a permis de perfectionner cette solution par l'emploi simultané de deux ou trois longueurs d'onde. On peut ainsi résoudre plus efficacement le problème des mélanges à plusieurs constituants. Le problème du choix des longueurs d'onde adéquates est développé dans le livre d'Osborne<sup>(1)</sup>. Toutefois ces méthodes rencontrent souvent des difficultés lorsque les différences entre spectres voisins sont faibles, ce qui est souvent le cas.

Le coût de l'informatique diminuant, on a pu envisager d'utiliser l'ensemble du spectre disponible ( $\rightarrow$  *full spectrum method*) plutôt qu'une sélection de quelques longueurs d'onde. Cette approche donne de meilleures performances et est aussi mieux adaptée au moyen infrarouge car elle permet d'exploiter pleinement la grande spécificité des spectres obtenus.

Les techniques de sélection de fréquences ont très vite évoluées avec les progrès de l'informatique vers des méthodes dites de condensation des données qui permettent des sélections plus spécifiques de l'information non plus en terme de fréquences, mais en terme d'informations utiles: l'idée est d'essayer de séparer les informations chimiques d'artefacts tels le bruit du spectre, la ligne de base... Les algorithmes les plus utilisés s'appellent méthode PCR ou méthode PLS.

Plus récemment la chimie analytique s'est enrichie d'une nouvelle approche qui consiste à essayer de relier une propriété expérimentale (un spectre infrarouge par exemple) à des

qualités plus subjectives du produit<sup>(9)</sup>. On peut ainsi évaluer le pouvoir de panification de différentes variétés de blé<sup>(10)</sup> ou classer par catégories des laits écrémés<sup>(11)</sup> selon le traitement thermique qu'ils ont subi. On parle d'analyse factorielle discriminante (→ *Factorial Discriminante Analysis, FDA*) ou encore d'analyse discriminante canonique.

D'autres développements encore plus sophistiqués commencent à apparaître avec l'utilisation des techniques d'intelligence artificielle: systèmes experts<sup>(12)</sup> ou encore théorie des réseaux neuronaux<sup>(13,14)</sup> (→ *neural networks*) qui a déjà permis à Long<sup>(15)</sup> et à Bos<sup>(16)</sup> de présenter des applications en analyse quantitative.

Contrairement aux méthodes d'analyse habituelles, les réseaux de neurones ne reconnaissent pas un spectre parce qu'il suit un modèle imposé *a priori* (loi de Beer-Lambert par exemple), mais en simulant ce que nous appelons "l'impression de déjà vu". De ce fait ils sont potentiellement moins sensibles au bruit ou aux fluctuations qui affectent les produits analysés comme par exemple la provenance de différents fournisseurs.

Ces possibilités nouvelles tant qualitatives que quantitatives sont le fruit de l'arrivée massive de l'informatique dans les laboratoires et des efforts conjoints des statisticiens, des informaticiens et des chimistes pour utiliser les données expérimentales de façon plus intensives.

Ce mode de pensée nouveau rend nécessaire une terminologie spécifique intégrant ces technologies récentes: on parle de chimiométrie (→ *chemometrics*). La chimiométrie est l'ensemble des techniques mathématiques de traitement de spectres destinées à améliorer leur interprétation qualitative ou quantitative et de réaliser cette interprétation.

## 2.4) GENERALITES

---

Une analyse quantitative s'effectue en trois étapes clefs:

- a) la calibration ou étalonnage
- b) la validation
- c) la prédiction

Cette dernière opération est le coeur de l'analyse car c'est maintenant que l'on va calculer la concentration inconnue  $C$  étant donnée la réponse  $R$  du système (l'absorbance à une longueur d'onde par exemple).

Auparavant il est indispensable d'établir la loi de calibration du système, c'est à dire de savoir calculer  $C$  connaissant  $R$ .

Le modèle théorique dépend d'un certain nombre de paramètres que l'étape de calibration va s'attacher à calculer. Les algorithmes de calcul sont nombreux, mais dans tous les cas il est nécessaire de disposer d'un ensemble de spectres témoins déjà étalonnés à l'aide d'une méthode de référence supposée exacte. Selon les auteurs ces spectres sont appelés références, standards ou spectres de calibration ( $\rightarrow$  *training set*). Le choix de ces standards n'est pas indifférent:

- \* l'étendue de la gamme de concentration des standards de calibration doit recouper celle rencontrée en phase de prédiction.
- \* les standards doivent être "physiquement" représentatifs des solutions rencontrées lors de la prédiction: on ne peut doser les constituants d'un jus de pomme en étalonnant avec un jus d'orange.
- \* on doit en outre éviter les situations où deux constituants se retrouvent en proportion constante l'un par rapport à l'autre dans chaque standard. Dans ce cas de figure, le programme quantitatif s'avèrerait incapable de les différencier et considérerait qu'il s'agit du même produit.

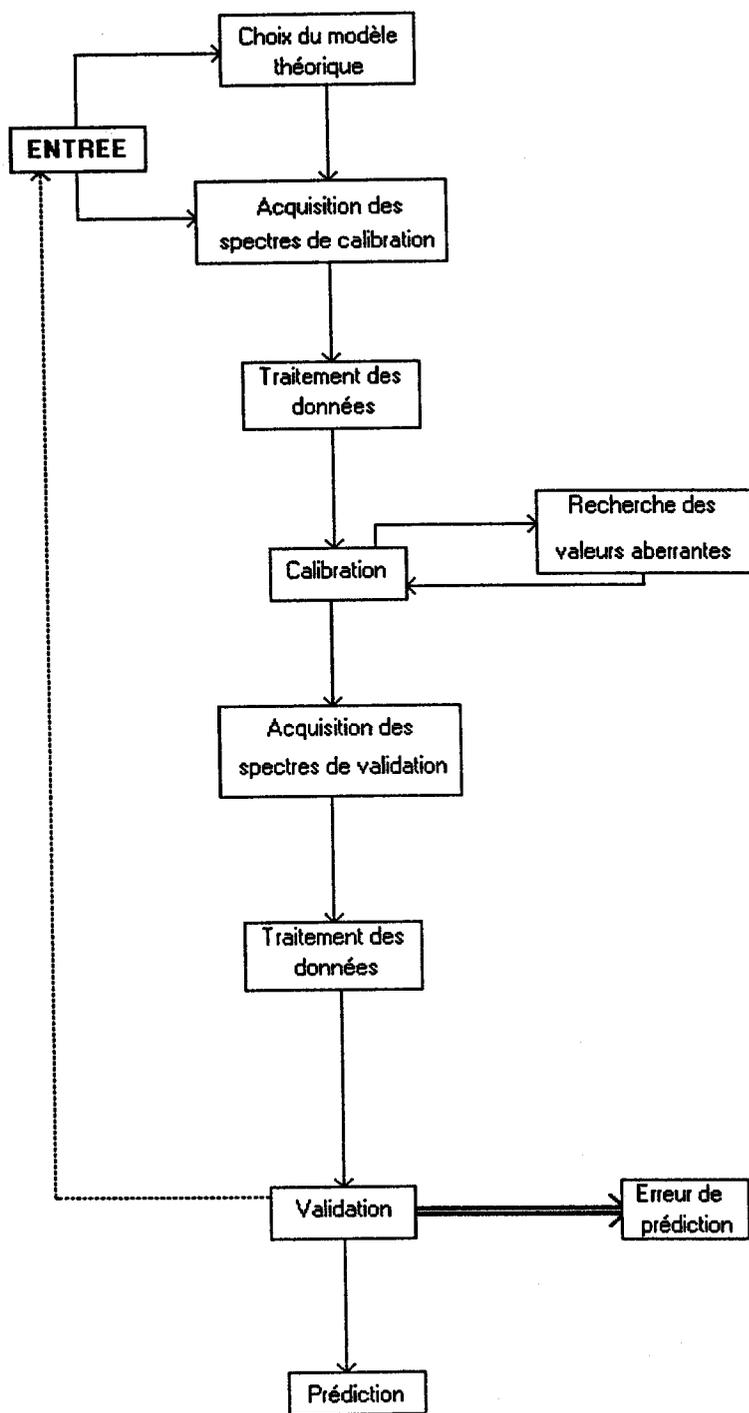
Les paramètres du modèle théorique ne peuvent être connus de façon exacte. On doit donc se contenter d'en trouver un estimateur. Dans le langage de la statistique on dit que l'on réalise un ajustement ( $\rightarrow$  *fitting*).

Puisque que seule une estimation de la loi de calibration réelle n'est accessible à l'expérience, il est nécessaire d'en tester sa fiabilité avant usage. Cette opération s'appelle la validation du système. Pour cela on va former un autre jeu de spectres totalemt indépendants des précédents. Le mot indépendant signifie que l'on ne peut tester une calibration à l'aide des spectres qui ont servi de standard (validation interne). Les spectres de validation sont eux aussi supposés parfaitement connus à l'aide de la méthode de référence. On pourra donc comparer les valeurs vraies aux valeurs calculées par la loi de calibration trouvée. On parle de validation externe. Il est important de signaler que les résultats de la calibration ne donnent *aucune* indication quand à la valeur prédictive du modèle ainsi construit.

La calibration est généralement compliquée par la présence de valeurs anormales ( $\rightarrow$  *outliers*). D'après la classification introduite par Beckman<sup>(17)</sup> une valeur peut être anormale pour deux raisons: elle s'écarte notablement de la courbe de régression ( $\rightarrow$  *discordant observation*), elle est située sur la courbe de régression, mais pas à la bonne place ( $\rightarrow$  *contaminant*). Conserver ou non la valeur anormale est un choix de l'utilisateur, pas un problème statistique. Il se peut qu'une valeur anormale ne doive rien à un accident de calibration, mais soit le reflet d'une variabilité mal prise en compte. Dans ce cas il sera peut être nécessaire de réviser la calibration ou même la méthode d'échantillonnage.

Les valeurs expérimentales brutes ne conviennent pas toujours directement pour l'analyse quantitative et nécessitent souvent un traitement préliminaire (lissage, linéarisation...). Ces opérations sont rassemblées sous le nom de traitement des données et sont la source d'une inépuisable littérature. Elles ne doivent pas être négligées car elles s'avèrent souvent décisives,

toutefois la qualité des spectres est bien plus importante: *aucun traitement de données ne permettra de trouver de l'information là où il n'y en a pas.*



**Figure 1:** Les principales étapes du développement d'une analyse quantitative.

## 2.5) LES TECHNIQUES DE TRAITEMENTS DE SPECTRES

---

De nombreux phénomènes contribuent à diminuer l'efficacité des algorithmes d'analyse quantitative. Les principaux sont:

- \* non linéarité des spectres
- \* problème de la ligne de base
- \* le problème du bruit
- \* la propagation des erreurs numériques
- \* le problème de collinéarité

Un des aspects importants d'une analyse quantitative consiste à tenter de corriger ces effets par un traitement adéquat en général préalable à la régression.

### **\* non linéarité des spectres**

Les spectres sont non linéaires lorsque la réponse expérimentale n'est pas proportionnelle à la concentration. Les raisons peuvent être: la non linéarité du détecteur, le choix de concentrations trop fortes, interférences chimiques... Pour tenter de corriger ces effets les stratégies classiques sont:

- linéariser par des transformations comme Kubelka-Munk, correction MSC,...
- dériver les spectres
- diluer les mélanges
- découper la population des échantillons en sous-ensembles balayant une gamme de concentration plus restreinte, on effectuera une calibration pour chaque sous-ensemble.

L'effet de la transformation de Kubelka-Munk et de la correction MSC sera abordé au chapitre IV relatif à la réflexion diffuse.

La dilution est souvent indispensable pour se placer dans un domaine de linéarité de la réponse spectrale, mais en diluant on diminue également le rapport signal sur bruit ce qui peut entraîner d'autres complications.

On a souvent tenté d'introduire des termes croisés ou quadratiques dans le modèle de régression<sup>(18,19)</sup> pour tenir compte des interférences entre constituants. Cette approche pose de nombreux problèmes théoriques et semble peu utilisée actuellement. Constatant que corriger un problème de linéarité n'est pas toujours facile, Alfaro<sup>(20)</sup> suggère dans sa thèse d'utiliser directement des modèles non linéaires au lieu d'imposer au système une linéarité complète et peut-être artificielle.

Une autre solution, plus originale, est proposée notamment par Long<sup>(15)</sup> avec l'emploi de réseaux neuronaux qui semblent moins sensibles au problème de linéarité et de bruit.

#### **\* le problème de la ligne de base**

La méthode habituelle de traitement consiste à soustraire au spectre un ou plusieurs segments de droite sensés représenter la ligne de base. Les algorithmes de recherche de ces segments ont fait l'objet de nombreuses publications. Malgré cela aucune technique n'est à notre connaissance suffisamment précise pour assurer une utilisation correcte en analyse quantitative.

Il est toutefois possible d'éliminer une ligne de base non curviligne en dérivant deux fois les spectres. En contrepartie on accentue l'effet du bruit sur le spectre. Cet inconvénient peut être atténué en effectuant la dérivation dans l'espace de Fourier. On peut alors contrôler le rapport signal/bruit par une fonction d'apodisation et gagner ainsi pratiquement un ordre de grandeur par rapport aux calculs réalisés dans l'espace réel. Notre expérience suggère toutefois que cette correction est rarement décisive dans le MIR. Par contre elle est fréquemment employée dans le proche infrarouge en raison sans doute d'un rapport signal/bruit plus favorable.

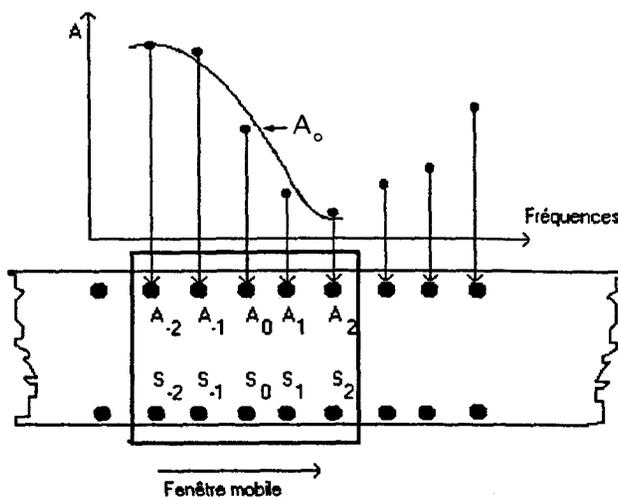
### \* le problème du bruit

La chaîne analytique est à chaque instant soumise à des perturbations extérieures dues à l'environnement du laboratoire (température, pression...) et aux contraintes physiques des instruments de mesure (stabilité de l'alimentation électrique, réglage optique...). L'ensemble de ces perturbations contribue au bruit du spectre.

Si l'on ne peut agir sur la chaîne de mesure, il est possible de réaliser des accumulations de spectres. On montre alors que le bruit diminue comme la racine carrée du nombre d'acquisitions.

On peut également diminuer le bruit sur les spectres par un lissage. Un des algorithmes de lissage les plus simple est celui dit de la moyenne mobile: on remplace chaque point par la moyenne des ordonnées des points qui l'entourent. Cette opération peut être décrite comme une convolution avec une fonction de convolution rectangulaire.

On peut également effectuer une moyenne pondérée, ce qui équivaut à modifier la forme de la fonction de convolution.



**Figure 2:** Illustration de la méthode de Savitzky et Golay pour 5 points de convolution.

La réponse  $A_0$  est corrigée en la remplaçant par la moyenne - pondérée par les  $S_i$  - des éléments spectraux repérés par la fenêtre mobile.

Une approche plus précise consiste à interpoler *localement* la courbe par une droite ou un polynôme selon une procédure des moindres carrés. Savitzky et Golay<sup>(21,22)</sup> ont montré que cette opération peut se ramener à une convolution, quel que soit le polynôme d'interpolation. Ce résultat rend cette approche tout à fait attractive pour une utilisation pratique. D'autres algorithmes existent. Dans une série d'articles<sup>(23)</sup>, Maddams en présentent quelques uns et compare leurs performances avec celles de l'algorithme de Savitzky et Golay pour l'amélioration de la résolution des bandes et l'analyse quantitative.

**\* la propagation des erreurs numériques**

Considérons le système d'équations  $A \cdot x = b$  dans lequel:

$$A = \begin{pmatrix} 3.56 & 1.92 \\ -1.92 & 2.44 \end{pmatrix} \quad b^t = (-20.0 \ 15.0)$$

la solution est  $x^t = (-4.0 \ 3.0)$ .

Supposons que  $b$  soit la cible d'une perturbation qui engendre une erreur relative de 2%,  $b^t$  devient  $(-19.6 \ 15.4)$ . Le nouveau vecteur solution est donc:  $(-3.7 \ 3.4)^t$ . On constate que la perturbation initiale de 2% conduit à une erreur finale de 10%. Il y a donc un effet d'amplification de l'erreur. Le facteur d'amplification dépend des matrices qui interviennent ainsi que de la méthode de calcul utilisée. Un algorithme est dit robuste ou bien conditionné, s'il est peu sensible aux incertitudes sur les données initiales. On montre<sup>(29)</sup> que la robustesse d'un algorithme de résolution d'un problème de type  $y = A \cdot x$  est liée à la valeur de  $\kappa_A = \|A\| \cdot \|A^{-1}\|$  que l'on appelle conditionnement de la matrice  $A$  ( $\rightarrow$  *condition number*). On démontre que quelle que soit  $A$ ,  $\kappa_A \geq 1$ . La résolution par la méthode des moindres carrés donne un facteur d'amplification des erreurs numériques dans lequel intervient  $\kappa_A^2$ .

Pour résoudre au mieux le problème des moindres carrés on devra par conséquent:

- \* essayer de réduire la taille des matrices (propagation des erreurs numériques).
- \* diminuer  $\kappa_A$  en prémultipliant les deux membres de l'équation par une matrice adéquate. Voir à ce sujet Jochum et al.<sup>(24)</sup>, ainsi que Brezinski<sup>(25)</sup>. Actuellement seules des solutions empiriques sont connues à ce problème.
- \* changer l'algorithme de résolution. Par exemple la résolution du problème des moindres carrés linéaires selon l'algorithme de Householder conduit à un facteur d'amplification dans lequel n'intervient plus que  $\kappa_A$ .
- \* Remplacer A par une matrice orthogonale dont le conditionnement vaut 1 (analyse factorielle, méthode PLS, algorithme de Householder, méthode de Jacobi...).

**\* le problème de collinéarité**

Les méthodes matricielles nécessitent l'inversion d'une ou plusieurs matrices. La question qui se pose est de savoir si elles sont inversibles, ce qui équivaut à se demander si les vecteurs constituant la matrice sont linéairement indépendants, c'est à dire non corrélés.

D'autre part les ordinateurs ne peuvent travailler sur les représentations exactes des valeurs qui interviennent dans les calculs: les résultats numériques sont limités par les erreurs d'arrondi et le mode de représentation binaire des données. Dans ce contexte l'expression "linéairement indépendant" doit donc s'étendre au cas de vecteurs numériquement très proches. Les matrices quasi-singulières ainsi créées ne seront inversées par le logiciel qu'au prix d'importantes erreurs numériques. C'est en cela que réside le problème de collinéarité<sup>(26,27)</sup> ( $\rightarrow$  *collinearity*).

On ne peut le résoudre qu'à l'aide d'algorithmes robustes comme par exemple l'algorithme de Gram-Schmidt<sup>(28)</sup>, la méthode de Householder (Voir Lascaux<sup>(29)</sup> et Sibony<sup>(30)</sup> pour une description détaillée ainsi qu'un organigramme) et surtout les techniques de condensation des données.

La corrélation qui existe entre les constituants du mélange est une des sources des problèmes de collinéarité et de non linéarité des spectres. Lindberg et *al.*<sup>(40)</sup> expliquent que chaque constituant réémettant sur toute la gamme spectrale, la lumière infrarouge reçue par chaque fraction du mélange n'est pas entièrement le fait de la source. Ainsi une variation de la concentration de l'un des constituants exerce une influence sur le spectre de tous les autres et donc sur la concentration calculée. On dit que les constituants sont corrélés.

## 2.6) FONDEMENTS THEORIQUES DES ALGORITHMES D'ANALYSE QUANTITATIVE

---

### 2.6.1) ESTIMATEURS DES PARAMETRES STATISTIQUES

On considère une variable aléatoire  $X$  dont la moyenne (ou espérance) est  $\mu$  et l'écart type (ou déviation standard) est  $\sigma$  ( $\rightarrow$  *standard déviation*). La variance de  $X$  est  $\sigma^2$  par définition. On dispose en outre de  $n$  réalisations  $X_1, X_2, \dots, X_n$  de  $X$ .

#### \* le biais

Un estimateur  $\hat{\mu}$  de  $\mu$  est dit non biaisé si et seulement si la moyenne de ses réalisations est égale à  $\mu$ , c'est à dire:  $E(\hat{\mu}) = \mu$ . On définit le biais par:

$$\Delta = E(\hat{\mu}) - \mu \quad (\text{Eq 1})$$

En instrumentation on parlera d'erreur systématique.

#### \* la moyenne

On pose:  $E(X) = \sum X_i/n$ . Si les  $n$  réalisations de  $X$  représentent l'ensemble de la population, alors  $E(X) = \mu$ , sinon  $E(X)$  est un estimateur *non biaisé* de  $\mu$ .

#### \* la variance

Si les  $n$  réalisations précédentes de  $X$  représentent l'ensemble de la population statistique, alors par définition:

$$\sigma^2 = \sum (X_i - \mu)^2 / n \quad (\text{Eq 2})$$

Dans le cas où l'on ne dispose que d'un échantillonnage, on ne peut qu'obtenir un estimateur. L'estimateur non biaisé de  $\sigma^2$  est:

$$\hat{\sigma}^2 = \sum (X_i - \hat{\mu})^2 / (n-1) \quad (\text{Eq 3})$$

Par contre l'estimateur de l'écart type  $\sigma$  calculé à partir de (Eq 3) n'est pas absolument correct. On peut montrer que les valeurs obtenues sont systématiquement trop faibles<sup>(31)</sup>, l'erreur relative étant approximativement égale à  $1/4(n-1)$ .

## 2.6.2) LES MODELES THEORIQUES

De nombreuses lois de calibration peuvent être imaginées. La plus utilisée est le modèle linéaire pour lequel la réponse du système est proportionnelle à la concentration ou à une fonction non paramétrée de la concentration. Ce modèle est le plus simple possible et il est le plus souvent suffisant. En outre il est statistiquement bien compris et est le seul qui soit traité par tous les programmes quantitatifs.

Le modèle linéaire de Beer-Lambert servira de support dans les chapitres suivants pour définir les notions de base relatives au fonctionnement des logiciels quantitatifs. La loi de Beer-Lambert dit que pour une fréquence donnée  $\nu$ , l'absorbance  $y$  mesurée est proportionnelle à la concentration molaire  $x$  de la solution:

$$y = \epsilon x l = K l \quad (\text{Eq 4})$$

$\epsilon$  est le coefficient d'absorption molaire à la fréquence  $\nu$  du constituant qui absorbe et  $l$  le chemin optique.  $K = \epsilon x$  est le coefficient d'absorption. On rajoute souvent une constante additive  $\epsilon_0$  qui constitue le blanc et exprime que même pour une concentration nulle on observera un signal résiduel. On parle alors de modèle additif. Du point de vue de la description mathématique du modèle, on interprète  $\epsilon_0$  comme le coefficient d'absorption d'un constituant fictif de concentration 1.

Dans le modèle décrit par (Eq 4), il n'y a qu'une seule variable: c'est  $x$ . On parlera alors de modèle linéaire simple. Si la solution étudiée contient  $m$  constituants différents, alors l'absorbance totale  $y$  (mesurée par le détecteur) s'écrit:

$$y = \sum_{j=1}^m \epsilon_j x_j \quad 1 \leq j \leq m \quad (\text{Eq 5})$$

Cette fois il y a  $m$  variables, il s'agit du modèle de régression multilinéaire ( $\rightarrow$  *MLR* ou *MLS*).

Le modèle théorique ainsi construit n'est jamais rigoureusement vérifié par les données expérimentales. On est donc conduit à introduire une erreur  $e$  qui est égale à la part de la valeur expérimentale non expliquée par le modèle:

$$e = y_{\text{exp}} - y_{\text{calc}} = y - \sum_{j=1}^m \epsilon_j x_j \quad (\text{Eq 6})$$

Soit encore:

$$y = \sum l\epsilon_j x_j + e \quad (\text{Eq 7})$$

La variable  $e$  est appelée résidu ( $\rightarrow$  *residual*). Les paramètres  $\epsilon_j$  sont les coefficients de régression, parfois appelés sensibilités ( $\rightarrow$  *sensitivity*).

### 2.6.3) ECRITURE MATRICIELLE

L'équation (Eq 7) traduit une relation entre la réponse  $y$  du détecteur et la concentration  $x$ . Pour qu'elle soit utilisable, on doit pouvoir calculer les  $m$  coefficients inconnus. C'est l'étape de calibration. Cela n'est possible qu'en écrivant au moins  $m$  équations. Ces équations sont obtenues en réécrivant (Eq 7) pour chacun des  $n$  standards de calibration disponibles.

On obtient le système d'équations suivant:

$$\left\{ \begin{array}{l} y_1 = l\epsilon_1 x_{11} + l\epsilon_2 x_{12} + \dots + l\epsilon_m x_{1m} + e_1 \\ y_2 = l\epsilon_1 x_{21} + l\epsilon_2 x_{22} + \dots + l\epsilon_m x_{2m} + e_2 \\ \cdot \\ \cdot \\ y_n = l\epsilon_1 x_{n1} + l\epsilon_2 x_{n2} + \dots + l\epsilon_m x_{nm} + e_n \end{array} \right. \quad (\text{Eq 8})$$

où  $x_{ij}$  est la concentration du constituant  $j$  dans la solution  $i$ .

Ce système d'équations peut s'écrire de façon plus compacte:

$$y = X \cdot b + e \quad (\text{Eq 9})$$

Avec:

$$y_{n1} = (y_1 \ y_2 \ \dots \ y_n \ \epsilon)^t$$

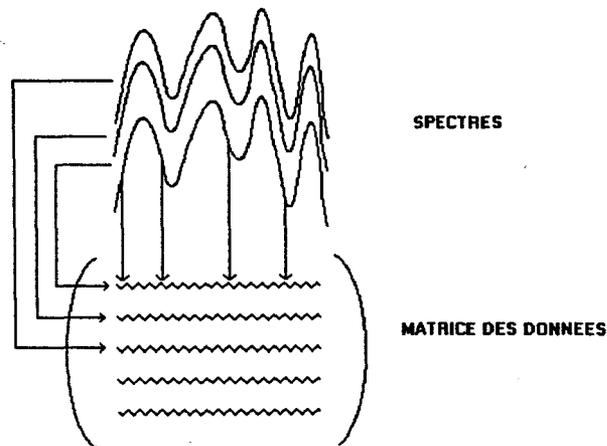
$$b_{m1} = (l\epsilon_1 \ l\epsilon_2 \ \dots \ l\epsilon_m)^t$$

$$e = (e_1 \ e_2 \ \dots \ e_n)^t$$

$$X_{nm} = (x_{ij}) \ 1 \leq i \leq n, \ 1 \leq j \leq m$$

Il est possible de généraliser (Eq 9) à chacune des  $p$  fréquences auxquelles on échantillonne les spectres. Pour cela il suffit d'écrire  $p$  systèmes d'équations analogues à (Eq 9) que l'on rassemble sous la forme matricielle:

$$Y = X \cdot B + E \quad (\text{Eq 10})$$



**Figure 3:** Construction de la matrice Y

## **2.6.4) LE CRITERE DES MOINDRES CARRES LINEAIRES**

### **2.6.4.1) LE THEOREME DE GAUSS-MARKOW ET LE CRITERE DES MOINDRES CARRES**

La résolution de (Eq 10) conduit à envisager trois cas de figure:

#### **\* $m > n$**

Il y a plus d'inconnues que d'équations, le système admet une infinité de solutions toutes différentes.

#### **\* $m = n$**

Le système admet soit une solution unique et alors  $E=0$ , soit aucune solution. Ce cas est rare en pratique et conduit à des solutions instables (peu robustes) puisque les perturbations dues à des valeurs anormales ne peuvent être compensées.

\*  $m < n$

Cette fois, on obtient autant de solutions exactes et différentes que de façon de choisir  $m$  équations parmi  $n$  autres. Il n'est donc pas possible de trouver une solution unique les vérifiant toutes simultanément. On doit se contenter d'une solution intermédiaire qui réalisera "approximativement" chacune des  $n$  équations et alors nécessairement  $E < > 0$ . Il existe une infinité de solutions réalisant ce programme, chacune d'elles est appelée un ajustement du modèle. On s'intéresse surtout à celles qui minimisent l'ensemble des résidus. Le critère de minimisation le plus utilisé est le critère des moindres carrés ( $\rightarrow$  *least squares fitting*). Par définition, il s'agit du programme qui minimise  $\sum (e_i)^2 = e^t \cdot e$  dans chacun des  $p$  systèmes d'équations. Il conduit à des calculs algébriques et statistiques simples, ce qui explique son succès.

D'autres possibilités existent, on peut consulter à ce sujet la bibliographie de Morineau<sup>(32)</sup>.

Au prix de certaines hypothèses, dites théorème de Gauss-Markow, cet ajustement est le meilleur possible pour la régression linéaire, ce qui ne signifie pas qu'il soit le plus robuste.

Etant données les notations suivantes:

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_m x_m + e$$

Les hypothèses<sup>(33)</sup> du théorème de Gauss-Markow sont:

- 1) les  $x_j$  sont connus de façon exacte (valeurs certaines ou déterministes).
- 2) les  $e_i$  et donc les  $y_i$  sont des variables aléatoires.
- 3) on cherche des estimateurs des  $\alpha_j$  non biaisés.
- 4) les résidus ont même variance et une espérance mathématique nulle.
- 5) les résidus ne sont pas corrélés entre eux (covariance nulle).

Ces hypothèses vérifiées, on démontre que l'on ne pourra faire mieux que les moindres carrés seulement avec des estimateurs biaisés et non linéaires. Pour pouvoir approfondir l'exploitation

statistique du modèle linéaire on rajoute souvent l'hypothèse supplémentaire que la distribution des résidus est gaussienne. Cette hypothèse n'est pas nécessaire au théorème de Gauss-Markow, elle présente seulement l'avantage de conduire à des calculs simples: c'est une hypothèse de *commodité* qu'il faudra toujours vérifier. Pour la suite c'est dans ce contexte que nous parlerons des hypothèses de Gauss-Markow.

Tukey<sup>(34)</sup> évoque ce que devient l'ajustement lorsque l'on abandonne une des conditions du théorème.

#### **2.6.4.2) RESOLUTION DE L'AJUSTEMENT DES MOINDRES CARRÉS**

On doit résoudre:

$$y = X \cdot b + e$$

en imposant à  $e^t \cdot e$  d'atteindre un minimum.  $X$  est appelée variable indépendante (ou explicative) et  $y$  est la variable dépendante (ou expliquée).

On démontre<sup>(33)</sup> que la condition d'extremum est:

$$X^t \cdot y = X^t \cdot X \cdot b \quad (\text{Eq 11})$$

et que cet extremum atteint un minimum. (Eq 11) s'appelle équation normale du système.

Puisque  $X^t \cdot X$  est une matrice carrée, alors (Eq 11) admet pour solution *unique*:

$$b = (X^t \cdot X)^{-1} X^t \cdot y \quad (\text{Eq 12})$$

La solution des moindres carrés de (Eq 9) est donc:

$$\hat{b} = (X^t \cdot X)^{-1} X^t \cdot y \quad (\text{Eq 13})$$

$$\hat{e} = y - X \cdot \hat{b}$$

L'estimateur *non biaisé* de la variance des résidus est:

$$\hat{\sigma}^2 = (\sum e_i^2) / (n-m) \quad (\text{Eq 14})$$

Et la matrice des covariances est donnée par:

$$S = \hat{\sigma}^2 (X^t \cdot X)^{-1} \quad (\text{Eq 15})$$

Ce paramètre  $\sigma$  est souvent dénommé SEE ( $\rightarrow$  *Standard Error of Estimate*) ou SEC ( $\rightarrow$  *Standard Error of Calibration*). Il n'a en principe de signification que si l'erreur est effectivement indépendante de la concentration. Dans le cas où le modèle additif est choisi, il faut diviser par  $n-m-1$ .

Le théorème de Gauss-Markov conduit à considérer  $b_j$  comme l'estimateur d'une variable aléatoire  $\beta_j$ . Alors  $(b_j - \beta_j)/s_{jj}$  suit une loi de Student à  $n-m$  degrés de liberté ( $n-m-1$  si modèle additif). Le paramètre  $s_{jj}$  est un élément de  $S$ , il s'agit d'un estimateur de la variance de  $b_j$ . Ce résultat permet de construire un intervalle de confiance pour un seuil donné sur les coefficients de régression.

