

DEVELOPPEMENT D'UNE METHODE D'ANALYSE QUANTITATIVE

SOMMAIRE

INTRODUCTION

VII - ETUDE BIBLIOGRAPHIQUE

7.1 - Les avantages des instruments à TF par rapport aux spectromètres dispersifs

7.2 - Utilisation des techniques infrarouge dans le domaine des explosifs

7.3 - Apport des techniques chimiométriques

7.3.1 - Analyse de données et méthodes chimiométriques

7.3.2 - L'analyse quantitative et l'apport des méthodes chimiométriques au traitement des spectres infrarouges

7.3.3 - Les méthodes de prédictions quantitatives multivariées ou étalonnages (calibrations)

7.3.3.1 - Régression linéaire et Méthode des moindres carrés

7.3.3.2 - La régression linéaire multiple et l'analyse multicomposants

7.3.4 - Les méthodes de compression de données

7.3.5 - La méthode des moindres carrés partiels PLS

7.3.5.1 - Le modèle

7.3.5.2 - L'offre logicielle

7.3.6 - Exemples d'applications de la méthode des moindres carrés partiels en analyse quantitative

CONCLUSION

VIII - ETUDE EXPLORATOIRE - DEMARCHE EXPERIMENTALE

8.1 - Domaine d'étalonnage et choix des mélanges étalons

8.2 - Conception de la matrice d'étalonnage

8.3 - Matrice d'étalonnage et plan de mélanges avec contraintes

8.4 - Construction de la matrice d'étalonnage à partir des points informatifs

8.5 - Réponses expérimentales

8.6 - Expérimentation

8.6.1 - Réalisation des mélanges étalons

8.6.2 - Techniques d'échantillonnage et de saisie des spectres IR

8.6.2.1 - Appareillage et Logiciels d'exploitation des données

8.6.2.2 - Echantillonnage

8.6.2.3 - Saisie des spectres

8.7 - Exploitation des résultats

8.7.1 - Méthodes d'analyse développées en « Externe »

8.7.1.1 - Exportation et mise en forme des données spectrales

8.7.1.2 - Exploitation des données spectrales

8.7.1.3 - Analyse des modèles retenus

8.7.2 - Méthodes d'analyse développées en « Interne » avec Turbo Quant

8.7.3 - Analyse de lots industriels

8.7.3.1 - Etude préliminaire

8.7.3.2 - Analyse d'un échantillonnage de 14 lots industriels

8.7.3.3 - Discussion des résultats pour l'analyse de la forme ϵ

8.7.3.4 - Discussion des résultats pour l'analyse de la forme α , β et γ

8.7.4 - Conclusion

IX - DEVELOPPEMENT DE LA METHODE D'ANALYSE QUANTITATIVE : APPROCHE DEFINITIVE

9.1 - Travaux préliminaires

9.1.1 - Analyse thermique différentielle

9.1.2 - Sensibilité spectrale de la forme α

9.1.3 - Influence de l'impureté pentanitrite

9.1.4 - Conclusion

9.2 - Mise au point d'une nouvelle technique de préparation de la pastille de KBr

9.2.1 - Facteurs expérimentaux et réponses

9.2.2 - Matrices d'expériences et plans d'expérimentations

9.2.3 - Mise en œuvre de l'expérimentation

9.2.3.1 - Saisie des spectres

9.2.3.2 - Résultats expérimentaux

9.2.4 - Exploitation des résultats

9.2.4.1 - Granulométrie après broyage (Y_1)

9.2.4.2 - Granulométrie après mélangeage (Y_2)

9.2.4.3 - Epaisseur de la pastille et énergie du faisceau IR absorbée (Y_3 et Y_4)

9.2.5 - Variabilité spectrale

9.2.6 - Conclusion

9.3 - Matrice d'étalonnage et plan de mélanges avec contraintes

9.4 - Construction de la Matrice d'étalonnage à partir des points informatifs

9.5 - Réponses expérimentales

9.6 - Expérimentation du second plan d'expériences

9.6.1 - Réalisation des mélanges étalons

9.6.2 - Techniques d'échantillonnage et de saisie des spectres IR

9.6.2.1 - Appareillage et Logiciels d'exploitation des données

9.6.2.2 - Echantillonnage

9.6.2.3 - Saisie des spectres

9.7 - Exploitation des résultats

9.7.1 - *Méthodes d'analyse développées en « Externe »*

9.7.1.1 - Exportation et mise en forme des données spectrales

9.7.1.2 - Exploitation des données spectrales

9.7.1.3 - Analyse des modèles retenus

9.7.2 - *Méthodes d'analyse développées en « Interne » avec Turbo Quant*

CONCLUSION

X - COMPARAISON Moyen IR / Proche IR

10.1 - Applications

10.2 - Expérimentation

10.3 - Etalonnage

10.4 - Résultats et discussion

10.5 - Conclusion

XI - VALIDATION INDUSTRIELLE

11.1 - Analyse de lots industriels

11.1.1 - *Analyse d'un échantillonnage de 10 lots industriels*

11.1.2 - *Discussion des résultats pour l'analyse de la forme polymorphe ε*

11.1.3 - *Discussion des résultats pour l'analyse des formes polymorphes α , β et γ*

11.1.4 - *Comparaison partie exploratoire plan n°1 (méthode TQ₃) et partie approche définitive plan n°2 (méthode TQ₄)*

11.1.5 - *Conclusion*

11.2 - Transfert de la méthode sur le site de production

11.2.1 - *Introduction*

11.2.2 - *Démarche expérimentale*

11.2.2.1 - Domaine d'étalonnage

11.2.2.2 - Construction de la matrice d'étalonnage à partir des points informatifs

11.2.3 - *Expérimentation*

11.2.3.1 - Réalisation des mélanges étalons

11.2.3.2 - Techniques d'échantillonnage et saisie des spectres IR

11.2.4 - *Exploitation des résultats*

11.2.4.1 - Méthodes d'analyse développées avec QUANT⁺

11.2.4.2 - Méthodes d'analyse développées avec TURBO QUANT

CONCLUSION

XII - SUIVI DE LA METHODE

12.1 - Introduction

12.2 - Etude exploratoire

12.2.1 - Discussion

12.2.2 - Conclusion

12.3 - Mise au point du contrôle

12.3.1 - Suivi de la méthode de dosage (TQ₄)

12.3.2 - Etude de l'origine des fluctuations

12.3.2.1 - Elaboration et réalisation de la matrice d'étalonnage sans étalon interne

12.3.2.2 - Méthode d'analyse développée avec TQ sans étalon interne

12.3.2.3 - Recherche d'un étalon interne

12.3.2.4 - Matrice d'étalonnage avec l'étalon interne

12.3.2.5 - Méthode d'analyse développée avec l'étalon interne

CONCLUSION GENERALE

PERSPECTIVES

Références bibliographiques

ANNEXES

INTRODUCTION

Rappelons que le cahier des charges industriel implique la commercialisation d'un produit dont la teneur en forme ϵ soit supérieure à 95% avec estimation de la teneur des autres formes et des impuretés présentes. Il est donc nécessaire de mettre en œuvre une méthode de dosage possédant une précision suffisante correspondant aux impératifs techniques et commerciaux.

Le choix de la spectrométrie moyen Infra-rouge ayant été fait, il s'agit dès lors de mettre au point, d'optimiser et de valider une méthode d'analyse quantitative. *Les outils chimio-métriques* ont été mis à profit pour concevoir notre expérimentation d'une part et pour traiter l'information expérimentale obtenue avec les spectres saisis aussi bien pour l'étalonnage de la méthode de dosage que pour sa validation.

VII - ETUDE BIBLIOGRAPHIQUE

A l'origine la spectrométrie IR en mode dispersif est une technique très utilisée dans les laboratoires de synthèses pour identifier rapidement les fonctions des structures obtenues. Dans les années 80, l'apport de la transformée de Fourier (TF) couplé aux outils informatiques a donné une deuxième jeunesse à cette technique. En effet le développement des spectromètres à transformée de Fourier construits autour d'un interféromètre de type Michelson et d'un ordinateur, liés aux progrès de la technologie dans le domaine des détecteurs, de l'optique, de l'électronique et de la microinformatique, lui ont permis de devenir l'une des techniques industrielles d'analyse de routine les plus couramment utilisées.

7.1 - Les avantages des instruments à TF par rapport aux spectromètres dispersifs

Les spectromètres à transformée de Fourier sont caractérisés par :

- une grande simplicité mécanique, seul le miroir mobile de l'interféromètre étant en mouvement ;
- une détection simultanée de toutes les fréquences à travers l'interférogramme, d'où une grande rapidité de mesure (obtention d'un spectre en 1sec à comparer aux 10min nécessaires à un appareil dispersif pour la même résolution) ;
- une absence de fentes, le détecteur reçoit toute l'énergie transmise par l'échantillon. Cet avantage est particulièrement utile dans le cas d'échantillons très absorbants et pour les techniques de réflexion (ATR, réflexion diffuse) ;

- un rapport signal/bruit équivalent à celui d'un appareil dispersif obtenu en un seul balayage. Ce rapport peut encore être amélioré par coaddition (accumulation) des spectres : grande énergie et faible rapport S/B conduisent à une haute sensibilité ;
- une grande précision sur la fréquence ($0,01\text{cm}^{-1}$) grâce à un laser He-Ne qui sert à repérer avec précision le déplacement (la position) du miroir mobile à chaque balayage ;
- une absence de lumière parasite qui permet une réponse photométrique variant linéairement jusqu'à une densité optique de 3.
- une absence d'échauffement de l'échantillon qui peut être soumis beaucoup moins longtemps à la radiation IR que dans le cas d'un dispersif classique. Cet avantage est précieux dans le cas des molécules énergétiques telles le CL20.
- enfin on n'observe aucun spectre d'émission de l'échantillon qui est placé optiquement après l'interféromètre. Les radiations IR émises par l'échantillon ne sont pas modulées et par conséquent non détectées.

En résumé, par rapport aux appareils dispersifs qui effectuent un balayage des longueurs d'ondes et ce, d'autant plus lentement que la résolution désirée est plus grande, les appareils à transformée de Fourier saisissent de manière simultanée l'intégralité de l'information spectrale. Il en résulte une grande rapidité d'acquisition des spectres et un accroissement considérable de la sensibilité par addition de spectres, d'où le développement des techniques de couplage et des études cinétiques. La précision des mesures des fréquences, une bonne qualité des mesures d'absorbance et les possibilités de retraitement des spectres numérisés via le développement des techniques d'analyse de données ont donné un nouvel essor à l'analyse quantitative IR. Pour une information détaillée sur l'IRTF, parmi les nombreux ouvrages consacrés à la spectrométrie IRTF on peut se référer à celui de GRIFFITHS et HASETH [1].

7.2 - Utilisation des techniques infrarouge dans le domaine des explosifs

L'emploi de la spectroscopie IRTF pour étudier les propergols remonte à 1979 quand W.W.SCHWARZ et O.C.SMITH [2] explicitent la raison du choix de cette technique analytique pour ce type d'étude. Depuis, la spectrométrie infrarouge a été utilisée pour l'étude de matériaux énergétiques tels que les propergols, notamment pour :

- l'analyse, la détermination et l'identification des résidus solides et des produits gazeux issus de leur décomposition ;

- la mesure d'absorption durant la combustion, le suivi de l'évolution des profils gazeux en fonction des profils de concentration et de température ;
- l'étude de la cinétique de décomposition, de la mesure de la vitesse de formation, la détermination des paramètres d'Arrhénius E_a et A (pour ces types d'analyses, l'infrarouge est parfois associé à d'autres techniques : spectrométrie de masse, chromatographie gazeuse... [3-12]) ;
- pour réaliser des contrôles qualité, de conformité d'un produit (cas du contrôle de propergols composites en association avec l'HPLC [13]) ;
- pour étudier la stabilité thermique d'un explosif dans des formulations [14] ;
- pour l'identification de produits inconnus issus d'une synthèse de produits (elle est alors souvent associée à la spectrométrie de masse) [15] ;
- pour la détermination de propriétés optiques de matériaux [16-18], ou de données spectrales (elle est alors associée à la spectrométrie RAMAN) [19] ;
- pour l'étude d'interactions telles que les interactions entre le perchlorate d'ammonium et des agents liants [20] ;
- pour la réalisation du diagramme de phase de produits énergétiques ou pour des études de solubilité [21, 22] ;
- pour l'identification de formes polymorphes [23-25]. H. OSTMARK et coll. [26] ont analysé et identifié les quatre formes polymorphes du CL20 ;
- pour l'étude cinétique d'une réaction telle que la réaction entre un HTPB et un IPDI [27] ;
- pour la détermination de la proportion d'un matériau pouvant être extrait d'une matrice : détermination de la quantité de PAG pouvant être extraite du liant dans un propergol [28] ;

La spectrométrie IRTF moderne permet le couplage d'un grand nombre de techniques d'observation, de séparation et/ou d'analyses thermiques. J.C. BOULOU [29] signale les potentialités du couplage IRTF-Analyse Thermique Différentielle via les apports technologiques associés à la sensibilité et à la stabilité des spectromètres.

- l'intérêt du couplage IRTF/TG (thermogravimétrie) pour l'analyse des gaz dégagés de combustibles solides est illustré par les travaux d'I. PITKANEN et coll. [30]. Ce couplage permet l'obtention d'informations sur l'évolution de la masse de l'échantillon et permet une identification qualitative des gaz dégagés lors de sa dégradation thermique.

- pour l'étude du comportement à la combustion d'un composé, V.V.SERUSHKIN et V.I.KOLESOV [5] ainsi que N.BINKE et coll. [31] ont utilisé la calorimétrie différentielle à balayage (DSC) associée au couplage IRTF/ATG.

