

50376
1997
336

N°2085

THÈSE

présentée à

L'UNIVERSITÉ DES SCIENCES ET TECHNOLOGIES DE LILLE

pour obtenir le titre de

DOCTEUR

Mention : SPECTROCHIMIE

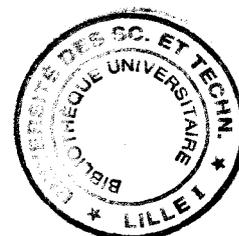
par

Ludovic DUPONCHEL

**VALIDATION PAR LES MÉTHODES CHIMIOMÉTRIQUES D'UN
ANALYSEUR SPECTROMÉTRIQUE DE LAIT TRAVAILLANT
DANS LE PROCHE INFRAROUGE.
TRANSFERT DE CALIBRATION EN VUE DE
DÉVELOPPEMENTS INDUSTRIELS.**

Soutenue le 13 Octobre 1997 devant la commission d'examen :

- Président : M. LEGRAND (Université des Sciences et Technologies de Lille).
Rapporteurs : M. MEURENS (Université Catholique de Louvain-la-Neuve, Belgique).
: M. FEINGERG (Institut National d'Agronomie de Paris Grignon).
Examineurs : M. HUVENNE (Université des Sciences et Technologies de Lille).
: Mme. CONSTANT (Université des Sciences et Technologies de Lille).
: M. DARIER (Société JOBIN-YVON).
: M. GRAPIN (I.N.R.A. Poligny).
: M. LECONTE (Fédération Française du Contrôle Laitier).



A Laurence et Guillaume.

Ce travail a bénéficié du financement CIFRE de l'A.N.R.T. entre la Société JOBIN-YVON et le Laboratoire de Spectrochimie Infrarouge et Raman du C.N.R.S. (Unité 2631L) dirigé par M. BOUGEARD, Directeur de Recherches.

J'exprime ma profonde gratitude à M. LEGRAND, Professeur de l'Université de Lille et Directeur de L'Ecole Universitaire D'Ingénieurs de Lille, de l'honneur qu'il me fait en acceptant la présidence de ce jury.

Cette recherche a été réalisée sous la direction de M. le Professeur HUVENNE, responsable du département IRTF du LASIR. Je tiens à lui exprimer toute ma reconnaissance pour la confiance qu'il m'a témoignée en m'acceptant dans son équipe en 1992. Tout au long de ces années, il a su me faire partager sa pédagogie, ses compétences en spectroscopie et en chimométrie.

J'adresse mes remerciements à M. MEURENS, Professeur à l'Université catholique de Louvain-la-Neuve en Belgique, pour l'honneur qu'il me fait en acceptant d'être rapporteur de ce travail.

Je suis très sensible à l'honneur qu'à bien voulu me faire M. FEINBERG, Directeur de Recherches à l'Institut National d'Agronomie de Paris-Grignon, en acceptant de faire partie de ce jury.

Je remercie vivement Mme. CONSTANT, Professeur de l'Université de Lille, d'avoir accepté d'examiner ce mémoire.

Mes remerciements vont à M. DARIER, Directeur du département optique de la Société JOBIN-YVON, pour l'intérêt qu'il a toujours porté à ce travail et l'honneur qu'il me fait en le jugeant.

J'exprime également toute ma reconnaissance à M. GRAPPIN, Directeur de la station de recherche INRA à Poligny, pour sa connaissance du lait indispensable à l'aboutissement de ce projet et sa présence dans ce jury.

Je remercie Melle DUPUY pour m'avoir initié à la spectroscopie et à la chimométrie dès mon entrée au LASIR.

Je tiens à exprimer mes remerciements à M. SOMBRET qui m'a fait partager ses compétences lors de nos fructueuses discussions sur l'instrumentation.

J'exprime ma reconnaissance à M. BAZIN, M. LECONTE et à Melle DEBRABANT, de la Fédération Française du Contrôle Laitier, pour les nombreuses réunions constructives qui ont guidé ce projet.

Je remercie M. LHOMME, Directeur du laboratoire régional du contrôle laitier de St-Pol sur Ternoise, pour son accueil et ses conseils avisés donnés lors de nos campagnes de mesures.

Que Messieurs THEVENON et CHAUVIN, Ingénieurs de Recherches, trouvent ici l'expression de ma profonde et sincère gratitude pour leurs compétences en instrumentation qu'ils ont su me faire partager. Ces remerciements seraient incomplets si j'omettais Messieurs DUPRE, LOZACK et VITU pour la bonne ambiance qu'ils assurent au sein de l'équipe informatique.

Je remercie toutes les personnes du LASIR et de la Société JOBIN-YVON pour la sympathie qu'ils m'ont témoignée.

Une mention spéciale revient à LYDIE et CHRISTINE, deux très bonnes amies ...

TABLE DES MATIÈRES

INTRODUCTION.....	6
I LE LAIT ET SA TECHNOLOGIE	9
1.1 LES CARACTÉRISTIQUES ESSENTIELLES DU LAIT	9
1.1.1 <i>Introduction.....</i>	9
1.1.2 <i>Production du lait.....</i>	10
1.2 LA CONSTITUTION DU LAIT	10
1.3 LES LIPIDES	11
1.3.1 <i>Structure chimique</i>	11
1.3.2 <i>Les globules gras.....</i>	12
1.3.3 <i>Dosage des matières grasses.....</i>	13
1.4 LES PROTIDES.....	14
1.4.1 <i>Structure chimique</i>	14
1.4.2 <i>Les micelles de protéines.....</i>	14
1.4.3 <i>Dosage des protéines</i>	15
1.5 LES GLUCIDES	16
1.5.1 <i>Structure chimique</i>	16
1.5.2 <i>Dosage des glucides.....</i>	17
1.6 LA TRAITE	17
II LA SPECTROSCOPIE PROCHE INFRAROUGE	19
2.1 PRINCIPES FONDAMENTAUX	19
2.1.1 <i>L'oscillateur harmonique continu.....</i>	19
2.1.2 <i>L'oscillateur harmonique quantifié.....</i>	21
2.1.3 <i>L'oscillateur anharmonique quantifié.....</i>	23
2.2 ATTRIBUTIONS DES BANDES SPECTRALES DANS LE NIR	28
2.3 L'INSTRUMENTATION PROCHE INFRAROUGE.....	31
2.3.1 <i>Développement chronologique de l'instrumentation.....</i>	31
2.3.2 <i>Histoire du proche infrarouge et des produits laitiers.....</i>	32
2.3.3 <i>L'instrumentation NIR.....</i>	33

2.3.3.1 Les analyseurs spectrométriques.....	34
2.3.3.1.1 Les filtres interférentiels	34
2.3.3.1.2 Les filtres opto-acoustiques.....	36
2.3.3.1.3 L'interférométrie.....	37
2.3.3.1.4 Les monochromateurs	39
2.3.3.2 Les fibres optiques	44
2.3.3.3 La chaîne électronique de détection.....	46
2.3.3.4 Les techniques d'échantillonnage.....	49
2.4 AVANTAGES ET INCONVÉNIENTS DE LA SPECTROSCOPIE NIR.....	52
III TRAITEMENTS MATHÉMATIQUES DES DONNÉES SPECTRALES	54
3.1 INTRODUCTION.....	54
3.2 NOTATIONS	54
3.3 PRINCIPE	55
3.4 LE PRÉ-TRAITEMENT DES DONNÉES SPECTRALES.....	58
3.4.1 <i>La dérivée</i>	59
3.4.2 <i>La normalisation par référence interne</i>	62
3.4.3 <i>La normalisation par la méthode « MSC »</i>	63
3.4.6 <i>La normalisation par la méthode « SNV »</i>	66
3.4.7 <i>La « normalisation simple »</i>	68
3.4.8 <i>La normalisation par la méthode « MEA »</i>	68
3.5 LES ALGORITHMES DE RÉGRESSION	69
3.5.1 <i>Introduction</i>	69
3.5.2 <i>La régression linéaire simple</i>	69
3.5.3 <i>Les régressions multilinéaires</i>	71
3.5.3.1 La méthode CLS	71
3.5.3.1 La méthode ILS	73
3.5.4 <i>Les méthodes de condensation des données</i>	76
3.5.4.1 <i>Introduction</i>	76
3.5.4.2 <i>La régression PCR</i>	78
3.5.4.2 <i>La régression PLS</i>	80
3.5.5 <i>Comparaison des algorithmes</i>	84
IV VALIDATION DE MAQUETTES DE FAISABILITÉ.....	85
4.1 INTRODUCTION.....	85

4.2	ÉTAT ACTUEL DES CONTRÔLES LAITIERS	85
4.3	LE CAHIER DES CHARGES.....	86
4.4	INTÉGRATION ET UTILISATION DE L'APPAREIL DANS L'ÉTABLE	87
4.5	LES APPORTS DE CETTE NOUVELLE MÉTHODE DE DOSAGE.	89
4.6	DESCRIPTION DE LA MAQUETTE À VALIDER	89
4.7	LA CHIMIOMÉTRIE : OUTIL DE VALIDATION DES PARAMÈTRES INSTRUMENTAUX.....	92
4.7.1	<i>Introduction</i>	92
4.7.2	<i>L'échantillon</i>	92
4.7.3	<i>La stabilité du signal</i>	93
4.7.4	<i>La linéarité du signal</i>	94
4.7.5	<i>L'analyse quantitative</i>	95
4.8	LA SONDE « TRANSMISSION PURE ».....	95
4.8.1	<i>Introduction</i>	95
4.8.2	<i>La répétabilité</i>	96
4.8.3	<i>La linéarité</i>	99
4.8.4	<i>Discussion</i>	102
4.9	LA SONDE «TRANSFLEXION »	104
4.9.1	<i>Introduction</i>	104
4.9.2	<i>La répétabilité</i>	107
4.9.3	<i>La linéarité</i>	109
4.9.4	<i>L'analyse quantitative</i>	111
4.9.4.1	Dosage de solutions modèles de matières grasses	111
4.9.4.2	Dosage de solutions modèles de protéines.....	113
4.9.4.3	Dosage de laits crus	116
4.10	LA SONDE 0-45°	123
4.10.1	<i>Introduction</i>	123
4.10.2	<i>Analyse quantitative de solutions modèles</i>	127
4.10.3	<i>Analyse quantitative de laits crus</i>	131
4.11	LA SONDE « DIFFUSION ».....	133
4.11.1	<i>Introduction</i>	133
4.11.2	<i>Analyse quantitative de solutions modèles</i>	136
4.11.3	<i>Analyse quantitative de laits crus</i>	138

4.12 LE MONOCHROMATEUR	141
4.12.1 Le filtre d'ordre	141
4.12.2 Les fentes	142
4.13 LA CHAÎNE ÉLECTRONIQUE DE DÉTECTION	143
4.14 CONCLUSION	146
V LE TRANSFERT DE CALIBRATION	148
5.1 INTRODUCTION	148
5.2 MISE EN PLACE D'UN TRANSFERT DE CALIBRATION	149
5.2.1 Les échantillons « standards »	149
5.2.1.1 Les critères de sélection	149
5.2.1.2 Les différents types de standards	150
5.2.1.3 Le nombre de standards	151
5.3 LA FAISABILITÉ DES ALGORITHMES	152
5.3.1 Introduction	152
5.3.2 Analyse quantitative sur l'instrument maître	156
5.3.3 Étude des algorithmes de transfert	157
5.3.3.1 Introduction	157
5.3.3.2 La correction des concentrations	158
i) Principe	158
ii) Application aux données synthétiques	158
5.3.3.3 La correction spectrale	161
5.3.3.3.1 La standardisation directe (DS)	161
i) Principe	161
ii) Application de la standardisation directe aux données synthétiques	163
5.3.3.3.2 La standardisation directe pas à pas (PDS)	164
i) Principe	164
ii) Application de la standardisation directe aux données synthétiques	165
5.3.3.3.2 La standardisation par la méthode de SHENK	167
i) Principe	167
ii) Application de la méthode de SHENK aux données synthétiques	170
5.3.3.3.3 Standardisation par un réseau de neurones	171
5.3.3.4 Synthèse des résultats	175
5.4 APPLICATION DU TRANSFERT AU DOSAGE D'HYDROCARBURES	176
5.4.1 Introduction	176
5.4.2 Analyse quantitative sur l'instrument maître	176

5.4.3 <i>Le transfert de la calibration</i>	180
5.5 TRANSFERT DE CALIBRATION SUR L'ANALYSEUR DE LAIT	182
5.5 CONCLUSION	185
CONCLUSION GÉNÉRALE	187
TABLE DES ILLUSTRATIONS	190
INDEX	195
BIBLIOGRAPHIE	199

INTRODUCTION

Dans le secteur agro-alimentaire, la connaissance rapide des caractéristiques chimiques des matières premières, des produits en cours de réalisation ou des produits finis, est devenue une nécessité. La mise en place de normes pour assurer la qualité des produits a largement contribué à cette tendance. Le proche infrarouge, longtemps considéré comme un domaine spectral inexploitable par les fondamentalistes, s'est révélée dans les années 70 grâce à l'essor de la micro-informatique et des techniques statistiques de régressions multivariées. Elle est alors devenue une méthode de choix pour l'analyse quantitative et qualitative dans le secteur de l'agro-alimentaire. En fait, la spectroscopie proche infrarouge est particulièrement bien adaptée à l'analyse spectrale sur les process. Les fibres optiques permettent tout d'abord de délocaliser le spectromètre du lieu d'analyse sur des distances pouvant atteindre un kilomètre. Les faibles extinctions des absorptions proche infrarouge autorisent l'étude d'échantillons sur des chemins optiques relativement importants jusqu'aux produits finis conditionnés. L'eau, omniprésente en agro-alimentaire, masque le spectre moyen infrarouge par de nombreuses bandes d'absorption sur un large domaine alors qu'elles sont très localisées en proche infrarouge.

Jusqu'à une date récente, le lait a été payé au volume ou au poids comme si sa composition était constante et sa qualité invariable. Ce paiement était injuste pour les éleveurs soucieux de la richesse et de la qualité du lait. A partir de 1969, de nouvelles lois sont adoptées. Le paiement est effectué à partir du dosage des matières grasses et des protéines. A l'heure actuelle, les quotas laitiers poussent d'autant plus les éleveurs à optimiser la qualité de la production laitière. Le contrôle des teneurs en matières grasses et protéines du lait est donc très important tant pour s'assurer de la qualité de sa valeur marchande que pour l'orienter vers l'unité de production la plus valorisante. Le contrôle du lait est actuellement effectué par prélèvement manuel d'un volume représentatif de la traite sur chaque animal du troupeau. Tous les échantillons sont ensuite centralisés pour l'analyse chimique dans un laboratoire qui est souvent régional. Ce protocole pourtant simple a de nombreux inconvénients. A la faible fréquence des contrôles s'ajoute en effet le problème de la représentativité de l'échantillon prélevé manuellement. Une inversion sur l'identification des flacons de prélèvement est encore possible. Mais l'inconvénient majeur reste le transport des échantillons qui provoque des secousses et développe ainsi un barattage. A ce stade, le lait n'a plus ses propriétés

physico-chimiques d'origine car les réactions de lipolyse et de dénaturation des protéines sont facilitées.

La volonté d'améliorer le contrôle a induit le développement d'un analyseur permettant de doser les matières grasses et les protéines du lait pendant la traite sur chaque vache du troupeau. Les instrumentalistes associés au projet ont conçu des maquettes de faisabilité de l'analyseur respectant un cahier des charges défini. L'objectif de notre étude était de contrôler les performances des appareils proposés pour l'analyse quantitative des matières grasses et des protéines. Chaque étape du développement de l'instrumentation a été validée par des résultats issus de la chimométrie. Dans une première partie, nous présenterons les domaines nécessaires à la conception de cet appareil. La pluridisciplinarité du sujet nous impose en effet la connaissance de la technologie laitière, de la zoologie, de la spectroscopie proche infrarouge et de l'analyse quantitative par les méthodes indirectes. Dans une seconde partie, les résultats de recherche sur les validations successives des modifications instrumentales sont présentés. La sonde optique, considérée comme la pièce maîtresse de l'analyseur, sera principalement étudiée. L'optimisation du monochromateur et de la chaîne électronique de détection sera développée plus succinctement.

La dernière partie de cette recherche est consacrée au transfert de calibration. En fait, l'utilisation des régressions multivariées en spectroscopie a fait apparaître deux problèmes. Le premier est le transfert de l'équation de calibration d'un premier appareil (dit « maître ») vers un ou plusieurs autres (dits « esclaves »). Le cas typique est le développement d'une analyse quantitative sur un appareil au laboratoire de recherche devant être transférée sur des analyseurs en ligne sur le process. Le deuxième problème survient lorsque les caractéristiques d'un appareil varient au cours du temps ou encore lorsqu'une panne nécessite le remplacement d'un composant de la chaîne de mesure. Dans ces deux cas, la réponse spectrale change et l'utilisation de la calibration initialement développée fait naître de larges erreurs pour le dosage des composés d'intérêt. Les algorithmes de transfert de calibration corrigent les concentrations ou les spectres afin de retrouver le niveau d'erreur initial évitant ainsi une nouvelle calibration de l'analyseur. L'objectif de cette recherche est l'étude systématique des nombreux algorithmes proposés dans la littérature dont on connaît peu l'efficacité relative. Nous exploiterons successivement des données spectrales synthétiques, des spectres de produits pétroliers et finalement des spectres issus du dosage du lait sur l'analyseur proche infrarouge afin de vérifier le potentiel de ces algorithmes. Nous indiquerons finalement

d'autres voies de recherche pour le transfert par la modification de certains algorithmes ou même la génération de nouveaux.

I LE LAIT ET SA TECHNOLOGIE

1.1 Les caractéristiques essentielles du lait

1.1.1 Introduction

Le lait est un liquide sécrété par les glandes mammaires des femelles mammifères, après la naissance du jeune. La fonction naturelle du lait est d'être l'aliment exclusif des jeunes mammifères pendant la période critique de leur existence, après la naissance, alors que la croissance est rapide et qu'il ne peut lui être substitué d'autres aliments. La grande complexité du lait répond à cette fonction⁽¹⁾. C'est une émulsion de matières grasses, sous forme globulaire, dans un liquide (matrice) qui présente des analogies avec le plasma sanguin. Cette matrice est elle-même une suspension de matières protéiques dans un sérum. Ce dernier est une solution neutre contenant principalement des sels.

Il y a donc dans le lait quatre types de constituants importants :

- les lipides, constitués essentiellement de graisses ordinaires (triglycérides),
- les protéines (caséines, albumines et globulines)
- les glucides (essentiellement le lactose),
- les sels.

Mais il y a aussi de nombreux constituants présents en quantité minime : lécithines, vitamines, enzymes, nucléotides, gaz dissous, etc., néanmoins importants pour l'activité biologique. De nombreux facteurs influent sur la composition du lait. Elle peut varier au cours du cycle de lactation^(2,3). A l'époque de la naissance, la mamelle sécrète le colostrum, liquide différent du lait principalement dans ses parties protéiques et salines. De façon générale, les taux en matières grasses et matières protéiques évoluent régulièrement au cours du cycle. L'état de santé de la vache influe également sur la qualité de son lait (laits pathologiques). L'alimentation peut être un facteur de variation important. En effet, les éleveurs optimisent les taux en matières grasses et protéines par des rations alimentaires contenant des éléments spécifiques⁽⁴⁾. Les saisons et la température sont aussi à prendre en compte. Le dernier facteur est génétique : la race.

1.1.2 Production du lait

L'homme a développé la sécrétion lactée des mammifères domestiques de façon intensive. Elle ne touche qu'un très petit nombre d'espèces (une douzaine sur 4000). Ainsi dans certaines espèces, il y a des « races laitières » formées par la sélection et les méthodes d'exploitation. Entre la production d'un individu inculte et celle d'une vache exceptionnelle, il peut y avoir un rapport de 1 à 20. En Asie par exemple, la production moyenne d'une vache n'atteint que 200 kg/an alors que dans les pays d'Europe occidentale, cette production se situe vers 4000 kg/an.

Les espèces laitières changent selon les contrées, mais l'espèce bovine est de loin la plus répandue et son aire de propagation s'accroît toujours. Le lait de vache est donc le plus important. A l'échelle mondiale la prédominance du lait de vache est écrasante. Le tableau ci-dessous établit les productions mondiales de lait en fonction des espèces.

Race	Production (millions de tonnes)	Part relative
Vache	420000	100
Buflesse	26000	6
Brebis	7000	2
Chèvre	7000	2

Tableau 1 : Production laitière mondiale (établie par le F.A.O en 1979).

1.2 La constitution du lait

Une des propriétés fondamentales du lait est d'être un mélange hétérogène. Mais l'hétérogénéité du lait est une notion vulgaire. On sait que le lait abandonné à lui-même à la température ambiante se sépare progressivement en trois parties :

- la crème : couche de globules gras rassemblés par effet de gravité ;
- le caillé : caséine coagulée, conséquence des activités microbiennes ;
- le sérum ou petit-lait, contenant les produits solubles, qui se sépare du caillé.

Le lait présente donc différentes phases. Rappelons que l'on appelle « phase » dans un milieu hétérogène, toute partie qui constitue une matière homogène, quel que soit son état de division. Dans le lait, nous pouvons donc en distinguer trois :

- l'émulsion de matière grasse, sous forme globulaire,
- la suspension de caséine, liée à des sels minéraux, sous forme de « micelles »
- la solution ou phase hydrique, qui forme le milieu massique, continu.

Le lait peut contenir des gaz dissous, mais ils ne forment pas réellement une phase gazeuse distincte. Le tableau ci-dessous donne la composition typique d'un lait de vache ainsi que les propriétés physiques des différents constituants.

Substances	Composition (g/l)	Etat physique des constituants
Eau	905	Eau libre (solvant) + eau liée (3,7%)
Glucides : lactose	49	Solution
Lipides	35	
Matières grasses proprement dite	34	Emulsion de globules gras
Lécithine (phospholipides)	0.5	
Partie insaponifiable (stéroïls, carotènes, tocophérols)	0.5	
Protides	34	Suspension micellaire
Caséine	27	
Protéines "solubles" (globulines, albumines)	5.5	
Substances azotées non-protéiques	1.5	
Sels	9	
de l'acide citrique	2	
de l'acide phosphorique	2.6	
de l'acide chlorhydrique	1.7	
Extrait sec total	127	
Extrait sec non gras	92	

Tableau 2 : Composition moyenne du lait de vache.

1.3 Les lipides

1.3.1 Structure chimique

La matière grasse (MG) du lait est principalement constituée de triglycérides. Ce sont des esters du glycérol et d'acides gras aliphatiques (Figure 1). Les radicaux acides R peuvent être identiques ou différents (glycérides homogènes ou mixtes). Ces esters peuvent redonner par hydrolyse leurs constituants de départ. Elle peut être effectuée par action enzymatique des lipases. C'est cette réaction qui intervient notamment dans le rancissement du lait. Les alcalins hydrolysent aussi les esters dans une réaction que l'on appelle saponification car le sel R-COONa est un savon.

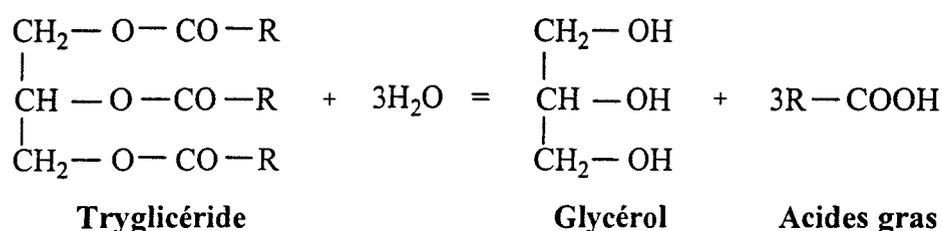


Figure 1 : Les lipides dans le lait

La nature des acides gras R et leurs proportions respectives différencient les divers corps gras et déterminent leurs propriétés. La matière grasse du lait présente environ 150 acides gras donnant ainsi plus de 1300 espèces de triglycérides différents⁽⁵⁾. Les proportions en acides gras saturés et insaturés sont respectivement de 2/3 et 1/3. La matière grasse est le constituant du lait le plus variable en proportion (20 à 60 g/Kg de lait).

1.3.2 Les globules gras

La matière grasse se trouve dans le lait sous forme de globules sphériques, visibles au microscope optique. Ils ont tendance à se réunir en grappe. Leur diamètre varie avec les espèces. En général, les races bovines produisant un lait riche en graisse donnent une proportion élevée de gros globules et inversement. Le diamètre moyen des globules tend à diminuer avec l'avancement du cycle de lactation. Au cours d'une traite, les premiers volumes contiennent des globules plus petits en moyenne que ceux recueillis à la fin.

Les dimensions extrêmes observées vont de 0.1 à 20 microns. En fait, le diamètre des globules gras est rarement inférieur à 2 microns ni supérieur à 12. Dans le lait de vache, le diamètre moyen se situe entre 3 et 5 microns.

Ces globules sont hétérogènes. Ils sont essentiellement constitués par une micro-goutte de triglycérides, partiellement cristallisés à température ambiante, entourée d'une enveloppe complexe ou film d'absorption^(6,7). Cette membrane empêche la réunion des globules gras en une phase grasse continue. Ces globules jouent un rôle très important dans le phénomène de diffusion de la lumière dans le lait.⁽⁴⁴⁾.

1.3.3 Dosage des matières grasses

Les matières grasses peuvent être dosées par un grand nombre de techniques analytiques⁽⁸⁾. Il est tout d'abord nécessaire de détruire l'état globulaire pour pouvoir séparer la matière grasse ou l'extraire par un solvant^(9,10). Les acides concentrés et chauds sont les plus employés. On obtient ainsi, outre la destruction de la membrane globulaire, la dissolution de la caséine et une bonne séparation des deux phases. On peut classer ces dosages en deux familles :

Les méthodes volumétriques :

On mesure simplement le volume de la phase grasse, séparée de la phase aqueuse par centrifugation, dans des appareils spécialement gradués appelés « butyromètres ». Ce sont des méthodes de routine d'exécution rapide et suffisamment précises.

La méthode acido-butyrométrique, selon GERBER⁽¹¹⁾ en Europe et BABCOK⁽¹²⁾ en Amérique, reste la plus utilisée pour l'analyse du lait malgré l'emploi de réactif dangereux tels que l'acide sulfurique auquel on ajoute de l'alcool amylique en vue d'aider à la séparation des deux phases. La matière grasse « GERBER » est caractérisée par la présence d'esters isoamyliques et une proportion élevée d'acides gras libres et de diglycérides, par suite d'une forte hydrolyse. Cette méthode bénéficie d'une très longue expérience et de nombreux efforts de standardisation.

Les méthodes pondérales :

La graisse est extraite par des solvants, en général l'éther éthylique et l'éther de pétrole. L'extraction est effectuée, soit d'une manière discontinue par décantations successives dans des ampoules, soit d'une manière continue dans des appareils spéciaux à épuisement comme le « SOXHLET ». Après évaporation du solvant qui n'est qu'un véhicule transitoire, la graisse est pesée. Ce sont des méthodes précises mais d'exécution relativement longue. Le dosage « ROSE-GOTTLIEB⁽¹³⁾ » (méthode éthéro-ammoniacale) est le plus connu de cette famille. Il est choisi comme méthode de référence par les instances internationales. Les lipides sont libérés par l'ammoniaque et l'alcool.

1.4 Les protides

1.4.1 Structure chimique

Les substances azotées forment la partie la plus complexe du lait. En effet, sa composition n'est pas encore parfaitement connue. Du point de vue nutritionnel, les protides constituent une partie remarquable du lait. Ce sont des constituants fondamentaux des cellules et des matériaux biologiques très importants.

Les protéines (MP) sont des macromolécules dans lesquelles se trouvent réunies par l'unique liaison peptidique, en enchaînement non ramifié des unités représentées par 20 acides α -aminés⁽¹⁴⁾ (Figure 2). La structure primaire de la caséine α est donnée en Figure 3. Leurs propriétés découlent à la fois de leur composition et d'une structure spatiale particulière.

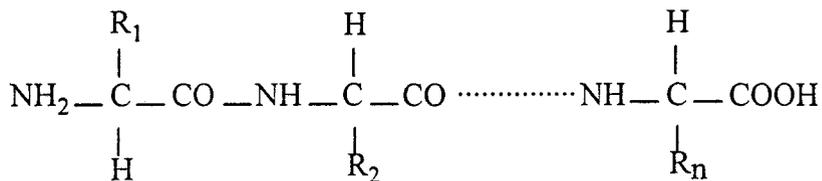


Figure 2 : Les protéines

MKLLILTCLVAVALARPKHPIKHQGLPQEVLNENLLRFFVAPFPEVFGKEKVNELSKD
 IGSESTEDQAMEDIKQMEAESISSSEEIVPNSVEQKHIQKEDVPSERYLGYLEQLLRK
 KYPQLEIVPNSAEERLHSMKEGIHAQQKEPMIGVNQELAYFYPELFRQFYQLDAYPSG
 AWYYVPLGTQYTDAPSFSDIPNPIGSENSEKTTMPLW

Figure 3 : Séquence des acides α -aminés de la caséine α .

Les protéines du lait sont essentiellement des caséines (78 % de la matière azotée). Elles possèdent un certain caractère acide et sont phosphorées. Dans le lait, la caséine se trouve engagée dans un complexe salin à l'état micellaire. Sa teneur peut varier en valeur absolue mais beaucoup moins que les matières grasses (20 à 45 g/kg de lait).

1.4.2 Les micelles de protéines

En présence d'ions comme le calcium, les caséines ont tendance à se regrouper pour

former des micelles^(15,16). On dit que les micelles sont des phosphocaséinates acides de calcium. Pour un lait normal, leur diamètre se situe entre 20 et 300 nm avec une moyenne de 100 nm. Comme les globules gras, ces micelles augmentent le phénomène de diffusion de la lumière.

1.4.3 Dosage des protéines

Il est pratiquement impossible de déterminer directement les matières azotées par l'isolement et la pesée.

La méthode de KJELDAHL

Il y a longtemps que la méthode de KJELDAHL^(17,18,19) est considérée comme la méthode de référence pour le dosage des protéines. Dans ce dosage, la matière organique est détruite par l'acide sulfurique et l'azote se retrouve sous forme de sulfate d'ammoniaque. Différents catalyseurs sont utilisés pour élever le point d'ébullition et accélérer la digestion. L'ammoniaque est déplacée par la soude, distillée et titrée par une liqueur acide déci- ou centi-normale. Le défaut de la méthode de KJELDAHL, qui est de ne pas doser complètement toutes les formes de l'azote, apparaît peu dans le cas du lait qui ne contient que très peu d'azote hétérocyclique. La méthode la plus exacte et la plus précise pour le dosage de l'azote est celle qui est utilisée dans l'analyse organique élémentaire : la méthode de DUMAS. Elle permet de doser toutes les formes de l'azote mais reste délicate à mettre en oeuvre. De ce fait, elle n'est plus utilisée dans les laboratoires spécialisés où l'on étudie les produits agro-alimentaires. La méthode de KJELDAHL est sûre, bien connue après une longue expérience dans ces deux versions micro- et macro-analytique.

De nombreuses variantes ont été proposées pour éviter la distillation, en titrant l'ammoniaque par colorimétrie : méthode iodométrique, méthode de NESSLER, méthode à la ninhydrine. Mais ces méthodes ne se sont pas répandues.

La méthode de KJELDAHL convient à des travaux précis portant sur un nombre d'échantillons restreint mais elle est difficile à mettre en place pour l'analyse de routine pour des raisons de temps et de coût.

La méthode au formol

Cette méthode revient à faire un titrage acidimétrique. Le volume de solution de soude titrée nécessaire pour neutraliser l'acidité résultant de l'addition de formol est proportionnel à la quantité de protéines et d'acides aminés libres présents. Cette méthode est rapide mais souffre d'un défaut de précision surtout sur le dosage de laits individuels.

La méthode de KOFRANY

Cette méthode a fait l'objet d'une application permettant le dosage en série. C'est une simplification du procédé KJELDAHL dont on conserve seulement la distillation alcaline, en supprimant la minéralisation acide. Tout l'azote des protéines n'est pas transformé en ammoniac, mais le rapport entre la quantité d'azote présente dans une protéine et celle qui est transformée en ammoniac est relativement constant. Les dosages obtenus sont de bonne qualité mais les méthodes colorimétriques sont moins onéreuses.

La méthode « Noir Amido »

La méthode colorimétrique « Noir Amido »^(20,21,22) a été longtemps recommandée pour les dosages en grandes séries. On fait réagir avec le lait un excès de solution du colorant Noir Amido dans un tampon phosphate-citrate à pH 2,45. Les volumes de lait et de colorant sont alors fixés une fois pour toutes. Après précipitation des protéines, on détermine l'excès du colorant non fixé en mesurant l'absorbance de la solution à 580 nm. Ainsi plus la quantité de protéines est grande plus l'absorbance de la solution sera faible. A partir de plusieurs étalons, nous pouvons établir une calibration entre les concentrations en protéines données par la méthode KJELDAHL et les absorbances des solutions à cette longueur d'onde. Pour un échantillon inconnu, la lecture de l'absorbance permet alors de prédire sa concentration.

1.5 Les glucides

1.5.1 Structure chimique

Le lactose est le seul glucide libre existant en quantité importante dans le lait⁽²³⁾. C'est un disaccharide constitué de glucose et de galactose. C'est aussi le composant majeur le plus simple et le plus constant en proportion. Le lait en contient entre 48 et 50g/l. Le lactose est un

facteur limitant pour la production du lait. En effet la quantité de lait dépend des possibilités de synthèse du lactose dans la mamelle. Le lactose se trouve tel quel en solution.

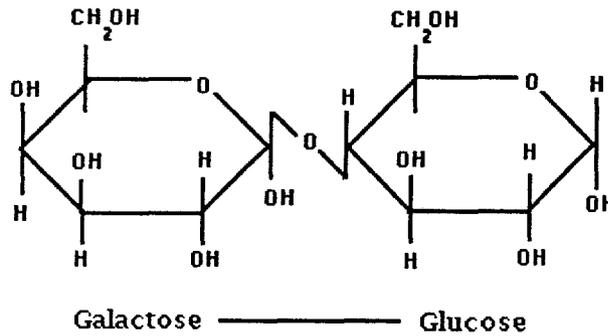


Figure 4 : Le lactose α

1.5.2 Dosage des glucides

Du fait de l'existence du groupe aldéhydrique libre, le lactose est un sucre réducteur. Il réduit notamment la liqueur cuprique alcaline de Fehling. Ce dosage est de plus en plus souvent remplacé par le dosage colorimétrique. Il est basé sur la réaction que donnent avec diverses substances organiques, les produits de dégradation des sucres réducteurs en milieu fortement acide (réaction de BIAL avec l'orcinol, réaction avec l'anthrone, etc.). Ces méthodes ont été à leur tour supplantées par les techniques chromatographiques en phase liquide ou gaz⁽²⁴⁾.

1.6 La traite

La connaissance de la traite sera indispensable à la compréhension du fonctionnement et de l'intégration de notre appareil de dosage dans les fermes. La figure ci-après schématise une salle de traite classique. Lors de la traite, l'éleveur fait entrer un premier groupe de vaches dans la salle. Par habitude, elles se placent souvent dans le même ordre sur les différentes places (stalles). Elles se trouvent le long de la salle en hauteur sur les quais de traite. L'éleveur travaille dans la fosse pour avoir les mamelles à sa hauteur. Il place aux mamelles de chaque vache un faisceau trayeur branché sur un système de pompe à vide. Une alternance de dépressions permet de simuler la succion et le massage du veau qui tète. Le lait éjecté est recueilli dans une canalisation en inox appelé lactoduc. L'ensemble de la traite se retrouve dans une citerne réfrigérée. Chaque faisceau trayeur possède un dispositif appelé compteur-

échantillonneur. Celui-ci ordonne tout d'abord au faisceau trayeur de se décrocher lorsque la traite de l'animal est terminée.

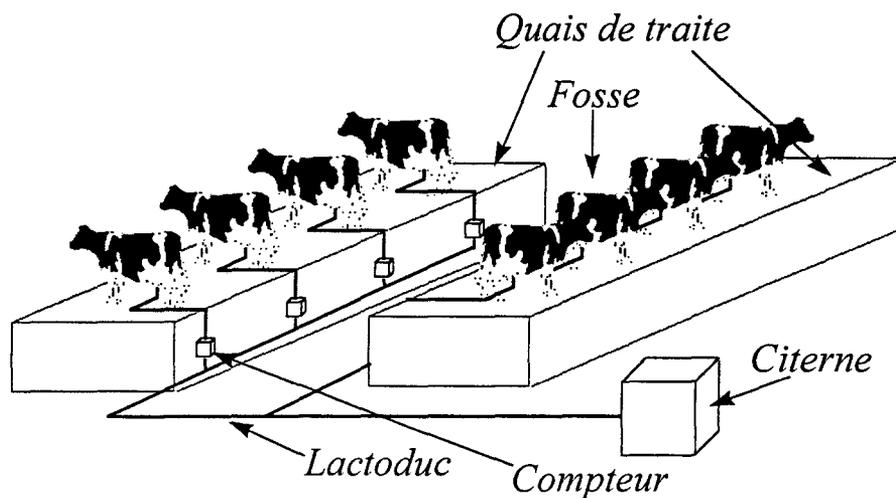


Figure 5 : Représentation d'une salle de traite.

Il mesure le volume total et la courbe d'éjection (volume en fonction du temps) du lait sécrété. Il prélève un faible volume de lait représentatif de la traite de l'animal considéré. Ce dernier volume de lait peut être utilisé pour le contrôle des teneurs en matières grasses et protéines. La traite du groupe est terminée lorsque tous les faisceaux sont décrochés. Les vaches sortent de la salle alors qu'un autre groupe les remplace.

II LA SPECTROSCOPIE PROCHE INFRAROUGE

2.1 Principes fondamentaux

Les liaisons chimiques sont élastiques. Ainsi, les atomes d'une molécule ne restent pas en position fixe les uns par rapport aux autres, mais au contraire, vibrent autour d'une position d'équilibre. La manière la plus simple de décrire cette vibration est de procéder par étapes en partant d'un modèle simple et en le complétant progressivement. Il est classique d'effectuer cette démarche en trois étapes correspondant aux modèles suivants :

- Une liaison chimique ressemble à un ressort qui obéit à la loi de HOOKE. C'est le modèle de l'oscillateur harmonique continu.
- Les états énergétiques sont en fait quantifiés, c'est-à-dire que l'énergie est absorbée par la molécule par sauts discontinus. C'est le modèle de l'oscillateur harmonique quantifié.
- Enfin, les molécules réelles n'obéissent pas exactement à la loi de HOOKE, ce qui conduit à étudier un modèle anharmonique quantifié.

2.1.1 L'oscillateur harmonique continu

Ce premier modèle fait appel à la mécanique classique. Pour modéliser une vibration moléculaire, considérons deux masses m_1 et m_2 attachées par un ressort. (Figure 6).

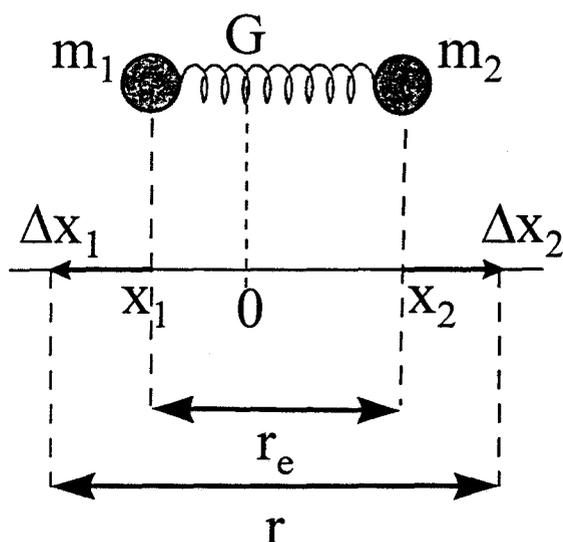


Figure 6 : Le modèle de l'oscillateur mécanique

Soit r la distance entre les deux masses, r_e cette même longueur à l'équilibre (pas de compression ni d'étirement du ressort), Δx_1 et Δx_2 les déplacements des masses m_1 et m_2 par rapport aux positions d'équilibre x_1 et x_2 . Introduisons tout d'abord Q la variable d'élongation du ressort telle que :

$$(éq. 1) \quad Q = r - r_e = \Delta x_1 - \Delta x_2$$

L'énergie cinétique du système en mouvement est donnée par :

$$(éq. 2) \quad T = \frac{1}{2} \sum_{i=1}^2 m_i v_i^2 \quad \text{avec} \quad v_i = \frac{dx_i}{dt} = \dot{x}_i$$

Par introduction du point G, centre de gravité du système, nous pouvons noter :

$$(éq. 3) \quad m_1 x_1 = m_2 x_2$$

$$\text{donc} \quad \frac{x_1}{m_2} = \frac{x_2}{m_1} = \frac{x_1 + x_2}{m_1 + m_2} = \frac{r_e}{m_1 + m_2}$$

$$\text{soit} \quad \frac{\Delta x_1}{m_2} = \frac{\Delta x_2}{m_1} = \frac{Q}{m_1 + m_2}$$

$$\text{d'où} \quad \Delta x_1 = \frac{m_2}{m_1 + m_2} Q \quad \text{et} \quad \Delta x_2 = \frac{m_1}{m_1 + m_2} Q$$

$$\text{donc} \quad \Delta \dot{x}_1 = \frac{m_2}{m_1 + m_2} \dot{Q} \quad \text{et} \quad \Delta \dot{x}_2 = \frac{m_1}{m_1 + m_2} \dot{Q}$$

Une substitution dans (éq. 2) nous donne finalement :

$$(éq. 4) \quad T = \frac{1}{2} \mu \dot{Q}^2 \quad \text{avec} \quad \mu = \frac{m_1 m_2}{m_1 + m_2}$$

La constante μ est appelée masse réduite du système. Après avoir déterminé l'énergie cinétique T par rapport à la coordonnée Q , il faut caractériser l'énergie potentielle du système notée V . Elle est définie par le travail des forces exercées sur le système. Pour notre modèle, la force exercée par le ressort est donnée par :

$$(éq. 5) \quad F = -kQ \quad \text{où } k \text{ est la constante de force de la liaison}$$

Ainsi
$$(éq. 6) \quad V = \int -F dQ = \frac{1}{2} k Q^2$$

Tout système en mouvement peut être décrit par l'équation de LAGRANGE, généralisation des équations de NEWTON ($\sum F = m\gamma$) :

$$(éq. 7) \quad \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{Q}} \right) + \frac{\partial V}{\partial Q} = 0$$

La substitution de V et T dans la dernière équation donne :

$$(éq. 8) \quad \mu \ddot{Q} + kQ = 0$$

On montre aisément que Q est de la forme :

$$(éq. 9) \quad Q = a \cos(\omega t + \phi)$$

Le système est donc en vibration avec une pulsation $\omega = \sqrt{\frac{k}{\mu}}$ ou une fréquence de ν

$= \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}$ exprimée en Hertz. Pour des raisons de commodité, on introduit souvent la

fréquence en nombre d'onde notée $\bar{\nu}$ telle que $\bar{\nu} = \frac{\nu}{c}$. Ainsi nous avons $\bar{\nu} = \frac{1}{2\pi c} \sqrt{\frac{k}{\mu}}$

exprimée en cm^{-1} si c est homogène à des cm.s^{-1} . Si on écarte donc les deux masses de leur position d'origine d'une distance arbitraire $Q = Q_{\text{max}}$, elles vibreront avec une énergie totale égale à $\frac{1}{2} k.Q_{\text{max}}^2$. Ce modèle prédit donc que l'oscillateur peut vibrer à n'importe quelle énergie.

2.1.2 L'oscillateur harmonique quantifié

Ce deuxième modèle fait appel à la théorie quantique. Au début du siècle, de nombreuses expériences ont permis de contredire ce premier modèle basé sur le postulat d'un continuum en énergie pour la matière. C'est par l'étude du rayonnement thermique émis par un corps porté à haute température que s'est introduite historiquement la notion de quantification de l'énergie⁽²⁵⁾.

En 1900, Max PLANCK proposa l'hypothèse selon laquelle l'énergie d'un système oscillant est discontinue. Tout changement d'énergie du système se caractérise donc par une transition entre deux niveaux discrets. Il en résulte une émission ou une absorption de quantités discrètes d'énergie, les quanta. Considérons une transition entre deux niveaux d'énergie E1 et E2. Max PLANCK a proposé d'associer à cette transition énergétique une onde électromagnétique de fréquence ν telle que :

$$(éq. 10) \quad \Delta E = E2 - E1 = h\nu \quad \text{où } h \text{ est la constante de PLANCK.}$$

Donc si on éclaire une molécule sur un niveau d'énergie E1 avec une lumière

polychromatique, un niveau E2 peut être atteint. Le spectre d'absorption de cette molécule donnera une raie unique de fréquence $\nu = \Delta E / h$.

Une molécule dans l'espace possède de nombreuses énergies (rotation, vibration...). Dans l'infrarouge, seuls les changements énergétiques liés à des états vibratoires ou rotatoires sont détectés. L'énergie de rotation est cependant beaucoup plus faible que l'énergie de vibration. De plus, les bandes de rotation ne sont visibles que pour les gaz. Pour les solides et les liquides la spectroscopie infrarouge ne fait intervenir que les modes de vibration.

Les énergies de vibration sont quantifiées comme toutes les énergies moléculaires. La résolution de l'équation de SCHRODINGER indépendante du temps utilisant l'hamiltonien vibrationnel d'une molécule diatomique (oscillateur harmonique) permet d'obtenir les niveaux d'énergie vibrationnelle E_v tels que :

$$\text{(éq. 11)} \quad E_v = (v + \frac{1}{2})h\nu \text{ avec } v = 0, 1, 2 \dots$$

où v le nombre quantique vibrationnel caractérise le niveau d'énergie du système et ν la fréquence du vibreur considéré. De plus, des restrictions sur les transitions entre niveaux d'énergie sont apportées par la règle de sélection :

$$\text{(éq. 12)} \quad \Delta v = \pm 1$$

Pour faire passer une molécule de son niveau fondamental ($v=0$) à son premier état excité ($v=1$) il faut apporter une énergie égale à $h\nu$. La fréquence de vibration de l'oscillateur est trop importante pour être mesurée ($\cong 10^{14}$ hertz) mais peut néanmoins être estimée indirectement. En effet, si on envoie sur l'oscillateur une onde électromagnétique de fréquence ν et qu'elle se trouve absorbée, on pourra dire que l'énergie apportée au système permet d'effectuer une transition énergétique. De cette façon, nous pourrions confondre par analogie la fréquence ν de l'onde électromagnétique et la fréquence du vibreur. La règle de sélection n'autorisant que des transitions entre deux niveaux successifs (n et $n+1$), on montre qu'il n'existe qu'une seule fréquence d'absorption pour toutes les transitions énergétiques (Figure 7).

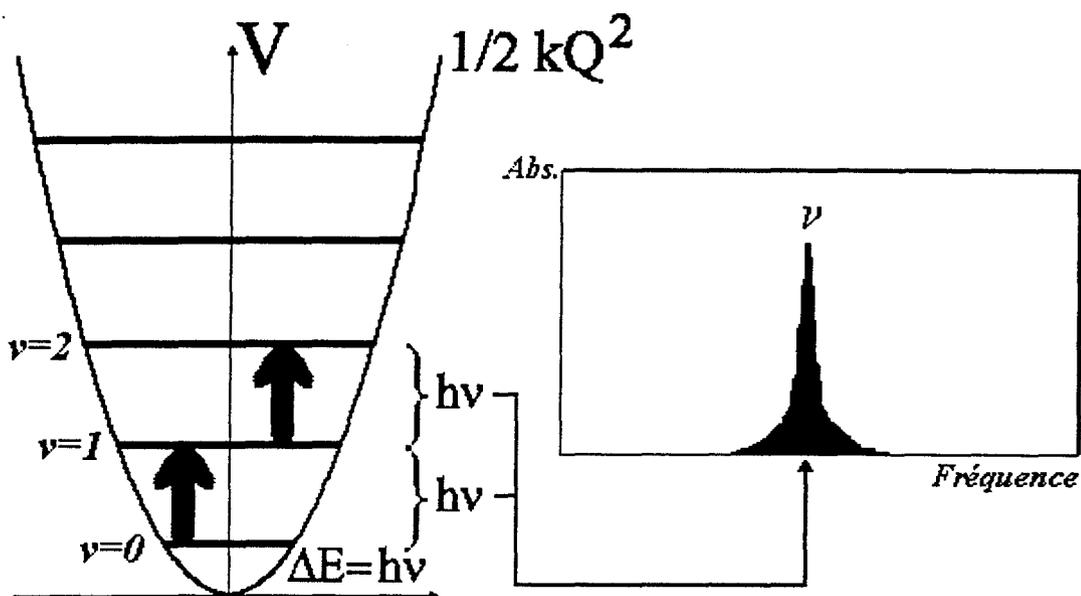


Figure 7 : Observations spectrales des transitions pour l'oscillateur harmonique.

Bien que ce traitement quantique de l'oscillateur harmonique permette d'interpréter l'observation des bandes fondamentales de vibration dans l'infrarouge moyen^(26,27), il n'explique pas la présence des bandes harmoniques dans le proche infrarouge (NIR) d'où l'introduction du modèle anharmonique quantifié.

2.1.3 L'oscillateur anharmonique quantifié

Les bandes du proche infrarouge observées à des fréquences voisines de 2ν , 3ν ... sont issues des transitions entre niveaux d'énergie où $\Delta\nu = \pm 2, \pm 3 \dots$. Ces transitions sont interdites par la règle de sélection du modèle précédent. Cette anomalie provient du fait que les molécules réelles n'obéissent pas exactement à la loi de l'oscillateur harmonique (Loi de HOOKE). En effet, lorsque l'on approche deux atomes l'un de l'autre, l'augmentation des forces de répulsion coulombiennes sont de plus en plus importantes. L'énergie potentielle croit alors plus vite que le prédit le modèle harmonique (Figure 8). De plus lorsque l'on écarte les atomes d'une distance suffisamment grande, le potentiel atteint une valeur seuil appelée énergie de dissociation E_d . Cet état de dissociation n'est pas géré par le modèle harmonique. MORSE a proposé une fonction empirique pour décrire ce nouveau profil d'énergie potentielle :

$$\text{(éq. 13) } V = E_d(1 - e^{-\alpha Q})^2$$

où α est une constante caractéristique de la molécule. Si on considère un développement de Taylor autour de la position d'équilibre pour l'énergie potentielle V nous avons :

$$(éq. 14) \quad V = V_e + \left(\frac{\partial V}{\partial Q}\right)_e \cdot Q + \frac{1}{2} \left(\frac{\partial^2 V}{\partial Q^2}\right)_e \cdot Q^2 + \frac{1}{6} \left(\frac{\partial^3 V}{\partial Q^3}\right)_e \cdot Q^3 + \dots$$

Dans cette expression, le potentiel absolu à la position d'équilibre ne peut pas être évalué. Par principe, on fixe ce potentiel à zéro. De plus, nous sommes sur un minimum d'énergie potentielle à l'équilibre donc $\left(\frac{\partial V}{\partial Q}\right)_e$ est nulle. La partie en Q^2 est appelée terme harmonique du potentiel. Dans l'hypothèse harmonique, c'est le seul terme que l'on conserve pour le développement de V . On retrouve ainsi $V = \frac{1}{2} \left(\frac{\partial^2 V}{\partial Q^2}\right)_e \cdot Q^2 = \frac{1}{2} k \cdot Q^2$. L'oscillateur anharmonique prend donc en compte le terme en Q^3 pour corriger la faiblesse du modèle harmonique soit :

$$(éq. 15) \quad V = \frac{1}{2} \left(\frac{\partial^2 V}{\partial Q^2}\right)_e \cdot Q^2 + \frac{1}{6} \left(\frac{\partial^3 V}{\partial Q^3}\right)_e \cdot Q^3$$

L'utilisation de cette nouvelle expression de V pour la résolution de l'équation de SCHRODINGER indépendante du temps permet d'obtenir les niveaux d'énergie vibrationnelle E_v tels que :

$$(éq. 16) \quad E_v = (v + \frac{1}{2})h\nu - (v + \frac{1}{2})^2 h\nu\chi$$

où χ est appelée constante d'anharmonicité. Comme pour le modèle harmonique, ce développement mathématique impose une règle de sélection :

$$(éq. 17) \quad \Delta v = \pm 1, \pm 2 \pm 3 \dots$$

La variation d'énergie ΔE associée à une transition entre un niveau v et $v + \Delta v$ est :

$$(éq. 18) \quad \Delta E = h\nu\Delta v[1 - (2v + \Delta v + 1)\chi]$$

Avec ce modèle, nous pouvons maintenant expliquer l'existence des harmoniques dans les spectres de vibration. Grâce à la statistique de MAXWELL-BOLTZMAN nous montrons que seules les trois premières transitions sont réellement observables. Cette statistique dit qu'à température ambiante, presque toutes les molécules se trouvent dans l'état fondamental $v = 0$ (99 %). Un faible nombre se trouve sur le niveau $v = 1$ (< 1%) et un nombre encore plus faible sur $v = 2$ (<< 1%).

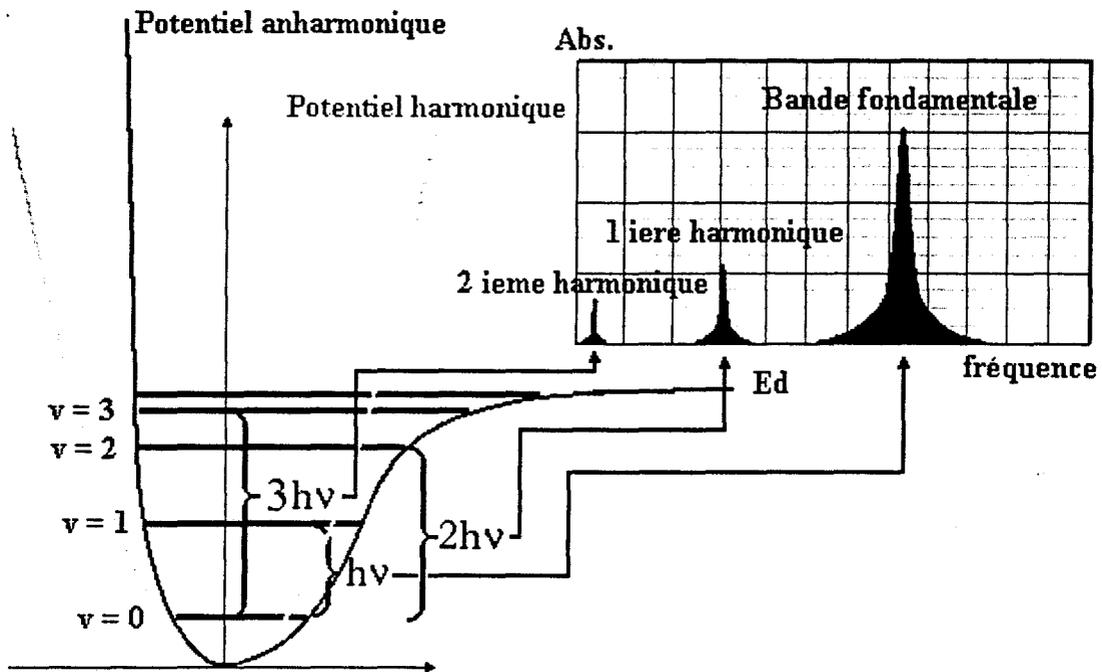


Figure 8 : Observations spectrales des transitions pour l'oscillateur anharmonique.

Comme nous pouvons le voir sur la Figure 8, trois transitions énergétiques peuvent être observées en absorption :

- De $v = 0$ à $v = 1$, $\Delta v = +1$, $\Delta E = hv(1-2\chi)$, harmonique de rang 1 communément appelée « fondamentale » visible dans l'infrarouge moyen (400 à 4000 cm^{-1}).
- De $v = 0$ à $v = 2$, $\Delta v = +2$, $\Delta E = 2 hv(1-3\chi)$, harmonique de rang 2 communément appelée « première harmonique » visible dans le proche infrarouge (900 à 2500 nm).
- De $v = 0$ à $v = 3$, $\Delta v = +3$, $\Delta E = 3 hv(1-4\chi)$, harmonique de rang 3 communément appelée « deuxième harmonique » visible dans le proche infrarouge.

Or la constante d'anharmonicité χ est comprise entre 0.01 et 0.05 . Ces bandes de vibration possèdent donc approximativement des fréquences ν , 2ν , 3ν .

Ce développement a été effectué sur la base d'un vibreur simple (molécule diatomique) ne possédant donc qu'une seule vibration. Le calcul peut être généralisé à une molécule polyatomique constituée de n atomes⁽²⁸⁾ :

3n coordonnées d'équilibre	3n coordonnées de déplacement
Atome 1.....Atome n	Atome 1.....Atome n
$x_1, y_1, z_1, \dots, x_n, y_n, z_n$	$\Delta x_1, \Delta y_1, \Delta z_1, \dots, \Delta x_n, \Delta y_n, \Delta z_n$

L'énergie cinétique du système est donnée par :

$$(éq. 19) \quad T = \frac{1}{2} \sum_{i=1}^n m_i v_i^2 = \frac{1}{2} \sum_{i=1}^n m_i (\Delta x_i^2 + \Delta y_i^2 + \Delta z_i^2) \text{ avec } m_i \text{ la masse de l'atome } i.$$

Posons $q_1 = \sqrt{m_1} \Delta x_1$, $q_2 = \sqrt{m_1} \Delta y_1$, $q_3 = \sqrt{m_1} \Delta z_1$, $q_4 = \sqrt{m_2} \Delta x_2$, ..., $q_{3n} = \sqrt{m_n} \Delta z_n$. Les q_i sont appelées coordonnées pondérées de déplacement. Nous obtenons donc :

$$(éq. 20) \quad T = \frac{1}{2} \sum_{i=1}^n \dot{q}_i^2$$

Afin de calculer l'énergie potentielle V du système, nous effectuons à nouveau un développement de TAYLOR au voisinage de l'équilibre :

$$(éq. 21) \quad V = V_e + \sum_{i=1}^{3n} \left(\frac{\partial V}{\partial q_i} \right)_e q_i + \frac{1}{2} \sum_{i=1}^{3n} \sum_{j=1}^{3n} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_e q_i q_j + \frac{1}{6} \sum_{i=1}^{3n} \sum_{j=1}^{3n} \sum_{k=1}^{3n} \left(\frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k} \right)_e q_i q_j q_k + \dots$$

Dans l'hypothèse harmonique, seul le terme quadratique est conservé :

$$(éq. 22) \quad V = \frac{1}{2} \sum_{i=1}^{3n} \sum_{j=1}^{3n} f_{i,j} q_i q_j \text{ avec } f_{i,j} = \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_e$$

L'équation de LAGRANGE peut être appliquée à chacune des $3n$ coordonnées de déplacement du système :

$$(éq. 23) \quad \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_i} \right) + \frac{\partial V}{\partial q_i} = 0 \text{ pour } i=1 \dots 3n.$$

Par substitution de T et V , on obtient facilement :

$$(éq. 24) \quad \ddot{q}_i + \sum_{j=1}^{3n} f_{i,j} q_j = 0 \text{ pour } i=1 \dots 3n.$$

Si on suppose q_i être de la forme $a \cos(\omega t + \phi)$, le système d'équations précédent peut s'écrire sous la forme matricielle suivante :

$$\begin{pmatrix} (f_{1,1} - \omega^2) & \dots & f_{1,i} & \dots & f_{1,3n} \\ \dots & \dots & \dots & \dots & \dots \\ f_{i,1} & \dots & (f_{i,i} - \omega^2) & \dots & f_{i,3n} \\ \dots & \dots & \dots & \dots & \dots \\ f_{3n,1} & \dots & f_{3n,i} & \dots & (f_{3n,3n} - \omega^2) \end{pmatrix} \cdot \begin{pmatrix} q_1 \\ \dots \\ q_i \\ \dots \\ q_{3n} \end{pmatrix} = \vec{0}$$

La résolution de ce système est donnée par le déterminant séculaire $\det(\mathbf{F} - \omega^2 \mathbf{I}) = 0$ où \mathbf{F} est la matrice des élément $f_{i,j}$ (i et $j=1 \dots 3n$) et \mathbf{I} la matrice identité⁽²⁹⁾. Les coordonnées de déplacement ne peuvent en effet être nulles. La diagonalisation de \mathbf{F} permet d'obtenir $3n-6$

valeurs propres différentes de zéro ($3n-5$ si la molécule est linéaire) notées λ_k . On en déduit les $3n-6$ fréquences fondamentales de la molécule notées ν_k car $\lambda_k = \omega^2 = 4\pi^2\nu_k^2$. Les vecteurs propres combinaisons linéaires des q_i sont appelés coordonnées normales de vibration notées Q_k . La résolution de l'équation de SCHRODINGER permet d'obtenir l'énergie de vibration du système qui est la somme des énergies de chaque mode normal⁽³⁰⁾ :

$$(éq. 25) \quad E_{\text{vib}} = \sum_{k=1}^{3n-6} \left(\nu_k + \frac{1}{2} \right) h\nu_k$$

On introduit donc $3n-6$ nombres quantiques ν_k . Si on considère l'anharmonicité, nous obtenons :

$$(éq. 26) \quad E_{\text{vib}} = \sum_{k=1}^{3n-6} \left(\nu_k + \frac{1}{2} \right) h\nu_k + \sum_{i=1}^{3n-6} \sum_{k \geq i} \left(\nu_i + \frac{1}{2} \right) \left(\nu_k + \frac{1}{2} \right) h\chi_{i,k}$$

Les $\chi_{i,k}$ sont les constantes d'anharmonicité. Comme pour la molécule diatomique, ce développement mathématique impose une règle de sélection mais qui est cette fois-ci généralisée pour chaque nombre quantique ν_k :

$$(éq. 27) \quad \Delta\nu_k = \pm 1, \pm 2 \pm 3 \dots$$

A partir de (éq. 25), on retrouve les harmoniques de chaque mode de vibration mais aussi des transitions dites de combinaison :

$$(éq. 28) \quad \Delta E_{\nu_k \rightarrow \nu_k + \Delta\nu_k} = \sum_{k=1}^{3n-6} \Delta\nu_k h\nu_k$$

Ainsi, un spectre proche infrarouge par la présence des bandes harmoniques et des combinaisons apparaît comme l'écho déformé de celui obtenu en infrarouge moyen.

Prenons l'exemple simple de la molécule d'eau. Elle présente trois modes fondamentaux de vibrations (Figure 9).

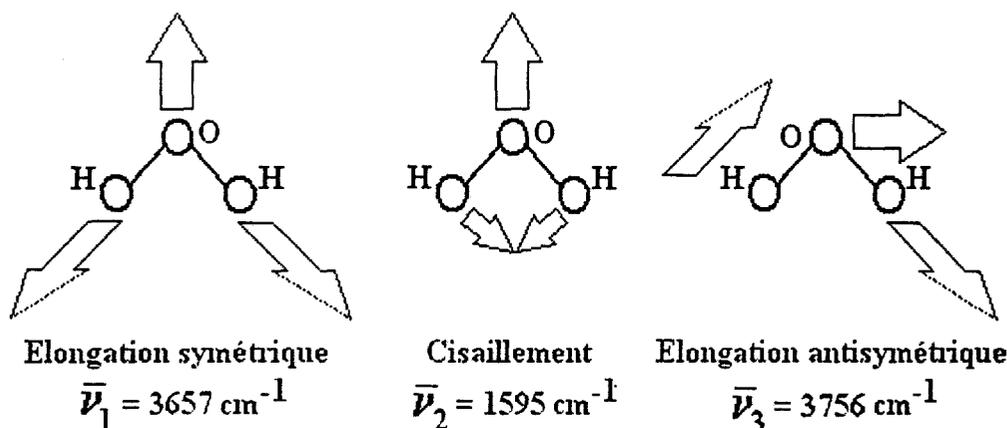


Figure 9 : Les modes fondamentaux de vibration de la molécule d'eau.

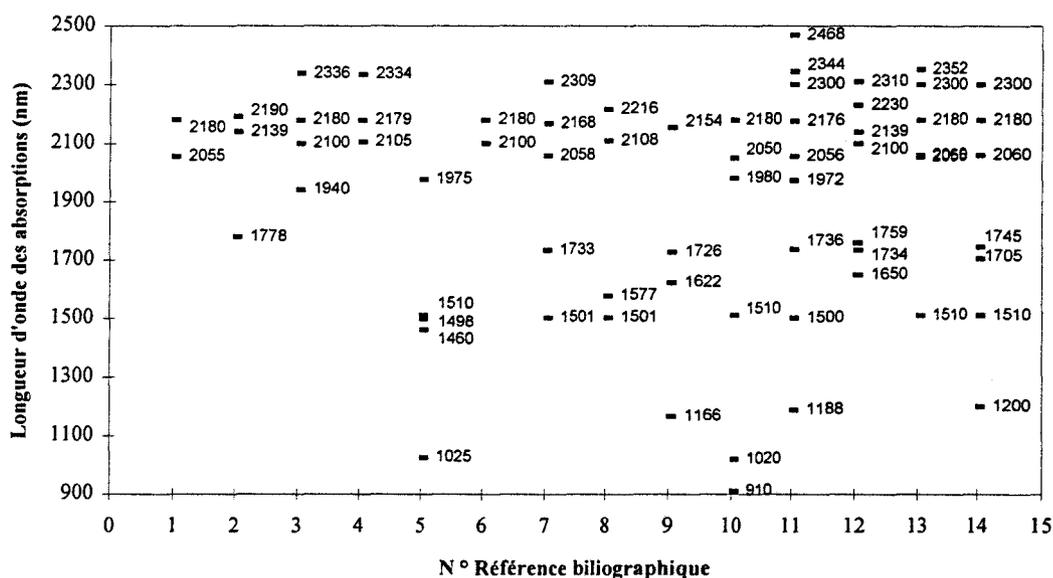
D'après IWAMOTO et al.⁽³¹⁾, une attribution fine des bandes visibles dans le NIR mesurées en longueur d'onde peut être effectuée à partir de ces modes :

- $\lambda = 960$ nm correspond à la combinaison $(2\nu_1 + \nu_2)$.
- $\lambda = 1150$ nm correspond à la combinaison $(\nu_1 + \nu_2 + \nu_3)$.
- $\lambda = 1440$ nm correspond à la combinaison $(\nu_1 + \nu_3)$.
- $\lambda = 1930$ nm correspond à la combinaison $(\nu_2 + \nu_3)$.

2.2 Attributions des bandes spectrales dans le NIR

Dans le proche infrarouge, les modes de vibrations chimiques les mieux observés correspondent aux élongations des liaisons contenant un atome léger telles que O-H, C-H, S-H, N-H... De nombreux auteurs ont tenté de recenser les attributions des bandes spectrales^(32,33,34) à partir de l'étude de nombreux produits.

Il nous est apparu important de faire ce recensement pour les bandes caractéristiques des protéines et des matières grasses sur lesquelles nous travaillerons (Tableau 1 et Tableau 2). Il est bien entendu que la liste présentée n'est pas exhaustive. Le tableau ci-dessous montre que le pointé des longueurs d'onde caractéristiques est souvent ambigu.



Références :

(1) D. L. Wetzel, *Anal. Chem.*, 55(10), 1172(a) (1983).

(2) J.J Workman, *Doctoral Dissertation*, Columbia University, University Microfilms International No. LD-00993 (1984).

- (3) J.S. Schenk, M.O. Westerhaus, M.R. Hoover, *J. Dairy Sci.*, 62(5), p810 (1979).
- (4) F.E. Barton, D. Burdick, *J. Agr. Food Chem.*, 27, p1248 (1979).
- (5) W. Kaye, *Spectrochem. Acta*, 6, p257 (1954).
- (6) J.S. Shenk, R.F. Barnes, *Proc. 34th Southern Pasture and Forage Crop Improvement Conf.*, p57-62 (1977).
- (7) J.S. Shenk, K.H. Norris, R.F. Barnes, G.W. Fissel, *XIII Internatl. Grassl. Cong.*, p1440-1441 (1977).
- (8) Burdick, F.E. Barton, B.D. Nelson, *Proc. 36th Southern Pasture and Forage Crop Improvement Conf.*, p81-86 (1979).
- (9) Coelho, F.G. Hembry, *Proc. of the Program and Research Progress Reports*, p195-199 (1982).
- (10) B.G. Osborne, T. Fearn, *Near Infrared spectroscopy of Food analysis*, Edition Longman Scientific and Technical, New York, p22 (1986).
- (11) D. Bertrand, *Thèse de Doctorat*, Université de Dijon (1988).
- (12) D.A. Burns, E.W. Ciurczak, *Handbook of Near-Infrared Analysis*, Ed Marcel Decker (1992).
- (13) J.S. Shenk, J.J. Workman, M.O. westerhaus, *Handbook of Near-Infrared Analysis*, Ed Marcel Decker (1992).
- (14) R.D. Rosendthal, *Analisis*, Neotec Instrument Inc. (1979).