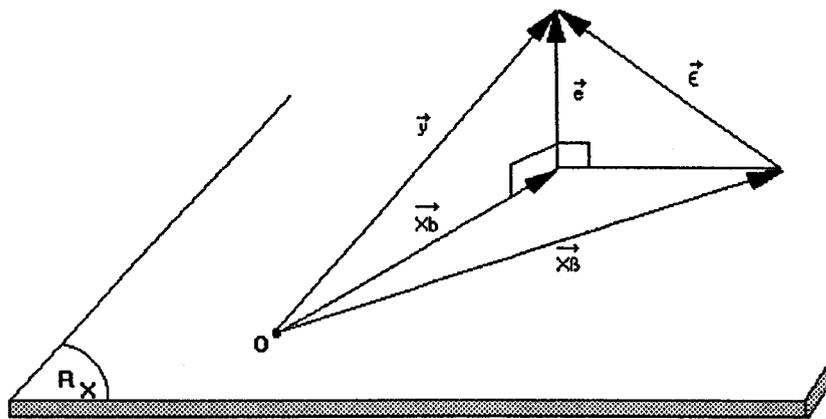


Figure 4: interprétation géométrique du critère des moindres carrés

L'ajustement des moindres carrés admet une interprétation géométrique simple. Si l'on reprend (Eq 9) et que l'on prémultiplie par  $X^t$ :

$$X^t \cdot y = X^t \cdot X \cdot b + X^t \cdot e \quad (\text{Eq 16})$$

D'où étant donnée l'équation normale (Eq 11):

$$X^t \cdot e = 0$$

Ce produit scalaire nul traduit l'orthogonalité du vecteur  $e$  avec le sous espace  $R_X$  engendré par les  $m$  vecteurs de  $X$ . Autrement dit l'ajustement des moindres carrés appliqué au modèle  $y = X \cdot \beta + \epsilon$  consiste à approcher  $y$  par sa projection orthogonale  $X \cdot b$  sur le sous espace  $R_X$ .

En résumé:

Pour un modèle multilinéaire du type:

$$y = X \cdot b + e.$$

Si  $m \leq n$  alors la solution des moindres carrés est:

$$\hat{b} = (X^t \cdot X)^{-1} X^t \cdot y$$

$$\hat{e} = y - X \cdot \hat{b}$$

$e$  étant le résidu.

Les hypothèses de Gauss-Markow sont:

- Les résidus sont des variables aléatoires non corrélées, de moyenne nulle et de même variance

$\sigma^2$ . Un estimateur non biaisé de  $\sigma^2$  est donné par:

$$\sigma^2 = (\sum e^2)/(n-m) \quad \text{pour le modèle non additif}$$

$$\sigma^2 = (\sum e^2)/(n-m-1) \quad \text{pour le modèle additif}$$

(Eq 17)

Ces résultats sont généralisables à (Eq 10).

*"Les questions de méthode priment sur toutes les autres."*

Descartes, Discours de la méthode.

## 2.7) APPLICATION A L'ANALYSE QUANTITATIVE

---

Il existe trois algorithmes de base pour résoudre le modèle linéaire:

- \* méthode CLS
- \* méthode ILS
- \* méthode des matrices Q

On les appelle encore méthodes matricielles. Nous ne présenterons que les résultats essentiels, les détails peuvent être trouvés dans la littérature<sup>(35,36,37,38,39)</sup>. Les notations employées sont définies au chapitre 2.1.

### 2.7.1) LA METHODE CLS

On l'a également appelée régression linéaire ordinaire, méthode directe, calibration totale ou méthode des matrices K. Elle consiste tout simplement à appliquer la loi de Beer-Lambert:

$$Y = X \cdot B + E \quad (\text{Eq 18})$$

**B** apparaît donc comme la matrice des spectres des produits purs à la concentration d'une unité.

#### \* calibration

Connaissant **X** et **Y**, on calcule **B** par régression linéaire à partir de (Eq 18):

$$\hat{B} = (X^t \cdot X)^{-1} X^t \cdot Y$$

$$\hat{E}_c = Y - X \cdot \hat{B}$$

$\dim(\hat{B}) = m \times p$  ou  $(m+1) \times p$  pour le modèle additif

$\dim(\hat{E}_c) = n \times p$

\* validation-prédiction

On connaît  $\hat{B}$  et  $Y_u$ , on calcule  $\hat{X}_u$ :

$$\hat{X}_u = Y_u \cdot \hat{B}^t \cdot (\hat{B} \cdot \hat{B}^t)^{-1}$$

$$\hat{E}_p = Y_u - \hat{X}_u \cdot \hat{B}$$

\* remarques

- La méthode CLS exige de connaître tous les constituants du mélange et de les inclure dans la calibration.
- Il s'agit d'une méthode de spectres entiers ( $\rightarrow$  *full spectrum method*), il n'y a donc pas besoin de faire une sélection des fréquences utilisées. Cette méthode pourra donc fournir une meilleure précision que celles qui se contentent de quelques longueurs d'onde<sup>(40)</sup> et est plus efficace pour la détection des spectres aberrants.
- La méthode CLS permet d'intégrer une ligne de base<sup>(41)</sup> aux spectres.
- Anderson et Griffiths<sup>(42)</sup> ont calculé les variations de la hauteur réelle d'un pic en fonction de la hauteur théorique pour différentes formes de la fonction d'appareil. Ils ont ainsi montré qu'à partir de 0.7 DO la loi de Beer-Lambert n'est plus suivie et préconisent donc de travailler en dessous de ce plafond pour les applications quantitatives.

2.7.2) LA METHODE ILS

On l'appelle également méthode des matrices P, régression inverse ( $\rightarrow$  *Inverse Least Squares*), calibration partielle ou méthode MLR bien que les trois méthodes présentées ici reposent sur des algorithmes des régression linéaire multiple.

Le modèle théorique est le même que celui de la méthode CLS si ce n'est que l'on permute X et Y. L'algorithme de régression s'écrit:

$$X = Y \cdot B + E \tag{Eq 19}$$

Cette méthode semble avoir été proposée pour la première fois par Krutchkoff<sup>(43)</sup> en 1967, elle consiste à choisir Y comme variable indépendante plutôt que X.

\* calibration

$$\hat{\mathbf{B}} = (\mathbf{Y}^t \cdot \mathbf{Y})^{-1} \mathbf{Y}^t \cdot \mathbf{X}$$

$$\hat{\mathbf{E}}_c = \mathbf{X} - \mathbf{Y} \cdot \hat{\mathbf{B}}$$

$\dim(\hat{\mathbf{B}}) = pxm$  ou  $px(m+1)$  si modèle additif

$\dim(\hat{\mathbf{E}}_c) = nxm$  ou  $nx(m+1)$

\* validation-prédiction

$$\hat{\mathbf{X}}_u = \mathbf{Y}_u \cdot \hat{\mathbf{B}}$$

$$\hat{\mathbf{E}}_p = \hat{\mathbf{X}}_u - \mathbf{Y}_u \cdot \hat{\mathbf{B}}$$

\* remarques

- La méthode ne peut traiter un spectre complet, il sera donc nécessaire de choisir les fréquences sur lesquelles on travaillera. Ce problème de réduction de la taille des données peut aussi être résolu de façon beaucoup plus efficace par les techniques de compression des données décrites au paragraphe 2.7.4. A partir de la méthode ILS on pourra alors développer les algorithmes PCR et PLS.

- On montre qu'au voisinage de la valeur moyenne des concentrations, le prédicteur inverse est plus performant que le prédicteur direct<sup>(44)</sup>. De plus la méthode inverse est mieux stabilisée à l'égard du bruit sur les données expérimentales que la méthode directe. Ces résultats sont toutefois théoriques.

- Contrairement à la méthode CLS, il n'est plus indispensable de connaître tous les constituants du mélange car la calibration ne fait pas explicitement intervenir le nombre  $m$  de constituants. On pourra donc effectuer une analyse même en présence d'impuretés parasites et mal connues dans le mélange<sup>(45)</sup>.

Par exemple si l'on dose le  $j^{\text{ème}}$  constituant, (Eq 19) devient :

$$x_j = \mathbf{Y} \cdot \hat{\mathbf{b}}_j \tag{Eq 20}$$

que l'on résout par moindres carrés de la façon habituelle.

### 2.7.3) METHODE DE LA MATRICE Q

Cette méthode peu connue a été développée par Mc Clure<sup>(46)</sup>. L'idée consiste à régresser directement  $Y_u$  sur  $Y$ :

$$Y_u = Q \cdot Y + E \quad (\text{Eq 21})$$

#### \* calibration

$$\hat{Q} = Y_u \cdot Y^t \cdot (Y \cdot Y^t)^{-1}$$

$$\hat{E}_c = Y_u - \hat{Q} \cdot Y$$

#### \* validation-prédiction

La loi de Beer-Lambert s'écrit:  $Y = X \cdot B$ .

En remplaçant dans (Eq 21) on trouve:

$$X_u \cdot B = \hat{Q} \cdot X \cdot B$$

d'où:

$$X_u \cdot B \cdot B^t = \hat{Q} \cdot X \cdot B \cdot B^t$$

et en inversant  $B \cdot B^t$ :

$$\hat{X}_u = \hat{Q} \cdot X$$

qui est l'équation de prédiction.

#### \* remarques

- Nous n'avons qu'une seule matrice de dimension  $n \times n$  à inverser. Pour améliorer la précision on devra donc disposer de plus de standards que lors des méthodes précédentes.
- Cette méthode s'avère à l'usage très sensible à la similarité des spectres des produits purs. Elle nécessite donc l'emploi d'algorithmes de calculs numériques robustes. On peut ainsi s'attendre à des performances très variables selon la nature des mélanges.

### 2.7.4) LES TECHNIQUES DE CONDENSATION DES DONNEES

Les difficultés abordées au chapitre 2.5 au sujet des traitements de données sont en grande partie résolues par les techniques de condensation des données.

Lorsque l'on dose une solution contenant  $m$  constituants, il est intuitivement clair que seulement  $m$  caractéristiques de ce mélange devraient suffire pour effectuer l'analyse. Pourtant l'expérience montre que pour maintenir une précision suffisante, on doit effectuer un nombre de mesures très supérieur à  $m$ . On interprète ce résultat en disant qu'il faut tenir compte du bruit sur les spectres, des erreurs possibles d'échantillonnage, des mesures erronées... En d'autres termes on cherche à marginaliser dans la masse des résultats l'effet des valeurs anormales et donc à améliorer la robustesse du système.

Cela ne se fait pas sans inconvénient dont un accroissement exponentiel du temps de traitement et un risque accru de propagation d'erreurs numériques.

Une meilleure approche consiste à extraire de la masse d'informations, celles qui sont redondantes (corrélées), anormales ou nuisibles (bruit) afin de ne garder que les meilleures pour l'analyse quantitative. Cette idée de caractériser un ensemble complexe par un nombre restreint de variables n'est pas nouvelle: par exemple en 1879 Bertillon montra que la connaissance de seulement quelques mesures morphologiques est suffisante pour retrouver un individu inscrit dans un fichier de police.

Les techniques qui réalisent cette idée sont dites méthodes factorielles ou méthodes de condensation des données.

Dans le cas de l'analyse quantitative, celle-ci interviendront notamment sur les points suivants:

- \* amélioration du conditionnement et de la robustesse du problème de régression.
- \* réduction de la masse de données.
- \* détection et élimination des valeurs anormales.
- \* réduction du bruit du spectre.
- \* résolution du problème de collinéarité.
- \* amélioration de la linéarité des données.

Mathématiquement, le problème de toute méthode factorielle se formule ainsi: étant donné une matrice rectangulaire  $M$  de dimension  $l.c$ , est-il possible de reconstituer chacun des  $l.c$  coefficients à partir d'un plus petit nombre de valeurs numériques?

Supposons qu'il existe un vecteur  $t_1$  à  $l$  composantes et un vecteur  $p_1$  à  $c$  composantes tels que  $M = t_1 \cdot p_1^t$ . On aura reconstitué les  $l.c$  coefficients de  $M$  à l'aide de seulement  $l+c$  valeurs numériques (dans ce cas  $M$  est de rang 1). En pratique ce cas de figure est extrêmement rare et l'on cherche plutôt une approximation de rang  $h$ :

$$M = t_1 p_1^t + t_2 p_2^t + \dots + t_h p_h^t + E \quad (\text{Eq 22})$$

$E$  étant une matrice résiduelle dont les termes sont suffisamment petits pour que l'on puisse considérer la reconstitution comme suffisante. (Eq 22) peut également s'écrire de façon plus compacte:

$$M = T \cdot P^t + E \quad (\text{Eq 23})$$

$T$  est appelée matrice des scores ( $\rightarrow$  *scores matrix*) et  $P^t$  est la matrice des loadings ( $\rightarrow$  *loadings matrix*). Les loadings sont rangés selon les lignes de  $P^t$  et les scores selon les colonnes de  $T$ . Les *scores* sont également appelés: *abstract factors*, *factors* ou encore *eigenspectra of M*. On dit que l'on approxime  $M$  par la matrice  $T$  de dimensions plus petites. Les loadings forment une base vectorielle de l'espace de projection et les scores sont les nouvelles coordonnées de  $M$  dans cet espace. L'espace de projection est appelé espace factoriel par opposition à l'espace réel initial.

Cette décomposition n'est pas unique et fait l'objet de nombreux algorithmes. La méthode CLS en est un exemple simple: la matrice  $Y$  équivaut à une description des spectres fréquence par fréquence, soit  $n \times p$  valeurs. Rapprochons (Eq 18) de (Eq 23) en faisant  $Y=M$ . La matrice  $B$  est la matrice des spectres des constituants purs. On est donc passé d'une représentation par fréquences (matrice  $Y$ ) à une représentation par concentrations (matrice  $X$ ) soit  $n \times m$  valeurs. Le nombre de constituants  $m$  étant inférieur au nombre de fréquences échantillonnées on a bien réalisé une réduction du nombre de données. Dans ce contexte les loadings sont les spectres des produits purs et les scores les concentrations. L'intérêt du nouvel espace factoriel est que les coordonnées des

spectres coïncident avec les concentrations recherchées, mais ce n'est pas nécessairement le cas des autres méthodes factorielles.

D'autres techniques de condensation existent, ce sont par exemple les méthodes de régression multiple pas à pas ( $\rightarrow$  *SMLR*), la régression de Hruschka ou la régression de Fourier (voir Martens<sup>(35)</sup> pour un exposé détaillé). Un groupe de méthodes particulièrement utilisées sont les méthodes dites bilinéaires pour lesquelles les matrices **P** et **T** sont issues ni d'une sélection sur les données initiales (*SMLR* par exemple), ni d'une transformation mathématique fixée (régression de Fourier), mais estimées à partir de **X** et/ou **Y**. Les deux algorithmes les plus utilisés en analyse quantitative sont la méthode PCR et la méthode PLS. Ces procédures conduisent à une projection des matrices selon des axes orthogonaux, ce qui suffit pour résoudre la plupart des insuffisances des méthodes matricielles. En contrepartie les axes obtenus n'ont plus de significations physiques contrairement à ceux de la méthode CLS. Les deux variantes principales pour l'analyse quantitative se distinguent les unes des autres par la méthode de construction de l'espace factoriel. La méthode PCR débute par une analyse en composantes principales de **Y** suivie d'une régression de type ILS sur la matrice des scores. Elle est mieux conditionnée que la méthode ILS seule, mais elle n'optimise que la représentation de **Y** sans tenir compte des concentrations. On a donc aucune garantie que l'espace factoriel ainsi construit soit optimal pour la prédiction des concentrations. On peut construire des espaces factoriels en utilisant également les informations chimiques: c'est le but de la méthode PLS.

Si les bases théoriques existent parfois depuis le début du siècle, seule l'arrivée d'une informatique bon marché a permis d'en exploiter pleinement les acquis. Ces méthodes sont donc tout à fait représentatives de l'approche chimiométrique du problème quantitatif.

#### **2.7.4.1) L'ANALYSE EN COMPOSANTES PRINCIPALES ET LA METHODE PCR**

La méthode PCR est, tout comme les méthodes matricielles précédemment décrites, constituée d'une étape de calibration, de validation puis de prédiction. La différence est que l'espace vectoriel dans lequel se réaliseront ces étapes n'est pas l'espace canonique issu des données expérimentales, mais un espace construit spécialement par une procédure purement mathématique. Cet algorithme est surtout utilisé dans le cadre de la méthode ILS, en raison des avantages qu'elle procure.

##### **\* construction du nouvel espace vectoriel**

On recherche une décomposition de  $Y$  du type (Eq 23) dont les axes soient orthogonaux. Parmi l'infinité de solutions possibles, la méthode PCR s'intéresse à la décomposition en composantes principales de la matrice  $Y$ , c'est à dire aux vecteurs propres de  $Y^t \cdot Y$  qui correspondent aux valeurs propres ( $\rightarrow$  *eigenvalues*) les plus élevées (en valeur absolue).  $Y^t \cdot Y$  est encore parfois appelée matrice de dispersion. Du point de vue informatique on peut effectuer la décomposition directement sur  $Y$ , c'est à dire sans devoir calculer  $Y^t \cdot Y$ , à l'aide de divers algorithmes comme la décomposition en valeurs singulières (méthode SVD) ou l'algorithme NIPALS qui est une méthode itérative mise au point par Fisher dans les années 30. Une bonne description en est donnée par Geladi et Kowalski<sup>(47)</sup> ainsi que par Martens<sup>(35)</sup>. Récemment Donahue a présenté une technique dite "orthogonalisation par moyennes successives" qui serait encore plus performante<sup>(48)</sup>.

On peut se demander quel est l'intérêt du choix des vecteurs propres de  $Y^t \cdot Y$  plutôt que de n'importe quel autre ensemble de vecteurs orthogonaux.

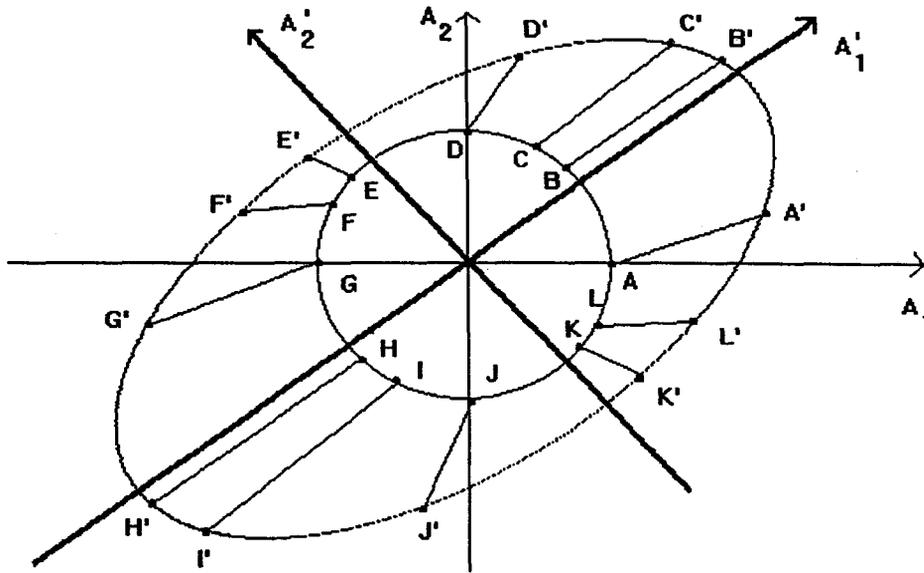
La résolution du problème des moindres carrés par la méthode ILS est équivalent à la résolution de l'équation normale de (Eq 20) :

$$Y^t \cdot Y \cdot b_j = Y^t \cdot x_j \quad (\text{Eq 24})$$

$Y^t \cdot Y$  est une matrice carrée d'ordre  $p$ , on peut donc la considérer comme la matrice d'un endomorphisme  $\Phi$ . Résoudre (Eq 24) équivaut donc à résoudre l'équation vectorielle:

$$\Phi(b_j) = Y^t \cdot x_j \quad (\text{Eq 25})$$

La fonction  $\Phi$  ne transforme pas le vecteur  $b_j$  (ou le point de coordonnées  $b_j$ ) de la même façon selon la direction choisie. On le vérifie facilement en dimension 2 (voir figure 5) lorsque l'on trace l'image par  $\Phi$  d'un cercle.



**Figure 5:** transformation d'un cercle en ellipse par la fonction  $Y^t \cdot Y$ . (D'après De Lagarde<sup>(49)</sup>).

Le cercle est devenu une ellipse. La fonction  $\Phi$  tend donc à "étaler" les données préférentiellement dans les directions  $A'_1$  et  $A'_2$  qui sont les axes de symétrie de l'ellipse. Choisissons alors les axes  $A'_1$  et  $A'_2$  comme nouveau repère. Dans ce nouveau système d'axes, pour transformer le cercle initial en ellipse (son image par  $\Phi$ ) il suffit de multiplier les abscisses par un réel  $\lambda_1$  et les ordonnées par un réel  $\lambda_2$ . La matrice de  $\Phi$  dans  $(A'_1, A'_2)$  s'écrit donc:

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

$\lambda_1$  et  $\lambda_2$  apparaissent comme les valeurs propres (ou valeurs latentes) de  $\Phi$  et  $A'_1$  et  $A'_2$  sont donc les axes de vecteur propre ou composantes principales.

Les valeurs propres apparaissent donc comme une mesure de l'étalement des données (c'est à dire de la variance) le long de l'axe correspondant. Dans les cas extrêmes il peut même arriver que l'expansion selon  $A_2$  soit négligeable devant celle selon  $A_1$ , c'est à dire que  $\lambda_2$  est très faible devant  $\lambda_1$ .

On ne commettra donc pas une erreur importante en négligeant ce qui se passe dans la direction  $A_2$ . On est donc ramené à un problème en dimension 1 plus simple que celui de dimension 2. Ces résultats sont généralisables à une dimension quelconque. On obtient alors  $p$  composantes principales (dimension de  $Y^t.Y$ ). Dans ce cas l'ellipse devient une ellipsoïde appelée ellipsoïde d'inertie. On ne conserve que les  $h$  vecteurs propres unitaires jugés significatifs selon un certain critère de rejet comme par exemple la méthode des validations croisées<sup>(50,51)</sup> ( $\rightarrow$  *cross-validatory method*). Ces vecteurs propres correspondent aux valeurs propres les plus élevées. L'algorithme PCR utilise ces  $h$  vecteurs en tant que loadings. Si  $T$  désigne la matrice des scores:

$$Y = T \cdot P^t + E \quad (\text{Eq 26})$$

$$\text{Dim}(T) = nxh$$

$$\text{Dim}(P^t) = hxp$$

$P^t$  est la matrice des loadings, celle-ci est orthonormale ( $P \cdot P^t = P^t \cdot P = I$ ), tandis que  $T$  est une matrice orthogonale. L'expression ainsi obtenue constitue la décomposition en composantes principales de  $Y$ ,  $T$  étant la projection de  $Y$  dans le repère des loadings. Le calcul de  $T$  se fait par régression linéaire, mais avec une simplification due au fait que  $P^t$  est une matrice orthonormale, on trouve alors:

$$T = Y \cdot P \quad (\text{Eq 27})$$

Un aspect important de cette décomposition est que les axes choisis correspondent à une dispersion maximale des données ce qui permet de repérer les points proches avec une meilleure précision. Ce résultat peut être encore amélioré en centrant les données<sup>(40)</sup>. Traditionnellement on standardise également  $Y$ . Cette opération est surtout utile lorsque la

dimension des grandeurs utilisées n'est pas homogène, ce qui ne concerne guère l'analyse quantitative infrarouge.

Cette façon de décomposer une matrice constitue l'analyse en composantes principales (→ PCA) dont nous résumons les principales étapes:

#### ANALYSE EN COMPOSANTES PRINCIPALES DE LA MATRICE Y

- \* On centre et éventuellement standardise Y.
- \* On calcule les vecteurs propres unitaires de  $Y^t \cdot Y$ .
- \* On sélectionne les h vecteurs propres les plus significatifs afin de construire la matrice des loadings  $P^t$ . Cette matrice définit le nouveau repère de projection.
- \* On calcule la matrice des scores  $T = Y \cdot P$ . Cette matrice est la projection de Y dans le nouveau repère. On l'utilisera comme un estimateur de la matrice Y.

#### \* calibration

On dispose d'une matrice de réponses Y et d'une matrice de concentrations X.

Nous pouvons alors calculer la projection T de Y dans l'espace des composantes principales.

La méthode PCR est basée sur l'algorithme ILS, on écrit donc:

$$X = T \cdot B + E \quad (\text{Eq 28})$$

B est calculée par régression linéaire:

$$\hat{B} = (T^t \cdot T)^{-1} T^t \cdot X \quad (\text{Eq 29})$$

$$\hat{E}_c = X - T \cdot \hat{B}$$

On remarque que T étant une matrice orthogonale,  $T^t \cdot T$  est nécessairement inversible (pas de problèmes de collinéarité).

**\* validation-prédiction**

$$\hat{T}_u = Y_u \cdot P$$

$$\hat{X}_u = \hat{T}_u \cdot \hat{B}$$

$$\hat{E}_p = \hat{X}_u - \hat{T}_u \cdot \hat{B}$$

**\* remarques**

- La méthode PCR combine les avantages de la méthode CLS (utilisation de tout le spectre) avec ceux de la méthode ILS (analyse possible en ne connaissant la concentration que de certains constituants dans les standards), mais sans en reprendre les inconvénients.
- Si les acquisitions sont faites dans de bonnes conditions, l'allure du spectre doit provenir davantage des constituants du mélange que du bruit. On peut donc prévoir que les valeurs propres les plus faibles correspondront aux phénomènes parasites comme le bruit qui sera de ce fait éliminé (les valeurs propres mesurent l'étalement des valeurs le long des loadings).

**2.7.4.2) LA METHODE PLS**

L'algorithme PLS a été développé dans les années 70 pour résoudre des problèmes économiques. Il fut ensuite repris en 1983 par différents groupes<sup>(40,52,53)</sup> pour être appliqué au problème de l'analyse quantitative. Comme la méthode PCR il s'agit également d'une méthode factorielle, c'est à dire que l'on cherche une décomposition des matrices de données du type (Eq 23). L'algorithme de décomposition est par contre très différent: on réduit les informations non plus dans un espace vectoriel constitué des dimensions dominantes en terme de variance, mais dans les directions les plus pertinentes en terme de prédiction des concentrations. Elles ne sont pas nécessairement identiques. Les informations spectroscopiques (Y) aussi bien que les informations chimiques (X) vont donc participer à l'étape de calibration.

### \* Construction du nouvel espace factoriel et calibration

On recherche une décomposition en scores et en loadings de X et Y:

$$X = T \cdot P^t + E_X$$

$$Y = U \cdot Q^t + E_Y$$

en imposant une relation interne entre scores qui est généralement linéaire, c'est l'équation de calibration:

$$T = U \cdot B + E \quad (\text{Eq 30})$$

E est une matrice de résidus et B la matrice des coefficients de régression. L'idée la plus simple serait d'utiliser la décomposition en composantes principales de X et Y puis de résoudre (Eq 29) par régression multilinéaire. Cette approche est artificielle et n'apporte rien de plus que la méthode PCR. Il est plus intéressant de rechercher des décompositions dont le critère de sélection soit directement la minimisation de  $\|E\|$ , sans se préoccuper spécialement de minimiser  $\|E_X\|$  ou  $\|E_Y\|$ . Tout comme la méthode PCR il existe deux types principaux d'algorithmes de recherche: la décomposition en valeurs singulières<sup>(54)</sup> et un algorithme itératif dit NIPALS<sup>(47)</sup>. Ces algorithmes se répartissent en deux catégories selon que l'on estime chaque  $x_j$  séparément (PLS1) ou bien simultanément (PLS2). Contrairement au cas de la méthode PCR le choix n'est pas indifférent puisque les  $x_j$  interviennent dans la construction de l'espace de réduction. La question du choix le mieux adapté est débattue par Martens<sup>(35)</sup>.

### \* validation-prédiction

Des relations du chapitre précédent on extrait l'équation de prédiction sachant que  $P^t$  et  $Q^t$  sont des matrices orthonormales:

$$\hat{U}_u = Y_u \cdot \hat{Q}$$

$$\hat{X}_u = \hat{U}_u \cdot \hat{B} \cdot P^t \quad (\text{équation de prédiction})$$

$$\hat{T}_u = \hat{X}_u \cdot P$$

$$\hat{E}_p = \hat{T}_u - \hat{U}_u \cdot \hat{B}$$

$\hat{Q}^t$ ,  $\hat{B}$  et  $P^t$  sont obtenues lors de l'étape de calibration.  $Y_u$  est la matrice des spectres inconnus (valeurs expérimentales) tandis que  $\hat{X}_u$  est la matrice des concentrations que l'on cherche à estimer.

**\* remarques**

- Les propriétés statistiques de la méthode PLS sont actuellement moins bien comprises que celles de la méthode PCR. Par exemple l'expression exacte du paramètre SEE n'est pas encore établie de façon exacte, des approximations doivent être utilisées.
- Malgré sa complexité apparente, PLS est plus rapide que les autres méthodes matricielles, y compris PCR.

## 2.8) COMPARAISON DES METHODES CLS, ILS, PCR ET PLS

---

Les trois méthodes de spectres complets (CLS, PCR et PLS) sont dans la plupart des situations plus performantes que la méthode ILS. La raison en est qu'en travaillant sur tout le spectre, on peut espérer diminuer l'influence de non-linéarités locales. Toutefois on peut aussi imaginer une situation extrême où seule une petite partie du spectre obéit à une loi du type Beer-Lambert. Dans ce cas, la méthode ILS intégrant moins de données, il est probable qu'elle délivrera de meilleurs résultats que ses concurrentes. La méthode ILS est aussi moins performante pour la recherche des valeurs aberrantes puisque qu'elle utilise des informations spectrales moins détaillées que les techniques de spectres complets.

Les méthodes PCR et PLS conduisent à des matrices orthogonales. Il n'y a donc plus de risque de collinéarité. En outre elles travaillent sur des matrices de dimension plus faible, mieux conditionnées et utilisent des données de meilleure qualité (moins de bruit, de valeurs redondantes...). Ces méthodes doivent donc normalement obtenir de meilleurs résultats que la méthode CLS. Le principal point faible de ces algorithmes résulte dans le choix du nombre correct de loadings. On appelle "*overfitting*" les situations où trop de loadings sont retenus et "*underfitting*" les cas contraires.

En général la méthode PLS offre des performances comparables à celles de la méthode PCR. Elle est toutefois moins sensible à la présence de phénomènes comme une ligne de base aléatoire sur les spectres. La raison en est que la décomposition de type PCR utilise uniquement la variance calculée sur les informations spectrales sans tenir compte des concentrations. Puisque la variance due à la ligne de base peut être importante, la décomposition PCR peut donc être significativement influencée par ce type de fluctuation qui n'est pas corrélable aux concentrations. Par contre, cet effet sera plus facilement relativisé dans le cas de la méthode PLS. Au chapitre 5.4 nous présenterons un exemple concret dans lequel la méthode PLS fonctionne correctement, mais pas la méthode PCR.

Une comparaison détaillée des performances des méthodes CLS, ILS, PCR et PLS pour différentes conditions opératoires peut être trouvée dans un article de Thomas et Haaland<sup>(55)</sup>. Cette étude montre que selon les situations (bruit sur les concentrations, sur le spectre, nombre de standards...) les résultats donnés par ces différentes méthodes sont très variables. Ce travail confirme la supériorité des méthodes de spectres complets, mais montre également que l'écart entre CLS d'une part et PCR ou PLS d'autre part est souvent relativement faible, tout au moins dans les situations étudiées dans cet article.

## 2.9) VALIDATION D'UNE ANALYSE

---

Cette rubrique a permis d'introduire dans la littérature une inflation de sigles dont nous reproduisons les plus fréquemment utilisés:

MSE	Mean Square Error
MSEP	MSE of Prediction
PRESS	PREdiction Sum of Squares
RMSD	Root Mean Square of Differences
RMSE	Root Mean Square of Errors
RMSEP	RMSE of Prediction
RMSP	Root Mean Square of Prediction
RSD	Relative Standard Deviation (→ <i>coefficient de variation</i> )
SEE	Standard Error of Estimate
SEP	Standard Error of Prediction

Faute de définitions normalisées, ces paramètres peuvent recevoir des expressions littérales différentes selon les auteurs.

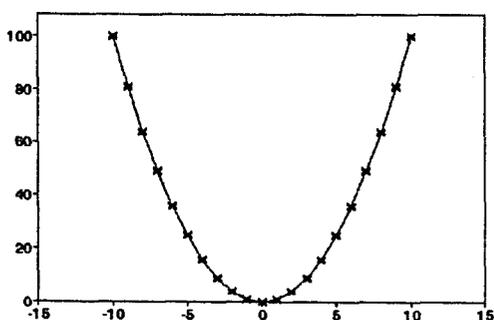
Il n'est peut être pas inutile de rappeler que le but essentiel de l'étape de validation est:

- vérifier que le modèle construit par la calibration fonctionne correctement.
- estimer l'erreur réalisée lors de la prédiction.

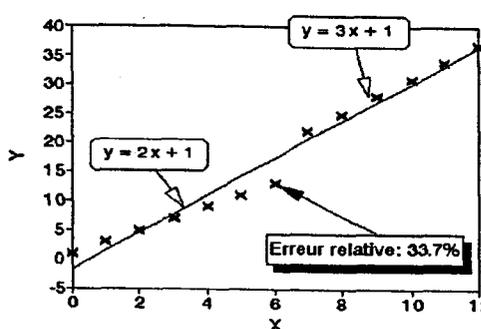
Ces deux objectifs peuvent être remplis *seulement* en étudiant le comportement en prédiction de spectres connus et *différents* des standards de calibration. Nous allons décrire rapidement les techniques les plus fréquemment utilisées en s'intéressant plus spécialement à leurs effets secondaires:

**\* Le coefficient de corrélation (r)**

Cette statistique a été introduite au début du siècle par Karl Pearson. Son objet n'est pas d'estimer l'erreur de prédiction, mais de vérifier si le modèle de calibration est plausible. Un paramètre dérivé est le coefficient de détermination  $r^2$  qui mesure la part de la variance totale expliquée par le modèle. Malgré une apparente simplicité le coefficient de corrélation est en pratique très délicat à manier comme l'attestent les deux exemples suivants:



Cas de la parabole:  $r = 0$ , pourtant  $x$  et  $y$  sont parfaitement corrélées, la corrélation n'est tout simplement pas linéaire.



Deux segments de droite:  $r = 0,987$ . Cette valeur est très bonne, pourtant la prédiction conduit à des erreurs importantes dans la partie médiane.