- le couplage CPV-IRTF a été utilisé par M.BENCHABANE et G.NADEAU [32] dans le cas de petits échantillons de propergols solides à base de nitrate d'ammonium brûlés dans une chambre fermée.

- le couplage HPLC-IRTF a été utilisé pour la détermination de traces de soluté dans un solvant. D'après G.K.FERGUSON et J.A.HASETH [33] la détermination de la composition chimique des produits de combustion de propergols et les limites de détections sont bonnes.

- le couplage Microscopie-IRTF a été utilisé par J.D.LOUDEN et coll. [34] et par B.VOGELSANGER et coll. [35] pour l'étude de profils de diffusion de produits.

- la microscopie infrarouge est un accessoire d'échantillonnage universel et souple. Elle est désignée pour l'analyse de petits échantillons, de films monomoléculaires sur substances aqueuses par exemple. D'après F.J.BERGIN et coll. [36] cette technique présente l'avantage d'être rapide et non destructive.

- la réalisation de thermolyse rapide confinée (CRT) associée à la spectrométrie IRTF permet notamment d'étudier la décomposition thermique de matériaux par la mesure des produits gazeux dégagés qui rend possible une éventuelle détermination de la vitesse de décomposition du produit en phase condensée [37-39].

- A.ISHITANI et coll. [40] présentent l'association de la spectroscopie électronique pour l'analyse chimique et de la spectrométrie IRTF pour l'analyse de l'état de surface.

- S.DESIET et F.PERREAULT [13] comparent le couplage HPLC-IRTF avec la spectroscopie RMN ^1H pour le contrôle de propergols composites à base de GAP (glycidyl azide polymère).

- l'association des différentes méthodes d'échantillonnage IRTF (microscopie, ATR, couplage Désorption-CPV-IRTF) a permis à R.A.PESCE RODRIGUEZ [41] de caractériser des propergols solides avec des résultats comparables à ceux de la thermogravimétrie.

- T.LINDBLOM [42] aborde les différentes utilisations de la spectroscopie IRTF dans l'industrie des explosifs tout en argumentant par des exemples.

7.3 - Apport des techniques chimiométriques

Néologisme apparu aux début des années 70, nous pouvons définir la Chimiométrie comme l'ensemble des méthodes mathématiques appliquées à la genèse, à la collecte et au traitement de l'information expérimentale, leur mise en œuvre étant rendue possible grâce à l'outil informatique. Leur utilisation systématique et raisonnée constitue la base de la Méthodologie de la Recherche Expérimentale pratiquée dans notre Laboratoire.

7.3.1 - Analyse de données et méthodes chimiométriques

L'ensemble des techniques d'analyse de données relevant de la Chimiométrie est présenté d'une part par l'organigramme reporté figure 54, d'autre part par le tableau 9 qui rassemble les techniques et les domaines d'applications correspondants. Nous pouvons ainsi justifier le choix de la méthode PLS pour l'étude du polymorphisme du CL20.

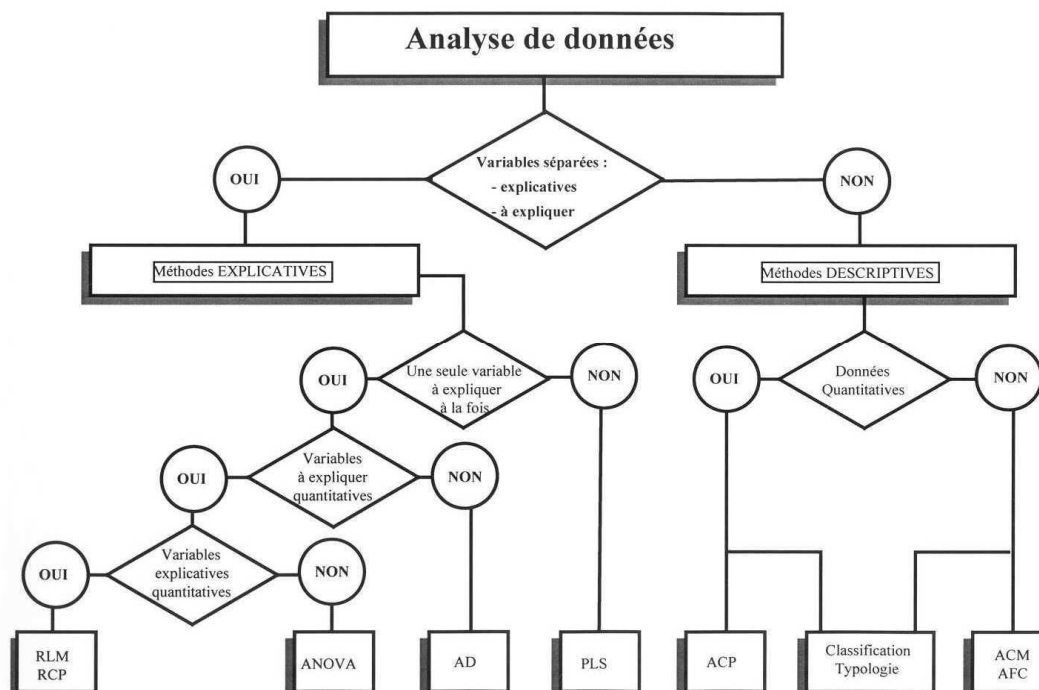


Figure 54 : techniques d'analyse de données

Tableau 9 : méthodes chimiométriques

THEMES	METHODES	DOMAINES D'APPLICATIONS
<i>Analyse de données</i>	<ul style="list-style-type: none"> - analyse statistique - analyse factorielle - analyse en composantes principales (ACP) - régression PLS – PCR - réseaux neuronaux - relations structure-propriétés 	Analyses NIR, IR-TF, UV Chromatographies HPLC Spectrométrie de Masse, Fluorescence Cinétique, Propriétés physicochimiques Analyses des eaux, Archéologie, Biotechnologie, Géochimie Huiles essentielles, Pharmacie, Pollution
<i>Calibration</i>	<ul style="list-style-type: none"> - régression RLM, PCR, PLS - réseaux neuronaux - ACP - validation croisée 	UV, IR-TF, NIR Fluorescence Médicament, Analyse chimique

Certains outils sont présentés dans ce tableau sans être commentés dans le texte, car ils n'ont pas servi pour l'étude du CL20

7.3.2 - L'analyse quantitative et l'apport des méthodes chimiométriques au traitement des spectres infrarouges [43]

Comme nous l'avons vu, le développement des techniques spectroscopiques IR quantitatives est dû aux possibilités offertes par l'IRTF associées au développement permanent des systèmes informatiques. La qualité des analyses quantitatives a beaucoup progressé en raison d'une meilleure précision des appareils (spectre acquis numériquement, appareils à rapport S/B élevé, bonne linéarité sur un large domaine). C'est cette numérisation des spectres qui autorise l'emploi de toutes les techniques d'analyse de données rendues possibles par l'offre logicielle actuelle. A noter qu'avec cette numérisation les possibilités de traitements correctifs des spectres (correction de la ligne de base, dérivée d'ordre 1 à 3 etc.) sont nombreuses.

Ces méthodes chimiométriques permettent l'analyse multicomposants et l'exploitation de la plus grande partie de l'information sans émettre d'hypothèses trop restrictives sur le plan de la théorie spectrale. En amont, elles permettent l'optimisation des essais en garantissant la qualité spectrale recueillie. En aval, dans le cas des analyses quantitatives, elles permettent la description des données sous forme synthétique (moyennes, variances...) et la quantification des résultats (méthodes prédictives) à l'aide des méthodes de régression permettant la compression des données comme la régression en composantes principales (PCR) et la méthode des moindres carrés partiels (PLS).

Nous concevons alors mieux l'importance de ces méthodes dans l'extraction de l'information quantitative contenue dans les spectres IR, dans la mesure où les spectromètres IRTF génèrent rapidement une multitude de données spectrales que les spectroscopistes ont du mal à traiter, par exemple pour notre étude, nous avons eu à traiter environ 200 spectres chacun d'entre eux étant décrit par de l'ordre de 200 valeurs d'absorbances !

7.3.3 - Les méthodes de prédictions quantitatives multivariées ou étalonnages (calibrations)

D'une manière générale, **Etalonner (Calibrer)** consiste à utiliser des données empiriques et une connaissance antérieure pour déterminer comment **prévoir une information quantitative inconnue Y à partir de mesures disponibles X, via une certaine fonction mathématique de transfert**. Dans le cas d'une méthode d'analyse quantitative, il s'agit de la phase d'étalonnage qui va conditionner la « qualité » de la méthode. Quand on veut prédire une propriété Y (ici une concentration en un analyte donné) à partir d'un ensemble de variables X_j (ici les absorbances aux longueurs d'ondes échantillonnées), les problèmes les plus fréquents que l'on rencontre sont :

- **le manque de sélectivité** : aucune des variables X_j n'est suffisante à elle seule pour expliquer Y. Ici, pour l'analyse d'un milieu complexe en spectroscopie infrarouge quantitative, aucune absorbance à une longueur d'onde déterminée n'apporte une information suffisante en soi sur l'un des constituants du mélange, on est donc condamné à prendre en compte l'ensemble des variables X_j .

- **les colinéarités (ou dépendances)** : il peut y avoir redondance d'information entre certaines des variables X_j (on dit aussi qu'il y a *intercorrélation*). Ce phénomène est inhérent à l'analyse spectrale puisque bien évidemment les absorbances échantillonnées sur un même pic sont corrélées. De ce fait on ne peut pas postuler que chaque variable X_j apporte une information unique concernant Y tel que le suppose la méthode de régression linéaire multiple traditionnelle. Cette redondance, qui est ici subie, a pour avantage de diminuer la sensibilité de Y au "bruit" et de faciliter la détection des points aberrants, qualités essentielles que doivent présenter les méthodes d'étalonnage.

- **le manque de connaissance** : notre compréhension "a priori" des mécanismes vrais cachés derrière les données peut être partielle ou même fausse . En spectroscopie, on ne peut

pas savoir à priori quels phénomènes d'interférences peuvent se produire dans un milieu où les différents composés peuvent présenter des possibilités d'interactions.

Etalonner, avons nous dit, c'est utiliser des données expérimentales et une connaissance antérieure pour déterminer comment *prévoir une information quantitative* inconnue Y à partir de *mesures disponibles* X, via une certaine *fonction mathématique de transfert*. Le modèle mathématique postulé peut être une droite si Y varie linéairement avec X, sinon ce peut être un polynôme de degré convenable. Dans le cas d'une analyse spectrale quantitative, on est en général dans les conditions d'application de la loi de Beer-Lambert pour laquelle l'absorbance A à une longueur d'onde donnée est une fonction linéaire de la concentration c en analyte dans le domaine considéré :

$$A = k.c .$$

Théoriquement, si on a à doser un seul analyte, on se trouve face à un problème d'étalonnage classique pour lequel on dispose alors de deux approches :

1. l'étalonnage direct avec Absorbance = **a** Concentration + **b**
2. l'étalonnage inverse avec Concentration = **a'** Absorbance + **b'**.

Ces deux approches ne sont pas équivalentes dans la mesure où les incertitudes sur les concentrations étalons et celles sur les absorbances mesurées expérimentalement sont de nature très différente.

Le présent travail ressort de l'étalonnage inverse.

Dans le cas de l'étalonnage inverse qui nous intéresse ici : si Y représente la teneur en analyte et X le résultat donné par la méthode d'analyse (ici l'absorbance à une longueur d'onde donnée), on postule $Y = aX + b$. *L'expérimentateur ne peut déterminer les coefficients a et b de ce modèle qu'à partir de données expérimentales qu'il acquiert pendant la phase d'étalonnage*. Ces données sont toujours en nombre limité, elles ne représentent donc qu'un échantillon de la population de toutes les teneurs en analyte que l'on pourrait étudier. Il en découle que *l'on ne peut avoir qu'une estimation des coefficients a et b du modèle postulé*.

7.3.3.1 - Régression linéaire et Méthode des moindres carrés

Cette estimation des coefficients est réalisée par *régression linéaire* au sens des *moindres carrés*. Cette méthode consiste à mesurer l'efficacité d'un ajustement par la somme des

carrés des écarts r_i (écarts appelés "résidus") entre la valeur vraie de la concentration en analyte et sa valeur estimée par la courbe d'étalonnage. Faire un ajustement c'est **minimiser** la "distance" $S = \sum [Y_i - f(X_i)]^2 = \sum r_i^2$. Dans le cas de l'analyse d'un seul composé à partir d'une seule bande d'absorption, la concentration n'est liée qu'à une seule variable, l'absorbance de la bande considérée : $[Analyte] = a \text{ Absorbance} + b$.

On parle alors d'analyse monovariée et de *régression linéaire simple*.

Dans le cas où on utilise plus d'une bande d'absorption et/ou que l'on est en présence de plusieurs composés, la concentration est liée à plusieurs variables, les absorbances des bandes considérées, on parle alors d'analyse multivariée et de *régression linéaire multiple* :

$$[Analyte] = \mathbf{a} \text{ Absorbance1} + \mathbf{b} \text{ Absorbance2} + \dots + \text{cste.}$$

De même, pour une absorbance à une longueur d'onde λ_i donnée, quand on est en présence de plusieurs composés, nous avons :

$$A_{\lambda_i} = \mathbf{a}' [Analyte1] + \mathbf{b}' [Analyte2] + \dots + \text{cste.}$$

Que la régression soit simple ou multiple, il s'agit toujours du même outil de calcul.