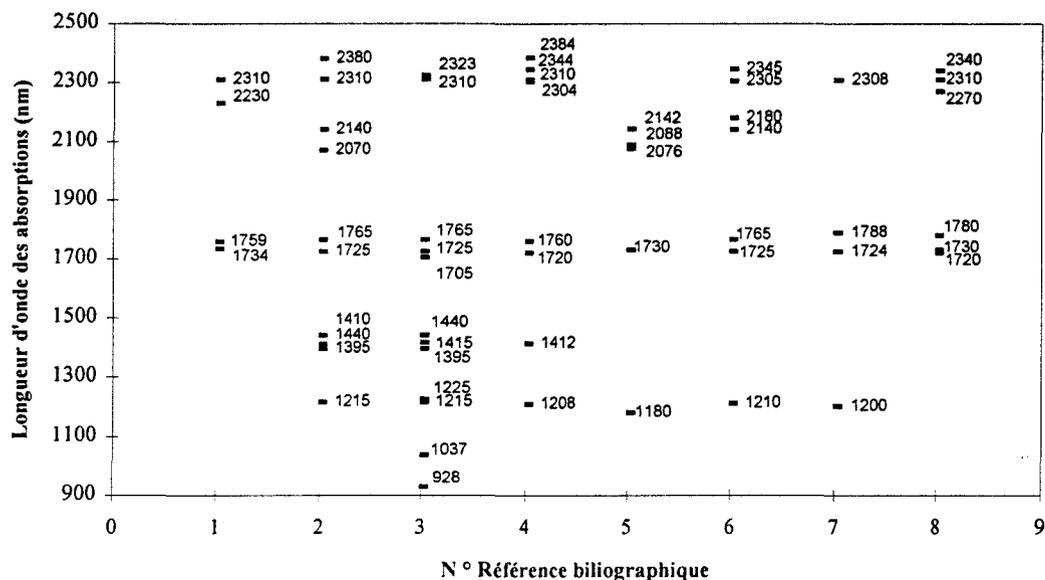
Tableau 3 : Longueurs d'onde caractéristiques des absorptions « protéine ».

Le pointé des bandes d'absorption reste difficile, car les bandes sont larges et souvent faibles en intensité. De plus, les protéines sont en présence d'autres espèces qui peuvent interférer et occasionner des déplacements en longueur d'onde. Néanmoins, les auteurs semblent d'accord sur l'attribution des bandes suivantes:

- $\lambda \cong 1020 \text{ nm} \rightarrow$ Combinaison ($2 \times \text{NH}_{\text{elongation}} + 2 \times \text{Amide I}$)
- $\lambda \cong 1510 \text{ nm} \rightarrow$ 1^{ère} Harmonique NH
- $\lambda \cong 1980 \text{ nm} \rightarrow$ Combinaison ($\text{NH}_{\text{elongation Antisymétrique}} + \text{Amide II}$)
- $\lambda \cong 2050 \text{ nm} \rightarrow$ Combinaison ($\text{NH}_{\text{elongation Symétrique}} + \text{Amide II}$)
- $\lambda \cong 2180 \text{ nm} \rightarrow$ Combinaison ($2 \times \text{Amide I} + \text{Amide III}$)

L'appellation Amide I, II, III est utilisée pour coder les vibrations du groupement amide. Ces fréquences proviennent de sa structure mésomère et des masses assez proches des atomes C, N et O. Les fréquences de vibration des liaisons C-O et C-N sont alors couplées.

De la même façon, nous pouvons faire un recensement pour les matières grasses (tableau 2).



Références :

- (1) D.A. Burns, E.W. Ciurczak, *Handbook of Near-Infrared Analysis*, Ed Maecel Decker (1992).
- (2) J.S. Shenk, J.J. Workman, M.O. Westerhaus, *Handbook of Near-Infrared Analysis*, Ed Maecel Decker, p394 (1992).
- (3) B.G. Osborne, T. Fearn, *Near Infrared spectroscopy of Food analysis*, Edition Longman Scientific and Technical, New York, p22 (1986).
- (4) D. Bertrand, *Thèse de Doctorat*, Université de Dijon (1988).
- (5) E. Diaz-Carrillo, *Journal of Near Infrared Spectroscopy*, 1, p144 (1993).
- (6) R.D. Rosenthal, *Analisis*, Neotec Instrument Inc. (1979).
- (7) G. Downey, P.Robert, D. Bertrand, P.M. Kelly, *Applied Spectroscopy*, 44(1), p150 (1990).
- (8) R.J. Baer, J. F. Franck, M. Loewenstein, *J. Assoc. Anal. Chem.*, 64(4), p858 (1983).

Tableau 4 : Longueurs d'onde caractéristiques des absorptions « matière grasse ».

Nous pouvons alors attribuer les différentes vibrations :

- $\lambda \cong 928 \text{ nm} \rightarrow 3^{\text{ème}}$ Harmonique CH
- $\lambda \cong 1037 \text{ nm} \rightarrow$ Combinaison ($2 \times \text{CH}_{\text{élongation}} + 2 \times \text{CH}_{\text{déformation}} + (\text{CH}_2)_n$)
- $\lambda \cong 1215 \text{ nm} \rightarrow 2^{\text{ème}}$ Harmonique CH
- $\lambda \cong 1415 \text{ nm} \rightarrow$ Combinaison ($2 \times \text{CH}_{\text{élongation}} + \text{CH}_{\text{déformation}}$)
- $\lambda \cong 1725 \text{ nm} \rightarrow 1^{\text{ère}}$ Harmonique CH
- $\lambda \cong 2310 \text{ nm} \rightarrow$ Combinaison ($\text{CH}_{\text{élongation}} + \text{CH}_{\text{déformation}}$)
- $\lambda \cong 2323 \text{ nm} \rightarrow$ Combinaison ($\text{CH}_{\text{élongation}} + \text{CH}_{\text{déformation}}$)
- $\lambda \cong 2380 \text{ nm} \rightarrow$ Combinaison ($\text{CH}_{\text{élongation}} + \text{CC}_{\text{déformation}}$)

Ces résultats montrent un grand nombre de domaines spectraux exploitables pour l'analyse quantitative des matières grasses et des protéines.

2.3 L'instrumentation proche infrarouge.

2.3.1 Développement chronologique de l'instrumentation.

L'histoire de la spectroscopie proche infrarouge (NIR) commence au début des années 1800 avec l'astronome W. HERSCHEL^(35,36,37). Il décomposa la lumière du soleil avec un prisme puis déplaça un thermomètre dans les différentes couleurs. La température resta stable sur tout le spectre et augmenta au moment où il dépassa la lumière rouge du spectre. Une énergie était donc présente au-dessous du rouge et comme il ne pouvait pas la voir il l'appela « Infrarouge ». Ce phénomène resta très longtemps une curiosité, car il était difficile de faire des mesures dans ce domaine. Il était plus aisé de travailler dans le visible. L'invention de la plaque photographique a même permis d'étendre l'étude aux ultraviolets (U.V.). Il y avait tellement de choses à apprendre dans ces deux domaines que l'infrarouge resta longtemps inexploité.

C'est au début du XX^{ème} siècle que W.W. COBLENTZ construit le premier spectromètre infrarouge (1 à 15 microns). Les mesures se font longueur d'onde après longueur d'onde, l'étude d'un échantillon pouvant prendre plus d'une journée. Après l'étude de nombreux produits, il découvre que les bandes d'absorption sont caractéristiques de groupements chimiques. L'étude structurale est en marche. Mais pendant de nombreuses années, les chimistes préférèrent le moyen infrarouge (MIR). Il considère le proche infrarouge (NIR) comme une espèce de « No-man's land » où il n'existe pas de bandes fines. De plus, les problèmes de ligne de base sont fréquents et le recouvrement des bandes est important. Il y a aussi une perte dramatique de la sensibilité par rapport au MIR (2 à 3 ordres de grandeur). Les théoriciens ont très vite déclaré que l'attribution des bandes était difficile par la présence de nombreuses combinaisons et harmoniques. Toutes ces raisons étaient suffisantes pour freiner le développement de la spectroscopie NIR.

En effet, si on regarde le nombre de publications sur le NIR⁽³⁸⁾ l'explosion tardive ne se situe qu'au début des années 80 (Tableau 5). Les premières références bibliographiques en NIR sont celles de KUBELKA *et al.*⁽³⁹⁾. Elles présentent la technique de réflexion diffuse.

Mais jusqu'en 1950, les chercheurs ignorent cette technique au profit de la transmission^(40,41,42). C'est à partir de 1960 que l'on reconnaît le potentiel de la spectroscopie NIR pour l'analyse quantitative. Karl NORRIS, considéré comme le père de la spectroscopie NIR, développe alors de nombreuses analyses quantitatives en agro-alimentaire^(43,44).

Période	Nombre de publications
1930-40	3
1940-50	4
1950-60	23
1960-70	23
1970-80	152
1980-90	>1000

Tableau 5: Évolution du nombre de publications dans le NIR.

Au fil du temps, les appareils vont se perfectionner. Les traitements mathématiques se développeront grâce à l'explosion de la micro-informatique. La chaîne de détection sera de moins en moins bruitée pour atteindre aujourd'hui un rapport signal sur bruit de plusieurs milliers. L'acquisition d'un spectre sera de plus en plus rapide. Les spectromètres modernes donnent un spectre en moins d'une minute. Cette présentation succincte peut être complétée par de nombreuses publications^(45,46).

2.3.2 Histoire du proche infrarouge et des produits laitiers.

Les premières mesures NIR sur les constituants du lait et des produits laitiers ont été effectuées par GOULDEN en 1957⁽⁴⁷⁾. Cependant, on peut considérer que BEN GERA et NORRIS furent les premiers à développer des applications NIR sur le lait comparables à celles d'aujourd'hui⁽⁴⁴⁾. En 1978, CASADO *et al*^(48,49) ont appliqué la spectroscopie NIR à l'analyse des constituants majoritaires dans les laits en poudre entiers et écrémés. Cet axe de recherche a ensuite été développé par BAER *et al*⁽⁵⁰⁾ pour l'analyse de la composition du petit lait en poudre, et par VILDER *et al*⁽⁵¹⁾ pour le lait en poudre. De très bons résultats ont aussi été obtenus par FRANKHUIZEN *et al*^(52,53) pour les dosages de l'humidité, des protéines, du lactose, des matières grasses et des cendres sur des laits en poudre écrémés provenant de plusieurs pays de la communauté européenne. Un peu plus tard, WOOLARD *et al*⁽⁵⁴⁾ ont pu appliquer ces derniers dosages à d'autres produits tels que le lait entier en poudre, le babeurre,

les caséines. Ces résultats ont d'ailleurs induit l'adoption de la spectroscopie proche infrarouge comme méthode officielle de dosage des matières grasses, de l'humidité, des protéines, du lactose, de l'acidité et des cendres sur ces derniers produits en Nouvelle Zélande. Dans cette même période, TOOHEY *et al*⁽⁵⁵⁾ ont développé l'analyse quantitative des matières grasses et de l'humidité sur le process de vaporisation du lait. Inspirés par ces bons résultats, DOWNEY *et al*⁽⁵⁶⁾ ont présenté une classification de laits en poudre sur la base de la sévérité du traitement thermique à partir de la spectroscopie NIR et de l'analyse factorielle discriminante. De la même manière, GIANGIACOMO *et al*⁽⁵⁷⁾ ont conçu une méthode de détection des ajouts frauduleux dans les laits en poudre.

En 1977, ABOSHAMAA *et al*^(58,59) sont les premiers à analyser du fromage par la spectroscopie NIR. D'autres chercheurs améliorent ensuite l'analyse^(60, 61). Dans cette même période, on voit les premières tentatives d'utilisation des fibres optiques en spectroscopie proche infrarouge pour l'analyse des fromages de type « bleu ». Malheureusement, le faible pourcentage de transmission des fibres NIR développées dans cette période n'a pas permis d'analyser les différents constituants de ces fromages. En 1988, WEHLING *et al*⁽⁶²⁾ déterminent le taux d'humidité du fromage de Cheddar.

Ces mêmes domaines spectraux sont utilisés pour la détermination des concentrations des constituants majoritaires du lait liquide^(63,64,65). Ces études montrent l'effet négatif de l'absorption de l'eau qui est beaucoup plus importante que celles des constituants à doser. Elles témoignent aussi de l'effet parasite de la diffusion des globules gras sur le spectre NIR. Une équipe japonaise a aussi développé, à partir du proche infrarouge, le dosage du beurre⁽⁶⁶⁾ ainsi que la détection de crèmes liquides frelatées⁽⁶⁷⁾. Régulièrement, de nouvelles applications du proche infrarouge sur les produits laitiers apparaissent.

2.3.3 L'instrumentation NIR.

Nous pouvons classer les instruments d'analyse proche infrarouge en fonction de leurs configurations optiques et des techniques d'échantillonnage qu'ils utilisent. Ces caractéristiques seront développées dans les parties suivantes.

2.3.3.1 Les analyseurs spectrométriques

Cinq catégories regroupent les instruments à base de filtres interférentiels, de réseaux de diffraction, de filtres opto-acoustiques, d'interféromètres ou de diodes émettrices. Cette dernière famille ne sera pas étudiée, car son utilisation dans le proche infrarouge est encore trop marginale⁽⁶⁸⁾.

2.3.3.1.1 Les filtres interférentiels

Une bonne analyse spectrale n'est possible que si le système de sélection des longueurs d'onde se fait sur des bandes passantes faibles. On assure ainsi la sensibilité de la technique. Les filtres de « Fabry-Perot », appelés communément filtres interférentiels sont facilement fabriqués pour des bandes passantes inférieures à 10 nm. L'étendue de faisceau les avantage énormément, car l'énergie lumineuse qui frappe la surface d'un filtre est beaucoup plus importante que la surface d'une fente de monochromateur par exemple. Un filtre interférentiel est constitué d'un film transparent (souvent du MgF_2) pris en sandwich entre deux plaques de verre semi-transparentes elles-mêmes recouvertes par des particules métalliques (Figure 10).

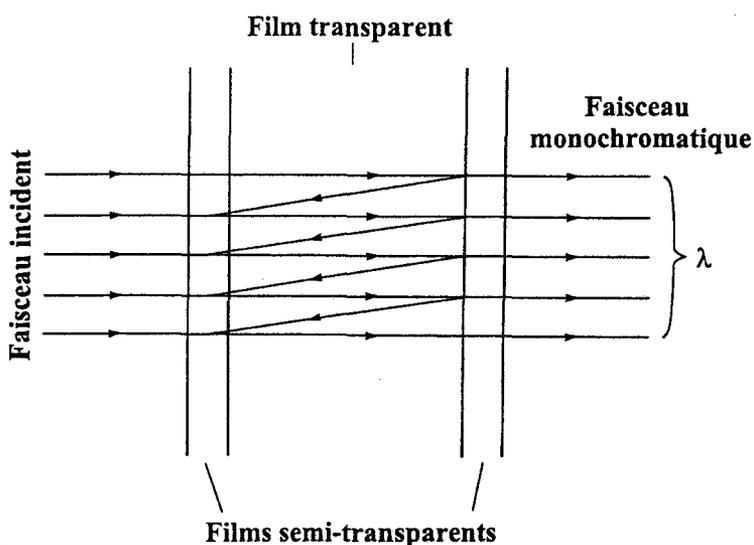


Figure 10 : Le filtre interférentiel

On montre que la radiation λ transmise par un filtre est fonction de l'épaisseur de diélectrique notée l , de son indice de réfraction n , de l'ordre de l'interférence α et de l'angle d'incidence θ du faisceau qui le frappe (par rapport à la normale).

$$(éq. 29) \quad \lambda = (2.l.n.\cos \theta)/\alpha$$

Cette sélection est issue d'un phénomène de réflexions multiples dans le diélectrique (interférences constructives pour λ et destructives pour les autres longueurs d'onde). Pour l'analyse quantitative, les filtres sont sélectionnés pour des longueurs d'onde caractéristiques des constituants à doser. Dans une configuration classique (la plus ancienne), les filtres sont placés sur la circonférence d'un disque tournant (tourelle). Le faisceau incident est perpendiculaire au filtre. L'acquisition est séquentielle (Figure 11).

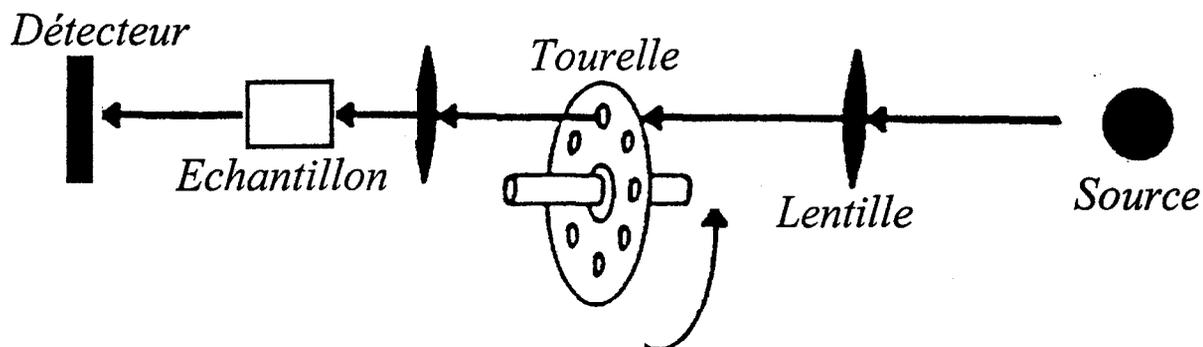


Figure 11 : Montage classique de filtres interférentiels.

Une autre configuration dite des filtres oscillants utilise la dépendance de λ par rapport à l'angle d'incidence θ du faisceau incident. Dans la figure ci-dessous deux systèmes sont présentés. De cette façon, il est possible de couvrir plusieurs domaines spectraux avec quelques filtres⁽⁶⁹⁾. Les systèmes à filtres sont certainement les moins coûteux. Néanmoins, cette configuration est peu flexible car l'étude de nouvelles espèces chimiques à de nouvelles longueurs d'onde demande l'ajout ou le remplacement de filtres sur la tourelle.

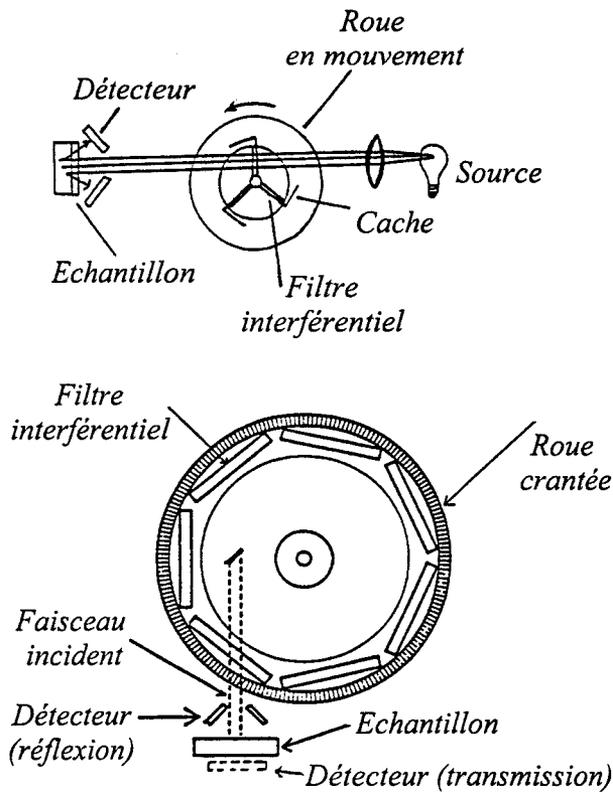


Figure 12 : Systèmes de filtres oscillants.

2.3.3.1.2 Les filtres opto-acoustiques

Les filtres opto-acoustiques (AOTF) constituent l'un des développements les plus importants de la spectroscopie NIR des années 90. La figure ci-dessous représente l'un de ces filtres.

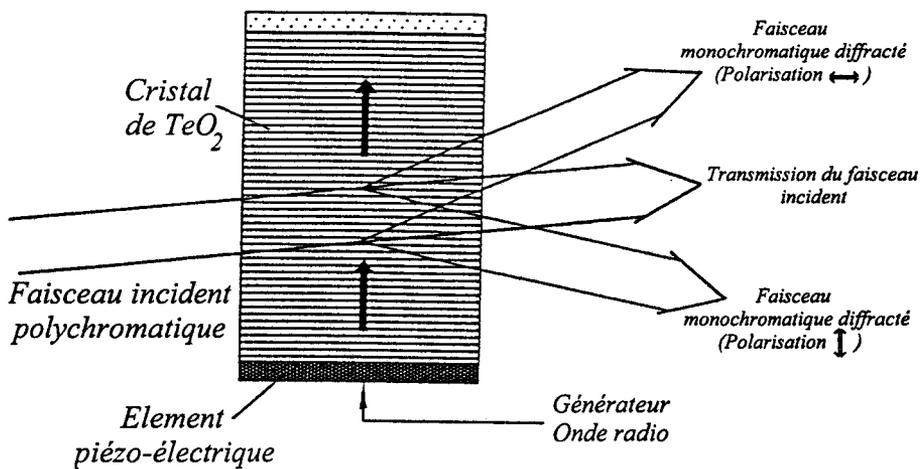


Figure 13 : Filtre opto-acoustique

La lumière polychromatique est dirigée vers un cristal d'oxyde de tellure (TeO_2). Un matériau piézo-électrique (LiNbO_3) est placé sur le côté du cristal. Sous l'effet d'un générateur de fréquence radio, ce matériau transmet une onde acoustique de haute fréquence (30 à 200 MHz) qui se propage dans le cristal (flèche noire) et change son indice de réfraction. Le cristal est usiné de telle sorte qu'il devienne une cavité de résonance acoustique. En interaction avec le cristal résonant, le rayon incident se divise en trois parties. Au centre, nous avons un faisceau polychromatique identique au rayon incident (pas d'absorptions de TeO_2 dans le NIR). Sur les côtés, il y a deux faisceaux monochromatiques identiques en longueur d'onde mais à polarisation verticale et horizontale. Cette longueur d'onde est reliée à la vitesse et à la fréquence de l'onde acoustique, à la biréfringence du cristal ainsi qu'à sa forme. Le générateur d'ondes radio permet donc de sélectionner la longueur d'onde recherchée. Ce système a de nombreux avantages. Il n'y a pas de perte de temps entre la sélection de deux longueurs d'onde puisque le changement de fréquence est d'origine électronique (opposition aux systèmes mécaniques). La bande passante peut être de 1 nm et la transmission de 98 % pour les deux faisceaux polarisés. Ce filtre peut être considéré comme un monochromateur sans parties mobiles donc très robuste. L'acquisition d'un spectre peut être très rapide. L'inconvénient majeur de ces filtres est leur faible ouverture numérique.

2.3.3.1.3 L'interférométrie

Nous pouvons considérer que la transformée de Fourier appliquée au proche infrarouge date des années 80⁽⁷⁰⁾. Les premiers travaux n'ont pas porté sur l'interférométrie dans l'instrumentation NIR mais sur l'utilisation de la transformée de Fourier de spectres provenant d'appareils dispersifs. Ils ont montré l'efficacité de la condensation des données, du filtrage numérique, de la correction de diffusions parasites (taille de particules différentes), de la déconvolution^(71,72,73). C'est donc un peu plus tard que l'on a découvert le potentiel des appareils à transformée de Fourier. A l'heure actuelle, ils donnent le meilleur rapport signal sur bruit et la meilleure répétabilité spectrale en analyse proche infrarouge. Cette supériorité peut s'expliquer par les trois avantages de FELLGETT, JACQUINOT et CONNES. Le premier est connu sous le nom d'avantage de FELLGETT ou avantage multiplex⁽⁷⁴⁾. En spectrométrie classique, si on souhaite analyser l'absorbance d'un échantillon sur n longueurs d'onde sur une durée totale t, chaque élément spectral est observé pendant un temps t/n. En interférométrie par contre, toutes les longueurs d'onde sont observées simultanément. De ce

fait, chaque élément spectral est observé pendant un temps t . Le signal et le bruit seront respectivement proportionnels à t et \sqrt{t} . Le gain de rapport signal sur bruit entre un appareil à transformée de Fourier et un appareil dispersif sera de \sqrt{n} . Ce gain est par exemple de 17 pour une analyse spectrale entre 1000 et 2500 nm pour une résolution de 5 nm. L'avantage de JACQUINOT est attribué à l'importance de flux entrant dans l'interféromètre. Pour les monochromateurs par exemple, une grande partie de l'énergie lumineuse est perdue à cause des fentes. L'avantage de CONNES est attribué à l'utilisation d'un laser dans l'interféromètre. Il permet de repérer les longueurs d'onde avec une très grande précision et une excellente reproductibilité. La figure ci-après donne le schéma de principe de l'interféromètre de MICHELSON⁽⁷⁵⁾. La source éclaire une lame séparatrice semi-transparente. Elle divise le faisceau en deux parties égales. L'un des deux faisceaux se propage vers le miroir fixe et l'autre vers le miroir mobile qui se déplace linéairement. Après réflexion sur les deux miroirs, les deux composantes se recombinaient sur la séparatrice en créant des interférences car elles n'ont pas parcouru les mêmes trajets optiques. Le détecteur enregistre donc l'interférogramme dont l'intensité est fonction du déplacement du miroir mobile. La transformée de Fourier de celui-ci donne finalement le spectre en fréquence.

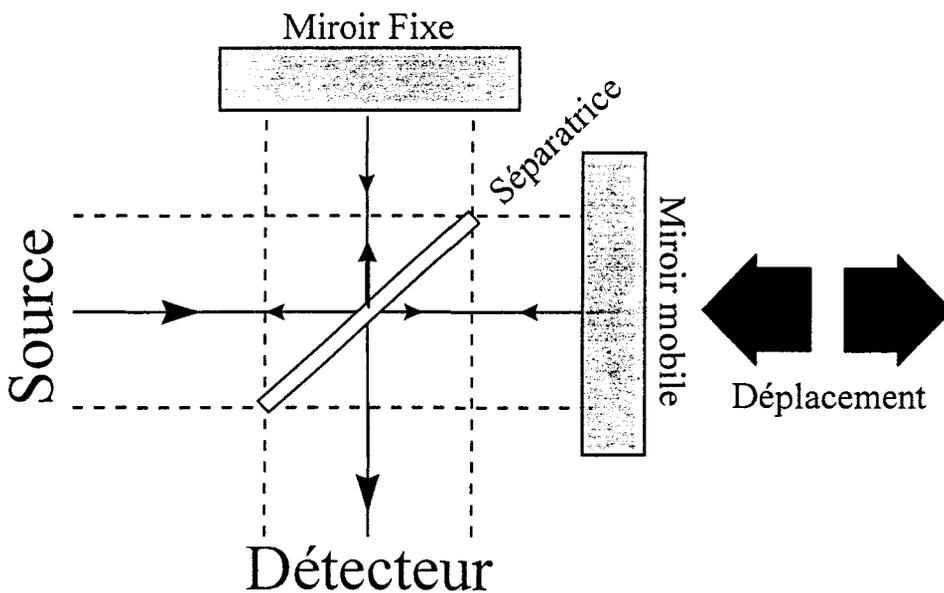


Figure 14 : Interféromètre de MICHELSON.

2.3.3.1.4 Les monochromateurs

Les systèmes optiques constitués de réseaux de diffraction sont appelés monochromateurs. Ils contiennent toujours une fente d'entrée, une fente de sortie, un filtre d'ordre ainsi que des miroirs de collimation et de focalisation.

Le réseau de diffraction est la pièce maîtresse d'un monochromateur car il conditionne en grande partie la fonction d'appareil^(76,78). Les réseaux sont dits « gravés » ou « holographiques ». Dans le premier cas, une pointe diamant vient graver de nombreux traits parallèles sur une plaque de verre (600 traits par millimètre en moyenne). En holographie, on utilise les franges d'interférences générées par deux lasers que l'on envoie à la surface du réseau recouvert d'un matériau photosensible. L'efficacité d'un réseau de diffraction est le rapport entre l'énergie diffractée et l'énergie initiale qui le frappe. Elle est souvent exprimée en pour-cent et dépend d'un grand nombre de facteurs tels que le type de réseau (gravé ou holographique), le nombre de traits ou encore la longueur d'onde de « blaze ». La Figure 15 représente les courbes d'efficacité d'un réseau gravé et d'un réseau holographique.

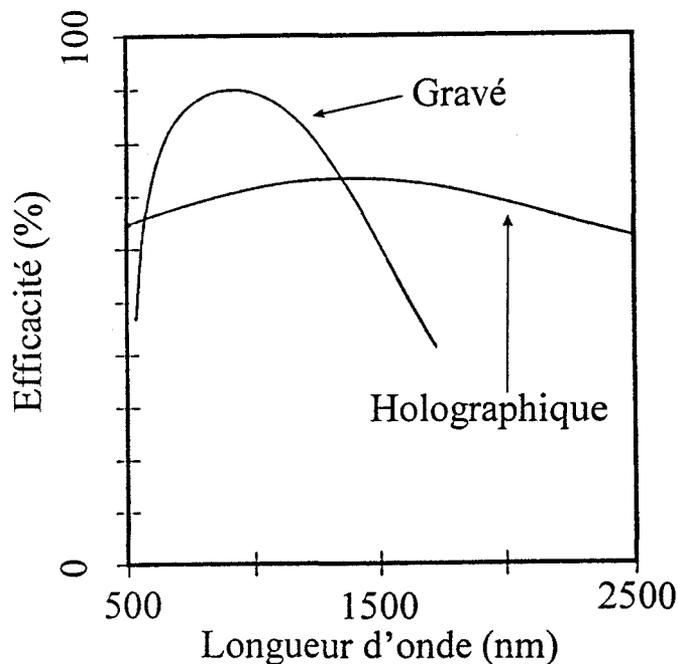


Figure 15 : Efficacité des réseaux holographiques et gravés.

Le réseau gravé possède une efficacité plus importante sur un faible domaine de longueurs d'onde. Elle chute néanmoins rapidement pour des longueurs d'onde faibles. Le réseau

holographique a une efficacité plus faible qu'il conserve sur un domaine spectral large. Le rapport signal sur bruit d'un réseau gravé est important mais celui d'un réseau holographique l'est encore plus malgré un signal plus faible. En fait, le bruit occasionné en partie par le phénomène de « ghost » (gravure périodique de traits parasites) sur les réseaux gravés, n'existe pas sur les réseaux holographiques (Figure 16).

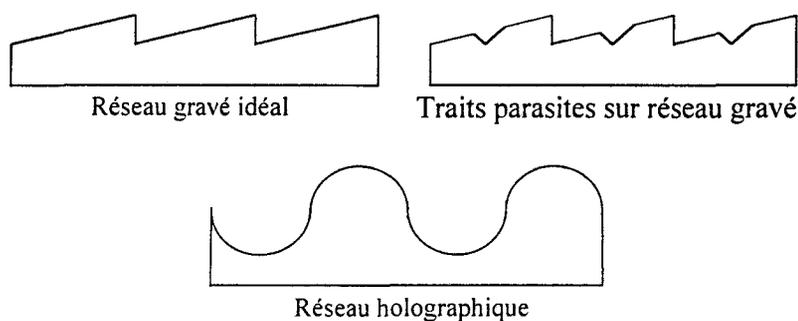


Figure 16 : Profils des réseaux de diffraction.

Le nombre de traits d'un réseau est un facteur très important. Un réseau de 1200 traits/mm utilisera par exemple des fentes deux fois plus larges qu'un réseau de 600 traits/mm pour atteindre la même bande passante. Le gain en signal sera donc très important pour les densités de traits élevées. Le domaine spectral balayé sera néanmoins plus restreint pour une mécanique de monochromateur donnée (dispersion plus faible). La longueur d'onde de « blaze » est le dernier paramètre sur lequel nous pouvons agir afin d'optimiser le signal sortant du monochromateur. Le profil d'un réseau peut être optimisé afin d'obtenir une efficacité maximale à une longueur d'onde donnée appelée longueur d'onde de blaze (λ_B). Une loi empirique donne alors une efficacité de 50 % aux longueurs d'onde $2\lambda_B/3$ et $2\lambda_B$. La Figure 17 présente les courbes d'efficacité de deux réseaux gravés blazés à 800 et 1250nm. Nous observons qu'un réseau blazé à une longueur d'onde importante ne permet pas d'étudier un domaine spectral aux longueurs d'onde faibles (coupure rapide du signal). Réciproquement, une longueur d'onde de blaze faible donne une efficacité limitée aux longueurs d'onde élevées. Toutes ces constatations montrent que le choix du réseau est délicat.

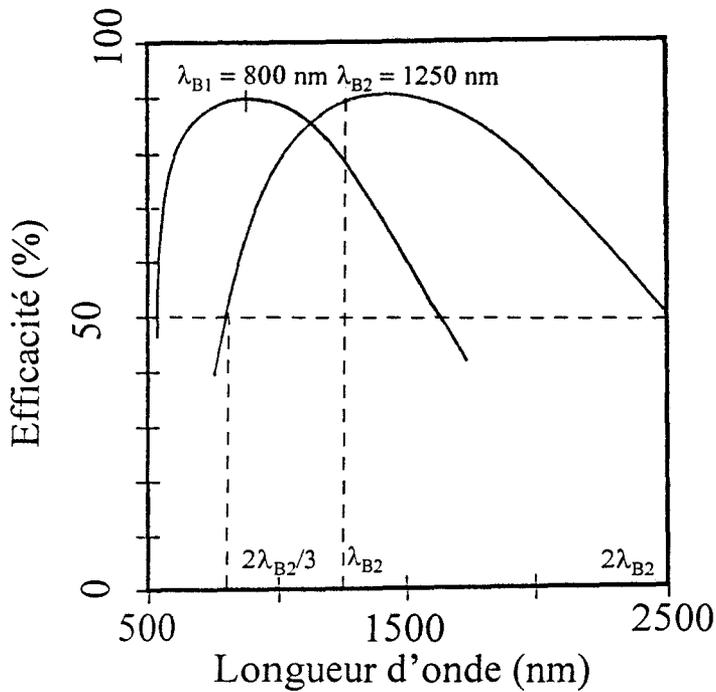


Figure 17 : Influence du blaze sur la courbe d'efficacité.

Lorsqu'un faisceau polychromatique frappe un réseau de diffraction, il est décomposé en longueurs d'onde discrètes par un jeu d'interférences constructives et destructives. La figure ci-dessous représente un système simple constitué d'une fente d'entrée, d'un réseau et d'une fente de sortie (Figure 18).

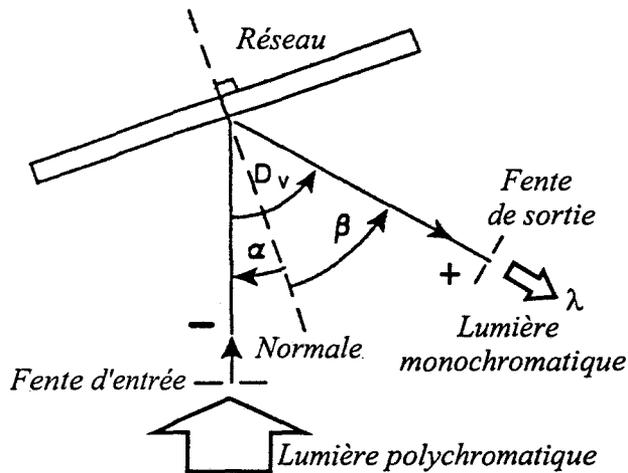


Figure 18 : Sélection des longueurs d'onde par un monochromateur.

On montre que le faisceau passant par la fente de sortie a une longueur d'onde λ (nm) telle que :

$$(éq. 30) \quad \sin \alpha + \sin \beta = 10^{-6} \cdot k n \lambda$$

avec α l'angle du faisceau incident avec la normale du réseau, β l'angle du faisceau diffracté avec la normale, k l'ordre de la diffraction et n la densité de trait du réseau (traits/mm)^(76,77). Or dans les configurations ordinaires, les fentes sont fixes et le réseau tourne autour d'un axe situé en son centre. On peut donc dire que la quantité $D_v = \beta - \alpha$ est constante pour une configuration optique donnée. Ainsi par rotation du réseau, les variations des angles α et β permettent de sélectionner des longueurs d'onde différentes. De cette façon, nous pouvons obtenir le spectre d'un échantillon.

Du point de vue fondamental, la résolution détermine la capacité du monochromateur à séparer deux éléments spectraux adjacents. Elle dépend de nombreux facteurs tels que les phénomènes de diffraction, les aberrations, la qualité des composants optiques, leur alignement et la largeur des fentes d'entrée et de sortie du monochromateur. Pour un système optique donné, l'élargissement ou le rétrécissement des fentes permet de couvrir une large gamme de résolution. La résolution doit être choisie avec attention. Si elle est trop faible, les spectres auront tendance à perdre leurs informations spectrales (phénomène de lissage). Ce phénomène est observé lorsque les fentes sont trop ouvertes. Une résolution trop importante diminue rapidement le rapport signal sur bruit. Les fentes sont fines, le signal diminue rapidement. Prenons l'exemple d'un monochromateur aux fentes d'entrée et de sortie identiques. Si elles sont suffisamment larges, nous pouvons considérer qu'il n'y a pas d'aberrations optiques. La résolution à une longueur d'onde donnée est alors obtenue par multiplication de la dispersion du réseau à la longueur d'onde considérée par la largeur de fente. Le flux sortant est alors proportionnel au carré de la surface de la fente. Le Tableau 6 présente l'effet de la résolution spectrale sur le flux sortant du monochromateur pour une dispersion de 10 nm/mm et des hauteurs de fente supposées fixes. Nous avons attribué un flux arbitraire de 100 pour la résolution de 10 nm afin de faciliter la comparaison.

Résolution (nm)	Largeur de fente (mm)	Flux (Unité arbitraire)
2	0,2	4
6	0,6	36
10	1,0	100
20	2,0	400

Tableau 6 : Influence de la résolution spectrale sur le flux sortant du monochromateur.

Le filtre d'ordre constitue la dernière pièce principale d'un monochromateur. Il permet comme son nom l'indique de sélectionner un ordre de diffraction d'un réseau. La Figure 19 représente le phénomène de diffraction d'un faisceau polychromatique arrivant sur le réseau sous un angle α par rapport à la normale.

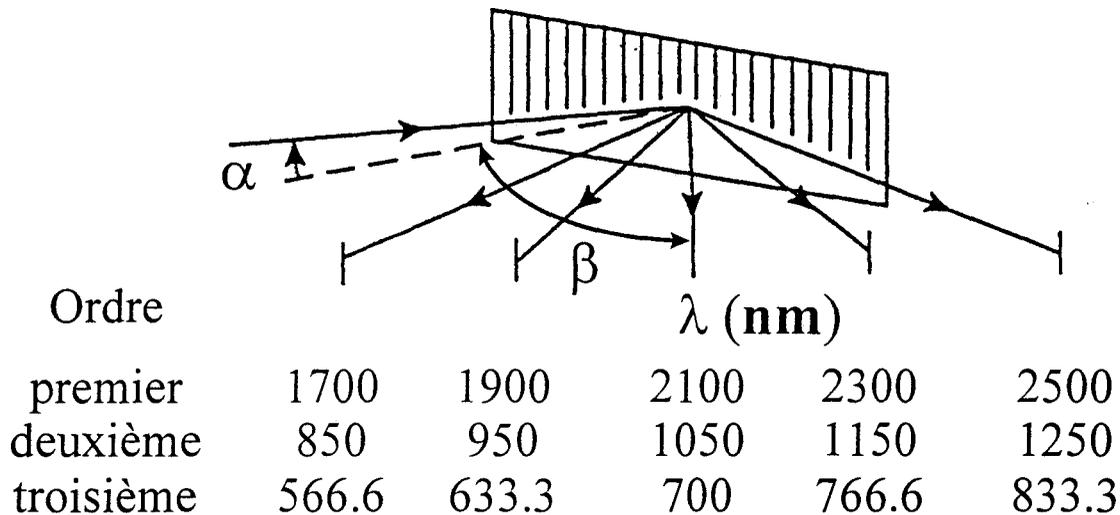


Figure 19 : Superposition des ordres de diffraction d'un réseau.

Dans cet exemple, le réseau de diffraction permet d'obtenir un spectre de 1700 à 2500 nm au premier ordre. Pour un réseau (n défini) et des valeurs de α et β données nous obtenons à partir de (éq. 30) :

$$(éq. 31) \quad k\lambda = \text{constante}$$

Ainsi, un détecteur placé sous un angle β afin d'analyser le signal à la longueur d'onde $\lambda = 2500$ nm au premier ordre, analysera aussi la longueur d'onde $\lambda/2$ soit 1250 nm au second ordre, la longueur d'onde $\lambda/3$ soit 833.3 nm au troisième ordre etc... Si un filtre est placé à l'entrée du monochromateur afin de supprimer toutes les longueurs d'onde inférieures à 1300 nm contenues dans le spectre continu de la lumière polychromatique, il n'y aura plus de recouvrement d'ordre. Le détecteur analysera uniquement la contribution à 2500 nm. L'absence de ce filtre augmente donc considérablement le recouvrement spectral.

Les monochromateurs à réseaux tournants sont souvent très lents mais nous verrons que des applications analytiques rapides sont possibles moyennant quelques modifications instrumentales. Il reste néanmoins le système le moins coûteux pour obtenir le spectre complet d'un échantillon. Leur fiabilité a été démontrée depuis longtemps. Les

monochromateurs sont d'ailleurs utilisés par un grand nombre de constructeurs de systèmes NIR^(26,38).

2.3.3.2 Les fibres optiques

Une fibre optique permet de transporter une énergie lumineuse entre deux points. Elle équipe donc les instruments d'analyses spectrométriques délocalisés du lieu de la mesure. Une fibre optique est caractérisée par sa capacité à transmettre le signal optique sur un certain domaine spectral, sur une certaine distance et par son ouverture numérique.

Afin de définir l'ouverture numérique, revenons sur le principe de transmission d'un faisceau lumineux dans ce guide d'ondes. Les fibres utilisées en proche infrarouge sont des fibres à sauts d'indice. Elles sont constituées d'un cœur de silice pur d'indice de réfraction n_1 (« core ») et d'un enrobage de silice dopée d'indice n_2 (« cladding ») (Figure 20). La transmission du signal optique repose sur une réflexion totale interne à l'interface des deux milieux silicés⁽⁷⁸⁾.

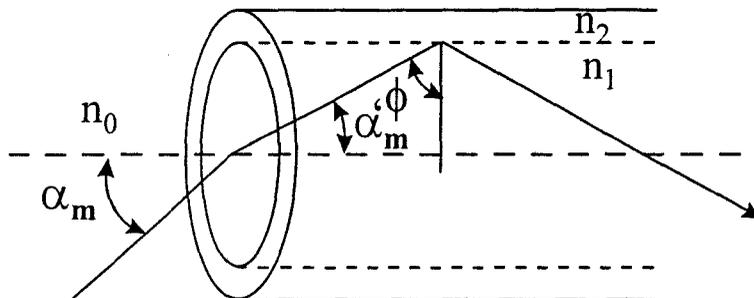


Figure 20 : Transmission du signal dans une fibre optique.

A partir des lois de l'optique géométrique, nous pouvons montrer qu'une onde se propage par réflexion totale dans la fibre si l'angle ϕ est supérieur à ϕ_c avec :

$$\text{(éq. 32) } \phi_c = \text{Arcsin} (n_2 / n_1)$$

Pour un angle ϕ inférieur à ϕ_c le faisceau lumineux sera absorbé par l'enrobage de la fibre optique. A partir de la dernière expression, nous allons pouvoir définir l'angle maximum α_m sous lequel un rayon lumineux peut entrer ou sortir d'une fibre optique et effectuer des réflexions multiples. Considérons un faisceau lumineux arrivant d'un milieu extérieur à la fibre d'indice n_0 sous un angle α_m . La loi de SNELL et la trigonométrie classique nous donnent :

$$(éq. 33) \quad n_0 \cdot \sin \alpha_m = n_1 \cdot \sin \alpha'_m = n_1 \cdot \cos \phi_c = n_1 \cdot (1 - \sin^2 \phi_c)^{1/2} = (n_1^2 - n_2^2)^{1/2} = NA$$

Nous définissons ainsi l'ouverture numérique par le paramètre NA. Les fibres en silice communément utilisées en proche infrarouge possèdent une ouverture numérique variant de 0,2 à 0,3. L'angle maximum sous lequel un faisceau lumineux peut entrer ou sortir de la fibre varie donc de 11 à 17°. Ce paramètre intrinsèque à la fibre est une limitation car l'éclairage ou la collection de l'énergie lumineuse est très localisée. La sonde optique devra donc être développée dans le but de minimiser cet effet.

La transmission du signal optique à travers la fibre est un paramètre très important. Les matériaux qui constituent ces guides d'ondes sont choisis pour leur transparence sur le domaine spectral considéré. Seules les fibres de silice permettent de couvrir le domaine du proche infrarouge. La figure ci-dessous donne l'atténuation du signal optique passant par une fibre de silice sur le domaine 0,4-2,5 µm soit 400-2500 nm.

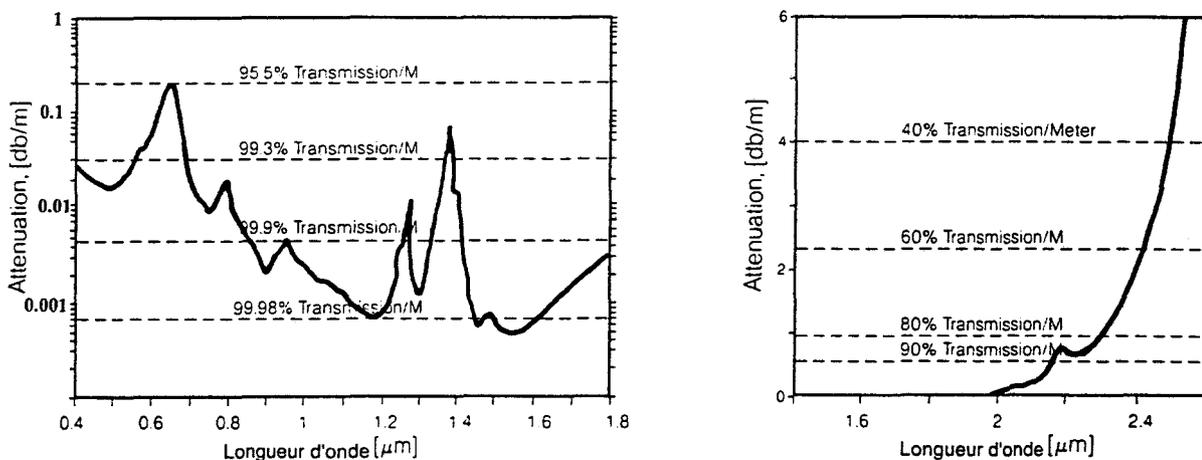


Figure 21 : Atténuation du signal optique dans une fibre de silice.

Nous rappelons qu'une atténuation exprimée en décibels est donnée par :

$$(éq. 34) \quad \text{Atténuation (dB / m)} = \left(\frac{10 \cdot \log \left[\frac{\text{Puissance entrant dans la fibre}}{\text{Puissance sortant de la fibre}} \right]}{\text{Longueur(m)}} \right)$$

On remarque que la fibre est particulièrement transparente entre 900 et 1800 nm puisqu'elle permet une transmission minimum de 99.4 % par mètre sur ce domaine. Aux alentours de 1400 nm, nous observons des absorptions dues aux groupements OH présents dans la fibre. C'est la maîtrise du faible taux en OH qui a permis de concevoir des fibres en silice performantes. Au-dessus de 2000 nm, nous observons une atténuation importante rendant

l'analyse spectrale très difficile sur une longue distance⁽⁷⁹⁾. En effet, à 1800 nm nous avons une atténuation de 0,003 dB soit 99.93 % de transmission pour un mètre. Pour une analyse déportée de 20 m, la transmission sera de 0,9993²⁰ soit 98.6 %. Par contre à 2300 nm, nous avons une atténuation de 0,7 dB soit 85 % de transmission pour un mètre. Sur une distance de 20 m, la transmission chute à 3.87 %. A cette atténuation qui est intrinsèque à la fibre, il faudra bien entendu ajouter l'absorption de l'échantillon. Dans ces conditions, le signal qui transite par la fibre est largement atténué pour les grandes longueurs d'onde. Cet exemple simple montre la limitation des fibres qu'il faudra surmonter dans cette recherche. Il existe des fibres en chalcogénure vitreux⁽⁸⁰⁾ non absorbantes au delà de 2000 nm utilisées dans le moyen infrarouge (longueur d'onde supérieure à 2500 nm). Malheureusement, elles sont très onéreuses à cause de la difficulté de fabrication. Leur utilisation dans l'étable serait aussi hasardeuse car elles sont fragiles. Dans le proche infrarouge, on utilise quelques fois des fibres en fluorure de Zirconium mais elles coupent aussi rapidement le signal. La fibre de silice reste à ce jour la plus adaptée aux applications proche infrarouge.

2.3.3.3 La chaîne électronique de détection

L'efficacité croissante de la spectroscopie proche infrarouge a été principalement obtenue grâce à l'augmentation du rapport signal sur bruit⁽³⁸⁾. En fait, nous pouvons considérer que les systèmes et les composants électroniques sont responsables de ce phénomène⁽⁸¹⁾. La chaîne électronique de détection est constituée d'un détecteur, d'un amplificateur et d'une détection synchrone.

La détection synchrone

Le spectre de la lumière du jour a des contributions dans le domaine du proche infrarouge. Si une partie de cette énergie, aussi faible soit elle, arrive sur le détecteur de l'analyseur, elle sera comptabilisée à tort comme provenant de la source. Le bruit du système devient plus important car des fluctuations externes sont ajoutées. La modulation en fréquence de la source supprime ce problème. En effet, les contributions spectrales extérieures à l'analyseur ne sont pas modulées. La détection synchrone effectue la démodulation du signal qui arrive sur le détecteur et filtre ainsi le signal utile. Elle est d'ailleurs utilisée pour de nombreux montages où le signal est faible dans un bruit important⁽⁸²⁾. Il existe à l'heure

actuelle deux types de détection synchrone : les « analogiques » et les « numériques ». Les détections analogiques utilisent des composants électroniques classiques. Un jeu de filtres, d'amplificateurs et de diviseurs démodulent le signal. Les détections numériques travaillent avec des microprocesseurs spécialisés pour le traitement du signal (*DSP Digital Signal Processing*). Le signal analogique qui arrive du détecteur est numérisé par un convertisseur. Le microprocesseur filtre numériquement le signal modulé en utilisant l'algorithme de la transformée de Fourier. Les détections numériques surpassent de loin les détections analogiques par leur qualité de filtrage et leur rapidité⁽⁸²⁾ mais elles sont néanmoins plus onéreuses. De plus, de nombreux spectromètres proches infrarouges, basés sur la détection synchrone analogique ont déjà assuré des analyses quantitatives de qualité⁽³⁸⁾. Dans ces conditions, la nécessité d'une détection synchrone numérique doit être systématiquement étudiée lors du développement d'un spectromètre.

Le détecteur

Depuis de nombreuses années, le détecteur au sulfure de plomb (PbS) tient la première place pour l'analyse du domaine proche infrarouge⁽⁸³⁾. Il appartient à la famille des détecteurs photoconductifs au même titre que les détecteurs de type PbSe (Sélénure de plomb). Il est construit à partir de matériaux homogènes dont le schéma de principe est donné ci-dessous (Figure 22).

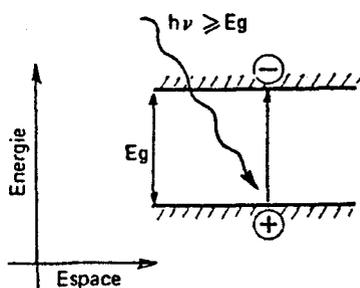


Figure 22 : Schéma de principe d'un détecteur photoconductif.

La lecture de l'effet photoélectrique, matérialisé par l'apport d'électrons dans la bande de conduction (ou de trous dans la bande de valence) est effectuée par la mesure des variations de la résistance du semi-conducteur. Le succès du détecteur PbS est principalement dû au fait qu'il est le seul à pouvoir couvrir le domaine 900-2500 nm avec une bonne détectivité. Sa surface active peut aussi atteindre 1 cm². Il reste néanmoins relativement lent puisque son

temps de réaction peut varier de 50 à 300 μs . Le deuxième rang des détecteurs proche infrarouge est tenu par l'InGaAs⁽⁸⁴⁾ (Arséniure de Gallium et d'Indium). Il fait partie des détecteurs photovoltaïques au même titre que les détecteurs Ge (germanium), InAs (Arséniure d'indium), InSb (Antimoineure d'Indium) et MCT (Tellure de mercure et de cadmium). Ils sont construits sur la base d'une jonction dont le schéma de principe est donné sur la Figure 23. Lorsqu'une énergie $h\nu$ arrive sur le détecteur, il y a création d'une paire électron-trou dans la région n. Le trou est ensuite drainé à travers la jonction vers la région p où les trous sont majoritaires. Il y a création d'un courant entre les bornes du détecteur.

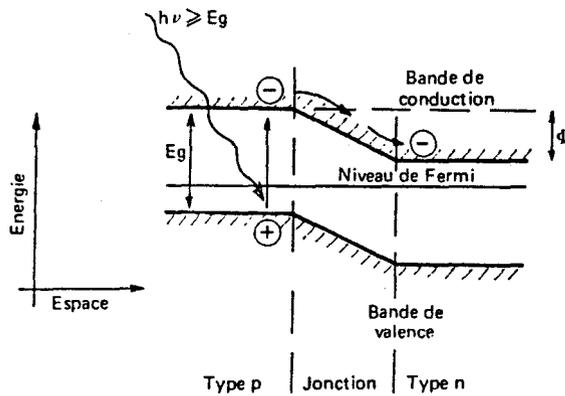


Figure 23 : Schéma de principe d'un détecteur photovoltaïque.

Ces détecteurs sont très rapides puisque leur temps de réaction varie de 1 à 10 μs . Néanmoins, la surface active des détecteurs InGaAs est souvent très faible (1 mm^2 en moyenne). La Figure 24 donne une comparaison des détectivités des détecteurs PbS et InGaAs.

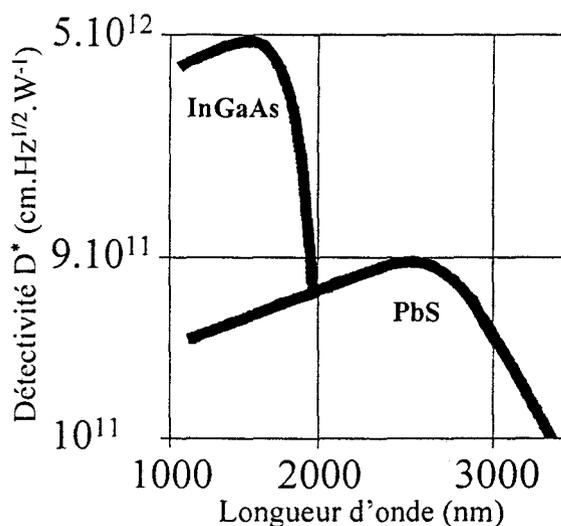


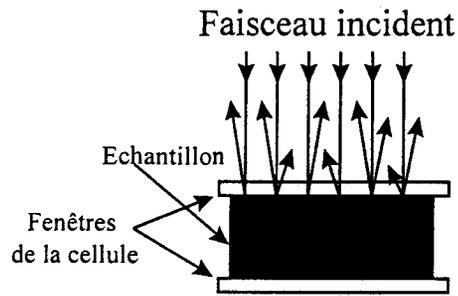
Figure 24 : Détectivité des détecteurs PbS et InGaAs.

La détectivité, mesurée par le paramètre D' , caractérise le rapport signal sur bruit obtenu pour une énergie incidente de 1 Watt, normalisé par une surface de 1 cm^2 et une bande passante de bruit de 1 Hz. Le détecteur InGaAs montre une détectivité excellente aux alentours de 1500 nm mais qui chute complètement à partir de 1700 nm. Le détecteur PbS donne une détectivité moyenne sur une large zone. Toutes ces observations montrent la nécessité d'étudier les effets du détecteur sur la qualité spectrale lorsqu'il est intégré dans un système spectrométrique donné.

2.3.3.4 Les techniques d'échantillonnage

Avant de commencer l'analyse spectrale d'un échantillon, il convient de choisir une technique d'échantillonnage. Cette étape est très importante, car le mode de présentation de l'échantillon conditionne la faisabilité d'une analyse quantitative par la répétabilité, la sensibilité et la spécificité des informations spectrales. Par ce choix, nous définissons l'éclairage de l'échantillon mais aussi la collection de la lumière transmise, diffusée ou réfléchi. La configuration souvent figée des instruments NIR commerciaux ne nous permet pas de changer de technique. Lorsque l'on met en place une analyse proche infrarouge, il faut se demander quel est le phénomène optique dominant dans l'échantillon. La figure ci-dessous représente trois cas typiques (Figure 25). Le premier échantillon analysé est très diffusant(a), le second l'est modérément (b) alors que le troisième l'est très peu, limpide comme un solvant pur (c). Dans le dernier cas, l'analyse spectrale est facile. C'est la transmission du faisceau qui domine. Pour des chemins optiques raisonnables, le phénomène de diffusion sur les faces d'entrée et de sortie de la cellule est négligeable. En plaçant un détecteur derrière la cellule, nous pouvons obtenir facilement un spectre NIR de qualité. Le premier cas (a) représente les échantillons très diffusants tels que les poudres. Nous pouvons considérer que même sur des chemins optiques faibles, il n'y a pas de transmission derrière la cellule.

(a) Réflexion diffuse



(b) Réflexion et transmission diffuses (c) Transmission



Figure 25 : Interactions rayonnement-matière.

Pour ce type d'échantillon, il existe deux modes d'échantillonnage : la géométrie 0-45°⁽⁸⁵⁾ et la sphère d'intégration^(86,87) (Figure 26). Ils utilisent tous les deux un faisceau monochromatique qui est envoyé sous une incidence normale par rapport à la surface de l'échantillon. Dans le premier dispositif, la lumière diffusée dans tout le demi-espace est collectée par 2 ou 4 détecteurs placés à 45° toujours par rapport à cette normale.

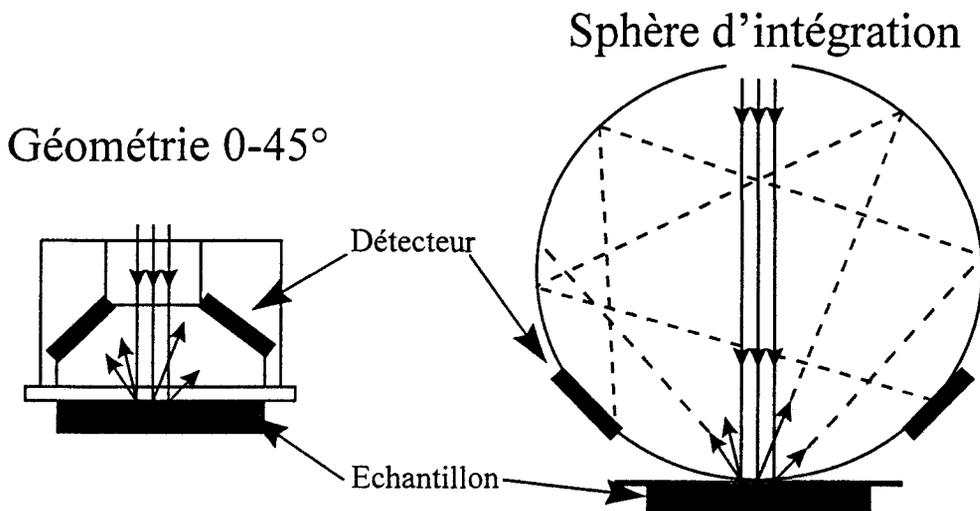


Figure 26 : Systèmes de récupération du signal diffusé.

Pour le deuxième dispositif, la lumière diffusée effectue des réflexions multiples à la surface de la sphère d'intégration. Le signal est intégré par des détecteurs collés sur cette même surface. Les deux systèmes donnent de bons résultats mais la sphère d'intégration reste supérieure. Elle intègre mieux l'espace de diffusion (moins de pertes de signal) et prend moins en compte les réflexions spéculaires. Finalement, l'échantillon le plus difficile à analyser en proche infrarouge est le liquide hétérogène (Figure 25b). S'il est très diffusant, une analyse efficace peut être effectuée avec les deux dispositifs précédents. Par contre, si la diffusion est moyenne, une partie de l'énergie incidente est transmise. Les dispositifs précédents sont inutilisables, car la diffusion qui arrive sur les détecteurs est trop faible. A ce stade, on utilise la transflexion (Figure 27). La lumière transmise ou diffusée vers l'arrière de la cellule est renvoyée vers l'avant par un matériau réfléchissant. Les dispositifs précédents sont à nouveau utilisables.

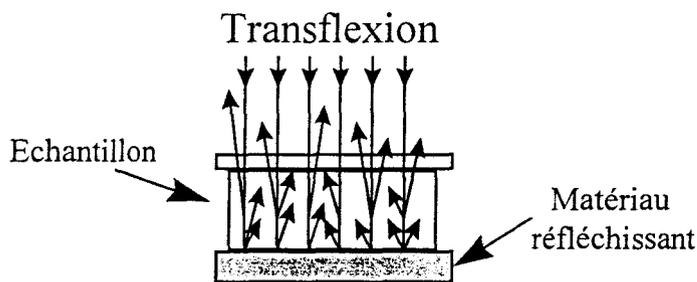


Figure 27 : La transflexion.

En proche infrarouge, un spectre d'absorbance s'exprime différemment en fonction de la technique d'échantillonnage utilisée. Pour le mode transmission simple, nous avons :

$$(éq. 35) \quad A = \log \left(\frac{1}{T} \right) \quad \text{avec} \quad T = \frac{S_{\text{Echantillon}}}{S_{\text{Référence}}}$$

où T est la transmission de l'échantillon, $S_{\text{Echantillon}}$ et $S_{\text{référence}}$ (simple faisceau) les signaux qui arrivent au détecteur lorsqu'on analyse respectivement l'échantillon et la référence. Pour la réflexion diffuse et la transflexion, on introduit la réflectance de l'échantillon notée R :

$$(éq. 36) \quad A = \log \left(\frac{1}{R} \right) \quad \text{avec} \quad R = \frac{S_{\text{Echantillon}}}{S_{\text{Référence}}}$$

Pour des échantillons très diffusants (cas des poudres), la forme logarithmique de l'absorbance ne suit plus la loi de Beer-Lambert. On utilise la fonction de KUBELKA-MUNK^(39,88,89) :

$$\text{(éq. 37)} \quad A = \frac{1 - R^2}{2R}$$

2.4 Avantages et inconvénients de la spectroscopie NIR

La spectroscopie NIR offre de nombreux avantages pour l'analyse quantitative alors qu'elle présente un intérêt limité pour l'étude structurale. En effet, il est impossible de donner une interprétation complète des bandes dans un spectre proche infrarouge. Le recouvrement des bandes larges et faibles en intensité ne nous aide en rien.

Le premier avantage est propre à l'instrumentation. La source NIR possède un bon rendement énergétique. En effet le maximum d'émission de ces sources se trouve dans le domaine spectral étudié. Il en résulte une énergie importante et une lumière parasite faible. Ces facteurs combinés aux capacités des détecteurs permettent d'obtenir un rapport signal sur bruit de plus de 10000.

Les matériaux tels que le verre ou le quartz sont transparents dans le proche infrarouge. Il est donc possible d'analyser des échantillons conditionnés en fin de process^(90,91,92). De plus, des fibres optiques en silice permettent de délocaliser l'analyse sur de grandes distances. Les détecteurs NIR ont un temps de réponse très court de l'ordre de 1 à 500 μ s. Des analyses très rapides peuvent ainsi être effectuées. Mais l'avantage certainement le plus important pour la spectroscopie proche infrarouge est la diminution de l'absorbance pour les bandes harmoniques successives. Ainsi, contrairement à l'infrarouge moyen, on peut analyser avec moins de contraintes des échantillons modérément plus concentrés sur des chemins optiques plus importants⁽⁹³⁾.

Les bandes harmoniques et les combinaisons ont des profils larges et arrondis. C'est un avantage considérable pour l'analyse quantitative. En effet un léger déplacement de bande, pour une raison non contrôlée aura une plus faible influence sur l'intensité de la lumière absorbée à la longueur d'onde considérée. La spectroscopie proche infrarouge permet d'analyser des échantillons in situ dans leurs états naturels sans les détruire^(94,95). Le niveau d'absorption des solvants dans le proche infrarouge reste acceptable. Pour obtenir une mesure spectrale ayant un sens, il est indispensable qu'une quantité non négligeable de la lumière incidente ne soit pas absorbée (signal non saturé). Pour l'eau par exemple les bandes sont exploitables et bien localisées. Son dosage est effectué sur de nombreux produits organiques.

En spectroscopie infrarouge, nous sommes souvent gênés par des phénomènes de diffusion parasite. C'est le cas des produits agro-alimentaires qui sont assez hétérogènes. Ces interférences sont très importantes lorsque la longueur d'onde du faisceau incident est proche de la taille des particules étudiées. Or les longueurs d'onde du proche infrarouge (0,7 à 2,5 microns) sont très éloignées de la taille des particules (10 à 100microns)⁽⁹⁶⁾. Ces interférences sont donc moindres. Pour toutes ces raisons la spectroscopie NIR à un grand potentiel d'analyse de produits agro-alimentaires^(97,98).

III TRAITEMENTS MATHÉMATIQUES DES DONNÉES SPECTRALES

3.1 Introduction

L'objectif de l'analyse quantitative en spectroscopie est de créer un modèle mathématique qui permet de relier le spectre d'un échantillon aux concentrations des constituants que l'on veut doser. Comme nous avons pu le voir dans la partie historique, c'est le développement des techniques de régressions multivariées qui a permis une exploitation de la spectroscopie NIR. Ce chapitre va nous permettre d'apprécier la méthodologie à suivre lors de la mise en place d'une analyse quantitative dans le proche infrarouge. Nous insisterons sur l'importance de chacune des phases pour l'obtention d'une analyse de bonne qualité. Après une description globale de la procédure de calibration, nous nous attarderons sur les concepts mathématiques utilisés dans ces analyses.

3.2 Notations

Il n'existe pas à l'heure actuelle de normalisation officielle en chimiométrie aussi bien pour la terminologie employée que pour les notations. Celles rencontrées dans la littérature sont donc aussi variées que contradictoires malgré une tentative de normalisation de MARTENS *et al*⁽¹⁴⁾. Les notations employées dans ce mémoire sont les suivantes :

- Les variables scalaires sont notées par des lettres minuscules normales.

Exemple : a , b , c .

- Les matrices seront représentées par des lettres majuscules en gras, à l'exception des vecteurs.

Exemple : \mathbf{A} , \mathbf{H} , \mathbf{T} .

- Les vecteurs (matrices lignes ou colonne) sont représentés par des caractères minuscules gras.

Exemple : \mathbf{t} , \mathbf{a} , \mathbf{e} .

- La matrice transposée de \mathbf{A} est notée \mathbf{A}^t .
- La matrice inverse de \mathbf{A} est notée \mathbf{A}^{-1} .

3.3 Principe

Une analyse quantitative se divise en trois étapes : la calibration, la validation et la prédiction. La Figure 28 donne le schéma de son développement.

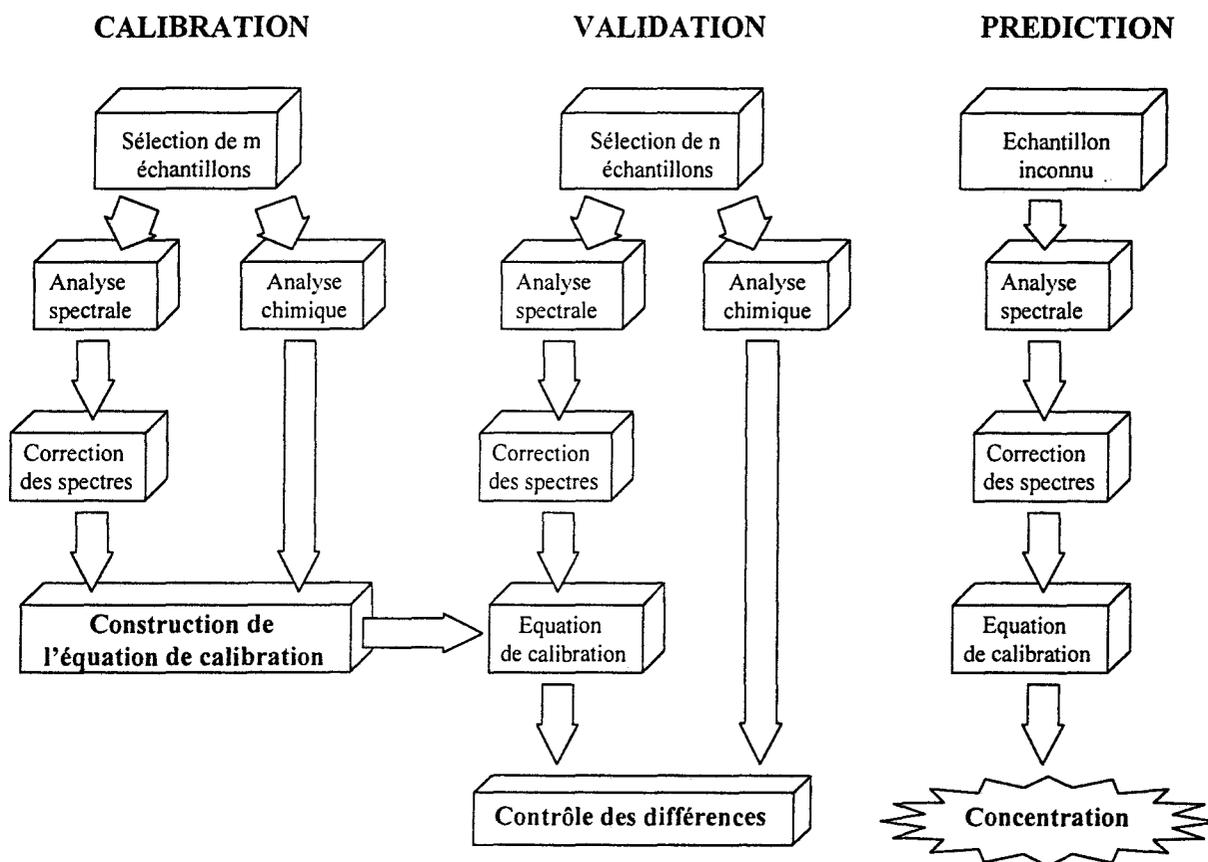


Figure 28 : Schéma de principe d'une analyse quantitative.