Le problème provient du manque de sensibilité du coefficient de corrélation à la répartition effective des erreurs. En principe, un examen visuel pourrait permettre d'identifier immédiatement le problème, mais à condition de ne pas être dans un contexte multilinéaire. Pour cette raison nous pensons que le coefficient de corrélation n'est pas un paramètre très bien adapté aux applications pratiques. Par exemple Van Arendonk<sup>(56)</sup> présente un exemple extrait de l'analyse d'une publication où l'examen du seul coefficient de corrélation a conduit à une interprétation totalement fautive d'un phénomène physique.

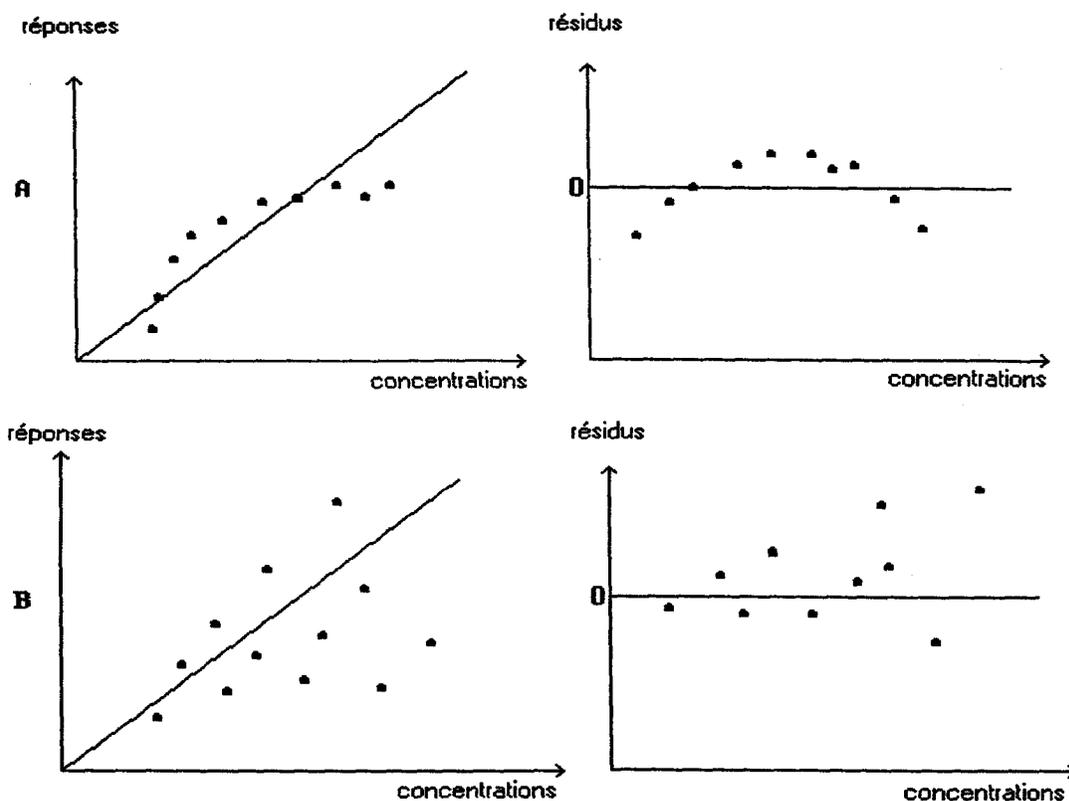
**\* analyse des résidus**

On calcule le résidu obtenu pour chaque concentration:

$$e = x - \hat{x}$$

Si l'on accepte les hypothèses de Gauss-Markow, alors les résidus doivent avoir même variance, même moyenne (égale à 0) et être non corrélés. En pratique il est surtout facile de vérifier une conséquence de ces conditions qui est que les résidus soient indépendants de

la concentration. Le plus simple est de faire un test visuel en traçant les résidus (éventuellement standardisés) en fonction des concentrations correspondantes. On vérifie alors l'absence de regroupements de résidus de même signe et l'homogénéité de la distribution de part et d'autre de 0 (absence de biais). L'apparition d'une nette tendance (Fig 6-A) ou d'une dépendance de la valeur du résidu en fonction de la concentration (Fig 6-B) peut traduire un mauvais ajustement ou bien que les hypothèses de Gauss-Markow ne soient pas vérifiées. Une analyse du résidu aurait par exemple permis de suspecter les deux ajustements précédents.

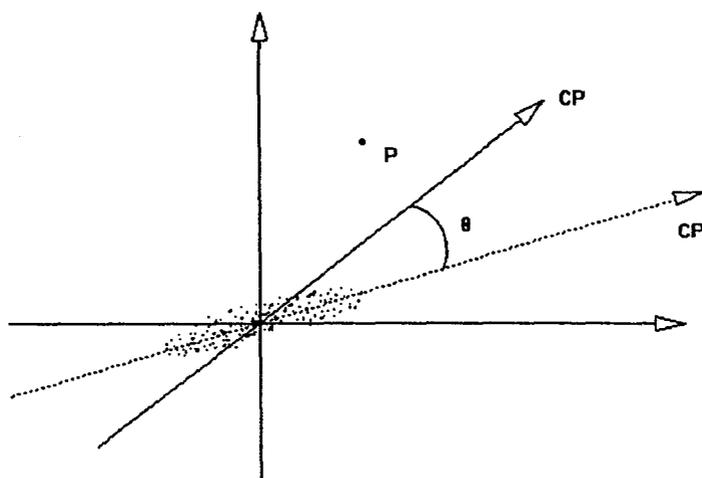


**Figure 6:** Illustration de quelques répartitions atypiques des résidus. A gauche les points expérimentaux, à droite les résidus correspondants.

Lorsque le nombre d'observations est faible, il est nécessaire d'être très prudent dans ses conclusions.

### \* intervalle de confiance

L'estimation d'un intervalle de confiance peut être lié au calcul d'un paramètre souvent calculé par les logiciels quantitatifs. Il s'agit d'une distance au sens mathématique du terme, mais qui n'a rien à voir avec la distance euclidienne mesurée avec une règle. Cette distance est employée dans deux contextes différents qui sont la recherche de valeurs aberrantes et le calcul d'un intervalle de confiance de prédiction. La littérature américaine emploie souvent l'expression de "*leverage*" au lieu de "Distance de Mahalanobis" en raison de l'interprétation physique simple qu'a ce paramètre:



**Figure 7:** Interprétation géométrique de la distance de Mahalanobis

L'axe CP est un des axes factoriels qui décrivent le nuage de point. Le point P est en dehors du nuage, son influence sur le calcul de CP peut être mesurée par l'angle  $\theta$  de rotation entre CP et l'axe CP' obtenu en l'absence de ce point. La distance de Mahalanobis sera d'autant plus forte que cet angle sera important.

Ce paramètre ne suffit toutefois pas pour décider si un point est aberrant: un point avec une distance forte, mais situé près de l'axe factoriel, ne fait que le renforcer. La distance de Mahalanobis fonctionne plutôt comme une alarme pour désigner les points qui influencent le plus les résultats. La présence d'une valeur anormale parmi eux devra donc être particulièrement surveillée. Dans le contexte d'une analyse quantitative, on calcule la

distance de Mahalanobis par rapport au nuage de point des standards de calibration. La distance de Mahalanobis intervient aussi directement dans l'évaluation d'un *intervalle de confiance de prédiction*. Plaçons nous dans le cadre du modèle ILS. La variable  $\hat{x}_p$  est la valeur prédite de  $x_p$ , concentrations des constituants du spectre  $y_p$ . On a donc:

$$\hat{x}_p = y_p^t \cdot B$$

Y désignant la matrice des standards de calibration, on montre<sup>(33)</sup> dans le cadre du critère des moindres carrés que:

$$s_p^2 = (SEE)^2 (1 + y_p^t (Y^t \cdot Y)^{-1} y_p)$$

est un estimateur sans biais de la variance de l'erreur de prédiction. Le paramètre  $D^2 = y_p^t (Y^t \cdot Y)^{-1} y_p$  est l'expression de la distance de Mahalanobis du spectre  $y_p$  par rapport aux standards de calibration. Par conséquent:

$$s_p^2 = (SEE)^2 (1 + D^2) \quad (\text{Eq 31})$$

Dans le modèle statistique de Gauss-Markow et en supposant une distribution gaussienne des résidus, la variable  $(\hat{x}_p - x_p)/s_p$  suit une loi de Student à n-p degré de liberté. C'est de cette propriété que l'on peut calculer un intervalle de confiance sur la valeur prédite. Ce résultat montre l'importance d'avoir un  $D^2$  faible, mais ne signifie pas qu'un spectre avec une distance de Mahalanobis élevée soit aberrant.

#### \* moyenne des erreurs quadratiques (MSE)

L'erreur quadratique moyenne (MSE) est définie par:

$$MSE = E((x - \hat{x})^2) = E(e^2) \quad (\text{Eq 32})$$

Ce paramètre présente l'inconvénient d'avoir la dimension d'une concentration au carré, ce qui est peu intuitif à utiliser. C'est pourquoi on définit également le RMSE (ou RMSD):

$$RMSD = RMSE = (MSE)^{1/2}$$

Il s'agit de la définition donnée par Hruschka<sup>(4)</sup>. La question est donc d'estimer MSE. Deux philosophies sont possibles selon que l'on fasse ou non une étape de validation.

a) validation externe

Lorsque MSE est estimé à partir d'un jeu de spectres de validation, il s'appelle MSEP. On définit de la même façon le paramètre RMSEP (ou RMSP). MSEP est calculé selon l'équation (Eq 32) à partir des spectres de validation.

b) validation interne

Plusieurs stratégies sont possibles. Calculer directement MSE par (Eq 32) appliquée aux standards de calibration ne donne pas de résultat représentatifs car c'est le paramètre que minimise justement la procédure des moindres carrés. Une meilleure estimation peut être obtenue par la méthode des validations croisées: du jeu de standard on extrait un spectre qui est utilisé en validation tandis que les n-1 autres servent de standard pour la calibration. On recommence ainsi n fois. Lorsque chaque spectre aura servi une fois en prédiction on peut alors estimer le MSE comme précédemment (Eq 32). Il est clair que cette procédure n'a de sens que si l'on dispose d'un grand nombre de standards. D'autres techniques existent, Martens et Naes en présentent quelques une dans leur ouvrage<sup>(4)</sup>.

\* erreur standard de prédiction (SEP)

Le SEP (erreur standard de prédiction) est un paramètre souvent utilisé. Un estimateur est donné par:

$$SEP = \sqrt{\frac{\sum(\hat{x} - E(\hat{x}))^2}{(M - 1)}} \quad (\text{Eq 33})$$

M est le nombre de spectres où le constituant considéré apparaît à la concentration x. Cette définition est en pratique difficile à utiliser ce qui fait que très souvent on estime le SEP par le RMSEP<sup>(57)</sup>:

$$SEP = \sqrt{\text{PRESS}/M} \quad (\text{Eq 34})$$

$$\text{PRESS} = \sum e^2$$

M désignant le nombre de spectres disponibles cette fois. Cette formulation n'a évidemment de sens que si les résidus sont indépendants de la concentration.

Ces paramètres n'ont généralement de signification statistique que si l'on dispose de suffisamment de spectres de validation pour couvrir toute la variabilité des échantillons et toute l'étendue de la gamme de concentration. Ce n'est pas toujours possible, dans ce cas on devra soit se contenter de les évaluer sur un ensemble plus faible de spectres en étant conscient des problèmes de précision et de fiabilité.

Parfois ils ne sont interprétables que dans un contexte statistique précis (intervalle de confiance). Il sera donc nécessaire de vérifier les hypothèses faites, ce qui est en pratique souvent difficile. Le théorème de Gauss-Markow s'appuie sur des idées inspirées du théorème de la limite centrale. Or la réalité est souvent beaucoup plus complexe. C'est pourquoi l'usage actuel tend de plus en plus à employer les statistiques précédentes dans un contexte non paramétrique. On peut par exemple considérer que l'erreur sur une mesure est égale à  $3 \cdot \text{SEP}$ , sans faire intervenir la notion de risque statistique. Cette notion présentant toutefois quelques avantages, nous avons adopté dans ce travail une approche intermédiaire qui consiste à tracer une courbe de fréquence cumulée des erreurs relatives de prédiction. Cette courbe donne directement un intervalle de confiance pour un risque donné sans faire intervenir explicitement de loi de probabilité: il s'agit d'un intervalle de confiance *expérimental*.

## **CHAPITRE III**

### **III- ACQUISITION ET TRANSFERT DES DONNEES**

### 3) ACQUISITION ET TRANSFERT DE DONNEES

Les spectres ont été réalisés dans le moyen infrarouge sur un spectromètre IFS 48 de la société SADIS BRUKER (voir figure 8).

Il s'agit d'un spectromètre de routine muni d'un compartiment échantillon purgé à l'air sec. Le spectromètre est équipé d'un détecteur MCT refroidi à l'azote liquide, la source est un globar.

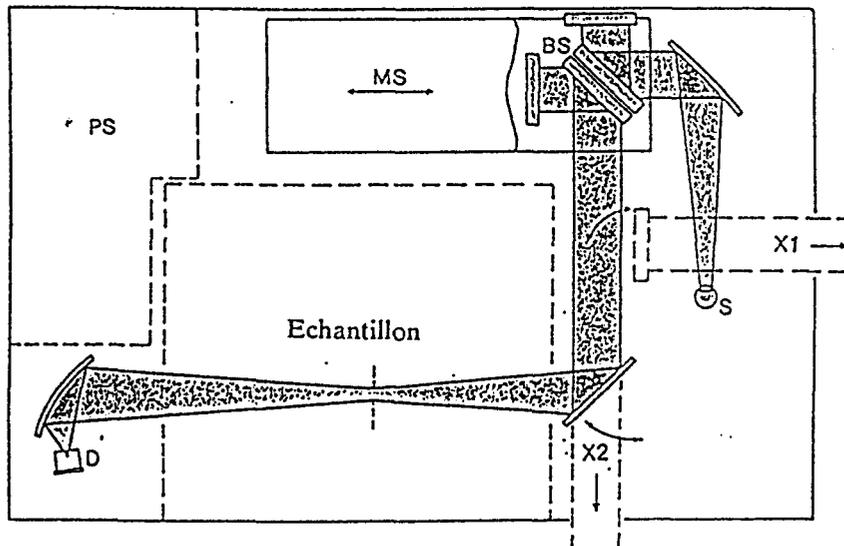
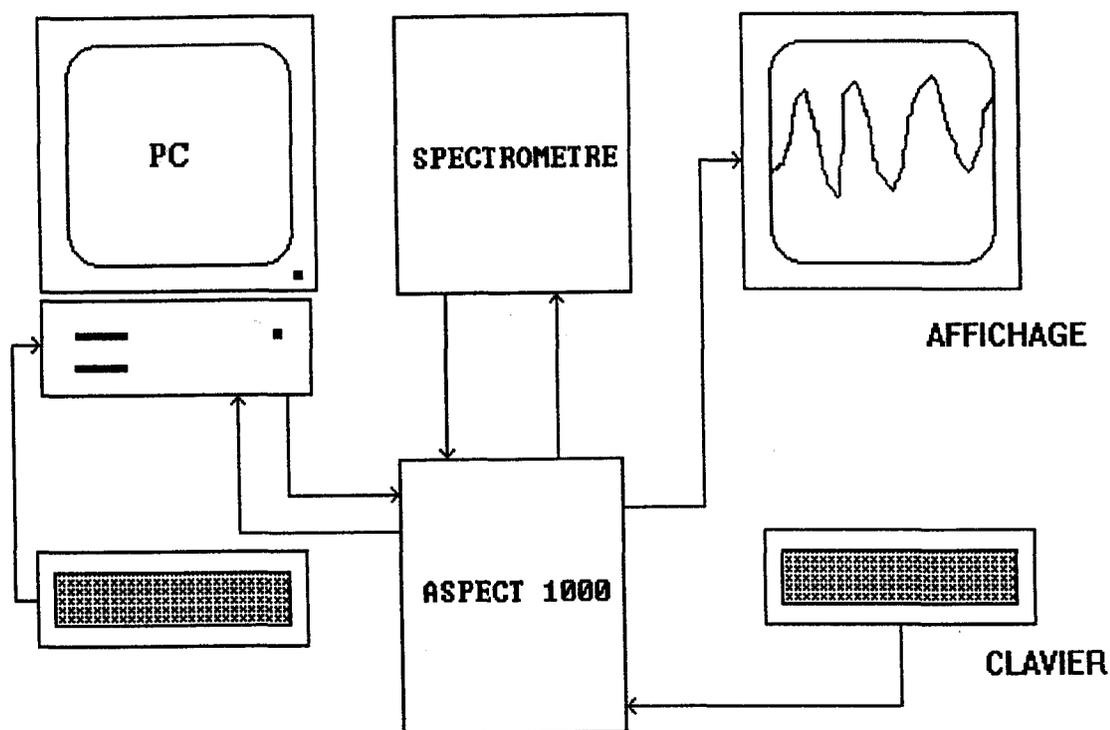


Figure 8: Schéma de l'IFS 48 (Document constructeur).

S: Source, BS: séparatrice, MS: miroir mobile, D: détecteur, PS: alimentation, X1, X2: faisceau externe

Le spectromètre est commandé par une informatique BRUKER dite Aspect 1000. La chaîne de mesure complète est représentée dans la figure suivante. Nous avons réalisé une liaison série entre l'Aspect et un IBM PC afin de pouvoir échanger des spectres dans les deux sens. Le transfert est assuré par un logiciel écrit en TURBO PASCAL et en assembleur que nous avons dénommé SPECTRA.



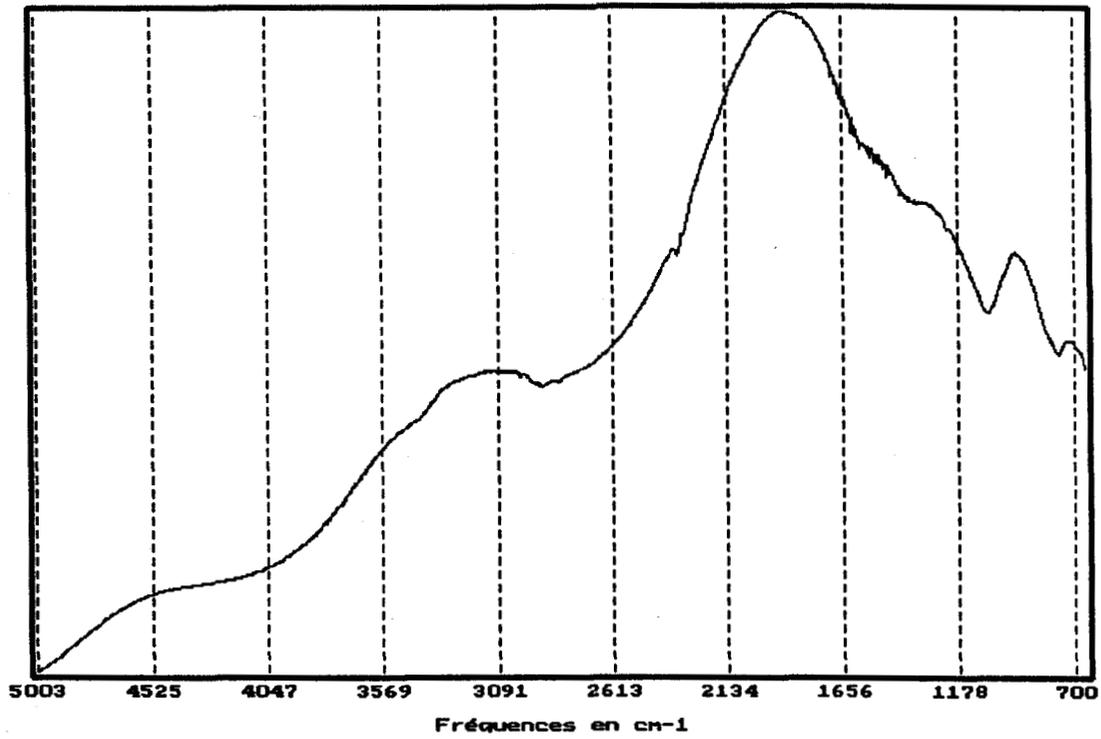
**Figure 9:** Schéma de la chaîne de mesure.

L'Aspect travaille en format binaire 24 bits. Puisqu'aucun compilateur sur PC n'est capable d'exploiter directement ce format, une option de SPECTRA permet d'effectuer la conversion des fichiers Aspect vers un format 16 bits de sorte à pouvoir être exploité plus facilement par la suite.

Sauf exception, les spectres ont été réalisés à une résolution de  $4 \text{ cm}^{-1}$ . Un mauvais rapport signal/bruit est une cause importante d'interférences et de non linéarité dans les problèmes de dosage. Les spectromètres IRTF offrant la possibilité de faire des acquisitions multiples, on montre que le rapport signal/bruit augmente comme la racine carrée du nombre d'acquisitions.

Toutefois l'accumulation d'un grand nombre d'interférogrammes se traduit très rapidement par des durées d'analyse importantes. Un compromis fréquemment choisi se situe entre 200 et 300

accumulations. Cette valeur dépend surtout du type d'échantillon étudié. Si le rapport signal/bruit est mauvais, on devra augmenter le nombre d'acquisitions.



**Figure 10:** spectre simple faisceau de la source (MCT).

## **CHAPITRE IV**

### **IV- ETUDE CRITIQUE DE DEUX METHODES D'ECHANTILLONNAGE**

- 1. La réflexion diffuse**
- 2. Réflexion totale atténuée**

## 4) ETUDE CRITIQUE DE DEUX METHODES D'ECHANTILLONNAGE

### 4.1) LA REFLEXION DIFFUSE

---

Les techniques de réflexion diffuse sont connues depuis longtemps en UV et dans le visible. Elles ont également été étendues très tôt au proche infrarouge.

Pour des raisons technologiques elles ont par contre longtemps été ignorées dans le moyen et le lointain infrarouge. Ce n'est qu'avec le développement de l'IRTF et des travaux de Griffiths<sup>(58)</sup> que cette spectroscopie a pu démontrer ses potentialités dans le moyen infrarouge pour l'analyse qualitative et quantitative.

Le principal intérêt qui lui fut reconnu est un échantillonnage plus simple par rapport aux méthodes classiques comme le pastillage. D'autre part la réflexion diffuse n'impose pas de contraintes thermodynamiques importantes aux matériaux, il y a donc peu de risque de réactions chimiques parasites.

Toutes ces qualités font de cette technique une méthode *a priori* intéressante pour l'analyse quantitative des matériaux solides.

#### 4.1.1) INTRODUCTION THEORIQUE

On distingue deux spectrométries de réflexion selon que la lumière pénètre ou non à l'intérieur du matériau.

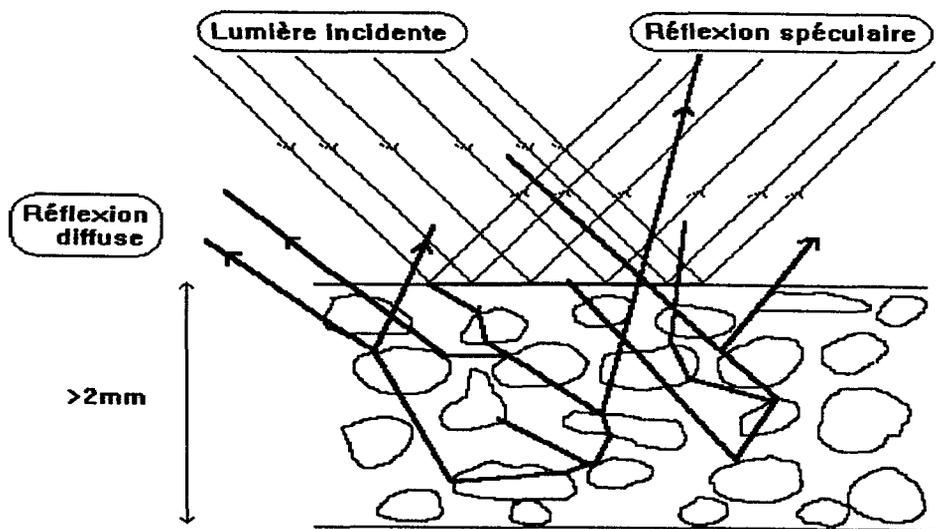
Dans ce dernier cas on parle de réflexion de Fresnel qui est décrite par les lois du même nom. Celle-ci est dite spéculaire lorsque la lumière incidente  $I_0$  est réfléchie dans le plan d'incidence et selon le même angle que la lumière incidente (direction spéculaire) et diffuse lorsqu'elle est réfléchie dans toutes les directions. Si  $I$  désigne la lumière réfléchie, on définit la réflectance  $R$  par:

$$R = I/I_0 \quad (\text{Eq 35})$$

Pour des particules significativement plus petites que la longueur d'onde du faisceau incident, la réflectance n'est plus décrite par les équations de Fresnel car la lumière incidente peut pénétrer dans le matériau et y subir simultanément des réfractions, des diffractions et des absorptions selon la loi de Beer-Lambert (voir figure 11). La lumière ainsi réfléchi est recueillie dans toutes les directions: on parle de spectrométrie de réflexion diffuse.

Si la taille des particules demeure plus large que la longueur d'onde de la radiation une partie de la lumière est réfléchi sur les faces orientées au hasard des grains. La réflectance diffuse est donc la contribution de deux composantes distinctes: la réflectance de Fresnel (dite encore réflectance spéculaire) et la réflectance de Kubelka-Munk. La contribution de ces deux phénomènes est ce que l'on nomme communément la réflexion diffuse.

La probabilité pour que la radiation pénètre dans l'échantillon et ressorte en n'ayant subi que des réflexions de Fresnel est faible, la réflexion spéculaire est donc essentiellement un phénomène de surface.



**Figure 11:** Les deux types de réflexion sont schématisés. L'échantillon doit avoir au moins 2 mm d'épaisseur pour que les conditions de Kubelka-Munk soient réalisées.

Ces deux composantes obéissent à des lois physiques qui leurs sont propres. Il est donc nécessaire de pouvoir en éliminer l'une ou l'autre lors d'applications pratiques: un spectre dans lequel l'une de ces composantes n'est pas nettement prépondérante est à notre connaissance inutilisable.

La composante diffuse est régie, à l'échelle microscopique, par la loi de Beer-Lambert. Ce n'est plus le cas au niveau macroscopique car le chemin optique de la radiation n'est plus défini. Le problème de modélisation mathématique du phénomène a donné lieu à de nombreux travaux dont on peut trouver un aperçu dans un article de Hecht<sup>(59)</sup>.

La théorie la plus simple et la plus employée (dans le moyen infrarouge) est celle de Kubelka-Munk<sup>(60)</sup>: pour un milieu isotrope, homogène de profondeur infinie depuis l'interface air/solide (semi-infini) et *en l'absence de réflexion spéculaire*, la réflectance  $R_\infty$  est liée au coefficient d'absorption  $K$  et au coefficient de diffusion  $S$  par:

$$F(R_\infty) = (1 - R_\infty)^2 / 2R_\infty = K/S \quad (\text{Eq 36})$$

$F$  est appelée fonction de Kubelka-Munk.

Aucun diffuseur idéal n'est connu dans le moyen infrarouge, on mesure donc en fait le rapport:

$$R_{\text{exp}} = I_m / I_{\text{ref}} \quad (\text{Eq 37})$$

$I_m$  étant le spectre simple faisceau du mélange étudié, tandis que  $I_{\text{ref}}$  est celui d'un matériau de référence le plus proche possible d'un diffuseur idéal.

On a alors:

$$R_\infty = I_m / I_0 = (I_m / I_{\text{ref}}) \cdot (I_{\text{ref}} / I_0) = R_{\text{exp}} \cdot R_{\text{ref}}$$

$$R_{\text{exp}} = I_m / I_{\text{ref}} \text{ est la réflectance relative de l'analyte.}$$

$$R_{\text{ref}} = I_{\text{ref}} / I_0 \text{ est la réflectance absolue de la référence.}$$

La fonction de Kubelka-Munk s'écrit alors:

$$F(R_\infty) = (1 - R_{\text{exp}} \cdot R_{\text{ref}})^2 / 2R_{\text{exp}} \cdot R_{\text{ref}} \quad (\text{Eq 38})$$

Si le matériau de référence est suffisamment proche d'un diffuseur idéal, alors  $R_{\text{ref}}$  peut être choisie égale à 1. Reinecke et al.<sup>(61)</sup> exposent dans leur article une méthode simple pour calculer  $R_{\text{ref}}$  au

niveau d'une bande d'absorption. Un des diffuseurs les plus utilisés est KBr qui sert également à diluer l'analyte.

Un spectre exprimé en unité de Kubelka-Munk présente la même allure qu'en unité d'absorbance (à condition que la contribution de la réflexion spéculaire soit négligeable). En rapprochant (Eq 36) de (Eq 4) on remarque que  $1/S$  s'interprète simplement comme l'équivalent du chemin optique de la loi de Beer-Lambert. Des expériences menées dans l'UV ont montré que  $1/S$  était sensiblement proportionnel à la taille moyenne des particules<sup>(62)</sup>. Il est donc logique de rechercher une relation du type loi de Beer-Lambert entre  $F(R_{\infty})$  et la concentration.

Avec les notations suivantes pour l'analyte et KBr respectivement:

$k_a, k_{KBR}$  : coefficients d'absorption molaire

$s_a, s_{KBR}$  : coefficients de diffusion molaire

$C_a, C_{KBR}$  : concentrations molaires

Hattori et al.<sup>(63)</sup> écrivent:

$$F(R_{\infty}) = (k_a.C_a + k_{KBR}.C_{KBR}) / (s_a.C_a + s_{KBR}.C_{KBR}) \quad (\text{Eq 39})$$

En général on a  $s_{KBR} \gg s_a$ . Si de plus on suppose:

-  $C_a \ll C_{KBR}$

-  $C_{KBR}$  constant

Alors (Eq 39) se met sous la forme:

$$F(R_{\infty}) = A.C_a + B \quad (\text{Eq 40})$$

Avec:  $A = k_a/s_{KBR}.C_{KBR}$  et  $B = k_{KBR}/s_{KBR}$

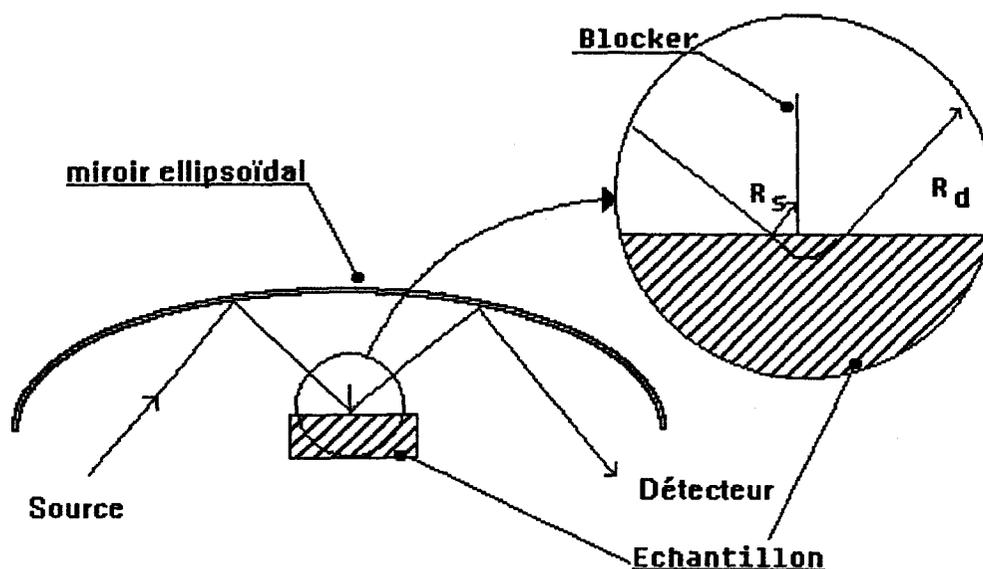
Cette relation montre comment utiliser les spectres de réflexion diffuse en analyse quantitative.

#### 4.1.2) INSTRUMENTATION

En toute rigueur, seule une sphère d'intégration permet d'obtenir un spectre de réflexion diffuse. Ce dispositif est très utilisé en UV, dans le visible et le proche infrarouge<sup>(64)</sup>.

Dans le moyen infrarouge il est techniquement très délicat à construire par manque de revêtements bien adaptés pour la sphère, de plus on est limité par la détectivité des détecteurs disponibles en MIR. De ce fait la sphère d'intégration est surtout réservée à des fins de standardisation de méthodes ou de matériaux.

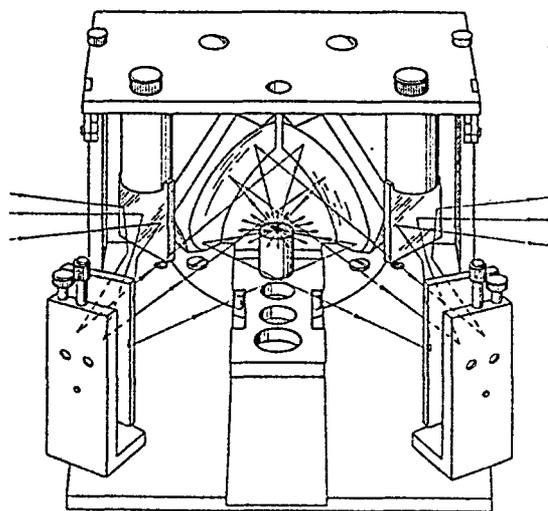
Depecker présente un certain nombre d'alternatives dans sa thèse<sup>(65)</sup>. Un des montages les plus simple est le "Collector" de BARNES (figure 12). Cet accessoire, relativement peu coûteux, est parfois complété par un dispositif dit "blocker", mis au point par BARNES<sup>(66)</sup>, sensé éliminer la réflexion spéculaire.