Expression matricielle de la Régression : dans le cas d'une seule variable X, le **modèle** s'écrit : $\eta = \beta_0 + \beta_1 X + r$ avec r = terme d'erreur représentant la partie du phénomène non expliquée par le modèle. En posant $X_0 = 1$, le modèle devient : $\eta = \beta_0 X_0 + \beta_1 X + r$. Pour chacune des valeurs X_i de la variable X, l'**estimation** de η sera $Y = b_0 + b_1 X_i$; pour n points d'étalonnage (pour n valeurs étalons de X), si on adopte la notation matricielle, on pourra écrire de manière compacte le *système de n équations linéaires* qui lient les réponses observées aux variables choisies par $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{r}$ avec :

$$\mathbf{X} = \begin{bmatrix} X_{01} & X_1 \\ X_{02} & X_2 \\ \vdots & \vdots \\ X_{0i} & X_i \\ \vdots & \vdots \\ X_{0n} & X_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_i \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_i \\ \vdots \\ r_n \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

matrice du modèle vecteur des réponses vecteur des erreurs (résidus) vecteur des coefficients du modèle

De manière générale, si on a k facteurs, on aura pour un modèle additif :

$$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_j X_j + \dots + \beta_k X_k + r$$

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_j X_j + \dots + b_k X_k$$

Pour n essais on va obtenir un système de n équations à $k+1$ inconnues : de la même manière que pour définir une droite il faut deux points au minimum, dans le cas général il faut $n \geq k+1$. Dans la notation ci-dessous on constate que la matrice du modèle X représente la matrice des coefficients des inconnues de ce système d'équations (les X_{ij} sont des nombres, ce sont les valeurs des variables explicatives, ici les absorbances, pour les essais considérés, les inconnues sont les b_i).

$$X = \begin{bmatrix} X_{01} & X_{11} & X_{21} & \vdots & X_{j1} & \vdots & X_{k1} \\ X_{02} & X_{12} & X_{22} & \vdots & X_{j2} & \vdots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{0i} & X_{1i} & X_{2i} & \vdots & X_{ji} & \vdots & X_{ki} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{0n} & X_{1n} & X_{2n} & \vdots & X_{jn} & \vdots & X_{kn} \end{bmatrix} Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{bmatrix} r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_i \\ \vdots \\ r_n \end{bmatrix} \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_k \end{bmatrix}$$

On a donc un système *linéaire par rapport aux coefficients du modèle* avec toujours pour écriture matricielle : $Y = X\beta + r$. En symbolisant par X' la matrice transposée de X , si $X'X$ n'est pas singulière (c'est à dire s'il est possible de calculer la matrice inverse correspondante), on aura l'estimation du vecteur des coefficients du modèle au sens des moindres carrés par :

$$B = (X'X)^{-1}X'Y$$

Conséquence : ce calcul matriciel qui implique l'inversion d'une matrice n'est possible (donc la résolution de ce système d'équations n'est possible) que s'il n'existe aucune corrélation entre les colonnes de la matrice X , c'est à dire *aucune corrélation entre les variables X_i* .

7.3.3.2 - La régression linéaire multiple et l'analyse multicomposants

Toute méthode d'analyse quantitative multicomposants classique repose donc sur la régression linéaire multiple or *deux obstacles s'opposent à l'utilisation de cette technique dans le cas de l'analyse spectrale* : le grand nombre de variables explicatives X_i et leurs corrélations.

Le nombre : lorsqu'on analyse un mélange de composés, *aucune absorbance à une longueur d'onde déterminée n'apporte une information suffisante* en soi sur l'un des constituants du mélange. En effet, dans le cas général, plusieurs composés si ce n'est tous, absorbent plus ou moins à une fréquence donnée. Pour obtenir une information suffisante, précise

et pertinente, *il est absolument nécessaire d'utiliser* les absorbances d'un grand nombre de longueurs d'ondes (le maximum théorique étant chaque point d'échantillonnage du spectre), *ce sont ces absorbances qui sont les variables explicatives* X_i de la concentration en analyte Y_i recherchée.

Si nous prenons l'analyse des formes polymorphes du CL20 dans ce contexte, pour la fenêtre spectrale $1200-600\text{cm}^{-1}$ qui semble être la zone informative permettant de les différencier, avec seulement une résolution de 8cm^{-1} , on obtient un échantillonnage de 75 absorbances (donc de 75 variables explicatives X_i). Il en résulte que, pour en déduire une «courbe» d'étalonnage dont l'équation comporterait 76 coefficients à estimer (75 coefficients d'absorbance + 1 constante), il faudrait disposer au minimum d'un système de 76 équations obtenues à partir de 76 mélanges étalons de composition connue !

Les corrélations : pour un analyte et une bande d'absorption donnés, les variations d'un grand nombre d'absorbances successives sont naturellement *fortement corrélées entre elles* puisque quand l'intensité de cette bande croît ou décroît, les absorbances de deux points voisins varient de la même manière (et ceci, à intervalles répétés dans le spectre !).

Compression des données : pour surmonter ces deux contraintes, nous devons utiliser une technique mathématique qui nous permette de *condenser toute l'information expérimentale* en un nombre plus réduit de *nouvelles variables explicatives non corrélées entres-elles*.

7.3.4 - Les méthodes de compression de données

L'idée générale est de postuler que l'information contenue dans un grand nombre de variables expérimentales contrôlées (ou variables réelles) peut être *concentrée* dans quelques variables sous-jacentes que l'on peut définir comme des variables *latentes* (que l'on trouve également appelées *scores* dans la littérature anglo-saxonne ou "composants" ou "facteurs de régression" et même "facteurs" tout court). Ce sont ces *variables latentes* ou *scores* qui seront utilisées dans l'équation de régression.

Par exemple le point expérimental i est défini par les k variables réelles X_j décrites par le vecteur ligne $\mathbf{x}' = (x_1, x_2, \dots, x_k)$, ces k variables réelles sont "concentrées" en A variables latentes $t : t_1, t_2, \dots, t_A$ avec une *relation linéaire entre ces nouvelles variables et les "vraies"* :

$$(t_1, t_2, \dots, t_A)' = h_1[(x_1, x_2, \dots, x_k)'].$$

Ce sont ces *nouvelles variables* qui sont *utilisées comme régresseurs dans l'équation de régression* vis à vis des réponses étudiées. Si maintenant on considère un ensemble de j réponses (y_1, y_2, \dots, y_j) , on postule la relation :

$$(y_1, y_2, \dots, y_j)' = h_2 [(t_1, t_2, \dots, t_A)'] + \mathbf{f}'$$

relation dans laquelle \mathbf{f}' représente la partie de y qui ne peut pas être expliquée par les variables t . C'est l'ensemble des deux fonctions h_1 et h_2 qui forme le prédicteur désiré de y :

$$\hat{y} = f(x) \text{ avec } f(x) = h_2(h_1(x)).$$

En résumé, la compression des données d'un grand nombre de variables x en quelques variables latentes t simplifie la calibration statistique en réduisant le nombre de paramètres du modèle à estimer pour la régression entre X et Y . La compression des données permet également de simplifier l'interprétation des résultats en mettant en évidence les principales relations entre les X : les quelques premiers facteurs t_i , par exemple les deux premiers, peuvent "ouvrir une fenêtre" à deux dimensions sur l'espace à k dimensions des X_j .

Pour revenir aux relations entre X et Y , il est habituel de les approcher par un modèle linéaire via au minimum, un centrage des variables : à partir des variables centrées $X = x_{\text{obs.}} - \bar{x}$ et $Y = y_{\text{obs.}} - \bar{y}$ dans le cas où les systèmes d'unité sont homogènes (par ex. compositions centésimales en % pour les X , unités d'absorbance pour les Y , ce qui est le cas d'une analyse spectrale quantitative), à partir des variables centrées réduites $X = (x_{\text{obs.}} - \bar{x})/\sigma_x$ et $Y = (y_{\text{obs.}} - \bar{y})/\sigma_y$ quand les variables naturelles sont caractérisées par des unités différentes.

Remarque : ne pas oublier que ces opérations de centrage ne sont pas neutres, l'opération de centrage simple a pour conséquence d'augmenter le "poids" des variables les plus dispersées alors que le passage aux variables centrées réduites donne la même échelle de variation à toutes les variables, donc à celles qui varient très peu et qui, de ce fait, peuvent n'avoir "physiquement" aucune influence sur le phénomène étudié.

Le modèle de compression linéaire des données peut alors être écrit sous la forme matricielle suivante qui est la plus générale possible :

$$\mathbf{T} = \mathbf{XV} \quad (1) \quad \text{suivi de} \quad \mathbf{Y} = \mathbf{TQ}' + \mathbf{F} \quad (2)$$

La matrice \mathbf{V} représente le système de relation linéaire qui existe entre les variables \mathbf{t} et les variables \mathbf{x} , la matrice \mathbf{Q}' représente le vecteur des coefficients du modèle linéaire reliant Y aux variables latentes (aux scores) \mathbf{t} , \mathbf{F} représentant toujours la part de variation de Y non expliquée par le modèle (bruit).

Après la détermination de \mathbf{V} et l'estimation de \mathbf{Q} , on obtient le prédicteur de Y par la relation : $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{V}\hat{\mathbf{Q}}$ (avec \mathbf{X} et \mathbf{Y} centrés). Si on avait utilisé les variables non centrées, il aurait fallu écrire : $\hat{\mathbf{Y}} = \mathbf{1}\mathbf{b}_0 + \mathbf{X}\hat{\mathbf{B}}$ (3) ou $\hat{\mathbf{B}} = \mathbf{V}\hat{\mathbf{Q}}$ (4) et $\mathbf{b}_0' = \bar{\mathbf{y}}' - \bar{\mathbf{x}}'\hat{\mathbf{B}}$ (5)

Il est important de noter que, contrairement au prédicteur de la régression multilinéaire classique (RML), *les coefficients b_{kj}* (k pour les X et j pour les Y) *ne sont pas estimés indépendamment par les méthodes de compression de données*. L'explication provient de ce que nous avons résolu les problèmes de dépendances (de corrélation) entre certaines variables X par l'extraction des variables latentes et de ce fait nous avons rendu l'estimation des b_{kj} fonction des estimations de \mathbf{V} et $\hat{\mathbf{Q}}$.

Notons également que toutes les méthodes de compression de données convergent vers la RML quand le nombre A de variables latentes prises en compte devient égal au nombre k des variables réelles. Bien entendu, dans ce dernier cas, on a perdu la notion de «compression», il en résulte que la valeur optimale de A est inférieure à k.

Il existe un certain nombre de méthodes pour extraire les variables latentes \mathbf{t} des variables \mathbf{x} , en d'autres termes pour calculer la matrice \mathbf{V} . Les performances de ces méthodes varient avec les structures des données, elles sont donc fonction des problèmes à traiter. Leur choix dépend en principe de la connaissance que l'on a des phénomènes sous-jacents mis en jeu, des relations que l'on peut supposer entre les X, du niveau de bruit etc. Nous ne mentionnerons ici que la méthode relevant de notre contexte : *la méthode des moindres carrés partiels* (PLS ou Partial Least Square regression).

Pour la méthode PLS, les A variables latentes *résumant ensemble* les informations issues de X pertinentes pour modéliser Y, il en résulte que *les deux tableaux de données X et Y sont utilisés* dans l'estimation de \mathbf{V} (lorsqu'il n'y a qu'une réponse Y, PLS et PCR sont équivalents). Rappelons que la méthode PCR, elle, n'utilise que les informations apportées par X

pour estimer V , les données Y ne servent qu'à sélectionner les composantes principales de X retenues pour le modèle final.

L'analyse du CL20 correspondant au dosage de quatre composés (les concentrations des 4 formes polymorphes représentent le tableau des Y), *nous avons utilisé la méthode PLS*.

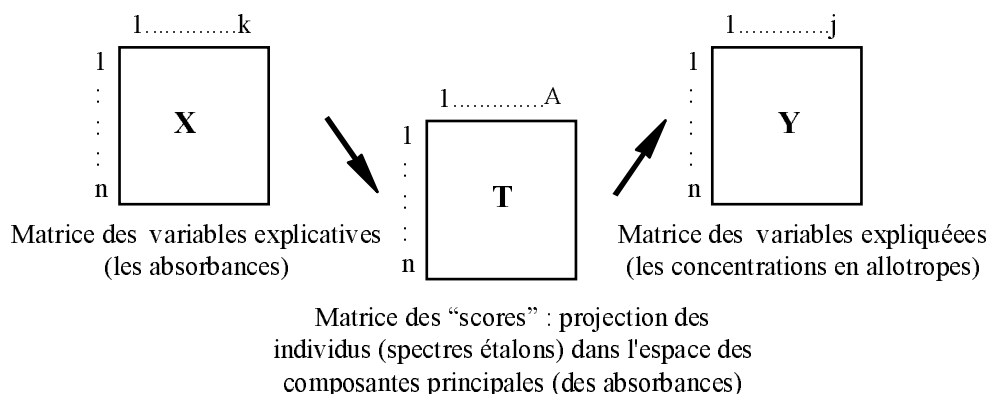
7.3.5 - La méthode des moindres carrés partiels PLS

C'est le domaine des études socio-économiques qui connaît les premiers développements de la méthode des moindres carrés partiels (PLS). L'algorithme NIPALS proposé par H.WOLD [44] est à l'origine du développement de la régression PLS. Son utilisation remonte aux années 80 avec les travaux de S.WOLD et coll. [45-46]. La méthode PLS est depuis largement utilisée, principalement dans le domaine des industries chimiques, pétrolières et agroalimentaires [47].