L'étape de calibration commence par la sélection de m échantillons. Ils sont choisis de telle sorte que les gammes de concentrations qu'ils couvrent reflètent bien les échantillons inconnus qu'il faudra analyser par la suite. Dans une population d'échantillons naturels comme les laits de vache, les concentrations sont réparties de manière sensiblement gaussienne. Ainsi un choix d'échantillons complètement au hasard risque de limiter les gammes de concentrations à des valeurs proches des valeurs moyennes. Cette sélection ne conduit pas à des équations de calibration satisfaisantes⁽⁹⁹⁾. Les concentrations des échantillons de calibration doivent donc être régulièrement réparties sur la plage de mesure. Ils doivent aussi tenir compte de tous les facteurs de variabilité pouvant influencer le spectre NIR. Finalement, ils se doivent d'être « physiquement » représentatifs de ces mêmes échantillons

inconnus : une calibration pour le dosage des protéines dans la viande est inutilisable sur les protéines du lait par exemple. Parallèlement à la sélection des échantillons par rapport à leurs concentrations, il existe de plus en plus de méthodes basées sur les données spectrales^(100,101,102,103,104,232). De cette façon, la sélection se fait en fonction de la variance de spectres. Lorsque les échantillons sont synthétiques, les plans d'expériences permettent de sélectionner un faible nombre d'échantillons couvrant toutes les gammes de concentrations⁽¹⁰⁵⁾. Néanmoins, la complexité des échantillons naturels ou industriels rend cette technique souvent inutilisable. La deuxième étape de la calibration est la mesure des concentrations des constituants que l'on voudra doser par le modèle mathématique. A ce stade, il faut choisir une méthode de dosage dite de référence. Elle doit être pertinente et précise si nous voulons avoir des chances de créer une analyse quantitative efficace^(106,259). En effet, une analyse quantitative par spectroscopie NIR ne peut obtenir une erreur de prédiction inférieure à l'erreur de la méthode de référence⁽¹³⁵⁾. Il faut aussi considérer l'effet d'amplification des erreurs par les régressions multivariées. Les méthodes de dosage retenues pour l'analyse sont reconnues par les instances internationales⁽¹⁰⁷⁾. Parallèlement à ce dosage de référence, on effectue l'analyse spectrale de chaque échantillon. Une grande attention aura été préalablement portée sur la préparation de l'échantillon, sur le choix de la technique d'échantillonnage, sur l'optimisation des paramètres d'acquisition des spectres. Cette étape est très importante, car les techniques de régression ne peuvent pas être utilisées pour l'analyse quantitative si des informations spectrales spécifiques ne sont pas observées. De façon systématique, la répétabilité et la linéarité des mesures spectrales sont vérifiées pour un protocole d'analyse donné^(108,109,110). La répétabilité est primordiale pour la réussite de toute méthode d'analyse quantitative. Elle se définit comme l'écart type des données spectrales issues d'un protocole d'analyse effectué plusieurs fois sur le même échantillon, dans le même laboratoire, par le même manipulateur et à des intervalles de temps courts. Elle est caractérisée, pour chaque longueur d'onde, par le coefficient relatif de déviation standard RSD⁽¹¹¹⁾ exprimé en pour-cent.

$$(éq. 38) \quad RSD = \frac{\sigma}{\bar{a}} \times 100 \quad \text{et} \quad \sigma = \sqrt{\frac{\sum_i^s (a_i - \bar{a})^2}{s}}$$

avec s le nombre de spectres utilisés pour la répétabilité, a_i l'absorbance du spectre i à la longueur d'onde considérée, \bar{a} l'absorbance moyenne des s spectres toujours à cette longueur

d'onde et σ la déviation standard de ces absorbances. On vérifie ensuite la linéarité du signal optique en fonction des concentrations des espèces à doser (Loi de Beer-Lambert).

Lorsque des phénomènes parasites comme la diffusion viennent perturber cette linéarité, on applique des algorithmes de correction aux données spectrales. Ils seront largement développés au paragraphe suivant. Cette première phase se termine par l'étape de construction de l'équation de calibration. Les spectres corrigés et les concentrations de référence des échantillons de calibration sont placés respectivement dans la matrice des données spectrales notée **A** et dans la matrice des concentrations notée **C**. Le but de l'équation de calibration est de reproduire le mieux possible les concentrations de la matrice **C** à partir des spectres de la matrice **A**. Schématiquement, une procédure de calibration se résume à la recherche de la matrice **K** qui relie les matrices **A** et **C** par l'équation matricielle $\mathbf{C} = \mathbf{K} \cdot \mathbf{A}$. De nombreux algorithmes de régression seront exposés à partir de la page 69. Nous insisterons sur leur efficacité par rapport aux conditions d'utilisation⁽¹¹²⁾. Lorsque l'équation de calibration est construite, on calcule deux paramètres statistiques : l'erreur standard de calibration et le coefficient de corrélation. L'erreur standard de calibration^(113,135) notée SEC (*Standard Error of Calibration*) est donnée par l'équation ci-dessous.

$$(éq. 39) \quad SEC = \sqrt{\frac{\sum_{j=1}^m (c_j - c'_j)^2}{m - 1 - q}}$$

avec c_j la concentration de référence de l'échantillon j , c'_j sa concentration prédite par le modèle, m le nombre d'échantillons utilisés pour la calibration et q le nombre de variables indépendantes dans la régression. Elle est exprimée dans l'unité du dosage de référence. Faute de définitions normalisées, les erreurs de calibration sont quelquefois exprimées par les paramètres PRESS (*PREDiction Sum of Squares*) ou RMSD (*Root Mean Square of Differences*). Le coefficient de corrélation noté r a été introduit par la statistique de K. PEARSON au début du siècle⁽¹¹⁴⁾. Il mesure la part de variance totale expliquée par la régression.

$$(éq. 40) \quad r = \sqrt{\frac{\sum_{j=1}^m (c'_j - \bar{c}')^2}{\sum_{j=1}^m (c_j - \bar{c})^2}}$$

avec \bar{c} la moyenne des m concentrations de référence et \bar{c}' la moyenne des m concentrations prédites par le modèle. L'erreur standard de calibration et le coefficient de corrélation ne permettent pas d'apprécier la qualité prédictive du modèle construit. Une erreur standard de calibration SEC faible et un coefficient de corrélation r^2 proche de 1 ne constituent qu'une condition nécessaire mais non suffisante au développement d'une bonne analyse quantitative. C'est donc l'étape de validation qui jouera ce rôle. En premier lieu, on sélectionne n échantillons différents des m utilisés en calibration (Figure 28). Ils sont suffisamment importants en nombre afin d'obtenir une bonne représentativité du comportement prédictif du modèle. Leurs concentrations couvrent tous les domaines fixés par la calibration. On effectue les dosages par les méthodes de référence ainsi que l'analyse spectrale de chaque échantillon conformément au protocole d'échantillonnage utilisé en calibration. Les n spectres sont alors transformés par la méthode de correction fixée dans la calibration. Ils constituent une nouvelle matrice de données spectrales notée A^u . La matrice K de l'équation de calibration donne finalement les concentrations des échantillons notées C^u par la relation $C^u = K \cdot A^u$. La qualité de l'analyse quantitative se mesure par les différences entre les concentrations prédites par le modèle et les concentrations de référence. A cet effet, on introduit l'erreur standard de prédiction notée SEP (*Standard Error of Prediction*)

$$(éq. 41) \quad SEP = \sqrt{\frac{\sum_{j=1}^n (c_j - c'_j)^2}{n-1}}$$

avec c_j la concentration de référence de l'échantillon j , c'_j sa concentration prédite par le modèle, n le nombre d'échantillons utilisés pour la validation. Si l'erreur est en accord avec le niveau de précision d'analyse fixé au départ, l'équation de calibration peut être utilisée dans la dernière phase de prédiction d'échantillons inconnus.

3.4 Le pré-traitement des données spectrales

Si le spectre NIR n'était pas affecté par la taille des particules et les variations de chemins optiques, nous pourrions les utiliser tels quels pour l'analyse quantitative. La première façon de maîtriser la diffusion de la lumière, cause principale de ces variations, est de standardiser la préparation de l'échantillon afin d'obtenir des propriétés optiques égales sur toutes les données⁽¹¹⁵⁾. Une deuxième alternative, pour éliminer ou du moins diminuer ces

interférences, est l'application d'un pré-traitement des spectres pour rendre au signal optique son caractère linéaire par rapport aux concentrations. Les traitements les plus utilisés sont développés ci-dessous. Il est impossible de dire par avance si tel ou tel autre traitement est le meilleur pour un problème donné. La seule technique est de tester toutes les méthodes et de retenir celle qui donne l'erreur standard de prédiction la plus faible^(116,154).

3.4.1 La dérivée

La dérivée a été la première transformation mathématique utilisée en spectroscopie proche infrarouge. Elle est appliquée aux données de nombreuses spectroscopies (UV, MIR, fluorescence...)^(117,118). Elle réduit les variations de ligne de base^(119,120,121) et sépare plus clairement les bandes d'absorption^(122,123). On exalte ainsi des bandes cachées dans le recouvrement observé sur les spectres proche infrarouge⁽¹²⁴⁾. La Figure 29 représente bien ce phénomène. Sur ces spectres d'absorbance, on peut voir de très faibles différences entre l'eau et le lait⁽¹²⁵⁾. Les informations spectrales dues aux matières grasses à 2304 et 2343 nm sont exaltées par l'utilisation de la dérivée. Néanmoins, un degré de dérivation trop important affecte le rapport signal sur bruit.

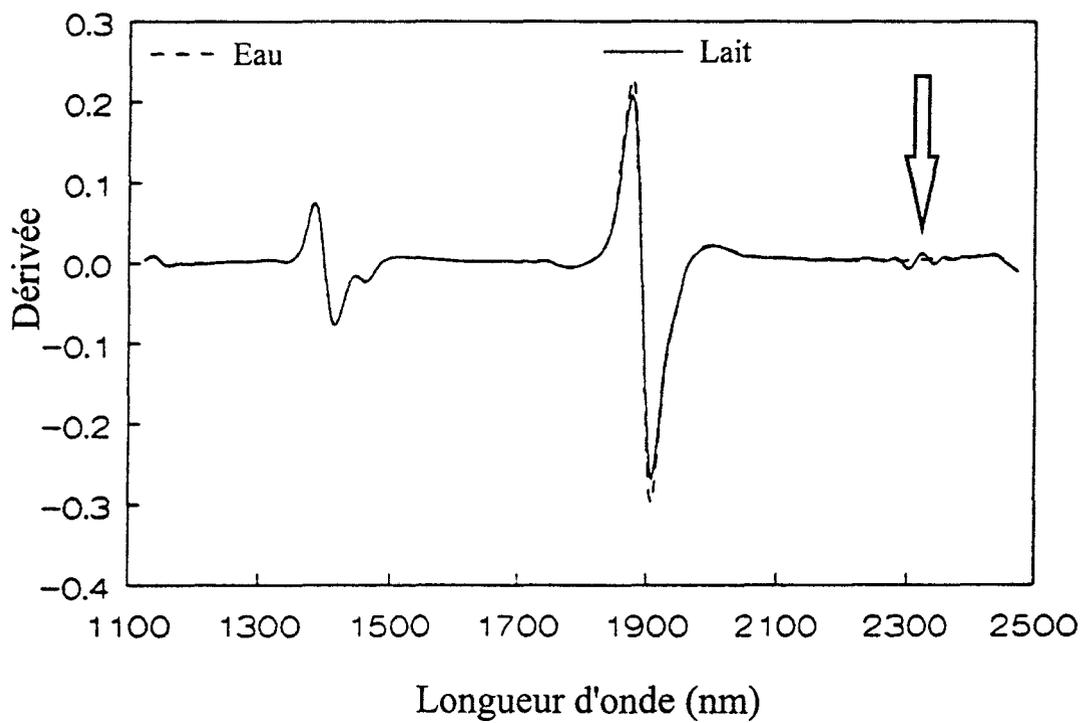
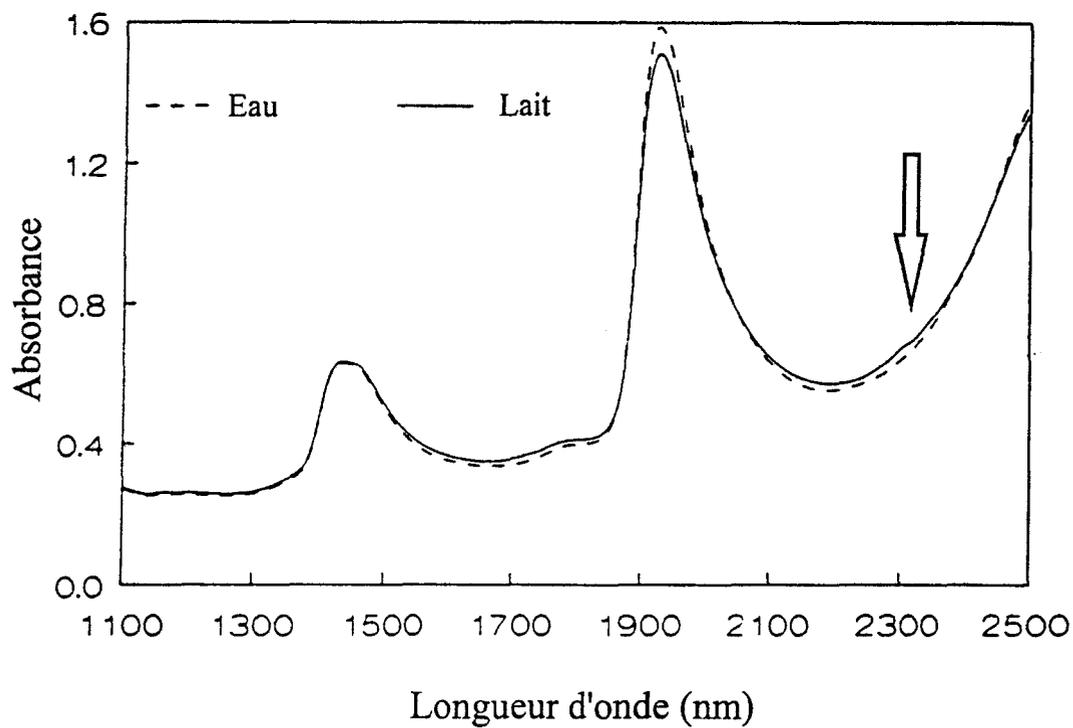


Figure 29 : Exaltation des informations spectrales par la différentiation

Il existe trois types de dérivées :

La dérivée « différence simple »

La méthode « différence simple » utilise une soustraction entre points successifs pour générer le spectre de dérivée première. Ainsi pour un spectre d'absorbance à n longueurs d'onde (a_1, a_2, \dots, a_n) on obtient le spectre dérivé (d_1, d_2, \dots, d_n) par l'opération sur chaque longueur d'onde i :

$$\text{(éq. 42)} \quad d_i = a_{i+1} - a_i$$

La dérivée seconde est obtenue de la même façon à partir de la dérivée première.

La dérivée « gap »

La méthode « gap »^(126,127,128) est inspirée de la méthode « différence ». La différence ne se fait plus sur des points successifs mais sur des points distants d'un intervalle (le gap) fixé par l'utilisateur. Prenons l'exemple d'un spectre proche infrarouge possédant un incrément entre longueurs d'onde de 5 nm. Si on définit un gap de 20 nm, la dérivée à 900 nm est obtenue par la différence d'absorbance entre les longueurs d'onde 920 et 900.

La dérivée selon « SAVITSKY-GOLAY »

C'est la dérivée la plus utilisée en spectroscopie⁽¹²⁹⁾. Elle est basée sur la technique de convolution de SAVITSKY et GOLAY⁽¹³⁰⁾. Elle effectue une régression linéaire par les moindres carrés entre un polynôme de degré k et au moins $k+1$ points du spectre autour de chaque longueur d'onde où on calcule la dérivée. La dérivée en un point est alors calculée par la dérivée du polynôme « fitté » en ce point. Les dérivées d'ordre 1 à 9 peuvent ainsi être obtenues. Ce calcul utilise le formalisme matriciel de STEINER⁽¹³¹⁾ et MADDEN⁽¹³²⁾ décrit ci-dessous :

Soient d l'ordre de la dérivée,

k le degré du polynôme,

s le nombre de points devant être modélisés par le polynôme,

$m = (s-1)/2$ (nombre de points de convolution),

θ une matrice colonne à k éléments,

A la matrice contenant les données spectrales.

$$\text{Soit } \mathbf{X} = \begin{pmatrix} 1 & \dots & 0 & \dots & 1 \\ -m & \dots & 0 & \dots & +m \\ \dots & \dots & \dots & \dots & \dots \\ (-m)^{k+1} & \dots & 0 & \dots & (+m)^{k+1} \end{pmatrix}$$

On recherche alors la matrice θ telle que $\mathbf{A} = \mathbf{X}\theta$. Par la méthode des moindres carrés, la solution est donnée par la minimisation de la matrice $\mathbf{S} = (\mathbf{A}-\mathbf{X}\theta)^t(\mathbf{A}-\mathbf{X}\theta)$ donc $\frac{\partial \mathbf{S}}{\partial \theta} = 0$.

Finalement $\theta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{A} = \mathbf{T}\mathbf{A}$ avec $\mathbf{T} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. La dérivée $d^{\text{ième}}$ est finalement donnée par la multiplication par $d!$ de la convolution entre la $(d+1)$ ième ligne de \mathbf{T} et le spectre considéré (éq. 43).

$$\text{(éq. 43)} \quad \frac{d^d y}{dx^d} = (d!) \sum_{j=-m}^{+m} y_i T_{d+1, i-j}$$

Il faut noter que la convolution tronque le spectre dérivé de m points aux deux extrémités du domaine spectral. Les publications en spectroscopie montrent l'intérêt porté par les chercheurs pour cet algorithme dans la dérivation^(133,134). Malgré des bandes d'absorption larges en proche infrarouge, les avis sont divisés sur la supériorité d'un algorithme de différentiation. Les trois types de dérivation donnent des résultats d'analyse quantitative sensiblement identiques. Pour l'infrarouge moyen où les bandes sont plus fines, il faudra utiliser la dérivée suivant SAVITSKY-GOLAY sous peine d'une détérioration importante de rapport signal sur bruit.

3.4.2 La normalisation par référence interne

Une façon simple de normaliser un spectre proche infrarouge pour éviter les problèmes d'interférences est de diviser toutes les absorbances de ce spectre par l'absorbance à une longueur d'onde dite de référence⁽¹³⁵⁾. En spectroscopie proche infrarouge une bande d'absorption du solvant est souvent pris comme référence interne si les problèmes de diffusion augmentent artificiellement le chemin optique par exemple. Cette normalisation est aussi utilisée lorsque des constituants interfèrent dans la même zone d'absorption^(136,137). On prend alors comme référence l'absorption d'une bande isolée d'un des composés pour effectuer l'analyse quantitative sur l'autre. Pour illustrer cette opération prenons l'exemple d'un spectre

de mélange de deux composés purs A et B (Figure 30). Le composé A possède deux bandes d'absorption à 12 et 61 en unité arbitraire et le composé B deux autres à 55 et 83. Si on veut effectuer le dosage du composé B à la longueur d'onde 55, nous sommes gênés par le recouvrement partiel de la bande à 61 du composé A. Il faut donc prendre la bande à 12 comme référence.

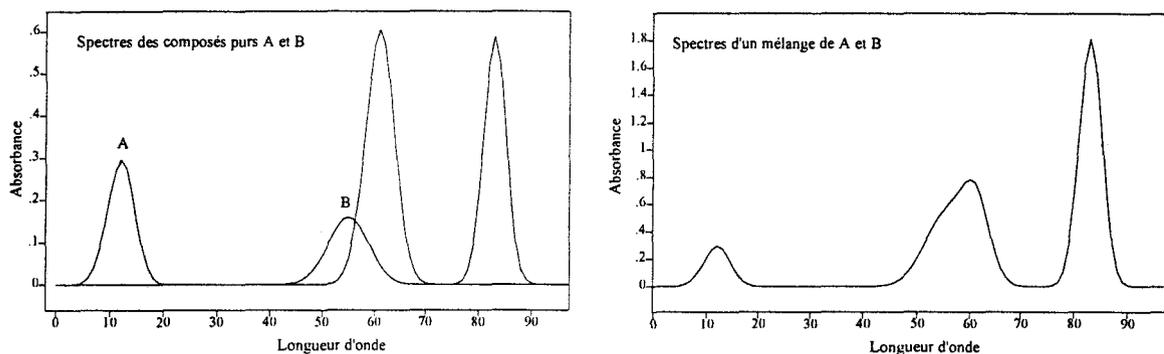


Figure 30 : Sélection d'une longueur d'onde de référence

3.4.3 La normalisation par la méthode « MSC »

Pour un spectre proche infrarouge l'énergie réfléchie provenant d'un échantillon est un mélange de réflexions spéculaires et diffuses. Celles-ci dépendent de la nature diffusante et des caractéristiques d'absorption de l'échantillon⁽¹³⁸⁾. La diversité en taille des particules au sein d'un échantillon affecte la diffusion qui est la cause majeure de variation en spectroscopie NIR⁽¹³⁹⁾. Cette diffusion de nature additive et multiplicative varie donc d'un échantillon à l'autre⁽¹⁴⁰⁾. Pour corriger cet effet parasite très important en mode réflectance, H. MARTENS *et al.* ont développé la correction « MSC » (*Multiplicative Scatter Correction*)^(141,142,143). Depuis peu d'autres auteurs ont renommé cette méthode « Multiplicative Signal Correction » car elle est applicable à d'autres types de données telle que la chromatographie^(144,135). Elle permet de séparer les variations multiplicatives issues de la diffusion des variations additives provenant des concentrations. Pour cette correction, il serait nécessaire d'avoir un spectre de référence appelé spectre « idéal ». A défaut, on prend le spectre moyen des spectres de calibration⁽¹⁴⁵⁾. Considérons s spectres d'absorbance utilisés pour la calibration soit s matrices ($w \times 1$ soit w lignes, 1 colonne) notées a_i ($i = 1 \dots s$) où w représente le nombre de longueurs d'onde. Soit \bar{a} le spectre moyen des a_i . Le principe de la « MSC » est la régression de chaque

spectre a_i sur le spectre moyen \bar{a} . Les coefficients f_i et g_i de cette régression linéaire nous permettent d'accéder aux facteurs additifs et multiplicatifs de la diffusion (éq. 44).

$$(éq. 44) \quad a_i = \mathbf{1} \cdot f_i + \bar{a} \cdot g_i + e_i$$

Le vecteur $\mathbf{1}(w \times 1)$ n'est constitué que d'éléments égaux à 1. Les erreurs $e_i (w \times 1)$ sont en grande partie attribuables aux informations chimiques. Ainsi l'absorbance corrigée a_{ik} à la longueur d'onde k et pour l'échantillon i est obtenu par l'équation suivante :

$$(éq. 45) \quad a_{ik, \text{ corrigé}} = (a_{ik, \text{ non corrigé}} - f_i) / g_i$$

Il est entendu que tout changement dans les spectres utilisés pour la calibration (suppression d'éléments aberrants) modifiera le spectre moyen \bar{a} . La correction devra alors être recalculée. Cette transformation est une technique dépendante du lot de données. La « MSC » a permis d'améliorer les résultats d'analyses quantitatives sur de nombreux problèmes^(140,146,147,148). Une analyse sur des produits pétroliers a même montré que la MSC permettait d'obtenir une calibration moins sensible aux décalages en longueur d'onde⁽¹⁴⁹⁾. Mais les résultats certainement les plus spectaculaires pour cette correction sont issus d'une étude détaillée menée par ISAKSSON *et al.*⁽¹⁴⁵⁾ sur le dosage des trois principaux constituants des produits agro-alimentaires, les protéines, les graisses et l'eau. La Figure 31 illustre bien l'effet de la correction sur les spectres NIR de farines possédant une distribution en taille de particules différente⁽¹⁵⁰⁾.

Il existe finalement une dernière version la « PMSC » (*Piece-wise MSC*) développée par ISAKSSON *et al.*⁽¹¹⁵⁾. Alors que la MSC corrige les effets linéaires additifs et multiplicatifs du signal, la PMSC corrige les non-linéarités. L'idée est d'effectuer la régression linéaire non plus sur tout le spectre mais sur un sous-ensemble de longueurs d'onde. Ainsi pour un spectre continu on remplace le spectre a_i dans (éq. 44) par un vecteur $d_{ik} (m+n+1 \times 1)$ avec m et n les nombres de longueurs d'ondes à prendre à gauche et à droite de k . On effectue en fait une MSC pas à pas sur une fenêtre de longueurs d'onde qui se déplace régulièrement sur l'ensemble du domaine spectral.

$$(éq. 46) \quad d_{ik} = [a_{i, k-m}, a_{i, k-m+1}, \dots, a_{ik}, \dots, a_{i, k+n-1}, a_{i, k+n}]'$$

La taille de la fenêtre est $m+n+1$. De la même façon on applique cette fenêtre au spectre moyen remplaçant \bar{a} par \bar{d}_k :

$$(éq. 47) \quad d_{ik} = \mathbf{1} \cdot f_{ik} + \bar{d}_k \cdot g_{ik} + e_{ik}$$

Par analogie à la MSC le spectre corrigé est donné par l'équation suivante :

$$(éq. 48) \quad a_{ik, \text{ corrigé}} = (a_{ik, \text{ non corrigé}} - f_{ik}) / g_{ik}$$

Afin d'éviter la troncature des spectres transformés (phénomène habituel des méthodes utilisant les fenêtres mobiles), les m premières et les n dernières longueurs d'ondes du domaine spectral sont corrigées par une fenêtre fixe de largeur $(m+n+1)$.

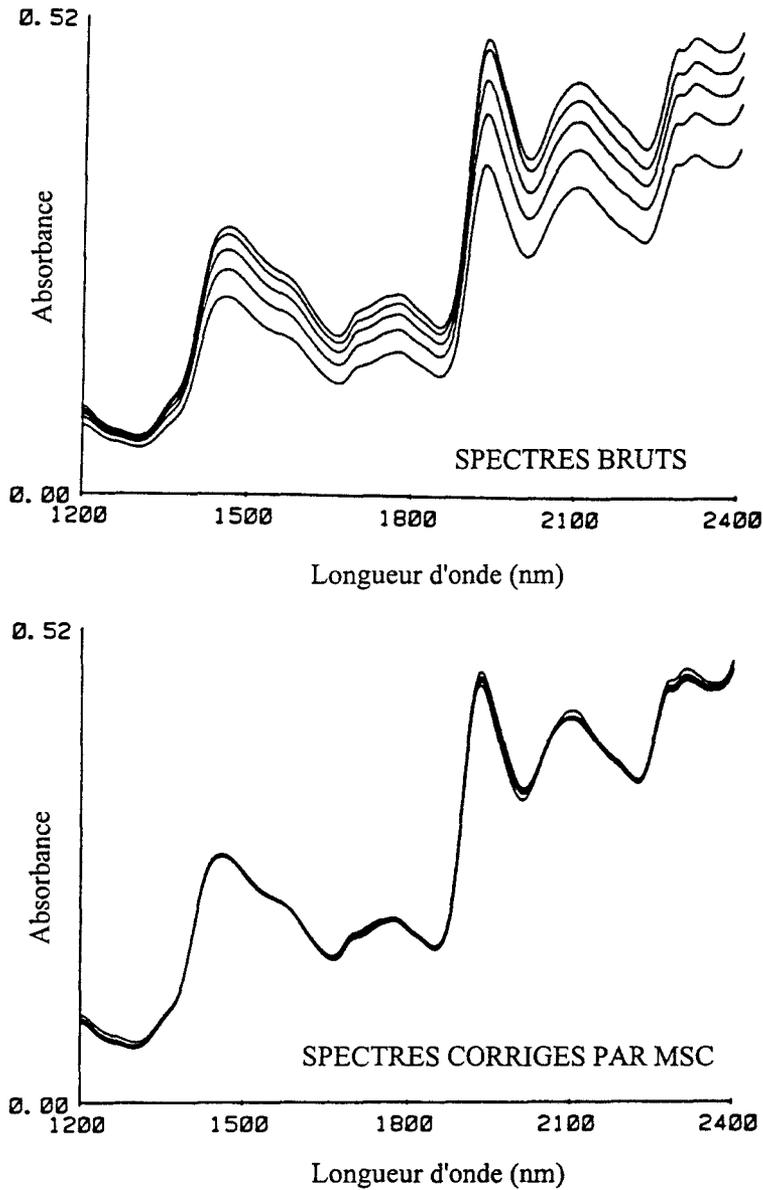


Figure 31: Effet de la MSC sur les spectres NIR

Lorsque les non-linéarités sont importantes cette méthode peut amener de grandes améliorations. L'application de la PMSC a permis d'améliorer les résultats du dosage des protéines, des matières grasses et de l'eau dans la viande de bœuf⁽¹⁵¹⁾, dans les mélanges

viande de bœuf / viande de porc⁽¹⁵²⁾ et finalement dans les saucisses fumées⁽¹⁵³⁾. Le Tableau 7 permet d'apprécier les améliorations sur ces trois analyses.

Viande de Bœuf			
	Spectres bruts log(1/R)	MSC	PMSC
SEP Protéines	0.33	0.27	0.21
Amélioration (%)	-	18	36
SEP Mat. grasses	0.75	0.42	0.34
Amélioration (%)	-	44	55
SEP Eau	0.46	0.42	0.29
Amélioration (%)	-	9	37
Mélange BœufPorc			
	Spectres bruts log(1/R)	MSC	PMSC
SEP Protéines	0.49	0.41	0.39
Amélioration (%)	-	16	20
SEP Mat. grasses	1.21	1	0.76
Amélioration (%)	-	17	37
SEP Eau	0.98	0.77	0.62
Amélioration (%)	-	21	37
Saucisses Fumées			
	Spectres bruts log(1/R)	MSC	PMSC
SEP Protéines	0.3	0.29	0.32
Amélioration (%)	-	3	-7
SEP Mat. grasses	0.9	0.89	0.85
Amélioration (%)	-	1	6
SEP Eau	0.55	0.51	0.51
Amélioration (%)	-	7	7

Tableau 7 : Influence des méthodes MSC et PMSC sur l'analyse quantitative

3.4.6 La normalisation par la méthode « SNV »

La méthode « SNV » (*Standard Normal Variate*) est une transformation de normalisation par la déviation standard^(140,154). Cette technique est appliquée à chaque spectre pris séparément sans référence à l'ensemble des échantillons de calibration. Contrairement à la méthode « MSC » elle est indépendante de l'ensemble des spectres utilisés pour l'analyse quantitative⁽¹⁴⁷⁾.

Considérons un spectre d'absorbance S constitué de w longueurs d'onde. L'absorbance à une longueur d'onde i est notée a_i . On calcule la moyenne des absorbances de S notée \bar{a}

$$(éq. 49) \bar{a} = \frac{\sum_{i=1}^w a_i}{w}$$

Celle-ci est retranchée à l'absorbance sur chaque longueur d'onde i . C'est l'étape de centrage des données. On norme finalement par la déviation standard σ des absorbances du spectre S pour obtenir l'absorbance corrigée à la longueur d'onde i notée SNV_i :

$$(éq. 50) SNV_i = (a_i - \bar{a}) / \sqrt{\frac{\sum_{i=1}^w (a_i - \bar{a})^2}{w - 1}} \quad \text{Pour } i = 1 \dots w$$

Cette transformation a permis d'améliorer les résultats de nombreuses analyses quantitatives^(155,156). La Figure 32 permet d'apprécier l'effet de la transformation SNV sur les spectres NIR de fourrage⁽¹⁴⁷⁾. Les variations de signal dues aux phénomènes physiques diminuent laissant apparaître les informations chimiques.

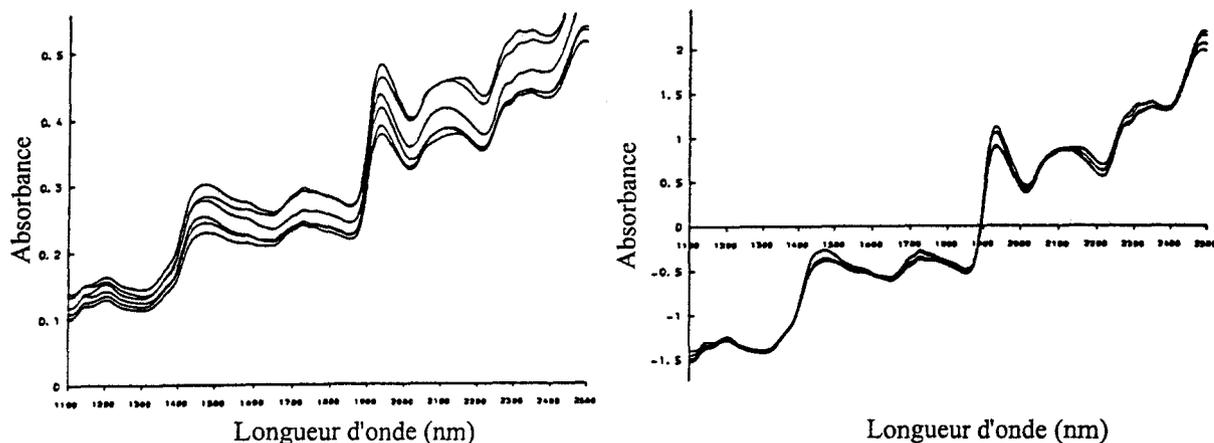


Figure 32 : Effet de la SNV sur les spectres NIR

Sur les spectres NIR, il est commun d'observer des lignes de base incurvées. Cette déformation nuit souvent à l'analyse quantitative. Afin de redresser les lignes de base, BARNES *et al.* ont proposé une modification de la SNV appelée « SNVD » (SNV De-trending)⁽¹⁵⁷⁾. Ainsi pour chaque spectre la correction SNV est toujours appliquée mais la ligne de base est redressée par un polynôme du second degré. L'équation du polynôme est obtenue par minimisation des moindres carrés.

Il existe une version généralisée de la SNV^(158,159). La seule différence avec la méthode classique est l'utilisation de tous les spectres du lot de calibration (s au total). Dans ce cas l'absorbance à une longueur d'onde i de l'échantillon j est notée a_{ij} . La moyenne des

absorbances des spectres à la longueur d'onde i est notée \bar{a}_i , et la déviation standard des spectres à la longueur d'onde i notée σ_i .

$$(éq. 51) \quad \bar{a}_i = \frac{\sum_{j=1}^s a_{ij}}{s} \quad \text{et} \quad \sigma_i = \sqrt{\frac{\sum_{j=1}^s (a_{ij} - \bar{a}_i)^2}{s-1}}$$

L'absorbance du spectre corrigé j à la longueur d'onde i notée $CENT_{ij}$ est donnée par :

$$(éq. 52) \quad CENT_{ij} = (a_{ij} - \bar{a}_i) / \sigma_i$$

Cette normalisation est quelquefois appliquée pour l'analyse en composantes principales. Elle permet d'extraire plus facilement les informations spectrales. Pour l'analyse quantitative, elle reste d'un intérêt limité.

3.4.7 La « normalisation simple »

Cette transformation est simple puisqu'elle fait appel à la norme euclidienne d'un vecteur⁽¹⁶⁰⁾. Elle est appliquée sur chaque spectre pris séparément. Considérons un spectre d'absorbance constitué de w longueurs d'onde. L'absorbance à la longueur d'onde i est notée a_i . On calcule tout d'abord la norme du spectre notée $Norm$:

$$(éq. 53) \quad Norm = \sqrt{\sum_{i=1}^w a_i^2}$$

L'absorbance corrigée à la longueur d'onde i notée $EUCL_i$ est donnée par :

$$(éq. 54) \quad EUCL_i = a_i / Norm$$

Cette correction est souvent utilisée pour diminuer l'effet de taille des particules sur les spectres NIR⁽¹⁶¹⁾. Elle est aussi largement utilisée par les réseaux de neurones pour uniformiser les données d'apprentissage⁽¹⁶²⁾.

3.4.8 La normalisation par la méthode « MEA »

Cette méthode que nous appellerons « MEA » a été proposée par ROBERT *et al.*⁽¹⁶³⁾. Elle utilise des opérations mathématiques simples. Elle reste rarement utilisée en analyse quantitative car les améliorations sont faibles. Considérons a_j l'absorbance de l'échantillon j

à la longueur d'onde i , s le nombre total de d'échantillons et w le nombre de longueurs d'onde. L'absorbance corrigée à la longueur d'onde i du spectre j notée $MEA_{i,j}$ est donné par :

$$(éq. 55) \quad MEA_{i,j} = a_{i,j} - \frac{1}{w} \left(\sum_{i=1}^w a_{i,j} \right) - \frac{1}{s} \left(\sum_{j=1}^s a_{i,j} \right) + \frac{1}{sw} \left(\sum_{i=1}^w \sum_{j=1}^s a_{i,j} \right)$$

Pour chaque spectre on soustrait la moyenne de ses absorbances et le spectre moyen de l'ensemble des échantillons.

3.5 Les algorithmes de régression

3.5.1 Introduction

Il existe de nombreux algorithmes de régression pour la construction d'une équation de calibration à partir de données spectrales. Nous présenterons dans cette partie les principales familles telles que la régression linéaire simple, les régressions multivariées et les techniques basées sur la condensation des données. Après le développement théorique de ces différents algorithmes, nous mettrons en évidence leurs avantages et leurs inconvénients. Une comparaison des aptitudes à la résolution d'analyses quantitatives sera aussi effectuée.

3.5.2 La régression linéaire simple

Lorsque l'on veut développer une analyse quantitative à partir de données spectrales, on part d'un modèle qui permet de relier l'absorbance à la concentration. De nombreux auteurs ont tenté de relier ces deux paramètres par une forme quadratique ou d'autres fonctions plus compliquées⁽¹⁶⁴⁾. Mais le modèle le plus utilisé est simplement linéaire. Il est connu sous le nom de loi de Beer-Lambert. A une longueur d'onde donnée i , l'absorbance a_i d'un constituant est directement proportionnelle à sa concentration molaire c (éq. 56).

$$(éq. 56) \quad a_i = \varepsilon.l.c = k_i.c$$

ε est le coefficient d'extinction molaire à la longueur d'onde considérée et l le chemin optique utilisé pour l'analyse spectrale. Prenons l'exemple du dosage de deux constituants A et B dans un mélange (Figure 33). Sur le plan spectral, on voit que ces deux constituants A et B n'interfèrent pas. L'absorbance à la longueur d'onde λ_1 est uniquement due au constituant A. L'absorbance à la longueur d'onde λ_2 est uniquement due au constituant B. L'équation

précédente peut donc être appliquée indépendamment sur chaque longueur d'onde λ_1 et λ_2 pour le dosage de A et B (éq. 57).

$$(éq. 57) \quad a_{\lambda_1} = k_{a,\lambda_1} \cdot c_a \quad \text{et} \quad a_{\lambda_2} = k_{b,\lambda_2} \cdot c_b$$

La prédiction de la concentration en A notée c_a pour des échantillons inconnus passe par la détermination de k_{a,λ_1} . La remarque est identique pour c_b et k_{b,λ_2} .

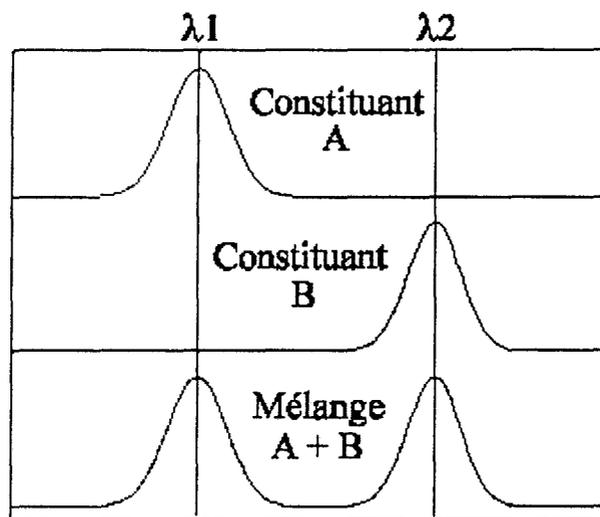


Figure 33 : Dosage de deux constituants sans interférence spectrale.

A ce stade, il suffit d'effectuer une régression linéaire simple par les moindres carrés à partir des données de calibration (spectres et concentrations de référence). La constante k est obtenue par minimisation de la différence entre l'absorbance réelle des échantillons et l'absorbance prédite par le modèle. Cette méthode très simple est malheureusement inutilisable sur des spectres de mélanges où des bandes d'absorption se superposent. En effet, considérons cette fois-ci le cas de deux constituants A et B dont les spectres d'absorption se recouvrent (Figure 34). On peut avec la procédure précédente déterminer des valeurs pour k_{a,λ_1} et k_{b,λ_2} mais les modèles ainsi constitués donneront des erreurs de prédiction pour A et B très importantes. En effet, dans l'absorbance a_{λ_1} , nous n'aurons pas pris en compte la contribution du composé B, de même pour a_{λ_2} et le composé A. Or la deuxième partie de la loi de Beer-Lambert énonce que les absorbances de plusieurs constituants à une longueur d'onde sont additives. Pour le système en présence, nous pouvons donc écrire :

$$(éq. 58) \quad a_{\lambda_1} = k_{a,\lambda_1} \cdot c_a + k_{b,\lambda_1} \cdot c_b$$

$$(éq. 59) \quad a_{\lambda_2} = k_{a,\lambda_2} \cdot c_a + k_{b,\lambda_2} \cdot c_b$$

C'est la généralisation de ce nouveau système d'équations qui a entraîné le développement des techniques multilinéaires que nous développerons par la suite.

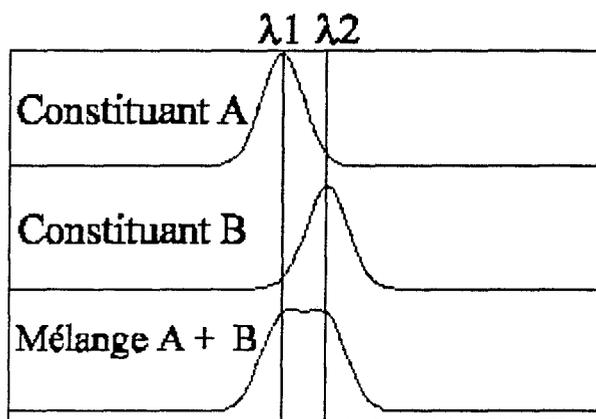


Figure 34 : Dosage de deux constituants avec interférences spectrales.

En conclusion, il est impossible d'utiliser la régression linéaire simple sur des données spectroscopiques car les recouvrements spectraux sont omniprésents. L'étude d'échantillons biologiques complexes en spectroscopie NIR conforte cette idée.

3.5.3 Les régressions multilinéaires

3.5.3.1 La méthode CLS

Le nom CLS^(165,166,167) (*Classical Least Squares*) vient de l'application de la régression multilinéaire à la forme classique de la loi de Beer-Lambert :

$$\text{(éq. 60) } a = k \cdot c$$

Elle est aussi appelée méthode de la matrice **K**. Nous avons vu précédemment que pour un spectre de mélange, la loi de Beer-Lambert donne le système d'équations suivant :

$$a_1 = k_{1,1} \cdot c_1 + \dots + k_{1,j} \cdot c_j + \dots + k_{1,n} \cdot c_n$$

.....

$$a_i = k_{i,1} \cdot c_1 + \dots + k_{i,j} \cdot c_j + \dots + k_{i,n} \cdot c_n$$

.....

$$a_w = k_{w,1} \cdot c_1 + \dots + k_{w,j} \cdot c_j + \dots + k_{w,n} \cdot c_n$$

où a_i est l'absorbance à la longueur d'onde i , k_{ij} la constante d'absorption du constituant j à la longueur d'onde i , et c_j la concentration de ce même constituant. Nous avons au total w

longueurs d'onde et n constituants. Ce modèle peut être appliqué pour un nombre de longueurs d'onde au moins égal au nombre de constituants dans les mélanges analysés. Cette méthode peut d'ailleurs être appliquée sur des spectres entiers. Si on adopte une écriture matricielle d'algèbre linéaire, le système d'équations précédent devient :

$$\text{(éq. 61)} \quad \mathbf{A} = \mathbf{K} \cdot \mathbf{C}$$

où \mathbf{A} est la matrice des données spectrales de dimension $(w \times s)$ avec s le nombre de spectres des échantillons de calibration suivant les colonnes, \mathbf{K} la matrice des constantes d'absorption de dimension $(w \times n)$ et \mathbf{c} la matrice des concentrations de dimension $(n \times s)$. La calibration se traduit par la recherche de la matrice \mathbf{K} . On multiplie donc l'équation précédente par la matrice transposée de \mathbf{C} notée \mathbf{C}^T :

$$\text{(éq. 62)} \quad \mathbf{A} \cdot \mathbf{C}^T = \mathbf{K} \cdot \mathbf{C} \cdot \mathbf{C}^T$$

Pour obtenir \mathbf{K} , il suffit de multiplier à nouveau l'expression par l'inverse de $\mathbf{C} \cdot \mathbf{C}^T$ notée $(\mathbf{C} \cdot \mathbf{C}^T)^{-1}$:

$$\text{(éq. 63)} \quad \mathbf{K} = \mathbf{A} \cdot \mathbf{C}^T \cdot (\mathbf{C} \cdot \mathbf{C}^T)^{-1}$$

L'inverse de la matrice $\mathbf{C} \cdot \mathbf{C}^T$ ne peut être calculée que si elle n'est pas singulière. En d'autres termes, il doit être impossible d'exprimer une colonne ou une ligne de \mathbf{C} par une combinaison linéaire d'autres lignes ou colonnes de cette même matrice. Ce problème survient par exemple lorsque les échantillons de calibration sont préparés par dilutions successives ou lorsque la somme des concentrations des constituants est égale à une valeur constante. Cette méthode a l'avantage d'utiliser de larges zones spectrales ou même le spectre entier. De cette façon, on obtient un effet de moyenne sur les erreurs de prédiction des constituants à doser. Mais le principal inconvénient de cette méthode de régression est la nécessité de connaître tous les constituants des mélanges pour la calibration. En effet, l'absorbance à une longueur d'onde est donnée par la somme des concentrations des constituants multipliées par les constantes d'absorption associées. Si la concentration d'un constituant est omise, le calcul de \mathbf{K} est inexact et la valeur prédite compromise. Ainsi pour des mélanges complexes (échantillons naturels) ou des échantillons contenant des contaminants inconnus à la calibration, le modèle donnera des erreurs de prédiction très importantes.

Lorsque la calibration est terminée, nous pouvons utiliser la matrice \mathbf{K} pour déterminer les concentrations d'échantillons inconnus (matrice $\mathbf{C}_{\text{prédiction}}$) à partir de leurs spectres (matrice $\mathbf{A}_{\text{prédiction}}$). Nous avons :

$$\mathbf{A}_{\text{prédiction}} = \mathbf{K} \cdot \mathbf{C}_{\text{prédiction}}$$

$$\text{donc } K^T \cdot A_{\text{prédiction}} = K^T \cdot K \cdot C_{\text{prédiction}}$$

$$\text{donc } C_{\text{prédiction}} = (K^T \cdot K)^{-1} \cdot K^T \cdot A_{\text{prédiction}}$$

En résumé :

Méthode CLS

Avantages :

- Les calculs sont rapides.
- La calibration ne nécessite pas une sélection de longueurs d'onde. Le spectre entier peut même être utilisé. Seul le nombre de longueurs d'onde utilisées doit être supérieur au nombre de constituant à doser.
- Un grand nombre de longueurs d'onde permet d'avoir un effet de moyenne sur les erreurs de prédiction. Le modèle construit est moins sensible au bruit des spectres.

Inconvénients :

- Nécessite la connaissance de tout les constituants des échantillons pour la calibration.
- Grande sensibilité aux variations de ligne de base puisque l'absorbance à une longueur d'onde est considérée comme provenant uniquement des constituants.

3.5.3.1 La méthode ILS

Sur des échantillons complexes tels que les produits naturels, il est très inhabituel et même impossible de connaître les concentrations de tout les constituants. D'ailleurs dans la plupart des cas seules les concentrations de certaines espèces nous intéressent. La méthode de régression ILS^(127,168) (*Inverse Least Squares*) tire son nom de l'application de la régression multilinéaire à la forme inverse de la loi de Beer-Lambert :

$$\text{(éq. 64) } c = p \cdot a$$

Elle est aussi appelée méthode de la matrice P ou plus communément connue sous le nom de méthode MLR (*Multilinear Regression*)⁽¹⁶⁹⁾. Par analogie à la méthode CLS, nous pouvons définir un système d'équations pour un spectre de mélange :

$$c_1 = p_{1,1} \cdot a_1 + \dots + p_{1,i} \cdot a_i + \dots + p_{w,1} \cdot a_w$$

$$\dots$$

$$c_j = p_{1,j} \cdot a_1 + \dots + p_{i,j} \cdot a_i + \dots + p_{w,j} \cdot a_w$$

$$\dots$$

$$c_n = p_{1,n} \cdot a_1 + \dots + p_{i,n} \cdot a_i + \dots + p_{w,n} \cdot a_w$$

où a_i est l'absorbance à la longueur d'onde i , $p_{i,j}$ la constante d'absorption du constituant j à la longueur d'onde i , et c_j la concentration de ce même constituant. Comme pour CLS, la résolution de ce système d'équations impose un nombre de longueurs d'onde plus grand que le nombre de constituants à doser. Par ce système d'équations, on remarque que la connaissance de toutes les concentrations du mélange n'est pas nécessaire. Si des constituants sont omis, les éléments $p_{i,j}$ peuvent tous être calculés. Par contre, le premier inconvénient d'ILS est la sélection de longueurs d'onde caractéristiques des éléments à doser sur le spectre. Cette recherche n'est pas toujours évidente même si des algorithmes de sélection sont régulièrement développés⁽¹⁷⁰⁾.

Avec une notation algébrique, le système d'équations précédent devient :

$$\text{(éq. 65)} \quad C = P \cdot A$$

Le second inconvénient de cette méthode est le nombre de longueurs d'onde sélectionnées qui doit être inférieur au nombre d'échantillons utilisés pour la calibration. De cette façon, la méthode CLS, considérée comme une régression de spectres entiers, est souvent mise en opposition avec la méthode ILS. En effet, une analyse quantitative avec ILS sur des spectres entiers contenant 500 longueurs d'onde demanderait au moins 500 mélanges pour la calibration. En théorie, il est possible d'augmenter le nombre d'échantillons de calibration afin de sélectionner plus de longueurs d'onde mais un nouveau problème survient. Les absorbances d'un spectre augmentent et diminuent ensemble en fonction des concentrations des échantillons. Cet effet, connu sous le nom de colinéarité, donne une prédiction de concentrations instable. Le deuxième inconvénient est un phénomène d'overfitting qui apparaît lorsque de nombreuses longueurs d'onde sont utilisées. En général, on commence par sélectionner quelques longueurs d'onde, puis d'autres sont ajoutées régulièrement afin d'améliorer la calibration. Néanmoins, à partir d'un certain nombre, la capacité prédictive du modèle commence à se détériorer. En effet, plus la sélection progresse, moins on a de chances de trouver des longueurs d'onde bien corrélées aux concentrations du composé à doser. Lorsque trop d'informations sont introduites dans le modèle, la calibration n'utilise plus des longueurs d'onde significatives pour le constituant à doser mais du bruit. L'effet de moyenne des erreurs obtenu par la sélection de nombreuses longueurs d'onde pour la régression CLS, est perdu pour la régression ILS. L'étape critique d'une régression ILS est donc la sélection des longueurs d'onde. Il y a un compromis à faire entre sélectionner suffisamment de

longueurs d'onde pour obtenir une bonne calibration et ne pas trop en sélectionner pour éviter les colinéarités et le bruit. La variable dépendante (la concentration dans notre cas) est calculée à partir d'un système multiple de variables indépendantes (les absorbances aux longueurs d'onde sélectionnées).

Si on revient à la dernière équation, la calibration pour ILS se traduit par la recherche de la matrice P (éq. 65). On multiplie donc cette équation par la matrice transposée de A notée A^T :

$$(éq. 66) \quad C \cdot A^T = P \cdot A \cdot A^T$$

Pour obtenir P , il suffit de multiplier à nouveau l'expression par l'inverse de $A \cdot A^T$ noté $(A \cdot A^T)^{-1}$:

$$(éq. 67) \quad P = C \cdot A^T \cdot (A \cdot A^T)^{-1}$$

L'inverse de la matrice $A \cdot A^T$ ne peut être calculée que si elle n'est pas singulière. Cette condition est moins restrictive comparée à la méthode CLS car les spectres utilisés sont toujours bruités. Néanmoins, il est conseillé de ne pas introduire des spectres trop ressemblants dans une calibration ILS. L'inverse de la matrice $A \cdot A^T$ pourra être calculé mais la nature colinéaire des données introduira des erreurs importantes. Lorsque la calibration est terminée, nous pouvons utiliser directement la matrice P pour déterminer les concentrations d'échantillons inconnus (matrice $C_{\text{prédiction}}$) à partir de leurs spectres (matrice $A_{\text{prédiction}}$). Nous avons :

$$(éq. 68) \quad C_{\text{prédiction}} = P \cdot A_{\text{prédiction}}$$

En résumé :

Méthode ILS

Avantages :

- Les calculs sont rapides.
- Le modèle autorise les calibrations sur des produits complexes puisque seules des connaissances sur les produits à doser sont requises.

Inconvénients :

- La sélection de longueurs d'onde caractéristiques des constituants à doser peut être longue et difficile. Il faut éviter les longueurs d'onde colinéaires.
- En général, un grand nombre d'échantillons est nécessaire à l'établissement d'une bonne calibration.

3.5.4 Les méthodes de condensation des données

3.5.4.1 Introduction

Les méthodes de condensation des données sont des techniques d'extraction de facteurs. Un espace de facteurs est un système particulier de coordonnées qui offre des avantages de calcul. Pour travailler dans un espace de facteurs, il suffit de transformer les données du système de coordonnées d'origine vers un nouveau système. On effectue donc un changement de base vectorielle. L'idée de la condensation des données est de caractériser des données complexes par un nombre restreint de variables. Pour comprendre facilement le problème, prenons l'exemple de six échantillons caractérisés par trois variables V1, V2 et V3 pouvant être l'absorbance à trois longueurs d'onde. Dans une représentation à trois dimensions chaque échantillon est un point (Figure 35). Dans cet exemple volontairement trivial, nous voyons que les échantillons sont alignés sur une droite de direction P1.

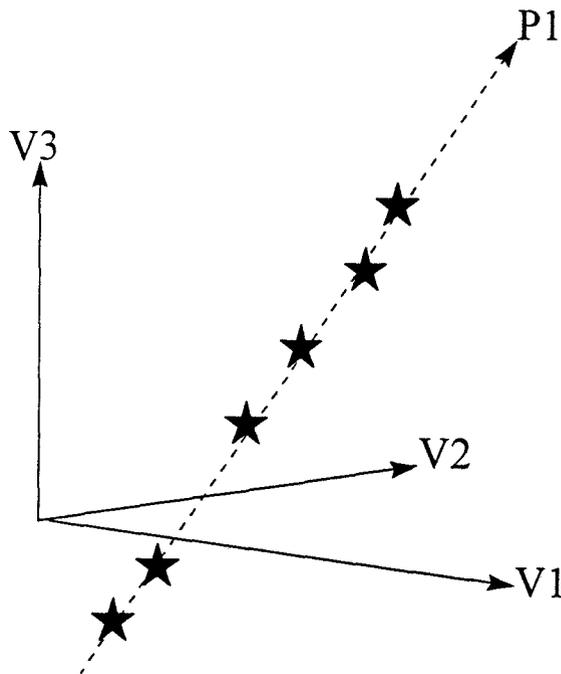


Figure 35 : La condensation de données.

Le facteur P1 traduit de façon complète la variance de tous les échantillons. Leurs projections sur ce vecteur donneront une seule variable dans le nouvel espace de travail. La condensation des données fait passer le nombre de variables de 3 à 1 en conservant l'information significative. Considérons maintenant des échantillons caractérisés par un grand nombre de variables comme un spectre de 500 longueurs d'onde. De la même façon, ils sont assimilés à

un point dans un espace à 500 dimensions. Au même titre que P1, nous pouvons extraire à partir de ces points représentant les spectres, des facteurs qui expriment toute la variance des échantillons. Pour les applications en spectroscopie, il est commun d'extraire un maximum de 15 facteurs. Ainsi dans cette nouvelle base vectorielle, chaque échantillon est caractérisé par 15 valeurs appelées *scores* qui sont les projections suivant ces nouveaux axes.

Cinq exemples d'espace de facteurs sont donnés ci-dessous :

- Les séries de Fourier.
- L'orthogonalisation de Gram-Schmidt.
- Les séries de Taylor.
- Les vecteurs propres dans l'extraction de composantes principales (PCA ou PCR).
- Les facteurs latents dans la méthode PLS.

Ce nouveau concept a supplanté les méthodes classiques CLS et ILS. Il y a de nombreuses raisons pour lesquelles on préfère à l'heure actuelle travailler dans un espace de facteurs plutôt que sur les coordonnées physiques naturelles (c'est-à-dire sur les spectres dans notre cas) :

- Le conditionnement numérique : En projetant les données d'un espace physique sur un espace de facteurs appropriés, il est possible d'éliminer les problèmes causés par des données hautement colinéaires. C'est le cas de calibrations effectuées à partir de spectres très similaires. Cette méthode réduit aussi les erreurs d'arrondis dans les calculs et permet même de développer des calibrations impossibles dans l'espace des coordonnées d'origine.
- Le filtrage numérique : En travaillant dans un espace de facteurs appropriés, on effectue un filtrage des données. Le bruit qui n'est pas corrélé aux caractéristiques des constituants est supprimé.
- Une nouvelle perception des données : La projection des données dans un nouvel espace de facteurs permet de révéler une structure qui était cachée dans l'espace d'origine. Elle peut quelques fois nous indiquer le nombre de constituants réellement présents dans les échantillons, et nous dire s'ils appartiennent ou non à un ensemble de propriétés physico-chimiques similaires.

Dans les parties qui vont suivre, seront présentées deux méthodes de régression qui utilisent la condensation des données par la recherche d'espaces de facteurs.

3.5.4.2 La régression PCR

La méthode de régression PCR (*Principal Component Regression*) combine en deux étapes la condensation des données par une analyse en composantes principales (PCA) et la régression inverse pour créer une analyse quantitative de mélanges complexes⁽¹⁷¹⁾. La phase PCA va donc rechercher à partir de la matrice des données spectrales A (espace d'origine), les directions de variances maximales qui créeront la nouvelle base vectorielle pour les spectres. Ces directions sont appelées facteurs, composantes principales ou encore loadings. L'extraction de ces facteurs est progressive puisqu'on recherche d'abord la direction qui exprime le maximum de variance. La seconde direction exprimera la variance résiduelle des données et devra être orthogonale à la première. Les composantes successives expriment toujours la variance résiduelle de l'extraction précédente et sont perpendiculaires deux à deux. On démontre que cela revient à dire que les composantes principales sont les vecteurs propres de la matrice des données spectrales A ⁽⁹⁹⁾. Mais la décomposition en valeurs propres et vecteurs propres nécessite une matrice carrée ce qui est rarement le cas pour A . Le calcul s'effectue donc sur le produit $A \cdot A^T$ que l'on appelle parfois matrice de dispersion ou matrice de covariance. Cette astuce permet de convertir une matrice rectangulaire en une matrice carrée sans perdre d'informations. L'extraction des vecteurs propres peut se faire par de nombreuses méthodes mathématiques. La plus connue est la décomposition en valeurs singulières (SVD)⁽¹⁷²⁾. Une deuxième, plus ancienne est basée sur l'algorithme NIPALS^(172,173). Les composantes principales recherchées sont donc les vecteurs propres possédant les valeurs propres associées les plus importantes. On conçoit facilement qu'un grand nombre de composantes principales puisse ainsi être extrait. Le but de l'analyse en composantes principales n'est pas d'extraire un maximum de facteurs, mais de conserver les plus significatifs pour la régression qui suivra. C'est à cet instant que le filtrage des données s'effectue. Des techniques statistiques telles que la méthode des validations croisées (*cross-validation*) permettent de cerner le nombre optimal de facteurs à utiliser^(174,175,176).

Considérons maintenant que l'analyse en composantes principales de $A(s \times w)$ (s spectres à w longueurs d'onde) a permis d'extraire f vecteurs loading significatifs contenus dans la matrice $V(f \times w)$. Nous pouvons obtenir la projection de la matrice des données spectrales sur la nouvelle base vectorielle de dimension f . Ces projections sont contenues dans la matrice des scores notée $T(s \times f)$ (Figure 36).

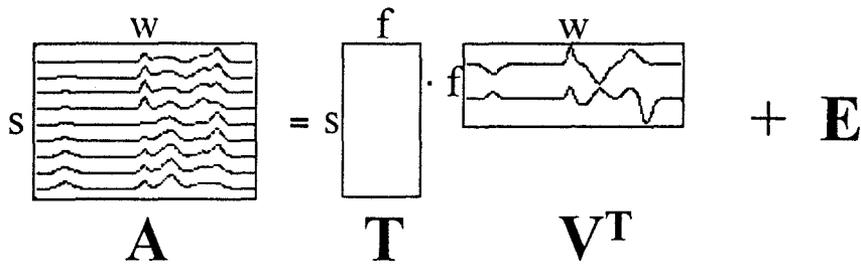


Figure 36 : Décomposition en scores et loadings des données spectrales.

Ainsi nous avons :

$$(éq. 69) \quad \mathbf{A} = \mathbf{T} \cdot \mathbf{V}^T$$

$$\text{donc } \mathbf{A} \cdot \mathbf{V} = \mathbf{T} \cdot \mathbf{V}^T \cdot \mathbf{V}$$

or la matrice \mathbf{V} est orthogonale donc $\mathbf{V}^T \cdot \mathbf{V} = \mathbf{V} \cdot \mathbf{V}^T = \mathbf{I}$

$$\text{d'où } (éq. 70) \quad \mathbf{T} = \mathbf{A} \cdot \mathbf{V}$$

Pour la calibration, nous effectuons finalement une régression inverse entre \mathbf{T} qui contient les informations spectrales des échantillons dans la nouvelle base vectorielle et la matrice des concentrations \mathbf{c} . La calibration se résume par la recherche de la matrice \mathbf{P} telle que :

$$(éq. 71) \quad \mathbf{c} = \mathbf{P} \cdot \mathbf{T}$$

$$\text{donc } (éq. 72) \quad \mathbf{c} \cdot \mathbf{T}^T = \mathbf{P} \cdot \mathbf{T} \cdot \mathbf{T}^T$$

$$\text{d'où } (éq. 73) \quad \mathbf{P} = \mathbf{c} \cdot \mathbf{T}^T \cdot (\mathbf{T} \cdot \mathbf{T}^T)^{-1}$$

Lorsque la calibration est terminée, nous pouvons utiliser la matrice \mathbf{P} pour déterminer les concentrations d'échantillons inconnus (matrice $\mathbf{c}_{\text{prédiction}}$) à partir de leurs spectres (matrice $\mathbf{A}_{\text{prédiction}}$). On projette tout d'abord la matrice des données spectrales $\mathbf{A}_{\text{prédiction}}$ sur l'espace vectorielle des composantes principales \mathbf{V} :

$$(éq. 74) \quad \mathbf{T}_{\text{prédiction}} = \mathbf{A}_{\text{prédiction}} \cdot \mathbf{V}$$

$$\text{puis } (éq. 75) \quad \mathbf{c}_{\text{prédiction}} = \mathbf{P} \cdot \mathbf{T}_{\text{prédiction}}$$

La régression n'est malheureusement pas sans inconvénient. Il faut en fait rappeler qu'elle se construit en deux étapes : la recherche des facteurs loading et des scores par une analyse en composantes principales puis la régression de ces projections avec les concentrations par une régression inverse. Or ce type de régression construit des calibrations de qualité si les variables sélectionnées (les longueurs d'onde) sont significatives des propriétés (concentrations) que l'on veut modéliser. L'analyse en composantes principales puis le calcul des scores se font de manière complètement indépendante par rapport aux concentrations. Les directions sélectionnées représentent les variations les plus larges contenues dans les données spectrales. Elles sont présumées rendre compte des variations de concentrations dans les

échantillons mais rien ne le garantit. En fait, la régression PCR introduit souvent plus de facteurs qu'il n'est nécessaire, car certains d'entre eux ne sont pas corrélés aux constituants qui nous intéressent. Comme pour la régression inverse, la qualité prédictive des régressions PCR souffre de la colinéarité des concentrations.

En résumé :

Méthode PCR

Avantages :

- Le conditionnement numérique : l'analyse en composantes principales supprime les colinéarités spectrales.
- Une nouvelle perception des données : l'analyse en composantes principales permet quelques fois de révéler des structures cachées de données spectrales.
- Le filtrage numérique : diminution du bruit et conservation de l'information significative.
- Ne demande pas de sélection de longueurs d'onde. On utilise souvent les spectres entiers ou de larges zones spectrales.
- Un grand nombre de longueurs d'onde donne un effet de moyenne sur les concentrations prédites.
- Comme pour la régression ILS, la calibration ne porte que sur les constituants d'intérêt rendant possible le dosage de produits complexes.
- Peut quelques fois être utilisée pour prédire des échantillons contenant un constituant absent (contaminant) dans les mélanges de calibration.

Inconvénients :

- Il n'y a aucune garantie sur la corrélation des composantes principales extraites et les concentrations des constituants à prédire.

3.5.4.2 La régression PLS

L'algorithme PLS a été développé dans les années 70 pour résoudre des problèmes en sciences économiques et sociales⁽¹³⁵⁾. Ses premières applications à l'analyse quantitative remontent aux années 80^(177,178). Comme la méthode PCR, la régression PLS extrait de ses données spectrales des facteurs constituant une nouvelle base vectorielle de travail. La particularité de cet algorithme est l'extraction de facteurs à partir de la covariance

spectres/concentrations. La condensation des données ne se fait plus suivant des directions de variance maximale mais sur les directions les plus pertinentes en terme de prédiction des concentrations⁽¹⁷⁹⁾. De cette façon, on efface l'inconvénient majeur de PCR puisque les données spectroscopiques A et les concentrations c sont exploitées de la même manière dans la calibration. Toute composante spectrale extraite des données spectrales correspond en même temps à une variation sur l'espace des concentrations. Lorsque l'on veut doser n constituants d'un mélange avec PLS, on effectue une régression sur chaque constituant pris séparément (algorithme PLS1) ou sur toutes les concentrations en même temps (PLS2). Néanmoins, la régression PLS1 comparée à la régression PLS2 donne très souvent de meilleurs résultats⁽¹³⁵⁾. C'est donc PLS1, plus communément utilisée, qui sera étudiée sous le nom de PLS.

Il existe deux façons d'effectuer la régression PLS : la technique originale appelée PLS orthogonale qui a été développée par WOLD⁽¹⁷⁷⁾ et la PLS non-orthogonale développée plus tard par MARTENS⁽¹³⁵⁾. Nous développerons la première méthode qui est la plus simple et la plus répandue.

La calibration :

Les données spectrales sont contenues dans la matrice A alors que les concentrations sont dans la matrice c . Nous allons tout d'abord choisir un nombre maximum de facteurs loading à extraire noté n_{\max} (souvent 15). La procédure d'extraction sera effectuée par les étapes C1 à C6 pour chaque facteur identifié par l'indice i ($i=1 \dots n_{\max}$).

Étape C1 : On recherche le facteur loading w_i (*loading weight*) qui donne la covariance maximum entre A et c mais avec la contrainte $w_i \cdot w_i^T = 1$. Ce qui revient à dire que w_i est unitaire et qu'il rend le produit $w_i^T \cdot A^T \cdot c$ maximum.

$$\text{(éq. 76)} \quad A = c \cdot w_i^T$$

une résolution par les moindres carrés donne $w_i = k \cdot A^T \cdot c$

où $k = \sqrt{c^T \cdot A \cdot A^T \cdot c}$ est une constante qui permet d'avoir w_i unitaire.