**Figure 12:** dispositif BARNES représenté avec son blocker.

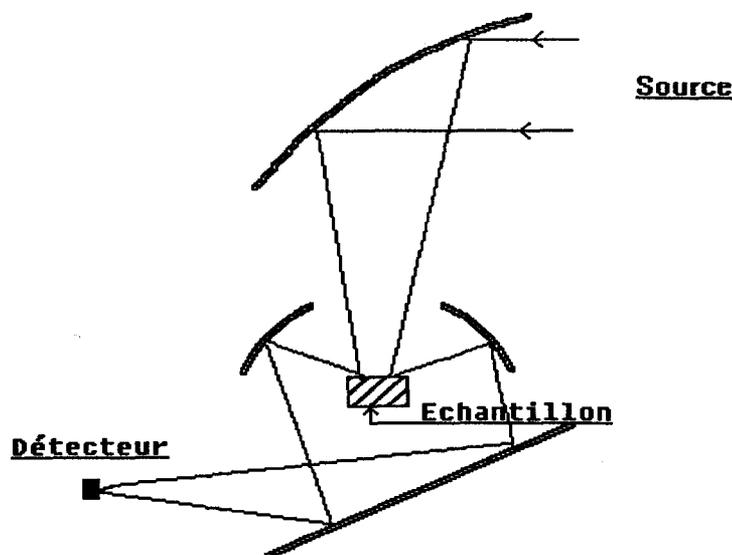
L'alignement de l'accessoire s'effectue en jouant sur les miroirs et sur la hauteur de la coupelle contenant l'échantillon, le miroir ellipsoïdal restant fixe. Sur certains modèles (le Baseline de SPECTRA-TECH) les miroirs et l'échantillon sont immobiles tandis que l'ellipsoïde seule se déplace en hauteur à l'aide d'un système à vis: cela en fait un accessoire très simple à régler et de ce fait bien adapté aux analyses de routine.

D'autres dispositifs plus complexes existent comme l'accessoire HARRICK (figure 13). L'idée consiste à collecter la lumière en dehors de la direction spéculaire ( $\rightarrow$  *off axis device*). Le dispositif est plus coûteux et beaucoup plus complexe à régler que le Collector.



**Figure 13:** montage HARRICK (*praying mantis*).

Un dernier accessoire intéressant est le montage de Fuller et Griffiths<sup>(58)</sup> (figure 14). Le faisceau arrivant perpendiculairement à l'échantillon, celui-ci élimine en théorie complètement la composante spéculaire. A notre connaissance il n'est pas commercialisé en raison peut-être d'un encombrement vertical qui rend difficile la construction d'un accessoire universel adaptable à tout type de spectromètre.



**Figure 14:** accessoire de Fuller et Griffiths.

Des variantes peuvent être imaginées avec l'ajout de polarisateurs croisés en entrée et en sortie pour filtrer la composante spéculaire<sup>(69)</sup>. Cette optique est peu utilisée car les accessoires de réflexion diffuse ne récupérant qu'une faible partie de la lumière incidente, même avec un détecteur MCT il devient difficile d'obtenir de bons spectres. Par exemple, l'optique HARRICK utilisée pour ce travail restitue sur poudre de KBr de l'ordre de 5% de l'énergie qui entre dans l'accessoire.

#### **4.1.3) OBTENTION DES SPECTRES INFRAROUGES PAR REFLEXION DIFFUSE**

Les hypothèses fondamentales de validité de (Eq 40) sont:

- \* échantillon semi-infini
- \* réflexion spéculaire négligeable
- \* coefficient de diffusion constant d'un point à l'autre de l'échantillon
- \* concentration en analyte pas trop importante

Expérimentalement les aspects à étudier sont:

- diminution de l'influence de la réflexion spéculaire
- contrôle de la linéarité de la loi de Kubelka-Munk

- contrôle du diamètre des particules
- contrôle de la concentration de l'analyte
- nature de la référence
- contrôle de la compacité du produit
- maîtrise de la répétabilité de l'échantillonnage

#### **4.1.3.1) REFLEXION SPECULAIRE ET NON LINEARITE**

La fraction de la réflectance totale que l'on peut attribuer à la réflexion spéculaire dépend des propriétés physiques du matériau (granulométrie...) ainsi que de l'optique de l'accessoire de réflexion diffuse<sup>(67,68,69)</sup>.

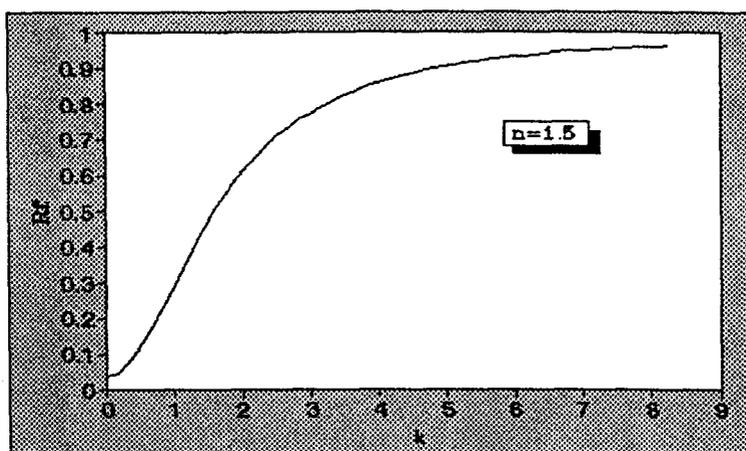
La présence de réflexion spéculaire peut se traduire par des inversions de bandes (→ *reststrahlen bands*), par de légers décalages des pics ou l'apparition de profils de dérivée première sur les spectres (→ *anomalous dispersion features*). Ces phénomènes peuvent être expliqués, au moins qualitativement, de la façon suivante:

Pour une radiation non polarisée de longueur d'onde  $\lambda$ , perpendiculaire à une surface plane, le taux de réflexion de Fresnel  $R_F$  est donné par la relation:

$$R_F = [(n-1)^2 + n^2k^2] / [(n+1)^2 + n^2k^2] \quad (\text{Eq 41})$$

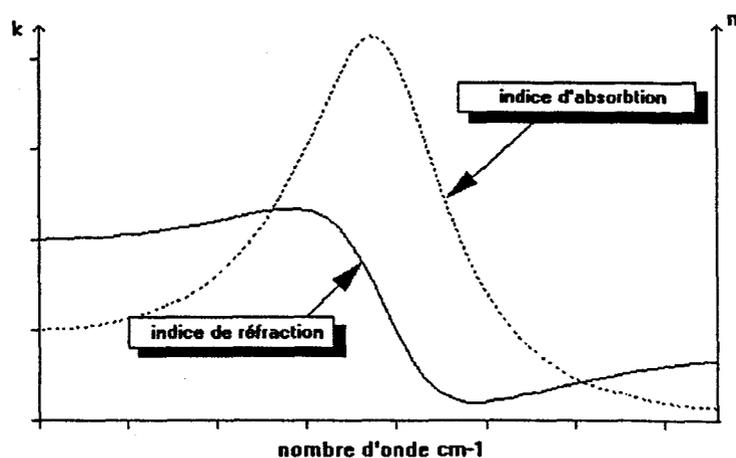
$n$ : indice de réfraction de l'échantillon

$k$ : indice d'absorption qui est proportionnel au coefficient d'absorption



**Figure 15:** variations de  $R_F$  en fonction de l'indice d'absorption. Le tracé est effectué pour  $n=1.5$ .

Pour les bandes à faible coefficient d'absorption, la réflexion spéculaire de surface est faible ( $R_F$  se rapproche du minimum de la courbe). L'allure de la bande suivra donc la loi de Kubelka-Munk. Lorsque le coefficient d'absorption augmente,  $R_F$  se rapproche de 1, la part de la lumière absorbée diminue donc au profit de celle directement réfléchi. Cela entraîne une diminution de la hauteur des bandes ne devant rien à des variations de concentration, le spectre tend donc à devenir non linéaire. Pour les valeurs élevées de  $k$  ( $R_F$  proche de 1) ces bandes peuvent finir par s'inverser ( $\rightarrow$  *reststrahlen bands*). Pour les coefficients d'absorption intermédiaires on obtient une combinaison de bandes d'absorption et de réflexion spéculaire. On sait qu'au niveau d'une bande d'absorption l'indice  $n$  varie très rapidement:



**Figure 16:** Allure en dérivée première de la dispersion anormale de l'indice de réfraction.

Ce phénomène est appelé dispersion anormale de  $n$  ( $\rightarrow$  *anomalous dispersion*). Il provoque donc des variations rapides de  $R_F$  (équation (Eq 41)) ce qui explique qualitativement l'apparition de figures de dérivée première ou de décalages en fréquence du pic d'absorption. Un tel décalage peut également intervenir en raison d'interactions chimiques entre les constituants.

#### **4.1.3.2) INFLUENCE DE LA DILUTION**

La dilution peut affecter la qualité du spectre de quatre façons:

- a) Certains détecteurs (comme les MCT) ne sont linéaires que sur une plage restreinte de signal. La dilution du matériau permet donc de s'y positionner.
- b) La validité de (Eq 38) n'est assurée qu'en milieu dilué.
- c) La dilution permet de diminuer l'effet de la réflexion spéculaire et ainsi d'améliorer la linéarité de la fonction de Kubelka-Munk<sup>(70)</sup>. Les figures 17-A à 17-D montrent le spectre de réflexion diffuse du saccharose pour différentes concentrations dans KBr. La figure 17-C montre une comparaison entre le spectre à 5% de saccharose avec un autre à 50%. La granulométrie est inférieure à 50  $\mu\text{m}$  pour chaque produit.

On voit que la forme des spectres dépend très fortement de la concentration et que l'importance des déformations n'est pas la même selon la fréquence observée. Ces déformations s'atténuent lorsqu'on dilue le sucre dans KBr et finissent par disparaître totalement vers 10%. Cette observation est générale et montre que la réflexion spéculaire provient principalement de l'analyte. En pratique on évite de travailler au-delà de 20% (en masse) lorsque l'analyte présente des bandes à forte absorbance.

On peut interpréter ce résultat de la façon suivante: en diluant, la proportion de KBr à la surface de l'échantillon augmente au détriment du saccharose ce qui diminue le taux de réflexion spéculaire due à l'analyte. En outre, on augmente la probabilité de rencontrer une

interface KBr/analyte et diminue celle de l'interface air/analyte. Puisque la différence entre l'indice de réfraction de l'analyte et celui de KBr est plus faible que celle entre l'air et l'analyte, le taux de réflexion de Fresnel diminuera donc également.

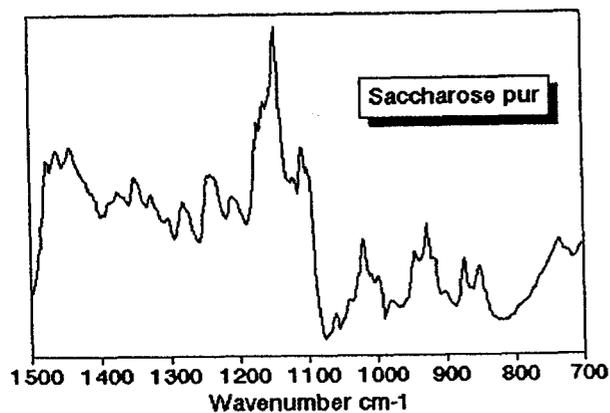


Figure 17-A

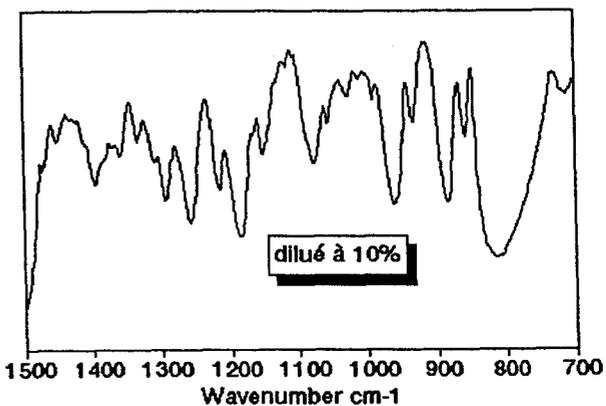


Figure 17-B

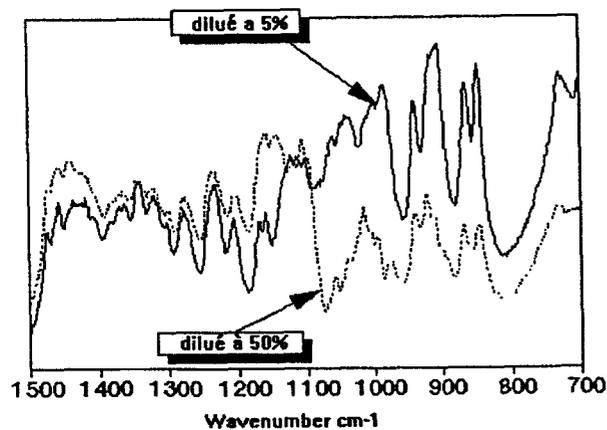


Figure 17-C

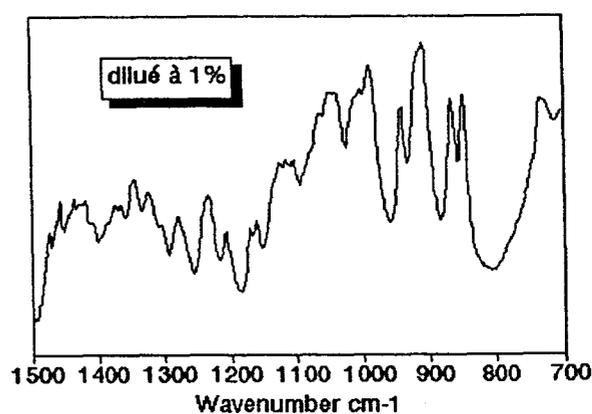
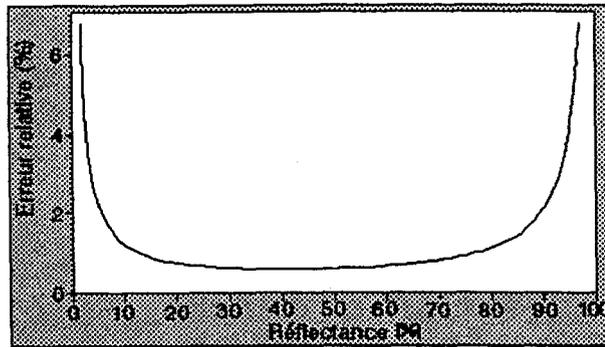


Figure 17-D

Spectre du saccharose pour différentes dilutions dans KBr.

d) La dilution permet également de se placer au minimum de l'erreur théorique de la fonction de Kubelka-Munk. Une incertitude sur la mesure de R (bruit...) produit sur  $F(R_\infty)$  une erreur relative:

$$\Delta F/F = [(1+R_\infty)/(1-R_\infty)] \cdot \Delta R_\infty/R_\infty$$

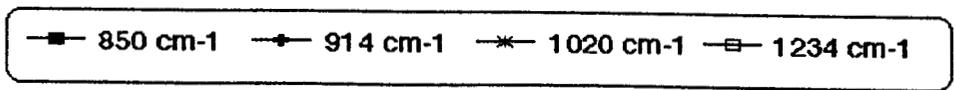
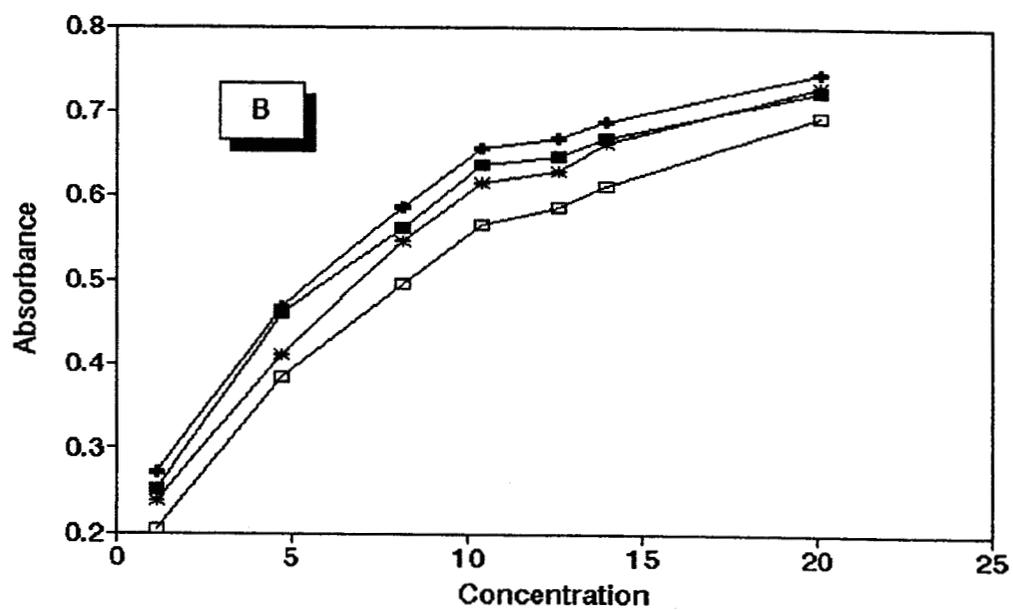
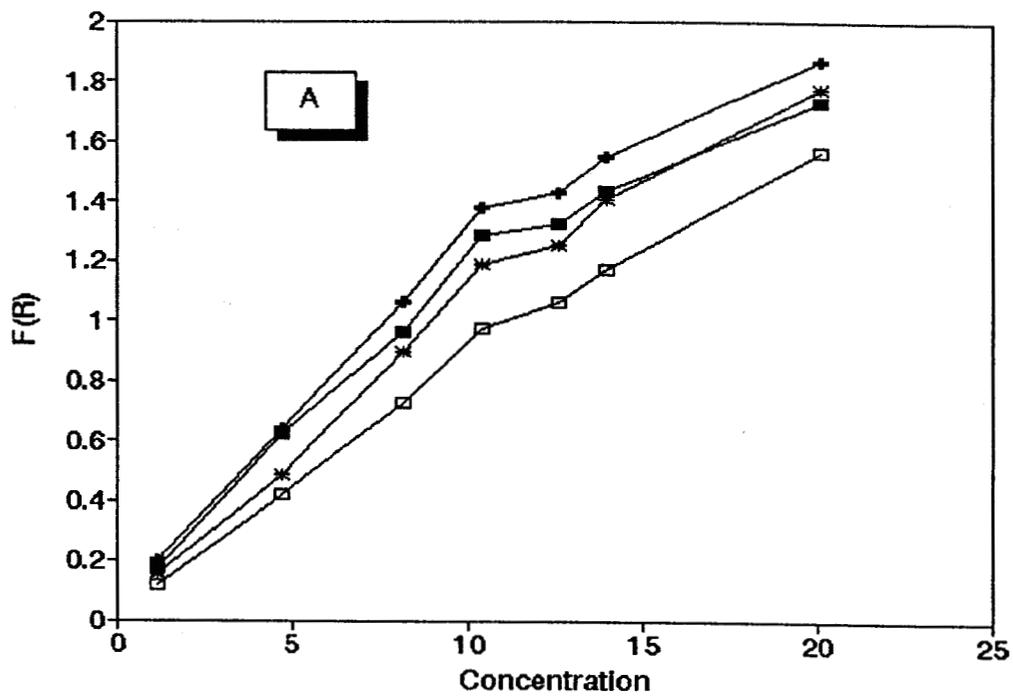


**Figure 18:** Tracé pour  $\Delta R = 10^{-3}$ . Une incertitude sur  $R_{\infty}$  produit un effet minimal sur  $F(R_{\infty})$  lorsque la réflectance est comprise entre 20 et 60%.

Cette courbe passe par un minimum pour  $R_{\infty} \approx 0,4$ . Elle montre que le domaine optimal se situe entre 20 et 60% en réflectance, soit 0.1 et 1.6 KM et qu'aux dilutions faibles ou bien élevées, la linéarité de la relation de Kubelka-Munk peut être affectée.

La figure 19 montre les variations de la hauteur de certaines bandes du saccharose dilué dans KBr en fonction de la concentration dans la plage 0-20%. Nous avons tracé la même figure en unité de Kubelka-Munk (figure 19-A) et en unité d'absorbance (figure 19-B).

La linéarité n'est toutefois pas totale et commence à diminuer à partir de 15% environ. Une compilation de différentes publications nous a montré que ce résultat est tout à fait général pour les matériaux organiques. Dans le cas de concentrations plus fortes ou bien de matériaux d'indice élevé, la cause principale de non linéarité est souvent la réflexion spéculaire. Brimmer et Griffiths<sup>(67,68)</sup> ont montré que l'utilisation d'une optique hors axe ( $\rightarrow$  *off axis*) ou bien de polariseurs permet d'étendre la zone de linéarité.



**Figure 19:** la transformation de Kubelka-Munk permet d'améliorer la linéarité des bandes avec les variations de concentration.

#### **4.1.3.3) NATURE DU DIFFUSEUR**

Le diffuseur idéal n'existe pas, pour s'en rapprocher un produit doit satisfaire aux contraintes suivantes:

- matrice non absorbante
- taux de réflexion de Fresnel inférieur à 4%
- indicatrice proche de celle du diffuseur isotrope
- chimiquement inerte

Dans le moyen infrarouge on travaille surtout avec les halogénures alcalins. On rencontre parfois des matériaux comme ZnSe, CaF<sub>2</sub>, Si, Ge ou la poudre de diamant<sup>(71)</sup>. Leur prix élevé les exclut des applications analytiques de routine. Toutefois leur emploi peut s'avérer nécessaire lorsque le produit étudié est susceptible d'échanges d'ions avec l'halogénure<sup>(72)</sup>.

Parmi les halogénures on choisit souvent KBr ou KCl. Le chlorure de potassium est plus coûteux que KBr ce qui limite son utilisation, par contre il a moins tendance à s'agglomérer. Certains auteurs ont testé des mélanges mixtes (KCl/CsI par exemple) sans obtenir de résultats meilleurs qu'avec KBr<sup>(73,74)</sup>.

Le tableau suivant résume les caractéristiques de quelques matériaux classiques.

	solubilité dans l'eau g/l à 25°C	région spectrale utile (cm <sup>-1</sup> )	Prix indicatif
Quartz	Non	10 000-5 000	*
Silicium	Non	8 300-660	**
Diamant	Non	MIR	****
KBr	654.5	40 000-340	*
NaCl	350.5	40 000-590	*
KCl	342.9	MIR	**
CsI	779.5	40 000-200	**
Polyéthylène	Non	600-30	*
ZnSe	Non	20 000-450	***
CaF <sub>2</sub>	2.4 10 <sup>-2</sup>	50 000-1140	**

Note: la région spectrale utile s'arrête lorsque le taux de transmission tombe à 45% pour une cellule constituée de 2 fenêtres de 2 mm d'épaisseur.

#### 4.1.3.4) INFLUENCE DE LA GRANULOMETRIE

La réflexion spéculaire est favorisée par la présence de surfaces de dimensions supérieures à la longueur d'onde de la radiation. Il est donc indispensable de broyer et tamiser les produits avant leur utilisation.

La figure 20 montre l'évolution du spectre du saccharose pour différentes granulométries moyennes de KBr et du sucre. Dans chaque cas la concentration est de 4.7% en masse. Deux répartitions granulométriques sont représentées. Elles sont repérées par le lettres A (0-50 μm) et B (50-90 μm). La première lettre correspond à la granulométrie du KBr.

On constate que l'intensité des bandes diminue fortement lorsque la taille moyenne des grains de l'analyte augmente, mais elle varie peu en fonction de celle du diffuseur. Ce qui montre l'importance de surveiller la granulométrie de l'analyte.

De façon qualitative, lorsque le diamètre moyen des particules de saccharose augmente, la distance moyenne entre les grains s'accroît puisque la concentration reste inchangée. Par conséquent la radiation traversera statistiquement moins de saccharose, la hauteur des bandes du sucre diminuera donc (figure 20-A,B).

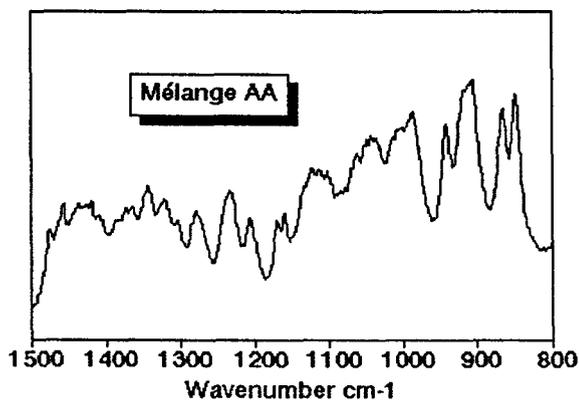


Figure 20-A

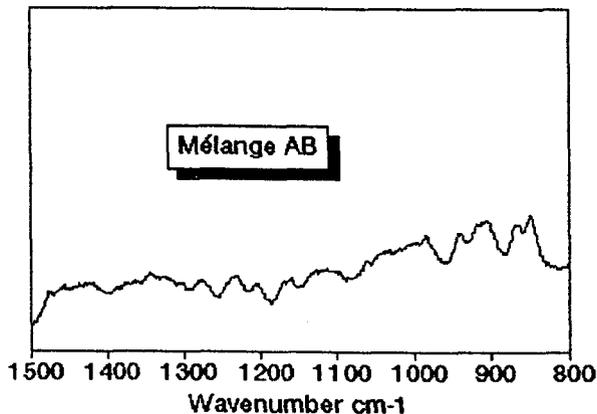


Figure 20-B

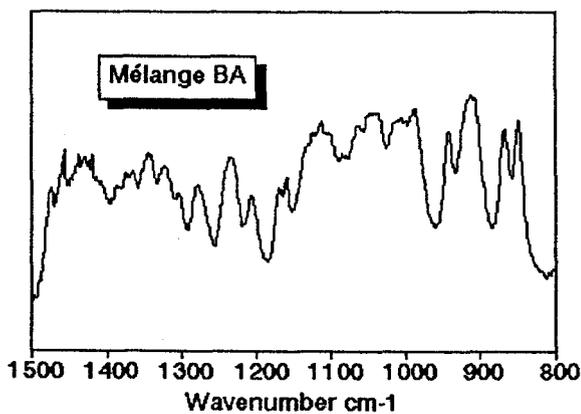


Figure 20-C

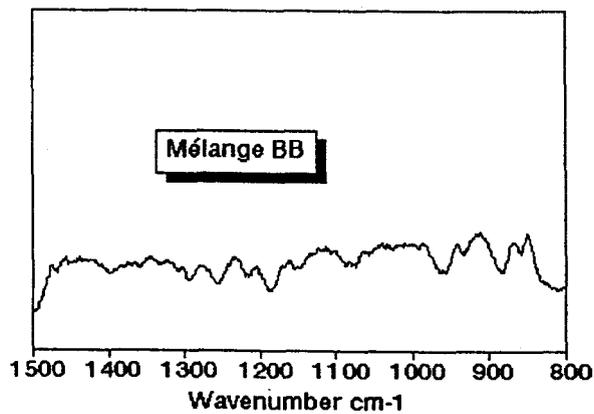


Figure 20-D

Effet de la granulométrie sur les spectres de réflexion diffuse.

(Les tracés sont réalisés à la même échelle pour chaque figure)

La distribution statistique de la taille des particules autour de la valeur moyenne joue sur la valeur du coefficient de diffusion. Pour maintenir une répartition reproductible il est préférable d'éviter le broyage manuel.

Pour de petites quantités nous avons utilisé un broyeur à billes. L'obtention d'une granulométrie reproductible nécessite un bon contrôle de la durée du broyage, de la fréquence de vibration et du remplissage (pas plus de 30 à 40% du mortier).

Nous avons également testé l'utilisation d'un broyeur électroménager KRUPS qui permet un broyage plus rapide des produits. Pour 60g de KBr nous obtenons typiquement la courbe de répartition suivante au bout d'une minute:

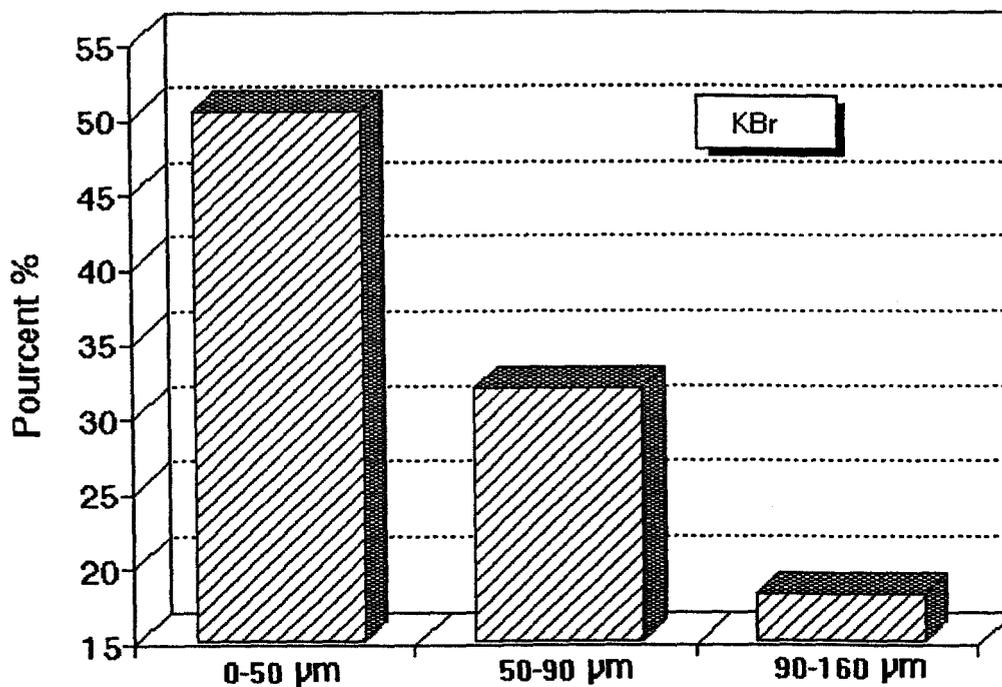


Figure 21: répartition de la granulométrie de KBr après broyage avec un broyeur électroménager.

Pour les sucres on trouve de l'ordre de 30% dans chaque catégorie.

Les essais réalisés montrent qu'en aucun cas on ne doit broyer le sucre et le KBr simultanément. Le sucre est plus mou que le KBr et de ce fait a tendance à rester collé sur les parois

du broyeur ce qui rend difficile le contrôle de la concentration. Le mode opératoire que nous avons adopté consiste donc à broyer les produits séparément puis à mélanger manuellement des masses bien déterminées de produit. Le problème de la préparation des produits pour la réflexion diffuse a été étudié plus en détail par Hamadeh et *al.*<sup>(75)</sup>.

Quelle que soit la technique de broyage, un tamisage est indispensable pour éliminer complètement les particules résiduelles de gros diamètre.

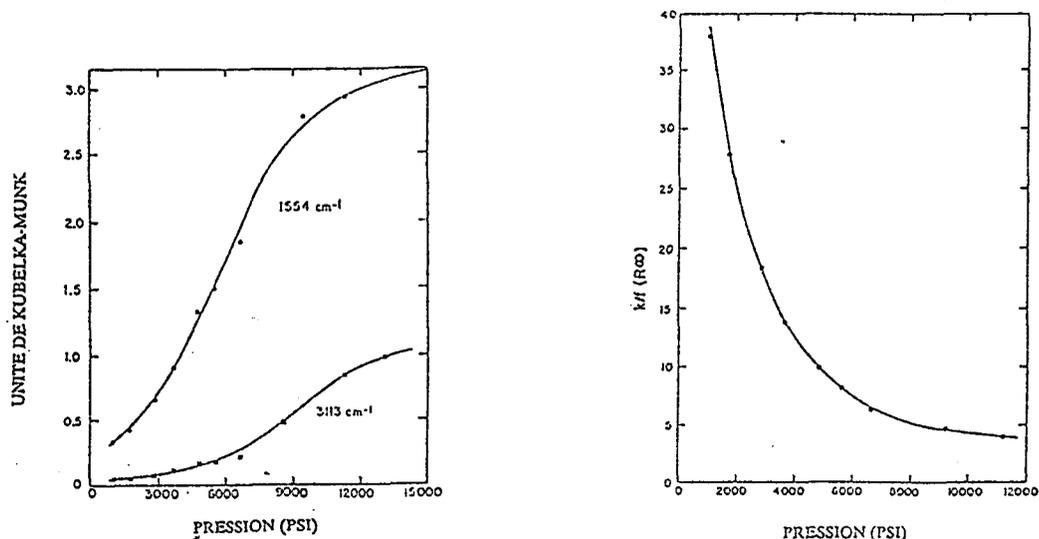
Le tamisage est une opération relativement longue ce qui est gênant pour les produits hygroscopiques. Expérimentalement nous avons trouvé qu'un mouvement circulaire des tamis dans le plan horizontal est le plus efficace en terme de rapidité: les 60 g précédents sont tamisés en moins d'une demi-heure contre une demi-journée sur une tamiseuse à vibrations.

Les produits sont systématiquement déshydratés à l'étuve pendant 24h avant toute utilisation (120 °C pour KBr et 60 °C pour les sucres).

#### **4.1.3.5) COMPACTITE ET REPETABILITE**

Dans une expérience de réflexion diffuse nous définissons la compacité ( $\rightarrow$  *packing*) comme la quantité de matière par unité de volume dans la coupelle.