Un grand nombre de travaux et d'ouvrages décrivent la méthode PLS dans le domaine qui est le leur. On retiendra ceux de : S.WOLD et coll. [48], de K.ESBENSEN et coll. (« Multivariate Analysis in Practice » (1994)), de A.HOSKULDSSON (« Prediction Methods in science and Technology, Vol 1 Basic Theory » (1996)). Pour aborder l'utilisation de la méthode PLS en spectroscopie infrarouge quantitative, nous retenons particulièrement l'ouvrage de H.MARTENS et T.NAES consacré à la Calibration multivariée [49] et la mise au point de M.P.FULLER et coll. [43].

Comme nous l'avons exposé dans le cadre général de la régression linéaire, cette méthode prédictive est construite spécifiquement pour l'étude de la compréhension et de la description des relations souvent très complexe entre X et Y en l'absence d'un modèle théorique. Elle prend en compte la multicollinéarité, ce que ne fait pas la régression multiple classique.

Cette méthode reprend la philosophie de base de la régression en composantes principales mais, *dans la méthode PLS, la décomposition en composantes principales est réalisée simultanément et de manière interdépendante entre les deux tableaux X et Y* : la matrice Y sert de "guide" à la décomposition de X . Reprenons sous forme schématique les différentes matrices que l'on manipule :



La matrice des "scores" T [50] représente les coordonnées des individus dans le nouveau système de coordonnées défini par les composantes principales de X . Les relations entre les variables X et Y et leurs propres composantes, sont des relations linéaires matérialisées par des "matrices de passage" appelées "loadings" [50] : ce sont les coefficients de régression de chaque variable stockés dans des matrices P et Q ; P est obtenue à partir de X et Q à partir de Y ; notons E et F les matrices des résidus $E = X - X_{\text{préd.}}$, $F = Y - Y_{\text{préd.}}$: le nombre optimal de facteurs (de composantes prises en compte) est celui qui conduit aux résidus les plus faibles. L'ensemble du processus correspond à la figure 55 :

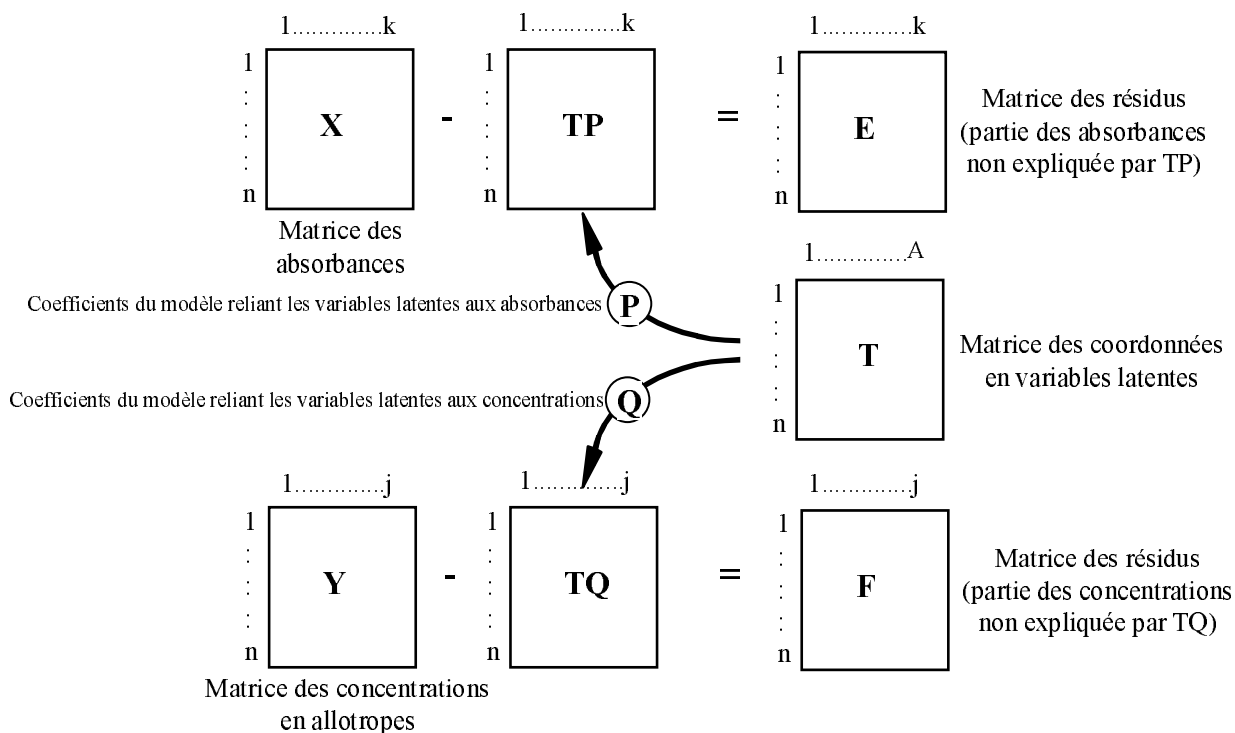


Figure 55 : processus

En résumé : *la méthode PLS diffère de la méthode PCR par l'utilisation active des variables Y pendant la décomposition linéaire de X*. En équilibrant les informations issues de X et de Y, la méthode *minimise l'impact d'importantes fluctuations dans X* dues à des effets "d'interférents", variations non pertinentes dans le modèle de calibration. C'est ainsi que pour des jeux de données peu précis, on peut obtenir des résultats acceptables avec la méthode PLS alors que la méthode PCR s'avère inutilisable (elle ne conduit à aucune corrélation significative).

Dans la méthode PLS, si on désire bien avoir une matrice **T** des scores orthogonale, la matrice **V** est difficile à interpréter et à comparer à celle de la méthode PCR. De plus, une matrice supplémentaire de loadings est nécessaire : la matrice **W**. Soit A_{\max} le nombre maximum de composantes (de facteurs PLS) que l'on a à calculer : ce nombre doit être supérieur au nombre de phénomènes que l'on suppose exister dans X de manière à se réserver quelques degrés de liberté pour modéliser des phénomènes inconnus.

Selon le nombre de réponses à étudier, il existe deux algorithmes : l'algorithme *PLS1* est utilisé quand *nous n'avons qu'une réponse à modéliser*, l'algorithme *PLS2* est mis en œuvre quand il y a *simultanément plusieurs réponses* à estimer (par exemple les concentrations de plusieurs composés présents dans l'échantillon à analyser). Il n'est pas nécessaire d'exposer, même sommairement, ces algorithmes ici. Pour une analyse détaillée se reporter à l'ouvrage de M.TENENHAUS [51].

Pour évaluer l'efficacité de la méthode PLS, de nombreux travaux ont été consacrés à la comparaison des résultats des diverses méthodes chimiométriques pour une même analyse [52-54]. En particulier D.M.HAALAND et coll. [55] montrent que la construction d'une calibration multivariée avec la méthode PLS peut être expliquée par la méthode des moindres carrés classique (CLS).

Des « variantes » de la méthode PLS sont développées pour essayer encore d'améliorer les résultats. B.K.LAVINE [56] et L.NORGAARD et coll. [57] donnent des exemples d'application des variantes iPLS et ALS.

7.3.5.1 - Le modèle

Il s'agit en général d'un simple modèle additif du premier degré dont les coefficients sont estimés par la méthode des moindres carrés partiels. Ce modèle qui doit représenter le mieux possible la réalité est de bonne qualité s'il est robuste, à une bonne validité et une bonne prédictivité (figure 56). Pour qu'un modèle soit robuste, il doit tenir compte de toutes les variabilités qu'il est possible d'avoir au cours de l'analyse (lieu, température, opérateur...). De nombreux facteurs influent sur les qualités du modèle : largeur de bande, fenêtre spectrale, résolution... [58-61].

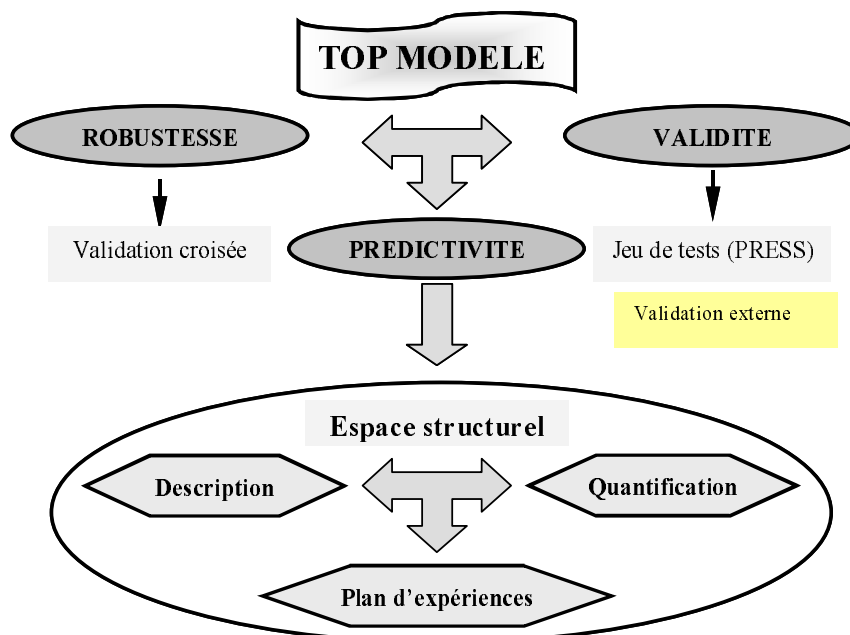


Figure 56 : le modèle

Validation croisée ou cross validation : opération qui consiste à enlever un étalon de la matrice d'étalonnage de taille m et à élaborer le modèle PLS avec les $m-1$ étalons restants. Puis, on applique le modèle à l'étalon rejeté. On répète l'opération pour les m étalons de la matrice. Chaque étalon possède alors une valeur réelle et une valeur estimée. L'écart entre ces deux valeurs est appelé erreur ou résidu et noté r_i . Le calcul de la somme des carrés des résidus permet de définir le PRESS.

PRESS : Predictive Residual Error Sum of Squares = $\sum (r_i^2)$. Une valeur de PRESS faible indique une bonne prédiction.

Validation externe : opération qui consiste à effectuer de nouveaux étalons, dont on connaît la valeur réelle et à tester le modèle élaboré en les analysant, par comparaison des valeurs prédites avec les valeurs réelles. L'étude des différents descripteurs (coefficient de corrélation, régression, variance, résidus) permet d'estimer la validité des modèles.

7.3.5.2 - L'offre logicielle

Le marché des logiciels scientifiques offre des produits de plus en plus conviviaux et puissants. Parmi les logiciels généralistes de statistiques offrant une bonne approche de la méthode PLS, on peut retenir : SIMCA, UNSCRAMBLER, LUMIERE, STATBOX. En ce

qui concerne l'infrarouge, la plupart des constructeurs d'appareils incluent un module d'analyse quantitative utilisant la méthode PLS dans le logiciel gérant le spectromètre (par exemple TURBO QUANT chez Nicolet, QUANT⁺ chez Perkin-Elmer ou le module de OPUS chez Bruker). Néanmoins, pour rendre accessible cette analyse quantitative à un opérateur non familier de l'analyse multivariée, il s'agit presque toujours d'un logiciel « boîte noire » qui masque les outils mathématiques et les méthodes chimiométriques mises en œuvre. Cela peut être dangereux car on peut très aisément obtenir des résultats aberrants non détectés par un analyste n'ayant pas un minimum de connaissances théoriques en analyse multivariée. Pour pallier ce problème, il existe des manuels d'utilisation des outils chimiométriques pour l'analyse quantitative de spectres [62, 63].

Un certain nombre de travaux [64-67] impliquant une analyse quantitative infrarouge mentionnent l'utilisation des logiciels dont nous nous sommes servis pour effectuer la quantification du CL20 (OMNIC - TURBO QUANT et SIMCA).

7.3.6 - Exemples d'applications de la méthode des moindres carrés partiels en analyse quantitative

Les exemples rapportés par M.TENENHAUS [51] illustrent les avantages de la régression PLS par rapport aux méthodes anciennes, la variété des domaines d'application et l'efficacité de cette méthode de régression dans le cas des données spectroscopiques. De nombreuses études d'analyse quantitative récemment publiées dans différents domaines (détergence, agroalimentaire, médecine...) utilisent la méthode des moindres carrés partiels [43, 50, 68-73]. Dans le cas précis de l'analyse quantitative infrarouge nous avons retenu les travaux suivants [74-79].

Il semble que la méthode PLS ait été peu utilisée dans le domaine des explosifs et plus précisément, à notre connaissance, aucune étude quantitative du polymorphisme du CL20 par spectrométrie infrarouge n'a été décrite dans la littérature.

CONCLUSION

En conclusion de cette étude bibliographique, l'objet de notre travail étant de mettre au point, d'optimiser et de valider une méthode d'analyse quantitative du CL20 polymorphe, le

choix de la spectrométrie IRTF *associée aux ressources des méthodes chimiométriques* nous paraît tout à fait judicieux avec une chance raisonnable de succès.

En faisant la synthèse des informations rassemblées dans cette étude bibliographique, on peut dresser l'organigramme général de la méthode d'analyse spectrale quantitative que nous devons développer (figure 57).