Étape C2 : On projette ensuite les données spectrales A sur ce facteur w_i pour obtenir les scores t_i .

$$(éq. 77) \quad \mathbf{A} = \mathbf{t}_i \cdot \mathbf{w}_i^T$$

$$\text{donc (éq. 78) } \quad \mathbf{t}_i = \mathbf{A} \cdot \mathbf{w}_i \quad \text{car } \mathbf{w}_i^T \cdot \mathbf{w}_i = 1$$

Étape C3 : On extrait ensuite le facteur loading spectral noté \mathbf{p}_i à partir de \mathbf{A} et \mathbf{t}_i :

$$(éq. 79) \quad \mathbf{A} = \mathbf{t}_i \cdot \mathbf{p}_i^T$$

une résolution par les moindres carrés donne

$$(éq. 80) \quad \mathbf{p}_i = \mathbf{A}^T \cdot \mathbf{t}_i \cdot (\mathbf{t}_i^T \cdot \mathbf{t}_i)^{-1}$$

Étape C4 : De la même façon, on extrait le facteur loading concentration noté \mathbf{q}_i à partir de \mathbf{c} et \mathbf{t}_i :

$$\mathbf{c} = \mathbf{t}_i \cdot \mathbf{q}_i$$

une résolution par les moindres carrés donne

$$(éq. 81) \quad \mathbf{q}_i = \mathbf{c}^T \cdot \mathbf{t}_i \cdot (\mathbf{t}_i^T \cdot \mathbf{t}_i)^{-1}$$

Étape C5 : On soustrait finalement la contribution de la composante extraite \mathbf{p}_i de la matrice de données spectrales \mathbf{A} pour obtenir une nouvelle matrice notée $\mathbf{A}_{\text{résiduelle}}$:

$$(éq. 82) \quad \mathbf{A}_{\text{résiduelle}} = \mathbf{A} - \mathbf{t}_i \cdot \mathbf{p}_i^T$$

Étape C6 : De même, on soustrait la contribution de la composante extraite \mathbf{q}_i de la matrice de concentration \mathbf{c} pour obtenir une nouvelle matrice notée $\mathbf{c}_{\text{résiduelle}}$:

$$(éq. 83) \quad \mathbf{c}_{\text{résiduelle}} = \mathbf{c} - \mathbf{t}_i \cdot \mathbf{q}_i$$

A la fin de cette étape, on réitère le calcul en utilisant $\mathbf{A}_{\text{résiduelle}}$ et $\mathbf{c}_{\text{résiduelle}}$ comme données de départ dans l'étape C1. Une autre composante \mathbf{w}_i peut être extraite perpendiculaire à \mathbf{w}_{i-1} . Le nom de PLS orthogonale vient d'ailleurs du fait que les \mathbf{w}_i sont deux à deux orthogonaux (même chose pour les \mathbf{t}_i). L'extraction est terminée lorsque $i = n_{\text{max}}$. A ce stade, comme pour la PCR, il faut choisir le nombre optimal n_{opt} ($1 < n_{\text{opt}} < n_{\text{max}}$) de facteurs loading significatifs. Si n_{opt} est trop faible, des informations significatives ne seront pas prises en compte par le modèle (*underfitting*). D'autre part, si n_{opt} est trop grand, le bruit des données est introduit dans le modèle et détériore sa qualité prédictive (*overfitting*). Toute l'efficacité de l'algorithme PLS repose sur le choix judicieux de ce paramètre. Des techniques statistiques nous aident dans cette recherche^(180,181). Lorsque le choix de n_{opt} est arrêté, on crée la matrice \mathbf{W} qui contient les vecteurs $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{n_{\text{opt}}}$ suivant ses colonnes et la matrice \mathbf{P} qui contient aussi les vecteurs $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{n_{\text{opt}}}$ suivant ses colonnes. Nous avons finalement à la fin de la calibration :

$$(éq. 84) \quad \mathbf{c} = \mathbf{A} \cdot \mathbf{H} \quad \text{avec } \mathbf{H} = \mathbf{W} \cdot (\mathbf{P}^T \cdot \mathbf{W})^{-1} \cdot \mathbf{q}$$

La prédiction :

Il y a deux façons de prédire les concentrations $c_{\text{prédiction}}$ d'échantillons inconnus à partir de leurs spectres contenus dans $A_{\text{prédiction}}$. La méthode la plus rapide utilise directement l'équation précédente :

$$\text{(éq. 85)} \quad c_{\text{prédiction}} = A_{\text{prédiction}} \cdot W \cdot (P^T \cdot W)^{-1} \cdot q$$

La seconde méthode se fait étape par étape comme pour la calibration mais sur un spectre à la fois. Soit $a_{\text{prédiction}}$ le spectre de l'échantillon à partir duquel on veut prédire la concentration $c_{\text{prédiction}}$ du constituant à doser.

Étape P1 : Nous projetons le spectre $a_{\text{prédiction}}$ sur le vecteur w_1 :

$$\text{(éq. 86)} \quad h_1 = a_{\text{prédiction}}^T \cdot w_1$$

Étape P2 : On extrait ensuite la contribution du premier facteur dans le spectre $a_{\text{prédiction}}$ pour obtenir le nouveau spectre $a_{\text{résidu}}$:

$$\text{(éq. 87)} \quad a_{\text{résidu}} = a_{\text{prédiction}} - h_1^T \cdot w_1^T$$

On réitère les étapes P1 et P2 pour toutes les composantes de matrice W . Ainsi par exemple, la contribution de la deuxième composante est extraite de la matrice des données spectrales par :

$$\text{(éq. 88)} \quad h_2 = a_{\text{résidu}}^T \cdot w_2$$

Lorsque le spectre a été projeté sur toutes les directions de la base vectorielle, nous obtenons la concentration par :

$$\text{(éq. 89)} \quad c_{\text{prédiction}} = \sum_{i=1}^{n_{\text{opt}}} h_i \cdot q_i$$

où q_i est le $i^{\text{ème}}$ élément de la matrice loading concentration q déterminée en calibration.

En résumé :

Méthode PLS

Avantages :

- Tous les avantages de la régression PCR peuvent être repris puisque PLS utilise aussi la condensation des données.
- Les facteurs loading extraits reflètent directement les variations des concentrations des constituants à doser contrairement à PCR qui travaille uniquement sur la variance spectrale.

Inconvénient :

- Il faut choisir judicieusement le nombre optimal de loading sous peine d'obtenir un underfitting ou un overfitting.

3.5.5 Comparaison des algorithmes

La méthode CLS ne figure pas dans les remarques qui suivent car elle est considérée comme inutilisable pour le dosage de produits complexes. Il n'existe pas de règle absolue sur la supériorité d'un algorithme de régression particulier. Seules les conditions d'utilisation permettent de donner des tendances. Si le nombre de variables de régression (longueurs d'onde dans le cas de spectres) est important, PLS et PCR donnent de meilleurs résultats que MLR. De plus, la méthode MLR est très sensible au recouvrement et au bruit spectral⁽²⁵⁹⁾. Si ces mêmes variables de régression ont un caractère très colinéaire (spectre NIR), la régression PLS est plus efficace^(182,183). Elle l'est aussi lorsque des variations non corrélées aux concentrations se trouvent sur les spectres (variation de la ligne de base). En effet, si ces variations sont importantes, la régression les prendra en compte au même titre que les variations chimiques des échantillons. Par contre, la méthode MLR peut être plus efficace que PCR ou PLS si le nombre de variables de régression est faible (moins de 10)^(120,291). La régression qui semble être la plus appropriée pour l'analyse quantitative à partir de spectres proche infrarouge est PLS.

IV VALIDATION DE MAQUETTES DE FAISABILITÉ

4.1 Introduction

Les instrumentalistes associés au projet ont conçu des maquettes d'analyseur de lait respectant le cahier des charges présenté ci-après. L'objectif de notre étude était de contrôler les performances des appareils proposés pour l'analyse quantitative des matières grasses et des protéines. Chaque étape du développement de l'instrumentation a été validée par des résultats issus de la chimométrie. L'approche est originale, car dans de nombreux cas, l'analyse quantitative est développée sur un appareil commercial qui n'est pas dédié au dosage des constituants d'intérêt. De ce fait, l'analyseur n'est pas nécessairement optimisé. Afin de comprendre les choix instrumentaux effectués pour cet appareil, nous définirons les parties à valider ainsi que les outils pour y parvenir.

4.2 État actuel des contrôles laitiers

A l'heure actuelle, le contrôle du lait ne se fait pas sur le site. Un technicien du contrôle laitier vient régulièrement prélever un volume de lait représentatif de la traite du jour pour chaque vache du troupeau d'une exploitation. Les échantillons conditionnés dans des flacons sont envoyés au laboratoire régional qui les analyse sur des appareils complètement automatiques. Bien que les meilleurs d'entre eux effectuent les dosages MG et MP de 350 laits en une heure, la fréquence des contrôles n'est pas très importante au niveau de chaque exploitation. Ce mode de prélèvement introduit aussi des erreurs dans le dosage. En effet, le volume prélevé n'est pas toujours représentatif de la traite. Une inversion sur l'identification de deux flacons est encore possible. De plus, le transport des échantillons de la ferme jusqu'au laboratoire peut devenir un facteur d'erreur important pour les dosages car les secousses développent le barattage du lait. La lipolyse accélère la détérioration des protéines et des matières grasses du lait. La chaleur de l'été favorise ces réactions. Dans ces conditions, certains échantillons arrivent au laboratoire sous forme de lait caillé impropre à l'analyse. La mise en place de l'analyseur vise à la diminution ou à la suppression de ces erreurs.

4.3 Le cahier des charges

Le cahier des charges permet de définir un cadre pour la conception de l'appareil de dosage. Il regroupe des considérations chimiométriques, instrumentales et économiques.

- L'instrument doit doser les matières grasses (MG) et les protéines (MP). Les gammes de concentrations à couvrir sont :

◇ MG : de 25 à 60 g/kg.

◇ MP : de 25 à 45 g/kg.

La précision demandée est de 1 g/kg pour les erreurs standards de prédiction en MG et MP.

- La cadence de l'analyse doit être compatible avec la cadence de la traite. Il a été décidé que l'analyse spectrale devait prendre 30 secondes par échantillon.
- L'analyseur doit être capable de couvrir la zone spectrale 900-2500 nm afin de permettre l'évolution de l'appareil pour le dosage de nouvelles espèces chimiques sans aucune modification instrumentale. Ce point a permis d'écarter les configurations à base de filtres interférentiels.
- L'homogénéisation du lait n'est pas autorisée malgré son utilisation classique en spectroscopie infrarouge^(184,185,186,187,188). Elle corrige une grande partie des variations des données spectrales en éclatant les globules gras du lait harmonisant ainsi leurs tailles⁽¹⁸⁹⁾. De cette façon, cette diffusion parasite est diminuée. L'obstacle à son utilisation est propre au devenir du lait dosé par l'analyseur. En effet, après l'analyse spectrale, le lait repart dans le tanker qui collecte l'ensemble de la traite. Or un échantillon de lait homogénéisé dans le tanker augmente la vitesse de lipolyse et dénature son contenu⁽¹⁹⁰⁾.
- L'instrument doit être capable d'effectuer des mesures délocalisées à 12 endroits différents dans la salle de traite. Cela impose l'utilisation des fibres optiques.
- Le système doit résister aux conditions susceptibles d'être rencontrées pendant ou en dehors de la traite :
 - ◇ Taux d'humidité pouvant atteindre 95 %.
 - ◇ Température de -15 à +40°C.
 - ◇ Vibrations.
 - ◇ Projections d'eau et de déjections.
 - ◇ Nettoyage du lactoduc par des acides forts et bases fortes concentrés ainsi que par de l'eau à 90°C.

Ces points permettent d'écarter les systèmes interférométriques qui sont très sensibles aux variations de température et à l'humidité.

- L'analyseur doit fonctionner en 24 volts et résister aux micro-coupures souvent présentes dans les systèmes électriques des exploitations.
- Son poids ne devra pas dépasser 20 kg pour qu'il soit considéré comme portable.
- Le coût fixé est limité par rapport aux appareils NIR du marché.
- L'analyseur sera capable d'effectuer le dosage de lait provenant d'autres races laitières comme les caprins.

L'impossibilité de trouver un appareil commercial susceptible de remplir toutes ces conditions a motivé la conception d'un analyseur proche infrarouge spécifique. L'intégration et l'utilisation de l'analyseur dans l'étable permettent de comprendre les contraintes imposées par le cahier des charges.

4.4 Intégration et utilisation de l'appareil dans l'étable

Avant d'aborder l'intégration de l'appareil dans la salle de traite, il convient de donner son schéma de principe. La Figure 37 donne une représentation simplifiée de l'analyseur conforme au cahier des charges. Une analyse plus précise sera donnée lors de la présentation de la maquette de faisabilité. La lumière polychromatique de la source passe dans une première fibre optique (fibre d'éclairage). De cette façon, la lumière est guidée jusqu'au lieu de l'analyse dans la sonde optique. Le lait qui circule dans cette sonde absorbe et diffuse la lumière qui l'éclaire. La deuxième fibre récupère cette énergie qui est renvoyée dans l'analyseur pour effectuer l'analyse spectrale (monochromateur). Lorsque l'analyseur est intégré dans la salle de traite, il est placé dans la fosse (Figure 38). Pour chaque poste de traite, il y a une sonde couplée au compteur-échantillonneur. Ainsi, une salle constituée de n stalles sera équipée de n sondes optiques, de n fibres optiques d'éclairage qui partent de l'analyseur et de n fibres optiques de collection qui y arrivent. L'introduction de l'analyseur dans la salle de traite ne change pas le déroulement classique des opérations. Les vaches entrent sur les quais. L'identification de la vache est effectuée automatiquement grâce à la détection d'une puce électronique portée par l'animal. L'éleveur qui est dans la fosse branche un faisceau trayeur sur chaque vache. Lorsque la traite d'une vache est terminée le compteur-échantillonneur décroche automatiquement le faisceau. Il transmet à l'analyseur de

nombreuses données telles que le volume de la traite, la courbe d'éjection du lait et le numéro d'identification de la vache.

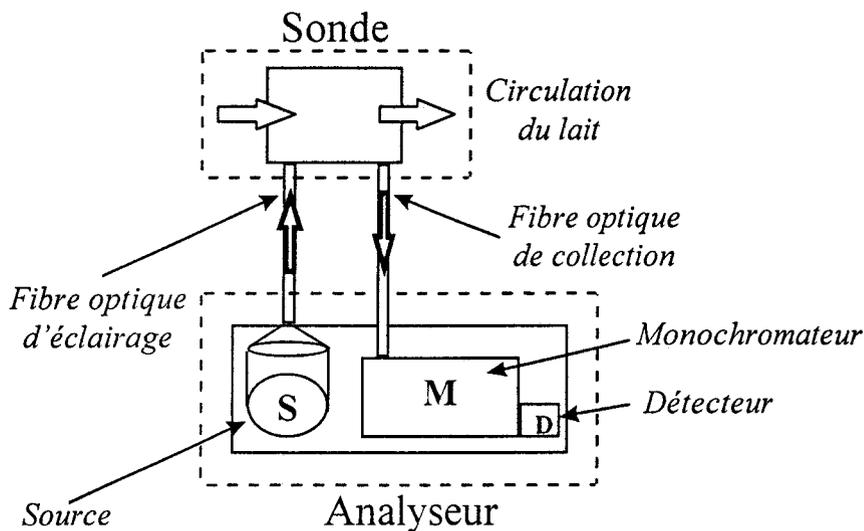


Figure 37 : Schéma de principe de l'analyseur.

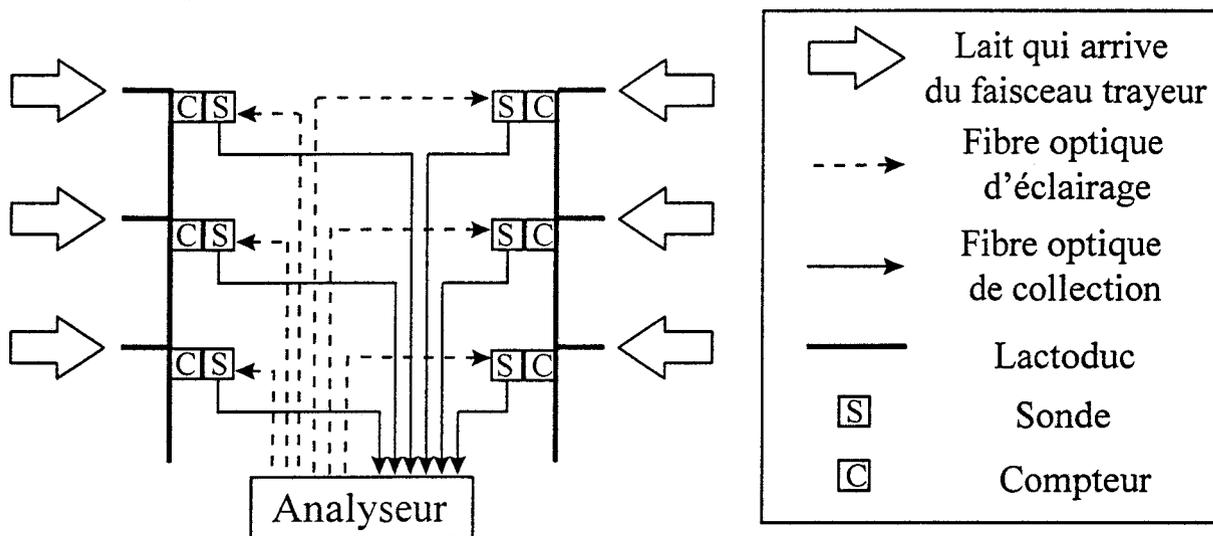


Figure 38 : Intégration de l'analyseur dans la salle de traite.

Le compteur annonce finalement à l'analyseur que l'analyse spectrale peut commencer sur la sonde associée. Le volume de lait représentatif de la traite qui a été prélevé par le compteur circule dans la sonde. Au bout de 30 secondes, l'acquisition du spectre est terminée. Au même titre que les premières données, les spectres sont stockés sur le disque dur de l'analyseur. Une deuxième analyse peut commencer lorsqu'une autre sonde annonce la fin de la traite sur un poste différent. Un deuxième groupe de vaches succède au premier. Lorsque la traite du

troupeau est terminée, l'analyseur calcule les teneurs MG et MP pour chaque vache à partir du spectre proche infrarouge.

4.5 Les apports de cette nouvelle méthode de dosage.

L'intégration de l'appareil dans l'étable permet une fréquence de contrôle du lait plus importante. Sur les grosses exploitations, il peut rester d'ailleurs à demeure. Une analyse quotidienne est donc effectuée sur chaque animal améliorant ainsi le contrôle de la qualité du lait. Le dosage régulier est important pour les éleveurs qui optimisent les concentrations MG et MP afin d'atteindre le prix maximum au litre de lait. L'analyse in situ élimine les erreurs d'échantillonnage introduites lors des prélèvements manuels. L'identification automatique de la bête supprime aussi toutes les erreurs d'attribution des bulletins d'analyses. Le barattage dû au transport est supprimé. L'analyse rapide du lait écarte la détérioration des protéines et des matières grasses au cours du temps. Les bactéries ne se développent pas. L'analyseur permet finalement d'avoir un contrôle laitier sur les nouvelles machines à traire complètement automatiques (robots de traite). Pour ce nouveau concept, les vaches viennent se faire traire lorsqu'elles le veulent à n'importe quelle heure de la journée. Lorsqu'une vache se place sur le quai de traite, un système de détection ultrasons/laser permet de localiser les pis de la vache. Un bras robotisé place le faisceau trayeur. Lorsque la traite est terminée, la vache repart à l'étable. Le couplage de notre appareil sur ce système est primordial, car le caractère irrégulier de la traite rend le contrôle classique difficile voire impossible.

4.6 Description de la maquette à valider

La maquette de départ a été conçue dans l'esprit d'un montage modulaire où des modifications instrumentales pourraient être apportées sans avoir à remettre en cause l'ensemble de l'instrument. Les éléments développés par la Société JOBIN-YVON intègrent bien ce concept. La figure ci-après décrit en détail la configuration de la maquette à valider par la chimométrie (Figure 39). Elle est tout d'abord constituée d'un bloc source. La lampe tungstène est la seule source d'énergie lumineuse qui permette de couvrir de manière continue le domaine spectral de 900 à 2500 nm⁽¹⁹¹⁾. A ce stade, il est nécessaire d'avoir une alimentation très stabilisée (alimentation à découpages) sous peine de fluctuations importantes

de l'énergie lumineuse. Un miroir sphérique de couplage amène la lumière vers le modulateur. Ce dernier est tout simplement un disque à palettes qui tourne à une fréquence précise. Nous décrirons sa fonction lors de l'étape de démodulation sur la détection synchrone. Une optique de focalisation permet ensuite de condenser le faisceau lumineux sur l'entrée de la fibre optique. L'énergie transite donc par la fibre d'éclairage pour parvenir à la sonde qui est en contact avec le lait à analyser.

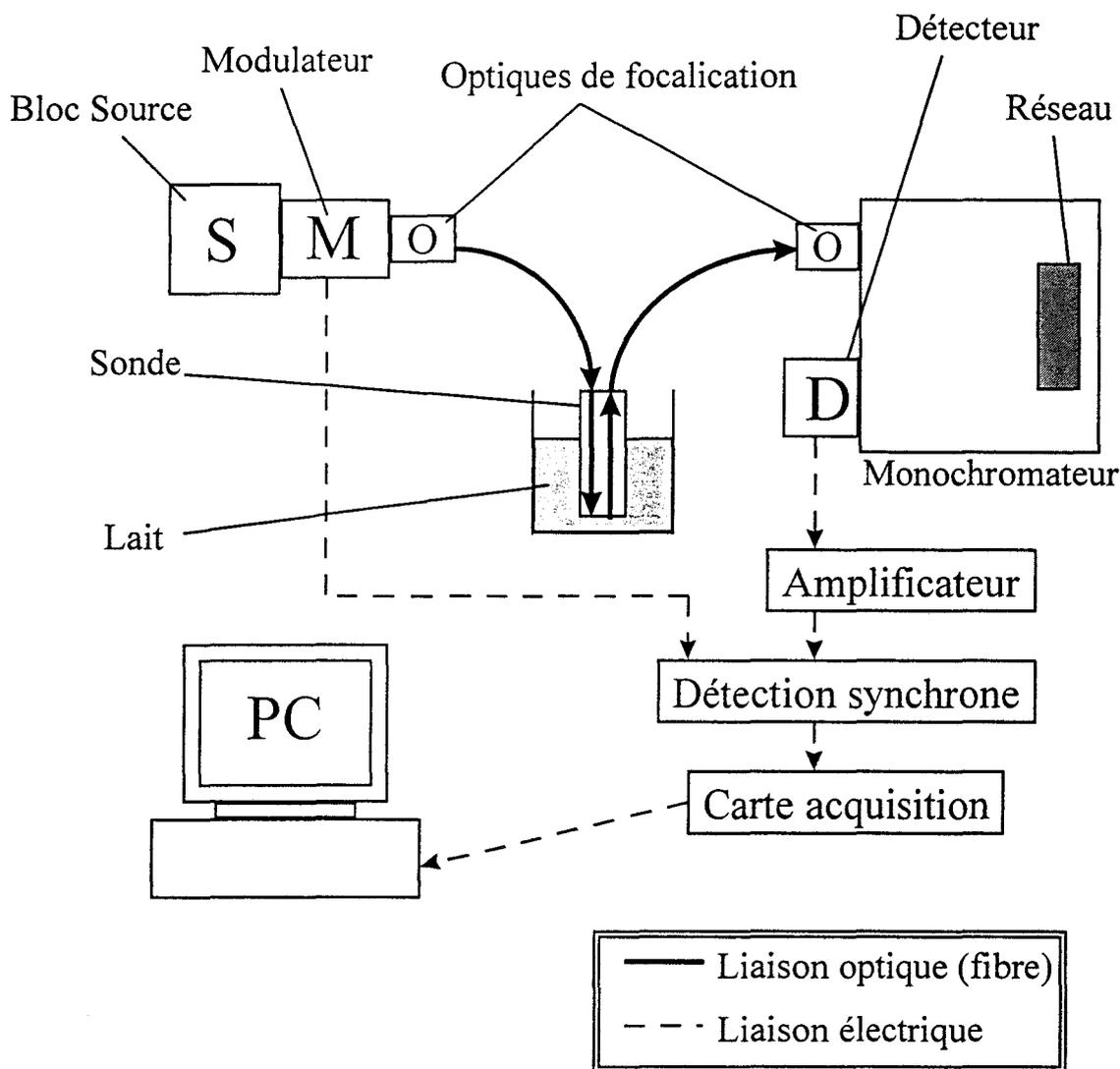


Figure 39 : Détail de la maquette de faisabilité.

A cet endroit, la fibre de collection récupère l'énergie non absorbée. La majeure partie de notre validation portera d'ailleurs sur la recherche de la sonde la plus performante pour l'analyse quantitative du lait. C'est en effet ce dispositif qui conditionne toutes les analyses délocalisées en proche infrarouge. Une deuxième optique de focalisation conjugue l'image de sortie de la fibre sur la fente d'entrée du monochromateur. Le monochromateur utilisé pour

cette maquette est un appareil de recherche du type 270M de la Société JOBIN-YVON. Il est conçu suivant une configuration optique asymétrique de CZERNY-TURNER. Les fentes d'entrée et de sortie sont motorisées permettant un ajustement variable. De cette façon, il est possible de jouer sur la bande passante. Un système de tourelle permet de commuter deux réseaux dispersifs optimisés pour des domaines spectraux différents. Le déplacement du réseau et les autres commutations sont entièrement pilotés par ordinateur. Ce monochromateur n'est pas adapté au cahier des charges de l'analyseur car il est trop lourd, trop volumineux et trop lent. Il a été néanmoins choisi pour ses nombreux réglages possibles.

Les impératifs de temps d'analyse, de poids et d'encombrement nous ont incité à développer, parallèlement à l'étude des sondes, deux nouvelles générations de monochromateur chez JOBIN-YVON. Le premier est un système dit à balayage rapide. Ce système révolutionnaire permet d'effectuer plusieurs spectres en une seconde sur un large domaine spectral (900 à 2500 nm par exemple). Il sera adaptable à un grand nombre de monochromateurs déjà existants dans la gamme JOBIN-YVON. Un brevet en cours d'acceptation nous empêche de donner plus de détails. C'est le deuxième système dit à entraînement direct qui sera intégré dans l'analyseur final. Le réseau est placé directement sur l'axe d'un moteur pas à pas. Malgré le temps passé pour ces développements, nous ne nous attarderons pas sur l'instrumentation des monochromateurs, car ce sujet ne concerne pas directement le thème de recherche posé.

L'énergie qui pénètre donc par la fente d'entrée du monochromateur est diffractée par le réseau. Une partie de cette énergie est envoyée sur la fente de sortie. Le détecteur qui se trouve derrière la fente transforme l'information lumineuse qui arrive en information électrique. C'est une représentation très schématique car les détecteurs ne fonctionnent pas tous de la même manière. Le signal électrique est amplifié puis injecté dans la détection synchrone (*lock-in*). Elle discrimine le signal modulé provenant de la source des radiations parasites. Un signal électrique est finalement envoyé sur une carte électronique d'acquisition en communication avec un ordinateur qui stocke le spectre pour l'analyse quantitative. Tout les constituants de cette maquette ont une influence sur la capacité prédictive de l'analyseur. La validation systématique des paramètres instrumentaux portera principalement sur l'élaboration des sondes optiques, car elle est la clef de voûte de la réussite de ce projet. Les autres paramètres seront développés en arrière-plan.

4.7 La chimiométrie : outil de validation des paramètres instrumentaux

4.7.1 Introduction

Cette partie permet d'apprécier la procédure mise en œuvre pour la validation et l'optimisation de l'analyseur. Elle peut être appliquée au développement de toute analyse quantitative par une méthode indirecte telle que la spectroscopie. Dans les paragraphes suivants, nous définirons les échantillons à utiliser, les procédures de contrôle de répétabilité et de linéarité spectrale. Nous terminerons sur les analyses quantitatives des gammes de laits reconstitués. Le contrôle final portera sur une analyse de laits de ferme.

4.7.2 L'échantillon

Comme nous avons pu le voir, le lait est un mélange complexe qui contient de nombreuses espèces chimiques en proportions variables. La première étape dans la mise en place d'une analyse quantitative est l'étude de « solutions modèles »⁽¹⁹²⁾. L'objectif est en fait de préparer des échantillons en matières grasses et protéines variables mais dans une matrice constante. L'analyse quantitative se retrouve donc simplifiée car les variations spectrales sont uniquement dues aux constituants à doser. Si les différents tests de validation ne sont pas positifs sur ces échantillons, l'analyse quantitative ne sera pas efficace sur les échantillons réels. Dans le cas du lait, la préparation d'échantillons reconstitués simplifie d'autant plus le problème que les structures globulaires et micellaires des matières grasses et des protéines sont fortement altérées ou détruites. Dans un premier temps, une partie de notre travail a consisté à rechercher une procédure de reconstitution de laits adaptée à nos besoins. En effet, devant le nombre important d'analyses quantitatives de contrôle à effectuer pour les nombreuses modifications instrumentales, cette reconstitution devait être la plus rapide et la moins coûteuse possible. L'unité de recherche INRA de Poligny maîtrise depuis longtemps la préparation d'échantillons à teneurs garanties en matières grasses et en protéines. Des centrifugations⁽¹⁾ et des ultra-filtrations⁽¹⁹³⁾ sur membranes permettent de séparer un lait entier en trois parties distinctes : deux solutions concentrées de matières grasses et de protéines et une solution de sérum limpide qui constitue la matrice. Après cette séparation, on effectue le dosage des matières grasses et des protéines dans les deux solutions. Les échantillons sont

ensuite reconstitués par mélange des trois solutions. Cette technique permet d'obtenir des échantillons reconstitués à matrice constante très proches des échantillons réels. Elle demande néanmoins un équipement spécifique et un temps de préparation important. Une méthode de reconstitution du lait à partir de lait en poudre écrémé, de matières grasses en solutions aqueuses (crème liquide) et d'eau distillée a finalement été adoptée pour sa simplicité, sa rapidité et son coût. Pour cette procédure, nous nous sommes directement inspirés de la méthode de reconstitution utilisée dans l'industrie agro-alimentaire⁽¹⁾. L'eau distillée est tout d'abord portée à 40°C. On ajoute progressivement sous agitation le lait en poudre écrémé. Le mélange est laissé sous agitation pendant une demi-heure afin d'obtenir une hydratation convenable. La crème liquide est finalement incorporée dans le mélange. Les échantillons préparés peuvent ensuite être conservés quelques jours à 4°C. L'information sur les concentrations des matières grasses et des protéines de la poudre de lait et de la crème permettent ensuite par pesée de reconstituer toutes les gammes de concentrations recherchées. Lorsque les tests de validation sur les solutions modèles sont terminés, ils sont renouvelés sur les laits de ferme. Chaque échantillon de lait frais utilisé pour la validation sera dit « individuel » car il proviendra d'un animal unique. Ces échantillons sont mis en opposition avec les laits dits de « mélanges ». Ce point est très important, car dans cette phase finale de la validation, les échantillons analysés doivent être représentatifs des laits susceptibles d'être rencontrés au moment de l'analyse donc de la traite dans notre cas. De plus, l'utilisation de laits de mélanges donnerait une mauvaise appréciation de la capacité prédictive de l'analyseur. En effet, il a été démontré que le dosage des matières grasses et des protéines du lait de chèvre dans le proche infrarouge donnait des erreurs de prédiction 5 fois moins importantes pour les laits de mélanges⁽³⁸⁾. Afin de simplifier l'analyse, les laits frais proviendront d'une même exploitation. Dans une étude ultérieure, il sera nécessaire de démontrer la nécessité ou non de développer une calibration pour chaque race bovine.

4.7.3 La stabilité du signal

Une bonne analyse quantitative passe par la maîtrise de la répétabilité du signal. Il n'existe pas de définition univoque de la répétabilité spectrale dans la littérature. La définition de MULTON⁽¹¹¹⁾ déjà présentée au paragraphe 3.2 (p 56) sera utilisée afin d'évaluer la bonne évolution du système au cours des modifications. Cette répétabilité est l'estimateur de l'erreur

systematique des données spectrales. C'est en fait l'écart type des données spectrales acquises successivement sur un même échantillon, par le même utilisateur, sur le même analyseur, dans le même laboratoire sur des intervalles de temps assez courts. Ainsi pour chaque modification, la répétabilité (RSD) sera calculée à partir de l'acquisition des spectres de 10 échantillons identiques provenant d'un même lait. Ce coefficient rendra compte des perturbations externes et internes à l'analyseur (variations de température, d'humidité, dérives électroniques et optiques...).

4.7.4 La linéarité du signal

Nous avons vu que les techniques de régressions multivariées utilisées pour cette recherche étaient basées sur une relation linéaire entre le signal optique et la concentration des analytes. Si cette relation est fragile ou même inexistante, les méthodes statistiques ne nous permettront pas d'arriver à une analyse quantitative de qualité. Nous pouvons même dire que les erreurs de prédiction seront très importantes. Dans ce cas, les techniques de pré-traitement des spectres seront appliquées afin de retrouver une relation de linéarité satisfaisante. La procédure de contrôle de la linéarité est très simple. Elle est d'ailleurs appliquée aux analyseurs de lait travaillant dans le moyen infrarouge⁽¹⁹⁴⁾. On prépare une dizaine d'échantillons par dilution d'un seul composé en solution aqueuse. Dans notre cas, nous pouvons préparer une gamme diluée de poudre de lait ou de crème liquide. L'étendue de la gamme des concentrations devra être en accord avec le cahier des charges soit 25-45 g/kg pour une dilution de protéines ou 25-60 g/kg pour les matières grasses. La définition de ces domaines d'utilisation est importante car des étendues de concentrations trop larges induisent des variations d'absorbance importantes et donc des non-linéarités. On effectue ensuite l'acquisition des spectres de la gamme. Une régression linéaire simple est calculée entre la réponse optique et la concentration aux longueurs d'onde caractéristiques du composé dilué. Une représentation graphique de ces deux paramètres permet rapidement d'observer des problèmes de non-linéarité.

4.7.5 L'analyse quantitative

L'analyse quantitative constitue l'aboutissement de la procédure de validation. Elle est effectuée tout d'abord sur une trentaine de solutions modèles couvrant bien les concentrations MG et MP fixées par le cahier des charges. Ce lot d'échantillons est divisé en deux groupes. Le premier est utilisé pour construire l'équation de calibration alors que le deuxième permet de tester la qualité prédictive de cette dernière. L'erreur standard de calibration et le coefficient de corrélation ne permettent pas d'apprécier la qualité prédictive du modèle construit. Une erreur standard de calibration (SEC) inférieure à 1 g/kg et un coefficient de corrélation (R^2) proche de 1 ne constituent qu'une condition nécessaire mais non suffisante au développement d'une bonne analyse quantitative sur l'analyseur. Finalement, l'erreur standard de prédiction (SEP) devra être inférieure à 1g/kg pour les deux espèces à doser. Si cette condition n'est pas remplie pour le dosage des solutions modèles, l'analyse quantitative de laits de ferme sera inefficace.

4.8 La sonde « transmission pure »

4.8.1 Introduction

Comme nous l'avons déjà dit, la sonde est une pièce maîtresse de la réussite de cette recherche. Chaque génération de configuration optique sera développée dans une partie distincte. Nous respecterons aussi la chronologie des modifications instrumentales.

Cette première sonde dite « transmission pure » a été conçue dans l'optique de la simplicité. En terme de coût, c'est aussi une configuration très avantageuse puisqu'elle ne nécessite que l'usinage d'une pièce métallique. Les deux fibres optiques de l'analyseur, c'est-à-dire la fibre d'éclairage et la fibre de collection, sont placées en vis-à-vis (Figure 40). Le chemin optique est caractérisé par la distance entre les extrémités des deux fibres où le lait circule. Nous avons fait remarquer page 44 que la transmission n'était adaptée qu'aux milieux peu diffusants. En fait, au début de l'étude nous ne pouvions pas présager que le lait l'était autant. Dans le pire des cas, nous pensions qu'un chemin optique faible (environ 200 μm) devait nous permettre de récupérer suffisamment d'énergie pour l'analyse spectrale. Ce type de montage a permis d'effectuer de nombreuses analyses quantitatives dans des domaines très

divers^(195,196,197,198,199). On peut d'ailleurs, à l'heure actuelle, considérer que l'analyse délocalisée sur de grandes distances n'est possible qu'avec ce type de sonde. L'exemple certainement le plus spectaculaire de délocalisation est le dosage d'alcools sur un process situé à 1 km de l'analyseur proche infrarouge soit 2 km aller et retour⁽²⁰⁰⁾. Dans ce cas l'analyse quantitative se prête bien car les alcools sont limpides et homogènes. La diffusion est donc négligeable même sur un chemin optique de 5 mm utilisé pour cette étude.

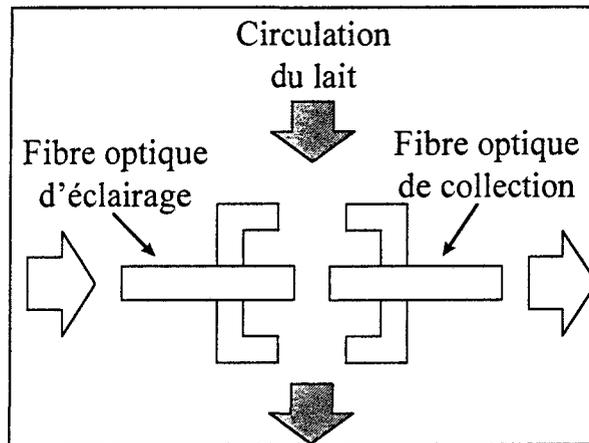


Figure 40 : Sonde "transmission pure".

Pour la validation de cette sonde, nous avons utilisé 5 m de fibre de silice d'un diamètre de 600 μm pour l'éclairage ainsi que 5 m pour la collection du signal optique. Le chemin optique est de 200 μm .

4.8.2 La répétabilité

L'acquisition d'un spectre d'absorbance commence par le choix d'une référence. En général, pour l'étude des produits aqueux comme le lait, un spectre de référence air ou eau est indifféremment utilisé^(85,195,201,202). Néanmoins, le choix d'une référence air est impossible pour notre analyseur qui est intégré dans un milieu humide. De plus, les cycles de lavage appliqués au système de traite et à la sonde utilisent des solutions aqueuses. Dans ces conditions, il serait impossible d'assurer une sonde complètement sèche pour l'acquisition d'une référence. La référence eau s'impose naturellement. Nous testons la répétabilité spectrale sur un échantillon de crème diluée contenant 40 g de matières grasses par kilogramme de lait reconstitué (noté 40 g/kg en MG). Un incrément spectral de 5 nm est choisi en accord avec la

bande passante du monochromateur qui est de 5 à 10 nm. Les bandes d'absorption sont en effet très larges. Sauf avertissement de notre part, ce paramétrage conforme à la bibliographie, restera constant tout au long de cette recherche. Les échantillons seront toujours portés à une température de 38 °C (température d'un lait fraîchement traité). Pour des températures plus basses, les globules gras ont tendance à se regrouper en grosses grappes inexistantes lors de la traite⁽¹⁾. Conformément à la procédure, nous effectuons l'acquisition d'un spectre pour dix sous-échantillons préparés à partir de la solution de base. La figure ci-dessous présente les spectres « simple faisceau » de l'air, de l'eau et d'un des 10 spectres de l'échantillon à 40 g/kg en MG. Une comparaison permet d'interpréter les contributions spectrales. Sur le spectre de l'air, on note une absorption à 1110 nm attribuable à une anomalie de WOOD⁽⁷⁶⁾. Ce phénomène classique caractéristique des réseaux de diffraction a été découvert en 1902. Il a pour origine un défaut dans la distribution d'énergie des réseaux.

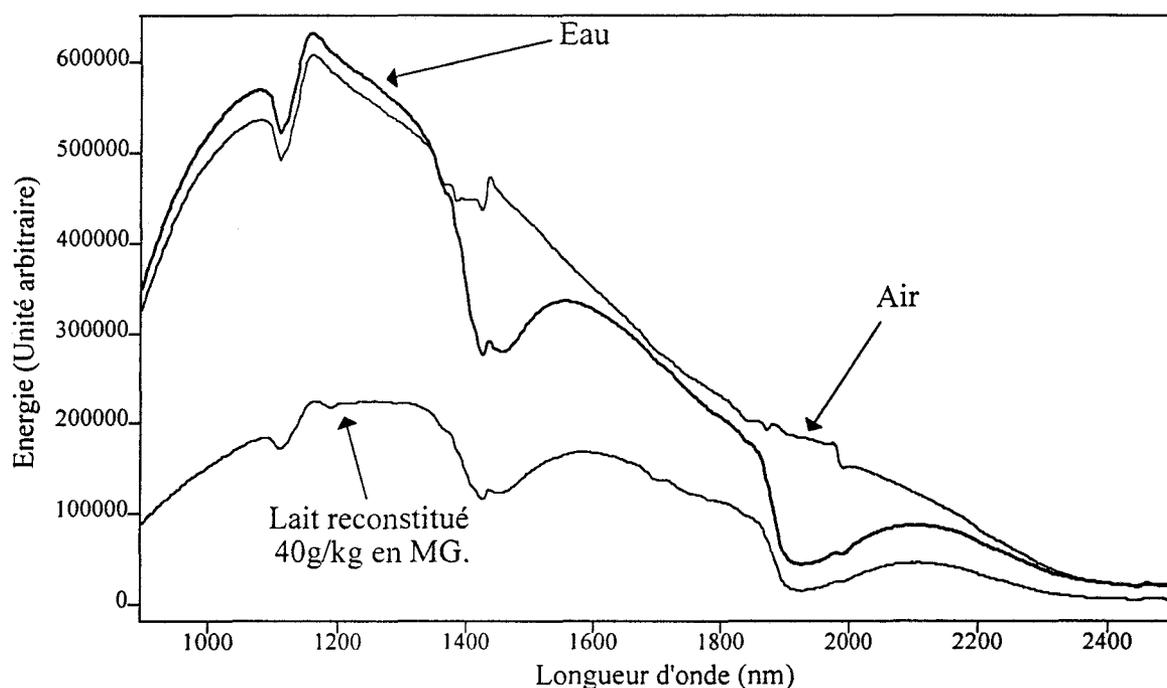


Figure 41 : Observation des contributions spectrales sur les spectres "simple faisceau".

Conformément à la Figure 21, nous retrouvons l'absorption des OH présents dans la silice de la fibre sur le spectre de l'air entre 1350 et 1435 nm. Sur le spectre de l'eau, nous remarquons les deux bandes larges caractéristiques des vibrations OH dans la molécule à 1460 et 1920 nm. Sur le spectre de lait reconstitué, de faibles contributions dans les zones 1680-1755 et 1420-1430 attribuables aux matières grasses s'ajoutent aux précédentes. Au delà de 2100 nm,

le signal qui arrive au détecteur est très faible. La Figure 42 représente les spectres d'absorbance ($A = \log 1 / T$) des 10 sous-échantillons utilisés pour la répétabilité. Afin de faciliter l'interprétation, l'un d'entre eux a été décalé de 0,1 UA. Nous remarquons que la ligne de base n'est pas rectiligne. Les courbures aux extrémités des domaines spectraux sont fréquentes en spectroscopie proche infrarouge lorsque des phénomènes de diffusion sont importants^(203,204,205). Le bruit est relativement important sur l'ensemble des spectres. Au-dessus de 2100 nm, le rapport signal sur bruit est plus que médiocre. En effet, nous avons vu sur les spectres « simple faisceau » que très peu d'énergie arrivait sur le détecteur.

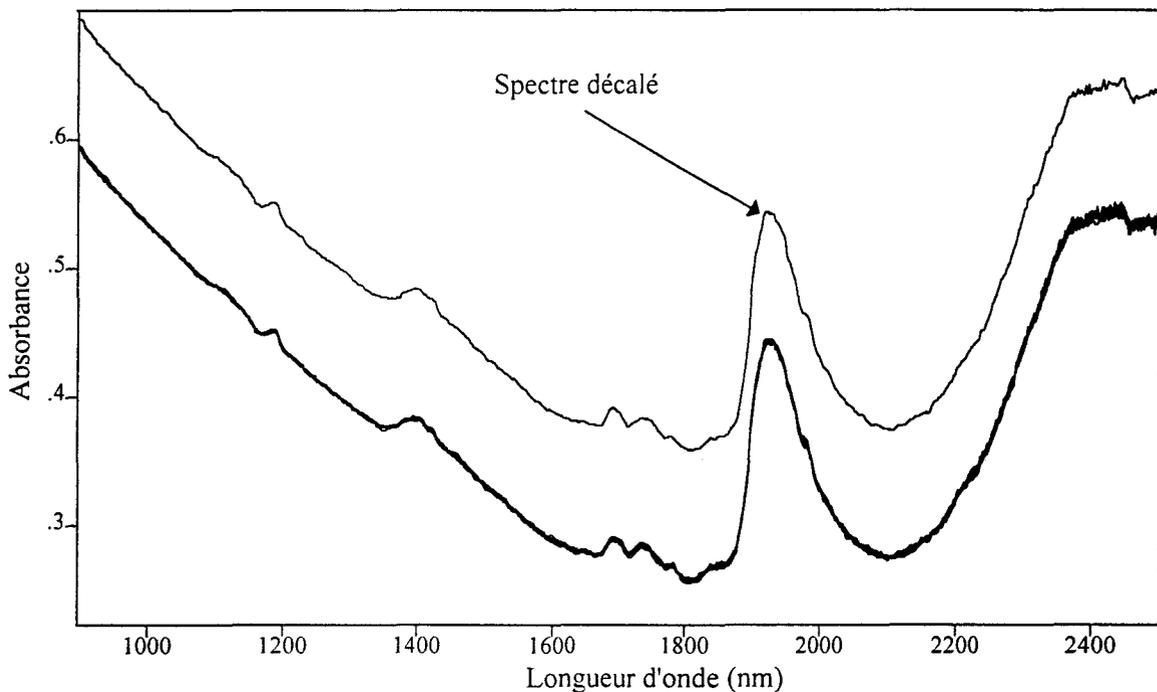


Figure 42 : Superposition des spectres utilisés pour l'évaluation de la répétabilité.

La bande la plus importante est centrée sur 1926 nm. Son intensité trop importante ne nous permet pas de l'attribuer à la troisième harmonique de la liaison CH dans les longues chaînes aliphatiques des matières grasses. Ce phénomène déjà observé est dû à la différence de température entre l'échantillon analysé et la référence eau⁽²⁰⁶⁾. En effet, la bande située à 1950 nm a tendance à se déplacer avec la température. L'épaulement visible à 1980 nm peut provenir de la troisième harmonique CH des matières grasses. Dans la zone 1695-1780 nm, trois absorptions sont visibles. Nous les attribuons aux premières harmoniques des elongations CH dans les groupements $-CH_2-$ ou $-CH_3$. Cette zone d'absorption est aussi caractéristique des troisièmes harmoniques de la déformation asymétrique de CH dans $-CH_2-$

et de la déformation asymétrique de CH dans C-CH₃. Une zone intéressante se situe entre 1400 et 1425 nm. Cette absorption peut provenir de la troisième harmonique de la vibration d'élongation de C=O dans les esters. La différence en température peut aussi faire apparaître une première harmonique de l'élongation OH dans l'eau. Une dernière bande à 1185 nm est attribuée à une deuxième harmonique de la vibration d'élongation CH. Après cette caractérisation, la répétabilité (RSD) peut être calculée sur ces différentes longueurs d'onde (Tableau 8).

Longueur d'onde (nm)	Répétabilité (RSD %)
1980	0,65
1780	0,49
1750	0,47
1685	0,46
1400	0,31
1185	0,22

Tableau 8 : Répétabilité de la sonde transmission pure

Une répétabilité de quelques pour-cent est acceptable en spectroscopie moyen infrarouge lors de la mise en place d'une analyse quantitative^(108,109,207). En spectroscopie proche infrarouge, le recouvrement spectral important nous oblige à avoir une meilleure répétabilité. Les résultats obtenus pour la sonde transmission sont bons. Pour la suite de l'étude, nous utiliserons ce paramètre de manière relative afin d'observer l'effet des modifications apportées à l'analyseur sur la qualité spectrale.

4.8.3 La linéarité

Afin de vérifier la linéarité du signal optique en fonction de la concentration, nous avons préparé une gamme de crèmes liquides diluées dont les teneurs sont consignées dans le tableau ci-dessous. Les spectres « simple faisceau » sont présentés dans la Figure 43.



Echantillon N°	Concentration MG (g.kg)
1	25,04
2	31,52
3	35,61
4	39,74
5	44,40
6	49,23
7	54,59
8	59,69

Tableau 9 : Gamme de linéarité en matières grasses.

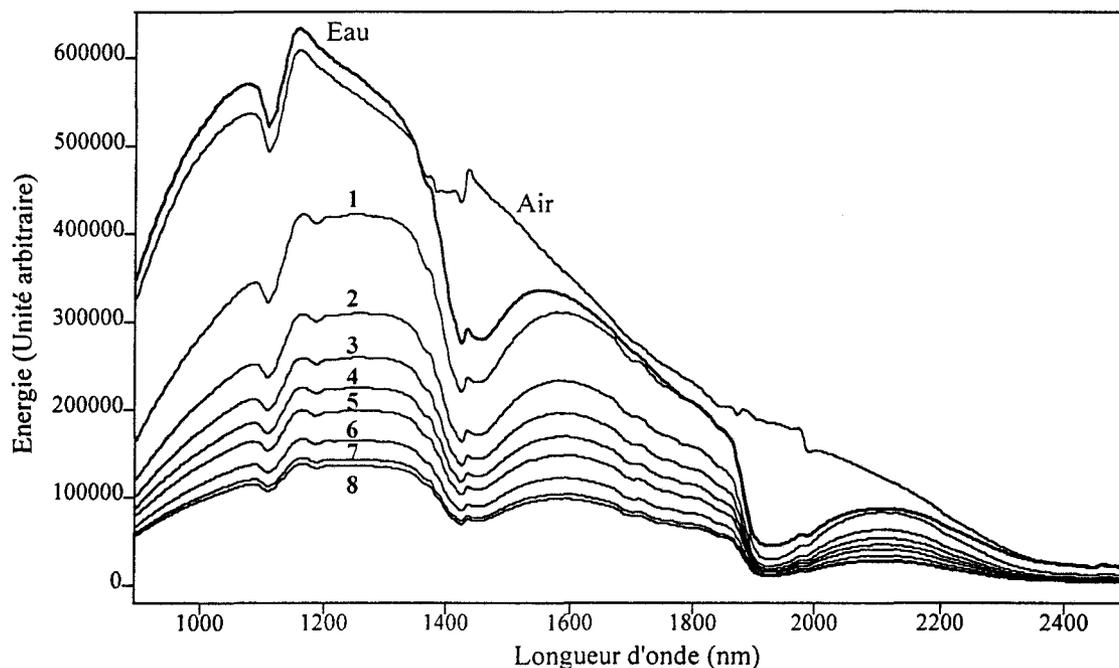


Figure 43 : Spectres "simple faisceau" pour la gamme de matières grasses.

Le numéro de l'échantillon repère chaque spectre. Au-dessus de 1950 nm, nous avons un signal complètement saturé c'est-à-dire que le détecteur mesure des énergies quasiment nulles. Les spectres d'absorbance sont donnés sur la Figure 44.

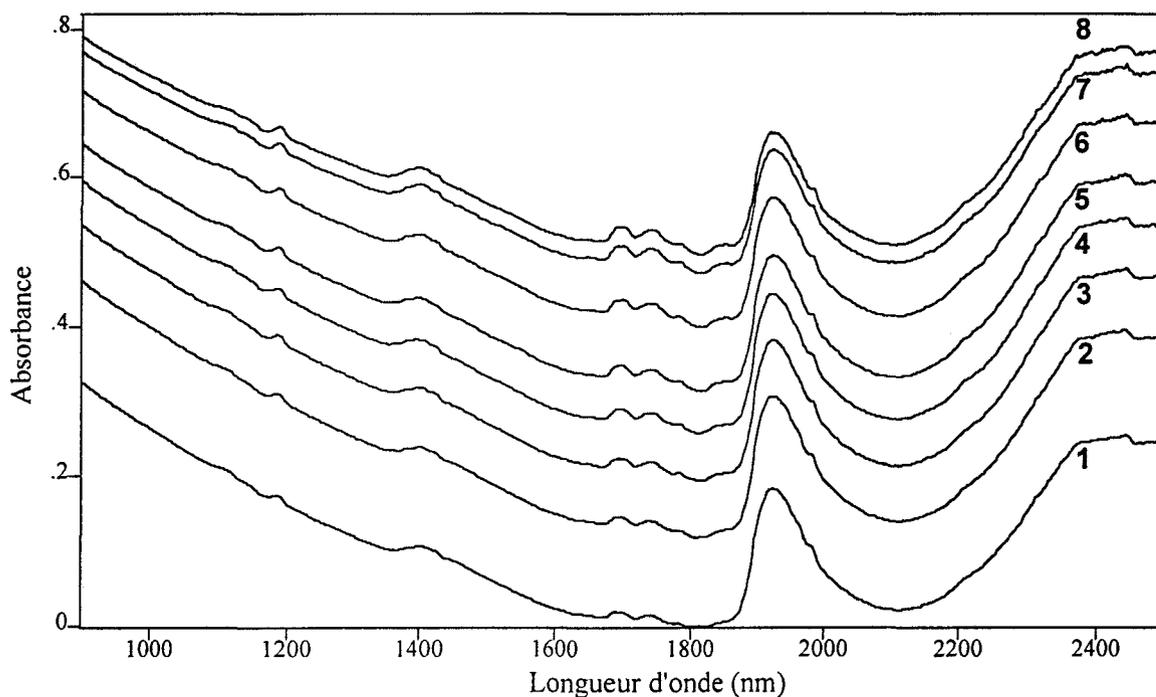


Figure 44 : Spectres d'absorbance pour la gamme de matières grasses.

Nous retrouvons bien les contributions spectrales présentées auparavant. Sur chaque bande, il y a une augmentation de l'absorbance en fonction du taux de matières grasses. Nous pouvons remarquer qu'un fond continu vient se greffer sur l'ensemble du spectre. Son importance est étroitement corrélée à la diffusion des rayons proches infrarouges par les matières grasses. L'évaluation de la linéarité passe par le calcul d'une régression linéaire simple entre les absorbances des échantillons de la gamme et leurs concentrations aux longueurs d'onde caractéristiques des matières grasses. La Figure 45 donne les résultats obtenus pour la longueur d'onde à 1685 nm.

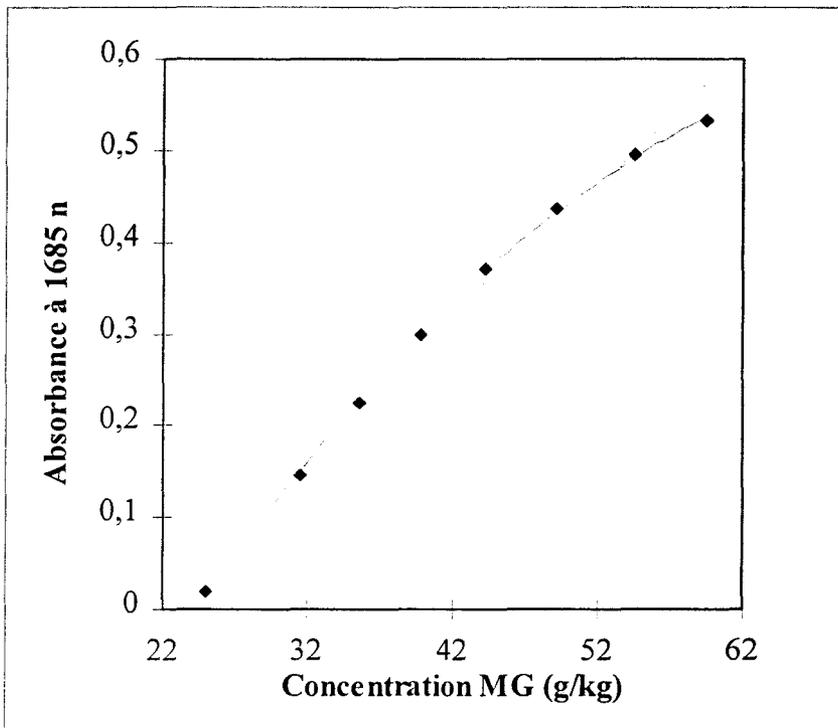


Figure 45 : Linéarité du signal optique à 1685 nm en fonction de la concentration MG.

Sur ce graphique, la courbe en pointillés représente la droite de régression linéaire obtenue par la méthode des moindres carrés. Si la linéarité est vérifiée, tous les points sont situés sur cette droite avec un coefficient de corrélation qui est très proche de 1. Nous voyons rapidement que la sonde transmission pure ne donne pas un signal optique linéaire par rapport à la concentration. Elle suit même plus facilement une courbe de régression polynomiale (trait plein). L'étude de la linéarité sur les autres longueurs d'onde du domaine spectral a donné des résultats strictement identiques. L'application des techniques de pré-traitement des spectres telles que la différentiation n'a pas permis de retrouver une relation linéaire. Cette dernière étude exclue définitivement le mode transmission. Implicitement, son utilisation en analyse quantitative donne des erreurs très importantes aux extrémités du domaine de concentration.

4.8.4 Discussion

Les problèmes de répétabilité et de linéarité peuvent être expliqués par la nature même des fibres optiques. En effet, nous avons vu à la page 44 qu'elles possédaient une ouverture numérique très faible. La configuration « transmission pure » qui utilise simplement ces fibres optiques en vis-à-vis n'atténue pas cet effet. La fibre de collection récupère donc une très

faible partie de la lumière diffusée par le lait dans les trois directions de l'espace (Figure 46). Ce défaut de signal et le facteur de dilution des constituants du lait donne un faible rapport signal sur bruit d'où une répétabilité médiocre. Les conditions opératoires trouvées dans la bibliographie ont motivé le choix arbitraire d'un chemin optique de 200 μm afin de commencer cette étude⁽⁸⁵⁾.

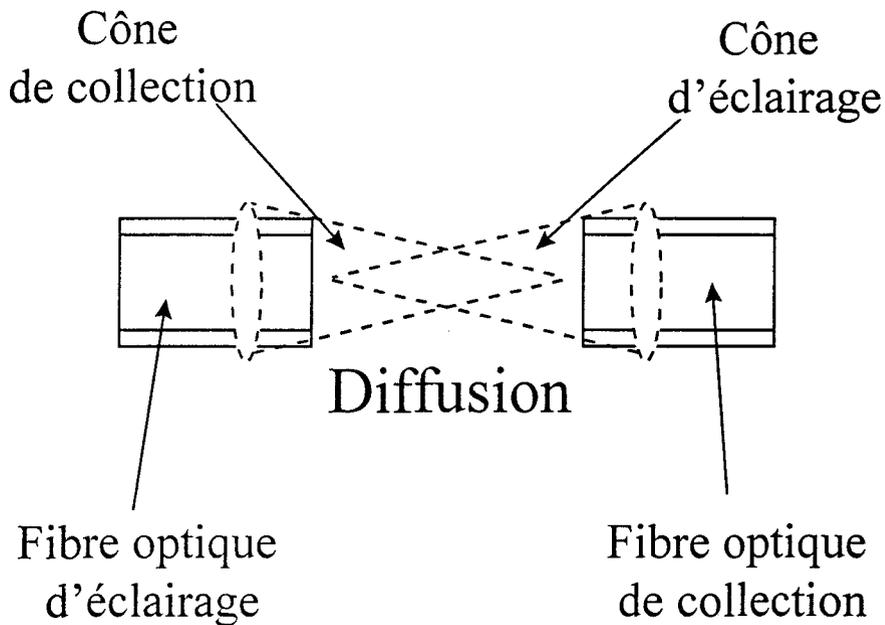


Figure 46 : Défaut de collection de la lumière diffusée en transmission pure.

Nous avons ensuite validé ce choix par une étude systématique du chemin optique. Pour les chemins optiques supérieurs, le rapport signal sur bruit était défavorable au développement d'une analyse quantitative. Les problèmes de linéarité étaient aussi plus importants. Un chemin optique faible permettait d'obtenir une meilleure linéarité. Néanmoins, la répétabilité était trop faible à cause du mauvais rapport signal sur bruit. En effet, de nombreuses bulles pouvaient se coincer dans le chemin optique. Un globule mesurant 20 μm et se déplaçant dans un chemin optique de 200 μm peut aussi faire varier le signal de façon conséquente⁽²⁰⁸⁾. Les problèmes de linéarité viennent principalement du manque d'intégration de la lumière diffusée⁽⁴³⁾. La lumière arrivant sur la fibre de collection est l'énergie qui n'a pas été absorbée par le lait. Or une grande partie de l'énergie est diffusée à l'extérieur du cône de collection. L'information absorbance est donc erronée. Une équipe de chercheurs, spécialisée dans l'analyse NIR de cellules vivantes, a même obtenu des spectres d'absorption comportant des bandes sans origine vibrationnelle⁽²⁰⁹⁾. Ce phénomène parasite a été attribué à la diffusion de la lumière sur ces corps qui était mal collectée. Pour ces cas d'analyses difficiles où la diffusion

est importante, il faut améliorer l'intégration de l'énergie venant de tout l'espace⁽²¹⁰⁾. Ces remarques ont donc motivé la conception de la sonde « transflexion ».

4.9 La sonde «transflexion »

4.9.1 Introduction

La technique de transflexion est largement utilisée pour les analyses de produits diffusants^(123,186,211,212,213). Elle tire son nom de la combinaison de la transmission et de la réflexion. Une configuration type est représentée sur la figure ci-dessous.

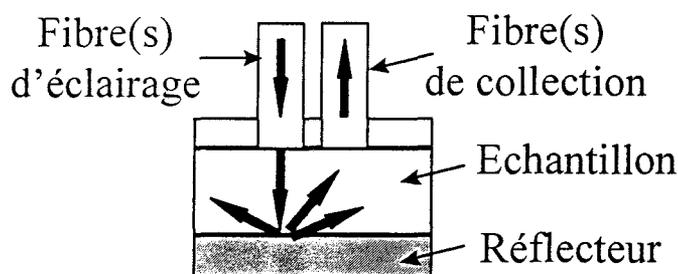


Figure 47 : La transflexion.

L'énergie lumineuse arrive grâce à une ou plusieurs fibres sur l'échantillon à analyser. Elle est tout d'abord transmise à travers celui-ci. Après une réflexion sur un miroir ou un matériau diffusant, la lumière est transmise à une ou plusieurs fibres de collection. La propagation de la lumière dans l'échantillon induit une diffusion qui est aussi récupérée par la ou les fibres de collection. Le premier avantage de cette configuration est la diminution de l'espace de diffusion d'un facteur deux. De cette façon, nous pouvons maîtriser plus facilement sa collection. Mais si on utilise une fibre d'éclairage et une fibre de collection, le problème de l'ouverture numérique reste entier. Pour éviter cela, les constructeurs de spectromètres proche infrarouge proposent une sonde transflexion multifibres. Plusieurs fibres, d'un diamètre de 100 μm par exemple, partent de la source et arrivent à l'échantillon. L'énergie lumineuse est transmise, diffusée vers une surface réfléchissante plane (souvent une pièce en acier inoxydable polie). Cette dernière renvoie l'énergie sur plusieurs fibres de collection. Ces faisceaux multifibres, communément appelés des « bundles », utilisent jusqu'à 300 fibres optiques dont le tiers est utilisé pour l'éclairage alors que le reste collecte l'énergie transmise ou diffusée^(214,215,216,217). Au niveau de la sonde, ces fibres sont distribuées de manière aléatoire⁽²¹⁸⁾ pour atteindre un diamètre total de plus d'un centimètre (Figure 48). Outre une

intégration du flux sur presque tout l'espace, cette sonde a pour avantage d'amener une énergie plus importante venant de la source. En fait, le seul moyen d'augmenter l'ouverture numérique est de multiplier le nombre de fibres optiques.

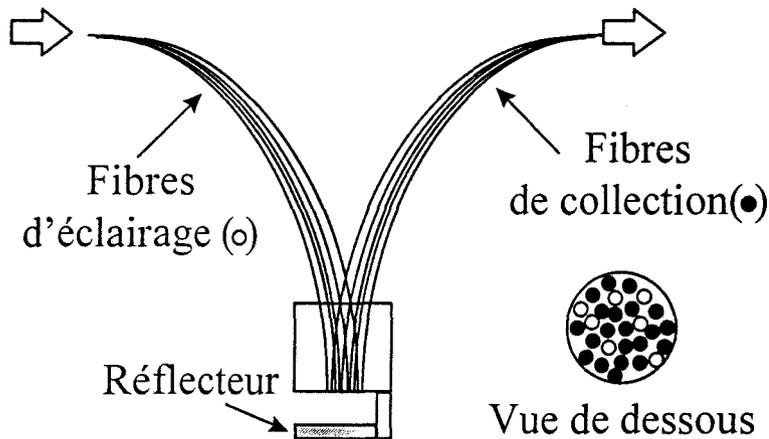


Figure 48 : Sonde à faisceau multifibre.

Ce nombre impressionnant de fibres utilisées donne vite un aperçu du coût d'un tel montage. D'ailleurs leurs longueurs dépassent rarement 1,5 mètres. Ces faisceaux sont aussi très rigides ce qui rend l'implantation impossible en salle de traite. Nous avons donc développé en premier lieu une sonde transflexion avec deux fibres et un réflecteur plan (identique à la Figure 47). Elle avait l'avantage d'être de conception très simple mais la répétabilité et la linéarité n'étaient pas améliorées par rapport à la sonde transmission. Pour un nombre de fibres donné, seul un nouveau réflecteur pouvait améliorer la qualité de la sonde. L'idée était d'utiliser un miroir concave afin d'obtenir l'image de la fibre d'éclairage sur la fibre de collection. La fragilité de la surface de ce nouveau réflecteur nous a obligé à l'isoler du lait. La première version de notre sonde transflexion est représentée sur la Figure 49. Nous retrouvons le système classique d'éclairage à deux fibres. Le lait à analyser circule entre le chemin optique délimité par les fibres et le hublot de verre. Derrière ce dernier, nous avons un espace vide suivi du miroir. Afin d'éviter les problèmes d'étanchéité, de condensation et de contraintes mécaniques, de nombreuses sondes se sont succédées pour arriver à la version définitive de la sonde transflexion (Figure 50).

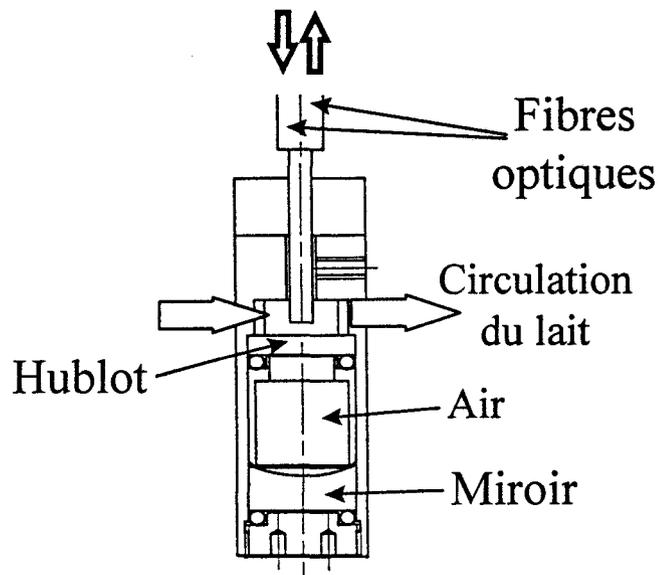


Figure 49 : La sonde transflexion.

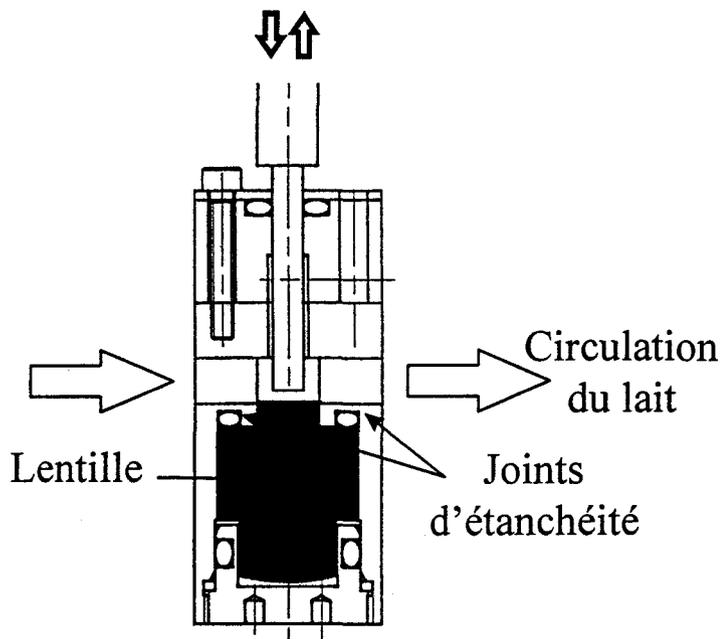


Figure 50 : Configuration finale de la sonde transflexion.

C'est donc le remplacement du hublot et du miroir par une pièce optique unique en verre qui a résolu les différents problèmes.

4.9.2 La répétabilité

Nous testons la répétabilité spectrale sur un échantillon de crème diluée contenant 40 g/kg de matières grasses. Le chemin optique est de 200 μm soit 100 μm entre l'extrémité des fibres et la lentille. En effet, pour la transflexion, il faut considérer que le faisceau lumineux traverse deux fois l'échantillon. Sauf précisions contraires, les spectres utilisés pour l'étude de la sonde transflexion, auront un incrément spectral de 5 nm, un temps d'intégration de 10 ms, et seront issus d'une moyenne de 5 acquisitions successives. Cette technique de la moyenne est souvent utilisée sur les échantillons complexes afin d'obtenir un spectre représentatif de l'échantillon^(26,38). La figure ci-dessous représente les spectres « simple faisceau » pour la référence et un des dix échantillons utilisés pour la répétabilité.