Expérimentalement on peut agir sur la compacité en appliquant une pression variable sur l'échantillon de façon à le tasser plus ou moins, mais sans aller jusqu'au pastillage du produit. Yeboah et *al.*<sup>(76)</sup> sont à notre connaissance les premiers à avoir étudié de façon systématique l'effet de la compacité sur les spectres de réflexion diffuse. Pour la bande à  $1554\text{ cm}^{-1}$  de la caféine diluée à 0.2% dans KCl ils ont obtenus les courbes suivantes:

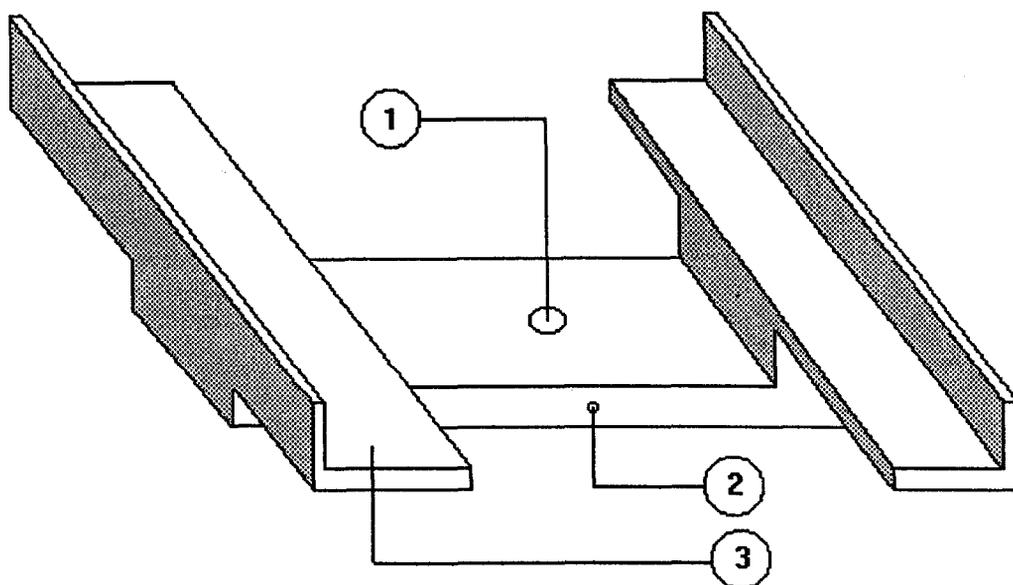


**Figure 22:** A gauche, variation de la hauteur de deux pics caractéristiques avec la pression. A droite, variation du coefficient de diffusion en fonction de la pression appliquée (pic à  $1554\text{ cm}^{-1}$ ).

Les échantillons sont normalement introduits dans la coupelle à l'aide d'une spatule, le surplus étant ensuite aplani. Cette procédure peut créer localement des zones où la compacité est plus forte qu'ailleurs. Cela se traduira par des variations du coefficient de diffusion  $S$  au sein de la coupelle, d'où des difficultés pour obtenir des spectres suffisamment répétables pour l'analyse quantitative. Il est donc très important d'effectuer un remplissage soigné de la coupelle. Même ainsi il est difficile d'obtenir des performances comparables sur une longue durée ou d'un manipulateur à l'autre. Lors de nos essais nous avons fréquemment même pas pu réussir de calibration avec un remplissage manuel des coupelles. Or, dans des applications pratiques il est indispensable de réussir à coup sûr le remplissage. C'est pourquoi il nous a semblé nécessaire d'étudier des montages réduisant au minimum le facteur humain.

Le premier montage étudié est inspiré d'une observation faite par Hembree<sup>(70)</sup> qui a noté que l'on peut améliorer la répétabilité en arasant le surplus de poudre avec une lame de rasoir glissant avec précaution et bien à plat sur le bord de la coupelle. Ainsi aucun effet de pression n'est à craindre. Des essais préliminaires ont révélé que cette procédure nécessite tout de même une

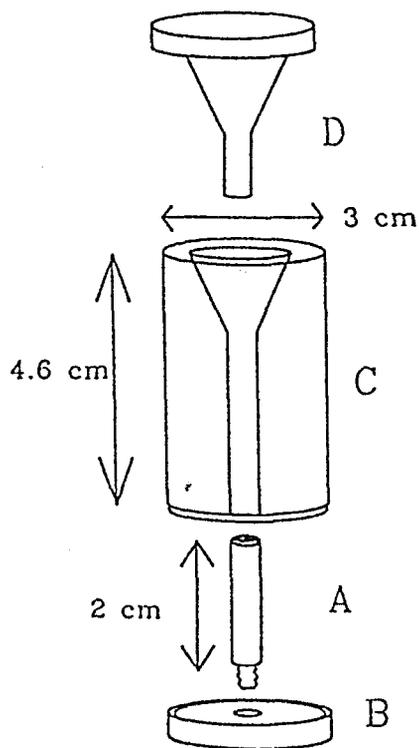
certaine dextérité pour réussir pleinement. Nous avons donc perfectionné la méthode en construisant le montage de la figure 23.



**Figure 23:** montage utilisé pour la méthode de la lame de rasoir.

Il s'agit d'une pièce métallique constituée d'une glissière (3) sur laquelle glisse la lame de rasoir. La coupelle vient se visser sur le filetage (1). A l'aide d'une petite traverse métallique posée sur la glissière, on peut ajuster la hauteur de la coupelle de sorte que son bord supérieur coïncide avec la hauteur de la glissière, la lame pourra ainsi araser le surplus en se déplaçant parfaitement à plat sur le rebord de la coupelle. Celle-ci peut être bloquée en introduisant une vis en nylon (pour ne pas endommager le filetage de la coupelle) dans (2). Nous avons eu des difficultés pour lisser le produit avec la lame sans que des cavités se forment à la surface de l'échantillon. Ce problème fut résolu en saupoudrant avec un tamis la poudre dans la coupelle. Cette procédure assure en outre un remplissage homogène de la coupelle, mais entraîne une perte importante en matériau.

Le second montage étudié est dérivé de celui présenté par TeVrucht<sup>(77)</sup> et fut utilisé par Yeboah pour ses travaux (figure 24).

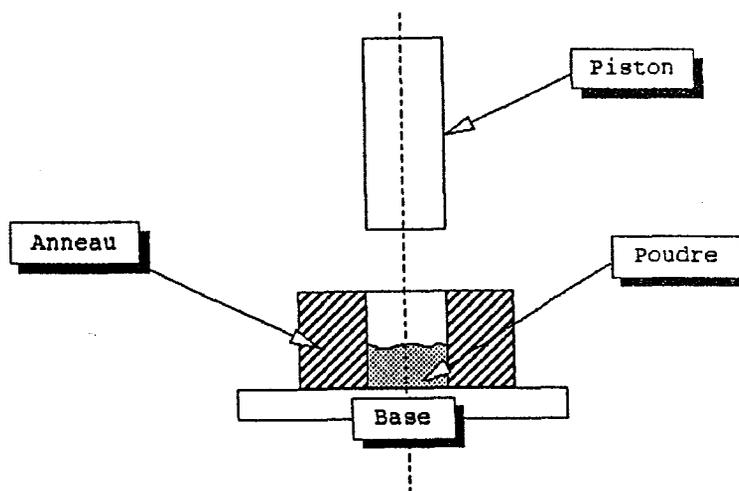


**Figure 24:** Montage de Griffiths pour l'échantillonnage en réflexion diffuse.

La coupelle (A) est vissée sur la base (B) puis insérée dans le cylindre (C). Après que la poudre ait été placée dans la coupelle, on introduit le couvercle (D) dans (C). A l'aide d'une presse il ne reste plus qu'à appliquer une pression connue sur l'échantillon par l'intermédiaire de (D). Ce montage nous semble toutefois peu efficace. Tout d'abord il présente une grande difficulté de fabrication. L'usinage doit être très précis pour que le couvercle (D) puisse s'appuyer de façon parfaitement horizontale sur le produit et sans offrir de jeu, autrement il devient impossible d'obtenir une surface régulière. En outre, il est indispensable de connaître avec précision la

quantité de produit nécessaire pour que, après le pressage, le bord supérieur de l'échantillon coïncide avec celui de la coupelle: s'il est trop haut on ne pourra pas démouler sans arracher le produit, s'il est trop bas on ne pourra pas prendre le spectre. Nous avons donc rapidement abandonné ce montage pour un autre ne présentant pas ces défauts.

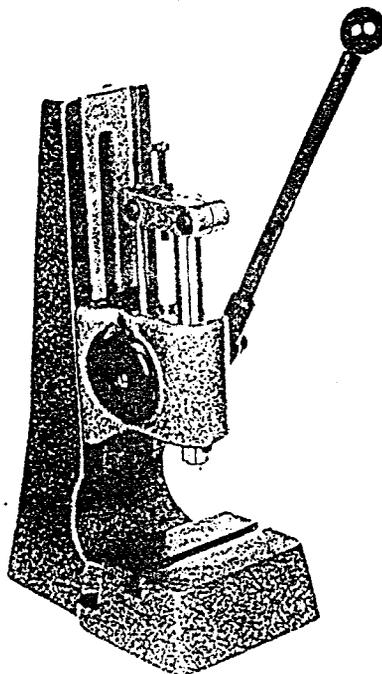
Le nouveau dispositif est constitué d'un simple anneau avec un perçage central de 10 mm (figure 25). L'anneau est posé sur un support lisse (la base). On le remplit alors d'une quantité suffisante de produit pour qu'après le pressage la galette obtenue fasse au moins 2 à 3 mm d'épaisseur. Un piston cylindrique en acier, fixé sur la broche de la presse, est alors inséré dans l'orifice central et peut transmettre une pression sur l'échantillon lorsque le bras de la presse est abaissé.



**Figure 25:** méthode de la presse, le piston est actionné par une presse à bras.

Si la pression est suffisante (quelques kilos) l'échantillon prend suffisamment de consistance pour former une galette capable de se soutenir d'elle-même sans s'effondrer lorsque l'on retourne l'anneau. La face jusqu'alors en contact avec la base est utilisée pour prendre le spectre en réflexion diffuse.

Cette fois il n'est plus nécessaire de peser avec précision le produit. En outre, le niveau de l'échantillon dans son support (l'anneau) est automatiquement ajusté à celui du rebord sans risque de déformations de sa surface. D'autre part on élimine aussi la nécessité de devoir réajuster la hauteur de l'échantillon à chaque mesure ce qui est source de beaucoup d'erreurs puisque une variation de 1 mm suffit à modifier le signal de 20%. Ce montage est de plus potentiellement automatisable plus aisément que le précédent.



**Figure 26:** La presse HP250Z.

Pour ces essais nous avons utilisé la presse HP250Z de la société ACME équipée d'une broche à crémaillère afin de s'assurer que la force fournie soit indépendante de la longueur de la course.

Notre dispositif ne permet pas de mesurer la pression exercée sur l'échantillon. Nous avons utilisé pour ce travail la pression créée par une masse d'environ 2.2 Kg fixée à l'extrémité du bras de la presse. Il est important de s'assurer que le contact entre le piston et l'échantillon se fasse doucement.

Nous avons pu comparer la compacité obtenue pour chaque échantillon sur une série de 10 essais réalisés en continu, les uns à la suite des autres. Le produit utilisé était un mélange de saccharose (4.7%) dilué dans KBr, les produits étant broyés à moins de 50  $\mu\text{m}$ . A titre de comparaison la compacité est de typiquement l'ordre de 2.5  $\text{g}/\text{cm}^3$  pour une pastille.

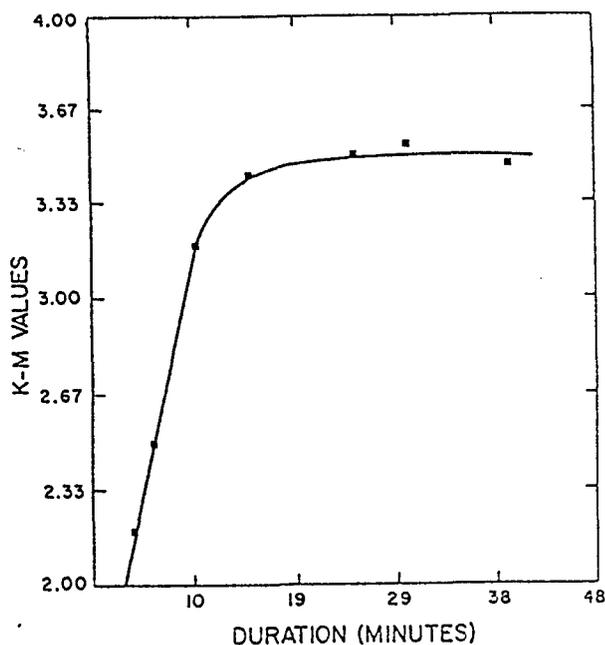
	Compacité (C) ( $\text{g}/\text{cm}^3$ )	S	S/C (%)
<b>Manuel</b>	0.74	0.01	1.9%
<b>Rasoir</b>	0.92	0.007	0.7%
<b>Presse</b>	1.68	0.012	0.7%

Note: S désignant l'écart type et S/C le coefficient de variation.

La comparaison des coefficients de variation montre que les deux montages spéciaux donnent des résultats nettement plus homogènes que le remplissage manuel, mais sans parvenir à se départager.

Le temps pendant lequel la pression est exercée a également de l'importance comme le montre la figure 27. La hauteur des bandes d'absorption augmente dans le temps, d'abord rapidement, puis se stabilise au bout de 15 minutes environ. Toutefois il n'est pas nécessaire d'attendre aussi longtemps, une durée intermédiaire est suffisante à condition qu'elle soit la même pour la référence et l'échantillon. Dans le cas d'un dispositif manuel, il est très difficile de mesurer

avec précision la durée pendant laquelle on presse. Pour cette raison nous avons adopté une durée assez longue de 3 minutes pour nos essais.



**Figure 27:** variation de la hauteur de la bande à  $1701\text{ cm}^{-1}$  de la caféine en fonction du temps de presse pour  $P = 400\text{ kg/cm}^2$ .

Nous avons ensuite étudié la répétabilité obtenue pour ces différents montages. Les produits utilisés sont le saccharose dilué dans KBr à 5%, tous deux sont broyés à moins de  $50\text{ }\mu\text{m}$ . La répétabilité peut être définie par le calcul du coefficient de variation sur la hauteur des bandes d'absorption (écart type rapporté à la moyenne). Les figures 28 et 29 sont obtenues à partir de la compilation des résultats obtenus sur une série de dix spectres en faisant varier la granulométrie de la même façon que pour les essais décrits au chapitre 4.1.3.4. En outre aucun lissage ni correction de ligne de base ou d'offset n'a été effectué. Les montages dits de la lame de rasoir et de la presse sont représentés. En ordonnée on lit le coefficient de variation calculé pour chaque fréquence et exprimé en pourcentage. Les mélanges utilisés sont les mêmes pour chaque montage. Pour faciliter les comparaisons, nous avons calculé les CV moyens pour chacune des 8 figures des pages 78 et 79, ces valeurs sont rassemblées dans le tableau suivant. Nous nous sommes de plus limité au domaine où le saccharose présente des bandes d'absorption.

	AA	AB	BA	BB
Rasoir	9.6%	18.6%	10.4%	15.0%
Presse	9.9%	17.3%	12.0%	12.9%

De ces différents éléments il ressort que les deux montages sont assez similaires du point de vue de leur répétabilité.

Mark et *al.*<sup>(78)</sup> ont présenté récemment une méthode destinée à améliorer la répétabilité des spectres. L'idée consiste à effectuer N spectres d'un même mélange et à utiliser le spectre moyen pour l'analyse. On peut espérer de cette façon moyennner certaines fluctuations. Il y a deux façons d'aborder cette méthodologie: soit on recommence le remplissage de la coupelle pour chaque spectre de la série ( $\rightarrow$  *repacking*), soit on se contente de repositionner la coupelle à chaque fois.

D'autres solutions ont également été testées pour essayer d'améliorer la répétabilité. Par exemple Velapoldi<sup>(79)</sup> propose de prendre les spectres de réflexion diffuse tout en faisant tourner l'échantillon de façon à moyennner les variations locales de compacité ou de l'état de la surface. De cette façon la précision serait améliorée d'un facteur 4 par rapport à un montage sans rotation.

METHODE DE LA LAME DE RASOIR

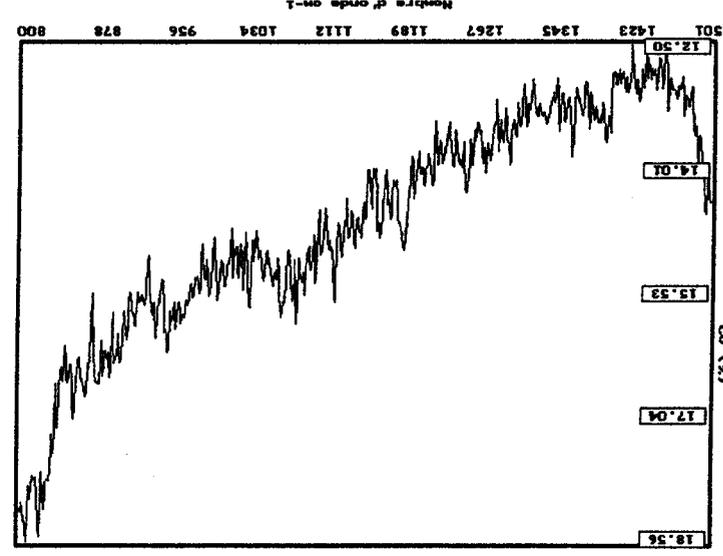
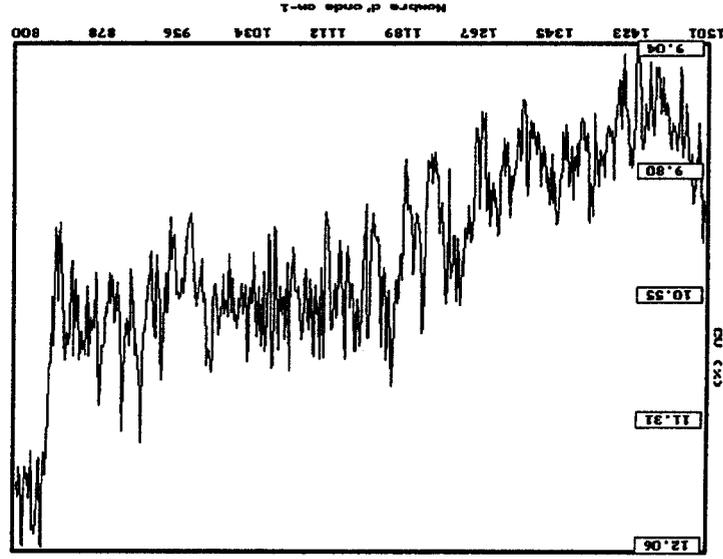
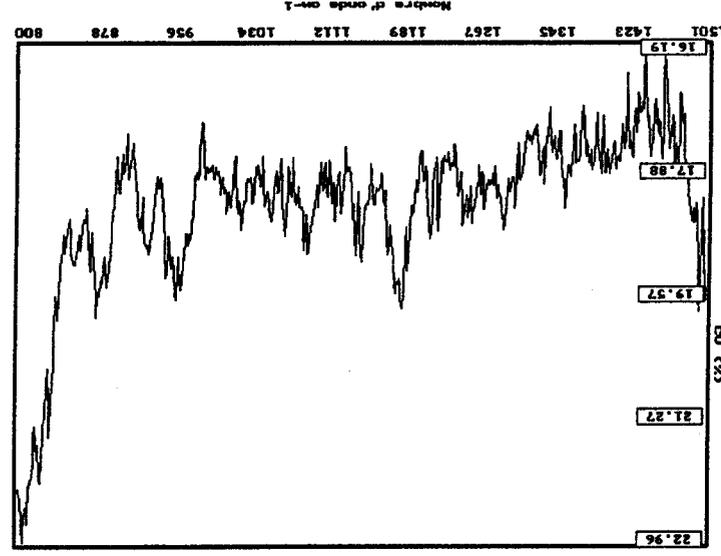
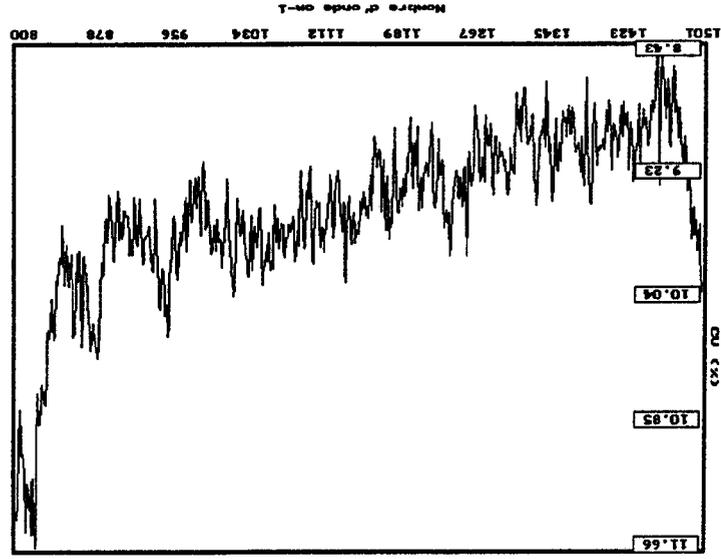


Figure 28: Coefficient de variation (répétabilité) en fonction de la fréquence.

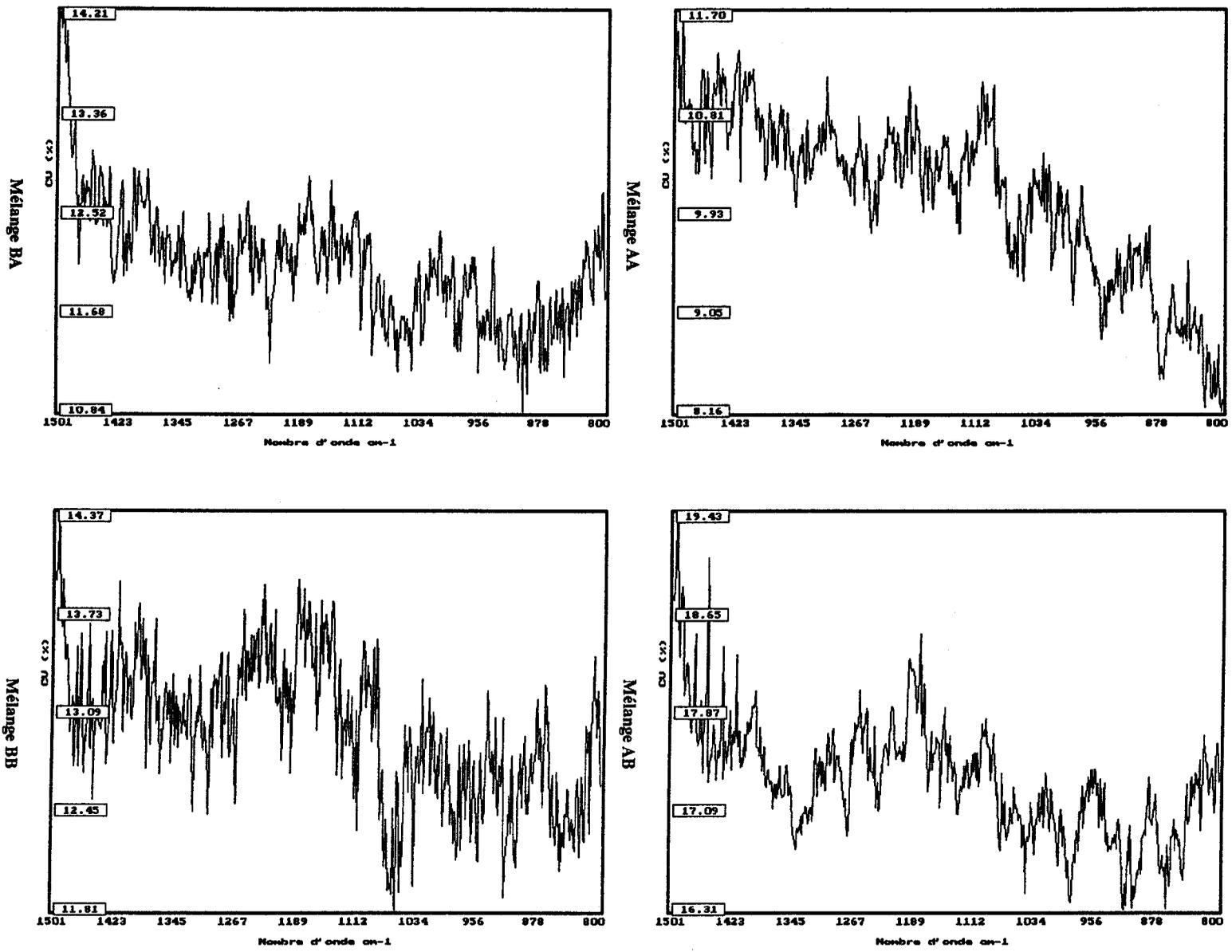
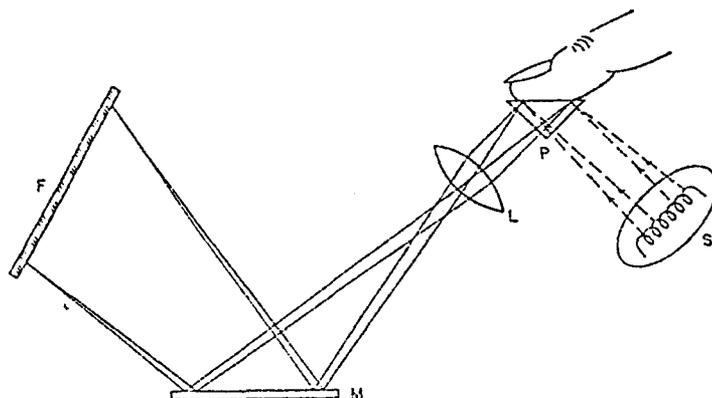


Figure 29: Coefficient de variation (répétabilité) en fonction de la fréquence.

## 4.2) REFLEXION TOTALE ATTENUÉE

### 4.2.1) INTRODUCTION THEORIQUE

Cette technique est parfois appelée réflexion interne ( $\rightarrow$  *Internal Reflexion Spectroscopy, IRS*). Lorsque la lumière passe d'un milieu  $M_1$  vers un milieu  $M_2$ , celle-ci est diffractée selon la loi des sinus. Si  $M_2$  est moins réfringent que  $M_1$ , alors il existe un angle  $\theta_c$ , dit angle critique, au delà duquel la lumière n'est plus réfractée, mais subit une réflexion totale. Contrairement aux apparences, cette réflexion n'est en réalité pas tout à fait totale. Ce phénomène peut être illustré par l'appareil décrit sur la figure 30.



**Figure 30:** Principe d'un appareil à relever les empreintes digitales.

Le doigt est pressé sur le prisme P éclairé par la source S. Une image de réflexion totale se forme sur la plaque photographique F par l'intermédiaire de la lentille L et du miroir M. D'après HARRICK<sup>(80)</sup>.

Contrairement à ce que l'on pourrait attendre si la réflexion était vraiment totale, la plaque photographique enregistre la trace des empreintes digitales du doigt.

Ce phénomène était déjà connu de Newton qui expliquait que la lumière pénètre dans  $M_2$ , puis retourne vers  $M_1$  selon une trajectoire parabolique due à l'attraction du milieu le plus dense. La réalité est toutefois plus complexe. A partir des équations de Maxwell, on établit l'existence d'une onde évanescente dans le milieu  $M_2$  qui est responsable du phénomène observé. Le milieu  $M_1$  est souvent appelé IRE ( $\rightarrow$  *Internal Reflexion Element*) dans la littérature. L'idée de

Fahrenfort consiste donc à mettre en contact un échantillon (milieu  $M_2$ ) avec un IRE d'indice plus élevé. Un montage analogue à celui de la figure 30 dans lequel on remplace la plaque photographique par un détecteur infrarouge permet d'obtenir des spectres ATR de l'échantillon.

On définit généralement la profondeur de pénétration  $d_p$  comme la distance nécessaire pour que l'amplitude de l'onde évanescente ne soit plus que  $1/e$  fois sa valeur à l'interface IRE/échantillon. On montre que pour un milieu  $M_2$  non absorbant:

$$d_p = \lambda / 2\pi n_1 (\sin^2 \theta - (n_2/n_1)^2)^{1/2} \quad (\text{Eq 42})$$

$\lambda$  est la longueur d'onde de la radiation,  $n_1$  et  $n_2$  sont les indices de l'IRE et de l'échantillon respectivement et  $\theta$  l'angle d'incidence.

D'autres définitions ont été proposées. Par exemple, Iwamoto<sup>(81)</sup> propose de définir la profondeur de pénétration par  $d_p/2$ .

Le tableau suivant donne les caractéristiques de quelques IRE fréquemment utilisés.

	Indice à $1000 \text{ cm}^{-1}$	$\theta_c$ ( $n_2 = 1$ )	$\theta_c$ ( $n_2 = 1.5$ )	Fréquence de coupure ( $\text{cm}^{-1}$ )
ZnSe	2.40	24.6	38.7	20 000
KRS-5	2.38	24.8	39.1	20 000
Ge	4.0	14.5	22.0	5 300
Si	3.4	17.1	26.2	8 000

**Note:** Les angles critiques sont calculés en degré.

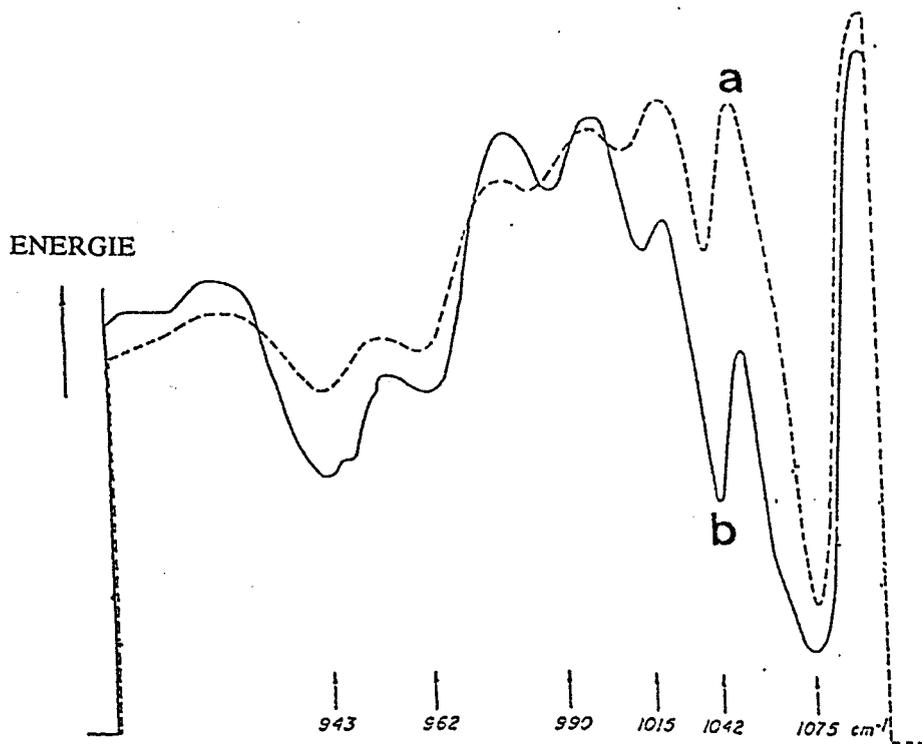
Ces matériaux sont tous insolubles dans l'eau à l'exception de KRS-5 qui présente une très faible solubilité, mais suffisante pour voiler le cristal à la longue. ZnSe est attaqué par les solutions acides et dégage  $\text{SeO}_2$  qui est toxique. L'emploi de liquides d'indice élevé a également été envisagé. L'intérêt en est très limité dans l'infrarouge.

La plupart des produits organiques ont un indice de réfraction de l'ordre de 1.5. Une application numérique de (Eq 42) avec  $\theta = 45^\circ$  et  $n_1 = 2.40$  (cristal de ZnSe) donne:

$$d_p \approx 0.20 \lambda$$

Seule une très fine couche de l'échantillon est donc concernée par le phénomène d'ATR. Ce résultat démontre l'intérêt de la technique pour les milieux fortement absorbants comme l'eau et soulève le problème du contact entre l'IRE et l'échantillon. Cette contrainte est surtout critique pour les poudres et les solides.

Il est possible d'augmenter le chemin optique en effectuant des réflexions multiples. Les accessoires les plus courants offrent jusqu'à 20 réflexions utiles. En ATR, le chemin optique est donc de l'ordre de la longueur d'onde étudiée. La figure 31 compare un spectre en transmission avec un spectre ATR. Fahrenfort a démontré que les spectres d'ATR exprimés en absorbance obéissent en général à la loi de Beer-Lambert. La technique est donc tout à fait praticable pour l'analyse quantitative et l'ATR est analogue à la spectrométrie de transmission pour ce qui concerne l'analyse qualitative. La profondeur de pénétration  $d_p$  est à l'ATR ce que l'épaisseur de la cellule est à la transmission. Il y a toutefois une différence importante puisqu'elle est proportionnelle à la longueur d'onde de la lumière. Tout se passe donc comme si l'on effectuait un spectre en transmission avec une longueur de cuve différente pour chaque fréquence. Pour les applications quantitatives il est possible d'effectuer une correction.

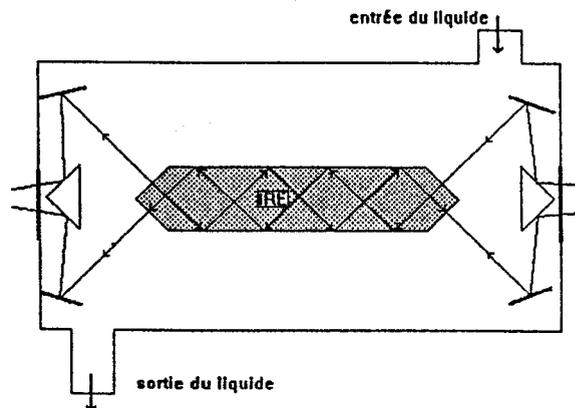


**Figure 31:** Spectre du dibutylphthalate.

A: en transmission, B en ATR (cristal d'AgCl)

#### 4.2.2) INSTRUMENTATION

De nombreux accessoires d'ATR ont été construits depuis 30 ans. HARRICK<sup>(80)</sup> en présente de nombreux dans son livre. Un des plus anciens est le montage dit CIRCLE (voir figure 32).