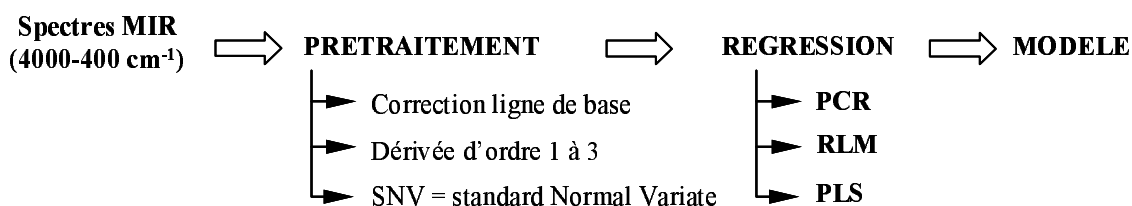


Figure 57 : analyse spectrale quantitative

VIII - ETUDE EXPLORATOIRE - DEMARCHE EXPERIMENTALE

8.1 - Domaine d'étalonnage et choix des mélanges étalons

L'étalonnage doit être construit à partir d'un jeu d'échantillons de taille suffisante, décrivant *tout le domaine d'intérêt y compris les limites, les extrapolations étant toujours dangereuses*. De plus, pour *prendre en compte des phénomènes d'interférences possibles* entre composés, donc pour espérer une précision maximale du modèle, *aucun composant ne doit être absent du jeu de calibration*, ce qui interdit le "0%". Nous avons retenu une concentration de 1% minimum pour chacun des composés.

Pour la présente étude, *après concertation* avec les personnes concernées par l'analyse du CL20, *en absence de toute information sur la composition réelle des lots industriels*, il a été décidé que les teneurs x_2 , x_3 et x_4 respectivement en chacune des formes polymorphes α , β et γ seraient comprises entre 1 et 10% ce qui implique que la teneur x_1 en forme ϵ soit comprise entre 70 et 97%. D'où le tableau des contraintes sur les composants du mélange :

$$\text{pour } \epsilon : 0,70 \leq x_1 \leq 0,97$$

$$\text{pour } \alpha : 0,01 \leq x_2 \leq 0,10$$

$$\text{pour } \beta : 0,01 \leq x_3 \leq 0,10$$

$$\text{pour } \gamma : 0,01 \leq x_4 \leq 0,10$$

Pour une première approche, la taille de la matrice de calibration peut être choisie à l'aide d'une règle empirique proposée par certains auteurs (en particulier par ceux qui ont

développé le logiciel Nicolet) : pour doser N composants, il faut disposer d'un ensemble *minimum de $3N+2$ échantillons* de calibration (dont un certain nombre serviront d'échantillons de validation). Pour notre problème nous devons donc disposer d'un minimum de 14 échantillons de compositions différentes.

Le tableau des compositions de ces échantillons s'appelle la matrice d'étalonnage.

8.2 - Conception de la matrice d'étalonnage

La matrice d'étalonnage doit répondre à la condition de «non corrélations» entre les variations de composition des différents mélanges étalons. Pour ce faire nous pouvons envisager deux approches :

Approche statistique classique via des Mélanges aléatoires : pour que les concentrations varient de façon non corrélée, elles sont déterminées de façon aléatoire pour l'ensemble du jeu d'échantillons.

Approche chimiométrique via des Mélanges structurés : on utilise un plan d'expériences spécifique aux problèmes de mélanges [80, 81] pour structurer la composition de la matrice d'étalonnage. Ces plans conduisent, par nature, à des variations des concentrations informatives aussi peu corrélées que possible. *C'est cette méthodologie que nous avons utilisée.*

8.3 - Matrice d'étalonnage et plan de mélanges avec contraintes

Les contraintes sur les concentrations des constituants du mélange des formes polymorphes du CL20 se matérialisent par un ensemble de troncatures du domaine expérimental qui conduisent à un polyèdre nommé «polyèdre des contraintes». Par extension de la stratégie de Scheffé aux domaines contraints [81], les mélanges les plus informatifs correspondent aux sommets, aux milieux des arêtes, aux centres de gravité des faces et au barycentre du polyèdre. Pour les contraintes définies ci-dessus nous obtenons 8 sommets, 12 arêtes, 6 faces et un barycentre, soit un ensemble de 27 points informatifs. C'est parmi ces points informatifs que l'on doit choisir les points d'étalonnage (ou de calibration). Pour 4 composants il est encore possible de représenter le polyèdre des contraintes, le domaine expérimental étant à trois dimensions (figure 58).

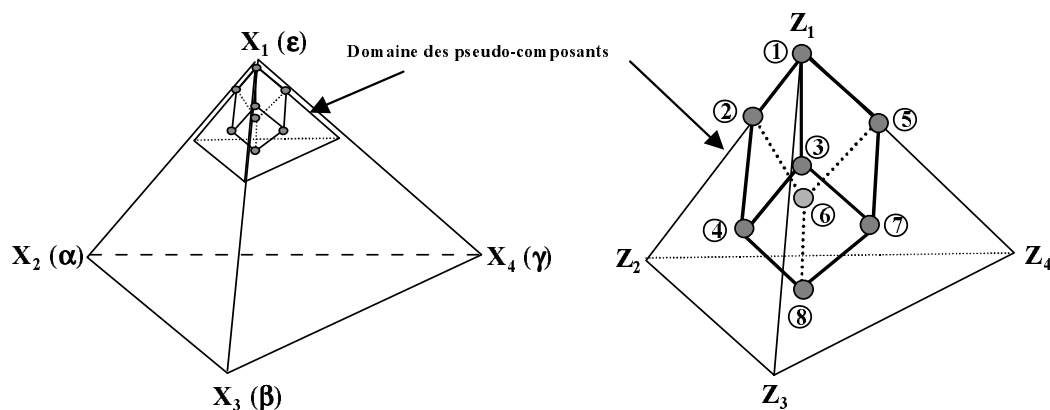


Figure 58 : domaine expérimental d'étalonnage

Comme mis à part le barycentre, tous les points informatifs sont situés à la «périphérie» du polyèdre, pour valider le modèle d'étalonnage, il est judicieux de choisir des *points tests internes au domaine*. Les points tests les plus pertinents (ici les compositions des échantillons de validation) se situent aux barycentres des polyèdres déterminés par chaque sommet et les milieux des arêtes issues de ce sommet. A partir des 8 sommets du polyèdre des contraintes, nous disposons de 8 points tests internes potentiels ce qui, avec les 27 points précédemment définis, constitue un ensemble de 35 points informatifs dont les coordonnées en variables naturelles et en pseudo-composants sont rassemblées dans le tableau 10.

8.4 - Construction de la Matrice d'étalonnage à partir des points informatifs

Pour relier une absorbance donnée aux concentrations des analytes et prendre en compte toutes les interférences spectrales possibles entre les quatre formes du CL20, nous avons postulé, a priori, un modèle cubique réduit [80, 81]. Pour 4 constituants, ce modèle comporte 14 coefficients à estimer, d'où un minimum de 14 points d'étalonnage. Comme nous disposons d'un ensemble de 35 mélanges informatifs candidats à la construction de la matrice d'étalonnage, nous devons sélectionner parmi eux les meilleurs points. *Cette sélection dépend* à la fois du modèle postulé, *du critère de décision* et du nombre N de mélanges que l'on se propose d'utiliser.

Pour le présent travail nous avons utilisé le critère de D-optimalité [82, 83] : pour une taille N donnée de la matrice d'étalonnage, la matrice du modèle X est choisie de telle manière que parmi toutes les C_{35}^N possibles, le déterminant de la matrice $X'X$ soit maximum. Le nombre optimum N de mélanges étalons (avec $N \geq 14$) a été déterminé à partir de la courbe de variation du déterminant réduit en fonction de N (figure 59). Tous calculs faits, *il en est résulté*

te une matrice d'étalonnage constituée de 21 mélanges étalons informatifs dont les coordonnées sont rassemblées dans le tableau 11.

Tableau 10 : coordonnées des points informatifs d'étalonnage

	N° Exp	Compositions naturelles				Pseudo-composants			
		Epsilon	Alpha	Beta	Gamma	Z ₁	Z ₂	Z ₃	Z ₄
Sommets	1	0.9700	0.0100	0.0100	0.0100	1.0000	0.0000	0.0000	0.0000
	2	0.8800	0.1000	0.0100	0.0100	0.6667	0.3333	0.0000	0.0000
	3	0.8800	0.0100	0.1000	0.0100	0.6667	0.0000	0.3333	0.0000
	4	0.7900	0.1000	0.1000	0.0100	0.3333	0.3333	0.3333	0.0000
	5	0.8800	0.0100	0.0100	0.1000	0.6667	0.0000	0.0000	0.3333
	6	0.7900	0.1000	0.0100	0.1000	0.3333	0.3333	0.0000	0.3333
	7	0.7900	0.0100	0.1000	0.1000	0.3333	0.0000	0.3333	0.3333
	8	0.7000	0.1000	0.1000	0.1000	0.0000	0.3333	0.3333	0.3333
Milieux des arêtes	9	0.9250	0.0550	0.0100	0.0100	0.8333	0.1667	0.0000	0.0000
	10	0.9250	0.0100	0.0550	0.0100	0.8333	0.0000	0.1667	0.0000
	11	0.9250	0.0100	0.0100	0.0550	0.8333	0.0000	0.0000	0.1667
	12	0.8350	0.1000	0.0550	0.0100	0.5000	0.3333	0.1667	0.0000
	13	0.8350	0.1000	0.0100	0.0550	0.5000	0.3333	0.0000	0.1667
	14	0.8350	0.0550	0.1000	0.0100	0.5000	0.1667	0.3333	0.0000
	15	0.8350	0.0100	0.1000	0.0550	0.5000	0.0000	0.3333	0.1667
	16	0.7450	0.1000	0.1000	0.0550	0.1667	0.3333	0.3333	0.1667
	17	0.8350	0.0550	0.0100	0.1000	0.5000	0.1667	0.0000	0.3333
	18	0.8350	0.0100	0.0550	0.1000	0.5000	0.0000	0.1667	0.3333
	19	0.7450	0.1000	0.0550	0.1000	0.1667	0.3333	0.1667	0.3333
	20	0.7450	0.0550	0.1000	0.1000	0.1667	0.1667	0.3333	0.3333
Centres de gravité des faces	21	0.8800	0.0100	0.0550	0.0550	0.6667	0.0000	0.1667	0.1667
	22	0.8800	0.0550	0.0100	0.0550	0.6667	0.1667	0.0000	0.1667
	23	0.8800	0.0550	0.0550	0.0100	0.6667	0.1667	0.1667	0.0000
	24	0.7900	0.1000	0.0550	0.0550	0.3333	0.3333	0.1667	0.1667
	25	0.7900	0.0550	0.1000	0.0550	0.3333	0.1667	0.3333	0.1667
	26	0.7900	0.0550	0.0550	0.1000	0.3333	0.1667	0.1667	0.3333
Barycentre	27	0.8350	0.0550	0.0550	0.0550	0.5000	0.1667	0.1667	0.1667
Points tests internes	28	0.9025	0.0325	0.0325	0.0325	0.7500	0.0833	0.0833	0.0833
	29	0.8575	0.0775	0.0325	0.0325	0.5833	0.2500	0.0833	0.0833
	30	0.8575	0.0325	0.0775	0.0325	0.5833	0.0833	0.2500	0.0833
	31	0.8125	0.0775	0.0775	0.0325	0.4167	0.2500	0.2500	0.0833
	32	0.8575	0.0325	0.0325	0.0775	0.5833	0.0833	0.0833	0.2500
	33	0.8125	0.0775	0.0325	0.0775	0.4167	0.2500	0.0833	0.2500
	34	0.8125	0.0325	0.0775	0.0775	0.4167	0.0833	0.2500	0.2500
	35	0.7675	0.0775	0.0775	0.0775	0.2500	0.2500	0.2500	0.2500

Géométriquement, les 8 sommets sont identifiés sur la figure 1.

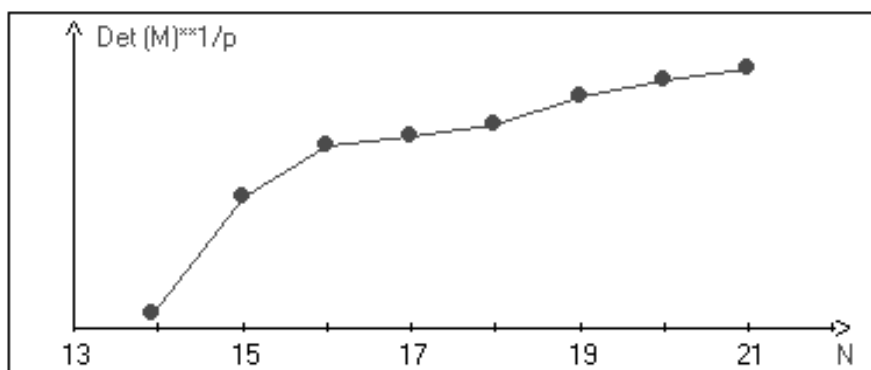


Figure 59 : variation du déterminant réduit de X'X en fonction du nombre de points

Tableau 11 : matrice d'étalonnage (théorique)

	N° Exp	Compositions naturelles				Pseudo-composants			
		Epsilon	Alpha	Beta	Gamma	Z ₁	Z ₂	Z ₃	Z ₄
Sommets	1	0.9700	0.0100	0.0100	0.0100	1.0000	0.0000	0.0000	0.0000
	2	0.8800	0.1000	0.0100	0.0100	0.6667	0.3333	0.0000	0.0000
	3	0.8800	0.0100	0.1000	0.0100	0.6667	0.0000	0.3333	0.0000
	4	0.7900	0.1000	0.1000	0.0100	0.3333	0.3333	0.3333	0.0000
	5	0.8800	0.0100	0.0100	0.1000	0.6667	0.0000	0.0000	0.3333
	6	0.7900	0.1000	0.0100	0.1000	0.3333	0.3333	0.0000	0.3333
	7	0.7900	0.0100	0.1000	0.1000	0.3333	0.0000	0.3333	0.3333
	8	0.7000	0.1000	0.1000	0.1000	0.0000	0.3333	0.3333	0.3333
Milieux des arêtes	9	0.9250	0.0550	0.0100	0.0100	0.8333	0.1667	0.0000	0.0000
	10	0.9250	0.0100	0.0550	0.0100	0.8333	0.0000	0.1667	0.0000
	11	0.9250	0.0100	0.0100	0.0550	0.8333	0.0000	0.0000	0.1667
	12	0.8350	0.1000	0.0550	0.0100	0.5000	0.3333	0.1667	0.0000
	13	0.8350	0.1000	0.0100	0.0550	0.5000	0.3333	0.0000	0.1667
	14	0.8350	0.0550	0.1000	0.0100	0.5000	0.1667	0.3333	0.0000
	15	0.8350	0.0100	0.1000	0.0550	0.5000	0.0000	0.3333	0.1667
	16	0.7450	0.1000	0.1000	0.0550	0.1667	0.3333	0.3333	0.1667
	17	0.8350	0.0550	0.0100	0.1000	0.5000	0.1667	0.0000	0.3333
	18	0.8350	0.0100	0.0550	0.1000	0.5000	0.0000	0.1667	0.3333
	19	0.7450	0.1000	0.0550	0.1000	0.1667	0.3333	0.1667	0.3333
	20	0.7450	0.0550	0.1000	0.1000	0.1667	0.1667	0.3333	0.3333
Centres de gravité des faces	23	0.8800	0.0550	0.0550	0.0100	0.6667	0.1667	0.1667	0.0000

La génération des points informatifs et la construction de la matrice D-optimale ont été réalisées à l'aide du Logiciel NEMROD

Points Tests : parmi les 8 points tests internes potentiels, nous en avons retenu 4 (les points 31 à 34) qui équilibrent les niveaux bas et haut des formes α et β (tableau 12).