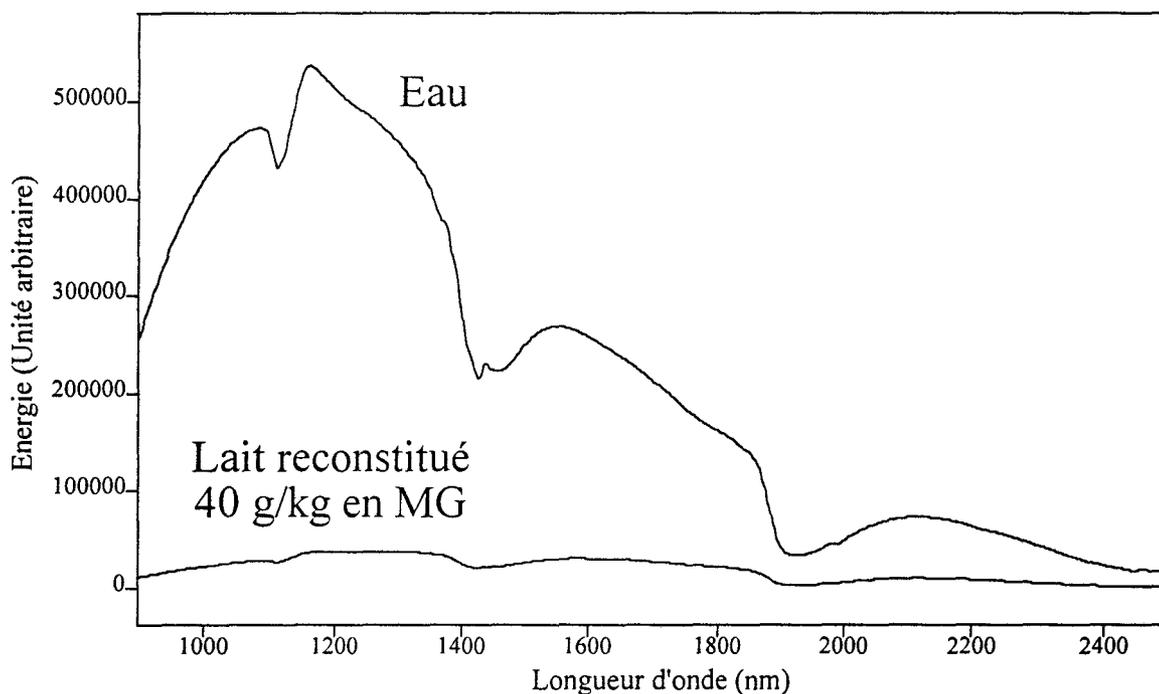


Figure 51 : Les spectres « simple faisceau » référence et échantillon.

Le spectre de l'eau conserve un profil identique à celui présenté pour la sonde « transmission pure ». Les contributions spectrales sont retrouvées. Par contre, le spectre de l'échantillon est très plat. Il y a très peu d'énergie qui arrive jusqu'au détecteur. Pour de tels niveaux de signal, les contributions spectrales sont indiscernables sur les spectres « simple faisceau ». La Figure 52 représente les spectres d'absorbance des échantillons utilisés pour la répétabilité.

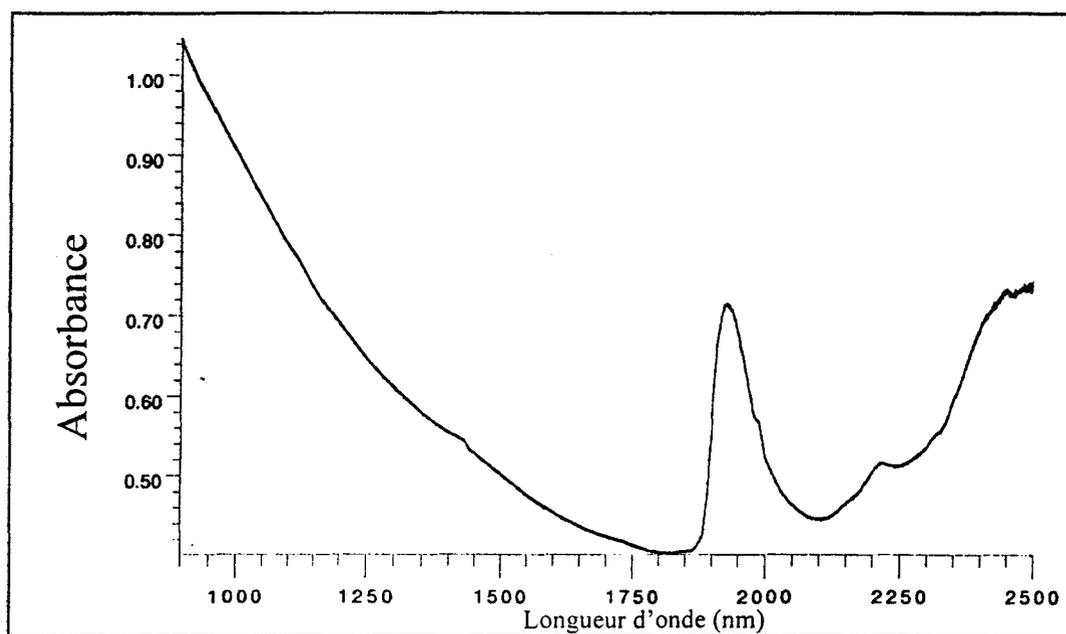


Figure 52 : Superposition des dix spectres utilisés pour la répétabilité.

La ligne de base reste incurvée. La bande d'absorption à 1925 nm est toujours présente. L'absorption située à 1400 nm est plus faible. Nous remarquons une contribution spectrale plus importante aux alentours de 2200 nm attribuable à la troisième harmonique de la vibration d'élongation C-O dans les longues chaînes des acides gras. D'autres absorptions caractéristiques des matières grasses sont visibles à 2305 et 2345 nm. Dans l'ensemble, l'intensité de ces bandes est beaucoup plus faible que celle observée en transmission pure. Ces spectres sont plus en accord avec les données bibliographiques rapportant des travaux sur des milieux aqueux très dilués comme le lait^(199,219,220). Le Tableau 10 donne la répétabilité de la sonde transflexion. Nous observons une légère détérioration par rapport à la sonde transmission pure. Il faut considérer que le niveau d'absorbance et le bruit sont plus importants. Cette sonde semble donner les mêmes signes de faiblesse.

Longueur d'onde (nm)	Répétabilité (RSD %)
2345	0,75
2305	0,74
2200	0,71
1980	0,75
1780	0,51
1750	0,55
1685	0,50
1400	0,34
1185	0,30

Tableau 10 : Répétabilité de la sonde transflexion.

4.9.3 La linéarité

Pour ce contrôle, nous avons préparé une gamme de crèmes liquides diluées proche de celle utilisée pour la sonde transmission pure afin de comparer les résultats plus facilement (Tableau 11). Les spectres d'absorbance sont représentés sur la Figure 53. Nous retrouvons les contributions observées lors de l'étude de la répétabilité. Malgré un faible rapport signal sur bruit au delà de 2200 nm, nous discernons entre autres la contribution des matières grasses à 2305 nm croissante avec la concentration. Une régression linéaire simple entre l'absorbance à 2305 nm et la concentration en matières grasses est représentée à la Figure 54.

Echantillon N°	Concentration MG (g.kg)
1	24.93
2	28.86
3	32.88
4	36.81
5	40.75
6	44.63
7	48.38
8	52.20

Tableau 11 : Gamme de linéarité en matières grasses.

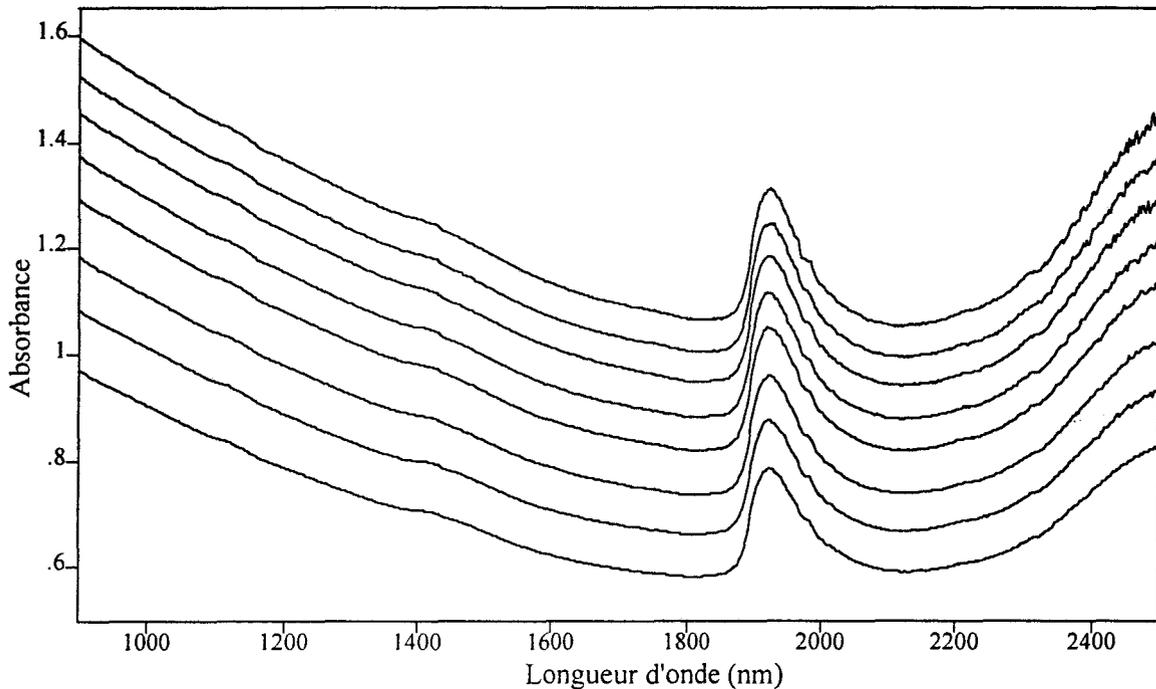


Figure 53 : Spectres d'absorbance pour la gamme de matières grasses.

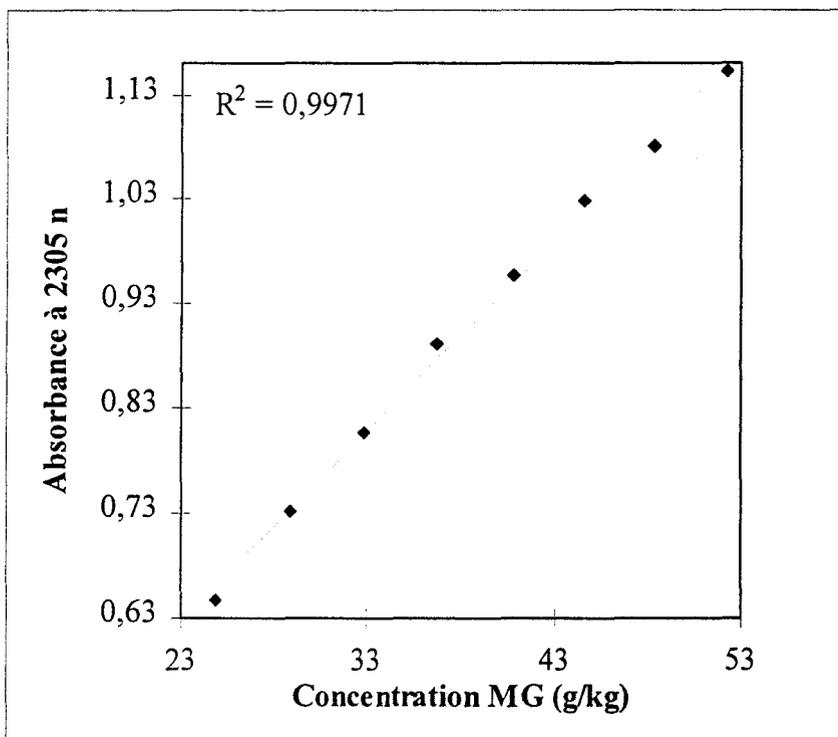


Figure 54 : Linéarité du signal optique à 2305 nm en fonction de la concentration MG.

La droite de régression donne un bon coefficient de corrélation pour la relation absorbance/ concentration. Le profil incurvé observé pour la sonde transmission pure a été largement

diminué. L'étude de la linéarité sur d'autres longueurs d'onde a donné de très bons résultats. Malgré une répétabilité médiocre, la linéarité du signal optique nous incite à étudier la faisabilité de l'analyse quantitative.

4.9.4 L'analyse quantitative

Cette partie permet d'évaluer l'influence de la répétabilité et de la linéarité sur la qualité prédictive en matières grasses et protéines de notre analyseur dans la configuration considérée. Nous traiterons successivement le dosage de solutions modèles en matières grasses, en protéines et finalement les laits crus de vache.

4.9.4.1 Dosage de solutions modèles de matières grasses

Ces solutions sont préparées conformément à la procédure de dilution de crème liquide dans l'eau. Les concentrations de référence sont obtenues à partir de la concentration en matières grasses de la solution mère. Pour l'analyse quantitative, nous préparons quatorze échantillons qui seront séparés en deux groupes pour la calibration et la prédiction (Tableau 12).

Echantillon N°	Utilisation [□]	Concentration MG (g/kg)
1	C	24.80
2	P	26.98
3	C	28.90
4	C	32.70
5	C	36.85
6	P	37.49
7	C	40.71
8	C	44.66
9	P	47.76
10	C	48.42
11	C	52.23
12	C	56.01
13	P	57.79
14	C	59.58

□ : C = Calibration, P = Prédiction.

Tableau 12 : Les solutions modèles de matières grasses.

La régression multivariée est développée à partir de la méthode PLS sur l'ensemble du spectre (900-2500 nm). Il a fallu extraire deux facteurs loading pour générer le modèle. La figure ci-dessous donne de façon classique les résultats de l'analyse quantitative sur les spectres bruts en calibration et en prédiction. La droite $y = x$ (trait noir) représente l'accord parfait entre les concentrations de référence et les concentrations prédites par le modèle.

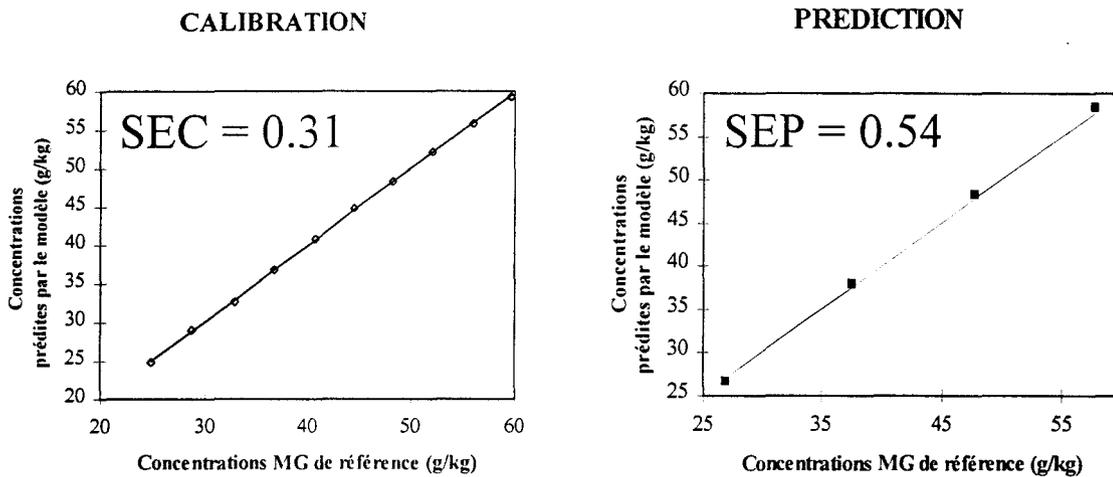


Figure 55 : Dosage des solutions modèles en matières grasses.

Les points sont bien dispersés de part et d'autre de la droite ce qui témoigne d'une absence d'erreur systématique (biais). Le dosage des matières grasses est de bonne qualité car l'erreur standard de prédiction est inférieure à 1 g/kg. Les coefficients de corrélation entre les concentrations de référence et les concentrations prédites pour la calibration et la prédiction sont proches de 0,99. Comme pour toute analyse quantitative, nous avons tenté d'améliorer les résultats par l'application de pré-traitements aux spectres. Tous ces algorithmes sont restés sans effet sur la qualité des dosages.

L'observation des loading est une pratique courante lors de la mise en place d'une analyse quantitative car ils représentent les contributions spectrales prises en compte dans le modèle^(221,222,223,224). La Figure 56 représente le premier loading pour le dosage des matières grasses à partir des spectres bruts. Il ne laisse pas apparaître de zones d'absorption particulières mais il est en tout point semblable à un spectre d'absorbance utilisé pour l'analyse (cf. Figure 53). Ce premier vecteur, exprimant 95 % de la variance totale, modélise la diffusion parasite qui se traduit par une variation globale de la ligne de base du spectre.

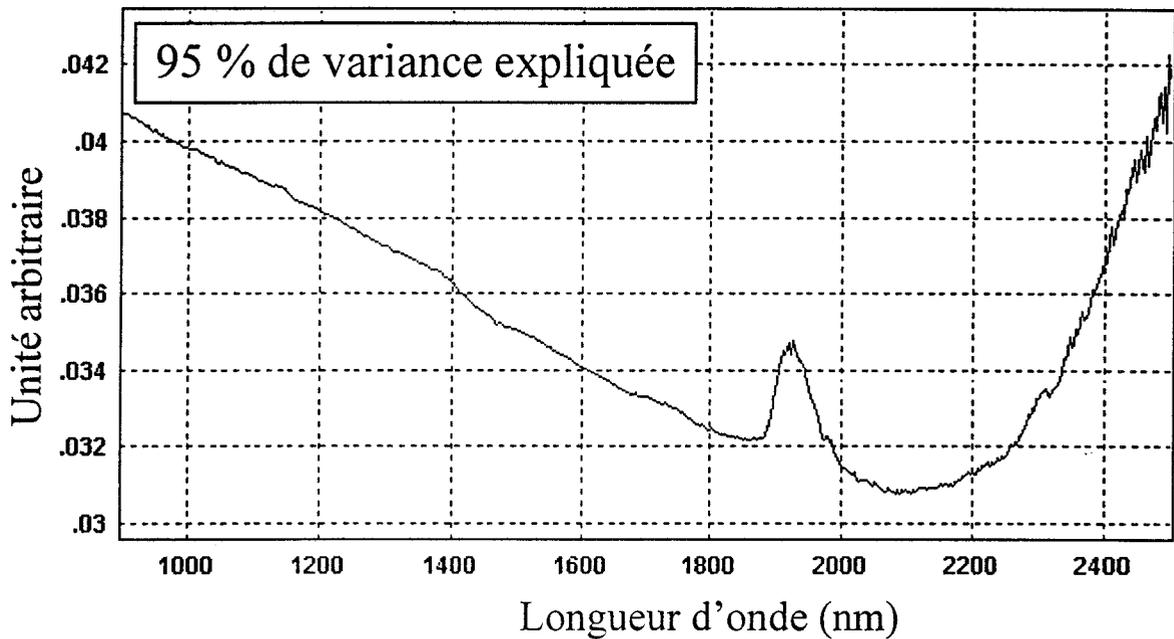


Figure 56 : Premier loading utilisé pour construire le modèle multivarié.

La discrimination des échantillons ne se fait pas sur des informations chimiques distinctes mais plutôt sur les propriétés optiques de l'échantillon et en particulier la diffusion de la lumière sur les globules gras. Les résultats du dosage des matières grasses sur ces solutions modèles sont acceptables.

4.9.4.2 Dosage de solutions modèles de protéines

Nous préparons un lot de 14 échantillons par dissolution de poudre de lait dans l'eau. La concentration en protéine couvre le domaine 25-40 g/kg. Comme pour le dosage des matières grasses, ce lot est divisé en deux pour la calibration et la prédiction. Les spectres utilisés pour la calibration sont représentés sur la Figure 57. Nous remarquons un rapport signal sur bruit médiocre et un niveau d'absorbance très faible. Sur des domaines de concentrations semblables, les protéines induiront donc des variations d'absorbance plus faibles que celles observées sur les matières grasses. Un « fond de diffusion » est encore présent sur les spectres mais son intensité est moindre.

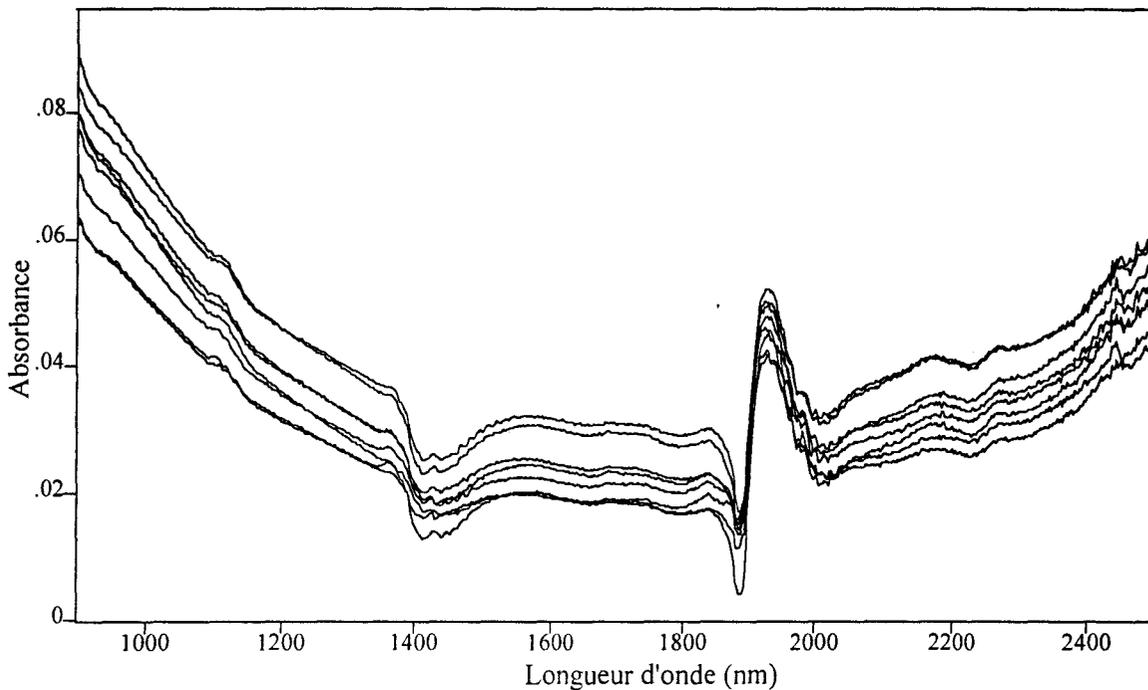


Figure 57 : Spectres de poudre de lait diluée.

Il a fallu sélectionner 4 facteurs loading pour construire le modèle multivarié à partir des spectres bruts (Figure 58). Les deux premiers facteurs possèdent une ligne de base incurvée comparable à celle d'un spectre en absorbance. Comme pour les matières grasses, ils modélisent la variation générale occasionnée par la diffusion. Sur le troisième facteur, nous observons les contributions chimiques prises en compte par le modèle à 2430, 1982, 1896, 1394 et 1102 nm. Ces absorptions sont attribuables aux vibrations des liaisons CN, CO et NH dans les protéines. Le modèle tire une information aux alentours de 1400 nm sur le quatrième facteur mais le bruit commence à être important sur l'ensemble du spectre. Le dernier facteur qui modélise du bruit n'est pas pris en compte par le modèle. L'analyse quantitative des protéines à partir des spectres bruts donne des résultats satisfaisants (Figure 59). Néanmoins, un système aussi simple (dissolution de poudre de lait dans l'eau) devrait être modélisé par un faible nombre de loading. Le modèle construit pour le dosage des protéines caractérise donc bien la difficulté d'extraire une information spectrale faible en intensité dans un bruit important. Lorsque les spectres sont dérivés, nous obtenons d'ailleurs une amplification des erreurs standards de prédiction (Tableau 13).

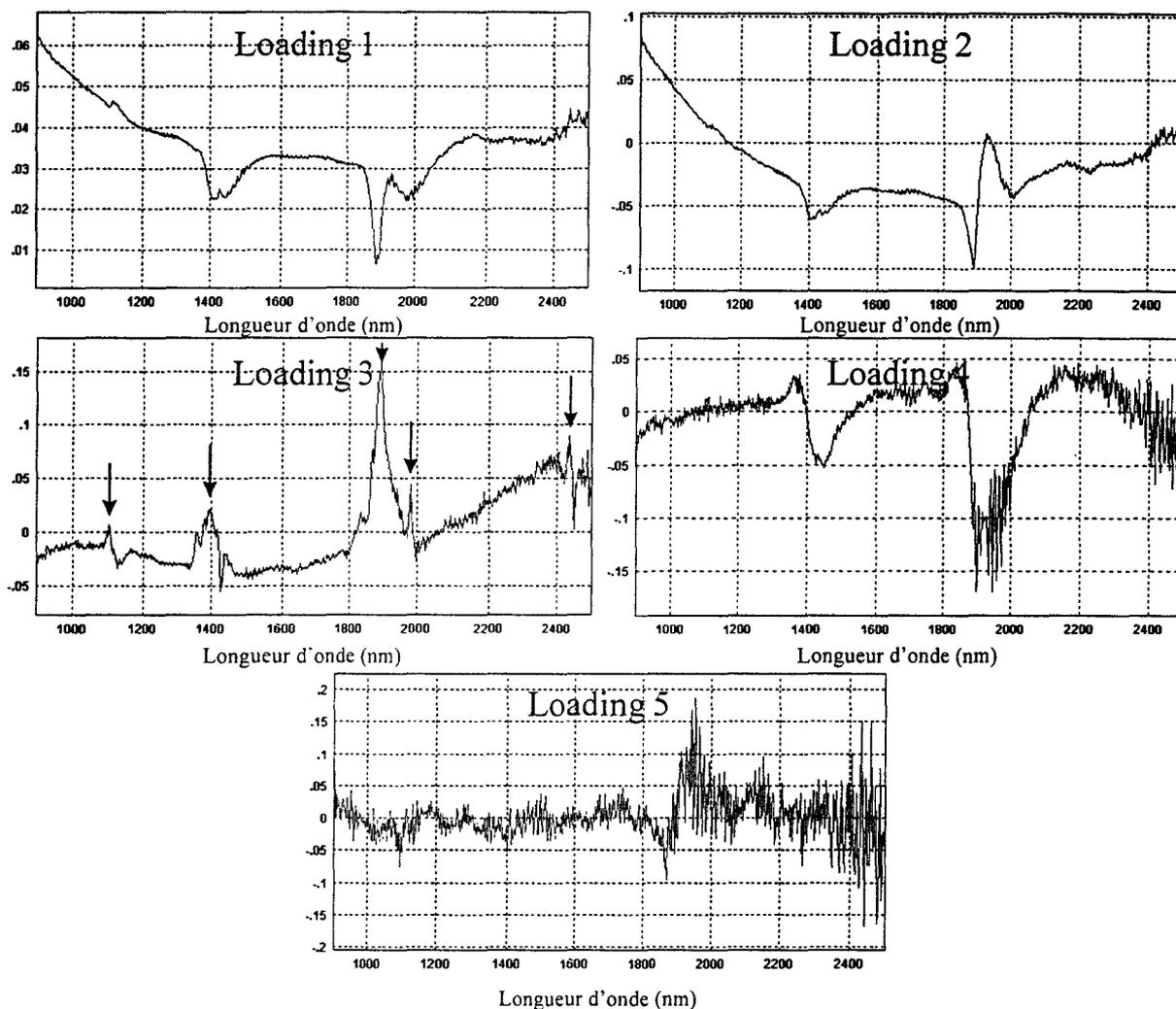


Figure 58 : Facteurs loading extraits pour le dosage des protéines.

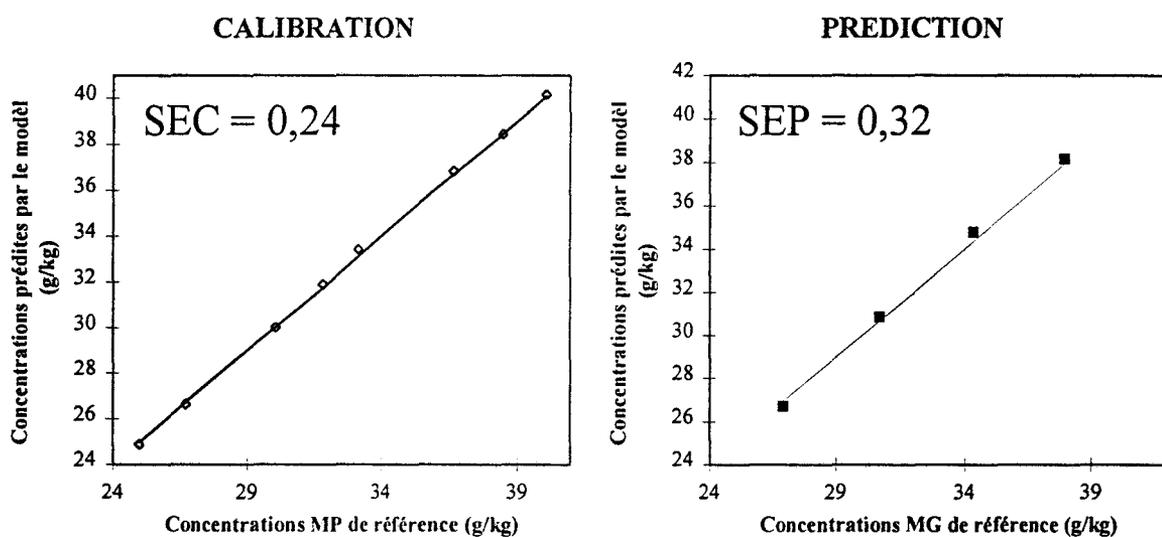


Figure 59 : Analyse quantitative des protéines à partir des spectres bruts.

Pré-traitement	SEC (g/kg)	SEP (g/kg)
Aucun	0.24	0.32
Dérivée 1ère (type "SAVITSKY-GOLAY")	0.60	1.00
Dérivée 2ème (type "SAVITSKY-GOLAY")	1.15	1.55

Tableau 13 : Influence du pré-traitement des spectres sur les erreurs standards.

Les solutions modèles de protéines sont dosées correctement. Les calibrations restent néanmoins très fragiles lors de l'application d'une différentiation.

4.9.4.3 Dosage de laits crus

Cette étape constitue le dosage de laits crus individuels tel que nous le pratiquerons en routine sur l'analyseur. Les laits proviennent d'un même troupeau de bovins. Les concentrations de référence en matières grasses et protéines sont données par un appareil automatique de dosage utilisant la spectrométrie MIR. Afin de tester la faisabilité de l'analyse quantitative, nous avons utilisé 118 échantillons de laits. Une première calibration a permis de détecter 20 échantillons dits aberrants (*outlier*) car leurs données spectrales s'écartaient significativement des autres. Pour la construction d'une calibration, il est fréquent de retirer quelques éléments du lot de données⁽¹³⁵⁾. En effet, ces éléments présents dans le modèle détériorent les erreurs standards de calibration et surtout les erreurs standards de prédiction. Néanmoins, la suppression d'un trop grand nombre d'éléments aberrants n'est pas acceptable pour l'établissement d'une bonne analyse quantitative. Les résultats qui suivent ne sont donc pas représentatifs des conditions réelles d'utilisation de l'analyseur. Ils permettront dans cette phase exploratoire de donner des tendances. La Figure 60 représente 30 des 98 spectres utilisés pour l'analyse quantitative. Le profil des spectres est très proche des solutions modèles de matières grasses. On observe toujours une ligne de base incurvée et une variation du niveau global du spectre. Le bruit est particulièrement important au-dessus de 2200 nm. Il ne provient pas d'un signal plus faible qui arrive sur le détecteur mais plutôt de la nature hétérogène du lait cru. En effet, les structures globulaires et micellaires n'étaient pas ou peu présentes dans les solutions modèles de matières grasses et de protéines. Leurs passages dans le chemin optique est un facteur important de variation de signal. Pour l'analyse quantitative, le lot d'échantillons a été divisé en deux groupes soit 88 pour la calibration et 10 pour la prédiction.

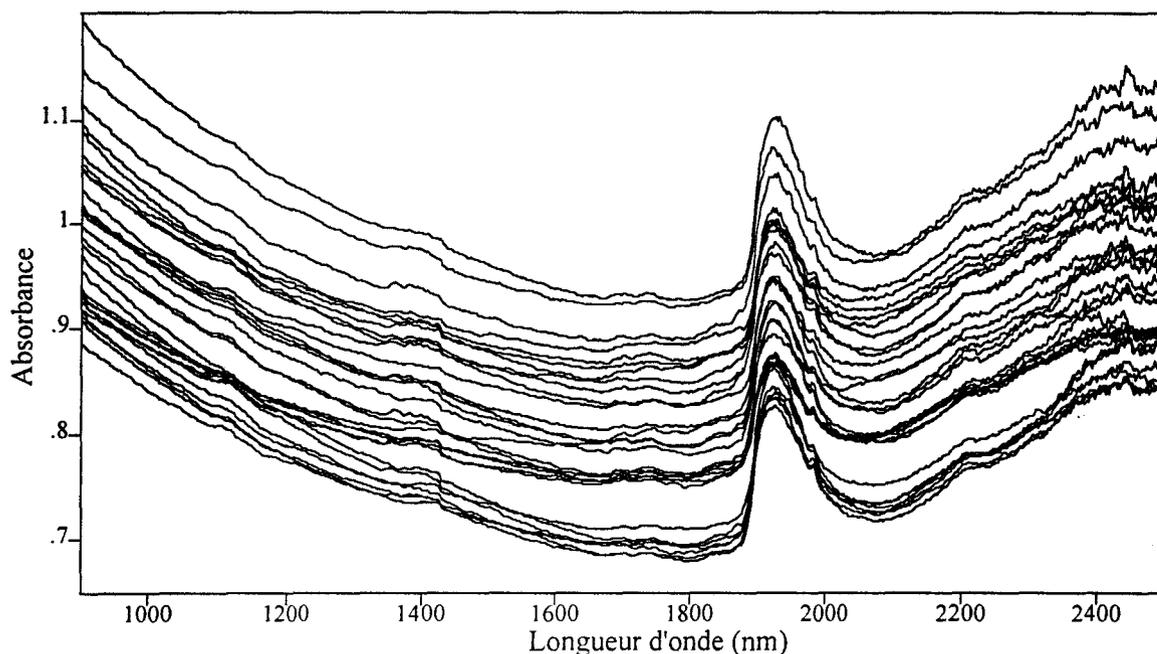


Figure 60 : Spectres de laits crus.

Dans cette première partie de l'étude, nous avons exploité l'ensemble du domaine spectral (900-2500 nm). Les résultats des calibrations et des prédictions sont présentés ci-dessous (Tableau 14). Le dosage des matières grasses à partir des spectres bruts est relativement satisfaisant.

MATIERES GRASSES							
Pré-traitement	CALIBRATION				PREDICTION		
	Loading	SEC (g/kg)	r2	Biais	SEP (g/kg)	r2	Biais
Aucun	5	0.65	0.99	<10e-2	1.05	0.94	-0.34
Dérivée 1ère (type "SAVITSKY-GOLAY")	3	1.15	0.95	<10e-2	1.16	0.93	0.06
Dérivée 2ème (type "SAVITSKY-GOLAY")	6	1.48	0.94	<10e-2	1.88	0.83	0.62

MATIERES PROTEIQUES							
Pré-traitement	CALIBRATION				PREDICTION		
	Loading	SEC (g/kg)	r2	Biais	SEP (g/kg)	r2	Biais
Aucun	5	1.26	0.74	<10e-2	1.27	0.63	0.08
Dérivée 1ère (type "SAVITSKY-GOLAY")	3	1.30	0.57	<10e-2	1.32	0.56	0.02
Dérivée 2ème (type "SAVITSKY-GOLAY")	6	1.10	0.80	<10e-2	1.15	0.83	0.56

Tableau 14 : Dosage des matières grasses et des protéines dans les laits crus.

(temps d'intégration : 0,0010 s)

Nous remarquons néanmoins un coefficient de corrélation qui s'écarte légèrement de 1 ainsi que la présence d'un biais (erreur systématique) en prédiction. Le dosage des protéines est inacceptable. Les faibles coefficients de corrélation pour la calibration et la prédiction témoignent d'une relation très fragile entre les concentrations de référence et les concentrations prédites par le modèle (Figure 61).

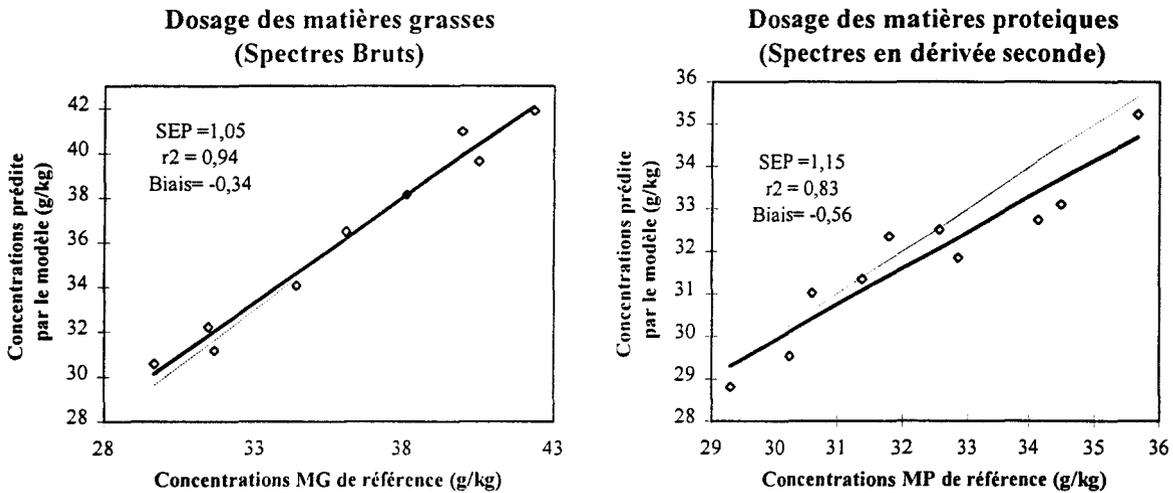


Figure 61 : Observation des corrélations et des biais pour le dosage des laits crus.

En spectroscopie NIR, les bandes d'absorption les plus intenses et les plus fines sont situées aux longueurs d'onde les plus grandes. En fait, les extinctions diminuent rapidement avec la longueur d'onde. L'utilisation du domaine spectral 900-2500 nm semble donc contradictoire dans le sens où les faibles longueurs d'onde ne doivent pas apporter d'informations significatives. Leur utilisation dans le modèle devrait donc détériorer les erreurs de prédiction. Afin de vérifier cet effet, nous avons effectué une étude systématique sur l'influence du domaine spectral pour l'analyse quantitative de ces laits (Tableau 15). Les résultats du dosage des protéines ne sont pas présentés car les corrélations encore très faibles nous empêchent de comparer les erreurs standards de prédiction. Contrairement à ce que l'on pouvait attendre, l'utilisation du domaine 1700-2500 nm ne donne pas une erreur standard de prédiction plus faible. La suppression de cette zone pour l'analyse quantitative permet même d'obtenir une analyse quantitative de qualité sur le domaine 900-1700 nm. Le rapport signal sur bruit est tellement défavorable au-dessus de 2000 nm que les très faibles bandes d'absorption en

dessous de 1700 nm donnent une analyse quantitative de meilleure qualité. L'amélioration des dosages passe donc par une augmentation du rapport signal sur bruit.

MATIERES GRASSES
(Spectres bruts)

Domaine spectral	CALIBRATION				PREDICTION		
	Loading	SEC (g/kg)	r2	Biais	SEP (g/kg)	r2	Biais
900-2500 nm	5	0.65	0.99	<10e-2	1.05	0.94	-0.34
1700-2500 nm	4	0.80	0.98	<10e-2	1.31	0.92	-0.53
900-1700 nm	5	0.78	0.98	<10e-2	1.04	0.95	-0.28

Tableau 15 : Influence du domaine spectral utilisé pour l'analyse quantitative.

Un temps d'intégration de 1 seconde a permis d'atteindre cet objectif. Nous avons néanmoins dû restreindre le domaine spectral (1700-2500 nm) et le nombre d'accumulations (deux par échantillon) afin de conserver un temps d'analyse compatible avec la manipulation précédente. L'influence du temps d'intégration sur le rapport signal sur bruit est observable sur les spectres d'un même échantillon de lait cru contenant 36 g/kg en MG et 31 g/kg en MP (Figure 62).

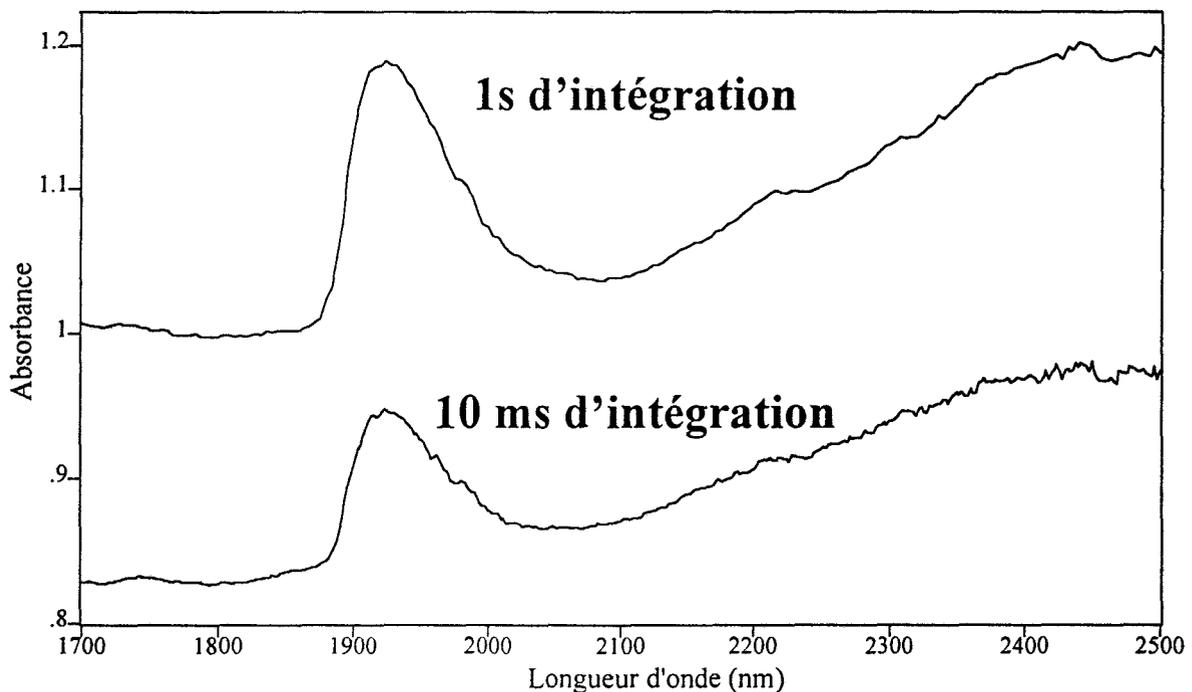


Figure 62 : Influence du temps d'intégration sur la qualité spectrale.

Cette nouvelle analyse quantitative a été développée sur 36 échantillons de lait cru soit 26 pour la calibration et 10 pour la prédiction. Aucun échantillon n'a été détecté comme élément aberrant. Les résultats sont consignés dans le Tableau 16. Nous observons une nette amélioration du dosage des matières grasses à partir des spectres bruts. Le dosage des protéines n'est pas satisfaisant.

Un autre moyen d'améliorer l'analyse quantitative était de mieux exploiter le domaine spectral 900-1700 nm par une augmentation du chemin optique. De cette façon, les absorptions sont plus importantes aux longueurs d'onde faibles et donc plus facilement détectables. Néanmoins, l'analyse spectrale aux longueurs d'onde élevées (> 2000 nm) devient impossible car l'énergie est complètement absorbée par l'échantillon.

MATIERES GRASSES							
Pré-traitement	CALIBRATION				PREDICTION		
	Loading	SEC (g/kg)	r2	Biais	SEP (g/kg)	r2	Biais
Aucun	6	0.29	0.99	<10e-2	0.62	0.99	-0.25
Dérivée 1ère (type "SAVITSKY-GOLAY")	5	4.56	0.18	<10e-2	5.35	0.5	4.97
Dérivée 2ème (type "SAVITSKY-GOLAY")	2	2.13	0.8	<10e-2	2.56	0.34	0.76

MATIERES PROTEIQUES							
Pré-traitement	CALIBRATION				PREDICTION		
	Loading	SEC (g/kg)	r2	Biais	SEP (g/kg)	r2	Biais
Aucun	2	0.55	0.93	<10e-2	1.55	0.42	-0.18
Dérivée 1ère (type "SAVITSKY-GOLAY")	4	1.05	0.64	<10e-2	1.13	0.62	-0.44
Dérivée 2ème (type "SAVITSKY-GOLAY")	2	1.01	0.67	<10e-2	1.04	0.28	-0.14

**Tableau 16 : Dosage des matières grasses et des protéines dans les laits crus.
(temps d'intégration : 1 s)**

Nous avons donc mis en place une analyse quantitative sur le domaine 900-1700 nm avec un temps d'intégration de 10 ms sur chaque élément spectral et un chemin optique de 1200 µm. Les résultats sont comparables à ceux présentés dans le Tableau 14. Le dosage des matières grasses n'est possible qu'à partir des spectres bruts (SEP = 1 g/kg). La différentiation n'améliore toujours pas les résultats. Le cas des protéines reste inchangé puisque les corrélations sont très faibles et les erreurs standards de prédiction très importantes. Devant ces résultats similaires sur les domaines 900-1700 nm (chemin optique 1200 µm) et 1700-2500 nm (chemin optique 200 µm), il est apparu intéressant de développer une sonde pouvant

exploiter simultanément les deux domaines spectraux. Ce dispositif original est appelé « sonde à chemin optique variable » (Figure 63). Elle conserve le mode transflexion. Le lait circule entre deux lentilles convexes. Le chemin optique est de 200 μm au centre et 1200 μm aux extrémités. De cette façon, nous avons un chemin optique suffisamment faible pour ne pas avoir un signal saturé aux longueurs d'onde élevées, et un chemin optique suffisamment important pour discerner des absorptions aux longueurs d'onde faibles (extinctions faibles).

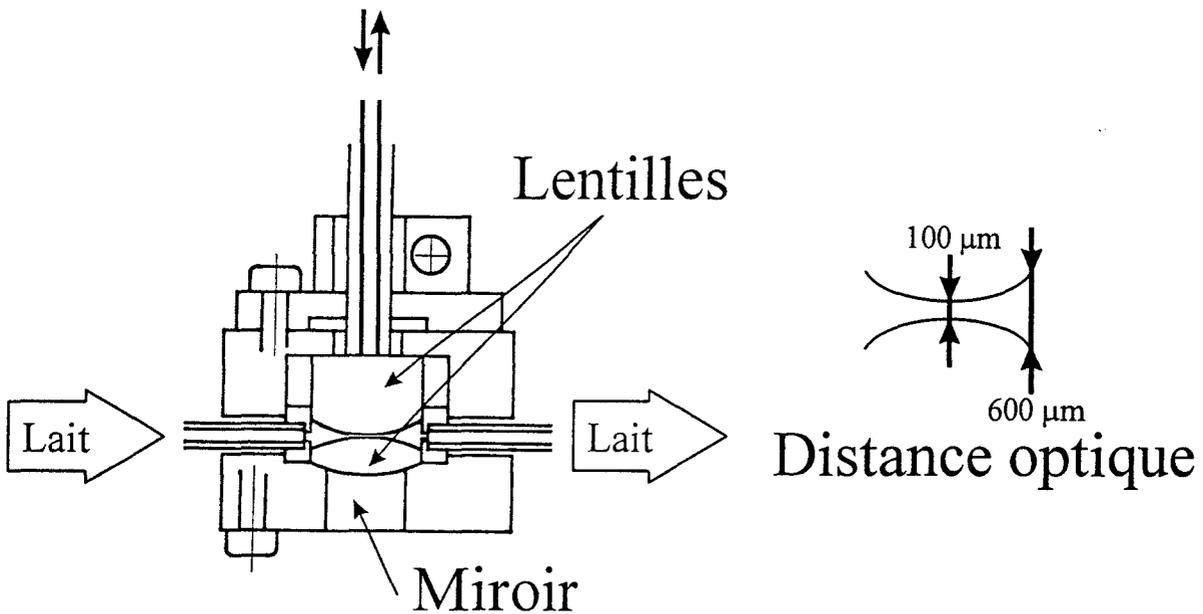


Figure 63 : Sonde à chemin optique variable.

La répétabilité a été évaluée à partir d'une solution modèle de protéines. Les spectres obtenus sont présentés sur la Figure 64. Leur profil est très perturbé par un bruit important apporté par la circulation du lait. En effet, la répétabilité est améliorée lorsque le débit diminue mais se détériore à nouveau si la circulation du lait est trop faible (migrations des globules dans le chemin optique). Au niveau du rétrécissement, il y a en fait création d'une surpression en amont et d'une dépression en aval du lait qui circule.

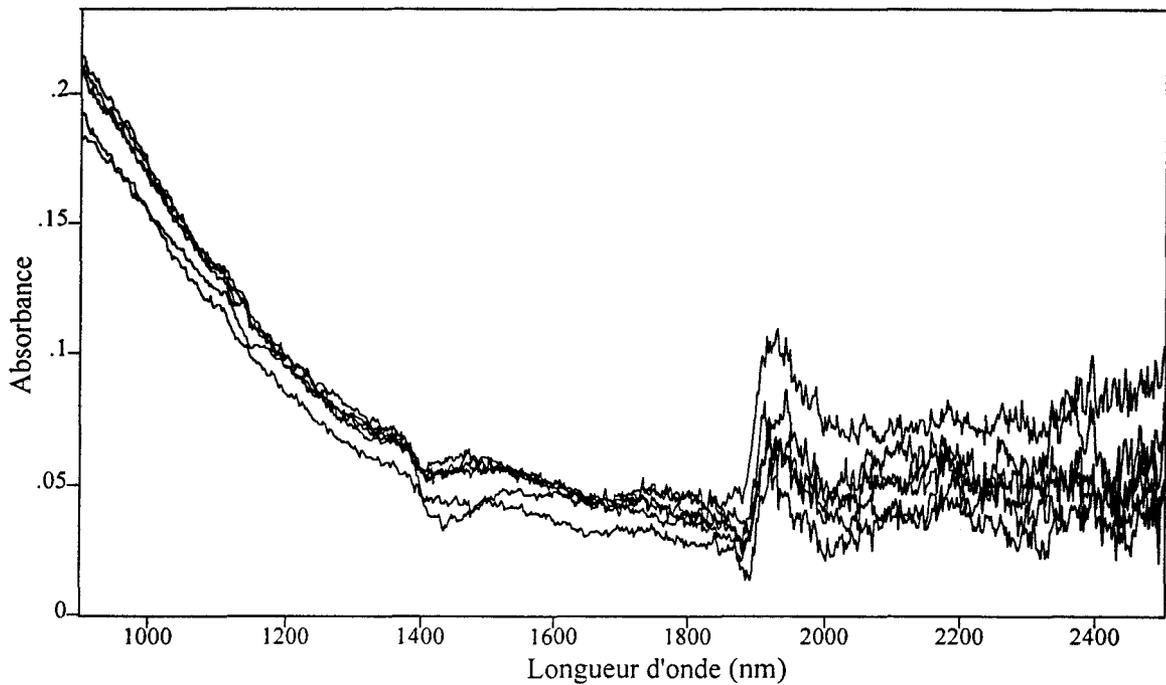


Figure 64 : Spectres de la solution modèle de protéine utilisée pour évaluer la répétabilité.

La mécanique des fluides montre qu'il y a accélération du liquide. L'écoulement n'est plus laminaire mais turbulent (Figure 65). Des changements locaux d'indice de réfraction sont observables⁽⁷⁸⁾.

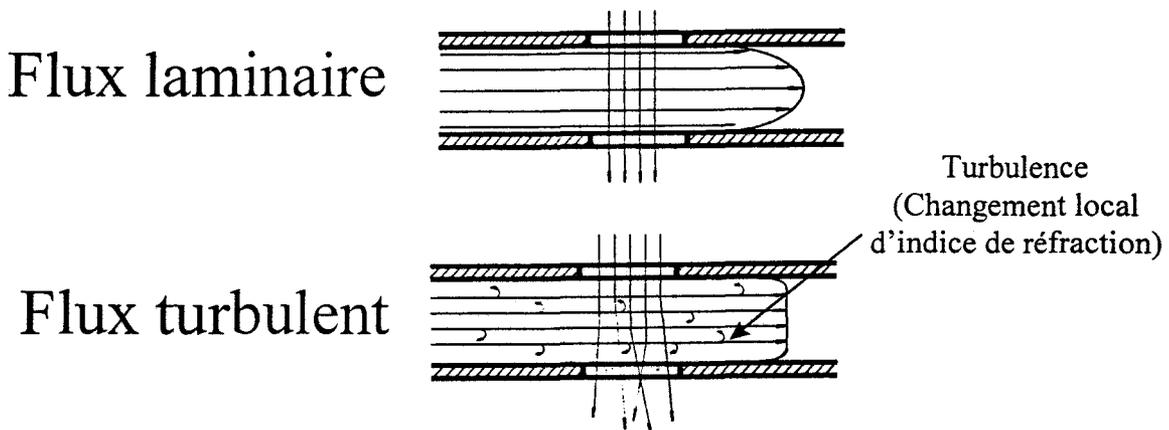


Figure 65 : Propriétés optiques des flux

Le chemin optique n'est donc pas constant. L'analyse en ligne par spectrométrie NIR demande d'ailleurs obligatoirement des systèmes de compensation des variations de pression

dans les conduits^(225,226). La qualité spectrale reste insuffisante pour mener une analyse quantitative.

En résumé, la première sonde transflexion a montré son potentiel pour l'analyse de solutions relativement homogènes telles que les dilutions de matières grasses et protéines. Sur les laits de ferme, la diffusion occasionnée par les nombreux globules gras est très importante. Le signal collecté par la fibre est trop faible comparé à l'énergie diffusée dans les trois directions de l'espace. Le rapport signal sur bruit est donc défavorable. Il faut obtenir une meilleure intégration du signal afin de caractériser la diffusion globale de l'échantillon.

4.10 La sonde 0-45°

4.10.1 Introduction

Sur la base d'un nombre donné de fibres optiques de collection, il n'existe pas de système optique qui permette d'augmenter l'ouverture numérique de la sonde. Le seul moyen est donc d'augmenter le nombre de fibres. Dans cette nouvelle configuration, la fibre optique centrale éclaire l'échantillon avec une lumière monochromatique (Figure 66).

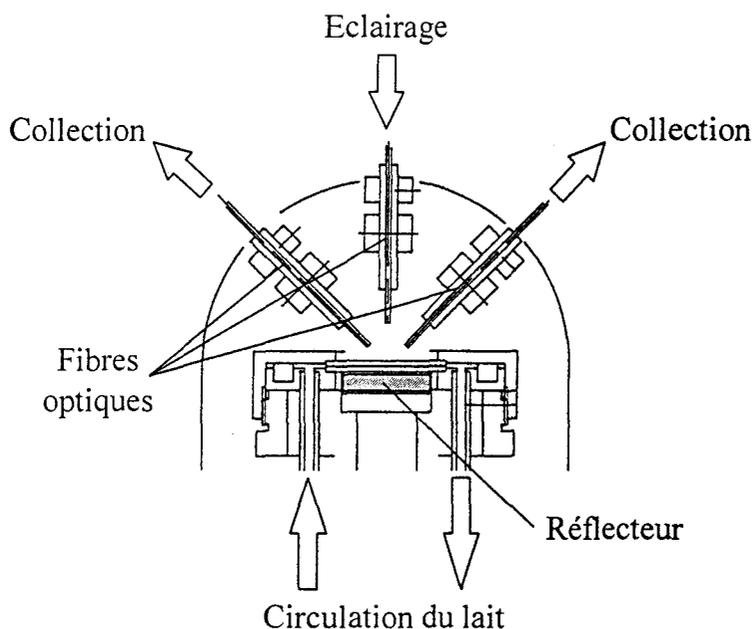


Figure 66 : La sonde 0-45°.

Le lait circule entre deux plaques de verre. L'énergie qui arrive jusqu'au réflecteur est rétro-propagée vers le demi-espace où se situent les deux fibres de collection. Contrairement aux sondes précédentes, le réflecteur est un polymère hautement réfléchissant (99 % de réflectance) appelé Spectralon™. Ce matériau recouvre la surface intérieure des sphères d'intégration récentes⁽²²⁷⁾. Nous nous sommes inspirés de la géométrie 0-45° (page 50) en remplaçant les détecteurs par les deux fibres optiques de collection. Mise à part le nombre de fibres optiques utilisées, l'inclinaison de 45° permet aussi de couvrir une zone plus importante au niveau de l'échantillon (Figure 67).

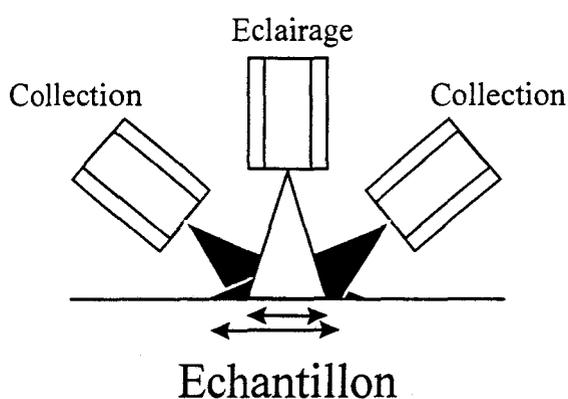


Figure 67 : Effet de l'inclinaison des fibres optiques sur la zone d'analyse.

Le principal défaut de la sonde transflexion était son développement dans l'esprit de l'optique géométrique où l'image de la fibre d'éclairage était la fibre de collection. Ce principe est intéressant sur les échantillons homogènes mais aberrant sur des échantillons aussi diffusants que le lait. Le gain en signal optique obtenu avec la sonde 0-45° est présenté sur la Figure 68. L'échantillon de lait est une solution modèle de matières grasses à 50 g/kg. Les paramètres d'acquisition sont identiques pour les deux sondes (10 ms d'intégration et 5 nm d'incrément spectral). Le chemin optique est de 200 μm . Les rapports d'intensité sont calculés à 2130 nm (longueur d'onde arbitraire) à partir des spectres simple faisceau. Si nous comparons le niveau de signal entre l'eau et le lait pour une sonde donnée, nous voyons que le rapport est plus intéressant sur la sonde 0-45°. En effet, il est plus difficile d'optimiser une chaîne électronique d'acquisition sur une dynamique de signal large (une décade pour la sonde transflexion). La collection de la diffusion est nettement améliorée puisque sur un échantillon peu diffusant comme l'eau, le gain est de 5 alors que sur le lait il est de 30. Ces observations induisent

immédiatement une meilleure qualité spectrale (Figure 69). Des bandes d'absorption apparaissent à 2305 et 2345 nm. Elles sont attribuables aux deuxièmes harmoniques des liaisons CH. Elles étaient faibles ou même masquées par le bruit avec la sonde transflexion. Sur l'ensemble du domaine spectral, la sonde 0-45° permet d'avoir un bon rapport signal sur bruit.

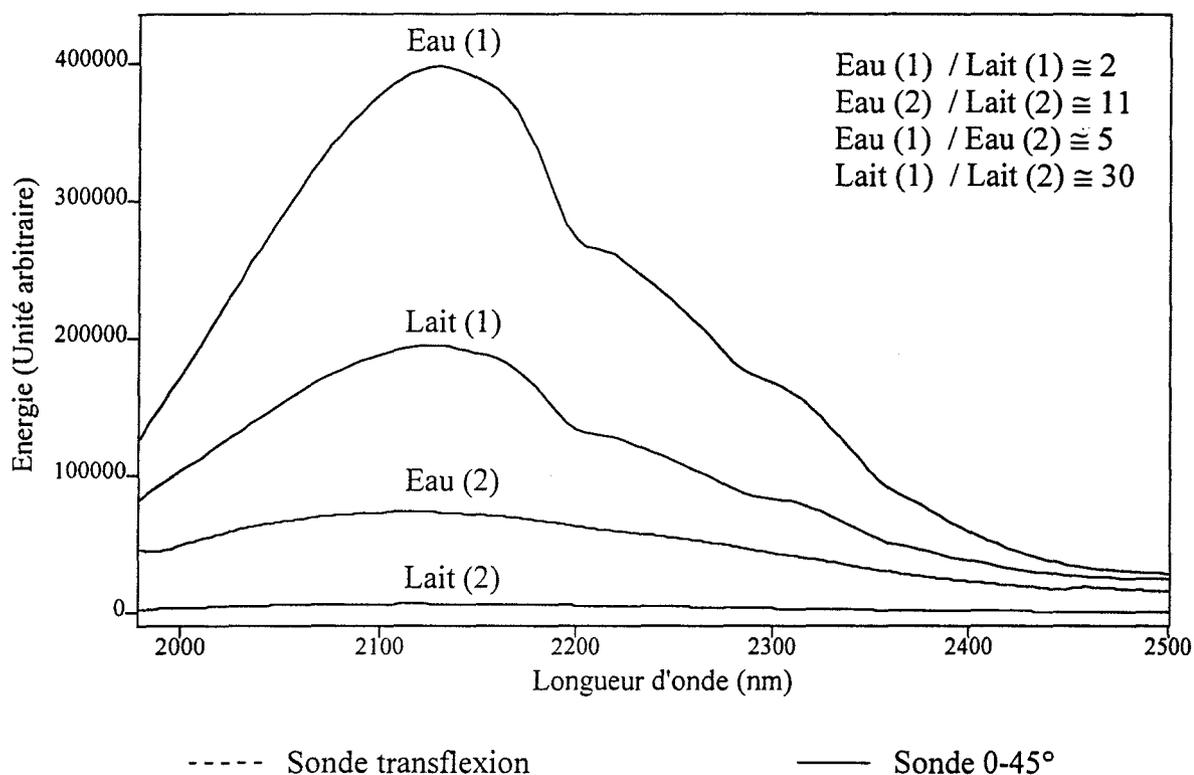


Figure 68 : Comparaison du signal optique des sondes transflexion et 0-45°.

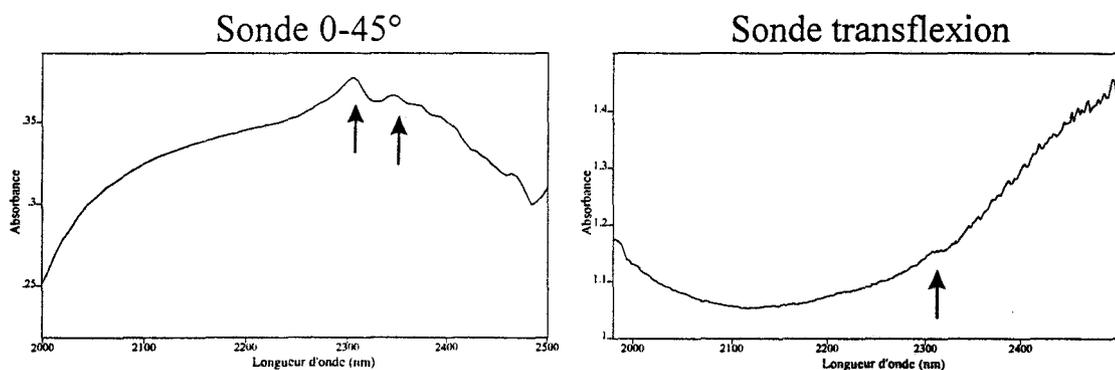
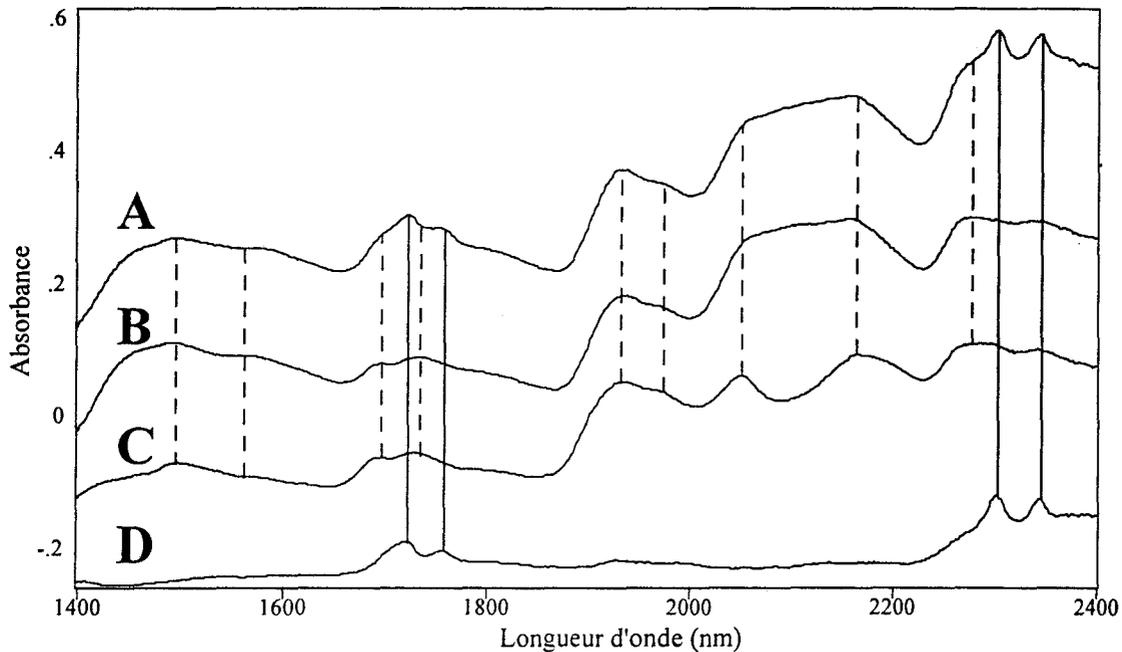


Figure 69 : Spectres d'absorption d'un échantillon sur les sondes transflexion et 0-45°.

Au vu de ces dernières données, il est apparu intéressant d'acquérir des spectres d'absorbance des composés de référence tels que la caséine, les poudres de lait demi-écrémé et écrémé

(Figure 70). Pour cela, il suffit de placer la poudre entre les deux plaques de verre. Le spectralon™ est toujours utilisé comme réflecteur. La sonde vide constitue la référence. Le chemin optique et les paramètres d'acquisition restent inchangés.



A : lait en poudre demi-écrémé C : caséine en poudre
 B : lait en poudre écrémé D : lait en poudre demi-écrémé - Lait en poudre écrémé

Figure 70 : Spectres de poudre de lait et de caséine.

Les spectres obtenus sont strictement identiques à ceux que l'on peut trouver dans la littérature⁽³⁸⁾. Les spectres de caséine et de poudre de lait écrémé sont très proches. Il ne faut pas oublier que les caséines constituent 80% des protéines totales dans le lait. Les différences aux alentours de 2100 et 1500 nm sont attribuables au lactose présent dans le lait en poudre. La différence des spectres de poudre de lait demi-écrémé et écrémé permet d'observer les contributions des matières grasses. Ces bandes d'absorption sont fines et intenses comparées à celles des protéines.

La répétabilité a été évaluée à partir d'une solution de matières grasses diluées à 40g/kg. Les résultats sont consignés dans le tableau ci-dessous. Nous observons une grande amélioration en partie due au gain de rapport signal sur bruit. La linéarité du signal optique en fonction de la concentration a été validée. Les résultats obtenus sont de bonne qualité. Ils sont semblables à la représentation donnée pour la sonde transflexion (Figure 54 page 110). Conformément à

la procédure suivie, ces premiers résultats nous autorisent à poursuivre la validation par l'analyse quantitative de solutions modèles.

Longueur d'onde (nm)	Répétabilité (RSD %)
2345	0,25
2305	0,22
2200	0,29

Tableau 17 : Répétabilité de la sonde 0-45°.

4.10.2 Analyse quantitative de solutions modèles

Afin de simplifier la procédure, nous avons préparé 16 solutions modèles à teneurs variables en matières grasses et en protéines. De cette façon, nous pouvons étudier le dosage des deux espèces sur le même lot d'échantillon. Pour chaque analyse quantitative, nous utilisons 12 échantillons pour la calibration et 4 pour la prédiction (Figure 71).

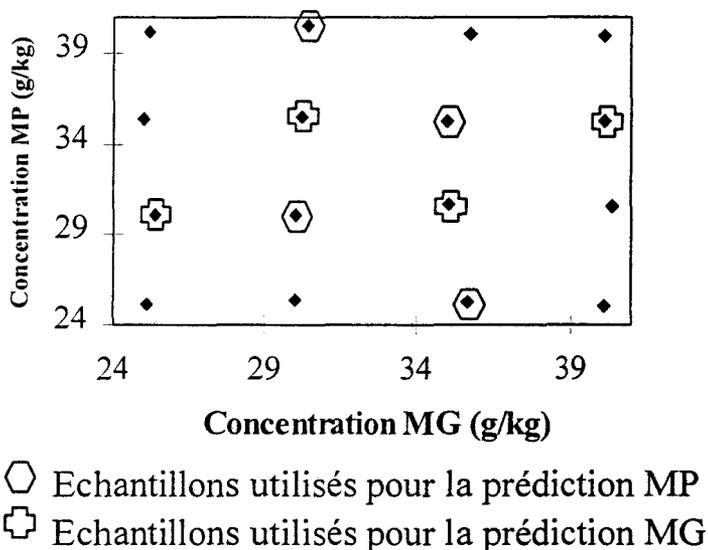


Figure 71 : Utilisation des échantillons pour la calibration et la prédiction.

L'acquisition s'effectue sur le domaine 1980-2500 nm avec une seconde de temps d'intégration sur chaque élément spectral et un chemin optique de 200 µm. Le spectre utilisé est la moyenne de trois acquisitions. Les meilleurs résultats sont obtenus à partir des spectres bruts (Tableau 18). Les erreurs sont du même ordre que celles présentées pour les solutions

modèles avec la sonde transflexion. Nous pouvons ainsi considérer que la sonde 0-45° est plus efficace car les solutions sont plus complexes. L'amplification des erreurs de prédiction par la différentiation est toujours présente mais son intensité diminue. Nous commençons donc à apercevoir les effets du rapport signal sur bruit sur les spectres dérivés.