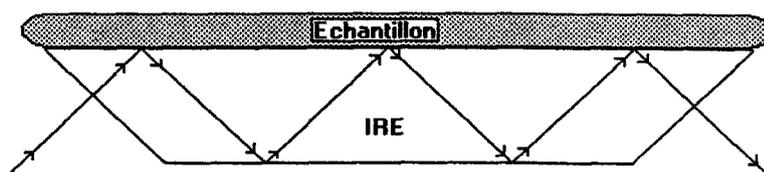


**figure 32:** Principe du CIRCLE.

L'IRE est un cylindre plongé dans une cuve contenant le produit à analyser. Le montage permet d'obtenir une très bonne sensibilité, mais il est coûteux et de maintenance difficile: l'étude de pâtes ou d'échantillons visqueux est impossible. De plus l'accessoire n'est pas adapté à l'analyse d'échantillons solides.

Il existe également des montages dits ATR vertical où le cristal est disposé verticalement en appui sur une plaquette métallique. Ce montage est particulièrement bien adapté pour l'étude d'échantillons solides, mais pas du tout pour les liquides.

Il y a deux ou trois ans est apparu sur le marché un ATR horizontal. L'IRE est posé horizontalement dans un cadre métallique en forme de cuvette (voir figure 33).



**Figure 33:** Principe de l'ATR horizontal.

L'échantillon liquide remplit la cuvette en recouvrant le cristal. On élimine ainsi les difficultés de nettoyage du CIRCLE et le coût de l'accessoire est aussi moins élevé. Le volume de liquide nécessaire pour recouvrir le cristal est de l'ordre de 2 ml, mais dépend beaucoup de sa viscosité. Dans la mesure où seule une très fine épaisseur est observée, il reste possible de travailler avec moins de solution si on l'étale en posant un support plat sur le cristal.

On peut utiliser l'accessoire pour les échantillons solides sous forme de poudre ou de film souple en montant une presse sur l'accessoire afin d'assurer le contact échantillon/IRE. A notre connaissance, les accessoires commerciaux sont tous conçus de façon à ce que le cristal ne soit maintenu que par ses bords. Il n'est donc pas possible d'appliquer une pression aussi importante que

pour les ATR verticaux sans risquer de le briser. On peut donc s'attendre à plus de difficultés pour obtenir de bons contacts avec un ATR horizontal.

#### Spectre de l'eau pure (ATR)

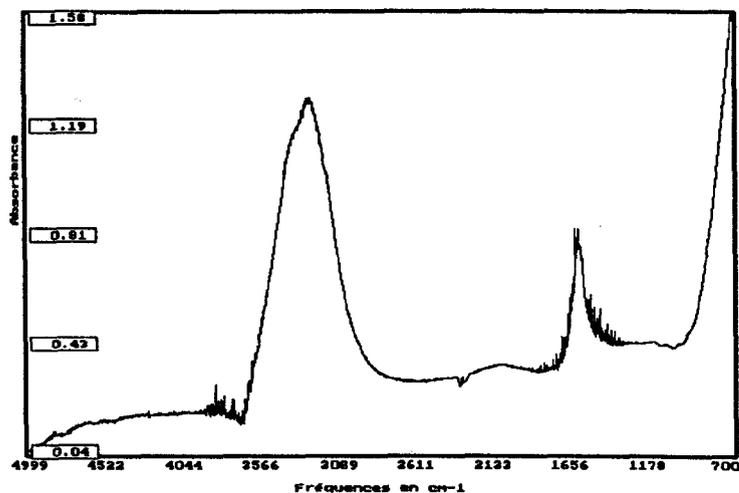


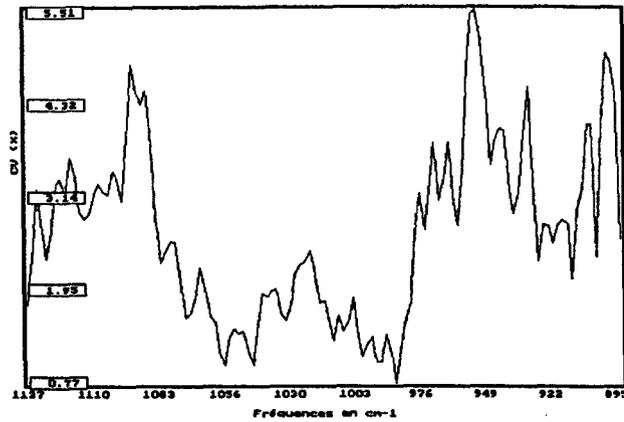
Figure 34: spectre de l'eau en ATR.

### 4.2.3) OBTENTION DE SPECTRES INFRAROUGES PAR ATR

Dans ce travail nous nous sommes limités à l'étude de l'ATR des liquides. L'accessoire utilisé est un ATR horizontal à 6 réflexions utiles (angle d'incidence 45°). L'échantillonnage par cette technique est particulièrement simple. Les aspects à surveiller plus particulièrement sont un bon recouvrement du cristal (éviter la formation de bulles d'air) ainsi que l'homogénéité et l'intégrité du liquide pendant la durée de l'analyse (dépôt de matières solides, décantation de phases liquides, évaporation du solvant...). Ces aspects sont trop spécifiques pour une étude générale de l'accessoire. Nous nous sommes donc limités à l'étude de la répétabilité afin d'effectuer une comparaison avec la situation déjà observée en réflexion diffuse.

#### 4.2.3.1) ETUDE DE LA REPETABILITE EN ATR DES LIQUIDES

Comme pour la réflexion diffuse, on définit la répétabilité par l'amplitude du coefficient de variation (CV) en fonction de la fréquence dans la gamme spectrale utilisée (voir figure 35).



**Figure 35:** Répétabilité de L'ATR en fonction de la fréquence observée.

Cette figure a été obtenue à partir de huit spectres de saccharose dilué dans de l'eau distillée. Nous nous sommes limité au domaine où les bandes du sucre n'interfèrent pas avec celles de l'eau.

La répétabilité de l'ATR est bien meilleure que celle de la réflexion diffuse. Le tableau suivant rassemble les CV moyens obtenus en ATR et en réflexion diffuse pour la meilleure configuration possible de chaque montage (spectres de type AA).

	<b>ATR</b>	<b>Presse</b>	<b>Rasoir</b>
<b>CV</b>	2.9%	9.9%	9.6%

## **CHAPITRE V**

### **V- APPLICATION A L'ANALYSE QUANTITATIVE**

- 1. Fonctionnement du logiciel de calcul**
- 2. La réflexion diffuse**
- 3. Analyses par ATR**
- 4. Application a des mélanges réels**

## 5) APPLICATION A L'ANALYSE QUANTITATIVE

Pour toute la suite, les spectres ont été réalisés pour une résolution de  $4 \text{ cm}^{-1}$  avec 300 accumulations pour la réflexion diffuse et 200 pour l'ATR.

Les sucres utilisés sont les suivants:

- \* D saccharose, A.C.S. Reagent, 99+%, ALDRICH
- \*  $\alpha$ -D glucose, A.C.S. Reagent, ALDRICH
- \* D fructose, A.C.S. reagent, 98%, ALDRICH
- \* maltose hydraté, 98%, SIGMA

Les spectres n'ont subi aucune correction de ligne de base.

## 5.1) FONCTIONNEMENT DU LOGICIEL DE CALCUL

---

Les traitements informatiques ont été assurés par un logiciel BRUKER implémentant la version présentée par Fredericks<sup>(62)</sup> de l'algorithme PCR sur l'Aspect 1000. Le modèle employé est l'équation transposée de (Eq 19), ce qui conduit à inverser le rôle des scores et des loadings par rapport à la version présentée au chapitre 2.7.4. On obtient ainsi une matrice de dispersion  $Z=Y.Y^t$  de dimension plus faible  $n \times n$  au lieu de  $p \times p$ . Cette présentation est donc particulièrement intéressante lorsque la recherche des composantes principales s'effectue en diagonalisant directement  $Z$ . Le logiciel permet de paramétrer:

- le nombre total de points utilisés sur les spectres (résolution).
- la dérivation (éventuelle) des spectres.
- une correction d'offset.
- le modèle additif ou non additif.
- la correction MSC.
- le choix automatique ou manuel des loadings.

Une fois la calibration effectuée, un test de Fisher est réalisé sur les coefficients de régression (la matrice  $B$  de l'équation (Eq 28)). Seuls les loadings dont les coefficients sont significativement différents de zéro sont retenus.

En fin de calibration, le logiciel édite un bulletin comprenant la concentration calculée, l'erreur relative (%), les  $n$  valeurs propres de la matrice de dispersion, les coefficients de régression (matrice  $B$ ) et la liste des loadings pour chaque constituant. Les tableaux suivants montrent un exemple de calibration obtenu sur une solution de saccharose, glucose et fructose à 300 g/l en ATR. Nous avons utilisé 8 standards, mais le logiciel peut traiter simultanément jusqu'à 25 spectres (calibration ou prédiction). Les concentrations sont exprimées en g/l.

Echantillon	Valeur réelle	Valeur calculée	Erreur (%)
1	68.57	68.31	0.293
2	77.14	77.23	0.146
3	85.71	85.54	0.170
4	94.29	94.46	0.262
5	102.86	102.43	0.392
6	111.43	111.26	0.099
7	128.57	128.74	0.156
8	72.86	73.11	0.412
<b>SEE</b> <b>Nb loadings</b>	0.38 4		

**Saccharose**

Echantillon	Valeur réelle	Valeur calculée	Erreur (%)
1	137.14	138.00	0.635
2	85.71	84.77	1.016
3	120.00	119.74	0.209
4	128.57	127.54	1.100
5	94.29	96.34	2.223
6	102.86	102.69	0.111
7	111.43	110.91	0.398
8	107.14	106.80	0.312
<b>SEE</b> <b>Nb loadings</b>	1.35 3		

**Glucose**

Echantillon	Valeur réelle	Valeur calculée	Erreur (%)
1	94.29	94.29	0.002
2	137.14	137.66	0.401
3	94.29	94.20	0.028
4	77.14	76.97	0.175
5	102.86	102.00	0.820
6	85.71	85.46	0.763
7	60.00	60.86	1.532
8	120.00	120.17	0.160
<b>SEE</b> <b>Nb loadings</b>	0.88 4		

**Fructose**

**Exemple de bulletin de calibration**

Le nombre de loadings dépend du critère de sélection des composantes principales. Il est nécessairement au moins égal au nombre de constituants qui participent aux spectres. Les loadings supplémentaires peuvent correspondre aux défauts du critère de sélection ou à l'influence d'effets parasites (la ligne de base, interactions chimiques...).

Le choix du nombre correct de loadings est le point critique des algorithmes de compression des données. S'il n'y en a pas assez la calibration devient impossible ( $\rightarrow$  *underfitting*), mais s'ils sont trop nombreux le système peut perdre ses propriétés prédictives dans certains cas ( $\rightarrow$  *overfitting*).

Les huit composantes principales de la calibration précédente sont représentées sur la figure 36. Nous avons précisé pour chaque loading sa contribution à la variance totale. On vérifie visuellement qu'ils n'ont effectivement pas de signification physique évidente. Chaque composante résulte de la contribution *simultanée* des effets dont la présence explique les différences entre les spectres (constituants chimiques, bruit, ligne de base, interactions entre constituants...). Il est possible que l'influence de l'un d'entre eux soit concentrée sur certaines composantes. C'est sur cette possibilité que sont fondées les techniques d'analyse discriminante (voir chapitre 2.3). C'est ainsi également que l'on explique la capacité de la méthode PCR à réduire le bruit du spectre. S'il est réalisé dans de bonnes conditions, le facteur dominant sera les informations purement chimiques. Par conséquent il est logique de penser que cet effet sera concentré sur les composantes de variance les plus élevées qui seules sont retenues par l'algorithme PCR.

En prédiction, le logiciel calcule la distance de Mahalanobis  $D^2$  de chaque spectre au noyau formé par les standards de calibration. Il calcule aussi la valeur de  $1.96 \cdot \text{SEE}$  pour chaque constituant analysé. Le 1.96 est la valeur d'une variable de Student avec un nombre infini de degré de liberté au risque de 5%. Il suffit donc de multiplier cette valeur par  $(1 + D^2)^{\frac{1}{2}}$  pour obtenir un intervalle de confiance sur la valeur prédite (voir (Eq 31)). Celui-ci sera donc d'autant plus large que la distance de Mahalanobis sera grande, mais cela ne signifie pas que la prédiction soit erronée. Il est important

de signaler que (Eq 31) n'est vraie que dans le cadre du modèle statistique de Gauss-Markow qu'il faudra donc vérifier au préalable.

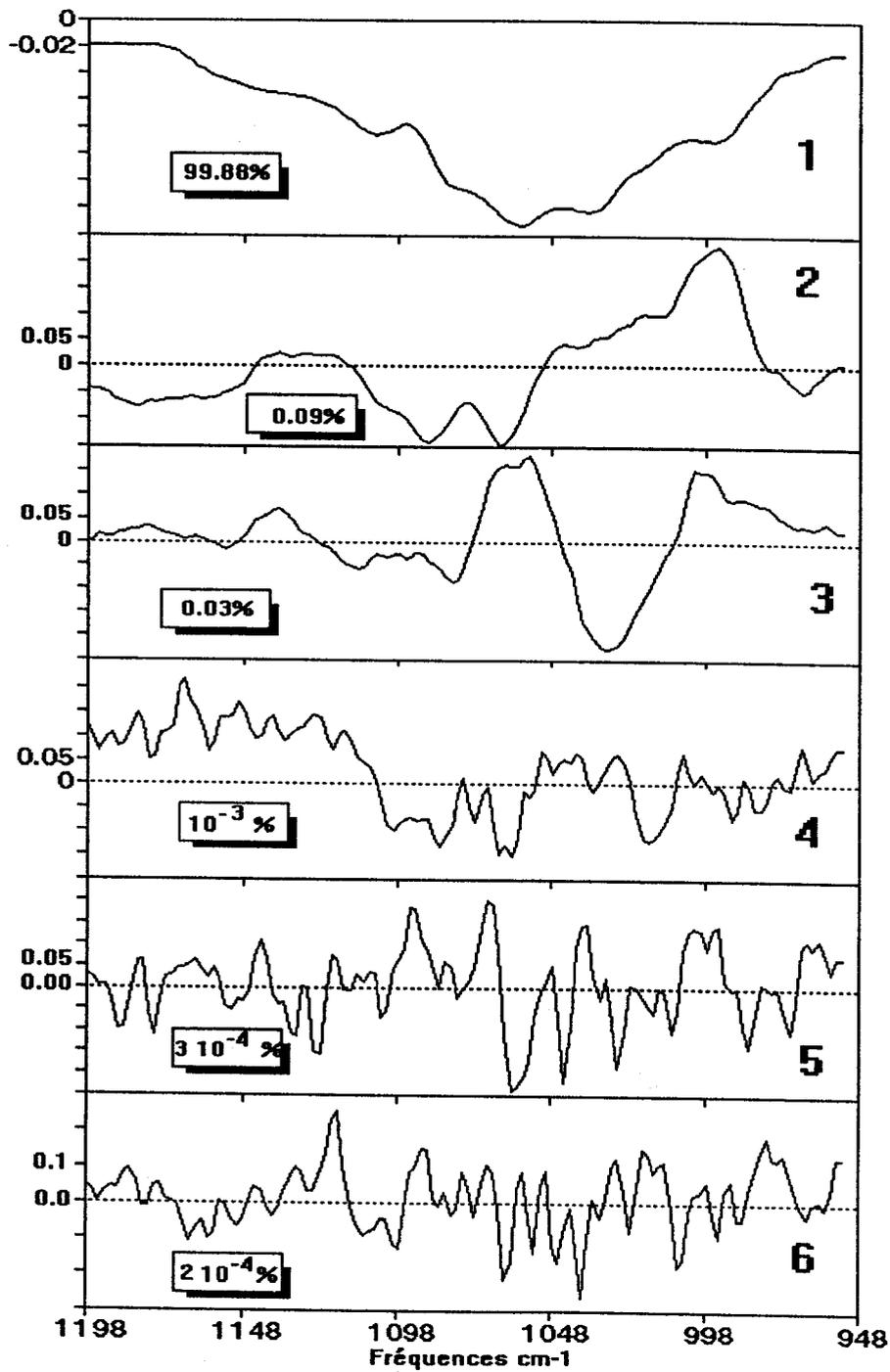
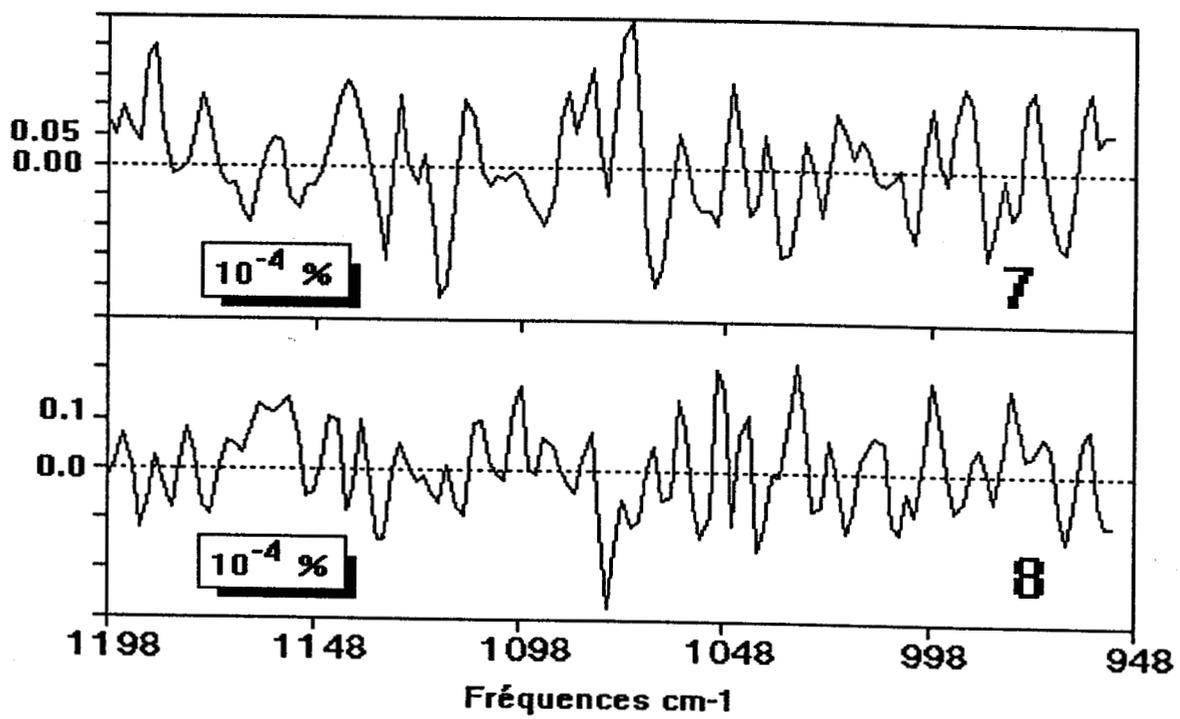


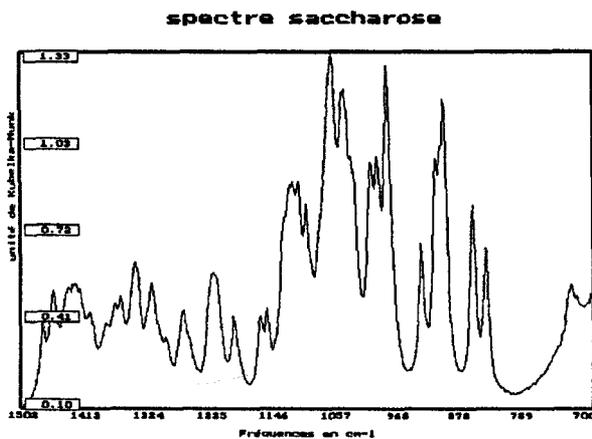
Figure 36-A: Les loadings de la calibration précédente.



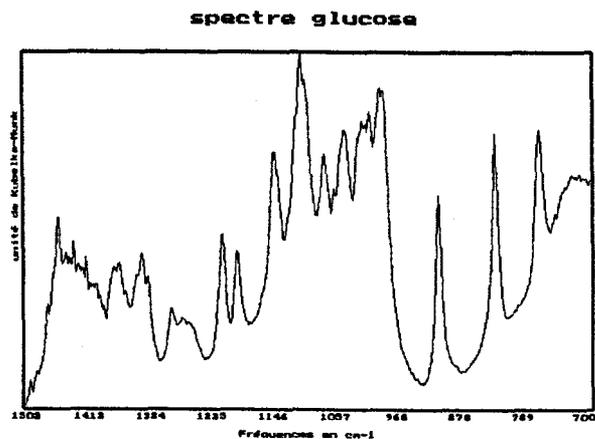
**Figure 36-B:** Les loadings de la calibration précédente.

## 5.2) LA REFLEXION DIFFUSE

Les résultats du chapitre 4.1 montrent que l'on peut espérer obtenir les meilleurs résultats lorsque les produits sont tamisés à moins de  $50 \mu\text{m}$ . Les deux montages précédemment étudiés étant difficiles à départager clairement, nous les utiliserons tous deux en analyse quantitative. Les essais de réflexion diffuse ont été réalisés sur des mélanges de saccharose et de glucose pour une concentration totale en sucre dans KBr de l'ordre de 5%. La gamme de concentration des sucres s'étend de 40 à 70%. Les analyses ont été effectuées sur la bande  $960\text{-}800 \text{ cm}^{-1}$  qui donne les meilleurs résultats. Nous reproduisons ci-dessous les spectres de réflexion diffuse de ces sucres.



**Figure 37-A:** Spectre de réflexion diffuse du saccharose



**Figure 37-B:** Spectre de réflexion diffuse du glucose.

On a montré aux chapitres précédents qu'il est en pratique difficile de maintenir le coefficient de diffusion  $S$  constant, ce qui n'est pas sans conséquence sur la linéarité de la fonction de Kubelka-Munk. L'équation (Eq 36) montre qu'une variation  $\Delta S$  de  $S$  produit un effet *multiplicatif* sur la fonction de Kubelka-Munk  $F$  (c'est à dire un effet que l'on ne peut éliminer par une dérivée). Un tel effet ne peut être corrigé par les modèles linéaires comme ceux décrits au chapitre 2.7.

La correction MSC ( $\rightarrow$  *Multiple Scatter Correction*) est souvent présentée pour apporter une correction à ce phénomène. Cette correction fut présentée simultanément par Geladi<sup>(83)</sup> et par Robert<sup>(84)</sup> en 1985. Ces derniers la dénomment "*méthode de l'écart à la régression*".

En réalité cette correction n'élimine pas l'effet multiplicatif de diffusion de la lumière, mais le ramène à une valeur identique à celle d'un spectre de référence  $Y_{ref}$ .

En l'absence de fluctuations spectrales dues aux variations de concentration de l'analyte et si l'on peut négliger le bruit des spectres, les  $n$  spectres  $Y_i$  seront tous identiques et la courbe  $Y_i = f(Y_{ref})$  sera représentée par la droite  $Y_i = Y_{ref}$ . En présence de variations de  $S$  et si l'effet de la diffusion ne dépend pas de la fréquence, on observera l'apparition d'une tendance générale affectant chaque fréquence. Dans le modèle MSC on suppose que cette tendance se traduit simplement par un changement de pente de la droite. On choisit généralement le spectre moyen  $Y_{mean}$  des  $Y_i$  comme spectre de référence et on écrit donc:

$$Y_i = b_i + a_i \cdot Y_{mean}$$

Les coefficients  $a_i$  et  $b_i$  sont calculés par régression linéaire pour chaque spectre, le spectre corrigé sera donc:

$$Y_{corr} = (Y_i - b_i)/a_i$$

Les variations d'origine chimique des spectres se traduisent par une distribution erratique des points autour de la droite: plus ces fluctuations seront fortes, moins la correction MSC sera précise. Elle s'applique donc plutôt à des spectres qui ne sont pas trop différents les uns des autres. Récemment Miller et Naes ont proposé une technique concurrente à base d'analyse factorielle qui éliminerait cet inconvénient<sup>(85)</sup>.

Pour l'analyse quantitative cette correction nous semble présenter de sérieux inconvénients: s'il est très possible que les effets multiplicatifs soient éliminés, on voit mal comment la linéarité des spectres peut être préservée par une telle correction. La figure 38 illustre l'effet de cette correction sur les courbes de la figure 19 à la page 65. Les fréquences à 850 et 1020  $cm^{-1}$  deviennent difficilement utilisables pour une analyse quantitative.

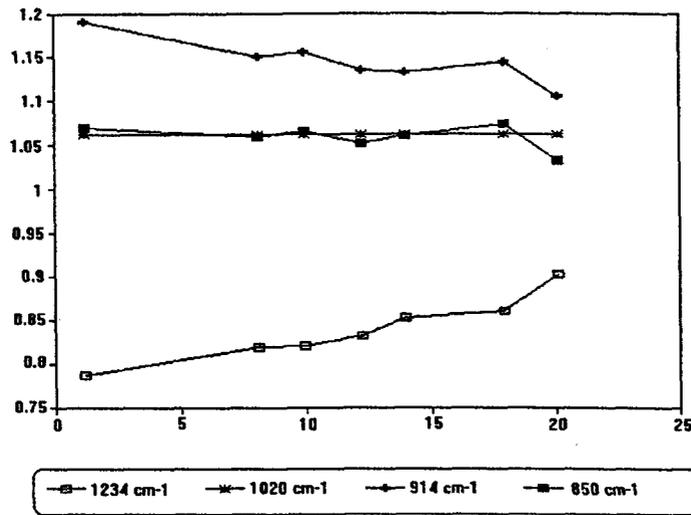


Figure 38: Effet de la correction MSC.

De fait les essais que nous avons pu réaliser confirment tout à fait qu'en général cette procédure rend les résultats nettement moins bons. Nous ne l'utiliserons donc pas dans ce travail.

La figure 39 représente une compilation des résultats obtenus sur 78 essais pour la méthode de la lame de rasoir et 230 essais pour la méthode de la presse. Il s'agit d'une courbe de fréquence relative cumulée, calculée pour chaque constituant et chaque montage. Ces courbes montrent par exemple que la probabilité qu'une mesure soit faite avec moins de 10% d'erreur est d'environ 0.9 pour la méthode de la presse.

Jusqu'à 10%, les courbes relatives à la méthode de la presse augmentent légèrement plus vite que pour la méthode de la lame de rasoir. Cette tendance s'inverse quasiment au-delà. Ce comportement montre que la distribution des erreurs relatives est légèrement plus décalée vers les faibles valeurs pour la méthode de la presse que pour celle de la lame de rasoir.

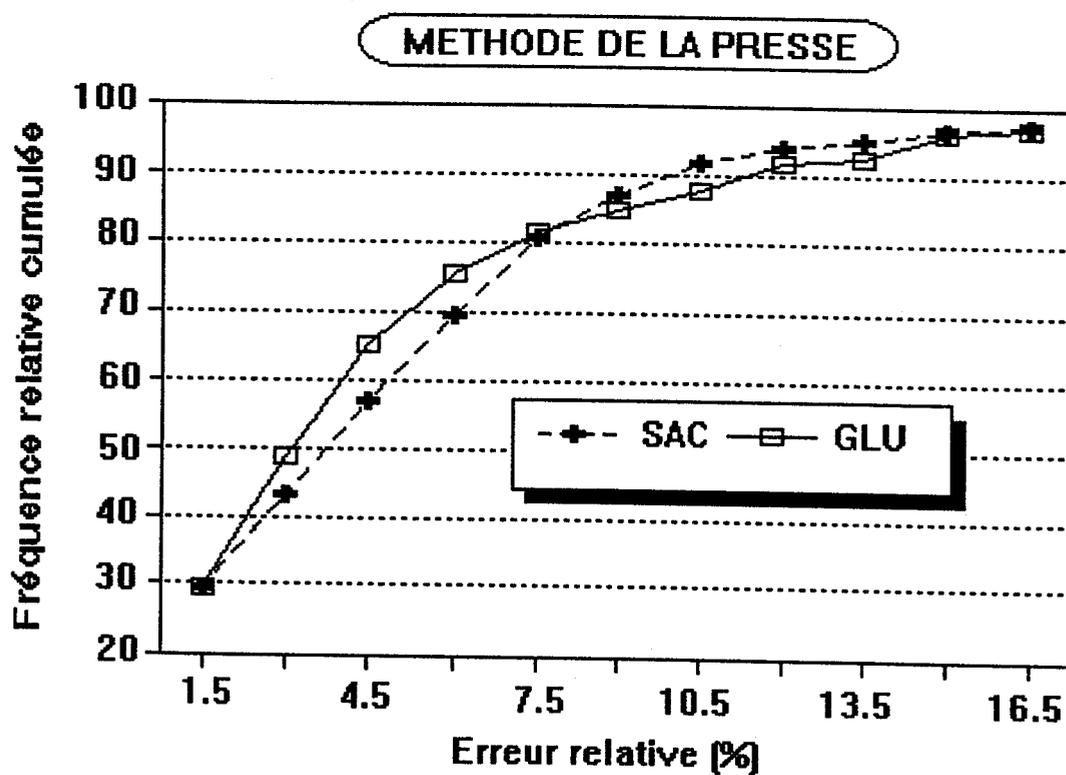
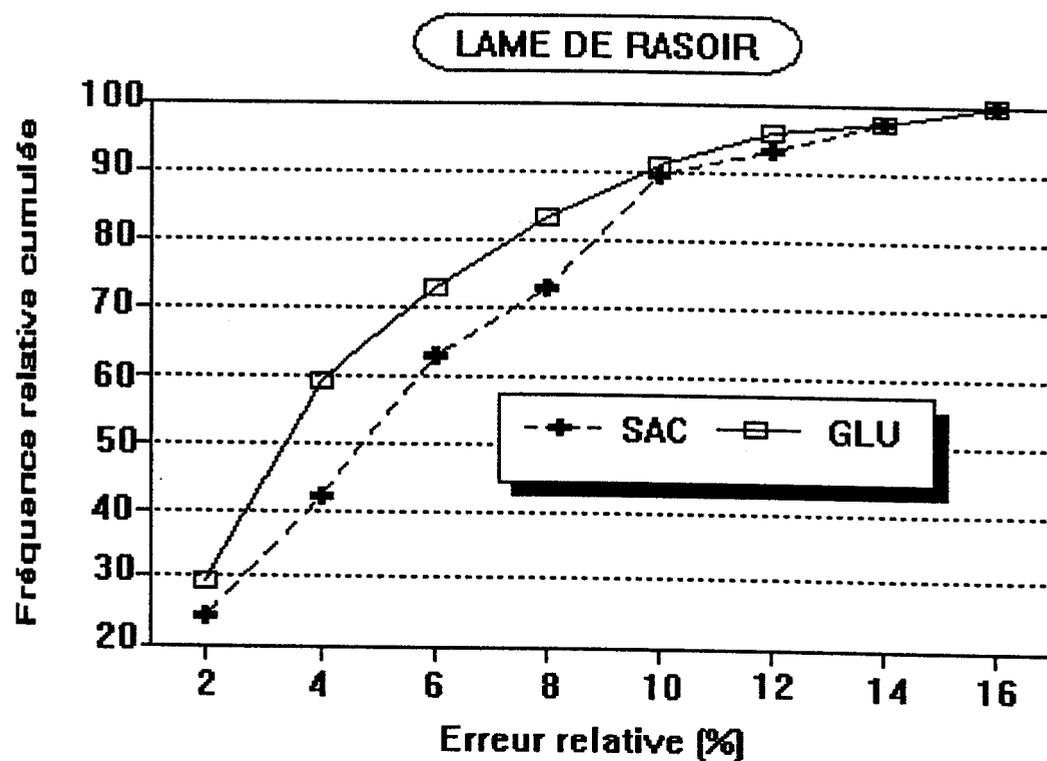


Figure 39: courbes de fréquences relatives cumulée obtenues par la méthode de la presse et de la lame de rasoir.

Nous avons obtenu les paramètres statistiques suivants:

Rasoir	Sac	Glu
Moyenne des erreurs relatives	5.34%	4.48%
Ecart type (%)	3.99%	3.59%
Résidus moyens	-0.94	0.58
Ecart type des résidus	2.91	2.88

Presse	Sac	Glu
Moyenne des erreurs relatives	4.51%	5.59%
Ecart type (%)	4.15%	4.26%
Résidus moyens	-0.35	-0.22
Ecart type des résidus	2.79	2.42

L'examen des deux tableaux précédents montrent que les deux montages sont assez proches l'un de l'autre, mais fournissent des résultats peu homogènes ce qui est confirmé par les histogrammes suivants:

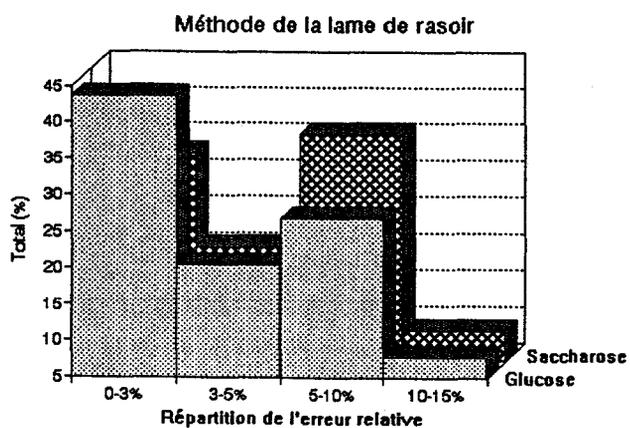


Figure 43-B

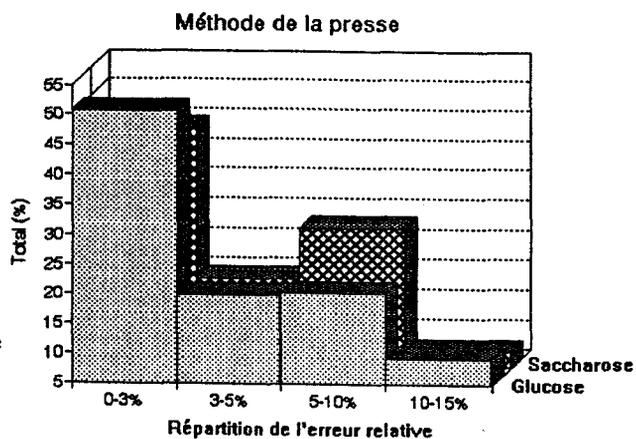


Figure 43-C

Fréquences relatives correspondants à différents intervalles d'erreur.