Tableau 12 : points tests

	N° Exp	Compositions naturelles				Pseudo-composants			
		Epsilon	Alpha	Beta	Gamma	Z ₁	Z ₂	Z ₃	Z ₄
Points tests internes	31	0.8125	0.0775	0.0775	0.0325	0.4167	0.2500	0.2500	0.0833
	32	0.8575	0.0325	0.0325	0.0775	0.5833	0.0833	0.0833	0.2500
	33	0.8125	0.0775	0.0325	0.0775	0.4167	0.2500	0.0833	0.2500
	34	0.8125	0.0325	0.0775	0.0775	0.4167	0.0833	0.2500	0.2500

8.5 - Réponses expérimentales

Globalement, la réponse expérimentale est le spectre infrarouge de la substance examinée, spectre saisi selon un protocole expérimental défini, fixé par la méthode d'analyse. En réalité, il s'agit d'un problème multiréponses : il y a autant de réponses individuelles (absor-

bances) qu'il y a de fréquences spectrales étudiées. De plus, pour l'exploitation numérique, on peut envisager ou non un traitement mathématique préliminaire du spectre brut enregistré par le spectromètre : par exemple correction de la dérive de ligne de base et/ou utilisation de la dérivée première du spectre (ou d'ordre supérieur).

8.6 - Expérimentation

8.6.1 - Réalisation des mélanges étalons

Nous avons réalisé 25 mélanges informatifs : les 21 points de la matrice d'étalonnage et les 4 points tests de validation. Compte tenu que l'on ne dispose pas de méthode de dosage alternative validée pour constituer une matrice d'étalonnage à partir d'échantillons industriels, l'étalonnage ne peut se faire qu'à partir de mélanges « reconstitués » à partir des formes polymorphes pures. Ces formes polymorphes ont été synthétisées au laboratoire de synthèse du CRB à partir du lot de CL20 réf. 5 titrant 99,1% (avec en particulier 0,7% d'impureté pentani-trée). Les méthodes de synthèse des formes polymorphes appartiennent au savoir-faire du groupe SNPE.

Les références des échantillons des formes polymorphes du CL20 utilisées pour effectuer les mélanges étalons sont rassemblées dans le tableau 13. Leur pureté chimique a été déterminée par HPLC (méthode de normalisation interne corrigée, validée par des travaux antérieurs) ; les teneurs observées étant très proche de 100%, les différentes formes polymorphes sont considérées comme «pures». Quant à la pureté monomorphique, elle a été estimée uniquement par IR (contrôle de conformité du spectre IR de chaque forme synthétisée par rapport aux spectres de référence issus de travaux antérieurs).

Tableau 13 : références des substances étalons

CL20	Référence	Pureté chimique
α	A2 amorphe	99,2 (0,7 % penta)
β	B2 amorphe	99,2 (0,7 % penta)
γ	G2 amorphe	99,2 (0,7 % penta)
ϵ	E2 amorphe	99,1 (0,7 % penta)

8.6.2 - Techniques d'échantillonnage et de saisie des spectres IR

8.6.2.1 - Appareillage et Logiciels d'exploitation des données

Pesées : balance de précision METTLER TOLEDO modèle AG245.

Matériel pour la réalisation des pastilles : utilisation de mortiers et pilons d'agate de marque MERCK, pastilleuse 13mm de marque SPECTRA TECH ; presse de marque BECKMAN (10 tonnes sur un moule de 13mm engendrent une pression de 750 453MPa, donnée constructeur).

Spectrométrie : spectromètre IRTF Nicolet 710 (figure 60) équipé d'un détecteur DTGS.

Logiciels Nicolet : OMNIC 3.1 pour l'acquisition et le traitement primaire des spectres et TURBO QUANT 5.0 pour l'analyse spectrale quantitative (méthode PLS).



Figure 60 : spectromètre Nicolet 710

Traitements « externes » des données : logiciels EXCEL et SIMCA-S (version 6). Ce dernier est un logiciel statistique généraliste permettant, outre l'utilisation des méthodes classiques, d'exploiter des données expérimentales par la méthode de régression des moindres carrés partiels (PLS, algorithme PLS1 ou 2) [51].

8.6.2.2 - Echantillonnage

Les mélanges synthétiques sont obtenus par pesée des formes polymorphes pures conformément aux compositions données par la matrice d'étalonnage et par dispersion dans du KBr cristallisé anhydre (conservé à l'étuve, de granulométrie et de qualité adaptées à la spectrométrie IR). Chaque mélange est réalisé selon la méthode d'échantillonnage dont la genèse ressort de travaux antérieurs :

- prendre 100mg de mélange des formes polymorphes du CL20, ajouter 900mg de KBr, puis broyer et mélanger manuellement pendant 20min. Pour pallier l'hygroscopie du KBr, mettre sous vide au dessiccateur (sous P₂O₅) pendant 24 heures.
- prendre 80mg de ce mélange primaire (CL20+KBr), ajouter 920mg de KBr, puis répéter le processus précédent (20min de broyage-mélangeage, 24h de dessiccation).
- les pastilles proprement dites sont réalisées à partir de 300mg de cette dispersion finale dans le KBr (sous vide, à l'aide d'une pastilleuse de 13mm de diamètre et d'une presse hydraulique).
- *pour chaque mélange étalon, on fabrique 3 pastilles* selon ce protocole.

Rappels : l'IRTF étant une technique spectrale mono-faisceau, chaque spectre est la résultante de l'absorption due à l'échantillon et de celle de «l'environnement» (système optique, atmosphère du banc optique et du compartiment échantillon, diffusion de la pastille et absorption du dispositif d'échantillonnage). Il en résulte que *le spectre de la substance étudiée est obtenu par soustraction* du «spectre du bruit de fond» appelé **Background**.

Le Background est réalisé en présence d'une pastille de 300mg de KBr pur (broyé pendant 5min pour atteindre un niveau granulométrique voisin de celui de l'échantillon et pastillé dans les mêmes conditions que les étalons).

8.6.2.3 - Saisie des spectres

Que ce soit pour le Background ou pour les échantillons, les spectres ont été saisis en transmission de 4000 à 400cm⁻¹ à l'aide d'un porte pastille de 13mm, avec les paramètres de réglage suivants du spectromètre : gain = 2 ; résolution = 4cm⁻¹ ; nombre de scans = 100.

Protocole de mesure : la surface de la pastille étant environ 4 fois plus grande que celle du faisceau infrarouge, *l'absorption mesurée n'est due qu'à une partie du CL20 présent* (la mesure de l'intégralité de la prise d'essai n'est pas effectuée). Pour pallier à la fois le manque d'homogénéité possible d'une pastille donnée et la dispersion probable des homogénéités entre pastilles différentes, la platine du porte pastille est légèrement décentrée et *chaque pastille est mesurée deux fois avec une rotation de 90° de la pastille* entre les deux mesures.

Cette procédure étant répétée pour chacune des trois pastilles d'un étalon donné, *nous disposons d'un ensemble de 6 spectres par étalon* qui peuvent être considérés comme une *répétition des points d'étalonnage*.

8.7- Exploitation des résultats

Comme nous l'avons exposé dans la partie théorique, seule la méthode de régression PLS est utilisable pour une analyse spectrale quantitative. Cette méthode, mathématiquement compliquée, est délicate à mettre en œuvre et à maîtriser par un utilisateur n'ayant pas les connaissances théoriques nécessaires. Pour pallier cette difficulté, les auteurs des logiciels intégrés aux spectromètres (tels que par exemple OMNIC et TURBO QUANT chez Nicolet) ont travaillé particulièrement leur « ergonomie » pour en faire des outils «boîte noire», utilisables par tout technicien supérieur chargé de l'analyse IR avec le minimum de formation. Il en résulte que *l'on n'a plus aujourd'hui accès à tous les paramètres* considérés comme intermédiaires ou trop « techniques » du point de vue mathématique. Le compromis est très acceptable pour une grande majorité de travaux pour lesquels les différences spectrales entre les composés à doser sont suffisantes, c'est à dire quand il s'agit d'une analyse de composés chimiquement différents.

Ce n'est plus du tout le cas pour le présent problème où nous avons affaire à *une seule structure chimique*, les différences structurelles entre les formes polymorphes ne se traduisant que par quelques différences vibrationnelles dans la zone spectrale 1200-600cm⁻¹. Or, pour utiliser une technique qui vraisemblablement est aux limites de ses possibilités et *pour déterminer les limites exactes de la méthode d'analyse* qui en découle, *il nous a paru insuffisant et dangereux* de se contenter des possibilités offertes par le logiciel «boîte noire » d'un fabricant d'appareil, bien que, in fine, *ce sera ce logiciel qui devra être utilisé en routine*. D'où la méthode de travail que nous avons choisie :

L'information spectrale sera «exportée» pour pouvoir être exploitée de manière «explicite» avec tous les outils nécessaires en particulier pour identifier les zones spectrales les plus informatives et les modèles les plus performants. Les paramètres optimaux seront ensuite «importés» dans le logiciel TURBO QUANT pour optimiser la méthode d'analyse qui sera utilisée en routine, directement à partir du spectromètre servant à saisir les spectres.

8.7.1 - Méthodes d'analyse développées en «Externe»

8.7.1.1 - Exportation et mise en forme des données spectrales

Par nature un «point» d'un spectre est défini par ses deux coordonnées : fréquence en abscisse et absorbance correspondante en ordonnée. Il y aura autant de «couples» fréquence-absorbance qu'il y a de points échantillonnés. Une saisie tous les 2cm^{-1} entre 4000 et 400cm^{-1} implique l'enregistrement de 1869 points !

Tout spectromètre IRTF crée un fichier de données dont le format est spécifique au fabricant avec une structure adaptée au logiciel d'exploitation. Pour rendre les données «transportables», les fabricants ont défini un format normalisé type texte (.CSV) sous lequel on peut sauvegarder les spectres à partir du logiciel de l'appareil (ici OMNIC). Sous ce format, le spectre défini ci-dessus se présentera sous la forme d'un tableau numérique comportant 2 colonnes et 1869 lignes. Si certains logiciels généralistes de statistique ayant prévu le traitement des données spectrales acceptent directement l'importation des données au format CSV, pour les autres *il est nécessaire de présenter les données sous un format acceptable* que ce soit pour la nature des variables affectées aux lignes et aux colonnes ou pour le format au sens informatique du terme. En revanche, tous les logiciels de traitement de données acceptent l'importation des données à partir d'un fichier EXCEL.

L'utilisation du logiciel SIMCA a nécessité une conversion, via EXCEL, des données brutes issues du spectromètre Nicolet 710. De plus, pour pouvoir être traité par ce logiciel, le tableau de données final doit présenter *les individus en lignes et les variables explicatives en colonnes* : par exemple, si pour l'établissement d'un modèle d'étalonnage nous devons analyser l'ensemble des 21 mélanges étalons, cela correspond au traitement de $21 \times 6 = 126$ spectres, dont nous devons présenter les données sous la forme d'un tableau d'absorbances constitué de 126 lignes (1 par spectre) et 1869 colonnes (1 par fréquence échantillonnée du spectre). Enfin, pour avoir la matrice des réponses (les concentrations en formes polymorphes), nous devons ajouter encore 4 colonnes pour les teneurs en chacune des formes polymorphes.

Nous devons satisfaire enfin à une dernière contrainte, celle du tableur *EXCEL* qui **limite à 256 le nombre de colonnes** d'un tableau (la limitation en nombre de lignes de plusieurs milliers est sans objet pour notre travail). Pour ce faire nous devons supprimer des fréquences pour passer de 1869 à moins de 256.

Mise en forme préliminaire

Pour l'ensemble du présent travail nous avons saisi environ 200 spectres que nous devons exploiter, d'où environ 200 fichiers à convertir. La première tâche consiste à partir des fichiers spectraux bruts et à les convertir en fichier texte sous OMNIC. Ces derniers sont ensuite repris sous EXCEL pour être convertis en fichiers unitaires numériques à partir desquels toutes les opérations de transformations sont possibles : ces fichiers «primaires» unitaires EXCEL comportent 2 colonnes de 1869 lignes (avec 1 colonne des fréquences et 1 des absorbances).