MATIERES GRASSES

Pré-traitement	CALIBRATION			PREDICTION	
	Loading	SEC (g/kg)	r2	SEP (g/kg)	r2
Aucun	2	0.45	0.97	0.52	0.99
Dérivée 1ère (type "SAVITSKY-GOLAY")	1	0.48	0.97	0.60	0.99
Dérivée 2ème (type "SAVITSKY-GOLAY")	3	0.47	0.99	1.14	0.97

MATIERES PROTEIQUES

Pré-traitement	CALIBRATION			PREDICTION	
	Loading	SEC (g/kg)	r2	SEP (g/kg)	r2
Aucun	5	0.09	0.99	0.30	0.99
Dérivée 1ère (type "SAVITSKY-GOLAY")	3	0.63	0.98	0.66	0.99
Dérivée 2ème (type "SAVITSKY-GOLAY")	5	0.48	0.99	0.95	0.99

Tableau 18 : Dosage des matières grasses et des protéines dans les solutions modèles.

Pour cette analyse quantitative, nous avons choisi le domaine spectral 1980-2500 nm afin de comparer la capacité des sondes 0-45° et transflexion. Il reste néanmoins limité car les fibres optiques absorbent fortement le signal sur ces longueurs d'onde. Nous avons donc mis en place une deuxième analyse quantitative sur le domaine 1500-2000 nm en conservant un chemin optique de 200 µm et des paramètres d'acquisitions identiques. Pour les deux espèces dosées, les erreurs standards de prédiction dépassent la limite des 1 g/kg quel que soit le traitement spectral appliqué. L'exploitation des absorptions situées en dessous de 2000 nm passe par l'augmentation du chemin optique.

Nous avons donc développé deux nouvelles analyses quantitatives à partir de deux autres lots d'échantillons synthétiques sur des chemins optiques de 500 et 900 µm. Le domaine spectral couvre la zone 1500-2500 nm. Les paramètres spectraux restent identiques. Les résultats de cette étude sont présentés sur les Tableaux 19A et B. Pour un chemin optique de 900 µm, seul le domaine 1500-1850 nm permet d'obtenir des erreurs inférieures à 1 g/kg. Elles sont néanmoins amplifiées par rapport au chemin optique de 200 µm. Toute contribution spectrale

prise en compte au dessus de 1850 nm détériore l'analyse. Les erreurs deviennent alors très importantes à cause d'un biais qui augmente et des coefficients de corrélation qui diminuent. Aux longueurs d'onde élevées, le faible signal qui arrive au détecteur décorrèle complètement l'information spectrale par rapport aux concentrations. Pour un chemin optique aussi important, l'éclairage de l'échantillon est beaucoup trop faible. Les résultats obtenus sur le chemin optique de 500 μm sont plus intéressants. Pour les matières grasses, nous obtenons une erreur de 0,36 g/kg à partir des spectres bruts en utilisant l'ensemble du domaine 1500-2500 nm. Le dosage est même amélioré par rapport au chemin optique de 200 μm . Si nous réduisons le domaine spectral à 1500-2025 nm, cette erreur reste stable puisqu'elle est de 0,41 g/kg. Ce point est particulièrement important puisqu'il montre que le dosage des matières grasses est possible sur un domaine n'exploitant pas la zone 2000-2500 nm. Si nous réduisons à nouveau le domaine à une zone 1500-1850 nm, nous observons une augmentation de l'erreur soit 0,60 g/kg. La zone 1825-2025 nm est donc primordiale pour obtenir un dosage des matières grasses de qualité sur un chemin optique de 500 μm . Pour les protéines, c'est le domaine 1500-2025 nm qui permet d'obtenir une erreur minimale de 0,32 g/kg. L'ajout de la zone 2025-2500 nm détériore fortement le dosage. La comparaison des erreurs obtenues sur les domaines 1500-1850 et 1500-2025 nm montre à nouveau l'importance du domaine 1850-2025 nm pour le dosage des protéines. En résumé, la sonde 0-45° permet d'effectuer les dosages des matières grasses et des protéines sur des solutions synthétiques en utilisant un domaine spectral compatible avec l'utilisation de grandes longueurs de fibre optique.

A)

Chemin optique : 500 μm

Domaines spectraux

	1500-1850 nm		1500-2025 nm		1500-2500 nm	
	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première
SEP MG (g/Kg)	0,60	0,64	0,41	0,95	0,36	1,28
SEP MP (g/Kg)	1,06	0,74	0,32	1,10	0,88	1,45

B)

Chemin optique : 900 μm

Domaines spectraux

	1500-1850 nm		1500-2025 nm		1500-2500 nm	
	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première
SEP MG (g/Kg)	0,69	0,96	1,19	2,51	Résultats	Résultats
SEP MP (g/Kg)	0,83	2,25	3,19	6,48	inexploitables	inexploitables

Tableaux 19A et B : Influence du domaine spectral utilisé pour l'analyse quantitative de laits reconstitués.

4.10.3 Analyse quantitative de laits crus

Cette partie a été développée à partir de 100 échantillons de lait soit 80 en calibration et 20 en prédiction. Les concentrations de référence sont données par un analyseur automatique MIR (*Bentley*TM). Les paramètres d'acquisition restent identiques à la manipulation précédente. La première analyse a porté sur un chemin optique de 500 μm conformément à l'optimisation sur les laits reconstitués. Les résultats sont présentés sur les Tableaux 20. Quel que soit le domaine spectral ou le pré-traitement appliqué, les erreurs standards de prédiction sont très importantes pour les protéines. Les coefficients de corrélation sont très faibles. Pour les courtes longueurs d'onde, la différentiation améliore légèrement ces résultats. Le dosage des matières grasses est plus en accord avec l'analyse des laits synthétiques. L'erreur la plus faible est obtenue sur le domaine 1500-2025 nm. La dérivée première améliore le dosage pour atteindre une erreur de 0,70 g/kg. L'introduction du domaine 2025-2500 nm détériore le modèle. De même, si nous limitons le domaine à 1500-1850 nm, nous avons à nouveau une amplification des erreurs. Pour un chemin optique de 500 μm , la zone 1850-2025 nm est nécessaire au dosage des matières grasses. L'ensemble des résultats montre que l'optimisation du chemin optique pour les laits synthétiques n'est pas transposable aux laits crus. Nous avons néanmoins montré qu'une diminution du chemin optique supprimait les erreurs importantes. En fait, la présence de nombreux globules gras dans le lait cru augmente de manière fictive le chemin optique. Pour retrouver les conditions d'analyse des laits reconstitués, il est logique de diminuer le chemin optique. Les 100 échantillons ont donc été analysés avec une sonde à 200 μm de chemin optique (Tableaux 20B). Nous observons une nette amélioration des erreurs de prédiction. Le dosage des matières grasses est très satisfaisant puisqu'il atteint une erreur de 0,58 g/kg en utilisant les spectres dérivés. Le domaine 1850-2025 nm est toujours décisif. Le dosage des protéines reste délicat puisque l'erreur minimale de 1,21 g/kg reste au dessus des spécifications attendues. Les coefficients de corrélation sont aussi très faibles ($< 0,8$).

A)

Chemin optique : 200 μm

Domaines spectraux

	1500-1850 nm		1500-2025 nm		1500-2500 nm	
	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première
SEP MG (g/Kg)	0,96	0,98	0,82	0,58	0,91	1,46
SEP MP (g/Kg)	1,35	1,31	1,27	1,21	1,73	1,82

B)

Chemin optique : 500 μm

Domaines spectraux

	1500-1850 nm		1500-2025 nm		1500-2500 nm	
	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première
SEP MG (g/Kg)	1,12	1,15	0,98	0,70	1,03	2,10
SEP MP (g/Kg)	1,55	1,49	1,53	1,46	1,98	2,18

Tableaux 20A et B : Dosage des matières grasses et des protéines sur le lait cru.

A ce stade de l'étude, les nombreux algorithmes de pré-traitement des spectres présentés dans la partie théorique ont été appliqués. Les résultats ne sont pas présentés car le dosage des protéines est resté en défaut. De plus, aucune amélioration n'a été apportée aux erreurs standards de prédiction des matières grasses.

Le passage de la sonde transflexion à la sonde 0-45° a permis d'augmenter considérablement le signal optique et par là même le rapport signal sur bruit. De cette façon, l'erreur standard de prédiction en matières grasses a été divisée par deux. Pour les protéines, le gain de signal n'a pas été suffisant pour exploiter les faibles bandes protéines masquées par le bruit. La solution réside donc dans une intégration du signal plus importante afin d'en améliorer à nouveau le niveau.

4.11 La sonde « diffusion »

4.11.1 Introduction

Les fibres optiques de collection ont été remplacées par un détecteur placé près de la vitre où circule le lait (Figure 72). De cette façon, l'étendue géométrique qui dépend de la surface et de l'angle solide sous lequel les rayons peuvent être absorbés, est largement plus importante :

- Surface du détecteur : 25 mm²
- Section d'une fibre optique de 600 µm de diamètre : 0,28 mm²
- Ouverture numérique d'un détecteur : $\cong 1$
- Ouverture numérique d'une fibre optique de 600 µm de diamètre : 0,2-0,3

De plus, l'absorption du signal par les fibres optiques pour les grandes longueurs d'onde est diminuée puisque la longueur totale utilisée est divisée par trois (comparaison avec la sonde 0-45°). Cette sonde impose une nouvelle configuration pour l'analyseur car chaque sonde contient un détecteur (Figure 73). La sélection des longueurs d'onde doit être effectuée avant d'arriver sur le détecteur afin d'obtenir un éclairage monochromatique de l'échantillon. Le signal qui repart de la sonde vers l'analyseur n'est plus optique mais électrique. On note toutefois quelques inconvénients pour cette configuration. Des travaux ont montré qu'il était plus intéressant en terme de flux arrivant sur le détecteur, d'éclairer un échantillon par une

lumière polychromatique intense puis d'analyser spectralement la lumière diffusée que de limiter directement l'éclairage par une lumière monochromatique⁽³⁸⁾.

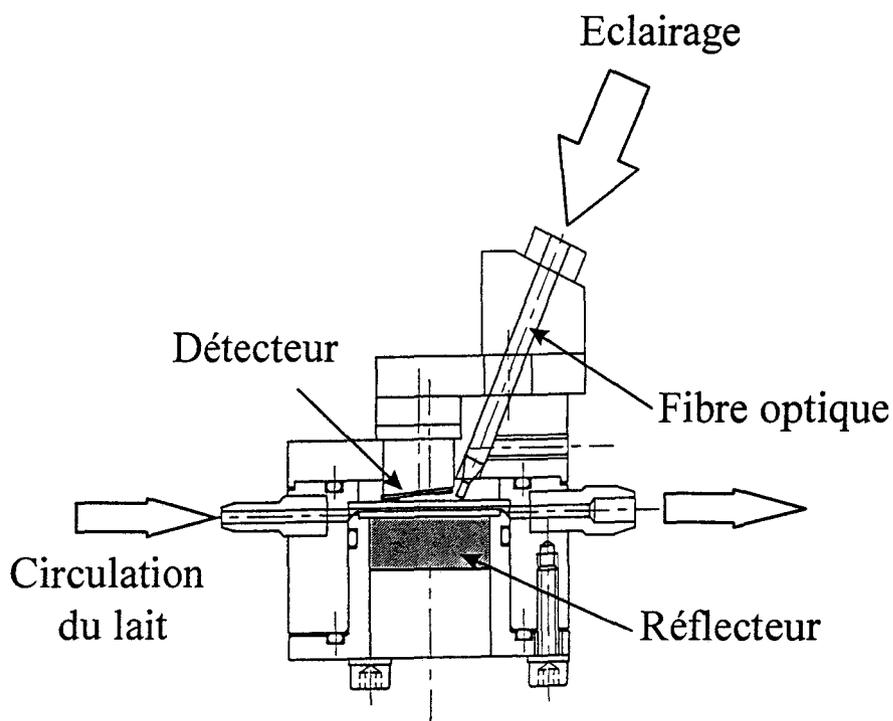


Figure 72 : La sonde « diffusion »

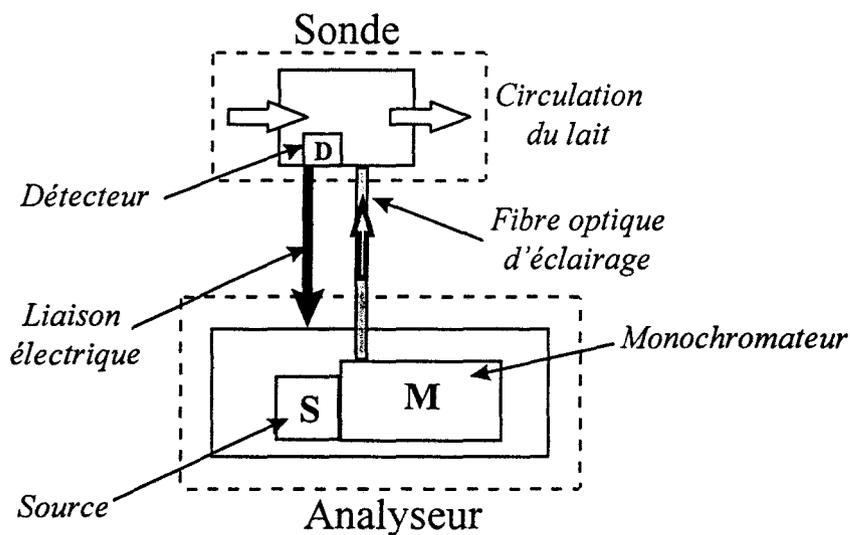


Figure 73 : Modification de l'analyseur pour l'intégration de la sonde « diffusion »

Le deuxième inconvénient est la proximité du détecteur par rapport au lait qui circule. Les détecteurs au sulfure de plomb (PbS) sont en effet très sensibles aux variations de

température^(228,229). Afin de stabiliser la signal nous avons mis en place un élément Peltier qui, par effet thermoélectrique, contrôle la température du détecteur^(230,231). La notion de stabilité du signal est très importante en proche infrarouge car les bandes d'absorption sont souvent très faibles. Le dernier inconvénient est lié à la transmission du signal électrique du détecteur jusqu'à l'analyseur. Même après une amplification, les courants et les tensions sortant du détecteur sont faibles. Sur de longues distances, il n'y a pas d'atténuation du signal comme nous pouvons le voir pour les fibres optiques mais un phénomène de parasitage. De cette façon, le bruit électronique peut prendre une place beaucoup plus importante par rapport à d'autres sources de bruit dans la chaîne de mesure.

Une comparaison des spectres simples faisceaux obtenus sur les sondes 0-45° et « diffusion » est donnée sur la Figure 74. Le lait cru analysé contient 35 g/kg en MG et 32 g/kg en MP. Les paramètres d'acquisition restent identiques soit 1 seconde d'intégration par élément spectral et 5 nm d'incrément.

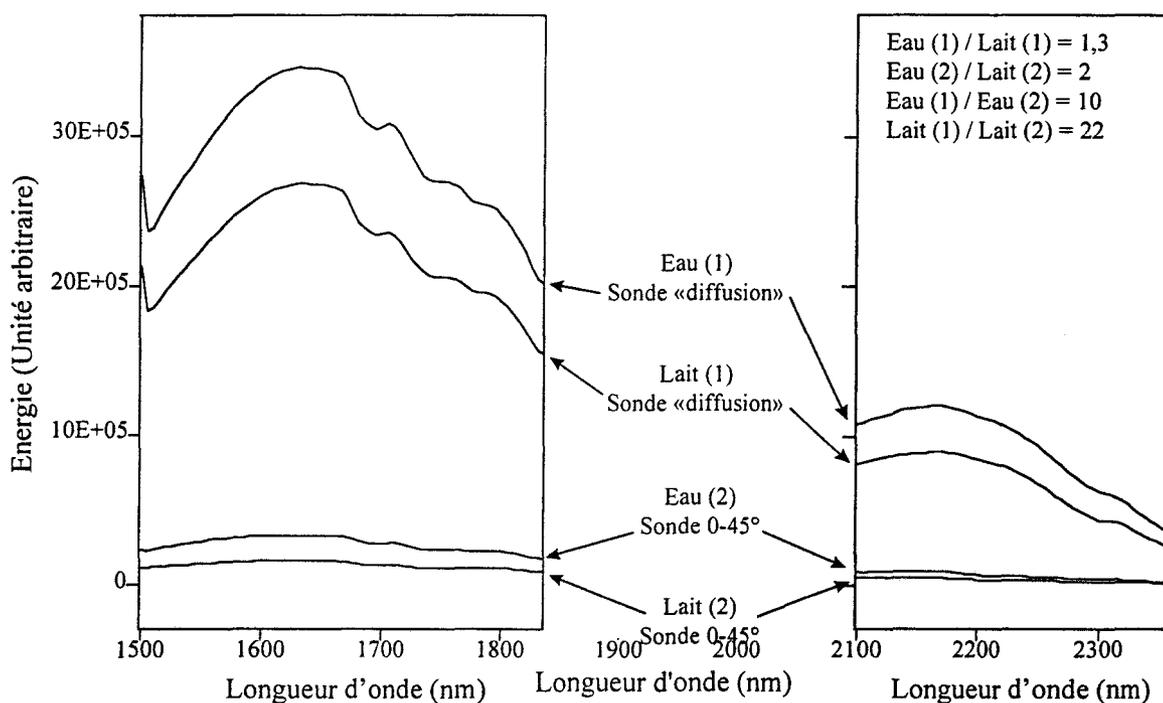


Figure 74 : Comparaison du niveau de signal entre la sonde 0-45° et la sonde diffusion

Les rapports d'intensité sont calculés à 1650 nm. Ils montrent un gain en signal d'autant plus important que l'échantillon analysé est diffusant. Depuis l'étude de la sonde transflexion, nous avons donc multiplié par 700 le signal collecté pour l'analyse du lait. La répétabilité a ensuite

été évaluée à partir d'une solution de matières grasses diluée à 40g/kg. Les résultats sont consignés dans le Tableau 21. Nous observons une légère amélioration par rapport à la sonde 0-45°. Cette inflexion de progression de la répétabilité, malgré un gain en signal important, montre que l'ouverture numérique n'est plus la source majeure de variation.

Longueur d'onde (nm)	Répétabilité (RSD %)
2345	0,21
2305	0,17
2200	0,23
1780	0,24
1750	0,27
1685	0,29
1400	0,30

Tableau 21 : Répétabilité de la sonde « diffusion »

De plus, la linéarité du signal optique en fonction de la concentration a été validée. Les résultats obtenus sont de bonne qualité. Ils sont semblables à la représentation donnée pour la sonde transflexion (Figure 54). Conformément à la procédure suivie, ces premiers résultats nous autorisent à poursuivre la validation par l'analyse quantitative de solutions modèles.

4.11.2 Analyse quantitative de solutions modèles

Toujours par soucis de rapidité, nous avons développé une analyse quantitative sur 16 solutions modèles à teneurs en matières grasses et protéines variables (procédure analogue à la sonde 0-45°). Les Tableaux 22A et B donnent les résultats des dosages en fonction du chemin optique utilisé. Le premier fait important est la diminution des erreurs par différenciation des spectres obtenue sur certaines analyses. C'est le témoignage d'un bon rapport signal sur bruit nettement amélioré par le gain en signal de la sonde « diffusion ». Pour un chemin optique de 200 μm , le domaine 1500-2500 nm donne les meilleurs résultats avec 0,50 et 0,43 g/kg pour les matières grasses et les protéines respectivement. Lorsque le domaine spectral s'éloigne des longueurs d'onde importantes, les erreurs de prédiction augmentent car le chemin optique devient beaucoup trop faible par rapport aux coefficients d'extinction. Ainsi par exemple, ces erreurs sont de 0,71 et 0,62 g/kg pour le domaine 1500-2025 nm. Elles sont relativement importantes pour le dosage de solutions aussi simples. De plus, la zone 2025-2500 nm est inexploitable dans le cadre de systèmes utilisant de grandes longueurs de fibres optiques.

A)

Chemin optique : 200 μm
Domaines spectraux

	1500-1850 nm		1500-2025 nm		1500-2500 nm	
	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première
SEP MG (g/Kg)	1.11	1.17	0.88	0.71	0.76	0.50
SEP MP (g/Kg)	0.83	0.85	0.69	0.62	0.67	0.43

B)

Chemin optique : 500 μm
Domaines spectraux

	1500-1850 nm		1500-2025 nm		1500-2500 nm	
	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première
SEP MG (g/Kg)	0.93	0.96	0.61	0.55	0.95	1.05
SEP MP (g/Kg)	0.81	0.89	0.51	0.46	0.91	0.97

C)

Chemin optique : 900 μm
Domaines spectraux

	1500-1850 nm		1500-2025 nm		1500-2500 nm	
	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première
SEP MG (g/Kg)	0.90	0.86	1.25	1.52	Résultats	Résultats
SEP MP (g/Kg)	0.81	0.74	1.19	1.37	inexploitables	inexploitables

Tableaux 22A, B et C : Dosages des solutions modèles avec la sonde « diffusion »

Le chemin optique de 200 μm semble donc très limité. En fait, le chemin optique optimum se situe aux alentours de 500 μm où les erreurs de prédiction sont de 0,55 et 0,46 g/kg pour la zone 1500-2025 nm sur les spectres dérivés. Nous observons dans ce cas une nette détérioration des dosages lorsque le domaine 2025-2500 nm est exploité pour les régressions multivariées. Un chemin optique et des coefficients d'extinction importants donnent alors un signal saturé. Pour un chemin optique de 900 μm , les effets précédents sont amplifiés. Les erreurs de 0,86 et 0,74 g/kg obtenues pour la zone 1500-1850 nm sur les spectres dérivés restent trop importantes en vue d'une exploitation. L'ensemble de ces résultats montre une grande similarité avec le dosage des solutions modèles par la sonde 0-45°. La différence de taille reste néanmoins l'amélioration des erreurs standards par l'emploi de la différentiation. A ce niveau de l'étude, le dosage des laits reconstitués était suffisamment fiable pour passer aux laits crus sans affiner la recherche du chemin optique.

4.11.3 Analyse quantitative de laits crus.

Conformément au protocole utilisé pour la sonde 0-45°, l'analyse quantitative a été développée à partir de 100 échantillons de laits soit 80 en calibration et 20 en prédiction. Les concentrations de référence sont toujours données par l'analyseur automatique MIR. Les paramètres d'acquisition restent identiques à la manipulation précédente. Les chemins optiques et les domaines spectraux utilisés sont identiques à ceux du dosage des laits reconstitués. Les résultats de l'analyse quantitative sont répertoriés dans les Tableaux 23A et B. Les coefficients de corrélation sont alors très faibles et les biais très importants. De nouveau, nous pouvons observer une amélioration de plusieurs analyses grâce à la différentiation. Le chemin optique de 500 μm qui était optimum pour les laits reconstitués ne l'est plus pour les laits crus. En effet, les erreurs minimales sont obtenues pour un chemin optique de 200 μm sur le domaine 1500-2025 nm soit 0,61 et 0,98 g/kg. Ce phénomène déjà observé s'interprète par une augmentation du chemin optique due à la diffusion de la lumière sur les globules gras. La prise en compte du signal saturé sur le domaine 2025-2500 nm détériore rapidement la régression. La zone 1825-2025 nm reste décisive pour le dosage des protéines dans le lait. A ce niveau de l'étude, nous pouvons considérer que les dosages des matières grasses et des protéines sont en accord avec le cahier des charges.

A)

Chemin optique : 200 μm

Domaines spectraux

	1500-1850 nm		1500-2025 nm		1500-2500 nm	
	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première
SEP MG (g/Kg)	0.92	0.96	0.79	0.61	0.85	1.24
SEP MP (g/Kg)	1.31	1.28	1.15	0.98	1.43	1.57

B)

Chemin optique : 500 μm

Domaines spectraux

	1500-1850 nm		1500-2025 nm		1500-2500 nm	
	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première
SEP MG (g/Kg)	1.05	1.01	1.31	1.46	Résultats	Résultats
SEP MP (g/Kg)	1.25	1.27	1.37	1.35	inexploitables	inexploitables

Tableaux 23A et B : Dosage de laits crus avec la sonde « diffusion »

Afin d'affiner les résultats nous avons effectué l'analyse sur un chemin optique de 350 μm (Tableau 24).

	Chemin optique : 350 μm			
	Domaines spectraux			
	1500-1850 nm		1500-2025 nm	
	Spectres bruts	Dérivée première	Spectres bruts	Dérivée première
SEP MG (g/Kg)	0.98	1.11	1.15	1.10
SEP MP (g/Kg)	1.29	1.30	1.27	1.21

Tableau 24 : Analyse quantitative de laits crus

Les résultats obtenus sont intermédiaires aux chemins optiques de 200 et 500 μm . Le chemin optique de 200 μm peut donc être considéré comme optimum pour l'analyse de laits crus par la sonde « diffusion ».

Une étude systématique sur l'influence des pré-traitements appliqués aux spectres est présentée sur le Tableau 25.

Pré-traitement	SEP MG (g.kg)	SEP MP (g.kg)
Spectres bruts	0.79	1.15
Dérivée première	0.61	0.98
Dérivée seconde	1.35	1.52
MSC	0.75	1.10
SNV	0.85	1.24
EUCL	1.23	1.39
MEA	1.17	1.31

Tableau 25: Influence du pré-traitement des spectres sur le dosage des matières grasses et des protéines.

Le domaine spectral utilisé est 1500-2225 nm. La dérivée première donne les meilleurs résultats. La correction de la diffusion parasite à partir de la méthode MSC (Cf. page 63) améliore légèrement l'analyse par rapport aux données brutes mais ne permet pas d'atteindre le niveau de la dérivée première. Une sélection précise des longueurs d'onde les plus corrélées aux concentrations n'a pas amélioré les résultats de l'analyse quantitative. Les autres méthodes de normalisation par la variance SNV (Cf. page 66), par la norme EUCL (Cf. page 68) et par la moyenne MEA (Cf. page 68), amplifient largement les erreurs standards. La

méthode de la référence interne n'a pas été appliquée à cause des bandes d'absorption larges et peu intenses sur les spectres du lait. Les quelques tentatives ont donné des résultats catastrophiques.

4.12 Le monochromateur

Le développement de l'analyseur de lait ne se résume pas à l'unique recherche de la sonde optimale. Les différentes configurations ont été présentées chronologiquement dans les premières parties car elles constituaient la pièce maîtresse du système. Nous avons tenté d'améliorer le dosage par quelques modifications instrumentales qui se sont intégrées entre ces étapes clefs que constituent les sondes optiques. A ce jour, tous les éléments de l'analyseur n'ont pas été optimisés. Nous présenterons donc les résultats de la validation de ces dernières modifications.

4.12.1 Le filtre d'ordre

Après avoir caractérisé l'inefficacité de la sonde transflexion pour le dosage des protéines dans le lait cru, nous avons développé une sonde « à chemin optique variable ». Le but était d'exploiter les informations spectrales sur des domaines très différents. Dans le même esprit, nous avons supprimé le filtre d'ordre afin d'obtenir plus d'informations spectrales sur une même longueur d'onde espérant ainsi détourner un phénomène parasite pour la spectroscopie en une éventuelle source d'information. Les chances d'efficacité étaient néanmoins très réduites car nous participions ainsi à l'amplification du recouvrement déjà important sur les spectres proche infrarouge.

L'analyse quantitative par la sonde transflexion des matières grasses sur une gamme de laits reconstitués a servi de support pour évaluer l'effet de cette modification. Les paramètres d'acquisition sont identiques au protocole page 111. Le domaine spectral couvre la zone 900-2500 nm. Le Tableau 26 donne les résultats de l'analyse quantitative sur les spectres bruts avec et sans le filtre d'ordre. Nous observons une détérioration importante de l'erreur standard de prédiction. Les hautes corrélations obtenues avec le filtre entre absorbances et concentrations ont largement diminué car le signal apporté par l'ordre 2 à cette longueur

d'onde n'était pas nécessairement dû au constituant d'intérêt. Le filtre d'ordre est donc indispensable à l'analyseur de lait.

	SEP MG (g/kg)
Avec filtre d'ordre	0.57
Sans filtre d'ordre	1.03

Tableau 26 : Influence du filtre d'ordre sur le dosage des matières grasses.

4.12.2 Les fentes

Lors de la présentation du monochromateur, nous avons montré que les fentes introduisaient deux effets contradictoires (Cf. page 42). En effet, une grande résolution induit un faible flux sortant du monochromateur et réciproquement. Mais une résolution minimale est nécessaire sous peine de pertes importantes des informations spectrales. De manière identique à l'étude du filtre d'ordre, les analyses quantitatives de laits reconstitués à teneurs variables en matières grasses et protéines avec la sonde transfexion ont servi de support pour cette nouvelle optimisation. Nous utilisons donc les spectres bruts sur le domaine 900-2500 nm. L'influence de la résolution spectrale sur les erreurs standards de prédiction est présentée sur le tableau ci-dessous.

Résolution (nm)	SEP MG (g/kg)	SEP MP (g/kg)
2	0.71	0.65
6	0.54	0.32
10	0.57	0.46
20	0.60	0.81

Tableau 27 : Influence de la résolution sur l'analyse quantitative de laits reconstitués (sonde transfexion).

Les résultats obtenus ne sont pas surprenants. Nous obtenons une erreur minimale pour une résolution intermédiaire sur les deux espèces à doser. En effet, lorsque la résolution est trop importante (2 nm), le rapport signal sur bruit diminue à cause du faible flux sortant du monochromateur. Pour les faibles résolutions, le flux est très important mais les contributions spectrales sont perdues par un effet de lissage. Une résolution voisine de 6 nm semble donc optimale pour l'analyseur. Ces premiers résultats permettent de donner une tendance. Il

conviendra de vérifier l'optimisation de ce paramètre pour le dosage de laits crus sur le prototype final.

4.13 La chaîne électronique de détection

L'objectif de notre étude était d'évaluer la nécessité de la détection numérique pour l'analyseur de lait. L'analyse quantitative de laits reconstitués par la sonde transmission pure a servi de support pour cette optimisation (MG variable uniquement). Tous les échantillons de calibration et de prédiction ont été analysés par une première chaîne de mesure équipée d'une détection synchrone analogique. Nous avons ensuite effectué une deuxième série de spectres sur la base d'une détection synchrone numérique. Le domaine spectral couvre la zone 900-2500 nm. La Figure 75 représente les spectres d'un échantillon contenant 35 g/kg de matières grasses acquis sur les deux systèmes.

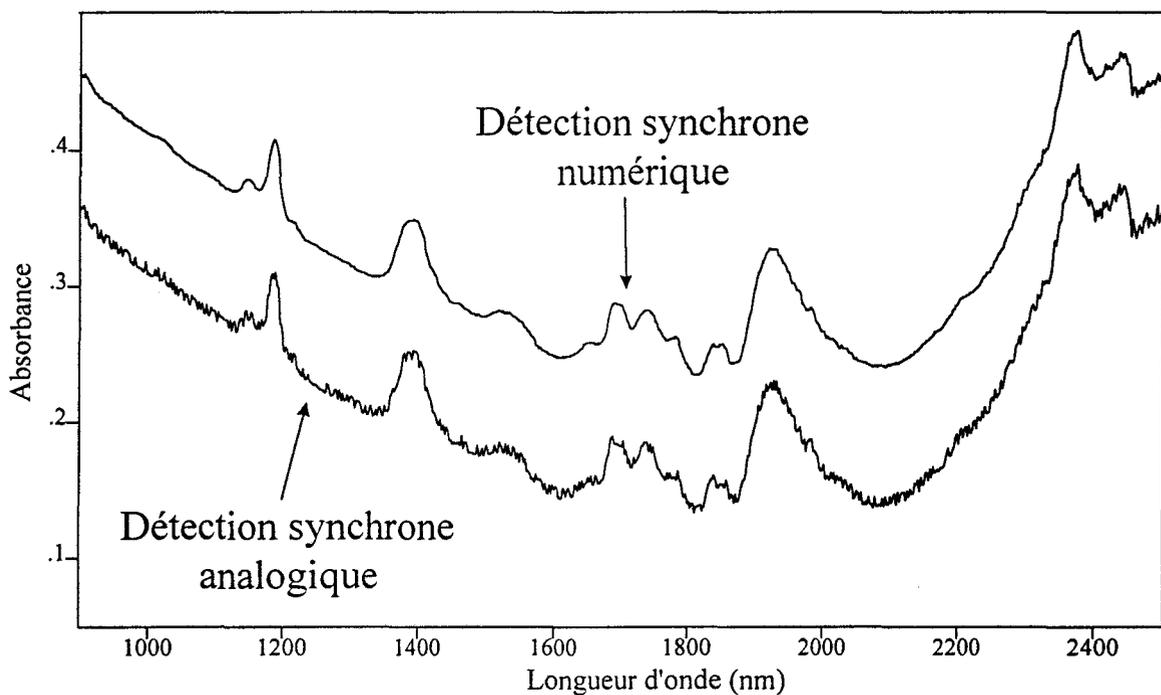


Figure 75 : Influence de la détection synchrone sur la qualité du signal.

Afin de faciliter la comparaison, le spectre acquis avec la détection numérique est décalé de 0,1 unité d'absorbance. Nous observons un rapport signal sur bruit beaucoup plus défavorable pour la détection synchrone analogique. Les résultats de la prédiction sur les deux systèmes sont donnés sur le Tableau 28. Les erreurs sont très importantes quel que soit le traitement

appliqué aux spectres. Cet effet est attribuable à la combinaison de la non-linéarité du signal optique (page 102) et du niveau de bruit trop important pour la spectroscopie proche infrarouge.

Pré-traitement	SEP MG (g/kg) Détection synchrone analogique	SEP MG (g/kg) Détection synchrone numérique
Aucun	1.55	1.31
Dérivée première	1.43	0.95
Dérivée seconde	1.59	1.35

Tableau 28 : Influence de la détection synchrone sur le dosage des matières grasses.

Une erreur minimale de 0,95 g/kg est obtenue avec la détection numérique sur les spectres en dérivée première (4 loadings). L'utilisation de la détection analogique amplifie considérablement l'erreur pour atteindre la valeur de 1,43 g/kg (5 loadings). La Figure 76 donne une comparaison des loadings pour ces deux analyses. Le premier loading est sensiblement équivalent pour les deux détections. A partir du deuxième loading, les différences sont considérables. La régression développée à partir de la détection numérique utilise 4 loadings montrant de nombreuses contributions spectrales. Pour la détection analogique, ces contributions sont masquées par un bruit important dès le deuxième loading. La deuxième régression construite sur la base de 5 loadings utilise des données très faiblement corrélées à la concentration en matières grasses d'où l'amplification des erreurs.

De manière indirecte, cette étude montre que la non-linéarité du signal est un réel obstacle au développement d'une analyse quantitative de qualité. Elle valide le rejet de la sonde transmission pure effectué lors de son étude. La détection synchrone numérique est indispensable au fonctionnement de l'analyseur. Cette nécessité n'a pas été démontrée pour le dosage de laits crus mais le niveau d'erreur rencontré devrait être équivalent.

Détection synchrone analogique

Détection synchrone numérique

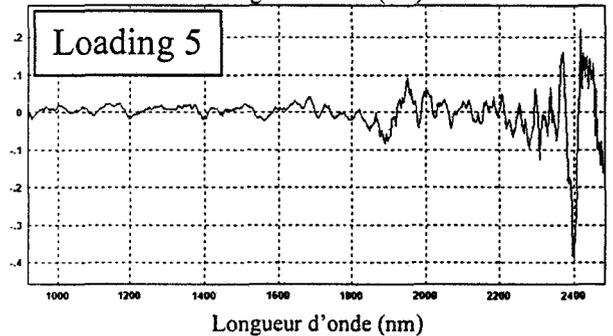
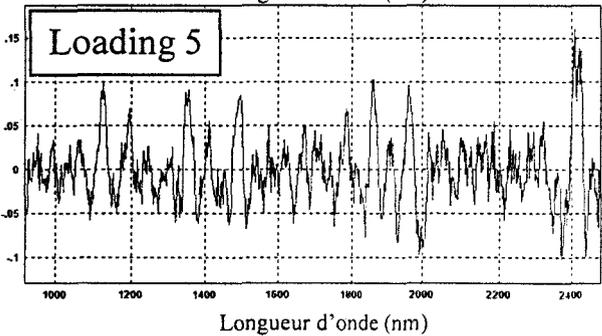
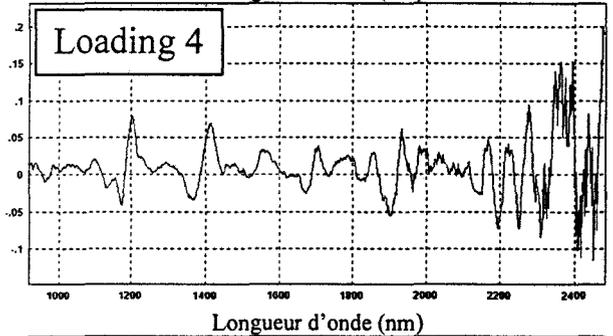
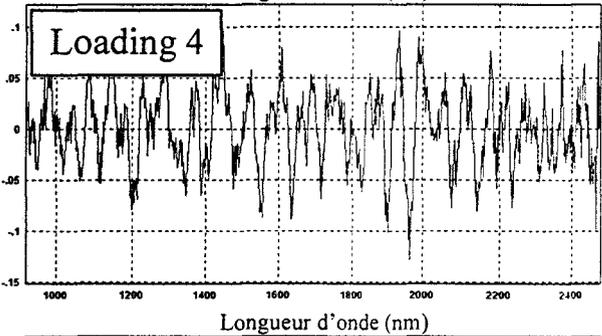
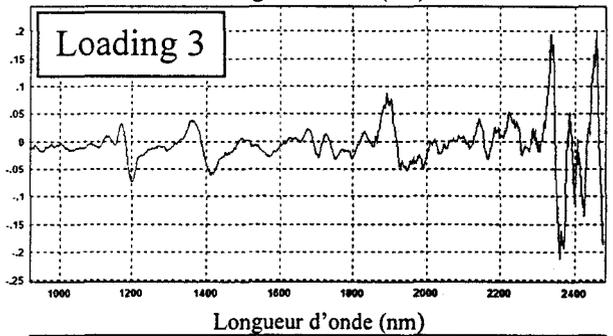
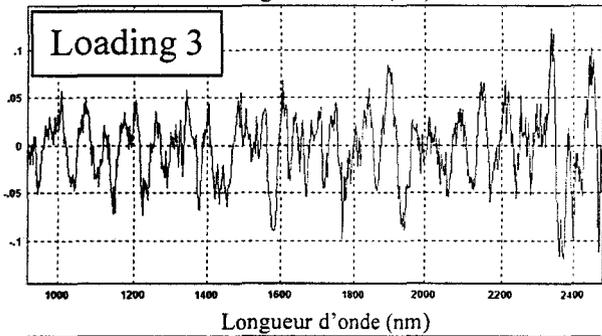
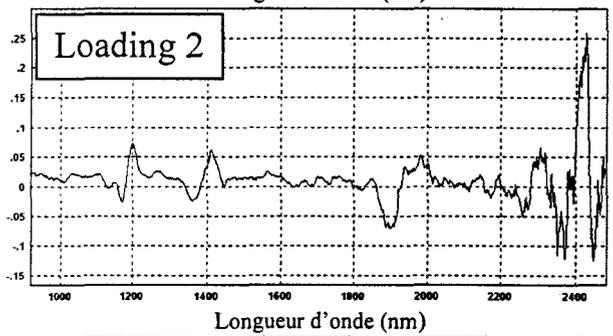
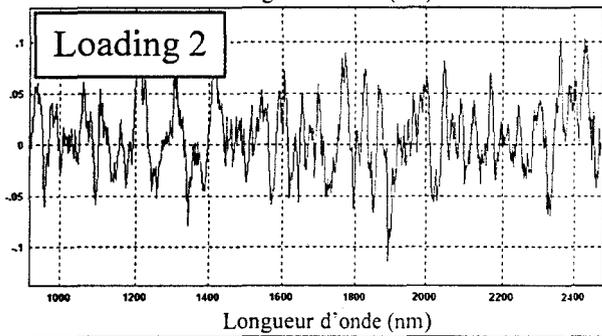
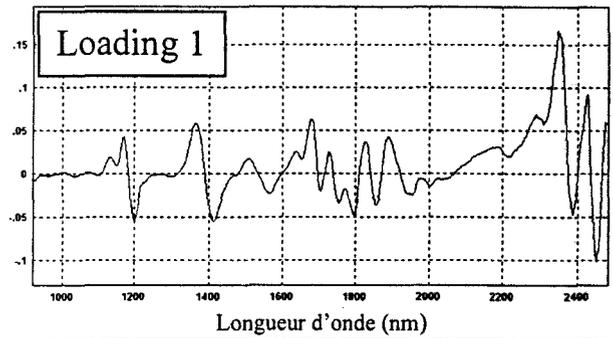
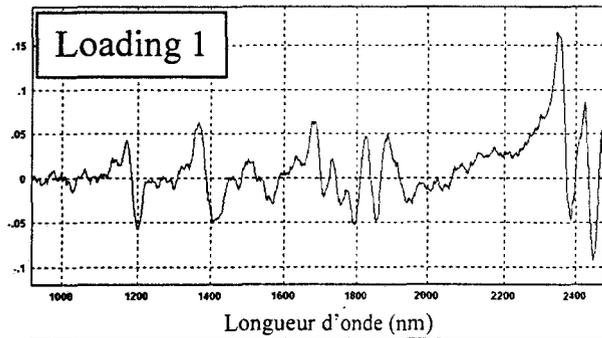


Figure 76 : Comparaison des loadings extraits pour les deux détections synchrones.

4.14 Conclusion

La chimiométrie illustre par cette étude sa capacité d'outil de validation pour l'instrumentation impliquée dans l'analyse quantitative. Tout au long du développement de l'analyseur, les erreurs de prédiction nous ont guidé pour l'optimisation des configurations optiques et électroniques. La sensibilité des résultats face aux changements apportés à la sonde optique montre toute l'importance qu'elle prend au sein de l'analyseur.

La sonde « transmission pure », conçue dans l'esprit de l'optique géométrique, s'est vite révélée inefficace. En fait, cette conception optique ne se prêtait pas à la diminution du principal défaut d'une fibre optique qui est sa faible ouverture numérique. La collection de la lumière diffusée par le lait dans les trois directions de l'espace était trop limitée. Ce manque d'intégration a induit une grande non-linéarité du signal et par là même le rejet de cette configuration optique. Afin de palier ce défaut de collection du signal, nous avons développé la sonde « transflexion ». L'ouverture numérique du système ne pouvait pas être améliorée puisque nous nous limitons à un nombre donné de fibres optiques. Par contre, l'ajout d'un réflecteur dans la sonde a permis de limiter la diffusion de l'échantillon dans un demi-espace d'où une augmentation du signal collecté. Une nette amélioration de la répétabilité et de la linéarité a suivi. Alors que la faisabilité des dosages des matières grasses et des protéines était démontrée pour des laits reconstitués, seul le dosage des matières grasses était satisfaisant pour les laits crus. A ce stade de l'étude, nous avons été sensibilisés sur la difficulté d'exploiter simultanément des domaines spectraux éloignés (diversité des extinctions) pour un chemin optique fixé. La sonde à chemin optique variable qui devait répondre à cet inconvénient est restée inexploitable à cause de la qualité spectrale obtenue. Afin de rompre définitivement avec l'esprit de l'optique géométrique inadaptée aux diffusions importantes, nous avons développé la sonde 0-45°. Un gain important en signal collecté a permis d'obtenir une bonne qualité spectrale. L'ensemble des analyses quantitatives a alors pu être amélioré laissant néanmoins le dosage des protéines au dessus des spécifications fixées par le cahier des charges. Mais le témoin d'une réelle amélioration de la qualité spectrale était la diminution de nombreuses erreurs standards de prédiction grâce à la différenciation de spectres jusque-là inefficace. La recherche constante d'une meilleure intégration du signal sur tout l'espace nous a finalement poussé à concevoir la sonde « diffusion » en remplaçant les fibres de collection de la sonde 0-45° par un détecteur posé directement sur la fenêtre de la sonde.

Un gain en signal collecté a de nouveau été obtenu grâce à une augmentation importante de l'ouverture numérique. Après la validation de la répétabilité et de la linéarité, une étude systématique sur le chemin optique et le pré-traitement des spectres a donné un dosage des matières grasses et des protéines conforme au cahier des charges. Mais le développement de l'analyseur ne s'est pas résumé à l'unique recherche de la sonde optimale puisque l'influence de certains composants du monochromateur et de la chaîne électronique sur les résultats des analyses quantitatives a été estimée. Nous avons ainsi démontré la nécessité du filtre d'ordre dans le monochromateur, de la détection synchrone numérique et d'une résolution spectrale voisine de 6 nm.

De nombreuses modifications instrumentales devraient nous permettre d'améliorer les capacités de dosage de l'analyseur. Sur le plan de l'optique, nous avons vu que le réseau de diffraction pouvait avoir une efficacité très différente en fonction de son profil (gravé ou holographique) et de la longueur d'onde de blaze. Il conviendra donc d'effectuer une étude systématique sur l'importance d'une efficacité moyenne pour un domaine spectral large ou d'une meilleure efficacité sur un domaine plus restreint, par rapport aux erreurs standards de prédiction en matières grasses et protéines. Au niveau de l'électronique, les détecteurs devraient apporter leur part à l'optimisation. Les détecteurs de type PbS ont une détectivité moyenne qu'ils conservent sur un large domaine spectral alors que les détecteurs InGaAs sont plus détectifs sur un domaine plus faible. Ils coupent en effet le signal optique au delà de 1700 nm. Par contre, les détecteurs PbS ont des surfaces actives beaucoup plus importantes. Une étude devra donc montrer l'importance relative de la détectivité par rapport à la surface active du détecteur au sein de la sonde diffusion pour le dosage des matières grasses et des protéines dans le lait cru. Les progrès réguliers des composants optoélectroniques ont donné naissance, à la fin de cette thèse, à une nouvelle génération de détecteur InGaAs dopé avec une longueur d'onde de coupure à 2100 nm⁽²²⁸⁾. Cette amélioration rend le choix du détecteur plus difficile et la nécessité d'une étude systématique plus grande encore. Finalement, un intérêt particulier pourra être porté sur l'échantillonnage pour la thermostatisation du lait, du corps de sonde et de l'analyseur afin d'augmenter la répétabilité.

V LE TRANSFERT DE CALIBRATION

5.1 Introduction

L'utilisation des régressions multivariées en spectroscopie a fait apparaître deux problèmes. Le premier est le transfert de l'équation de calibration d'un premier appareil (dit « maître ») vers un ou plusieurs autres (dits « esclaves »). Le cas typique est le développement d'une analyse quantitative sur un appareil au laboratoire de recherche devant être transférée sur des analyseurs en ligne sur le process. Dans notre cas, c'est le transfert de la calibration obtenue sur le prototype de dosage du lait vers les appareils de série. En considérant deux appareils A et B a priori identiques, l'utilisation directe de l'équation de calibration développée sur A aux spectres de B donnera des erreurs de prédiction très importantes. Cette amplification des erreurs est due aux différences de réponse spectrale de la chaîne de mesure entre A et B⁽²³²⁾. Il y a de nombreuses sources de variation, telles que les courbes d'émission des sources, les réponses spectrales des détecteurs, l'efficacité des réseaux de diffraction, les courbes de transmission des fibres optiques, les réglages optiques des monochromateurs (alignement, fentes, calibration en longueur d'onde) et les conditions d'acquisition des spectres. Malgré un effort constant des constructeurs pour la fabrication d'appareils aux caractéristiques identiques, les faibles différences qui subsistent entachent les prédictions de concentrations d'erreurs supplémentaires^(233,234). Il est de plus nécessaire d'effectuer un transfert de calibration, car pour des raisons évidentes de temps et coût, le développement d'une nouvelle analyse quantitative (préparation des échantillons, dosages de référence, acquisition des spectres, construction du modèle ...) sur chaque appareil est impossible⁽²³⁵⁾. Il faut aussi considérer que 12 sondes de mesure sont reliées par des fibres optiques de longueurs différentes à l'appareil de dosage. Une calibration développée sur la première ne pourra pas être utilisée avec efficacité sur les 11 autres, car les réponses spectrales sont différentes.

Le deuxième problème survient lorsque les caractéristiques d'un appareil varient au cours du temps⁽²³⁶⁾. On peut observer des variations de température, des dérives électroniques, des instabilités des longueurs d'onde et des détecteurs⁽²³⁷⁾. Les différents composants qui constituent l'appareil vieillissent aussi. De plus, lors du remplacement d'une pièce

défectueuse sur l'appareil il est courant d'obtenir des réponses spectrales très différentes. L'équation de calibration construite sur cet appareil à un temps donné ne pourra pas prédire avec une bonne précision les concentrations des produits à doser une année plus tard si les caractéristiques ont changé. Toutes ces raisons rendent le transfert de calibration indispensable pour le bon fonctionnement d'une analyse quantitative sur une période relativement longue.

5.2 Mise en place d'un transfert de calibration

Le but du transfert de calibration est de standardiser les réponses de l'appareil esclave pour obtenir des prédictions homogènes à celles obtenues avec l'appareil maître. Un ensemble d'échantillons appelés « standards » est analysé par les appareils maître et esclave. Ces données spectrales ou ces concentrations permettent alors de trouver la relation entre les deux appareils. Le choix des standards et de la méthode de transfert doit être attentivement étudiés afin d'obtenir une standardisation optimale.

5.2.1 Les échantillons « standards »

5.2.1.1 Les critères de sélection

En vue d'estimer, de la façon la plus précise, la différence entre la réponse spectrale de l'appareil maître et celle de l'esclave, il faut considérer avec beaucoup d'attention la stabilité et la représentativité des échantillons utilisés comme standard.

La stabilité : Les échantillons standards doivent avoir des propriétés physiques et chimiques stables au cours du temps. Si des modifications physiques ou chimiques des standards apparaissent entre l'acquisition des spectres sur l'appareil maître et celle de l'esclave, la différence spectrale sera bien sûr due aux différences instrumentales mais aussi à ces modifications. Si le transfert est fait à partir de ces données, les résultats de l'analyse quantitative sur l'instrument esclave seront erronés.

La représentativité : Les échantillons standards doivent être représentatifs des échantillons dont on veut prédire les concentrations avec la calibration. Si les standards n'appartiennent pas à la famille des échantillons étudiés, il faut s'assurer qu'ils possèdent des zones communes d'absorption.

5.2.1.2 Les différents types de standards

Sélection d'un sous-ensemble d'échantillons provenant de la calibration de l'appareil maître :

Cette première méthode extrait donc quelques standards du grand nombre d'échantillons utilisés pour la calibration de l'appareil maître. On effectue alors l'acquisition des spectres de ces standards sur l'instrument esclave.

L'avantage le plus important de cette méthode de sélection est que la nature des standards et des échantillons de calibration est identique. Si une bonne sélection est effectuée, on obtient une représentativité excellente. Différentes méthodes de sélection ont été proposées dans la littérature sur la base du caractère significatif d'un échantillon dans un lot⁽²³²⁾. Cette variable estime l'importance de l'échantillon sur les résultats de la calibration⁽²³⁸⁾. Cette technique s'est vite avérée médiocre car les standards ainsi sélectionnés n'étaient pas assez représentatifs des échantillons rencontrés en prédiction⁽²³⁹⁾. Une méthode de sélection plus appropriée basée sur l'algorithme de KENNARD et STONE⁽²⁴⁰⁾ a été proposée par DeNOORD *et al.*⁽²³⁹⁾. Cette technique utilise une sélection pas à pas qui a pour but de chercher l'échantillon dans la calibration le plus distant de l'ensemble des standards déjà extraits. Il a été démontré par BOUVERESSE *et al.* que l'ensemble des standards ainsi sélectionnés couvrait bien tout le domaine des concentrations⁽²³⁵⁾.

Malheureusement cette technique de sélection des standards au sein d'un ensemble d'échantillons de calibration ne peut pas être appliquée aux calibrations construites à partir d'échantillons extrêmement stables. Cette technique est donc inutilisable sur le lait.

Sélection de standards indépendants :

Cette approche a pour but d'estimer les différences entre les instruments maître et esclave en effectuant l'acquisition des spectres d'échantillons standards complètement indépendants même en nature des échantillons utilisés pour l'analyse quantitative. L'avantage le plus important de cette approche est l'utilisation de standards chimiquement et physiquement stables. Ces standards peuvent être synthétiques. Le deuxième avantage de ces standards est leur possible utilisation pour généraliser le transfert de calibration sur différentes analyses quantitatives⁽²³⁵⁾. Néanmoins, l'utilisation de ces échantillons standards indépendants pose des

problèmes de représentativité. Si les spectres des standards et des échantillons analysés sont trop différents, le transfert peut considérablement détériorer les erreurs de prédiction des composés à doser. Pour conserver cette bonne représentativité SHENK *et al.* ont proposé de nombreux ensembles de standards pour le transfert de calibration de produits agro-alimentaires^(241,242,243,244). L'ensemble de standards le plus approprié pour évaluer les différences entre instruments est un lot de 30 mélanges agro-alimentaires secs contenus dans des ampoules scellées en atmosphère inerte⁽²⁴²⁾. Ces mélanges possèdent une bonne représentativité pour les dosages de produits agro-alimentaires. Cet ensemble de standards a été utilisé avec succès par d'autres chercheurs tel que DARDENNE montrant en plus dans cette étude sa grande stabilité dans le temps^(245,246). Dans leurs travaux, un lot de standards inertes tels que les polymères ont été utilisés mais sans succès. KOWALSKI *et al.* ont proposé une solution intermédiaire mélangeant des échantillons génériques et des échantillons agro-alimentaires⁽²³⁶⁾. Une dernière solution proposée par BOUVERESSE *et al.*^(247,248) est l'utilisation d'échantillons organiques purs tels que des solvants dans des ampoules scellées. Les solvants organiques sont d'ailleurs utilisés comme standards de calibration en longueur d'onde pour la spectroscopie moyen infrarouge⁽²⁴⁹⁾.

5.2.1.3 Le nombre de standards

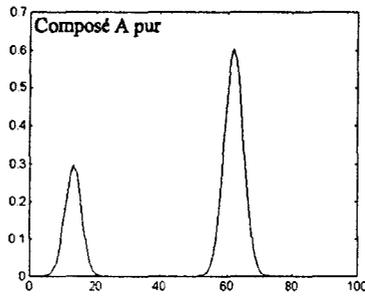
Le nombre de standards utilisés pour le transfert doit être attentivement choisi. En effet un nombre d'échantillons suffisamment important doit être sélectionné afin d'obtenir le maximum d'informations concernant les différences entre les deux instruments maître et esclave. Si ce nombre est trop faible, le transfert de calibration ne pourra pas assurer le bon fonctionnement des instruments esclaves en analyse quantitative. Néanmoins s'il est trop grand, le transfert devient inutile, car la construction d'une nouvelle calibration est possible. Les deux facteurs qui affectent le nombre de standards sont l'importance de la différence spectrale entre instruments maître/esclave et la méthode de standardisation. Des travaux assez contradictoires sont publiés sur la détermination de ce nombre d'échantillons qui semble pouvoir varier de quelques unités à plusieurs dizaines^(232,242,244).

5.3 La faisabilité des algorithmes

5.3.1 Introduction

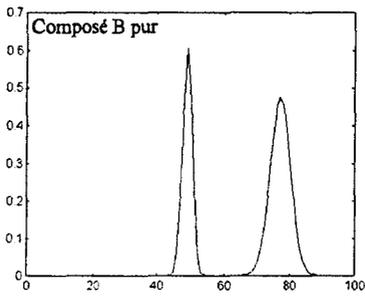
L'objectif de notre recherche était pour cette partie de tester les algorithmes de transfert afin d'établir de manière relative leur faisabilité au cas des analyseurs de lait. Pour réussir cette étude, il fallait dans un premier temps se placer en situation de totale maîtrise de tous les paramètres susceptibles d'influencer les spectres. Pour cela, nous avons travaillé sur des données synthétiques. Trois spectres de composés purs A,B,C ont été générés à partir de fonctions gaussiennes. A partir de ces trois spectres nous avons créé 125 spectres de mélanges par combinaison linéaire des composés A,B,C (intensité maximale de 0,6 en unité arbitraire) dont les concentrations en A, B et C sont c_a , c_b et c_c . Sur chaque spectre, nous avons ajouté une fonction de bruit de 1% suivant une loi normale par rapport à l'intensité maximale du spectre considéré (Figure 77). L'utilisation de spectres synthétiques est une pratique classique lorsqu'une étude systématique doit être mise en œuvre^(232,250). Ce lot de 125 spectres a été divisé en 3 sous-ensembles distincts : 75 pour la calibration, 30 pour la prédiction et 20 pour le transfert (Tableau 29). Ces spectres seront considérés comme provenant de l'instrument maître. Cet instrument sera utilisé pour construire l'équation de calibration. Afin de simuler des réponses spectrales différentes provenant d'autres appareils, nous avons appliqué sur chaque spectre de l'instrument maître une transformation spécifique simulant ainsi une fonction d'appareil différente par des décalages en longueur d'onde et en intensité. Cinq appareils esclaves ont ainsi pu être créés (Figure 78). Chaque algorithme de transfert sera évalué sur ces cinq appareils représentant les différentes variations susceptibles d'être rencontrées.

C_a*
(=1,2,3,4,5)



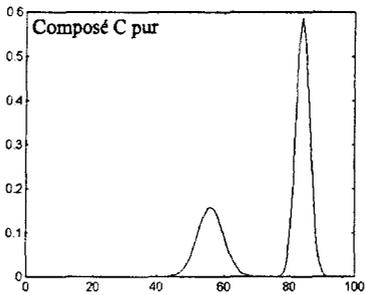
+

C_b*
(=1,2,3,4,5)



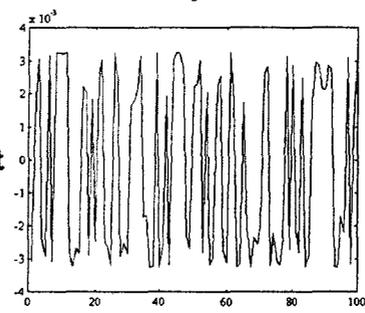
+

C_c*
(=1,2,3,4,5)

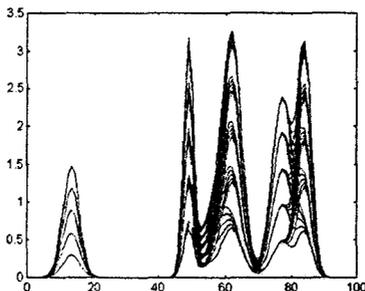


+

1% de Bruit



=



125 Spectres de mélange

Figure 77 : Création de spectres synthétiques

Spectre N°	Utilisation	Ca	Cb	Cc	Spectre N°	Utilisation	Ca	Cb	Cc	Spectre N°	Utilisation	Ca	Cb	Cc
1	C	1	1	1	51	C	3	1	1	101	P	5	1	1
2	C	1	1	2	52	P	3	1	2	102	C	5	1	2
3	C	1	1	3	53	T	3	1	3	103	C	5	1	3
4	P	1	1	4	54	C	3	1	4	104	P	5	1	4
5	C	1	1	5	55	C	3	1	5	105	C	5	1	5
6	P	1	2	1	56	P	3	2	1	106	C	5	2	1
7	C	1	2	2	57	C	3	2	2	107	P	5	2	2
8	P	1	2	3	58	C	3	2	3	108	P	5	2	3
9	C	1	2	4	59	T	3	2	4	109	C	5	2	4
10	C	1	2	5	60	P	3	2	5	110	C	5	2	5
11	T	1	3	1	61	C	3	3	1	111	C	5	3	1
12	P	1	3	2	62	C	3	3	2	112	P	5	3	2
13	C	1	3	3	63	C	3	3	3	113	T	5	3	3
14	C	1	3	4	64	P	3	3	4	114	C	5	3	4
15	C	1	3	5	65	T	3	3	5	115	C	5	3	5
16	P	1	4	1	66	C	3	4	1	116	P	5	4	1
17	T	1	4	2	67	C	3	4	2	117	C	5	4	2
18	C	1	4	3	68	P	3	4	3	118	C	5	4	3
19	C	1	4	4	69	C	3	4	4	119	T	5	4	4
20	P	1	4	5	70	C	3	4	5	120	P	5	4	5
21	C	1	5	1	71	T	3	5	1	121	C	5	5	1
22	C	1	5	2	72	P	3	5	2	122	C	5	5	2
23	T	1	5	3	73	C	3	5	3	123	C	5	5	3
24	P	1	5	4	74	C	3	5	4	124	C	5	5	4
25	C	1	5	5	75	C	3	5	5	125	C	5	5	5
26	C	2	1	1	76	P	4	1	1					
27	C	2	1	2	77	T	4	1	2					
28	P	2	1	3	78	C	4	1	3					
29	T	2	1	4	79	C	4	1	4					
30	C	2	1	5	80	P	4	1	5					
31	C	2	2	1	81	C	4	2	1					
32	P	2	2	2	82	C	4	2	2					
33	C	2	2	3	83	T	4	2	3					
34	C	2	2	4	84	P	4	2	4					
35	T	2	2	5	85	C	4	2	5					
36	P	2	3	1	86	C	4	3	1					
37	C	2	3	2	87	C	4	3	2					
38	C	2	3	3	88	P	4	3	3					
39	C	2	3	4	89	T	4	3	4					
40	P	2	3	5	90	C	4	3	5					
41	T	2	4	1	91	C	4	4	1					
42	C	2	4	2	92	P	4	4	2					
43	C	2	4	3	93	C	4	4	3					
44	P	2	4	4	94	C	4	4	4					
45	C	2	4	5	95	T	4	4	5					
46	C	2	5	1	96	P	4	5	1					
47	T	2	5	2	97	C	4	5	2					
48	P	2	5	3	98	C	4	5	3					
49	C	2	5	4	99	C	4	5	4					
50	C	2	5	5	100	P	4	5	5					

C Calibration
 Prédiction
 Transfert

Tableau 29 : Description des échantillons

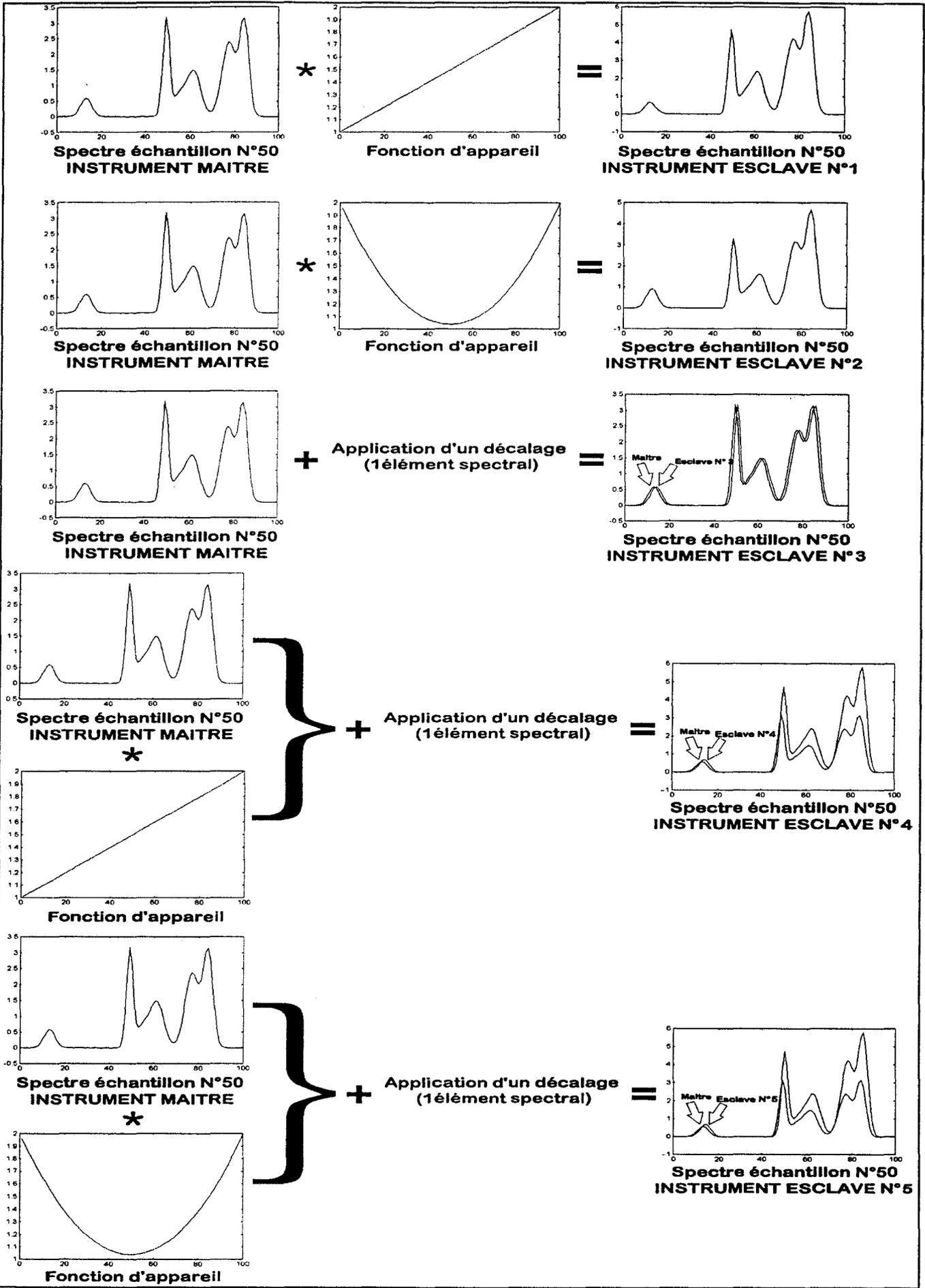


Figure 78 : Génération des spectres sur les instruments esclaves (Cas de l'échantillon N°50)

5.3.2 Analyse quantitative sur l'instrument maître

A partir des 75 échantillons de calibration et des 30 échantillons de prédiction, nous pouvons estimer la qualité prédictive de l'analyse. La méthode d'analyse multivariée est PLS. Le Tableau 30 donne les erreurs standards de prédiction (SEP) en fonction du nombre de facteurs « loading » et cela pour les trois constituants à doser.

Nombre de loading	Constituant A		Constituant B		Constituant C	
	SEC	SEP	SEC	SEP	SEC	SEP
1	0.1910	0.1870	0.2604	0.2691	0.3908	0.3653
2	0.0390	0.0400	0.0634	0.0630	0.0219	0.0230
3	0.0016	0.0018	0.0017	0.0016	0.0023	0.0024
4	0.0009	0.0021	0.0011	0.0021	0.0015	0.0025
5	0.0006	0.0025	0.0009	0.0022	0.0011	0.0027
6	0.0005	0.0027	0.0007	0.0023	0.0008	0.0030

Tableau 30 :Analyse quantitative sur l'instrument maître

On remarque aisément que le nombre de loading optimum pour les trois constituants est 3. Ce chiffre est aussi appuyé par une interprétation sans équivoque des loading et du test statistique de Fisher⁽¹³⁵⁾. Ces résultats montrent par ailleurs l'effet de l'overfitting lorsque le nombre de facteurs est trop important. Ce choix de 3 vecteurs loading permet de créer une équation de calibration pour chaque constituant à doser.

La nécessité d'un algorithme de transfert de calibration peut alors être facilement mise en évidence. En effet, si nous effectuons le dosage des constituants A, B et C à partir des spectres des échantillons de prédiction d'un appareil esclave et de l'équation de calibration de l'appareil maître, les résultats sont catastrophiques (Tableau 31). Les erreurs de prédictions sont largement amplifiées. Si on regarde le cas de l'instrument esclave 1 par exemple, le facteur d'amplification est d'au moins 1000. Sur des niveaux d'erreurs standards aussi importants, il serait hasardeux d'effectuer des comparaisons entre instruments esclaves. Ces variations provoquées sont bien plus importantes que celles que l'on peut réellement rencontrer mais elles permettront d'apprécier facilement la puissance des algorithmes de transfert.

Instrument	Constituant A	Constituant B	Constituant C
	SEP	SEP	SEP
Maître	0.0018	0.0016	0.0024
Esclave 1	1.7029	2.1294	2.8260
Esclave 2	0.4805	0.6259	1.5270
Esclave 3	0.0720	0.3959	0.2511
Esclave 4	1.6357	1.4950	3.0094
Esclave 5	0.4175	0.2046	1.6529

Tableau 31: Analyse quantitative sur les instruments esclaves

5.3.3 Étude des algorithmes de transfert

5.3.3.1 Introduction

Différentes approches de standardisation ont été proposées dans la littérature afin de transférer une calibration d'un appareil à un autre en corrigeant leurs différences. Il y a trois façons de procéder. La première est de corriger les concentrations prédites par l'appareil esclave donc sans toucher à la calibration ni aux spectres. La seconde est de transformer uniquement les spectres de l'appareil esclave afin de les faire ressembler à ceux de l'instrument maître. La dernière agit uniquement sur l'équation de calibration. D'autres techniques ont pu être étudiées comme la recherche de longueurs d'onde dites « iso-numériques » afin d'éviter l'utilisation de tout algorithme de transfert⁽²⁵¹⁾. Celles-ci sont choisies pour être insensibles aux variations spectrales entre instruments. De cette manière, la calibration est utilisée directement^(252,253). Cette méthode a de nombreux inconvénients comme l'impossibilité d'utiliser des algorithmes de condensation de données qui travaillent sur toutes les longueurs d'onde. Une dernière méthode proposée par De NOORD *et al.* est d'introduire toutes les variations possibles susceptibles d'être rencontrées au cours des acquisitions lors de la calibration sur l'instrument maître⁽²³⁹⁾. Mais la principale difficulté de cette approche est de définir exhaustivement ces variations. Ainsi, si une variation inattendue apparaît les résultats de l'analyse quantitative seront désastreux.