Il est difficile de savoir si les écarts entre les deux montages sont significatifs. Nous dirons que dans l'ensemble les résultats de l'analyse quantitative confirment les conclusions obtenues lors de l'étude de la répétabilité. D'autre part nous avons pu constater au cours de nos essais que le remplissage effectué par ces montages donne toujours des spectres suffisamment fiables pour effectuer une calibration et une validation dans des conditions raisonnables. Ce n'était pas toujours le cas lors du remplissage manuel des coupelles.

Trois causes peuvent *a priori* expliquer les difficultés rencontrées par la réflexion diffuse:

- \* Le bruit du spectre
- \* La répétabilité du compactage
- \* La répartition granulométrique

Les accessoires de réflexion diffuse ne restituent que 5% de lumière incidente sur KBr. La technique se caractérise donc par un rapport signal/bruit assez défavorable. Pour améliorer ce rapport on peut soit augmenter la résolution, soit réaliser des accumulations de spectres. Pour l'analyse quantitative, un compromis est fréquemment choisi avec une résolution de  $4 \text{ cm}^{-1}$ . Augmenter le nombre d'acquisition pose le problème de la stabilité du spectromètre, de l'environnement et de l'hydratation du KBr.

La linéarité des spectres de réflexion diffuse est un problème important. La loi de Kubelka-Munk est utilisée essentiellement parce qu'elle donne lieu à des calculs simples et rapides. Dans sa thèse Depecker<sup>(65)</sup> fait un inventaire de diverses transformations des spectres de réflexion diffuse en précisant leur domaine de validité. Si la transformation de Kubelka-Munk n'est pas toujours la plus efficace pour certains types de surface, elle représente globalement un choix assez correct. Le point critique de l'application de la fonction de Kubelka-Munk en analyse quantitative provient de la capacité que l'on a de maintenir le coefficient de diffusion  $S$  constant. Cela suppose de contrôler la répartition granulométrique des constituants ainsi que la répétabilité du compactage. On a montré



qu'à l'aide de techniques de remplissage semi-automatiques on pouvait améliorer ce dernier point, mais dans certaines limites.

En conclusion la réflexion diffuse dans le moyen infrarouge souffre de plusieurs inconvénients: les produits doivent être broyés et tamisés à des granulométries assez fines, la répétabilité est difficile à maîtriser, l'échantillonnage est relativement complexe (presse...). En contrepartie la technique peut analyser tout type d'échantillons solides (contrairement au pastillage) et le processus d'échantillonnage (broyage, tamisage et préparation) peut assez facilement être automatisé. Si la réflexion n'est certainement pas adaptée aux analyses de précision, elle peut sans doute trouver des utilisations pour des contrôles rapides de matière première par exemple.

### 5.3) ANALYSES PAR ATR

La figure 40 représente les spectres ATR de 4 solutions de sucres à la concentration de 300 g/l. La référence est l'eau pure. Les spectres ont été réalisés pour une résolution de 4  $\text{cm}^{-1}$  et 200 acquisitions sur un accessoire à 6 réflexions utiles.

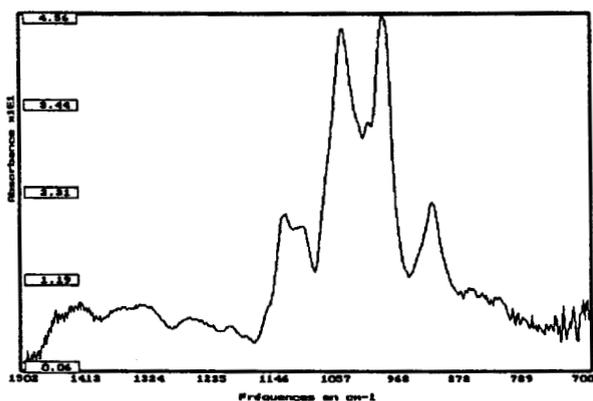


Figure 40-A: SACCHAROSE

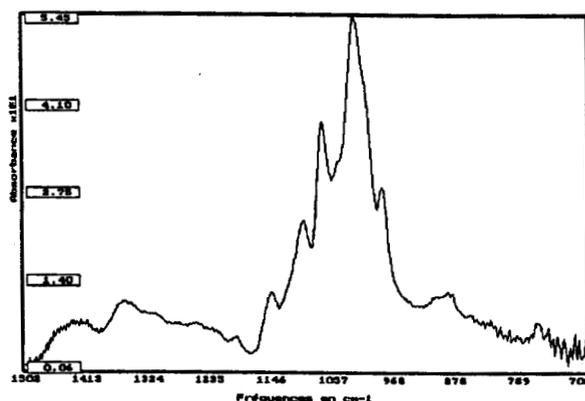


Figure 40-B: GLUCOSE

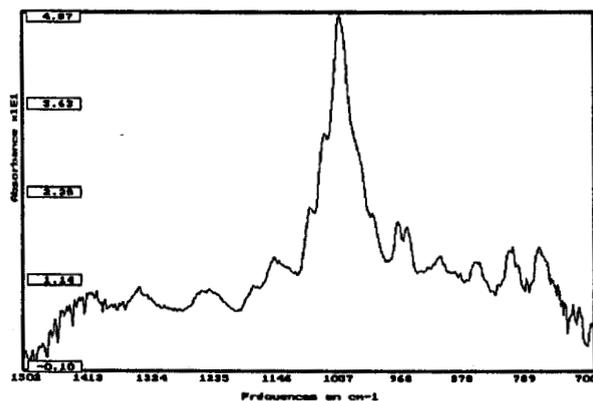


Figure 40-C: FRUCTOSE

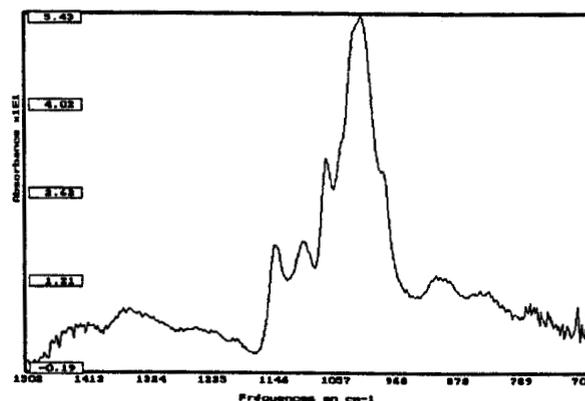


Figure 40-D: MALTOSE

Ces spectres sont très différents de ceux obtenus en réflexion diffuse. On observe une diminution importante de la résolution des bandes en raison des interactions avec les molécules d'eau. Certaines bandes disparaissent même presque entièrement. Pour l'analyse quantitative, nous nous sommes limité à la partie la plus significative du massif.

Nous avons réalisé une série de spectres de ces 4 sucres pour des concentrations comprises entre 60 et 300 g/l, puis calculé l'aire sous pic du massif principal du spectre d'absorbance sans

correction de ligne de base. Cette valeur est ensuite rapportée à la concentration pour obtenir les courbes des figures 41-A,B,C,D. Nous avons également reproduit les droites de régression linéaire.

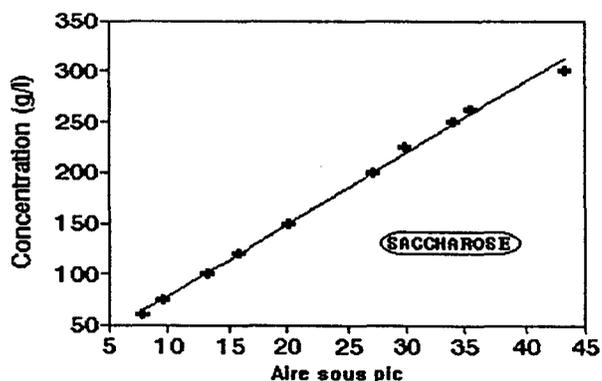


Figure 41-A

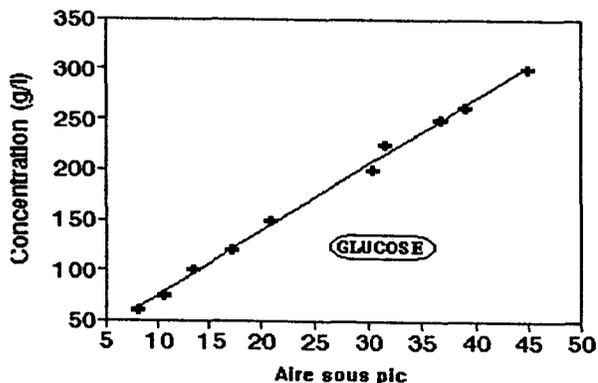


Figure 41-B

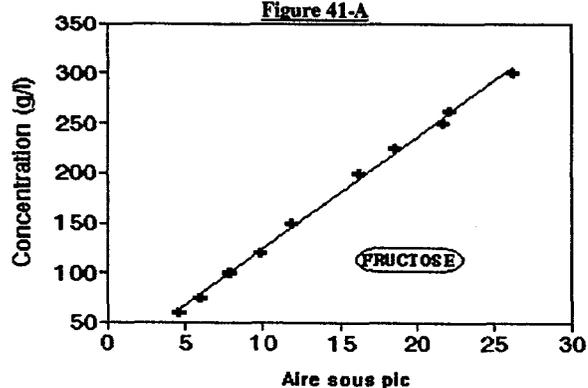


Figure 41-C

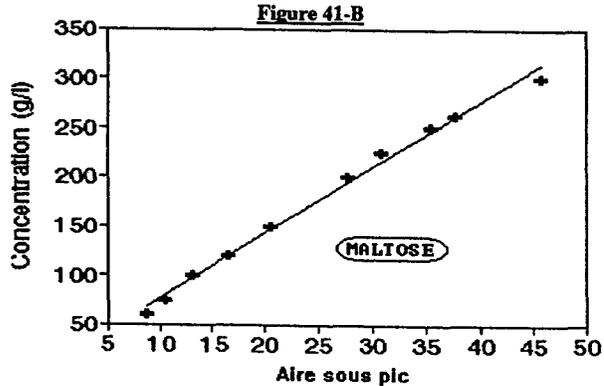


Figure 41-D

Ces courbes démontrent visuellement que les spectres ATR obéissent à une loi de type Beer-Lambert. En ce qui concerne l'analyse quantitative, l'ATR s'apparente donc aux techniques de transmission plutôt que de réflexion. On pourra donc légitimement modéliser les spectres d'absorbance selon un modèle linéaire.

De façon plus précise, nous avons calculé les paramètres suivants:

	Saccharose	Glucose	Fructose	Maltose
$r^2$	0.996	0.997	0.997	0.993
SEE (g/l)	5.6	5.3	4.9	7.6
Résidus moyens	0	0	0	0

Les coefficients de détermination sont très bons: le modèle linéaire explique presque 100% de la variance. La comparaison des variations de SEE avec l'ordre de grandeur des concentrations, de l'ordre de plusieurs centaines d'unités, est également très favorable pour les quatre séries. Cela confirme bien la bonne linéarité des spectres ATR.

L'étape suivante a consisté à généraliser ces résultats à des mélanges multicomposants de saccharose, glucose et fructose. Une étude préliminaire a démontré que les analyses sont optimales pour des solutions assez concentrées de l'ordre de 250 à 300 g/l de sucre total. De telles concentrations présentent deux avantages importants:

- on les prépare facilement par simple dissolution à froid.
- on obtient un bon rapport signal/bruit sans risquer des problèmes de non-linéarité du détecteur ou du mélange.

Une concentration de 300 g/l représente une valeur intermédiaire entre un sirop (700 g/l environ) et les jus de fruits naturels (de l'ordre de 100 g/l).

Notre démarche a été analogue à celle de la réflexion diffuse. La figure 42 représente la courbe de fréquence cumulée de l'erreur relative calculée sur 88 solutions.

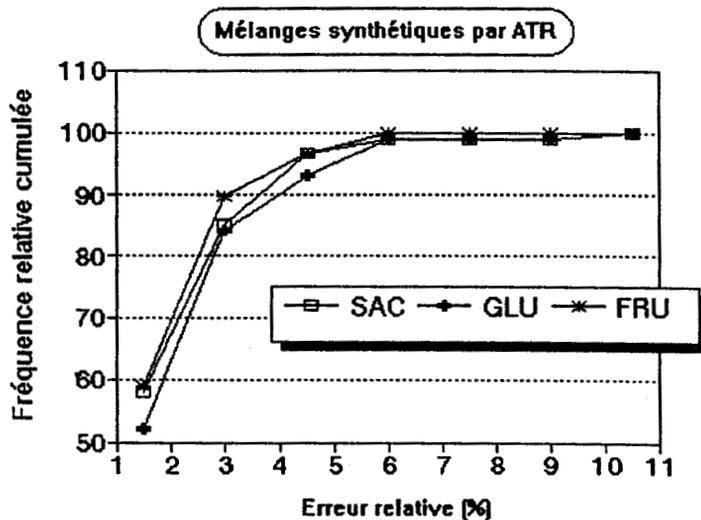
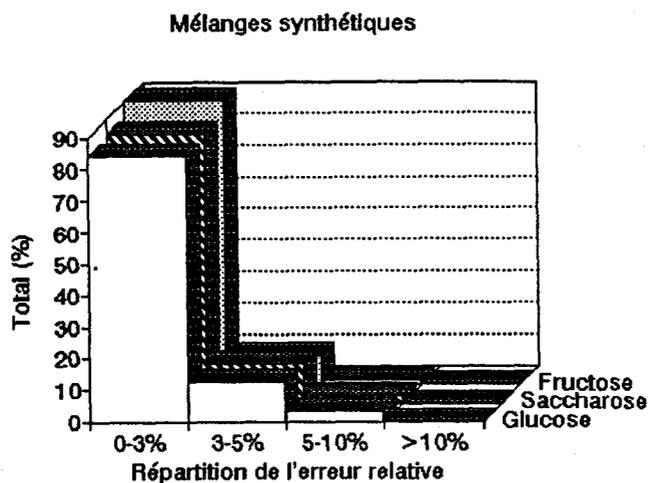


Figure 42

On observe une très forte croissance dans le domaine [0-3%], un ralentissement net sur la zone [3-6%], puis la courbe plafonne aux alentours de son maximum théorique. Cette structure est caractéristique d'une distribution fortement regroupée sous la barre des 3%: 90% des résultats ont moins de 3% d'erreur. On remarque en outre que la courbe suivie par chaque sucre est sensiblement équivalente: l'ATR donne des résultats beaucoup plus homogènes que la réflexion diffuse.

ATR	Sac	Glu	Fru
Moyenne des erreurs relatives	1.56	1.76	1.41
Ecart type (%)	1.52	1.59	1.21
Résidus moyens	-0.45	0.11	0.52
Ecart type des résidus	1.51	1.98	1.65

Ce tableau a été obtenu à partir des 88 mesures précédentes. Il montre que l'erreur relative moyenne et la dispersion sont en ATR de l'ordre du tiers des résultats obtenus en réflexion diffuse (voir tableau page 97). De tout ceci on conclut que l'ATR fournit des résultats plus précis que la réflexion diffuse.



**Figure 43:** histogramme montrant la répartition des fréquences relatives d'erreur pour différentes pages en ATR.

La comparaison entre la réflexion diffuse et l'ATR est plutôt favorable à cette dernière. Trois points nous semblent expliquer ce succès:

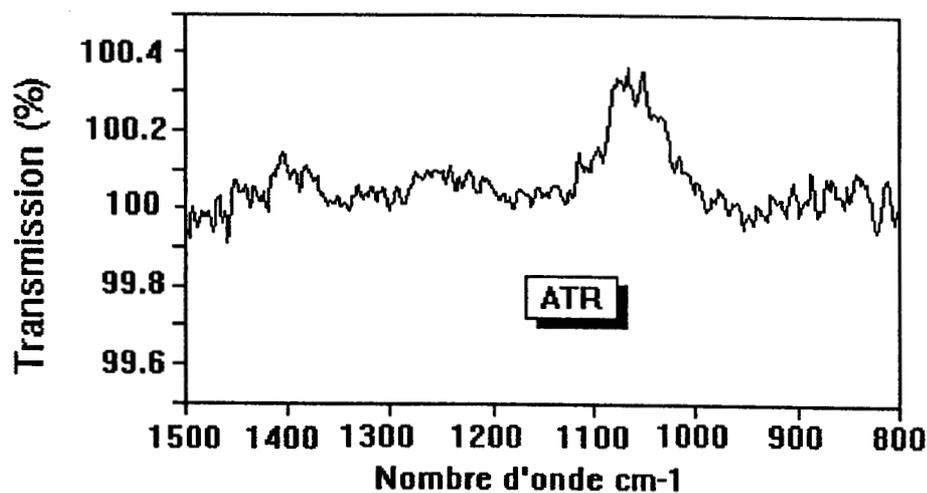
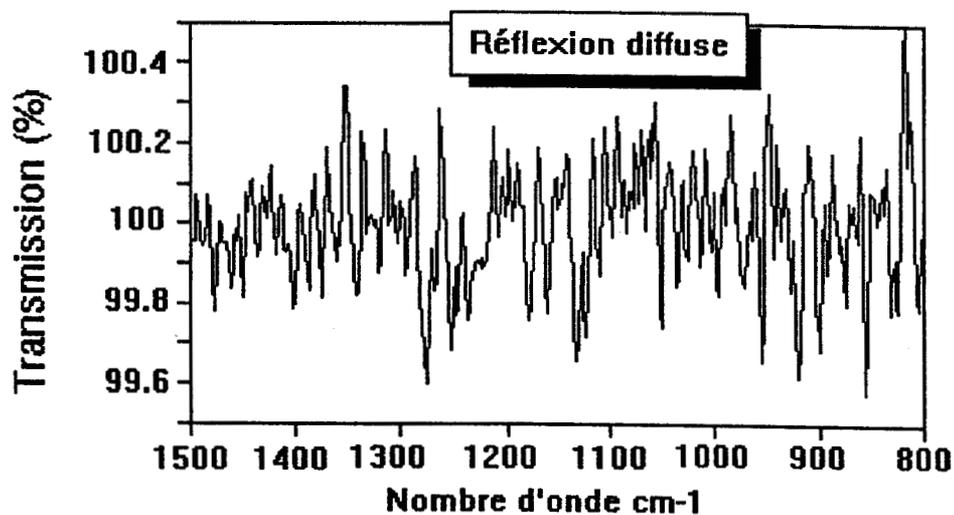
\* la répétabilité de la mesure est meilleure en ATR (3% environ) qu'en réflexion diffuse (de l'ordre de 10%).

\* Il est certainement plus facile d'échantillonner des liquides que des poudres. L'analyse des liquides est en fait le cas le plus favorable à la technique car il y a ni problème de contact avec le cristal, ni problème d'homogénéité. Notre accessoire (un ATR horizontal) n'est pas le plus efficace pour travailler sur des solides. C'est la raison pour laquelle nous avons étudié uniquement l'échantillonnage des liquides.

\* le rapport signal/bruit est plus élevé pour l'ATR: Un cristal d'ATR neuf peut restituer jusqu'à 30% de l'énergie incidente contre 5% pour la réflexion diffuse (sur KBr).

Parmi les inconvénients que l'on peut reprocher à l'ATR se trouvent:

- la fragilité du cristal: il résiste mal aux agressions chimiques et se brise ou se raye facilement, un cristal d'ATR n'est pas inusable.
- les difficultés pratiques pour travailler avec les poudres ou les solides en raison des problèmes de contact ou d'homogénéité de ces matériaux.
- le nettoyage du cristal bien que plus simple que pour le CIRCLE, notre expérience montre qu'en pratique il n'est pas toujours facile d'éliminer les derniers résidus de produit.



**Figure 44:** 100% effectués pour un montage de réflexion diffuse sur KBr ( $S/N = 692$ ) et en ATR sur air ( $S/N = 1234$ ).  
 Conditions: 200 scans, Zero Filling = 2, résolution =  $4 \text{ cm}^{-1}$ , APF = Norton-Beer.

## 5.4) APPLICATION A DES MELANGES REELS

---

### 5.4.1) JUS DE FRUITS

Nous avons réalisé des essais d'analyse quantitative sur des jus de fruits naturels. Différents exemples peuvent être trouvés dans la littérature, mais ceux-ci concernent essentiellement le proche infrarouge. Ces travaux soulignent la difficulté qu'il y a à déterminer les sucres présents en raison des interférences avec l'eau<sup>(86)</sup>.

Le moyen infrarouge est bien plus sensible à la présence d'eau que le proche infrarouge. Les raisons en sont les problèmes de saturation de bandes ainsi que la difficulté d'utiliser des cellules de transmission (contrôle de l'épaisseur de la cuve, recherche de matériaux inaltérables et transparents pour confectionner les fenêtres). Nous avons montré que ces problèmes peuvent être surmontés par l'ATR. Une autre possibilité est de cristalliser le mélange et de doser les sucres par réflexion diffuse. Cette solution pose de nombreux problèmes techniques et notamment l'évaporation complète de l'eau de façon rapide et sans risque de dégradation des constituants du mélange. En outre, nous avons montré que la réflexion diffuse ne permet pas de dosage avec une aussi bonne précision que l'ATR. Le choix de cette technique nous a donc semblé le plus approprié.

Le jus d'orange a été choisi comme modèle. Dans les jus naturels la concentration totale en saccharose, glucose et fructose est de l'ordre de 100 g/l. Industriellement ces jus sont préparés à partir de la variété *Citrus Sinensis*. Certains producteurs ajoutent de faibles quantités de *Citrus Reticulata* ou de *Citrus Aurantium*.

A l'initiative de l'Association Allemande des Industries des Jus de Fruits (VdF), les caractéristiques chimiques des jus commerciaux ont été tabulées, ce sont les valeurs RSK (Richtwert, Schwankungsbreite, Kennzahl). Il ne s'agit pas d'une norme légale, mais de critères pour évaluer l'authenticité des jus. Ces valeurs sont répertoriées dans le manuel "RSK-Values"<sup>(87)</sup>. Nous en avons extrait les renseignements suivants:

	Mini g/l	Maxi g/l	Valeur centrale g/l
Saccharose	-	47	33
Glucose	20	-	28
Fructose	22	-	30

Note: La valeur centrale n'est pas égale à la valeur moyenne. Il s'agit de la valeur la plus fréquemment rencontrée dans les produits industriels.

Nous avons pu disposer de 16 mélanges. Ces mélanges proviennent de différents pays producteurs et ont été préparés à partir de fruits récoltés durant les mois de Juillet, Août et Septembre des années 1988 et 1989. Les jus initialement concentrés et congelés ont été une première fois reconstitués par addition d'eau jusqu'à environ 11.2 degrés Brix. Ils ont alors été dosés selon les méthodes de référence préconisées par le manuel RSK-Values. Pour les sucres il s'agit de la chromatographie liquide HPLC. C'est sous cette forme que nous avons reçu les produits. Nous les avons reconcentrés à environ 300 g/l par évaporation partielle de l'eau au four à micro-ondes afin de se rapprocher de la forme initiale des produits et aussi pour se situer dans la gamme de concentration la plus favorable à l'ATR.

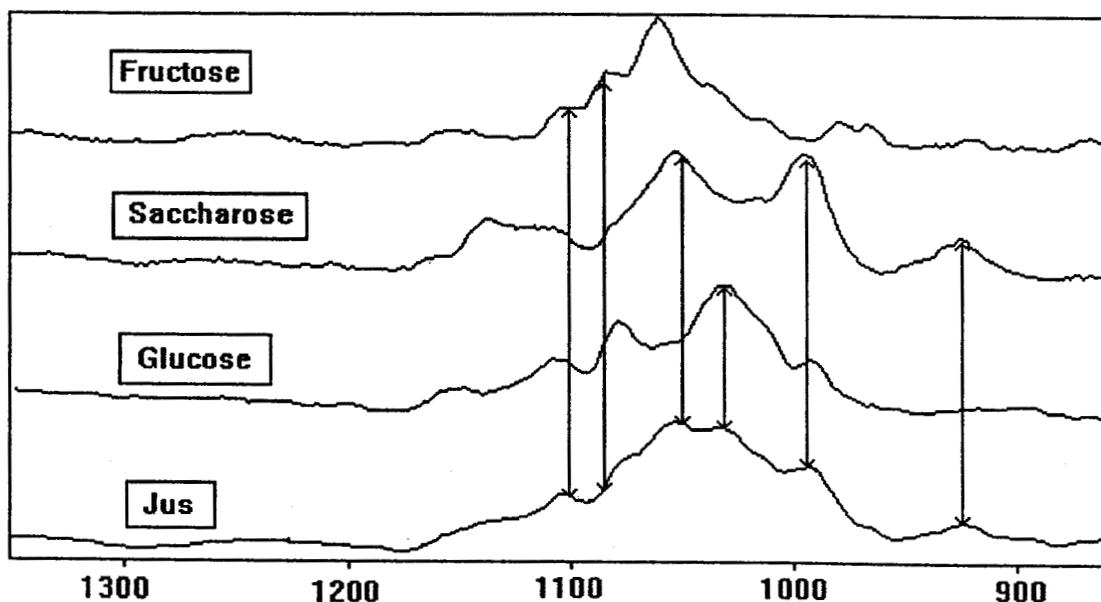


Figure 45: Le pointage est effectué sur les pics à 1101, 1085, 1048, 1031, 992 et 923  $\text{cm}^{-1}$ .

La figure 45 représente un spectre de jus d'orange avec en vis à vis les spectres du saccharose, du glucose et du fructose. Le pointage des pics montre clairement les similitudes qui existent entre le spectre du jus et ceux des sucres.

Ces 16 mélanges ont été répartis entre les standards de calibration et les standards de validation. La difficulté est de trouver pour chaque sucre une gamme de concentration suffisamment large pour établir une calibration. Il n'a pas été possible d'utiliser la même calibration pour chaque constituant. C'est la raison pour laquelle les tableaux suivants ne sont pas donnés avec le même nombre de spectres pour chaque mélange.

En conclusion les erreurs de dosages par ATR comprennent les erreurs introduites par les dilutions et concentrations successives, les erreurs de dosage par HPLC et les erreurs de dosage par ATR.

Nous avons rassemblés dans les tableaux suivants les résultats de la validation pour les sucres:

Saccharose	Réel g/l	Calculé g/l	Résidus g/l	Erreur relative %
	166.4	181.0	14.6	8.8
	170.6	185.2	14.6	8.6
	176.0	166.1	-9.9	5.7
	172.1	175.6	3.5	2.1
	169.7	168.5	-1.2	0.7
	168.8	169.5	0.7	0.4
	166.1	160.4	-5.7	3.4
	173.5	169.5	-4.0	2.3
	167.5	153.6	-13.5	8.3

Moyenne	170	-0.1	4.5%
Ecart type		9.8	3.4
Total		9	9

Glucose	Réel g/l	Calculé g/l	Résidus g/l	Erreur relative %
	93.8	93.3	-0.5	0.5
	97.7	95.0	-2.7	2.8
	90.6	91.9	1.3	1.5
	89.2	90.0	0.8	0.9
	92.7	88.6	-4.1	4.5
	83.1	90.8	7.7	9.3
	82.0	81.3	-0.7	0.9
	82.3	85.1	2.8	3.4
	99.0	92.2	-6.8	6.9
	105.5	98.9	-6.6	6.2

Moyenne	92		-0.9	3.7%
Ecart type			4.4	3.0
Total			10	10

Fructose	Réel g/l	Calculé g/l	Résidus	Erreur relative %
	99.8	100.5	0.7	0.8
	101.1	102.4	1.3	1.2
	96.1	99.1	3.0	3.1
	93.9	97.1	3.2	3.4
	98.5	95.5	-3.0	3.1
	89.8	97.8	8.0	9.0
	89.7	87.6	-2.1	2.3
	94.7	91.7	-3.0	3.1
	112.5	106.6	-5.9	5.2

Moyenne	97		0.2	3.4%
Ecart type			4.2	2.4
Total			9	9

Ces résultats sont sensiblement moins bons que ceux obtenus au chapitre 5.3 pour les solutions synthétiques. Les raisons sont probablement que ces dernières ont une composition mieux contrôlée, plus précisément:

- Les dilutions et les concentrations successives introduisent de nombreux facteurs d'erreur.

- Les jus d'oranges contiennent des résidus de pulpe en suspension malgré la centrifugation ainsi que des espèces chimiques mineures autres que les sucres. Ces espèces peuvent perturber les bandes d'absorption des sucres même si elles ne sont pas elles-mêmes actives en infrarouge.

Compte tenu de ces différents facteurs, ces résultats nous semblent encourageants.

D'autres constituants mineurs existent dans les jus et doivent également être dosés. Ce sont les acides organiques (citrique, malique...) ou les 20 acides aminés naturels. Le constituant mineur le plus concentré est l'acide citrique (30 g/l environ). Son spectre ATR est représenté sur la figure 46.

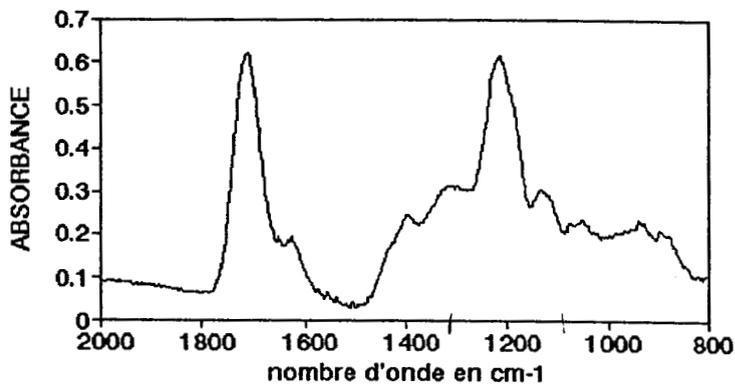


Figure 46: spectre ATR de l'acide citrique en solution.

Il dévoile deux zones intéressantes *a priori* pour tenter une analyse quantitative. Le tableau suivant résume l'analyse effectuée sur la bande 1300-1100  $\text{cm}^{-1}$ .

Ac. Citrique	Réel g/l	Calculé g/l	Résidus	Erreur relative %
	34.0	36.4	-2.4	7.0
	30.5	31.1	-0.6	1.9
	33.9	33.7	0.2	0.5
	44.4	40.6	-3.9	8.7

Moyenne	35	0.3	2.4%
Ecart type		2.7	3.8
Total		4	4

Les résultats sont numériquement assez bons. On remarque toutefois qu'il devient nécessaire d'utiliser un plus grand nombre de standards que pour les sucres. Nous pensons que la raison est que le domaine spectral choisi n'est pas celui où les bandes d'absorption de l'acide citrique sont le plus linéaires.

L'autre alternative est d'effectuer le dosage sur la bande des carbonyle entre 1750 et 1700  $\text{cm}^{-1}$ . Il se pose le problème de l'élimination de l'eau qui est le constituant majoritaire dans la solution. On sait que la forme des bandes de l'eau est très sensible à la composition chimique du milieu<sup>(88)</sup>. De part sa capacité à créer des liaisons hydrogènes, l'eau interagit facilement avec les groupements polaires du milieu. Cela se traduit par de légers décalages en fréquence des bandes caractéristiques des constituants actifs. De tout ceci il résulte que la soustraction du spectre de l'eau est très difficile à faire avec précision, en particulier pour observer un pic peu intense. De fait l'analyse de l'acide citrique ne donne pas de meilleurs résultats dans la gamme 1750-1600  $\text{cm}^{-1}$ .

Signalons pour finir une dernière limitation de l'ATR qui est la fréquence de coupure des IRE usuels. Celle-ci est de 750  $\text{cm}^{-1}$  pour le ZnSe. Le rapport Signal/Bruit diminue de façon très importante vers les basses fréquences rendant l'analyse quantitative très imprécise.

#### **5.4.2) SIROPS DE GLUCOSE**

Sous cette dénomination on désigne des mélanges de divers sucres autres que le saccharose. Le glucose n'y est pas nécessairement le produit majoritaire. On peut ainsi référencer 150 types de sirop selon leur composition qui se répartit dans les gammes suivantes:

	Minimum	Maximum
<b>Matière sèche</b>	70%	80%
<b>Glucose</b>	2%	95%
<b>Fructose</b>	0%	95%
<b>Disaccharide</b>	2%	65%

Le disaccharide est principalement le maltose. Ces sirops peuvent être classés en deux catégories selon qu'ils contiennent ou non du fructose.