Dans un deuxième temps, ces fichiers unitaires sont rassemblés en un seul fichier secondaire pour former un tableau où chaque ligne correspond à une fréquence et chaque colonne aux absorbances d'un spectre d'un mélange donné (pour 200 spectres on obtient un tableau de 1869 lignes et 200 colonnes).

Enfin, pour obtenir le tableau de données complet (tableau 14) rassemblant les variables explicatives (les absorbances) et les variables expliquées (les concentrations en formes polymorphes), nous rajoutons 4 lignes correspondant aux teneurs en chacune des formes polymorphes de chaque mélange étudié (matrice des compositions).

Tableau 14 : structure du tableau de données complet

Lignes	Fréquences (cm ⁻¹)	Absorbances					
		N° mél.					
1	399
.....	Matrice des absorbances			
1869	4000
1870	Forme α
1871	Forme β	Matrice des concentrations			
1872	Forme γ
1873	Forme ε

Réduction du nombre de fréquences informatives

Pour pouvoir être traité à l'aide du logiciel SIMCA, ce tableau de données doit être «transposé» : si, toujours pour 200 spectres, il n'y a aucun problème pour écrire les 201 colonnes sous la forme de 201 lignes, il n'en est pas de même pour les 1873 lignes qui deviendraient les 1873 colonnes du tableau transposé, car EXCEL n'accepte que 256 colonnes. Il faut réduire significativement le nombre de fréquences informatives. Pour ce faire nous allons procéder en deux étapes :

- éliminer les zones spectrales non informatives (c'est à dire ne présentant aucune bande d'absorption ou seulement des bandes non significatives) ;
- réduire la résolution du spectre : passer de 2 à 8cm^{-1} par exemple permet de diviser par 4 le nombre de fréquences informatives (une résolution de 8cm^{-1} est parfaitement raisonnable pour une analyse quantitative).

Pour la première étape, une analyse (visuelle) des spectres montre que nous pouvons supprimer deux zones spectrales qui ne présentent aucune absorption significative : la zone $4000 \rightarrow 3200\text{cm}^{-1}$ et la zone $2800 \rightarrow 1800\text{cm}^{-1}$. Cette suppression de 1800cm^{-1} s'accompagne de la suppression d'environ 934 lignes du tableau précédent. Cette suppression n'implique aucune manipulation particulière si ce n'est de sélectionner les groupes de lignes incriminés.

Il reste environ 939 lignes (dont 935 fréquences et 4 teneurs en chacun des formes polymorphes de chaque mélange), dont on ne peut diminuer le nombre qu'en diminuant la résolution. Diminuer la résolution de $2 \rightarrow 4\text{cm}^{-1}$ consiste à supprimer une ligne sur deux. Comme il est exclu de le faire manuellement sur un tableau de 935 lignes, nous avons imaginé une stratégie de tri adaptée à la structure de nos fichiers que nous avons programmée sous la forme d'une «MACRO» EXCEL.

En réitérant une seconde fois cette procédure, nous avons ***divisé par 4 le nombre initial de fréquences potentiellement informatives*** : finalement nous avons obtenu un tableau de 239 lignes et 213 colonnes. Ce tableau de données peut maintenant être transposé pour conduire à un tableau de 239 colonnes et 213 lignes, parfaitement compatible avec les contraintes d'EXCEL (< 256 colonnes).

8.7.1.2 - Exploitation des données spectrales

Identification des zones spectrales d'intérêt

Cette réduction importante, mais raisonnée, du nombre initial de variables explicatives (les absorbances), ne suffit toujours pas pour faire une régression multilinéaire classique puisque nous avons toujours plus de descripteurs (239) que d'individus (213). De plus, compte tenu que nous avons 6 spectres par mélange, les 213 individus ne correspondent en réalité qu'à trente cinq équations réellement différentes. D'où la nécessité d'utiliser la méthode de régression des moindres carrés partiels PLS.

Bien que la méthode de régression PLS tolère un certain niveau de corrélation entre les variables explicatives et qu'elle permette de limiter les conséquences des interférences spectrales et du « bruit » expérimental, il est néanmoins important d'éliminer au maximum les données qui n'apporteraient que du bruit.

Pour identifier les zones réellement informatives utilisables pour l'analyse quantitative nous devons utiliser *l'intersection des zones spectrales* correspondant simultanément à la *variance maximale* entre l'ensemble des spectres ET à la *corrélation maximale* entre absorbances et compositions centésimales.

Comme chaque fréquence étudiée correspond à une colonne du tableau EXCEL, il est facile de calculer la variance de chacune d'elle et de tracer l'histogramme des variances. De la même manière, l'utilitaire d'analyse d'EXCEL permet de calculer et d'afficher la matrice de corrélation entre variables. On identifie de cette manière les fréquences qui présentent le maximum de corrélations avec les concentrations en formes polymorphes.

La réunion et/ou l'intersection des fréquences identifiées selon les deux approches constituent les fenêtres spectrales qui présentent le maximum de probabilité d'être des zones informatives. Bien entendu le *compromis n'est pas univoque*. On peut envisager un certain nombre de combinaisons de fréquences à priori informatives.

Choix des mélanges de calibration et des mélanges de validation

La matrice de calibration (d'étalonnage) est constituée normalement des mélanges 1 à 20 et 23, les points 31 à 34 représentant les points de validation (points tests internes). Toute méthode de régression linéaire, la méthode PLS ne faisant pas exception à la règle, est très sensible aux points s'écartant le plus du modèle : ces points doivent-ils alors être considérés comme aberrants ? (sont-ils des « outliers » selon la terminologie anglaise trop souvent utilisée par les analystes). Nous mentionnerons nos choix au cas par cas.

La matrice d'étalonnage donne des valeurs théoriques des concentrations en formes polymorphes qu'il est souvent impossible de réaliser exactement expérimentalement. *La matrice des concentrations réelles est connue* à partir des données reportées dans le cahier de laboratoire de l'expérimentateur. En toute rigueur, seules les concentrations réelles doivent être utilisées dans les calculs, néanmoins les valeurs théoriques disponibles dans les fichiers de départ peuvent constituer une approximation suffisante.

Modèles de calibration

Nous avons réalisé **9 modèles prévisionnels** des quantités de chaque forme polymorphe du CL20. Chaque modèle a été construit en fonction des fréquences et des mélanges retenus pour le calcul. Ces modèles, obtenus à l'aide du logiciel SIMCA, sont référencés de MS₁ à MS₉. Les modèles MS₁ à MS₄ correspondent à l'utilisation des concentrations théoriques des mélanges du plan d'expériences, les modèles MS₅ à MS₉ correspondent à l'utilisation des concentrations réalisées expérimentalement qui en diffèrent quelque peu (on ne peut que *pe-ser exactement des quantités de matières approchant les valeurs théoriques* que l'on désire).

Les modèles MS₁ à MS₈ sont établis à partir de tous les spectres des mélanges 1 à 20 et 23 alors que le modèle MS₉ correspond à l'exclusion des spectres de la pastille 3 du mélange 7, cette dernière semblant être un point aberrant.

Aucun traitement des spectres n'a été effectué pour élaborer ces modèles (travail en absorbances brutes).

Modèles	Réf. fichiers	Conditions de modélisation (concentrations théoriques)
MS₁	M ₁ Variance	Suppression des fréquences ayant les plus petites variances
MS₂	M ₂ Corrélation	Suppression des fréquences ayant une corrélation avec les concentrations <1
MS₃	M ₃ Réunion	Suppression de l'ensemble des zones M ₁ + M ₂ .
MS₄	M ₂ Intersection	Suppression des zones communes à M ₁ et M ₂ .

Modèles	Réf. fichiers	Conditions de modélisation (concentrations réelles)
MS ₅	M ₅	Suppression des fréquences idem M ₄ .
MS ₆	M ₆ Variance2	Suppression des fréquences ayant une variance < 0,006
MS ₇	M ₇ Réunion Variance-Corrélation	Suppression de l'ensemble des zones M ₂ + M ₆ .
MS ₈	M ₈ Intersection Variance-Corrélation	Suppression de l'ensemble des zones de variance < 0,006 et de corrélation < 1.
MS ₉	M ₉ Intersection	idem M ₈ mais suppression du spectre du mélange 7 ₃ .

Nombre des composantes retenues pour chaque modèle

Le critère de détermination du nombre optimal de composantes à utiliser dans le modèle d'étalonnage correspond à la minimisation de la somme des carrés des résidus obtenus par validation croisée du modèle testé (PRESS, Q² ou test équivalent). Les résultats sont donnés tableau 15.

Tableau 15 : nombre de composantes retenus

Modèles	Nb de composantes	Modèles	Nb de composantes
MS ₁	9	MS ₅	9
MS ₂	8	MS ₆	9
MS ₃	6	MS ₇	8
MS ₄	9	MS ₈	8
		MS ₉	8

Examen des résidus [84]

Parmi les outils d'analyse de la qualité de l'étalonnage que nous avons utilisés, l'*examen de la distribution des résidus via la droite de Henry* présente l'avantage de la simplicité (méthode purement graphique : valeur estimée = f(valeur réelle)). En effet, comme pour toute régression, si le modèle d'étalonnage postulé ajuste correctement le phénomène physique, les écarts entre les valeurs prédites par le modèle et les valeurs expérimentales réelles (écarts appelés résidus) sont normalement distribués. La présence de résidus «anormalement» élevés (s'écartant significativement de la droite de Henry) peut correspondre à des points d'étalonnage aberrants (outliers). Après une étude critique de leur «aberration», *les points jugés aberrants doivent être éliminés de la matrice d'étalonnage.*

8.7.1.3 - Analyse des modèles retenus

La régression à partir des concentrations théoriques (modèles MS₁ à MS₄) ne présentant qu'un intérêt académique, nous ne reportons ici que l'étalonnage à partir des concentrations expérimentales. Tous traitements effectués, nous présentons ici les trois modèles les plus performants : MS₇, MS₈ et MS₉. Dans le tableau 16, nous avons rassemblé les résultats des dosages des points tests internes destinés à valider les 3 modèles postulés. Comme pour un

modèle donné et pour un mélange test donné nous disposons de 6 spectres, nous avons 6 évaluations de sa composition : pour chacune des formes polymorphes nous pouvons rendre la valeur moyenne de sa concentration estimée et un écart type représentant la variabilité de ces estimations individuelles.

Compte tenu qu'une régression PLS passe par des détours mathématiques complexes, *il n'est pas possible de donner ici, de manière accessible, un véritable intervalle de confiance sur une teneur prédite*. Néanmoins, nous pouvons examiner deux indicateurs statistiques simples : le *coefficient de variation* qui rapporte l'écart type d'une mesure à sa valeur moyenne et l'*erreur relative sur une teneur* en rapportant le résidu moyen à la valeur théorique de la teneur. Ces indicateurs simples sont rassemblés dans le tableau 17.

Tableau 16 : validation des modèles à partir des points tests internes

Méthodes	Pts valid.	Compositions réelles				Compositions estimées par les Méthodes de dosage							
		ϵ (%)	α (%)	β (%)	γ (%)	ϵ (%)	σ	α (%)	σ	β (%)	σ	γ (%)	σ
MS ₇	31	81.3	7.7	7.7	3.3	80.2	0.7	8.6	0.2	7.3	0.6	3.9	0.3
	32	85.6	3.3	3.3	7.7	85.1	0.7	4.6	0.4	1.1	0.5	9.1	0.5
	33	81.3	7.7	3.3	7.7	80.5	0.4	8.7	0.3	1.7	0.6	9.0	0.3
	34	81.3	3.3	7.7	7.7	79.8	0.9	3.3	0.3	7.8	1.2	9.0	0.3
MS ₈	31	81.3	7.7	7.7	3.3	81.1	0.7	8.0	0.4	7.2	0.6	3.5	0.5
	32	85.6	3.3	3.3	7.7	85.0	0.9	4.2	0.4	1.6	0.3	9.1	0.6
	33	81.3	7.7	3.3	7.7	81.0	0.7	8.5	0.4	1.7	0.4	8.7	0.5
	34	81.3	3.3	7.7	7.7	78.2	1.1	3.6	0.2	8.4	1.1	9.8	0.3
MS ₉	31	81.3	7.7	7.7	3.3	80.9	0.8	8.1	0.4	7.1	0.5	3.8	0.5
	32	85.6	3.3	3.3	7.7	84.5	0.8	4.2	0.4	1.7	0.3	9.5	0.5
	33	81.3	7.7	3.3	7.7	80.8	0.6	8.5	0.4	1.6	0.4	9.0	0.5
	34	81.3	3.3	7.7	7.7	78.0	1.1	3.5	0.2	8.6	1.1	9.9	0.3

Une première analyse globale montre que les trois modèles retenus présentent un intérêt équivalent et qu'ils peuvent servir de base à un développement de méthodes avec TURBO QUANT. L'examen du tableau 17 montre que *dans tous les cas on obtient une qualité de prédiction satisfaisante pour la teneur en forme ϵ* . Pour les autres formes polymorphes, d'une manière générale, les coefficients de variation et les erreurs relatives sont particulièrement élevés pour des teneurs de 3,3%, beaucoup plus acceptables pour celles de 7,7%.

On remarque également que *la forme polymorphe α semble mieux estimée* que β et γ , la forme β paraissant de loin la plus mal estimée.