Les deux premières familles d'algorithmes seront successivement testées sur nos données synthétiques, le but étant de retrouver des erreurs standards de prédiction sur les instruments esclaves aussi proches que celles obtenues sur l'instrument maître. Tous les algorithmes de transfert utilisés ont été programmés sous environnement WINDOWS95™ sur MATLAB™

version 4.0 pour la rapidité du développement. De plus, la compilation des programmes avant exécution rend les calculs matriciels plus fluides.

5.3.3.2 La correction des concentrations

i) Principe

Il n'existe qu'un seul algorithme travaillant sur les concentrations : la correction de « pente et de biais ». Cette approche est la plus simple et la plus rapide à mettre en œuvre. Elle est aussi la plus ancienne. La correction pente/biais était souvent utilisée avec la méthode MLR en analyse quantitative^(254,255). Dans une première étape, on prédit les concentrations des standards sur l'appareil maître ($c^{\text{Maître}}$) à partir des spectres NIR (matrice des données spectrales $A^{\text{Maître}}$) et de l'équation de calibration (caractérisée par la matrice H). Ces mêmes standards sont alors passés sur l'appareil esclave (matrice des données spectrales A^{Esclave}). On prédit ensuite à partir de ces spectres et de l'équation de calibration de l'instrument maître (H) les concentrations c^{Esclave} .

$$\text{(éq. 90)} \quad c^{\text{Maître}} = A^{\text{Maître}} \cdot H$$

$$\text{(éq. 91)} \quad c^{\text{Esclave}} = A^{\text{Esclave}} \cdot H$$

La dernière étape est le tracé des concentrations $c^{\text{Maître}}$ en fonction des concentrations c^{Esclave} . Une simple régression par les moindres carrés (sous ExcelTM par exemple) permet finalement d'obtenir l'équation de la droite reliant les deux concentrations (éq. 92). Le coefficient de corrélation R^2 de cette droite permet de contrôler l'efficacité de la régression.

$$\text{(éq. 92)} \quad c^{\text{Maître}} = (\text{biais}) + (\text{pente}) \cdot c^{\text{Esclave}}$$

Ainsi chaque concentration obtenue sur l'appareil esclave sera corrigée par cette équation d'où le nom de correction pente/biais.

ii) Application aux données synthétiques

La procédure est détaillée pour le transfert entre l'instrument maître et l'esclave 1. Nous l'appliquons donc sur les 20 échantillons de transfert. Les concentrations en constituant A prédites par l'instrument maître et par l'esclave 1 sont consignées dans le Tableau 32. Une droite de régression linéaire est calculée entre les concentrations prédites par l'instrument maître et l'esclave 1 (Figure 79). Cette transformation linéaire est ensuite appliquée aux 30

concentrations prédites par l'instrument esclave afin de calculer les concentrations prédites corrigées. Cette méthode a été reproduite pour chacun des trois constituants et pour les cinq instruments esclaves.

Concentration A Instrument Maître	Concentration A Instrument Esclave 1
1.0014	1.5048
1.0016	1.5023
1.0008	1.4775
1.0011	1.4454
2.0012	2.9708
2.0022	2.9431
2.0017	3.0359
2.0000	3.0089
3.0022	4.5260
2.9984	4.4934
3.0023	4.4710
2.9963	4.5709
4.0007	6.0816
3.9997	6.0530
4.0005	6.0295
4.0006	6.0022
5.0027	7.6436
5.0014	7.6223
5.0023	7.5870
4.9997	7.5623

Tableau 32 : Concentrations en constituant A pour les instruments maître et esclave

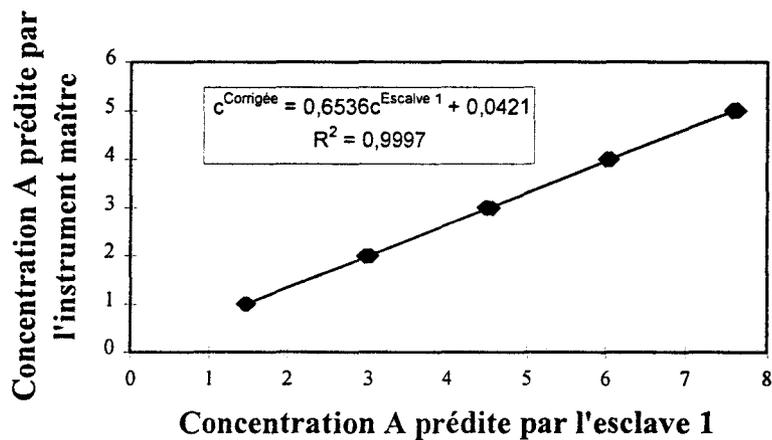


Figure 79 : Etablissement de la droite de régression

Les équations de correction sont présentées au Tableau 33 ainsi que les erreurs standards avant et après leur utilisation.

Instrument	Constituant A		Constituant B		Constituant C	
	Brut	SEP avec correction	Brut	SEP avec correction	Brut	SEP avec correction
Esclave 1	1.7029	0.0234 ($c^{\text{Corrigée}} = 0,6536.c^{\text{Esclave 1}} + 0,0421$) R2 = 0,9997	2.1294	0.0094 ($c^{\text{Corrigée}} = 0,6058.c^{\text{Esclave 1}} + 0,0212$) R2 = 0,9999	2.8260	0.0170 ($c^{\text{Corrigée}} = 0,5511.c^{\text{Esclave 1}} + 0,0475$) R2 = 0,9999
Esclave 2	0.4805	0.0530 ($c^{\text{Corrigée}} = 0,8523.c^{\text{Esclave 2}} + 0,0938$) R2 = 0,9988	0.6259	0.0197 ($c^{\text{Corrigée}} = 0,8347.c^{\text{Esclave 2}} + 0,0273$) R2 = 0,9990	1.5270	0.0340 ($c^{\text{Corrigée}} = 0,6949.c^{\text{Esclave 2}} + 0,0286$) R2 = 0,9994
Esclave 3	0.0720	0.0604 ($c^{\text{Corrigée}} = 1,0158.c^{\text{Esclave 3}} + 0,019$) R2 = 0,9985	0.3959	0.0768 ($c^{\text{Corrigée}} = 1,085.c^{\text{Esclave 3}} + 0,1339$) R2 = 0,9971	0.2511	0.2049 ($c^{\text{Corrigée}} = 1,0431.c^{\text{Esclave 3}} + 0,2589$) R2 = 0,9776
Esclave 4	1.6357	0.0448 ($c^{\text{Corrigée}} = 0,6652.c^{\text{Esclave 4}} + 0,0263$) R2 = 0,9992	1.4950	0.0906 ($c^{\text{Corrigée}} = 0,6561.c^{\text{Esclave 4}} + 0,1688$) R2 = 0,9961	3.0094	0.2054 ($c^{\text{Corrigée}} = 0,5755.c^{\text{Esclave 4}} - 0,2974$) R2 = 0,9781
Esclave 5	0.4118	0.0293 ($c^{\text{Corrigée}} = 0,8695.c^{\text{Esclave 5}} + 0,0752$) R2 = 0,9997	0.2046	1.0674 ($c^{\text{Corrigée}} = 0,9034.c^{\text{Esclave 5}} + 0,1886$) R2 = 0,9948	1.6529	0.2054 ($c^{\text{Corrigée}} = 0,7271.c^{\text{Esclave 5}} - 0,2655$) R2 = 0,9750
Valeurs optimales (Instrument Maître)		0.0018		0.0016		0.0024

Tableau 33 : Effet de la correction pente/biais sur les erreurs standards de prédiction

Nous remarquons rapidement que les gains sont faibles. Les erreurs sont loin des valeurs optimales. Lors de changement important des réponses spectrales, cette approche est inutilisable sous peine d'une lourde amplification des erreurs. Les seuls cas favorables de transfert à partir de cette correction ont été obtenus sur des variations spectrales faibles de type simple^(256,257). De plus cette méthode nous oblige à connaître les concentrations de référence des standards interdisant l'utilisation d'échantillons génériques. Les résultats du transfert n'étant pas satisfaisants nous nous sommes penchés sur les algorithmes effectuant des corrections spectrales.

5.3.3.3 La correction spectrale

Cette approche consiste à corriger le spectre de l'échantillon pris sur un instrument esclave pour le faire ressembler au spectre que l'on aurait pu prendre sur l'instrument maître. Les trois meilleures méthodes actuelles à savoir la standardisation directe (DS), la standardisation directe pas à pas (PDS), et la méthode de SHENK seront étudiées.

5.3.3.3.1 La standardisation directe (DS)

i) Principe

La standardisation directe a été introduite par WANG *et al.*⁽²³²⁾. Elle est basée sur la recherche d'une matrice dite de transfert **F**. Ainsi, si les spectres des standards pris sur l'appareil maître et sur l'esclave sont contenus respectivement suivant les lignes de $A^{\text{Maître}}$ et A^{Esclave} (matrices des données spectrales) nous pouvons écrire :

$$\text{(éq. 93) } A^{\text{Maître}} = A^{\text{Esclave}} \cdot F + E$$

La matrice **E** représente les résidus qui n'ont pu être modélisés. La matrice **F** qui est carrée peut être obtenue en multipliant la matrice inverse généralisée de $A^{\text{Maître}}$ par la matrice A^{Esclave} ⁽²⁵⁸⁾. Afin de calculer cette matrice inverse généralisée, nous pouvons employer la méthode PCR ou PLS. Comme nous l'avons déjà expliqué, c'est la technique PLS qui sera choisie car elle est la plus performante dans ces conditions d'utilisation.^(259, 260)

La Figure 80 permet de comprendre plus facilement la procédure de calcul de cette matrice de passage. Les matrices $A^{\text{Maître}}$ et A^{Esclave} contiennent autant de lignes que de spectres de standards (s au total). Ces spectres possèdent w longueurs d'onde. L'hypothèse de travail est la suivante : toutes les colonnes de $A^{\text{Maître}}$ sont indépendantes entre elles mais complètement corrélées à toutes les colonnes de A^{Esclave} .

$$\begin{array}{c}
 \text{Maître} \\
 \mathbf{A} \\
 \begin{pmatrix} a_{1,1}^{\text{Maître}} & \dots & a_{i,1}^{\text{Maître}} & \dots & a_{w,1}^{\text{Maître}} \\ \dots & \dots & \dots & \dots & \dots \\ a_{1,j}^{\text{Maître}} & \dots & a_{i,j}^{\text{Maître}} & \dots & a_{w,j}^{\text{Maître}} \\ \dots & \dots & \dots & \dots & \dots \\ a_{1,s}^{\text{Maître}} & \dots & a_{i,s}^{\text{Maître}} & \dots & a_{w,s}^{\text{Maître}} \end{pmatrix}
 \end{array}
 =
 \begin{array}{c}
 \text{Esclave} \\
 \mathbf{A} \\
 \begin{pmatrix} a_{1,1}^{\text{Esclave}} & \dots & a_{i,1}^{\text{Esclave}} & \dots & a_{w,1}^{\text{Esclave}} \\ \dots & \dots & \dots & \dots & \dots \\ a_{1,j}^{\text{Esclave}} & \dots & a_{i,j}^{\text{Esclave}} & \dots & a_{w,j}^{\text{Esclave}} \\ \dots & \dots & \dots & \dots & \dots \\ a_{1,s}^{\text{Esclave}} & \dots & a_{i,s}^{\text{Esclave}} & \dots & a_{w,s}^{\text{Esclave}} \end{pmatrix}
 \end{array}
 \cdot
 \begin{array}{c}
 \mathbf{F} \\
 \begin{pmatrix} F_{1,1} & \dots & F_{i,1} & \dots & F_{w,1} \\ \dots & \dots & \dots & \dots & \dots \\ F_{1,j} & \dots & F_{i,j} & \dots & F_{w,j} \\ \dots & \dots & \dots & \dots & \dots \\ F_{1,w} & \dots & F_{i,w} & \dots & F_{w,w} \end{pmatrix}
 \end{array}$$

Figure 80 : Calcul de la matrice de transfert F pour la standardisation directe

On effectue la régression multivariée PLS entre la longueur d'onde $i=1$ de la matrice $A^{\text{Maître}}$ et toutes les longueurs d'onde de la matrice A^{Esclave} . Ce calcul nous donne le coefficient de la régression soit la première colonne de F. Il suffit d'effectuer une itération du processus avec $i=1$ à w (w régressions PLS au total) pour obtenir la matrice F. Finalement après l'acquisition du spectre d'un échantillon, il suffit de le multiplier par F pour obtenir sa correction. L'inconvénient majeur de cet algorithme est le choix du nombre de vecteurs loading pour chaque régression. Mais l'utilisation de la décomposition en valeurs singulières (SVD)⁽¹³⁵⁾ et le critère de MOORE-PENROSE⁽²⁶¹⁾ pour la méthode PLS permettent une automatisation efficace de ce choix. Nous pouvons remarquer que pour cette phase de transfert aucune concentration n'est utilisée. Le transfert peut donc se faire avec des échantillons génériques⁽²³²⁾. Le polystyrène, longtemps employé pour évaluer les performances des

spectromètres infrarouge, est souvent pris comme standard générique⁽²³⁶⁾. Un autre avantage pour cette standardisation directe est la possibilité d'un transfert de la calibration entre deux instruments maître et esclave possédant un nombre de longueurs d'onde différent^(244,235) (monochromateur et appareil à filtres) ou encore des caractéristiques différentes (appareil de recherche à haute résolution et appareil de routine).

ii) Application de la standardisation directe aux données synthétiques

A partir des 20 standards, nous obtenons pour chaque appareil esclave le couple des matrices $A^{\text{maître}}$ et A^{esclave} (20×100) des données spectrales. La standardisation directe nous fournit pour chaque transfert la matrice de transfert F (100×100). Pour chaque instrument esclave, les 30 spectres de prédiction sont multipliés par la matrice de transfert associée. Ces 30 spectres alors corrigés sont prédits par l'équation de calibration. Les erreurs standards de prédiction sont consignées dans le tableau ci-dessous (Tableau 34).

Instrument	Constituant A SEP		Constituant B SEP		Constituant C SEP	
	sans correction	avec correction	sans correction	avec correction	sans correction	avec correction
Esclave 1	1.7029	0.0036	2.1294	0.0030	2.8260	0.0036
Esclave 2	0.4805	0.0037	0.6259	0.0034	1.5270	0.0038
Esclave 3	0.0720	0.0032	0.3959	0.0032	0.2511	0.0037
Esclave 4	1.6357	0.0032	1.4950	0.0033	3.0094	0.0034
Esclave 5	0.4118	0.0033	0.2046	0.0036	1.6529	0.0032
Valeurs optimales (Instrument Maître)		0.0018		0.0016		0.0024

Tableau 34 : Effet de la standardisation directe sur les erreurs standards de prédiction

Nous pouvons remarquer que cette méthode est de loin plus efficace que la correction de pente/biais. Malgré des différences spectrales abusives entre les différents instruments les erreurs standards entre maître et esclaves sont du même ordre. Nous pouvons attribuer cette efficacité au caractère multivarié de l'algorithme en comparaison avec la correction univariée précédente⁽²³⁵⁾. Une régression directe proposée dans la littérature⁽²⁵⁶⁾ de la matrice A^{esclave} sur toutes les colonnes de la matrice $A^{\text{maître}}$ avec la méthode PLS2 a été effectuée. Les résultats obtenus sont inutilisables. Ce résultat est en accord avec la supériorité déjà signalée de la méthode PLS1 sur la PLS2⁽¹³⁵⁾. Afin d'améliorer la standardisation directe WANG *et al.* ont proposé la standardisation directe pas à pas (PDS)⁽²³²⁾.

5.3.3.3.2 La standardisation directe pas à pas (PDS)

i) Principe

La différence entre DS et PDS est la façon de calculer la matrice de transfert. Cette nouvelle méthode considère que les informations spectrales à une certaine longueur d'onde sur l'instrument maître ne sont plus corrélées qu'à une fenêtre de petite dimension en longueur d'onde sur l'instrument esclave⁽²⁶²⁾. Ainsi pour chaque longueur d'onde i de l'instrument maître, on effectue une régression multivariée PCR ou PLS avec une fenêtre de longueurs d'onde $(i-k, i+k)$ sur l'instrument esclave soit entre $T_i^{\text{Maître}}$ et W_i (Figure 81).

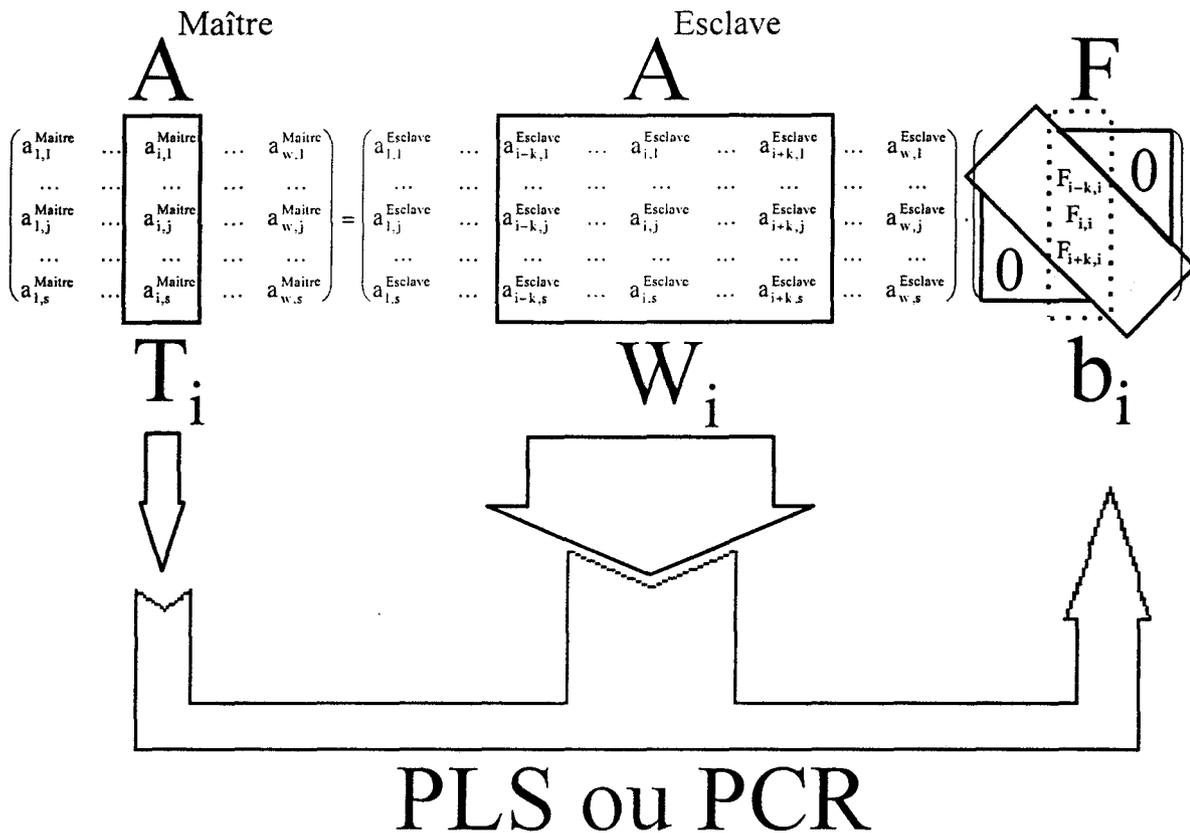


Figure 81 : Calcul de la matrice de transfert F pour la PDS

La taille de cette fenêtre est bien sûr choisie avant de commencer les calculs. Le coefficient de la régression b_i (vecteur colonne à $2k+1$ éléments) relie $T_i^{\text{Maître}}$ et W_i par la relation suivante :

$$\text{(éq. 94)} \quad T_i^{\text{Maître}} = W_i \cdot b_i + e_i$$

Les résidus non modélisés sont représentés par e_i . La largeur de cette fenêtre $(2k+1)$ reste souvent faible car les décalages en longueur d'onde dépassent rarement un élément spectral. Pour notre étude, cette fenêtre a été fixée à 5 longueurs d'onde soit $k=2$. Afin d'obtenir toutes les colonnes de la matrice de transfert F , ce calcul est effectué sur chaque longueur d'onde i du spectre avec un déplacement de la fenêtre W_i . La matrice F est diagonale en bande c'est-à-dire qu'elle possède des éléments non nuls uniquement suivant celle-ci. De manière identique à la standardisation directe, chaque spectre d'un instrument esclave est corrigé par la multiplication de la matrice de transfert F . Cette méthode a permis de transférer des calibrations entre instruments similaires ou différents⁽²⁶²⁾ en terme de configuration optique⁽²⁶³⁾, de rapport signal sur bruit, de température d'utilisation⁽²³⁷⁾, etc...

ii) Application de la standardisation directe aux données synthétiques

On utilise la même procédure que pour la standardisation directe, à savoir la recherche de la matrice de transfert F puis la correction des spectres de prédiction par multiplication. Pour le calcul de F nous avons testé les algorithmes PCR et PLS. L'analyse quantitative sur ces spectres corrigés donne finalement les erreurs standards suivantes pour les deux méthodes:

Instrument	Constituant A SEP		Constituant B SEP		Constituant C SEP	
	sans correction	avec correction	sans correction	avec correction	sans correction	avec correction
Esclave 1	1.7029	0.0064	2.1294	0.0025	2.8260	0.0031
Esclave 2	0.4805	0.0055	0.6259	0.0022	1.5270	0.0031
Esclave 3	0.0720	0.0038	0.3959	0.0029	0.2511	0.0026
Esclave 4	1.6357	0.0038	1.4950	0.0023	3.0094	0.0027
Esclave 5	0.4118	0.0031	0.2046	0.0024	1.6529	0.0025
Valeurs optimales (Instrument Maître)	0.0018		0.0016		0.0024	

Tableau 35 : Effet de la PDS sur les SEP (PCR pour calcul de F)

Instrument	Constituant A SEP		Constituant B SEP		Constituant C SEP	
	sans correction	avec correction	sans correction	avec correction	sans correction	avec correction
Esclave 1	1.7029	0.0029	2.1294	0.0021	2.8260	0.0035
Esclave 2	0.4805	0.0028	0.6259	0.0022	1.5270	0.0034
Esclave 3	0.0720	0.0022	0.3959	0.0021	0.2511	0.0027
Esclave 4	1.6357	0.0018	1.4950	0.0026	3.0094	0.0025
Esclave 5	0.4118	0.0021	0.2046	0.0022	1.6529	0.0027
Valeurs optimales (Instrument Maître)	0.0018		0.0016		0.0024	

Tableau 36 : Effet de la PDS sur les SEP (PLS pour calcul de F)

Pour le calcul de F , la méthode PLS semble légèrement plus efficace que la méthode PCR. De plus les erreurs standards de prédiction sont encore plus proches des valeurs cibles avec l'algorithme PDS qu'avec la méthode DS. Cette supériorité est due à l'utilisation d'un modèle multivarié local qui réduit le risque d'overfitting⁽²³⁵⁾ pour la PDS. Une deuxième explication peut être la présence de non-linéarités mieux modélisées par plusieurs régressions multivariées locales que par une régression multivariée globale unique. Néanmoins, lors du transfert de calibration avec PDS, les erreurs restent amplifiées par un coefficient proche de 1,5, valeur en accord avec la littérature⁽²³²⁾. Un centrage préalable des matrices $A^{\text{maître}}$ et A^{esclave} par leurs moyennes a permis à plusieurs chercheurs d'améliorer les résultats des transferts^(258,264,265). Son application sur le transfert par la méthode PDS basée sur la régression PCR nous a permis de réduire les erreurs standards de prédiction (Tableau 37).

Instrument	Constituant A SEP		Constituant B SEP		Constituant C SEP	
	sans correction	avec correction	sans correction	avec correction	sans correction	avec correction
Esclave 1	1.7029	0.0024	2.1294	0.0028	2.8260	0.0027
Esclave 2	0.4805	0.0030	0.6259	0.0026	1.5270	0.0027
Esclave 3	0.0720	0.0023	0.3959	0.0027	0.2511	0.0023
Esclave 4	1.6357	0.0024	1.4950	0.0025	3.0094	0.0025
Esclave 5	0.4118	0.0020	0.2046	0.0030	1.6529	0.0022
Valeurs optimales (Instrument Maître)	0.0018		0.0016		0.0024	

Tableau 37 : Effet de la PDS sur les SEP (Centrage des données et PCR pour calcul de F)

Ces derniers résultats montrent que les deux régressions PCR et PLS sont aussi performantes pour le transfert de calibration par la méthode PDS. Lors du calcul de la matrice F , il est important d'estimer pour chaque régression un nombre de vecteurs loading optimum sous peine de créer des artefacts sur les spectres corrigés⁽²³⁵⁾. Leur présence a pu être minimisée par une meilleure gestion des deux sources d'erreurs, la sélection des standards et le nombre de vecteurs latents (loading) à prendre en compte pour chaque régression PLS ou PCR⁽²⁶⁶⁾. Le manque de temps ne nous a pas permis d'utiliser ces derniers outils dans nos programmes. La sélection est améliorée par l'implémentation de l'algorithme de KENNARD *et al.*^(240,267,268) ou la méthode « MSID » développée par BOUVERESSE *et al.*⁽²⁶⁶⁾. Ces deux techniques utilisent la distance euclidienne alors que la méthode classique basée sur le « Leverage » utilise la distance de Mahalanobis. Pour le contrôle du nombre de vecteurs loading, BOUVERESSE *et al.*⁽²⁶⁶⁾ utilise un échantillon témoin proche du spectre moyen des échantillons standards. Son

spectre sur l'instrument esclave est transformé par la matrice de transfert puis retranché à son spectre acquis sur l'instrument maître. La distribution statistique de ce spectre de différence permet de détecter des longueurs d'onde où les calculs sont aberrants. Afin d'éviter cet overfitting, le nombre de loading pris pour la régression est réduit.

5.3.3.3.2 La standardisation par la méthode de SHENK

C'est le premier algorithme qui a utilisé une correction sur les spectres⁽²⁴²⁾. Il est d'un intérêt assez limité pour un futur développement industriel, car il est protégé par un brevet américain⁽²⁶⁹⁾. De plus seuls les transferts entre instruments identiques (nombre de longueurs d'onde) sont possibles. Cette approche effectue une correction univariée de l'ensemble du spectre en deux étapes successives, la correction des longueurs d'onde puis la correction des absorbances.

i) Principe

Considérons comme pour le cas de la standardisation directe, s échantillons standards. Les spectres de ces échantillons (w longueurs d'onde) acquis sur les instruments maître et esclave nous donnent les matrices des données spectrales correspondantes $A^{\text{Maître}}$ et A^{Esclave} . Comme pour la méthode PDS, nous allons considérer que l'information spectrale pour l'échantillon j à une longueur d'onde i notée $a_{i,j}^{\text{Maître}}$ dans $A^{\text{Maître}}$ est corrélée aux longueurs d'onde dans une fenêtre autour de i ($i-k, \dots, i, \dots, i+k$) notées $a_{i-k}^{\text{Esclave}}, \dots, a_i^{\text{Esclave}}, \dots, a_{i+k}^{\text{Esclave}}$ dans A^{Esclave} . Nous effectuons en premier lieu une régression linéaire simple entre T_i^M (vecteur colonne contenant $a_{i,1}^{\text{Maître}} \dots a_{i,j}^{\text{Maître}} \dots a_{i,w}^{\text{Maître}}$) et chacun des $T_{i-k}^E \dots T_i^E \dots T_{i+k}^E$ soit au total $(2k+1)$ régressions (Figure 82). Chacune d'elle permet d'obtenir un coefficient de corrélation noté r_{i-k}^2 entre T_i^M et T_{i-k}^E par exemple. La plus grande valeur des r^2 calculée précédemment donne la longueur d'onde (notée m avec $i-k \leq m \leq i+k$) de la fenêtre « esclave » la plus corrélée à la longueur d'onde i sur l'instrument maître. Néanmoins un décalage en longueur d'onde entre deux instruments maître et esclave n'est jamais un multiple d'incrément spectral. Nous pouvons juste affirmer que cette longueur d'onde optimale se situe entre $m-1$ et $m+1$.

Afin d'obtenir une estimation plus fine de cette valeur, nous modélisons les corrélations r_{m-1}^2 , r_m^2 et r_{m+1}^2 par une parabole d'équation $y^2 = ax^2 + bx + c$ où x est la longueur d'onde et y la corrélation correspondante (Figure 83). Ainsi pour $x = m-1$, $y = r_{m-1}^2$, pour $x = m$, $y = r_m^2$ et pour $x = m+1$, $y = r_{m+1}^2$. Les coefficients de cette équation se calculent par une minimisation par les moindres carrés. On obtient alors aisément la longueur d'onde λ_i où la corrélation est maximale. C'est cette longueur d'onde qui correspond le mieux à i sur l'instrument maître.

Il suffit de réitérer ces calculs pour chaque longueur d'onde i de $A^{\text{Maître}}$ avec un déplacement de la fenêtre sur A^{Esclave} . A chaque i sur l'instrument maître correspond un λ_i sur l'esclave. La correction spectrale se termine par la modélisation de λ_i par une autre parabole d'équation :

$$\text{(éq. 95)} \quad \lambda_i = ti^2 + ui + v \text{ (méthodes des moindres carrés)}$$

Ainsi l'absorbance à la longueur d'onde i dans A^{Esclave} notée T_i^E est remplacée par l'interpolation à la longueur d'onde λ_i (calculée à partir de i dans (éq. 95)) de $T_{(\text{Valeur entière de } \lambda_i)}^E$ et $T_{(\text{Valeur entière de } \lambda_i)+1}^E$. Toutes les absorbances de la matrice des données spectrales de l'instrument esclave sont ainsi corrigées (notation $A^{\text{Esclave corrigé}}$). Le transfert se termine par une correction des intensités. Une régression linéaire simple est appliquée sur chaque longueur d'onde i entre $T_i^{E,\text{Cor}}$ et T_i^M telle que $T_i^{E,\text{Cor}} = a_i \cdot T_i^M + b_i$ (Figure 84). Néanmoins comme nous avons pu le montrer sur la dérivée de SAVITSKY-GOLAY, l'utilisation d'une fenêtre mobile pour corriger les décalages rend les calculs impossibles sur les k premières et les k dernières longueurs d'onde des spectres. La procédure présentée n'est donc applicable que sur $1+k \leq i \leq w-k$. Pour les extrémités, nous effectuerons une régression multilinéaire (MLR) entre la longueur d'onde considérée sur $A^{\text{Maître}}$ et les $(k+1)$ longueurs d'onde les plus proches dans A^{Esclave} .

Exemples : Pour $k=2$,

Régression entre T_1^M et $T_1^{E,\text{cor}}$, $T_2^{E,\text{cor}}$, $T_3^{E,\text{cor}}$.

Régression entre T_2^M et $T_1^{E,\text{cor}}$, $T_2^{E,\text{cor}}$, $T_3^{E,\text{cor}}$.

Régression entre T_{w-1}^M et $T_{w-2}^{E,\text{cor}}$, $T_{w-1}^{E,\text{cor}}$, $T_w^{E,\text{cor}}$.

Régression entre T_w^M et $T_{w-2}^{E,\text{cor}}$, $T_{w-1}^{E,\text{cor}}$, $T_w^{E,\text{cor}}$.

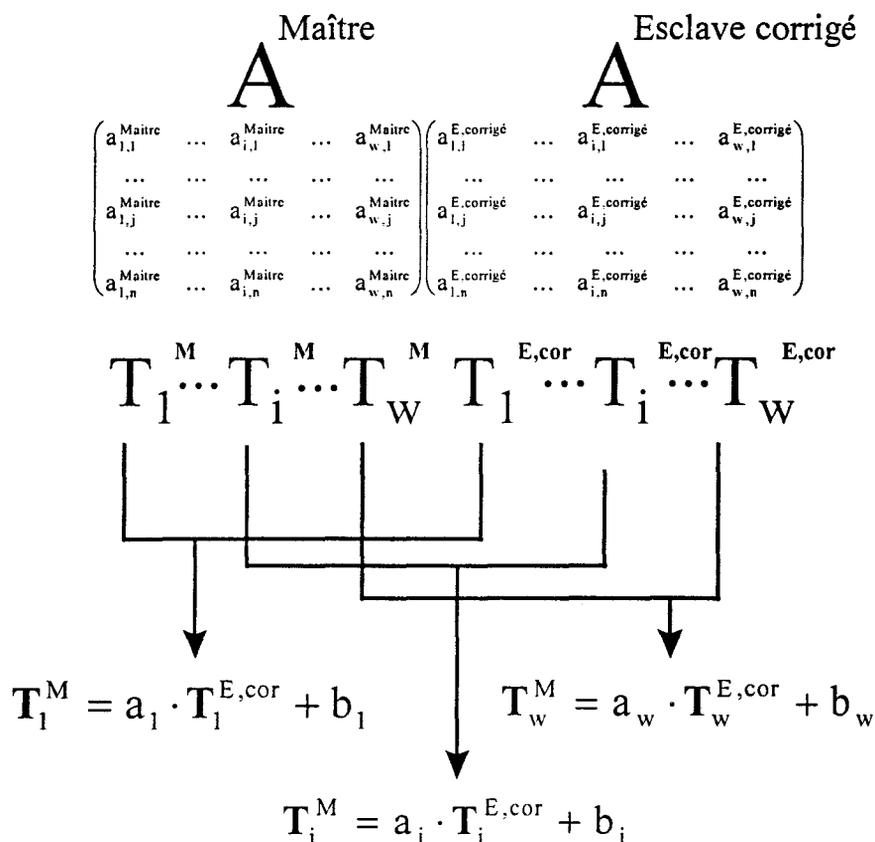


Figure 84 : Correction des intensités

Lorsque le calcul des paramètres de transfert est terminé, tout nouveau spectre acquis sur l'instrument esclave est d'abord corrigé par l'interpolation aux longueurs d'onde λ_i (valable pour les longueurs d'onde $1+k \leq i \leq w-k$). Chaque longueur d'onde du spectre corrigé est finalement transformée par les coefficients a_i et b_i correspondants. Les longueurs d'onde aux extrémités sont finalement corrigées séparément par les coefficients des régressions multilinéaires. Le spectre est utilisé dans l'équation de calibration de l'instrument maître pour prédire les concentrations. Cet algorithme de transfert est utilisé sur les instruments d'une firme mondialement connue qui est NIRSystem™.

ii) Application de la méthode de SHENK aux données synthétiques

Pour le test de ce dernier algorithme de transfert, la méthode est toujours la même. Les spectres de 20 standards des instruments maître et esclave nous permettent de calculer les paramètres de transfert à savoir la correction en longueur d'onde puis les différents coefficients des régressions a_i et b_i . Pour ces calculs, nous prenons $k=2$ soit une fenêtre de 5

longueurs d'onde. Les 30 spectres des échantillons de prédiction acquis sur les instruments esclaves sont corrigés. Les erreurs standards de prédiction sans et avec correction sont données dans le tableau ci-dessous (Tableau 38). Au vu des résultats la correction ne permet pas d'obtenir des erreurs suffisamment proches des valeurs optimales. Néanmoins, les erreurs obtenues sur l'instrument esclave 1 sont relativement faibles. Sur cet instrument, les variations spectrales sont linéaires sans décalage en longueur d'onde. Il semble donc que par sa nature univariée, cet algorithme ne puisse uniquement corriger que des variations simples d'absorbance. De plus toutes les longueurs d'onde sont corrigées indépendamment ce qui ne permet pas de prendre en compte le recouvrement spectral omniprésent en spectroscopie proche infrarouge. D'autres auteurs ont pu remarquer ces faits^(232,252).

Instrument	Constituant A SEP		Constituant B SEP		Constituant C SEP	
	sans correction	avec correction	sans correction	avec correction	sans correction	avec correction
Esclave 1	1.7029	0.0077	2.1294	0.0098	2.8260	0.0066
Esclave 2	0.4805	0.0320	0.6259	0.0330	1.5270	0.0350
Esclave 3	0.0720	0.0621	0.3959	0.0437	0.2511	0.1148
Esclave 4	1.6357	0.0517	1.4950	0.0445	3.0094	0.1321
Esclave 5	0.4118	0.0558	0.2046	0.0449	1.6529	0.1534
Valeurs optimales	0.0018		0.0016		0.0024	

Tableau 38 : Effet de la méthode de SHENK sur les SEP

Une équipe de chercheurs a tenté en vain d'améliorer cet algorithme⁽²⁴⁸⁾ en remplaçant la régression linéaire finale par une régression quadratique ($T_i^{E,cor} = a_i \cdot (T_i^M)^2 + b_i \cdot (T_i^M) + c_i$) ou par d'autres méthodes de régression telles que IRLS^(270,271,272) (*Iteratively Reweighted Least Squares*), LMS^(273,274,275) (*Least Median Squares*) ou encore LWR^(148,152,276,277) (*Locally Weighted Regression*). Le seul avantage de ces modifications est l'utilisation d'un nombre plus faible de standards.

5.3.3.3 Standardisation par un réseau de neurones

Une approche originale par la théorie des réseaux de neurones supervisés⁽²⁷⁸⁾ a été menée. Ce concept a permis de résoudre de nombreux problèmes quantitatifs ou qualitatifs en spectroscopie infrarouge^(279,280), en résonance magnétique nucléaire^(281,282), en spectroscopie de masse^(283,284). Cette recherche s'est orientée vers deux axes. Le premier est l'utilisation d'un réseau unique permettant de corriger les spectres de l'instrument esclave. Le second est une

modification de l'algorithme PDS par le remplacement des régressions PCR ou PLS par des réseaux de neurones de petites tailles.

Ainsi dans cette première partie, le but était de créer un réseau de neurones que l'on peut assimiler à une fonction dont la donnée d'entrée et la donnée de sortie sont respectivement un spectre de l'instrument esclave et le spectre du même échantillon sur l'instrument maître. Lorsque le réseau est construit, il est utilisé comme une boîte noire devant laquelle un spectre de l'instrument esclave est présenté pour obtenir sa correction. La Figure 85 représente l'architecture du réseau de neurones utilisé pour le transfert des données synthétiques.

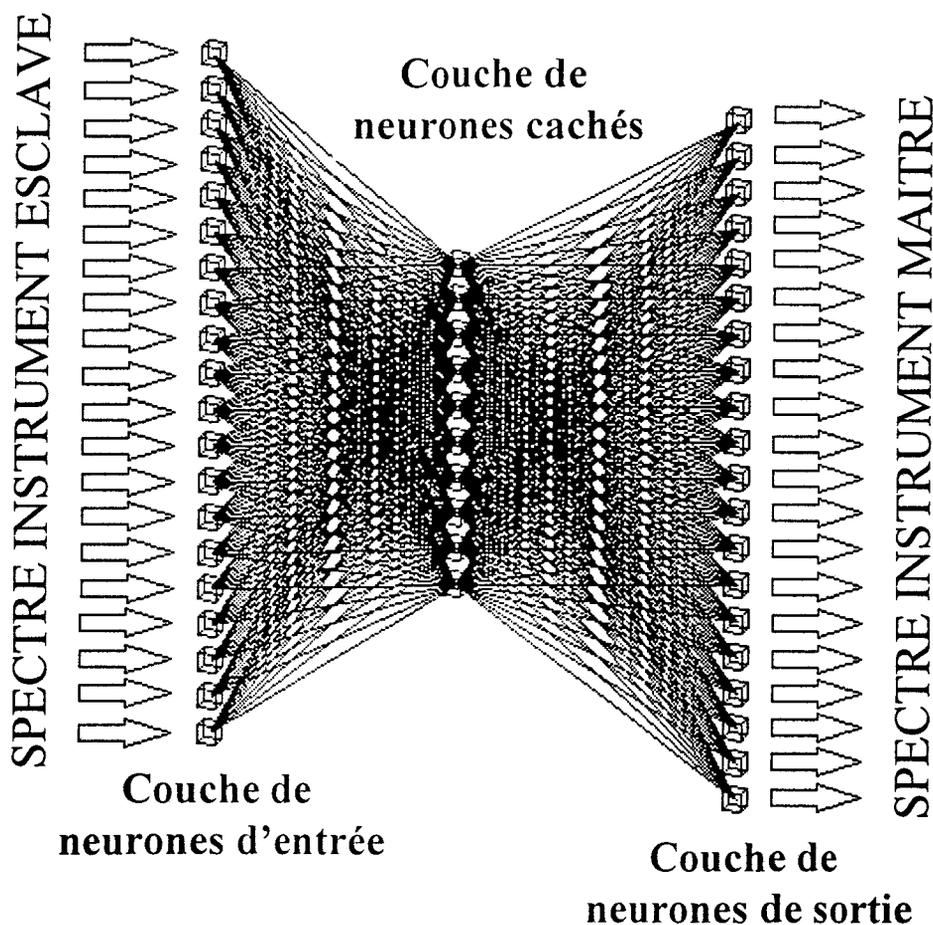


Figure 85 : Architecture du réseau utilisé pour le transfert.

Il est constitué dans notre cas de 100 neurones sur la couche d'entrée, 20 neurones sur la couche cachée et de 100 neurones sur la couche de sortie. Entre chaque neurone de deux couches successives, il y a une connexion possédant un poids. Celui-ci permet de transformer le signal qui transite entre les deux neurones considérés. Pour notre application, chaque

neurone de la couche d'entrée correspond à une longueur d'onde du spectre de l'instrument esclave. De la même façon, chaque neurone de la couche de sortie correspond à une longueur d'onde du spectre de l'instrument maître. La construction du réseau se divise en deux étapes : l'apprentissage et la validation. Pour l'apprentissage, le spectre infrarouge d'un échantillon est présenté devant la couche d'entrée du réseau. Le signal transite de neurone en neurone, affecté par les poids de chaque connexion. Les réponses des neurones de sortie sont ensuite comparées aux valeurs cibles associées au spectre initial c'est à dire au spectre du même échantillon mais acquis sur l'instrument maître. Cette différence est utilisée par la fonction d'apprentissage qui réajuste les poids de toutes les connexions afin de minimiser l'erreur en sortie de réseau. Le but est en fait de trouver la bonne combinaison de poids qui permet d'associer un spectre de l'instrument esclave au spectre de l'instrument maître. Après cette correction des poids, on présente à nouveau un spectre. Le processus est réitéré jusqu'à obtention d'un système stable. Les résultats du transfert des données synthétiques sont pour l'instant un peu décevants. Le nombre d'échantillons standards semble trop faible par rapport au nombre de connexions dans le réseau. Néanmoins, cette approche permet assez souvent d'obtenir de meilleurs résultats que les techniques multivariées classiques⁽²⁸⁵⁾. Elle gère plus facilement les problèmes de linéarité et génère des modèles souvent plus robustes face aux variations. De plus, ce type de réseau a déjà démontré son efficacité pour la correction des dérives sur un couplage pyrolyse/spectroscopie de masse⁽²⁸⁶⁾. Pour toutes ces raisons il semble intéressant de poursuivre dans cette voie.

La deuxième approche vise à remplacer les régressions PCR ou PLS dans l'algorithme PDS par un réseau de neurones de faible dimension (Figure 86). On conserve le système de fenêtre mobile. Seule la manière de relier les longueurs d'onde de l'instrument esclave à l'instrument maître change. Dans ce cas le nombre d'échantillons standards est plus en accord avec le nombre de connexions du réseau. Néanmoins, comme pour les régressions PCR et PLS, les réseaux peuvent être touchés par un overfitting. Afin de prévenir cet effet, il convient de maîtriser l'apprentissage sous peine d'une détérioration importante de l'analyse quantitative après le transfert. Toute l'efficacité de cette modification dans PDS réside sur ce problème.

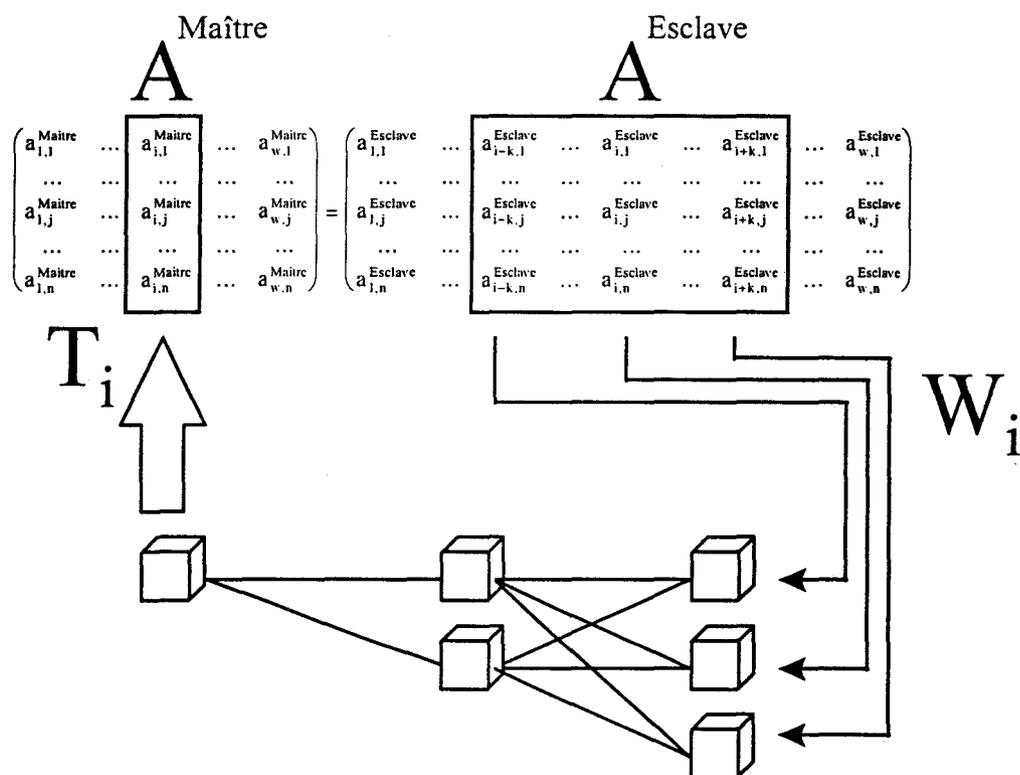


Figure 86 : Modification de l'algorithme PDS par les réseaux de neurones.

Les premiers résultats figurent dans le tableau ci-dessous (Tableau 39). La capacité de l'algorithme est encore limitée par rapport à la PDS basée sur PCR ou PLS mais il y a encore de nombreux paramètres à maîtriser.

	Constituant A SEP	Constituant B SEP	Constituant C SEP
PDS (Réseau de neurones)	0,0127	0,0238	0,0189
Valeur optimale (Instrument Maître)	0,0018	0,0016	0,0024

Tableau 39 : Erreur de prédiction pour la PDS modifiée par les réseaux de neurones.

Une étude systématique portant sur le nombre de neurones sur la couche cachée, les fonctions de transfert du signal entre neurones et les fonctions d'apprentissage doit permettre un transfert efficace et robuste.

5.3.3.4 Synthèse des résultats

Afin de rendre compte de la validité des transferts, le tableau ci-dessous regroupe pour les trois constituants les moyennes des erreurs standards des instruments esclaves en fonction de l'algorithme utilisé (Tableau 40).

Algorithme	Constituant A SEP moyen	Constituant B SEP moyen	Constituant C SEP moyen
Pente/Biais	0,0422	0,2528	0,1333
DS	0,0034	0,0033	0,0035
PDS (PCR) sans centrage	0,0045	0,0025	0,0028
PDS (PCR) avec centrage	0,0024	0,0027	0,0025
PDS (PLS)	0,0024	0,0022	0,0030
SHENK	0,0419	0,0352	0,0884
Valeur optimale (Instrument Maître)	0,0018	0,0016	0,0024

Tableau 40 : Synthèse des résultats du transfert sur les spectres synthétiques

On remarque clairement que l'algorithme le plus approprié est la méthode PDS basée sur la régression PLS ou sur PCR avec un centrage des données. Pour des variations spectrales importantes et complexes imposées dans notre étude, les méthodes pente/biais et SHENK sont inutilisables. Néanmoins de telles variations n'ont jamais cours, car les constructeurs font des efforts considérables pour concevoir des instruments similaires pour un même modèle (normes ISO). Dans ces conditions, tous les algorithmes ont permis d'obtenir des transferts relativement satisfaisants. Néanmoins, cette étude a montré la capacité du transfert PDS à gérer des variations complexes, ce qui rendra l'algorithme encore plus robuste pour des variations communément rencontrés.

5.4 Application du transfert au dosage d'hydrocarbures.

5.4.1 Introduction

Cette étude systématique sur des données synthétiques se devait d'être validée par des transferts de calibration avec la méthode PDS sur des données réelles. La première analyse quantitative qui a été choisie est celle des essences car l'industrie pétrolière exploite largement le domaine du proche infrarouge^(287,288,289). L'intérêt de cette série de spectres est son utilisation par de nombreux chercheurs en vue d'effectuer une comparaison systématique de nouveaux algorithmes ou leurs améliorations pour le transfert de calibration^(232,236,266,267). A cet effet, ils sont à disposition sur un site FTP d'Internet⁽²⁹⁰⁾. Les échantillons utilisés sont des essences reconstituées à partir d'un mélange de trois hydrocarbures saturés (notés S1, S2, S3) et de deux aromatiques (notés A1 et A2). Les spectres NIR ont été acquis sur deux spectromètres dispersifs provenant du même constructeur. Dans cette étude, le premier spectromètre noté instrument A sera considéré comme instrument maître alors que le second noté instrument B sera l'instrument esclave sur lequel on veut transférer la calibration. Le domaine spectral est défini de 800 à 1600 nm avec un incrément de 2 nm.

5.4.2 Analyse quantitative sur l'instrument maître

Pour effectuer ce transfert, trente échantillons ont été utilisés. Cet ensemble a été divisé en deux groupes, 24 échantillons pour la calibration et 6 pour son transfert. Les six échantillons standards ont été sélectionnés par la méthode « leverage ». Le Tableau 41 donne les concentrations des constituants à doser pour chaque échantillon ainsi que leur utilisation. Elles sont exprimées en pourcentage volume/volume. Pour l'analyse quantitative, nous travaillons avec la régression PLS sur les spectres bruts non dérivés. Les résultats de la calibration sur les cinq espèces à partir des 24 échantillons figurent dans le Tableau 42.

CONCENTRATIONS (Vol/Vol %)						
Echantillon N°	Utilisation [□]	S1	S2	A1	A2	S3
1	A	10.381	9.983	45.771	12.949	20.915
2	A	14.024	21.943	29.942	11.089	23.003
3	A	22.007	25.866	37.900	13.209	1.019
4	T	11.909	27.479	44.411	13.264	2.938
5	A	13.791	21.624	37.384	6.864	20.337
6	T	14.016	25.974	49.950	5.022	5.038
7	T	18.070	29.897	30.023	5.020	16.990
8	A	17.957	21.947	33.919	7.023	19.154
9	A	9.935	17.853	42.450	12.894	16.868
10	T	14.194	29.809	33.859	14.871	7.266
11	A	22.976	19.073	43.968	12.990	0.993
12	T	29.839	17.853	37.736	13.573	0.999
13	A	17.983	18.016	30.021	10.984	22.996
14	A	29.772	18.093	34.102	11.077	6.956
15	A	25.881	29.862	30.180	7.092	6.986
16	A	25.964	14.018	33.885	5.000	21.133
17	A	21.673	18.916	41.700	5.110	12.601
18	A	9.858	25.937	33.983	9.025	21.196
19	A	21.211	24.807	47.758	3.941	2.283
20	A	14.093	9.920	34.215	8.835	32.937
21	A	10.033	25.646	37.818	5.097	21.405
22	A	12.949	24.766	48.365	10.558	3.361
23	A	17.728	13.918	37.570	9.809	20.975
24	A	9.920	18.260	49.666	7.237	14.917
25	A	23.715	23.777	45.790	4.023	2.696
26	A	29.814	18.177	29.959	12.722	9.327
27	A	10.078	14.207	33.461	15.282	26.972
28	A	18.260	14.007	38.155	4.909	24.669
29	A	15.827	28.184	42.569	12.394	1.025
30	T	13.923	9.975	41.833	11.312	22.956
	Minimum	9.858	9.920	29.942	3.941	0.993
	Maximum	29.839	29.897	49.950	15.282	32.937
	Moyenne	17.698	20.613	38.365	9.072	12.689

□ : T = transfert, A = calibration ou prédiction.

Tableau 41 : Concentrations de référence pour les 30 échantillons d'essence.

Le nombre trop faible d'échantillons à notre disposition ne nous permet pas de caractériser la qualité prédictive de l'analyse sur des échantillons inconnus à la calibration. C'est donc l'erreur RMSD de prédiction de la validation croisée (*cross validation*) calculée sur les échantillons de calibration qui figure dans ce tableau. Pour chaque constituant, 9 vecteurs loading étaient nécessaires. La comparaison de ces erreurs aux gammes de concentration

montre que le dosage de ces essences se fait dans de bonnes conditions. Ce problème analytique est très intéressant, car les contributions spectrales des espèces à doser se situent sur des domaines limités. Le recouvrement est optimum.

	S1	S2	A1	A2	S3
Erreur de prédiction RMSD ² (cv)	0.4459	0.0974	0.0693	0.1404	0.3788
Nombre de loading	9	9	9	9	9

□ : Vol/Vol %

Tableau 42 : Erreurs de prédiction pour la calibration de l'instrument maître.

La Figure 87 permet d'apprécier les différences spectrales entre l'instrument maître (instrument A) et l'instrument esclave (instrument B) pour les échantillons N°1, 2 et 3. Nous remarquons des différences importantes pour les longueurs d'onde de 1450 à 1600 nm.

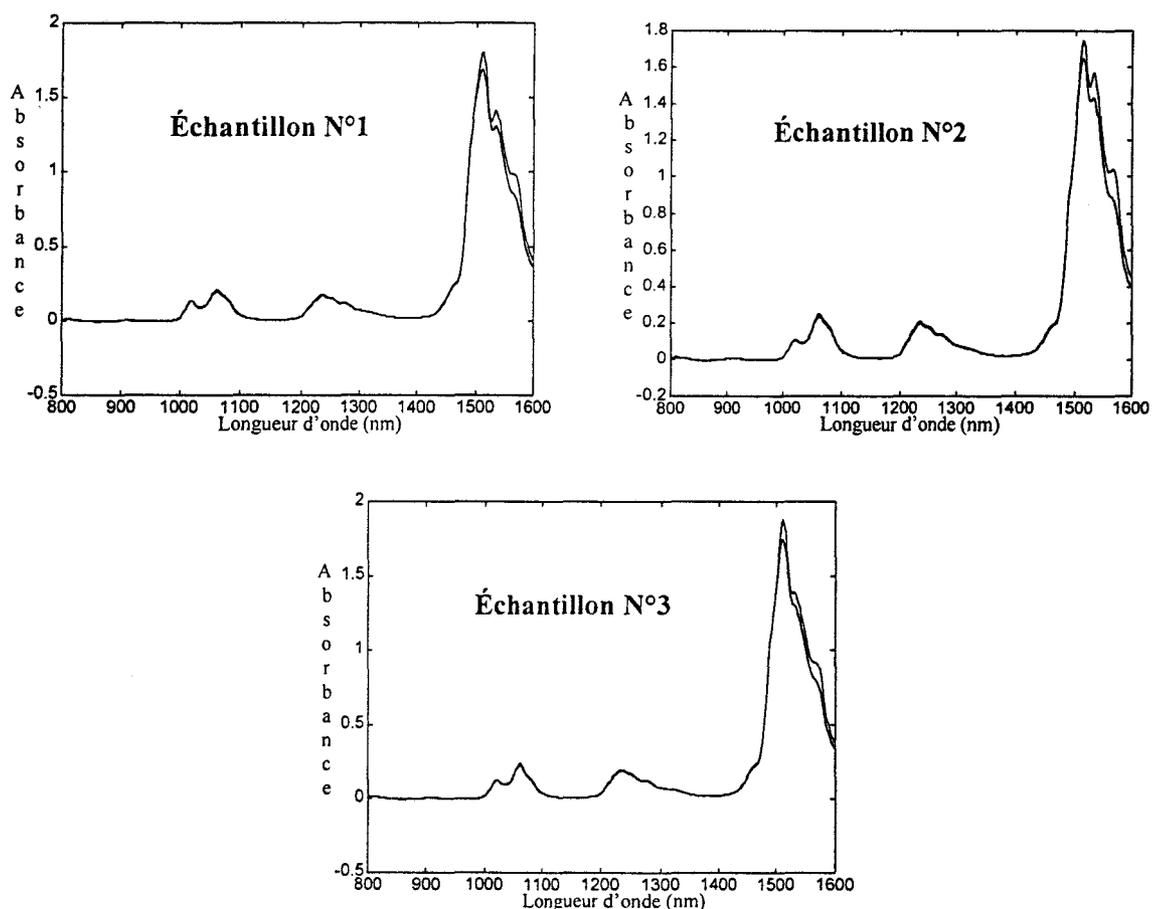


Figure 87 : Comparaison des réponses spectrales des instruments A et B.

Les différences aux longueurs d'onde plus faibles semblent moins importantes car les extinctions y sont moins fortes. Néanmoins, les spectres de différence entre instruments A et B pour les échantillons N°1,2 et 3 montrent des variations non négligeables sur les domaines 1000-1100 nm et 1200-1300 nm (Figure 88).

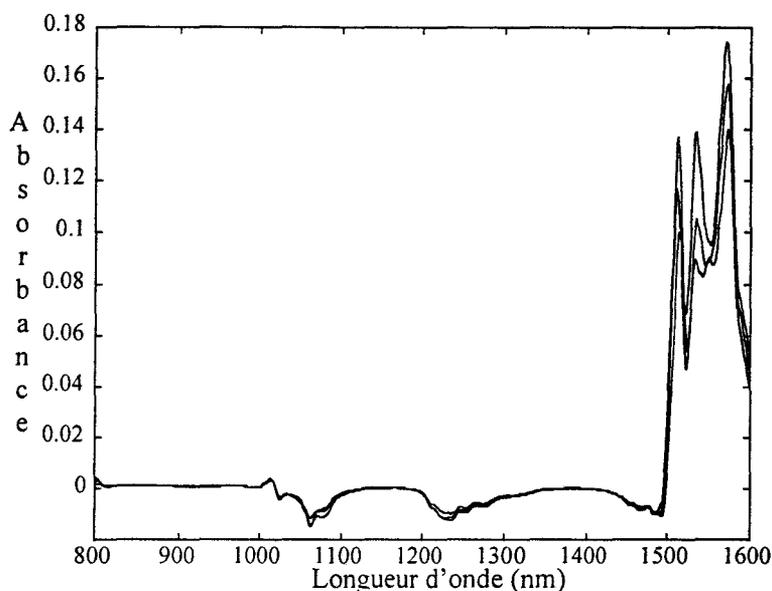


Figure 88 : Spectres de différence entre l'instrument A et B.

Ces différences sont d'ailleurs d'autant plus importantes que les variations spectrales sur ces domaines de longueur d'onde sont faibles même sur de fortes concentrations. Afin de montrer l'impact de ces variations sur l'analyse quantitative, nous avons prédit les concentrations des 24 échantillons à partir de leurs spectres acquis sur l'instrument B et de l'équation développée sur l'instrument A. Les résultats obtenus rendent le transfert de calibration indispensable (Tableau 43).

	S1	S2	A1	A2	S3
Erreur de prédiction Instrument A	0.4459	0.0974	0.0693	0.1404	0.3788
Erreur de prédiction Instrument B non corrigé	13.6510	3.7422	1.9572	4.9660	13.3200

Tableau 43: Effet des variations spectrales sur un instrument non corrigé.

5.4.3 Le transfert de la calibration

La première méthode qui a été testée sur ces données spectrales est la PDS basée sur la régression PCR avec un centrage préalable. Nous avons mis en place une étude systématique sur l'influence du nombre de standards et du nombre de longueurs d'onde à considérer pour la fenêtre. Pour chaque expérience, nous calculons la matrice de transfert qui sert à corriger les spectres de l'instrument B. Ceux-ci sont ensuite utilisés avec l'équation de l'instrument A pour obtenir finalement une erreur de prédiction. Les résultats de cette étude sont consignés dans le tableau ci-dessous.

Paramètres (v_ϕ)	S1 (RMSD)	S2 (RMSD)	A1 (RMSD)	A2 (RMSD)	S3 (RMSD)
(6_7)	0.8468	0.1665	0.1890	0.1949	0.8337
(6_5)	0.8336	0.1940	0.1202	0.1599	0.7746
(6_3)	1.2633	0.3351	0.1063	0.3320	1.3719
(5_7)	1.1615	0.1649	0.2091	0.2945	1.1497
(5_5)	0.8649	0.1745	0.1241	0.2300	0.8386
(5_3)	1.2607	0.3449	1.1160	0.3459	1.3579
(4_7)	1.1632	0.2123	0.2282	0.3380	0.9845
(4_5)	1.0490	0.2309	0.1262	0.2212	0.9028
(4_3)	1.3658	0.3409	0.0958	0.2971	1.4755
(3_7)	1.8233	0.3648	0.6299	1.2920	1.0273
(3_5)	3.4174	0.6507	0.8256	1.5405	2.3847
(3_3)	3.7498	0.9170	0.5690	1.0000	2.9262
Valeurs cibles	0.4459	0.0974	0.0693	0.1404	0.3788

(Erreur de prédiction instrument A)

(v_ϕ) : v = nombre d'échantillons standards pour le transfert, ϕ = taille de la fenêtre en longueur d'onde.

Tableau 44: Erreurs de prédiction (RMSD) après le transfert par la méthode PDS (basée sur PCR et le centrage des données).

Il est important de prendre un maximum d'échantillons pour le transfert sous peine d'une détérioration rapide des capacités prédictives de l'analyse quantitative. Le choix de la fenêtre optimale est difficile, car les résultats des dosages ne convergent pas vers le même nombre de longueur d'onde. Comme pour la mise en place d'une analyse quantitative, il conviendra d'effectuer une étude systématique afin de définir ces paramètres. Pour ce transfert, un

nombre de 6 échantillons standards et une fenêtre de 5 longueurs d'onde pour la PDS donnent un bon compromis. La Figure 89 montre les spectres de différence entre l'instrument A et l'instrument B corrigé par la PDS sur les 3 échantillons précédents. Une comparaison avec la Figure 88 permet d'observer des différences 100 fois moins importantes lorsque les spectres de l'instrument esclave sont corrigés par la PDS. De cette façon, nous pouvons dire que le transfert ajuste les spectres de l'instrument esclave pour ressembler à ceux de l'instrument maître.

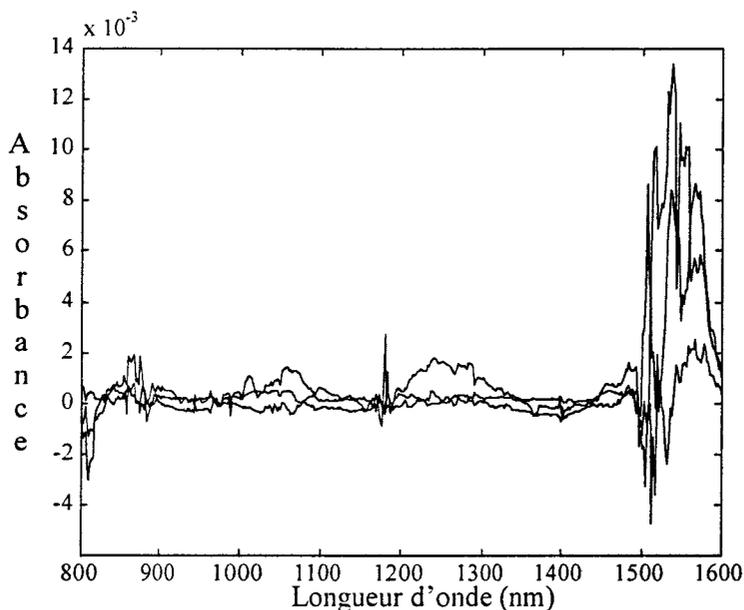


Figure 89 : Spectres de différence entre l'instrument A et l'instrument B corrigé.

Nous avons ensuite effectué une étude sur l'influence du centrage des données par leur moyenne pour la PDS (Tableau 45). Les résultats obtenus renforcent sa nécessité pour le transfert avec l'algorithme PDS.

Algorithme	S1 (RMSD)	S2 (RMSD)	A1 (RMSD)	A2 (RMSD)	S3 (RMSD)
PDS (PCR avec centrage)	0.8336	0.1940	0.1202	0.1599	0.7746
PDS (PCR sans centrage)	1.7372	0.2208	0.4066	0.9140	2.3961

Tableau 45: Influence du centrage des données sur la PDS basée sur PCR.

Sur les données synthétiques, la régression PLS pour l'algorithme PDS avait donné de très bons résultats. Il nous est donc apparu intéressant de comparer les deux types de régression pour PDS sur les données réelles. Le Tableau 46 donne les résultats de l'analyse quantitative

issuë du transfert à partir de 6 échantillons standards et un nombre variable de longueur d'onde dans la fenêtre. On remarque aisément que la PDS basée sur la régression PCR est largement plus efficace que la régression PLS pour le transfert de l'analyse quantitative des essences. Ces observations vont dans le sens d'une étude systématique menée par GARTHWAITE⁽²⁹¹⁾ *et al* sur les régressions multivariées. En effet, ils ont montré que pour un faible nombre de variables explicatives (fenêtre de longueur d'onde sur l'instrument esclave dans notre cas), la régression PCR donnait des erreurs plus faibles que la régression PLS.

	Paramètres (ν_ϕ)	S1 (RMSD)	S2 (RMSD)	A1 (RMSD)	A2 (RMSD)	S3 (RMSD)
PDS basée sur PCR	(6_7)	0.8468	0.1665	0.189	0.1949	0.8337
	(6_5)	0.8336	0.194	0.1202	0.1599	0.7746
	(6_3)	1.2633	0.3351	0.1063	0.332	1.3719
PDS basée sur PLS	(6_7)	1.9772	0.4119	0.2979	0.2862	1.7045
	(6_5)	2.2915	0.3177	0.3421	0.8011	1.8202
	(6_3)	3.0077	0.3608	0.41	1.1371	2.0515
Valeurs cibles (Erreur de prédiction instrument A)		0.4459	0.0974	0.0693	0.1404	0.3788

(ν_ϕ) : ν = nombre d'échantillons standards pour le transfert, ϕ = taille de la fenêtre en longueur d'onde.

Tableau 46 : Influence des régressions PCR et PLS sur la PDS.

Les résultats obtenus sur les données synthétiques semblent être en contradiction avec cette interprétation. Il n'en est rien, car ces chercheurs ont aussi démontré que la régression PCR donnait des erreurs plus importantes que la régression PLS sur des données spectrales bruitées. Le rapport signal sur bruit est effectivement beaucoup plus favorable sur les spectres proche infrarouge des essences que sur les données synthétiques.

5.5 Transfert de calibration sur l'analyseur de lait

Lorsque l'étude de faisabilité du transfert s'est terminée, il n'existait pas encore de copie du premier prototype afin d'effectuer un transfert de calibration sur un appareil de présérie. Les résultats qui suivent sont donc issus d'un même appareil ayant subis des modifications instrumentales incontrôlées pendant la recherche de la sonde optimale. Le premier lot de données utilisé provient de l'analyse quantitative de laits reconstitués sur la

dernière version de la sonde transflexion. Ces spectres seront considérés comme provenant de l'instrument maître pour le transfert. Le deuxième lot provient aussi de l'analyse quantitative de laits reconstitués sur la sonde transflexion mais après déplacement de l'analyseur pour une campagne de mesure à l'extérieur de notre laboratoire de recherche. Ces données seront considérées comme provenant de l'instrument esclave. Nous avons précédemment montré que la sonde transflexion permettait d'obtenir un dosage de qualité pour les matières grasses et les protéines des laits reconstitués à partir de spectres bruts (pages 112 et 115). Si ces équations de calibration sont appliquées aux spectres du deuxième lot d'échantillons, nous observons une détérioration importante des erreurs standards de prédiction (Tableau 47).

	SEP MG (g.kg)	SEP MP (g.kg)
Analyse quantitative avant le déplacement (Instrument maître)	0.54	0.32
Analyse quantitative après le déplacement (Utilisation des calibrations de l'instrument maître)	2.31	2.15
Analyse quantitative après le déplacement (Nouveau calibrage du système)	0.56	0.30

Tableau 47 : Amplification des erreurs due aux variations instrumentales.

Nous pouvons même dire que les régressions développées avant le déplacement deviennent inutilisables pour les dosages. Ces résultats indiquent une grande sensibilité des régressions multivariées face aux variations instrumentales. Néanmoins, la qualité spectrale n'est pas remise en cause car un nouveau calibrage permet de retrouver des erreurs homogènes par rapport aux dosages initiaux. Seules les variations de signal et les décalages en longueur d'onde sont incriminés. Pour ce cas de figure, le transfert de calibration est inutile car un nouveau calibrage de l'instrument est possible. La méthode de préparation est simple et rapide. Le nombre d'échantillons de calibration est restreint. Pour le dosage des matières grasses et des protéines dans le lait cru, nous ne pourrions pas nous permettre pour des raisons de temps et de coût d'effectuer à nouveau le calibrage de l'analyseur. Dans ce cas, le transfert de calibration devient obligatoire sous peine de non-conformité par rapport au cahier des charges. Les spectres de deux échantillons à teneurs en matières grasses et protéines

identiques mais acquis avant et après le déplacement, sont présentés sur la figure ci-dessous (Figure 90). Les spectres semblent se superposer mais une analyse plus fine montre un décalage en longueur d'onde d'environ 2 nm ainsi qu'une légère diminution du signal. Lors du transport, l'analyseur a subi de nombreuses vibrations favorisant le décalage du monochromateur.