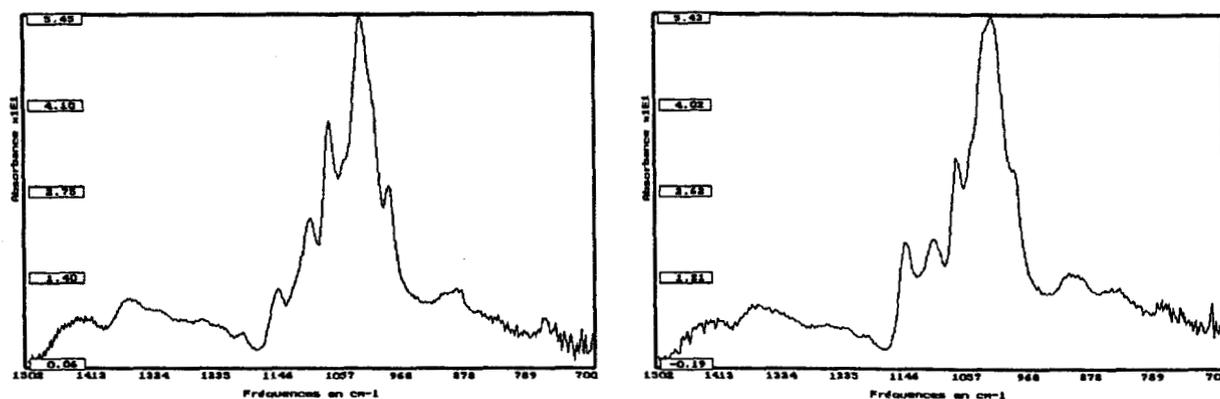
Les produits tests nous ont été fournis par la société Roquette Frères. En contrôle de qualité, le dosage des différents sucres est réalisé par chromatographie liquide, ce qui nécessite environ 30 minutes par échantillon.

Compte tenu de la nature des produits, la technique la mieux appropriée est *a priori* l'ATR. Nous avons pu disposer d'une série de mélanges dans chacune des deux catégories précédentes. Le mélange A sera constitué de glucose et de maltose, le mélange B contient en plus du fructose.

Nous avons dilué les produits pour nous situer dans une gamme d'absorbance compatible avec les mesures ATR en mélangeant 1 g de sirop avec 2 ml d'eau distillée. Le paramètre à mesurer est le pourcentage de sucre dans la matière sèche.

Les sirops ont une teneur en matière sèche variable, de plus il est en pratique impossible d'effectuer des dilutions rigoureusement identiques. La variable de calibration ne pourra donc pas être directement la proportion de sucre dans la matière sèche, mais la proportion de sucre dans le mélange dilué. Il sera donc nécessaire de pouvoir mesurer la quantité de matière sèche de façon indépendante pour remonter au bon paramètre.

Cet exemple est plus complexe que le précédent dans la mesure où deux de ses constituants majeurs ont des spectres très semblables (glucose et maltose) et où les constituants non dosés sont en concentration cette fois non négligeable (de l'ordre de 30%).



**Figure 47:** Les spectres du glucose (à gauche) et du maltose se ressemblent en première approximation à l'intensité des bandes près.

Les calculs ont été effectués sur un logiciel BRUKER implémentant l'algorithme PLS sur un PC. L'analyse est réalisée entre 1180 et 955  $\text{cm}^{-1}$  pour une résolution de 2  $\text{cm}^{-1}$ .

Les deux tableaux de la figure 48 contiennent les résultats de l'étape de validation effectuée pour chaque série. Une calibration spécifique à chaque constituant a été effectuée. Les lignes manquantes du tableau représentent les mélanges inclus dans la calibration. Dans le cas de la série (B), les mélanges 25 à 33 sont des mélanges artificiels créés uniquement pour élargir un peu la gamme de calibration.

Les spectres de calibration utilisés pour doser les séries A et B ne sont pas les mêmes pour chaque constituant afin d'optimiser le mieux possible la calibration. On a montré au paragraphe 2.7 que cela n'aurait pas été possible à partir de la méthode CLS car elle nécessite de doser *simultanément* tous les constituants du mélange. Cette possibilité de doser les analytes un par un est une des propriétés intéressantes des méthodes factorielles PCR et PLS.

MELANGE A

Mélange	GLUCOSE(H&H)	GLUCOSE(C&C)	Erreur(%)	DR&C(H&H)	DR&C(C&C)	Erreur(%)
1				38.3	36.4	4.8
2						
3				39	38.6	0.9
4	24.9	24.5	1.6	39	40.1	2.9
5	29.7	29.5	0.7	36.1	37.0	2.6
6	32.5	32.3	0.6	34.8	34.5	0.9
7	29.3	29.1	0.7	35	35.8	5.0
8	26.6	24.2	9.1	36.6	36.4	7.7
9						
10	31.9	31.1	2.5	34.4	37.9	10.2
11				45.6	47.6	4.4
12						
13	13.3	14.5	8.9	43.6	45.2	3.8
14	14.6	12.6	13.4	49.6	45.9	7.4
15	14.8	12.4	18.5	42.1	44.3	5.2
16				44.3	48.2	4.4
17	30.6	28.9	5.2	34.5	37.7	9.4
18	24.1	23.4	3.0			
19	28.3	28.7	1.5	35.5	38.3	7.8
20	28.2	27.6	2.3			
21	12.7	12.3	2.9	44	46.5	5.7
22	11.6	10.2	11.7	45.3	47.5	4.9
23	27.4	25.5	5.3			
25	25.7	26.7	4.1	38.3	39.2	2.3

Moyenne 23.9 5.2 39.8 5.0  
 Ecart type 4.9 4.9 2.7  
 Total 17 18

MELANGE B

Mélange	FRU(H&H)	FRU(C&C)	Erreur(%)	GLU(H&H)	GLU(C&C)	Erreur(%)	DR&C(H&H)	DR&C(C&C)	Erreur(%)
1	19.5	20.2	3.6	39.5	40.7	3.0	30.2	29.3	2.9
2	19.4	20.2	4.0	40.1	42.0	4.7	28.1	28.3	0.8
3	19.3	19.6	1.7	40.5	42.4	4.8	29.1	29.9	2.9
4	19.1	19.4	1.7	40.8	41.8	2.5	29	27.9	3.7
5	18.8	19.9	5.8	38.4	40.8	3.6	30.5	30.1	1.4
6	19.3	19.1	1.0	39.8	40.2	0.8	28.5	30.3	6.3
7	19.1	18.6	2.7	40.2	37.6	6.6	28.1	27.8	0.9
8	19.2	18.8	2.2	40.3	38.1	5.4	31.1	28.5	8.5
9	19.2	19.3	0.7	39.1	40.9	4.7	27.6	30.2	9.3
10	19.1	19.7	2.4	40.2	39.1	2.8	28.1	29.4	4.8
11	19.8	19.9	0.5	40.4	41.3	2.3	27.8	29.2	4.9
12	19.1	19.6	2.6	40.3	41.6	3.2	28.6	29.3	2.6
14	18.8	19.2	2.0	40.8	42.0	2.8	28.3	28.1	0.6
15	19	19.0	0.1	39.9	39.6	0.7	28.5	28.2	1.1
16	18.3	19.0	3.6	40.6	40.8	0.6	29	29.0	0.1
17	19	19.4	2.3	39.8	41.5	4.2	28.3	29.4	3.7
18	18.9	19.3	2.0	40.2	41.4	2.9	28.6	29.7	3.9
19	19.3	20.0	3.4	39	39.2	0.5	28.5	28.8	0.9
20	19.3	19.4	0.5	39.2	40.9	4.4	28.6	28.7	0.4
21	19.1	19.1	0.1	39.1	40.9	4.5			
22	20.1	20.1	0.2	40.2	39.9	0.7	28.6	27.4	4.1
23	18.6	19.0	2.2	39.7	40.9	3.1	29.7	28.1	5.4
24	18.7	18.9	1.1	39.4	41.7	5.7	29.5	29.7	0.6
25							25.6	24.9	3.8
26									
27									
28							17.8	17.8	0.2
29	15.8	15.8	0.3						
30							18.6	18.3	1.8
31							26.7	26.0	2.3
32									
33	15.4	15.4	0.1	33.8	33.7	0.3			

Moyenne 18.9 1.9 39.7 3.1 27.8 3.0  
 Ecart type 1.5 1.5 1.8 1.8 2.5  
 Total 25 24 26

Figure 48: Résultats de l'analyse des sirops de glucose.

Les erreurs relatives moyennes pour chaque série sont répertoriées dans la figure précédentes. les tableaux suivants sont relatifs aux résidus:

A	Résidu Moyen	Ecart type	Nombre de spectres
Glucose	-0.7	1.0	17
Maltose	1.3	1.8	19

B	Erreur Moyenne	Ecart type	Nombre de spectres
Glucose	0.7	1.3	25
Fructose	0.2	0.4	24
Maltose	0.0	1.1	26

La figure 48 montre qu'il existe des différences importantes en faveur de (B) entre les mélanges (A) et (B). Cela est surprenant puisque avec un constituant de moins, le dosage de (A) devrait être au contraire plus facile. N'ayant pas d'autres échantillons à notre disposition il n'a pas été possible de tester s'il s'agit d'une tendance générale ou bien d'un artefact lié au lot utilisé.

Les spécifications demandées sont d'au maximum 5% d'erreur. Dans la mesure où les solutions tests ont été étalonnées par chromatographie qui présente une incertitude du même ordre de grandeur, les résultats obtenus pour les mélanges (A) et (B) sont satisfaisants avec une légère réserve pour (A).

En conclusion cet essai démontre la *faisabilité* de l'ATR pour ce type de mélange. Il est maintenant nécessaire de réaliser d'autres essais à plus grande échelle. Les renseignements que l'on doit en tirer sont les suivants:

- Estimation de l'erreur de prédiction.

- Estimation de la fiabilité.
- Elucidation du cas posé par le mélange (A).
- Confirmation que les résultats corrects du mélange (B) traduisent une tendance générale et non pas le fait d'une gamme de concentration trop faible.
- Influence de l'origine des matières premières.

Il est important de noter que ces différents points n'ont de signification *que* statistique. Des essais à faible échelle comme les précédents ne peuvent que prouver la *faisabilité* de l'échantillonnage et de l'analyse. Un spectre infrarouge est la résultante complexe des espèces chimiques présentes dans le milieu, qu'elle soient ou non actives en infrarouge. Or la composition exacte de ce milieu peut varier en fonction de nombreux paramètres comme l'origine des matières premières ou l'époque de l'année pour un produit d'origine biologique. Il est donc nécessaire de savoir prendre en compte cette variabilité lors de l'analyse quantitative, ce qui peut nécessiter de réaliser une calibration spécifique à chaque situation rencontrée. Résoudre les problèmes de gestion que cela implique est un autre aspect important de l'analyse quantitative.

Nous avons expliqué au paragraphe 2.7 comment l'algorithme PCR construisait un espace de projection à l'aide des seules informations spectrales, tandis que PLS utilise simultanément les informations chimiques. Nous avons de plus remarqué la similitude des spectres ATR du glucose et du maltose en solution aqueuse. Il nous a donc paru intéressant d'étudier le comportement de ces deux algorithmes sur l'exemple mélange (A).

Les deux tableaux suivants résument les résultats de la calibration du mélange (A) pour le glucose.

PLS	Réel	Calculé	Erreur (%)
11	10.2	10.2	0.2
16	12.0	12.0	0.4
9	14.8	14.8	0.2
12	23.5	24.4	0.1
3	25.0	25.0	0.0
1	27.2	27.2	0.0
2	32.6	32.6	0.0

PCR	Réel	Calculé	Erreur (%)
11	10.2	10.2	0.4
16	12.0	12.1	1.2
9	14.8	14.6	1.2
12	23.5	23.6	0.4
3	25.0	23.5	0.1
1	27.2	27.2	0.1
2	32.6	32.5	0.3

Note: Modèle additif.

La calibration de la méthode PCR est donc tout à fait correcte. Le tableau suivant résume les résultats de la validation effectuée à partir de cette calibration. Les spectres sont les mêmes que ceux précédemment choisis pour la méthode PLS.

Mélange	Glucose (Réel)	Glucose (Calc.)	Erreur (%)
1			
2			
3			
4	24.9	25.7	3.4
5	29.7	15.7	47.2
6	32.5	-1.7	105.2
7	29.2	1.8	94.0
8	26.6	-8.7	132.8
9			
10	31.9	-8.0	125.1
11			
12			
13	13.3	-682.5	5231.4
14	14.6	28.7	96.7
15	14.8	20.4	37.8
16			
17	30.5	18.4	39.8
18	24.1	-0.2	101.0
19	28.3	12.5	55.7
20	28.2	13.9	50.6
21	12.7	8.4	33.8
22	11.6	22.8	96.3
23	27.4	17.4	36.4
25	25.7	-63.8	348.4

Certaines concentrations deviennent négatives tandis que les autres ne correspondent à rien: les résultats obtenus sont donc totalement inexploitable.

Deux conséquences intéressantes peuvent être tirées de cet exemple:

- Il est parfaitement possible qu'une calibration soit bonne, mais que le système soit incapable de fonctionner en prédiction. Comme nous l'avons souligné au chapitre 2.9, les résultats de calibration n'impliquent rien quand à la qualité de la prédiction. Ce comportement paradoxal est appelé *overfitting* par la littérature.

- L'exemple présenté illustre une limitation importante de l'algorithme PCR. Nous avons déjà remarqué que les spectres du glucose et du maltose sont très voisins. Or l'algorithme PCR n'utilise que les spectres pour construire l'espace factoriel. Dans ce cas particulier on peut imaginer qu'il ne dispose pas de suffisamment d'informations pour corrélérer de faibles variations à des concentrations. Le système pourra donc avoir des difficultés pour effectuer

une discrimination entre les deux constituants. La méthode PLS ne souffrant pas de cet inconvénient, il peut donc sembler logique qu'elle soit meilleure que l'approche PCR.

## **CHAPITRE VI**

### **VI- CONCLUSION GENERALE**

## 6) CONCLUSION GENERALE

L'objectif de ce travail a été d'essayer d'évaluer les possibilités du moyen infrarouge pour l'analyse quantitative appliquée à la résolution de problèmes industriels. Ce sont souvent des besoins de masse, par conséquent il est important de s'attacher à simplifier le plus possible la méthodologie afin d'optimiser au mieux la rapidité de la mesure, de réduire les risques d'erreur et de limiter la dépendance des résultats vis à vis de l'opérateur.

Dans un premier chapitre, nous avons fait le point sur les principaux algorithmes d'analyse quantitative en précisant leurs qualités et leurs défauts. Ce point est important car les logiciels d'analyse quantitative sont nombreux sur le marché et selon la méthode mathématique utilisée les performances peuvent être très différentes. Cette étude a souligné les performances supérieures des algorithmes d'analyse factorielle (PCR et PLS). Mais nous avons montré sur un cas pratique, qu'en certaines circonstances la méthode PCR peut devenir totalement inopérante.

Les deux techniques qui *a priori* nous semblaient répondre le mieux aux objectifs rappelés plus haut sont la réflexion diffuse et l'ATR horizontal. L'analyse quantitative est une application "haut de gamme" du moyen infrarouge, qui est de ce fait très sensible à la qualité des spectres. Nous avons donc étudié de façon détaillée les difficultés pratiques qui caractérisent ces deux approches ainsi que les moyens d'en limiter les effets.

Le point délicat de la réflexion diffuse est le manque de répétabilité qui pose de sérieux obstacles à son utilisation en analyse quantitative. Un remplissage manuel sans précautions ne permet pas de réussir avec certitude un spectre utilisable pour une analyse quantitative. Nous avons montré que le contrôle de la granulométrie et du compactage améliore fortement la répétabilité, mais la méthodologie se complique. Néanmoins il est possible d'automatiser facilement le montage proposé. Nous avons trouvé, toutefois, que les performances restent assez moyennes en terme

d'erreur de prédiction, ce qui confine au mieux la technique aux applications où une grande rapidité de mesure est nécessaire, sans qu'il soit indispensable d'obtenir une très haute précision.

La seconde approche étudiée est l'ATR horizontal. Cette technique est surtout adaptée aux liquides. Elle pose très peu de problèmes pratiques si ce n'est la relative fragilité de l'accessoire. Les performances pour l'analyse quantitative sont meilleures que pour la réflexion diffuse: la précision et la répétabilité sont améliorées d'un facteur 3. Les résultats ainsi obtenus font de l'ATR une technique d'analyse des liquides d'autant plus compétitive que la prise d'un spectre et d'une référence ne prend que 5 minutes. Le calcul des concentrations ne nécessite ensuite que quelques secondes.

A partir de l'exemple du dosage des sucres dans le jus d'orange et les sirops de glucose, nous avons confirmé la faisabilité suggérée par les mélanges synthétiques. Nous avons également mis en évidence deux des principales limitations de l'ATR:

- Il s'agit d'une méthode relativement peu sensible. Cela lui permet de s'astreindre des problèmes de saturation des bandes propres aux techniques par transmission, mais rend difficile le dosage des constituants mineurs.
- Les recouvrements avec les bandes du solvant, en particulier de l'eau, rendent difficile le dosage de certains constituants.

Une suite importante de ce travail préliminaire sera de préciser la robustesse de l'ATR par rapport à des fluctuations de la nature des échantillons (origine des matières premières...) ainsi que l'influence sur les résultats de l'opérateur et du spectromètre (erreur interlaboratoire).

Le problème de l'interférence de l'eau peut théoriquement être résolu soit par soustraction du spectre de l'eau pure, soit en réalisant des extraits secs. La soustraction du spectre de l'eau est techniquement difficile à réaliser avec précision pour observer les pics peu intenses des constituants mineurs. La méthode des extraits secs a été mise au point et expérimentée dans le proche infrarouge

par Meurens et son équipe<sup>(20)</sup> avec succès. Un travail actuellement en cours au laboratoire consiste à étudier des solutions permettant de généraliser cette approche au moyen infrarouge.

## **CHAPITRE VII**

### **VII- BIBLIOGRAPHIE**

## 7) BIBLIOGRAPHIE

- 1: Osborne B.G., Fearn T., *Near Infrared Spectroscopy in Food Analysis*. Longman, Wiley, New-York, 1986, 200 pages.
- 2: Goulden J.D.S., Shield J., Haswell R., *The Infrared Milk Analyser, J. of the Soc. of Dairy Techn.*, 1964, 17(1), p28-33.
- 3: Gilkison I.S., *Applications of Infrared to the Analyse of Milk and Milk Products, J. of the Science of Food and Agriculture*, 1983, 34(9), p1026-1027.
- 4: A.A.C.C., *Near Infrared Technology in the Agricultural and Food Industries*, Williams and Norris, St Paul (Minn.), 1987, 330 pages.
- 5: Griffiths P.R., Sloane H.J., Hannah R.W., *Interferometers vs Monochromators: Separating the Optical and Digital Advantages, Appl. Spect.*, 1977, 31(6), p485-495.
- 6: Griffiths P.R., De Haseth J.A., *Fourier Transform Infrared Spectrometry*, Wiley, New York, 1986.
- 7: Memelink A.A., *Infrared Quantitative Analysis of Trans Content in Fatty Acids and Derivatives, Laboratory Handbook for Oil and Fat Analysis*, Cocks, Van Rede Editors, Academic Press, p388-391.
- 8: Martens H., Wold S., *Suggested Recommendation for Notation and Terminology in Chemometrics*. Version 0.2, Norwegian Food Research Institute, 1984.
- 9: Bertrand D., *Thèse de doctorat 1988, Université de Dijon (France)*.
- 10: Devaux M.F., Bertrand D., Robert P., Morat J.L., *Extraction of NIR Spectral Information by Fast Fourier Transform and Principal Component Analysis, J. of Chemometrics*, 1987, 1, p103-110.
- 11: Downey G., Robert P., Bertrand D., Kelly P.M., *Classification of Commercial Skim Milk Powders According to Heat Treatment Using Factorial Discriminant Analysis of Near Infrared Reflectance Spectra. Appl. Spect.*, 1990, 44(1), p150-155.
- 12: Article collectif, *Chemometrics, Anal. Chem. Review*, 1986, 58(5), p294R-315R.

- 13: Lippmann R.P., **An Introduction to Computing with Neural Nets.** *IEEE ASSM Magazine.* Avril 1987, 4, p4-22.
- 14: Hoskins J.C., Himmelblau D.M., **Artificial Neural Network Models of Knowledge Representation in Chemical Engineering.** *Comput. Chem. Engng.* 1988, 12(9/10), p881-890.
- 15: Long J.R., Gregorion V.G., Gemperline P.J., **Spectroscopic Calibration and Quantification Using Artificial Neural Networks.** *Anal. Chem.* 1990, 62(17), p1791-1797.
- 16: Bos M., Bos A., Van Der Linden W.E., **Processing of Signals from an Ion-selective Electrode Array by a Neural Network.** *Anal. Chim. Acta*, 1990, 233, p31-39.
- 17: Beckman R.J., Cook R.D., **Outlier.....s,** *Technometrics*, 1983, 25(2), p119-163.
- 18: Maris M.A., Brown C.W., Lavery D.S., **Nonlinear Multicomponent Analysis by Infrared Spectrometry,** *Anal. Chem.*, 1983, 55(11), p1694-1703.
- 19: Kisner H.J., Brown C.W., Kavarnos G.J., **Multiple Analytical Frequencies and Standards for the Least-Squares Spectrometric Analysis of Serum Lipids,** *Anal. Chem.*, 1983, 55(11), p1703-1707.
- 20: Alfaro G., **Thèse de doctorat 1990, Université de Louvain-La-Neuve, (Belgique).**
- 21: Savitzky A., Golay M.J.E., **Smoothing and Differentiation of Data by Simplified Least Squares Procedures.** *Anal. Chem.*, 1964, 36(8), p1627-1639.
- 22: Steinier J., Termonia Y., Deltour J., **Comments on Smoothing and Differentiation of Data by Simplified Least Squares Procedure.** *Anal. Chem.*, 1972, 44(11), p1906-1909.
- 23: Maddams W.F., Mead W.L., **The Measurement of Derivative IR spectra (I/II),** *Spectrochimica Acta*, 1982, 38A(4), p437-457.
- 24: Jochum C., Jochum P., Kowalski B.R., **Error Propagation and Optimal Performance in Multicomponent Analysis,** *Anal. Chem.*, 1981, 53(1), p85-92.
- 25: Brezinsky C., **Algorithmique Numérique,** Ed. Ellipse 1988, 192 pages.
- 26: Besley D., Kuh E., Welsch R., **Regression Diagnostics: Identifying Influential Data and Sources of Collinearity,** Wiley, New-York 1980.

- 27: Wold S., Ruhe A., Wold H., Dunn W.J., **The Collinearity Problem in Linear Regression. The PLS Approach to Generalized Inverses.** *SIAM J. Sci. Stat. Comput.*, 1984, 5(3), p735-743.
- 28: Saarinen P., Kauppinen J., **Multicomponent Analysis of FT-IR Spectra,** *Appl. Spectr.*, 1991, 45(6), p953-963.
- 29: Lascaux P., Théodor R., **Analyse Numérique Matricielle Appliquée à l'Art de l'Ingénieur,** Ed. Masson, Paris 1986, tome 1, 400 pages.
- 30: Sibony M., Mardon J-Cl., **Systèmes Linéaires et non Linéaires,** Ed. Hermann 1988, tome 1 et 2.
- 31: Dagnelie P., **Théorie et Méthodes Statistiques,** Presses Académiques de Gembloux, Volume 1, 378 pages.
- 32: Morineau A., **Régressions Robustes, Méthodes d'Ajustement et de Validation.** *Rev. Stat. Appl.*, 26(3), p5-28.
- 33: Morineau A., Lebart L., Fénelon J-P., **Traitement des Données Statistiques,** Ed. Dunod, Paris 1982.
- 34: Tukey J.W., **Instead of Gauss-Markow Least Squares, What?** *App. Statistics*, Gupta Editor, North-Holland publishing company, 1975, p351-372.
- 35: Martens H., Naes T., **Multivariate Calibration,** Wiley, New-York 1988, 395 pages.
- 36: Haaland D.M., **Theoretical Comparison of Classical and Inverse Least-Squares Methods for Quantitative Infrared Spectroscopy,** *Proc. SPIE*, 1985, 553, p241-242.
- 37: Haaland D.M., Thomas E.V., **Partial Least-Squares Methods Spectral Analyses. 1: Relation to other Quantitative Calibration Methods and the Extraction of Qualitative Information.** *Anal. Chem.*, 1988, 60(11), p1193-1202.
- 38: Sharaf M.A., Illman D.L., Kowalski B.R., **Chemometrics,** Wiley, New-York, 1986.
- 39: Malinowski E.R., Howery D.G., **Factor Analysis in Chemistry,** Wiley, New-York, 1980.
- 40: Lindberg W., Persson J.-A., Wold S., **Partial Least Squares Method for Spectrofluorimetric Analysis of Mixtures of Humic Acid and Ligninsulfonate.** *Anal. Chem.*, 1983, 55(4), p643-648.

- 41: Haaland D.M., Easterling R.G., Improved Sensitivity of Infrared Spectroscopy by the Application of Least Squares Methods, *Appl. Spect.*, 1980, 34(5), p539-548.
- 42: Anderson R.J., Griffiths P.R., Errors in Absorbance Measurements in Infrared Fourier Transform Spectrometry because of Limited Instrument Resolution, *Anal. Chem.*, 1975, 47(14), p2339-2347.
- 43: Krutchkoff R.G., Classical and Inverse Regression Methods of Calibration, *Technometrics*, 1967, 9, p425-439.
- 44: Sundberg R., When is the Inverse Regression Estimator MSE-Superior to the Standard Regression Estimator in Multivariate Controlled Calibration Situations?, *Statistics and Probability Letters*, 1985, 3(2), p75-79.
- 45: Wetzel D.L., Near Infrared Reflectance Analysis: Sleeper among Spectroscopic Techniques, *Anal. Chem.*, 1983, 55(12), 1165A-1176A.
- 46: Nyden M.R., Forney G.P., Chittur K., Spectroscopic Quantitative Analysis of Strongly Interacting Systems: Human Plasma Protein Mixtures. *Appl. Spect.*, 1988, 42(4), p588-594.
- 47: Geladi P., Kowalski B.R., Partial Least Squares Regression: a Tutorial. *Anal. Chim. Acta*, 1986, 185, p1-17.
- 48: Donahue S.M., Brown C.W., Successive Average Orthogonalization of Spectral Data, *Anal. Chem.*, 1991, 63(10), p980-985.
- 49: De Lagarde J., *Initiation à l'Analyse des Données*, Dunod, 1983, 157 pages.
- 50: Wold S., Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics*, 1978, 20(4), p397-405.
- 51: Eastment H.T., Krzanowski W.J., Cross-Validatory Choice of the Number of Components from a Principal Component Analysis. *Technometrics*, 1982, 24(1), p73-77.
- 52: Martens H., Ph.D dissertation, Technical University of Norway, Trondheim (1985).
- 53: Frank I.E., Kalivas J.H., Kowalski B.R., Partial Least Squares Solutions for Multicomponent Analysis, *Anal. Chem.*, 1983, 55(11), 1800-1804.

- 66: Messerschmidt R.G., **Complete Elimination of Specular Reflectance in Infrared Diffuse Reflectance Measurements**, *Appl. Spect.*, 1985, 39(4), p737-739.
- 67: Brimmer P.J., Griffiths P.R., Harrick N.J., **Angular Dependence of Diffuse Reflectance Infrared Spectra. Part I: FT-IR Spectrogoniometer**, *Appl. Spect.*, 1986, 40(2), p258-265.
- 68: Brimmer P.J., Griffiths P.R., **Angular Dependence of Diffuse Reflectance Infrared Spectra. Part II: Effect of polarization**, *Appl. Spect.*, 1987, 41(5), p791-797.
- 69: Brimmer P.J., Griffiths P.R., **Angular Dependence of Diffuse Reflectance Infrared Spectra. Part III: Linearity of Kubelka-Munk Plots**, *Appl. Spect.*, 1988, 42(2), p242-247.
- 70: Hembree D.M., Smyrl H.R., **Anomalous Dispersion Effects in Diffuse Reflectance Infrared Fourier Transform Spectroscopy: a Study of Optical Geometries**, *Appl. Spect.*, 1989, 43(2), p267-274.
- 71: Brackett J.M., Azarraga L.V., Castles M.A., Rogers L.B., **Matrix Materials for Diffuse Reflectance Fourier Transform Infrared Spectrometry of Substances in Polar Solvents**, *Anal. Chem.*, 56(12), p2007-2010.
- 72: Iwaoka T., Wang S.H., Griffiths P.R., **Diffuse Reflectance Infrared Spectrometry of Inorganic Materials**, *Spectrochim. Acta*, 1985, 41A(1/2), p37-41.
- 73: TeVrucht M.L.E., Griffiths P.R., **Mixed Alkali Halide Dilution Matrices for DRIFT: Effect of Particle Size**, *Appl. Spect.*, 1989, 43(8), p1293-1294.
- 74: Ibbett R.N., Ph.D. 1987, *Université d'East Anglia (USA)*.
- 75: Hamadeh I.M., Yeboah S.A., Trumbull K.A., Griffiths P.R., **Preparation of Calibration Standards for Quantitative Diffuse Reflectance Infrared Spectrometry**, *Appl. Spect.*, 1984, 38(4), p486-491.
- 76: Yeboah S.A., Wang S., Griffiths P.R., **Effect of Pressure on Diffuse Reflectance Infrared Spectra of Compressed Powders**, *Appl. Spect.*, 1984, 38(2), p259-264.
- 77: TeVrucht M.L.E., Griffiths P.R., **A Simple Sample Packing Accessory for Diffuse Reflectance Infrared Spectrometry**, *Appl. Spect.*, 1989, 43(8), p1492-1494.

- 54: Lorber A., Wangen L.E., Kowalski B.R., A Theoretical Foundation for the PLS Algorithm. *J. of Chemometrics*, 1987, 1, p19-31.
- 55: Thomas E.V., Haaland D.M., Comparison of Multivariate Calibration Methods for Quantitative Spectral Analysis, *Anal. Chem.*, 1990, 62(10), p1091-1099.
- 56: Van Arendonk M.D., Skogerboe R.K., Grant C.L., Correlation Coefficient for Evaluation of Analytical Calibration Curves. *Anal. Chem.*, 1981, 53, p2349-2350.
- 57: Haaland D.M., Thomas E.V., Partial Least-Squares Methods for Spectral Analyses. 2. Application to Simulated and Glass Spectral Data, *Anal. Chem.*, 1988, 60(11), p1202-1208.
- 58: Fuller M.P., Griffiths P.R., Diffuse Reflectance Measurements by Infrared Fourier Transform Spectrometry, *Anal. Chem.*, 1978, 50(13), p1906-1910.
- 59: Hecht H.G., The Interpretation of Diffuse Reflectance Spectra, *J. of Research of the National Bureau of Standards*, 1976, 80A(4), p567-583.
- 60: Kubelka P., Munk F., Ein Beitrag zur Optik der Farbanstriche, *Z. Tech. Phys.*, 1931, 12, p593-601.
- 61: Reinecke D., Jansen A., Fister F., Schernau U., Quantitative Determination of Organic Compounds by Diffuse Reflectance Fourier Transform Infrared Spectrometry., *Anal. Chem.*, 1988, 60(11), p1221-1224.
- 62: Kortüm G., *Reflectance Spectroscopy*, 1969, Springer Verlag, Berlin, Heidelberg, New-York.
- 63: Hattori T., Shirai K., Niwa M., Murakami Y., Quantitative Interpretation of Infrared Diffuse Reflectance Spectra over Whole Concentration Range, *Bull. Chem. Soc. Jpn.*, 1981, 54(7), p1964-1967.
- 64: Richter W., Fourier Transform Reflectance Spectrometry between  $8000\text{ cm}^{-1}$  and  $800\text{ cm}^{-1}$  Using an Integrating Sphere., *Appl. Spect.*, 1983, 37(1), p32-38.
- 65: Depeker C., Thèse de doctorat 1988, *Université des Sciences et Techniques de Lille Flandres Artois*, (France).

- 78: Mark H., Workman J., **Effect of Repack on Calibrations Produced for NIR Reflectance Analysis**, *Anal. Chem.*, 1986, 58(7), p1454-1459.
- 79: Velapoldi R.A., Tvedt J.E., Christy A.A., **Sample Rotation in Diffuse Reflectance Fourier Transform Infrared Spectroscopy**, *Rev. Sci. Instrum.*, 1987, 58(6), p1126-1128.
- 80: Harrick N.J., **Internal Reflexion Spectroscopy**, Harrick Scientific Corporation, 1979, Ossining (New-York), 327 pages.
- 81: Iwamoto R., Ohta K., **Quantitative Surface Analysis by Fourier Transform Attenuated Total Reflection Infrared Spectroscopy**, *Appl. Spect.*, 1984, 38(3), p359-365.
- 82: Fredericks P.M., Lee J.B., Osborn P.R., Swinkels D.A., **Materials Characterization Using Factor Analysis of FT-IR Spectra. Part2: Mathematical and Statistical Considerations**, *Appl. Spect.*, 1985, 39(2), p311-316.
- 83: Geladi P., Mc Dougall D., Martens H., **Linearization and Scatter-Correction for NIR Reflectance Spectra of Meat**, *Appl. Spect.*, 1985, 39(3), p491-500.
- 84: Robert P., Bertrand D., **Réduction de l'Influence de la Granulométrie en Spectroscopie Proche Infrarouge par des Traitements Mathématiques**, *Science des Aliments*, 1985, 5, p501-517.
- 85: Miller C.E., Naes T., **A Pathlength Correction Method for NIR Spectroscopy**, *App. Spect.*, 1990, 44(5), p895-898.
- 86: Dull G. G., Giangiacoimo R., **Determination of Individual Simple Sugars in Aqueous Concentrations by Near Infrared Spectroscopy**, *Journal of Food Science*, 1984, 49, p1601-1603.
- 87: VdF (Association of the German Fruit Juice Industry), **RSK-Values, The Complete Manual, Guide Values of Specific Numbers, Including the Revised Methods of Analysis**, Verlag Flüssiges Obst Gmgh, D-5429 Schönborn, Germany, 1987.
- 88: Hannah R.W., **Introduction to Data Processing and Sources of Error**, *Advances in Applied Fourier Transform Infrared Spectroscopy*, 1988, Wiley and Sons, p1-42.