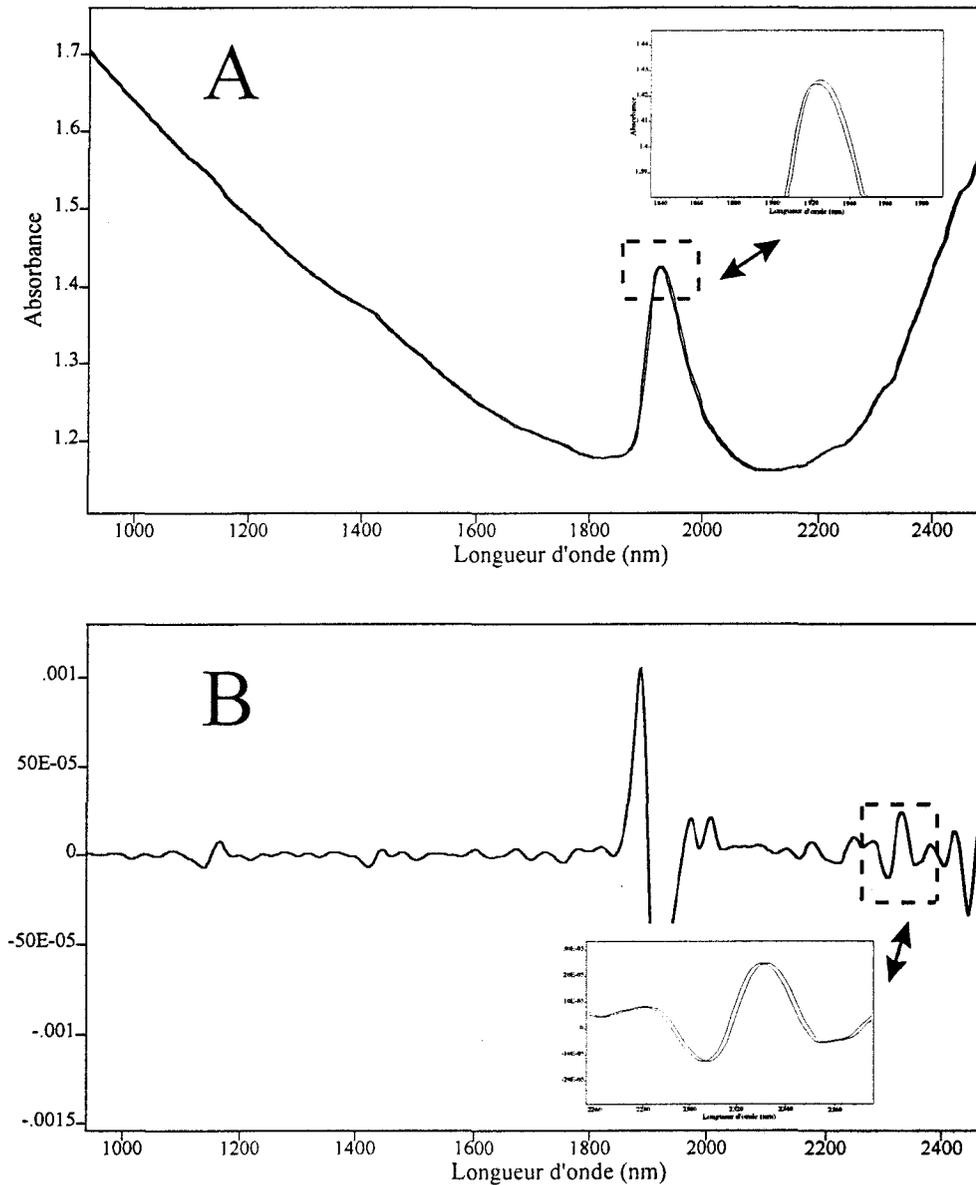


Figure 90 : Comparaison des spectres bruts (A) et de spectres dérivés (B) avant et après le déplacement de l'analyseur.

De plus, la sonde est régulièrement démontée afin d'assurer un nettoyage efficace entre chaque grande série d'analyses. La fibre optique n'est donc pas forcément replacée dans sa

position exacte. Nous avons appliqué la méthode PDS basée sur PLS et PCR avec ou sans centrage des données spectrales pour le transfert de calibration. Le calcul de la matrice de transfert F est mené sur 6 standards et une fenêtre spectrale de 5 longueurs d'onde. Il est en effet inutile et même périlleux (amplification des erreurs) de prendre une fenêtre plus grande alors que le décalage en longueurs d'onde est de 2 nm. L'effet de la correction des spectres par l'algorithme PDS est donné au Tableau 48. Nous observons à nouveau la nécessité du centrage des données spectrales pour la PDS basée sur PCR. Conformément à l'application précédente la régression PLS reste moins efficace que PCR pour le transfert de calibration. Seul l'algorithme PDS basée sur PCR avec un centrage des données spectrales est efficace puisque les erreurs standards sont conformes aux spécifications du cahier des charges. Il faut néanmoins compter sur une amplification non négligeable des erreurs. Le transfert de calibration est donc validé pour le dosage de laits reconstitués.

	SEP MG (g/kg)	SEP MP (g/kg)
Spectres non-corrigés	2.31	2.15
Spectres corrigés par PDS (basée sur PLS)	1.44	1.01
Spectres corrigés par PDS (basée sur PCR sans centrage)	1.93	1.26
Spectres corrigés par PDS (basée sur PCR avec centrage)	0.79	0.51

Tableau 48 : Transfert de calibration par la méthode PDS.

5.5 Conclusion

L'objectif de cette recherche était d'effectuer une étude systématique sur l'ensemble des algorithmes de transfert proposés dans la littérature. Après une présentation des conditions de mise en place d'un transfert de calibration, l'application aux données synthétiques a permis d'évaluer de manière relative le potentiel de chaque algorithme face à plusieurs types de variations spectrales. La correction appelée « pente/biais » est restée la plus inefficace face à des variations importantes comme les décalages en longueur d'onde. Pour les algorithmes de correction des spectres, la méthode de SHENK est la moins intéressante car l'amplification des erreurs de prédiction reste importante. Cette inefficacité est attribuée au caractère univarié des étapes de correction des décalages en longueur d'onde et des intensités. Les algorithmes de transfert basés sur la régression multivariée tels que la standardisation directe (DS) ou la standardisation pas à pas (PDS), ont donné des résultats satisfaisants. Un léger avantage a

néanmoins était donné à la méthode PDS qui modélise mieux les variations spectrales grâce à de nombreuses régressions sur une fenêtre en longueur d'onde mobile comparée à une régression unique sur tout le domaine spectral pour la méthode DS. Nous avons démontré la nécessité du centrage des données spectrales lorsque l'algorithme PDS utilise la régression PCR. A ce stade de l'étude, la supériorité de la régression PCR ou PLS dans l'algorithme de transfert PDS n'a néanmoins pas pu être déterminé. Deux nouvelles générations d'algorithmes de transfert ont ensuite été développées à partir des réseaux de neurones. La première utilise un réseau de neurones unique ayant pour couche d'entrée le spectre d'un échantillon acquis sur l'instrument esclave et pour couche de sortie ce même spectre corrigé homogène à l'instrument maître. La deuxième remplace les techniques de régressions classiques telles que PLS ou PCR dans PDS par un réseau de neurones. Les résultats restent à ce jour relativement médiocres à cause du manque de temps pour l'optimisation de nombreux paramètres qu'impose le développement d'un réseau de qualité. Ce travail aboutit à l'établissement de la faisabilité du transfert par l'algorithme PDS sur deux protocoles analytiques : le dosage de composés aromatiques et saturés dans les essences puis le dosage des matières grasses et des protéines dans les laits reconstitués sur l'analyseur.

Lors de la standardisation d'un appareil par rapport à un autre, il faudra garder à l'esprit l'amplification non négligeable des erreurs de prédiction. Compte tenu de cet effet, il convient de réduire au maximum les erreurs standards de prédiction initiales pour obtenir, après transfert, des résultats fiables. Une attention particulière devra être portée sur l'instrumentation afin d'éviter toutes variations liées au montage, démontage ou déplacement de l'analyseur. La suite de ce travail débouchera sur l'application et la vérification du bon fonctionnement de l'algorithme sur le dosage du lait cru dans les conditions d'utilisation journalière de l'analyseur.

CONCLUSION GÉNÉRALE

L'objectif de ce travail était double puisqu'il fallait dans un premier temps valider par la chimiométrie le développement d'un analyseur de lait travaillant dans le proche infrarouge puis étudier la faisabilité du transfert de calibration nécessaire aux appareils de présérie.

L'originalité de cette recherche se situe à plusieurs niveaux. Il est tout d'abord très rare de pouvoir suivre dans une telle recherche le développement d'un analyseur spectrométrique dédié à l'analyse quantitative de la définition du cahier des charges jusqu'à la conception du prototype final. De plus, l'analyse quantitative par spectrométrie exige une optimisation de la technique d'échantillonnage alors que l'acquisition des données spectrales est souvent effectuée sur des spectromètres commerciaux de série qui ne sont pas nécessairement optimisés pour l'étude des produits d'intérêt. La littérature montre aussi une grande richesse en analyses quantitatives développées en laboratoire de recherche sur des produits variés en se souciant très rarement du transfert de calibration nécessaire à son exploitation sur un autre site. L'augmentation des analyseurs spectroscopiques industriels inverse actuellement cette tendance. Nous avons finalement effectué une recherche sous contraintes dans le sens où le cahier des charges a défini les conditions d'utilisation de l'analyseur, mais aussi son coût et son intégration dans la salle de traite.

Le caractère pluridisciplinaire de ces travaux de recherche a imposé le développement d'une bibliographie pour les trois premiers chapitres sur la technologie laitière, les principes fondamentaux de la spectroscopie proche infrarouge, son instrumentation et les traitements mathématiques des données spectrales.

Pour la validation des maquettes de faisabilité par la chimiométrie, nous avons mis en place une procédure basée sur le contrôle de la répétabilité, de la linéarité et des capacités prédictives des calibrations sur le dosage des matières grasses et des protéines. Ce schéma de principe nous a guidé pour chaque étape du développement des modifications apportées aux configurations optiques et électroniques. Les résultats de ces différents tests ont rapidement montré l'importance que prenait la sonde optique au sein de l'analyseur. Le caractère diffusant du lait a compromis l'utilisation de sondes basées sur l'optique géométrique pure comme la sonde transmission. La volonté constante d'augmenter l'intégration du signal diffusé sur tout l'espace a engendré la sonde diffusion. Conformément au cahier des charges, le dosage des

matières grasses et des protéines a atteint avec cette dernière sonde des erreurs inférieures à 1 g/kg sur un domaine spectral compatible avec la transmission des fibres optiques. Bien que la sonde optique soit le centre de ce travail, l'analyseur doit aussi sa capacité prédictive à l'optimisation du monochromateur et de la chaîne électronique de détection. Nous avons en effet démontré la nécessité du filtre d'ordre, d'une résolution moyenne ou encore d'une détection synchrone numérique pour l'analyseur. L'ensemble de ces résultats permet de dire que le premier objectif a été atteint.

La dernière partie de ce mémoire a porté sur l'étude de la faisabilité du transfert de calibration pour l'analyseur. La création de données spectrales synthétiques a permis de maîtriser les variations d'intensité et de longueur d'onde entre les instruments visés pour le transfert. Une étude systématique des algorithmes de transfert a évalué les capacités relatives de chacun. Il n'existait pas dans la littérature de comparaison méthodique d'où l'intérêt de cette recherche. La correction dite de « pente et de biais », algorithme qui corrige les concentrations, s'est révélée inadaptée face aux variations spectrales importantes. Pour les algorithmes de corrections spectrales, la méthode de SHENK est restée la plus inefficace à cause de son caractère unilinéaire. La standardisation directe (DS ou PDS) a obtenu les résultats les plus intéressants malgré l'amplification des erreurs standards de prédiction. La faisabilité du transfert de calibration à partir de la standardisation pas à pas (PDS) a finalement été estimée sur des données réelles issues du dosage des hydrocarbures et des laits reconstitués sur l'analyseur. Les meilleurs résultats ont été obtenus grâce à l'utilisation simultanée de la régression PCR et du centrage des données dans l'algorithme. Nous pouvons finalement considérer que l'objectif du transfert a été atteint. Néanmoins, l'amplification des erreurs standards de prédiction toujours présente lors d'un transfert, relativise la qualité des résultats de dosage obtenue avec le prototype pour les laits crus sur la sonde diffusion. Dans ces conditions, les erreurs standards seraient en effet supérieures aux spécifications fixées par le cahier des charges après le transfert sur un appareil de présérie. Les solutions envisageables sont la diminution des erreurs standards de prédiction sur le prototype et l'amélioration de l'efficacité des algorithmes de transfert. Dans le premier cas, l'optimisation de l'analyseur doit se poursuivre par des études systématiques sur l'influence du réseau de diffraction, du détecteur, du système d'éclairage et même du type de fibre optique car le marché des chalcogénures est en pleine explosion. Dans le deuxième cas, nous comptons sur la

publication de nouveaux algorithmes plus performants ainsi que sur notre contribution à cette recherche par l'introduction des réseaux de neurones pour le transfert de calibration.

En résumé, nous pouvons dire que le prototype développé dans cette recherche a les performances fixées par le cahier des charges. Il faudra néanmoins vérifier son comportement face à une utilisation quotidienne en salle de traite avec toutes les contraintes que cela impose. Au niveau des analyses quantitatives, il faudra aussi étudier l'influence de la race sur les erreurs de prédiction. Les différentes dispersions en taille des globules gras nous obligeront-elles à développer plusieurs calibrations ? Cette question se posera aussi pour les saisons qui influent sur la matrice du lait.

La portabilité, la résistance à la température et à l'humidité, la délocalisation des analyses sur douze sites éloignés, et finalement le coût explique l'engouement des industries agro-alimentaires pour cette réalisation. A leur demande, des versions modifiées de ce prototype sont d'ailleurs actuellement à l'étude pour le contrôle en ligne de nombreux produits.

TABLE DES ILLUSTRATIONS

Figure 1 : Les lipides dans le lait.....	12
Figure 2 : Les protéines.....	14
Figure 3 : Séquence des acides α -aminés de la caséine α	14
Figure 4 : Le lactose α	17
Figure 5 : Représentation d'une salle de traite.....	18
Figure 6 : Le modèle de l'oscillateur mécanique.....	19
Figure 7 : Observations spectrales des transitions pour l'oscillateur harmonique.	23
Figure 8 : Observations spectrales des transitions pour l'oscillateur anharmonique.	25
Figure 9 : Les modes fondamentaux de vibration de la molécule d'eau.	27
Figure 10 : Le filtre interférentiel.....	34
Figure 11 : Montage classique de filtres interférentiels.	35
Figure 12 : Systèmes de filtres oscillants.	36
Figure 13 : Filtre opto-acoustique	36
Figure 14 : Interféromètre de MICHELSON.	38
Figure 15 : Efficacité des réseaux holographiques et gravés.	39
Figure 16 : Profils des réseaux de diffraction.....	40
Figure 17 : Influence du blaze sur la courbe d'efficacité.....	41
Figure 18 : Sélection des longueurs d'onde par un monochromateur.....	41
Figure 19 : Superposition des ordres de diffraction d'un réseau.	43
Figure 20 : Transmission du signal dans une fibre optique.....	44
Figure 21 : Atténuation du signal optique dans une fibre de silice.	45
Figure 22 : Schéma de principe d'un détecteur photoconductif.	47
Figure 23 : Schéma de principe d'un détecteur photovoltaïque.....	48
Figure 24 : Défectivité des détecteurs PbS et InGaAs.....	48
Figure 25 : Interactions rayonnement-matière.	50
Figure 26 : Systèmes de récupération du signal diffusé.....	50
Figure 27 : La transflexion.	51
Figure 28 : Schéma de principe d'une analyse quantitative.....	55
Figure 29 : Exaltation des informations spectrales par la différentiation.....	60

Figure 30 : Sélection d'une longueur d'onde de référence.....	63
Figure 31: Effet de la MSC sur les spectres NIR	65
Figure 32 : Effet de la SNV sur les spectres NIR.....	67
Figure 33 : Dosage de deux constituants sans interférence spectrale.....	70
Figure 34 : Dosage de deux constituants avec interférences spectrales.	71
Figure 35 : La condensation de données.	76
Figure 36 : Décomposition en scores et loadings des données spectrales.....	79
Figure 37 : Schéma de principe de l'analyseur.	88
Figure 38 : Intégration de l'analyseur dans la salle de traite.	88
Figure 39 : Détail de la maquette de faisabilité.....	90
Figure 40 : Sonde "transmission pure".....	96
Figure 41 : Observation des contributions spectrales sur les spectres "simple faisceau".	97
Figure 42 : Superposition des spectres utilisés pour l'évaluation de la répétabilité.....	98
Figure 43 : Spectres "simple faisceau" pour la gamme de matières grasses.	100
Figure 44 : Spectres d'absorbance pour la gamme de matières grasses.	101
Figure 45 : Linéarité du signal optique à 1685 nm en fonction de la concentration MG.....	102
Figure 46 : Défaut de collection de la lumière diffusée en transmission pure.	103
Figure 47 : La transflexion.	104
Figure 48 : Sonde à faisceau multifibre.....	105
Figure 49 : La sonde transflexion.....	106
Figure 50 : Configuration finale de la sonde transflexion.....	106
Figure 51 : Les spectres « simple faisceau » référence et échantillon.....	107
Figure 52 : Superposition des dix spectres utilisés pour la répétabilité.	108
Figure 53 : Spectres d'absorbance pour la gamme de matières grasses.	110
Figure 54 : Linéarité du signal optique à 2305 nm en fonction de la concentration MG.....	110
Figure 55 : Dosage des solutions modèles en matières grasses.	112
Figure 56 : Premier loading utilisé pour construire le modèle multivarié.....	113
Figure 57 : Spectres de poudre de lait diluée.	114
Figure 58 : Facteurs loading extraits pour le dosage des protéines.....	115
Figure 59 : Analyse quantitative des protéines à partir des spectres bruts.....	115
Figure 60 : Spectres de laits crus.....	117
Figure 61 : Observation des corrélations et des biais pour le dosage des laits crus.....	118

Figure 62 : Influence du temps d'intégration sur la qualité spectrale.....	119
Figure 63 : Sonde à chemin optique variable.....	121
Figure 64 : Spectres de la solution modèle de protéine utilisée pour évaluer la répétabilité.	122
Figure 65 : Propriétés optiques des flux.....	122
Figure 66 : La sonde 0-45°.....	123
Figure 67 : Effet de l'inclinaison des fibres optiques sur la zone d'analyse.	124
Figure 68 : Comparaison du signal optique des sondes transflexion et 0-45°.	125
Figure 69 : Spectres d'absorption d'un échantillon sur les sondes transflexion et 0-45°.	125
Figure 70 : Spectres de poudre de lait et de caséine.....	126
Figure 71 : Utilisation des échantillons pour la calibration et la prédiction.....	127
Figure 72 : La sonde « diffusion »	134
Figure 73 : Modification de l'analyseur pour l'intégration de la sonde « diffusion ».....	134
Figure 74 : Comparaison du niveau de signal entre la sonde 0-45° et la sonde diffusion	135
Figure 75 : Influence de la détection synchrone sur la qualité du signal.	143
Figure 76 : Comparaison des loadings extraits pour les deux détections synchrones.....	145
Figure 77 : Création de spectres synthétiques	153
Figure 78 : Génération des spectres sur les instruments esclaves (Cas de l'échantillon N°50)	155
Figure 79 : Etablissement de la droite de régression.....	159
Figure 80 : Calcul de la matrice de transfert F pour la standardisation directe.....	162
Figure 81 : Calcul de la matrice de transfert F pour la PDS	164
Figure 82 : Recherche de la corrélation maximale.....	168
Figure 83 : Affinement de la longueur d'onde optimale.....	168
Figure 84 : Correction des intensités.....	170
Figure 85 : Architecture du réseau utilisé pour le transfert.....	172
Figure 86 : Modification de l'algorithme PDS par les réseaux de neurones.	174
Figure 87 : Comparaison des réponses spectrales des instruments A et B.....	178
Figure 88 : Spectres de différence entre l'instrument A et B.....	179
Figure 89 : Spectres de différence entre l'instrument A et l'instrument B corrigé.	181
Figure 90 : Comparaison des spectres bruts (A) et de spectres dérivés (B) avant et après le déplacement de l'analyseur.	184

Tableau 1 : Production laitière mondiale (établie par le F.A.O en 1979).	10
Tableau 2 : Composition moyenne du lait de vache.	11
Tableau 3 : Longueurs d'onde caractéristiques des absorptions « protéine ».	29
Tableau 4 : Longueurs d'onde caractéristiques des absorptions « matière grasse ».	30
Tableau 5: Évolution du nombre de publications dans le NIR.....	32
Tableau 6 : Influence de la résolution spectrale sur le flux sortant du monochromateur.....	42
Tableau 7 : Influence des méthodes MSC et PMSC sur l'analyse quantitative.....	66
Tableau 8 : Répétabilité de la sonde transmission pure	99
Tableau 9 : Gamme de linéarité en matières grasses.....	100
Tableau 10 : Répétabilité de la sonde transflexion.....	109
Tableau 11 : Gamme de linéarité en matières grasses.....	109
Tableau 12 : Les solutions modèles de matières grasses.....	111
Tableau 13 : Influence du pré-traitement des spectres sur les erreurs standards.....	116
Tableau 14 : Dosage des matières grasses et des protéines dans les laits crus.....	117
Tableau 15 : Influence du domaine spectral utilisé pour l'analyse quantitative.....	119
Tableau 16 : Dosage des matières grasses et des protéines dans les laits crus.....	120
Tableau 17 : Répétabilité de la sonde 0-45°.....	127
Tableau 18 : Dosage des matières grasses et des protéines dans les solutions modèles.	128
Tableaux 19A et B : Influence du domaine spectral utilisé pour l'analyse quantitative de laits reconstitués.....	130
Tableaux 20A et B : Dosage des matières grasses et des protéines sur le lait cru.	132
Tableau 21 : Répétabilité de la sonde « diffusion »	136
Tableaux 22A, B et C : Dosages des solutions modèles avec la sonde « diffusion »	137
Tableaux 23A et B : Dosage de laits crus avec la sonde « diffusion ».....	139
Tableau 24 : Analyse quantitative de laits crus.....	140
Tableau 25: Influence du pré-traitement des spectres sur le dosage des matières grasses et des protéines.	140
Tableau 26 : Influence du filtre d'ordre sur le dosage des matières grasses.....	142
Tableau 27 : Influence de la résolution sur l'analyse quantitative de laits reconstitués (sonde transflexion).	142
Tableau 28 : Influence de la détection synchrone sur le dosage des matières grasses.	144
Tableau 29 : Description des échantillons.....	154

Tableau 30 :Analyse quantitative sur l'instrument maître	156
Tableau 31: Analyse quantitative sur les instruments esclaves.....	157
Tableau 32 : Concentrations en constituant A pour les instruments maître et esclave	159
Tableau 33 : Effet de la correction pente/biais sur les erreurs standards de prédiction	160
Tableau 34 : Effet de la standardisation directe sur les erreurs standards de prédiction.....	163
Tableau 35 : Effet de la PDS sur les SEP (PCR pour calcul de F).....	165
Tableau 36 : Effet de la PDS sur les SEP (PLS pour calcul de F)	165
Tableau 37 : Effet de la PDS sur les SEP (Centrage des données et PCR pour calcul de F). 166	
Tableau 38 : Effet de la méthode de SHENK sur les SEP	171
Tableau 39 : Erreur de prédiction pour la PDS modifiée par les réseaux de neurones.	174
Tableau 40 : Synthèse des résultats du transfert sur les spectres synthétiques	175
Tableau 41 : Concentrations de référence pour les 30 échantillons d'essence.....	177
Tableau 42 : Erreurs de prédiction pour la calibration de l'instrument maître.	178
Tableau 43: Effet des variations spectrales sur un instrument non corrigé.....	179
Tableau 44: Erreurs de prédiction (RMSD) après le transfert par la méthode PDS (basée sur PCR et le centrage des données).	180
Tableau 45: Influence du centrage des données sur la PDS basée sur PCR.....	181
Tableau 46 : Influence des régressions PCR et PLS sur la PDS.	182
Tableau 47 : Amplification des erreurs due aux variations instrumentales.....	183
Tableau 48 : Transfert de calibration par la méthode PDS.	185

INDEX

- A**
- aberrations optiques 42
 - acquisition 32,35,37,55,87,90, 93,95,123,125,126,130,134,138,141,148,149,150, 162
 - albumines 9
 - anharmonicité 24,25,27
 - AOTF 36
- B**
- BABCOK 13
 - bandes fondamentales 23
 - bandes harmoniques 23,27,52
 - barattage 84,88
 - biais 111,117,127,138,158,163,175
 - blaze 38,40,147
 - brevet 90,167
 - butyromètres 13
- C**
- cahier des charges 84,85,86,90,93,184,186
 - caillé 10
 - caséines 9,14,32,125
 - cellules 14,102
 - centrage 66,166,175,181,182
 - chemin optique 61,68,94,95,102,104, 106,115,119,120,121,123,125,126,127,130,135, 138,140,141
 - cladding 43
 - CLS 70,72,73,74,76,83
 - coefficient de corrélation 56,57,94,101,109,117, 158,167
 - colostrum 9
 - combinaison 27,28,71,103,143,152,173
 - contrôles laitiers 84
 - core 43
 - crème 10,92,93,95,106,110
 - cycle de lactation 9,12
- D**
- décalage 152,168,171,185,186
 - décomposition en valeurs singulières 77,162
 - démodulation 46,89
 - dérivée 58,59,60,61,130,140,144,170
 - détecteur 37,43,46,47,48,49, 51,90,97,99,115,128,132,133,147
 - détection synchrone 46,89,90,143,144
 - délectivité 47,48,147
 - déterminant séculaire 26
 - déviat ion standard 55,56,65,66
 - diagonalisation 26
 - différence simple 59
 - diffusion 12,15,33,49,50,52, 56,57,61,62,85,95,97,100,102,103,112,113,122, 123,132,133,134,135,137,138,139,140,147
 - dosages 13,15,16,32,57,84, 111,117,128,135,138,148,151,181,184
 - DS Voir standardisation directe
- E**
- eau 27,33,52,58,63,64, 85,92,95,96,97,106,110,112,113,123
 - échantillonnage 33,49,51,55,57,88
 - EMSC 63
 - émulsion 9,10
 - énergie cinétique 20,26
 - énergie de dissociation 23
 - énergie potentielle 20,23,24,26
 - énergie vibrationnelle 22,24
 - erreur standard de calibration 56,57,58,94
 - Excel 158

F		LAGRANGE.....	26
farines.....	63	Least Median Squares	172
fenêtre mobile	170,173	lécithines	9
fente.....	34,38,40,41,42,89,90	leverage.....	176
fibres optiques	33,43,52,85,86,94, 101,103,122,123,127,132,134,148	liaisons	19,28,29,113,124
filtre d'ordre	38,42,141,142	ligne de base.....	31,58,66,72,83,97,107,113,115
filtres	33,34,35,36,37,46,85,163	linéarité	55,56,91,93,98,100, 101,102,104,108,110,125,135,143,144,173
fonction d'appareil	38,152	lipides.....	9,11,12,13
fourrage	66	lipolyse.....	84,85
G		LMS	<i>Voir</i> Least Median Squares
gap.....	60	Locally Weighted Regression	172
gaz.....	9,11,17,22	LWR.....	<i>Voir</i> Locally Weighted Regression
GERBER.....	13	M	
globulaire	9,10,13	masse réduite.....	20
globulines.....	9	matières grasses.....	6,7,9,13,14,18,28, 29,30,31,32,58,64,84,85,88,91,92,93,95,96,97,99, 100,106,107,108,109,110,111,112,113,115,116, 119,122,123,125,126,127,128,130,131,132,135, 138,140,141,142,143,144,147,184
glucides	9,16,17	matières protéiques	9
glycérol	11	MATLAB.....	157
H		matrice de passage	162
Hétérogénéité	9	matrice inverse généralisée	161
homogénéisation	85	MG.....	<i>Voir</i> Matières grasses <i>Voir</i> Matières grasses
HOOKE.....	19,23	micelles	11,14,15
I		MIR.....	<i>Voir</i> moyen infrarouge
ILS.....	72,73,74,76,79	MLR.....	72,83,158,170
incrément spectral	95,106,123,168	moindres carrés	60,61,66,69,80,81,101,158,168,169
instrumentation.....	7,31,33,37,51,84,90,187	monochromateur	34,36,38,40,41,42, 43,86,89,90,96,141,142,163
interférométrie.....	37	moyen infrarouge.....	31,45,93,98,151
interpolation	169,170	MP.....	<i>Voir</i> Matières protéiques
IRLS.....	<i>Voir</i> Iteratively Reweighted Least Squares	MSC	62,63,64,65,140
Iteratively Reweighted Least Squares	172	N	
K		NESSLER	15
KJELDAHL	15,16	NIPALS.....	77
KUBELKA.....	31,51		
L			
lactose.....	9,16,17,32,125		

NIR..... *Voir* proche infrarouge
 NIRSystem 171
 normalisation..... 53,61,62,65,67,140
 normaliser..... 61

O

oscillateur harmonique 19,21,22,23,24
 ouverture numérique 37,43,44,101,103,122,135
 overfitting..... 73,81,156,166,167,174

P

PCR..... 76,77,79,81,83,161,
 164,165,166,172,173,174,175,181,182,183,185
 pente..... 158,163
 photoconductifs 47
 photovoltaïques 47
 PLANCK..... 21
 plasma sanguin 9
 PLS..... 76,79,80,81,83,111,
 156,157,161,162,164,165,166,172,173,174,175,
 176,182,183
 PLS1..... 80,163
 PLS2..... 80,163
 PMSC..... 63,64,65
 polychromatique..... 22,36,40,42,43,86,133
 prélèvements manuels 88
 process..... 7,33,52,95,148
 proche infrarouge 31
 production 6,10,17
 protéines 6,7,9,14,15,16,18,
 28,29,31,32,55,63,64,84,85,88,91,92,93,110,112,
 113,114,115,116,117,119,120,122,125,126,127,
 128,130,131,132,135,138,140,141,142,147,184

R

rapport signal sur bruit 32,37,39,41,46,
 48,52,58,61,97,102,108,112,117,122,124,125,127,
 132,135,142,143,165,183
 recouvrement spectral 43,98,171

référence interne.....61
 réflexion diffuse.....31,51
 règle de sélection.....22,23,24,27
 régression linéaire60,62,63,68,69,70,
 93,100,101,108,158,167,172
 régression quadratique172
 répétabilité.....37,49,55,91,92,95,
 97,98,101,104,106,107,108,110,120,121,125,134,
 147
 représentativité.....57,149,150,151
 réseaux de neurones67,172,174
 résolution ...22,24,26,37,41,42,68,73,80,81,142,163
 RMN172
 RSD..... *Voir* répétabilité

S

salle de traite17,18,85,86,87,104
 saucisses fumées64
 SAVITSKY60,61,170
 SCHRODINGER22,24,27
 sérum.....9,10,91
 SHENK151,161,167,171,175
 SNV65,66,140
 SNVD.....66
 SOXHLET13
 spectroscopie de masse172,173
 stabilité.....92,134,149,151
 standardisation directe..... 161,162,163,164,165,167
 standardisation directe pas à pas161,163,164
 standards85,113,115,117,119,127,
 130,132,138,142,147,149,150,151,156,157,158,
 159,160,161,162,163,165,166,167,171,172,173,
 174,175,176,181,182,183,184,186,187
 suspension9,11
 SVD *Voir* décomposition en valeurs singulières

T

théorie quantique.....21
 transfert7,148,149,150,151,152,

156,157,158,161,162,163,164,165,166,169,170,
 171,172,173,174,175,176,180,181,182,183,184,
 185
 transformée de Fourier 37,46
 transmission 32,33,36,43,44,45,
 49,51,94,95,98,101,102,103,104,106,107,108,134,
 143,144,148
 triglycérides..... 9,11,12

V

vache9,10,11,12,17,54,84,86,88,110
 viande de bœuf.....64
 viande de porc64

W

WANG150,161,163
 WINDOWS95157

BIBLIOGRAPHIE

- (¹) C. Alais, *Science du lait*, Edition SEPAIC (1983).
- (²) M. Journet, R. Jarrigue, *Ann. Zootechnie*, 9, p133 (1960).
- (³) K.F. Kwai-hang, J.F. Hayes, J.E. Moxley, H.G. Monardes, *J. Dairy Sci.*, 68, p1257 (1984).
- (⁴) R. Jarrigue, M. Journet, *Ann. Nutr. Alim.*, 13, p233 (1959).
- (⁵) L.J.R. Barron, G. Santa-Maria, *J. Dairy Res.*, 57, p517 (1990).
- (⁶) R.M. Dowben, J.R. Brunner, D.E. Phildott, *Biochim. Biophys. Acta*, 135, p1 (1962).
- (⁷) J.M. Prentice, *Dairy sci. Abstr.*, 21, p353 (1969).
- (⁸) M. Tuinstra-Lauwaars, E. Hopkins, S. Boelsma, *J. Assoc. Off. Anal. Chem.*, 68, p1235 (1985).
- (⁹) Groupe de travail CNERNA, *Revue Laitière Française*, 301, p677 (1972).
- (¹⁰) D.H. Kleyn, J. R. Trout, M. Weber, *J. Assoc. Off. Anal. Chem.*, 71(4), p851 (1988).
- (¹¹) Norme ISO N° 2446 :1976.
- (¹²) AOAC, *Association of Official Analytical Chemists*, bulletin N° 924.05 (1990).
- (¹³) AOAC, *Association of Official Analytical Chemists*, bulletin N° 905.02 (1990).
- (¹⁴) C. Alais, B. Blanc, *World review of Nutrition and Dietetics*, 20, p66 (1974).
- (¹⁵) J. Garnier, B. Ribadeau-Dumas, *J. Dairy Res.*, 37, p493 (1970).
- (¹⁶) C.W. Slaterry, *J. Dairy Res.*, 59, p1547 (1976).
- (¹⁷) AOAC, *Association of Official Analytical Chemists*, bulletin N° 920.105 (1990).
- (¹⁸) Norme ISO N° DP8968.
- (¹⁹) Bulletin IDF, *International Dairy Federation*, N° 20A :1986.
- (²⁰) AOAC, *Association of Official Analytical Chemists*, bulletin N° 972.16 (1990).
- (²¹) Norme ISO N° DP9622.
- (²²) Bulletin IDF, *International Dairy Federation*, N° 141 :1988.
- (²³) J. Montreuil, *Bull. Soc. Chim. Biol.*, 42, p1399 (1960).
- (²⁴) M. Lipp, *Food Chemistry*, 54, p213 (1995).
- (²⁵) B. Cagnac, J.C. Pebay-Peyroula, *Physique atomique*, Edition Dunod Université, tome 1, p307 (1983).
- (²⁶) B.G. Osborne, T. Fearn, *Near Infrared spectroscopy of Food analysis*, Edition Longman Scientific and Technical, New York, p22 (1986).



- ⁽²⁷⁾ P. Williams, K. Norris, *Near Infrared Technology in the Agriculture and Food industries*, American Association of Cereal Chemists, St Paul, MN (1987).
- ⁽²⁸⁾ I.N. Levine, *Molecular Spectroscopy*, Ed. John Wiley & Sons, New York, ISBN 0-471-53128-6 (1975).
- ⁽²⁹⁾ S. Califano, *Vibrational State*, Ed. John Wiley & Sons, New York, ISBN 0-471-12996-8 (1976).
- ⁽³⁰⁾ L. Pauling, *Introduction to quantum mechanics*, Ed. Dover, ISBN 0-486-64871-0 (1963).
- ⁽³¹⁾ M. Iwamoto, *Proc. Int. Near Infr. Diffuse Refl. Spectrosc. Conf.*, p1-6 (1986).
- ⁽³²⁾ R.J. Lauer, E.J. Rosenbaum, *Applied Spectroscopy*, 6(5), p29-46 (1954).
- ⁽³³⁾ K. Whetsel, *Applied Spectrometry Reviews*, 2, p1-67 (1968).
- ⁽³⁴⁾ R.F. Goddu, *Advances in Analytical Chemistry and Instrumentation*, 1, p347-417 (1960).
- ⁽³⁵⁾ W. Herschel, *Philos. Trans. R. Soc. London*, 90(XIII), p255 (1800).
- ⁽³⁶⁾ W. Herschel, *Philos. Trans. R. Soc. London*, 90(XIV), p284 (1800).
- ⁽³⁷⁾ W. Herschel, *Philos. Trans. R. Soc. London*, 90(XV), p293 (1800).
- ⁽³⁸⁾ D.A. Burns, E.W. Ciurczak, *Handbook of Near-Infrared Analysis*, Ed Maecel Decker (1992).
- ⁽³⁹⁾ P. Kubelka, F. Munk, *Zeits. Tech. Physik.*, 12, p593 (1931).
- ⁽⁴⁰⁾ J. Ellis, J. Bath, *J. Chem. Physik.*, 6, p723 (1938).
- ⁽⁴¹⁾ P. Barchewitz, *J. Chem. Physik.*, 45, p40 (1948).
- ⁽⁴²⁾ I. Barr, W. Harp, *Phys. Rev.*, 63, p457 (1943).
- ⁽⁴³⁾ I. Ben-Gera, K. Norris, *J. Feed Sci.*, 33, p64 (1968).
- ⁽⁴⁴⁾ I. Ben-Gera, K. Norris, *Israel J. Agr. Res.*, 18, p117 (1968).
- ⁽⁴⁵⁾ L.A. Butler, *Cereal Foods World*, 28, p238 (1983).
- ⁽⁴⁶⁾ M.S. Day, R.B. Fearn, *Lab. practice*, 31, p328 (1982).
- ⁽⁴⁷⁾ J.D.S Goulden, *J. Dairy Res.*, 24, p242 (1957).
- ⁽⁴⁸⁾ P. Casado, C. Blanco, A. Pozas, *Rev. Esp. Lecheria*, 108, p97 (1978).
- ⁽⁴⁹⁾ P. Casado, C. Blanco, A. Pozas, *Proc. Int. Dairy Congress*, p376 (1978).
- ⁽⁵⁰⁾ R.J. Baer, J.F. Franck, M. Loewenstein, *J. Food Sci.*, 48, p959 (1983).
- ⁽⁵¹⁾ J. De Vilder, R. Bossuyt, *Milchwissenschaft*, 38(2), p65 (1983).
- ⁽⁵²⁾ R. Frankhuizen, E.A.M. Boers, *Zuivelzicht*, 75(9), p210 (1983).
- ⁽⁵³⁾ R. Frankhuizen, E.A.M. Boers, *Zuivelzicht*, 75, p547 (1983).

- ⁽⁵⁴⁾ D.C. Woollard, *Proc. Int. Symp. Near Infrared Reflectance Spectrosc.*, 1st RACI, Cereal Chem. Div., Melbourne, Australie (1985).
- ⁽⁵⁵⁾ F.S. Toohey, *Proc. Int. Symp. Near Infrared Reflectance Spectrosc.*, 1st RACI, Cereal Chem. Div., Melbourne, Australie (1985).
- ⁽⁵⁶⁾ G. Dooney, P. Robert, D. Bertand, P.M. Kelly, *Appl. Spectrosc.*, 44, p150 (1990).
- ⁽⁵⁷⁾ R. Giangiaco, F. Braga, C. Galliena, *Proc. 4th Int. NIRS Conf.*, Aberdeen, Scotland, (1991).
- ⁽⁵⁸⁾ K. Aboshamaa, C.J. Washam, G.S. Birth, *72nd Annu. Meet. Am. Dairy Sci. Assoc.*, Ames, IA, USA (1977).
- ⁽⁵⁹⁾ G.S. Birth, C.J. Washam, *J. Dairy Sci.*, 60, p57 (1977).
- ⁽⁶⁰⁾ G.S. Birth, C.J. Washam, *Trans. ASAE*, p 78-3060, St Joseph, MI, USA (1978).
- ⁽⁶¹⁾ C.J. Washam, G.S. Birth, R. Giangiaco, *Proc. XX Int. Dairy Congr.*, Paris (Juin 1978).
- ⁽⁶²⁾ R.L. Wehling, M.M. Pierce, *J. Assoc. Off. Anal. Chem.*, 71(3), p571 (1988).
- ⁽⁶³⁾ H.R. Egli, U. Meyhack, *Spec. Publ. R. Soc. Chem.*, 49, p103 (1984).
- ⁽⁶⁴⁾ R. Frankhuizen, N.G. Van der Veen, *Neth. Milk Dairy J.*, 39(4), p191 (1985).
- ⁽⁶⁵⁾ T. Sato, M. Iwamoto, M. Hashizume, *Jpn. J. Zootech. Sci.*, 56(11), p878 (1985).
- ⁽⁶⁶⁾ T. Sato, M. Yoshino, *Jpn. J. Zootech. Sci.*, 59(9), p806 (1988).
- ⁽⁶⁷⁾ T. Sato, S. Kawano, M. Iwamoto, *2nd Int. NIRS Conf.*, Tokyo, JAPAN (1990).
- ⁽⁶⁸⁾ *Documentation commerciale*, Katrina Inc., Gaithersburg, MA, USA.
- ⁽⁶⁹⁾ B. Muller, *NIR News*, 3, p3 (1987).
- ⁽⁷⁰⁾ W.F. McClure, *Anal. Chem.*, 66(1), p43A (1994).
- ⁽⁷¹⁾ F.G. Giesbrecht, W.F. McClure, *Applied Spectroscopy*, 35, p210 (1981).
- ⁽⁷²⁾ F.G. Giesbrecht, A. Hamid, W.F. McClure, F.G. Weeks, *Applied Spectroscopy*, 38, p322 (1984).
- ⁽⁷³⁾ W.F. McClure, AM.C. Davies, *Mikrochim. Acta*, 1, p93 (1988).
- ⁽⁷⁴⁾ F.B. Fellgett, *J. Phys. Radium*, 19, p237 (1958).
- ⁽⁷⁵⁾ P.R. Griffiths, *Chemical Infrared Fourier Transform Spectroscopy*, Ed. John Wiley & Sons (1975).
- ⁽⁷⁶⁾ G.S. Hayat, *Handbook of diffraction gratings ruled and holographic*, JOBIN-YVON.
- ⁽⁷⁷⁾ J.M. Lerner, A. Thevenon, *The optics of spectroscopy, a tutorial*, Jobin-Yvon (1988).

- ⁽⁷⁸⁾ W.G. Driscoll, W. Vaughan, *Handbook of Optics*, Ed. McGraw-Hill, ISBN 0-07-047710-8 (1978).
- ⁽⁷⁹⁾ B.R. Buchanan, D.E. Honigs, *Applied Spectroscopy*, 41(8), p1388 (1987).
- ⁽⁸⁰⁾ D.A.C Compton, S.L. Hill, N.A. Wright, J. Piche, W.A. Stevenson, *Appl. Spectrosc.*, 42, p972 (1988).
- ⁽⁸¹⁾ P. Bousquet, *Spectroscopie Instrumentale*, Ed. Dunod Université (1969).
- ⁽⁸²⁾ *Stanford Research Technical Manual*, USA (1996).
- ⁽⁸³⁾ J.P. Chatard, *Nouvelle revue d'aéronautique et d'astronomie*, 4, p33 (1995).
- ⁽⁸⁴⁾ G. Broussaud, *Optoélectronique*, Ed. Masson, p81 (1974).
- ⁽⁸⁵⁾ G. Domjn, K.J. Kaffka, *Journal of Near Infrared Spectroscopy*, 2, p67 (1995).
- ⁽⁸⁶⁾ W.G. Egan, T. Hilgeman, *Applied Optics*, 14(5), p1137 (1975).
- ⁽⁸⁷⁾ J.A. Jacquez, H.F. Kuppenheim, *J. Opt. Soc. Am.*, 45(6), p460 (1955).
- ⁽⁸⁸⁾ P.C. Williams, *Cereal Chem.*, 54, p1264 (1975).
- ⁽⁸⁹⁾ C. Depecker, B. Sombret, P. Legrand, *Analisis*, 13(8), p349 (1985).
- ⁽⁹⁰⁾ J. Workman, *Journal of Near Infrared Spectroscopy*, 29, p1-38 (1994).
- ⁽⁹¹⁾ W.R. Seitz, *Spectr. Chem. Acta Rev.*, 15, p477-492 (1993).
- ⁽⁹²⁾ M.K. Alam, S.L. Stanton, *Process Control Qual.*, 4(4), p245-252 (1993).
- ⁽⁹³⁾ P.C. William, D.C. Sobering, *Journal of Near Infrared Spectroscopy*, 1(1), p25-32 (1993).
- ⁽⁹⁴⁾ R.G.J. Miller, H.A. Willis, *Journal of Applied Chemistry*, 6, p385-391, (1956).
- ⁽⁹⁵⁾ A. Elliot, W.E. Hanby, B.R. Malcolm, *British Journal of Applied Physics*, 5, p377-381, (1954).
- ⁽⁹⁶⁾ D. Honigs, *Analytical Instrumentation*, 14(1), p1-62 (1985).
- ⁽⁹⁷⁾ N. Pedretti, D. Bertrand, *Journal of Applied Chemistry*, 1(3), p174-184 (1993).
- ⁽⁹⁸⁾ D.G. Evans, C.N. Scotter, *Journal of Near Infrared Spectroscopy*, 1(1), p33-44 (1993).
- ⁽⁹⁹⁾ D. Bertrand, *Thèse de Doctorat*, Université de Dijon (1988).
- ⁽¹⁰⁰⁾ D. Bertrand, P. Robert, D. Launay, M.F. Devaux, *Proc. Euro Food Chem. 4th*, p253 (1987).
- ⁽¹⁰¹⁾ T. Isaksson, T. Naes, *Applied Spectroscopy*, 44(7), p1152 (1990).
- ⁽¹⁰²⁾ D. Honigs, *Analytical Instrumentation*, 14(1), p1 (1985).
- ⁽¹⁰³⁾ T. Naes, *Chemom. Intel. Lab. Syst.*, 5, p155 (1987).
- ⁽¹⁰⁴⁾ R.W. Gunderson, *Chemom. Intel. Lab. Syst.*, 3, p119 (1988).

- ⁽¹⁰⁵⁾ T. Naes, T. Isaksson, *Applied Spectroscopy*, 43(2), p328 (1989).
- ⁽¹⁰⁶⁾ A.H. Aastveit, P. Marum, *Applied Spectroscopy*, 45(1), p109 (1991).
- ⁽¹⁰⁷⁾ S. Boelsma, E. Hopkin, M. Lauwaars, *International Dairy Federation*, N°248 (1990).
- ⁽¹⁰⁸⁾ N. Dupuy, M. Meurens, B. Sombret, P. Legrand, J.P. Huvenne, *Applied Spectroscopy*, 47(4), p452 (1993).
- ⁽¹⁰⁹⁾ N. Dupuy, M. Meurens, B. Sombret, P. Legrand, J.P. Huvenne, *Applied Spectroscopy*, 46(5), p860 (1992).
- ⁽¹¹⁰⁾ N. Dupuy, L. Duponchel, B. Amram, J.P. Huvenne, P. Legrand, *Journal of Chemometrics*, 8, p333 (1994).
- ⁽¹¹¹⁾ J.L. Multon, *Techniques d'analyse et de contrôle dans les industries agro-alimentaires*, Vol. 1, Ed. Association pour la promotion Industrie-Agriculture.
- ⁽¹¹²⁾ H. Mark, *Anal. Chim. Acta*, 223, p75 (1989).
- ⁽¹¹³⁾ B.G. Osborne, T. Fearn, *Near Infrared spectroscopie of Food analysis*, Edition Longman Scientific and Technical, New York, p91 (1986).
- ⁽¹¹⁴⁾ F. De Lene Mirouze, *Thèse de doctorat*, Université de LILLE (1992).
- ⁽¹¹⁵⁾ T. Isaksson, B. Kowalski, *Applied Spectroscopy*, 47(6), p702 (1993).
- ⁽¹¹⁶⁾ J.S. Shenk, M.O. Westerhaus, *Analysis of Agriculture and Food Products by NIR Spectroscopy*, Ed. Infracsoft International, p103 (1993).
- ⁽¹¹⁷⁾ R.A. Dalteur, J.R. Hurtubise, *Anal. Chem*, 56, p819 (1984).
- ⁽¹¹⁸⁾ N. Dupuy, L. Duponchel, J.P. Huvenne, B. Sombret, P. Legrand, *Food Chem.*, 57(2), p245 (1996).
- ⁽¹¹⁹⁾ E. Diaz-Carrillo, *Journal of Near Infrared Spectroscopy*, 1, p144 (1993).
- ⁽¹²⁰⁾ L. Arakaki, D.H. Burns, *Applied Spectroscopy*, 46(12), p1919 (1992).
- ⁽¹²¹⁾ J.W. Wall, D.E. Grzybowski, S.L. Monfre, *Journal of Near Infrared Spectroscopy*, 1, p55 (1993).
- ⁽¹²²⁾ W.R. Hruschka, *Near Infrared Technology in the Agriculture and Food industries*, AACC, St Paul, MN, p35 (1987).
- ⁽¹²³⁾ Y. Ozaki, T. Miura, K. Sakurai, T. Matsunaga, *Applied Spectroscopy*, 46(5), p875 (1992).
- ⁽¹²⁴⁾ P. Levillain, D. Pompeydie, *Analysis*, 14, p1 (1986).
- ⁽¹²⁵⁾ J.D.S. Goulden, *J. Dairy Res.*, 31, p273 (1964).
- ⁽¹²⁶⁾ K.H. Norris, *Proc. IUFOST Symp.*, Oslo Norway (1982).

- (127) K.H. Norris, P.C. Williams, *Cereal Chem.*, 61(2), p158 (1984).
- (128) J.L. Rodriguez-Oterio, M. Hermida, A. Cepeda, *J. Assoc. Off. Anal. Chem.*, 78(3), p802 (1995).
- (129) C. Angulo, M. Analla, I. Jiménez, *NIR publications*, ISBN 0952866609, p604 (1996).
- (130) A. Savitsky, M.J.E. Golay, *Anal. Chem.*, 36(8), p1627 (1964).
- (131) J. Steiner, Y. Termonia, J. Deltour, *Anal. Chem.*, 44, p1906 (1972).
- (132) H.M. Madden, *Anal. Chem.*, 50, p1383 (1978).
- (133) C. Angulo, M. Analla, *Near Infrared Spectroscopy: The future Waves*, Ed. NIR publications, Chichester UK (1996).
- (134) M. Marjonieri, *Applied Spectroscopy*, 46(12), p1908 (1992).
- (135) H. Martens, T. Naes, *Multivariate Calibration*, Ed. John Wiley & Sons (1989).
- (136) J.W. Hall, K. Chan, *Int. Dairy Fed.*, IDF Seminar, cork p230 (1993).
- (137) P.J. Brimmer, J.W. Hall, *Applied Spectroscopy*, 40, p550 (1986).
- (138) T. Isaksson, T. Naes, *Applied Spectroscopy*, 42(7), p1273 (1988).
- (139) A.H. Aastveit, P. Marum, *Applied Spectroscopy*, 47(4), p463 (1993).
- (140) M.S. Dhanoa, S.J. Lister, R. Sanderson, *Journal of Near Infrared Spectroscopy*, 2, p43 (1994).
- (141) H. Martens, S.A. Jensen, P. Geladi, *Proc. Nordic Symp. Appl. Stat.*, p205 (1983).
- (142) J.L. Ilari, H. Martens, T. Isaksson, *Applied Spectroscopy*, 42, p722 (1988).
- (143) P. Geladi, D. MacDougall, H. Martens, *Applied Spectroscopy*, 39, p491 (1985).
- (144) H. Martens, E. Stark, *Journal of Pharmaceutical and biomedical Analysis*, 9(8), p625 (1991).
- (145) T. Isaksson, T. Naes, *Applied Spectroscopy*, 39, p491 (1988).
- (146) M. Manley, A.E.J. McGill, B.G. Osborne, *Journal of Near Infrared Spectroscopy*, 2, p93 (1994).
- (147) R.J Barnes, M.S. Dhanoa, J.S. Lister, *Appl. Spectrosc.*, 43(5), p772 (1989).
- (148) T. Naes, T. Isaksson, *Anal. Chem.*, 62, p664 (1990).
- (149) O.E. De Noord, *Chemom. Intel. Lab. Syst.* 23, p65 (1994).
- (150) B.G. Osborne, G.M. Barrett, *Milling*, 179(6), p26 (1986).
- (151) T. Isaksson, C.E. Miller, T. Naes, *Appl. Spectrosc.*, 46, p1685 (1992).
- (152) T. Naes, T. Isaksson, *Appl. Spectrosc.*, 46, p34 (1992).

- (153) T. Isakson, K.I. Hildrum, *Proc. Third Int. Conf. On Near Infrared Spectroscopy*, p202 (1991).
- (154) G. Sinnaeve, P. Dardenne, R. Agneessens, *Journal of Near Infrared Spectroscopy*, 2, p163 (1994).
- (155) P. Chandley, *Journal of Near Infrared Spectroscopy*, 1, p133 (1993).
- (156) D.G. Evans, A. Legrand, K. Jewell, C.N.G. Scotter, *Journal of Near Infrared Spectroscopy*, 1, p209 (1993).
- (157) M.S. Dhaona, S.J. Lister, R.J. Barnes, *Appl. Spectrosc.*, 49(6), p765 (1995).
- (158) D. Bertrand, M. Lila, V. Furtoss, P. Robert, *J. Sci. Food Agric.*, 41, p299 (1987).
- (159) T. Foucart, *Analyse factorielle*, Ed. MASSON (1982).
- (160) D. Bertrand, C.N. Scotter, *Appl. Spectrosc.*, 46(9), p1420 (1992).
- (161) G. Downey, P. Robert, D. Bertrand, P.M. Kelly, *Appl. Spectrosc.*, 44(1), p150 (1990).
- (162) T. Kohonen, *Self-Organization and Autoassociative memory*, Ed. Springer-Verlag (1988).
- (163) P. Robert, D. Bertrand, M.F. Devaux, *Anal. Chem.*, 59, p2187 (1987).
- (164) G. Alfaro Denus, *Thèse de doctorat*, Université Catholique de Louvain (1990).
- (165) C.W. Brown, P.F. Lynch, R.J. Obremski, D.S. Lavery, *Anal. Chem.*, 54, p1472 (1982).
- (166) D.M. Haaland, R.G. Easterling, *Appl. Spectrosc.*, 36, p665 (1982).
- (167) D.M. Haaland, R.G. Easterling, *Appl. Spectrosc.*, 34, p539 (1980).
- (168) H. Marck, *Anal. Chem.*, 58, p2814 (1986).
- (169) D.M. Haaland, *Spectroscopy*, 2, p56 (1987).
- (170) P.J. Brown, *J. Chemometrics*, 6, p151 (1992).
- (171) P.M. Fredericks, J.B. Lee, P.R. Osborn, D.A.J. Swinkels, *Appl. Spectrosc.*, 39, p303 (1985).
- (172) J.M. Sutter, J.H. Kalivas, P.M. Lang, *J. Chemometrics*, 6, p217 (1992).
- (173) Y. Miyashita, *J. Chemometrics*, 4, p97 (1990).
- (174) S. Wold, *Technometrics*, 20(4), p397 (1978).
- (175) H.T Eastment, W.J. Krzanowski, *Technometrics*, 24(1), p73 (1982).
- (176) G. Scarponi, I. Moret, G. Capodaglio, *J. Chemom.*, 4, p217 (1990).
- (177) W. Linberg, J.A. Person, S. Wold, *Anal. Chem.*, 55(4), p643 (1983).
- (178) I.E Franck, J.H. Kalivas, B.R. Kowalski, *Anal. Chem.*, 55(11), p1800 (1983).
- (179) P. Geladi, B.R. Kowalski, *Anal. Chim. Acta*, 185, p1 (1986).

- (180) M. Stone, *J. Roy. Stat. Soc.*, B, p111 (1974).
- (181) R.D. Snee, *Technometrics*, 19, p415 (1976).
- (182) K.G. Kowalski, *Chemo. Intel. Lab. Syst.*, 9, p177 (1990).
- (183) H.J. Luinge, E. Hop, E.T.G. Lutz, *Anal. Chim. Acta*, 284, p419 (1993).
- (184) R. Grappin, R. Jeunet, *Le Lait*, 558, p498 (1976).
- (185) D.A. Biggs, *J. Assoc. Off. Anal. Chem.*, 62(6), p1202 (1979).
- (186) T. Sato, M. Yoshino, S. Furukawa, *Jpn. J. Zootech. Sci.*, 58(8), p698 (1987).
- (187) E.B. Smith, D.M. Barbano, J.M. Lynch, J.R. Fleming, *Journal of AOAC International*, 77(5), p1217 (1994).
- (188) K.B. Strawbridge, E. Ray, F.R. Hallett, S.M. Tosh, D.G. Dalgleish, *Journal of colloid and interface science*, 171, p392 (1995).
- (189) K.B. Strawbridge, E. Ray, F.R. Hallett, *Journal of Colloid and Interface Science*, 171, p392 (1995).
- (190) F.R. Van der Voort, S. Kermasha, J.P. Smith, B.L. Mills, *J. Dairy Sci.*, 70, p1515 (1987).
- (191) G. Hayat, *Guide for spectroscopy*, JOBIN-YVON (1993).
- (192) N. Dupuy, *Thèse de doctorat*, Université de LILLE, N°722 (1993).
- (193) J.L. Maubois, G. Mocquot, *Le lait*, 51, p495 (1971).
- (194) E.B. Smith, D.M. Barbano, J.M. Lynch, *J. of AOAC Int.*, 76(6), 1300 (1993).
- (195) J. Lin, C. Brown, *Appl. Spectrosc.*, 47(1), p62 (1993).
- (196) M.G. Hansen, A. Khettry, *40 th Ann. Symp.*, Toronto, 28, p107 (1995).
- (197) K.L. Miller, D. Curtin, *Instrumentation for process industries : 49 th Ann. Symp.*, College sation, TEXAS USA, p53 (1994).
- (198) H. Lammers, M.P.G. Van Uum, J.P. De kleijn, *Macromol. Chem. Phys.*, 196, p2029 (1995).
- (199) P.B. Petersen, J.K. Andersen, J.T. Johansen, E.H. Roge, *Monogr. Eur. Brew. Conv.* Vol.20, p55 (1993).
- (200) R. Mackison, S.J. Brinkworth, R.M. Belchamber, *Appl. Spectrosc.*, 46(10), p1575 (1992).
- (201) J.T. Kuenstner, K.H. Norris, *Journal of Near Infrared Spectroscopy*, 2, p59 (1994).
- (202) H. Kamishikiryo, Y. Oritani, H. Takamura, T. Matoba, *J. of Food Sci.*, 59(2), p313 (1994).
- (203) D.M. Haaland, M.R. Robinson, G.W. Koepp, *Appl. Spectrosc.*, 46(6), p1020 (1992).

- (204) T. Isaksson, C.E. Miller, T. Naes, *Appl. Spectrosc.*, 46(11), p1685 (1992).
- (205) P. Robert, D. Bertrand, C. Demarquilly, *Animal Feed Science and Technologie*, 16, p215 (1986).
- (206) K.H. Hazen, M.A. Arnold, G.W. Small, *Appl. Spectrosc.*, 48(4), p477 (1994).
- (207) N. Dupuy, C. Ruckebush, L. Duponchel, J.P. Huvenne, P. Legrand, *Analytica Chimica Acta*, 17637, p1 (1996).
- (208) J. Birth, *Appl. Spectrosc.*, 36(6), p336 (1982).
- (209) P. Latimer, *Pl. Physiol.*, 32, XVII, (1957).
- (210) W.L. Butler, K.H. Norris, *Arch. Biochem. Biophys.*, 87(1), p31 (1960).
- (211) D.D. Archibald, C.E. Miller, L.T. Lin, *Appl. Spectrosc.*, 42(8), p764 (1988).
- (212) H. Maeda, Y. Ozaki, *Journal of Near Infrared Spectroscopy*, 3, p43 (1995).
- (213) R.D. Driver, K.P. Grim, G. Dewey, M.L. Brubaker, *SPIE*, Vol 2367, p159 (1995).
- (214) M. Bacci, B. Lanza, R. Linari, G. Tosini, *Appl. Spectrosc.*, 46(3), p510 (1992).
- (215) L.G Weyer, K.J. Becker, H.B. Leach, *Appl. Spectrosc.*, 41(5), p786 (1985).
- (216) R.L. Jones, G. Parkin, C.W.C. Harvey, *Sugar Processing Research*, Helsinki, Finlande, p224 (Août1994).
- (217) F.A. DeThomas, J.W. Hall, S.L. Monfre, *Talanta*, 41(3), p425 (1994).
- (218) P.J. Milne, P.K. Swart, *Appl. Spectrosc.*, 48(10), p1282 (1994).
- (219) R.T. Carl, *Fresenius J. Anal. Chem.*, 339, p70 (1991).
- (220) J. Lin, C. Brown, *Appl. Spectrosc.*, 46(12), p1809 (1992).
- (221) G. Downey, J. Boussion, *J. Sci. Food Agric.*, 71, p41 (1996).
- (222) M. Devaux, N. Natier, P. Robert, D. Bertrand, *Appl. Spectrosc.*, 49(1), p84 (1995).
- (223) M.R. Ellekjaer, K.I. Hildrum, T. Naes, T. Isaksson, *Journal of Near Infrared Spectroscopy*, 1, p65 (1993).
- (224) C.E. Miller, *Applied Spectroscopy*, 43(8), p1435 (1989).
- (225) M.G. Hansen, A. Khettry, *Polymer Engineering and Science*, 34(23), p1758 (1994).
- (226) J. Batra, A. Khettry, M.G. Hansen, *Polymer Engineering and Science*, 34(23), p1767 (1994).
- (227) Labsphere, *Catalogue produits ORIEL*, France (1996).
- (228) *Hamamatsu Technical Information*, SD-12 (1993).
- (229) *Sat Technical Information : Infrared Products*, Groupe Sagem, p27 (1997).

- (230) *Infrared Physics and Engineering*, Ed. McGraw-Hill, Londres (1993).
- (231) *Textron Technical Information : Infrared Detectors*, p2 (1997).
- (232) Y. Wang, D.J. Veltakamp, B.R. Kowalski, *Anal. Chem.*, 63, p2750 (1991).
- (233) D.J. Veltakamp, Y. Wang, B.R. Kowalski, *Technical Report*, N°78, Center for Process Analytical Chemistry, Université de Washington (1989).
- (234) I.S. Adhihetty, J.A. McGuire, D.M. Haaland, *Anal. Chem.*, 63, p2329 (1991).
- (235) E. Bouveresse, D.L. Massart, *Vibrational Spectroscopy*, 11, p3 (1996).
- (236) Y. Wang, B.R. Kowalski, *Appl. Spectrosc.*, 46(5), p764 (1992).
- (237) Y. Wang, B.R. Kowalski, *Anal. Chem.*, 65, p1301 (1993).
- (238) R.D. Cook, S. Weisberg, *Residuals and influence in regression*, Ed. Chapman and Hall (1982).
- (239) O.E. De Noord, *Proc. Int. Chemometrics Research Meeting*, p120 (1994).
- (240) R.W. Kennard, L.A. Stone, *Technometrics*, 11, p137 (1969).
- (241) J.S. Shenk, M.O. Westerhaus, W.C. Templeton, *Crop Sci.* 31, p159 (1985).
- (242) J.S. Shenk, M.O. Westerhaus, *Crop Sci.*, 31, p1694 (1991).
- (243) J.S. Shenk, *Proc. Third Int. Conf. On Near Infrared Spectroscopy*, p949 (1991).
- (244) J.S. Shenk, M.O. Westerhaus, *NIR news*, 4(5), p13 (1993).
- (245) P. Dardenne, R. Biston, *Proc. Third Int. Conf. On Near Infrared Spectroscopy*, p655 (1991).
- (246) P. Dardenne, R. Biston, G. Sinnaeve, *Near Infrared Spectroscopy : Bringing the Gap between Data Analysis and NIR Applications*, Ed. Ellis Horwood, Chichester, p453 (1992).
- (247) E. Bouveresse, D.L. Massart, P. Dardenne, *Anal. Chim. Acta*, 297, p405 (1994).
- (248) E. Bouveresse, D.L. Massart, P. Dardenne, *Anal. Chem.*, 67, p1381 (1995).
- (249) J.E. Bertie, C.D. Keeffe, R. N. Jones, *Appl. Spectrosc.*, 45(8), p1233 (1991).
- (250) Y. Wang, B.R. Kowalski, *Anal. Chem.*, 65, p1174 (1993).
- (251) H. Mark, J. Workman, *Spectroscopy*, 11, p28 (1991).
- (252) T. Dean, T. Isaksson, *NIR News*, 4(4), p14 (1993).
- (253) T. Dean, T. Isaksson, *NIR News*, 4(2), p8 (1993).
- (254) B. G. Osborne, T. Fearn, *J. Food Technol.*, 18, p453 (1983).
- (255) J.A. Jones, I. R. Last, B. F. Mac Donald, K.A. Prebble, *J. Pharm. Biomed. Anal.* 11, p1227 (1993).

- (256) J.L.M. Lebouille, W. C. Drost, *Monograph European Brewery Convention*, 20, p14 (1993).
- (257) N. B. Buchmann, S. Runfors, *Journal of Near Infrared Spectroscopy*, 3, p35 (1995).
- (258) M.A. Sharaf, D.L. Illman, B. Kowalski, *Chemometrics*, Ed. John Willey & Sons, ISSN 0069-2883, v.82 (1986).
- (259) E. V. Thomas, D.M. Haaland, *Anal. Chem.*, 62, p1091 (1990).
- (260) R. Marbach, H.M. Heise, *Chem. Intel. Lab. Syst.*, 9, p45 (1990).
- (261) S.R. Searle, *Matrix algebra useful for statistics*, Ed. John Wiley & Sons, New York (1982).
- (262) Y. Wang, M.J. Lysaght, B.R. Kowalski, *Anal. Chem.*, 94, p562 (1992).
- (263) C.S. Chen, C.W. Brown, S.C. Lo, *Applied Spectroscopy*, 51(5), p744 (1997).
- (264) Z. Wang, T. Dean, B.R. Kowalski, *Anal. Chem.*, 67, p2379 (1995).
- (265) L. Norgaard, *Chemo. Intel. Lab. Syst.*, 29, p283 (1995).
- (266) E. Bouveresse, D.L. Massart, *Chemo. Intel. Lab. Syst.*, 32, p201 (1996).
- (267) E. Bouveresse, C. Hartmann, D.L. Massart, *Anal. Chem.*, 68, p982 (1996).
- (268) B. Bourignon, P.F. De Aguiar, K. Thorré, D.L. Massart, *J. Chromatogr. Sci.*, 32, p144 (1994).
- (269) J.S. Shenk, M.O. Westerhaus, *US Pat. N° 4866644*, Septembre (1991).
- (270) F. Mosteller, J. Turkey, *Data analysis and Regression*, Ed. Addison-Wesley Reading (1977).
- (271) G.R. Phillips, E.R. Eyring, *Anal. Chem.*, 55, p1134 (1983).
- (272) P. Vankeerberghen, C. Vandenbosh, D. Massart, *Chemom. Intel. Lab. Syst.*, 12, p3 (1991).
- (273) P.J. Rosseeuw, *J. Am. Stat. Assoc.*, 79, p871 (1984).
- (274) P.J. Rosseeuw, A.M. Leroy, *Robust Regression and Outlier Detection*, Ed. John Wiley & Sons, New York (1987) ;
- (275) D.L. Massart, L. Kaufman, P.J. Rosseeuw, A.M. Leroy, *Anal. Chim. Acta*, 187, p171 (1986).
- (276) W.S. Cleveland, S.J. Devlin, *J. Am. Stat. Assoc.*, 83, p596 (1988).
- (277) Z. Wang, T. Isaksson, B.R. Kowalski, *Anal. Chem.*, 66, p249 (1994).
- (278) J. Zupan, J. Gasteiger, *Anal. Chim. Acta*, 248, p1 (1991).
- (279) M.E. Munk, M.S. Madison, E.W. Robb, *Mikrochim. Acta*, [Wien] p505 II (1991).

- ⁽²⁸⁰⁾ E.W. Robb, M.E. Munk, *Mikrochim. Acta*, [Wien] p131 I (1990).
- ⁽²⁸¹⁾ V. Kvasnicka, *J. Math. Chem.*, 6, p63 (1991).
- ⁽²⁸²⁾ J.U. Thomsen, B. Mayer, *J. Magn. Res.*, 84, p212 (1989).
- ⁽²⁸³⁾ J. Gasteiger, X. Li, V. Simon, *J. Mol. Struct.*, 292, p141 (1993).
- ⁽²⁸⁴⁾ B. Curry, D.E. Rumelhart, *Tetrahedron Comput. Methodol.*, 3, p213 (1990).
- ⁽²⁸⁵⁾ T.J. McAvoy, H. Te Su, N.S. Wang, M.He, *Biotechnology and Bioengineering*, 40, p53 (1992).
- ⁽²⁸⁶⁾ R. Gooddacre, D.B. Kell, *Anal. Chem.*, 68, p271 (1996).
- ⁽²⁸⁷⁾ A. Espinosa, M. Sanchez, *Oil and Gas Journal*, p49, Vol. Octobre (1994).
- ⁽²⁸⁸⁾ S.M. Maggard, *Fuel Reformulation*, 2(3), p71 (1992).
- ⁽²⁸⁹⁾ N. Asker, S. Kokot, *Applied Spectroscopy*, 45(7), p1153 (1991).
- ⁽²⁹⁰⁾ ftp ://ra.nrl.navy.mil
- ⁽²⁹¹⁾ P.H. Garthwaite, *Journal of American Statistical Association*, Vol. 94, N°425, p122 (1994).