Tableau 17 : coefficients de variation et erreurs relatives

Méthodes	Pts. valid	Teneurs théoriques				Forme ϵ		Forme α		Forme β		Forme γ	
		ϵ (%)	α (%)	β (%)	γ (%)	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %
MS ₇	31	81.3	7.7	7.7	3.3	0.8	1.3	2.5	11.0	7.8	5.3	7.9	17.6
	32	85.6	3.3	3.3	7.7	0.8	0.5	10.6	39.7	13.6	66.1	6.9	18.1
	33	81.3	7.7	3.3	7.7	0.5	0.9	3.6	12.5	17.6	47.3	3.2	16.6
	34	81.3	3.3	7.7	7.7	1.2	1.9	9.4	0.6	15.5	1.6	3.6	16.8
MS ₈	31	81.3	7.7	7.7	3.3	0.9	0.2	5.1	4.4	7.8	6.1	14.2	6.7
	32	85.6	3.3	3.3	7.7	1.1	0.7	12.4	28.2	9.1	51.2	8.3	18.6
	33	81.3	7.7	3.3	7.7	0.8	0.3	4.9	10.5	12.7	50.0	5.8	13.5
	34	81.3	3.3	7.7	7.7	1.4	3.8	5.8	8.2	14.3	8.7	4.0	26.9
MS ₉	31	81.3	7.7	7.7	3.3	1.0	0.5	4.8	5.1	7.0	7.4	14.8	14.5
	32	85.6	3.3	3.3	7.7	1.0	1.3	11.5	27.6	8.2	49.1	6.5	23.9
	33	81.3	7.7	3.3	7.7	0.8	0.6	5.1	10.9	13.3	51.2	6.6	16.6
	34	81.3	3.3	7.7	7.7	1.4	4.1	6.7	5.5	13.9	11.8	4.2	28.2

8.7.2 - Méthodes d'analyse développées en «Interne» avec Turbo Quant

Les premiers développements ont été basés sur les modèles MS₇ et MS₉ obtenus avec SIMCA, en essayant d'approcher au mieux leurs fenêtres spectrales avec TURBO QUANT. Dans ce contexte, nous avons retenu trois sélections spectrales qui ont conduit aux méthodes référencées dans ce document par TQ₁, TQ₂ et TQ₃ (leurs correspondances avec les fichiers de travail sont reportées dans le tableau 18). Néanmoins, à la *différence des modèles développés avec SIMCA*, pour ces modèles TQ₁, TQ₂ et TQ₃ nous avons *intégré le spectre de la forme ϵ* pure à la matrice d'étalonnage (donc le point [ϵ] = 100%, correspondant à un point 0 pour α , β et γ), ce qui se justifie *dans la mesure où les échantillons industriels ont des teneurs probablement > 97%*. Comme avec SIMCA, aucun traitement des spectres n'a été réalisé pour élaborer ces modèles (travail en absorbances brutes).

Tableau 18 : correspondances des méthodes avec les documents internes

Méthode	Référence interne	Fenêtres spectrales
TQ ₁	CL20 T Q PLS-3	5 fenêtres spectr. : 3713 – 3567 ; 1160 – 1076 ; 1027 – 1006 ; 985 – 935 ; 773 – 742 cm ⁻¹ .
TQ ₂	CL20 T Q PLS-6	8 fenêtres spectr. : les 5 précédentes + 692 – 653 ; 624 – 604 et 536 – 507 cm ⁻¹ .
TQ ₃	CL20 T Q PLS-6+	5 fenêtres spectr. : 1160 – 1076 ; 1027 – 1006 ; 985 – 935 ; 773 – 742 ; 695 – 599 cm ⁻¹ .

Pour chacun des composés analysés, un premier indicateur sur la validité du modèle d'étalonnage est donné par le coefficient de régression R². Nous avons rassemblé ces indicateurs dans le tableau 19.

Tableau 19 : coefficients de régression obtenus pour les modèles retenus

	TQ ₁	TQ ₂	TQ ₃
α	0.996	0.996	0.997
β	0.988	0.990	0.990
γ	0.996	0.996	0.996
ε	0.994	0.995	0.995

Dans le tableau 20, nous avons rassemblé les résultats des dosages des points tests internes destinés à valider les 3 modèles postulés. Comme pour les modèles développés avec SIMCA, pour chacune des formes polymorphes nous pouvons rendre la valeur moyenne de sa concentration estimée et un écart type représentant la dispersion de ces estimations individuelles.

Tableau 20 : validation des modèles à partir des points tests internes

Méthodes	Pts valid.	Compositions réelles				Compositions estimées par les Méthodes de dosage							
		ε (%)	α (%)	β (%)	γ (%)	ε (%)	σ	α (%)	σ	β (%)	σ	γ (%)	σ
TQ ₁	31	81.3	7.7	7.7	3.3	82,9	0,3	8.1	0,1	7,2	0,3	2,8	0,2
	32	85.6	3.3	3.3	7.7	86,7	0,4	3,7	0,2	3,3	0,4	7,1	0,2
	33	81.3	7.7	3.3	7.7	81,9	0,5	8,2	0,1	3,7	0,6	7,0	0,2
	34	81.3	3.3	7.7	7.7	80,3	0,3	3,9	0,1	8,2	0,2	7,4	0,1
TQ ₂	31	81.3	7.7	7.7	3.3	82,3	0,3	7,9	0,1	7,6	0,3	2,7	0,1
	32	85.6	3.3	3.3	7.7	86,2	0,4	3,5	0,2	3,7	0,4	6,9	0,1
	33	81.3	7.7	3.3	7.7	81,5	0,4	8,0	0,1	3,9	0,5	6,8	0,2
	34	81.3	3.3	7.7	7.7	80,7	0,3	4,0	0,1	8,1	0,2	7,2	0,1
TQ ₃	31	81.3	7.7	7.7	3.3	81,1	0,4	7,8	0,1	7,6	0,3	3,3	0,2
	32	85.6	3.3	3.3	7.7	85,1	0,5	3,2	0,2	3,7	0,4	7,5	0,2
	33	81.3	7.7	3.3	7.7	81,2	0,3	7,9	0,1	3,9	0,3	7,4	0,1
	34	81.3	3.3	7.7	7.7	81,2	0,4	4,0	0,1	7,8	0,3	7,2	0,2

De la même manière que précédemment, nous examinons les deux indicateurs coefficient de variation et erreur relative que nous avons rassemblés dans le tableau 21.

Comme pour les modèles développés avec SIMCA, l'examen de ce tableau montre que dans tous les cas on obtient une qualité de prédiction satisfaisante pour la teneur en forme ε . De même, pour les autres formes polymorphes, d'une manière générale, les coefficients de variation et les erreurs relatives sont particulièrement élevés pour des teneurs de 3,3%, beaucoup plus acceptables pour celles de 7,7%. On peut également remarquer la bonne prédiction globale pour le mélange test N°34.

Tableau 21 : coefficients de variation et erreurs relatives

Méthodes	Pts. valid	Teneurs théoriques				Forme ϵ		Forme α		Forme β		Forme γ	
		ϵ (%)	α (%)	β (%)	γ (%)	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %
TQ ₁	31	81.3	7.7	7.7	3.3	0.4	2.0	1.7	4.7	3.9	7.0	6.1	16.7
	32	85.6	3.3	3.3	7.7	0.5	1.3	5.5	11.5	13.3	1.2	2.6	8.3
	33	81.3	7.7	3.3	7.7	0.6	0.8	1.3	6.2	18.2	12.1	1.9	9.2
	34	81.3	3.3	7.7	7.7	0.3	1.2	1.5	19.4	1.9	6.8	1.2	3.8
TQ ₂	31	81.3	7.7	7.7	3.3	0.4	1.2	1.6	2.3	3.9	1.9	3.3	17.3
	32	85.6	3.3	3.3	7.7	0.5	0.7	4.5	6.1	13.0	10.9	1.7	10.6
	33	81.3	7.7	3.3	7.7	0.5	0.3	1.0	3.8	13.9	19.4	1.9	11.9
	34	81.3	3.3	7.7	7.7	0.4	0.8	0.9	21.8	2.7	5.5	1.4	7.1
TQ ₃	31	81.3	7.7	7.7	3.3	0.5	0.3	1.2	0.6	4.2	1.0	7.3	0.6
	32	85.6	3.3	3.3	7.7	0.6	0.6	6.4	2.7	12.1	12.1	2.6	2.1
	33	81.3	7.7	3.3	7.7	0.3	0.1	1.0	1.9	8.5	17.6	1.8	3.9
	34	81.3	3.3	7.7	7.7	0.5	0.1	0.6	22.4	3.9	1.0	1.9	6.2

Si les performances des modèles TQ_i sont globalement supérieures à celles des modèles MS_i, cela ne provient pas d'une différence de performance des logiciels où de leur utilisation, mais du fait que *la matrice d'étalonnage n'est pas la même* : la matrice utilisée pour développer les TQ_i correspond à *l'ajout du spectre de la forme ϵ pure*. La réflexion qui a conduit à cet ajout s'est révélé judicieuse, dans la mesure où *elle a amélioré la qualité globale des estimations*, en particulier pour les formes α et β .

8.7.3 - Analyses de lots industriels

8.7.3.1 - Etude préliminaire

Nous avons utilisé ces modèles d'étalonnage pour une première approche de l'analyse de lots industriels. Nous reportons dans le tableau 22 les résultats obtenus pour 5 de ces lots.

Tableau 22 : teneurs en forme ϵ de lots industriels

Références des lots industriels	Teneurs en ϵ estimées par les trois méthodes					
	TQ ₁		TQ ₂		TQ ₃	
	%	σ	%	σ	%	σ
3	97.8	0.9	97.9	1.1	95.8	1.9
6	100.3	1.0	101.7	0.9	100.1	1.2
7	98.3	0.9	100.1	0.3	98.7	0.3
8	98.1	1.1	100.0	1.8	98.9	1.3
15	97.9	1.6	99.6	1.3	99.6	1.6

Bien que l'on ne dispose pas d'analyses de référence pour juger de la validité de ces résultats, ils convergent pour suggérer que les lots industriels ont des teneurs en forme ϵ plus élevées que ce que l'on supposait au départ de cette étude (probablement >98%).

8.7.3.2 - Analyse d'un échantillonnage de 14 lots industriels

A l'issue de ces résultats, nous avons retenu seulement **deux modèles** pour examiner un **nombre plus important** de lots industriels. Nous avons retenu la transposition à Turbo Quant du modèle MS₈ et le modèle TQ₃ qui nous paraissent présenter les meilleurs compromis globaux. Nous reportons dans le tableau 23 les analyses chimiques, granulométriques et les résultats obtenus pour l'analyse de la forme ϵ des 14 lots industriels que nous avons étudiés et dans le tableau 24 les résultats obtenus pour l'analyse des trois autres formes.

Tableau 23 : teneurs en forme ϵ de 14 lots industriels

Références des lots industriels	Pureté chimique %	impureté pentani-trée %	Granulométrie μm	Teneurs en ϵ estimées			
				MS ₈		TQ ₃	
				%	σ	%	σ
1	98.9	1.0	34.5	99.5	0.7	98.1	0.3
2	98.6	1.2	29.4	98.4	0.6	98.3	0.4
3	99.2	0.2	35.0	94.4	2.6	95.8	1.9
4	98.9	0.5	28.0	94.6	2.0	96.4	1.9
5	98.9	0.7	36.0	94.6	1.5	97.2	1.2
6	99.2	0.6	111.0	99.8	0.4	100.1	1.2
7	99.2	0.7	5.5	98.1	1.9	98.7	0.3
8	99.1	0.8	27.0	97.0	1.9	98.9	1.3
9	98.5	1.1	≈ 300.0	91.3	0.6	90.3	0.6
10	98.4	1.2	31.0	99.4	1.0	100.9	0.7
11	98.6	1.2	32.0	99.6	0.7	99.5	1.0
12	98.9	1.0	151.0	95.9	1.5	98.7	0.9
13	98.5	1.3	-	-	-	>100.0	-
14	98.5	1.3	-	-	-	>100.0	-

- non mesuré

Tableau 24 : teneurs en formes α , β et γ de 14 lots industriels

Références des lots industriels	Teneurs en α estimées				Teneurs en β estimées				Teneurs en γ estimées			
	MS ₈		TQ ₃		MS ₈		TQ ₃		MS ₈		TQ ₃	
	%	σ	%	σ	%	σ	%	σ	%	σ	%	σ
1	0.0		1.1	0.2	3.0	0.4	<1.0		0.0		1.0	0.1
2	0.0		1.5	0.2	3.0	0.5	0.0		0.0		1.5	0.3
3	<1.0		2.1	0.8	3.7	2.4	<1.0		1.7	1.4	<1.0	
4	<1.0		1.6	0.6	2.5	1.7	<1.0		2.4	1.5	<1.0	
5	1.4	0.5	2.1	0.5	1.8	1.7	<1.0		2.3	0.4	<1.0	
6	<1.0		1.4	0.5	0.0		0.0		0.0		0.0	
7	0.0		<1.0		<1.0		0.0		2.0	0.9	<1.0	
8	<1.0		2.2	0.4	2.3	1.6	0.0		0.0		<1.0	
9	8.6	0.5	11.1	0.3	1.9	0.5	0.0		0.0		<1.0	
10	0.0		1.6	0.5	1.7	0.3	0.0		0.0		<1.0	
11	0.0		1.9	0.3	1.4	1.3	0.0		0.0		<1.0	
12	1.4	0.4	1.6	0.2	1.4	0.8	0.0		1.3	0.7	<1.0	
13			0.0				0.0				0.0	
14			0.0				0.0				0.0	

Les lots 13 et 14 sont obtenus par recristallisation du 9.

8.7.3.3 - Discussion des résultats pour l'analyse de la forme ϵ

Pour un même lot, la *comparaison directe des différences observées entre les estimations des deux méthodes d'analyse n'est pas possible* «arithmétiquement» car il faut tenir compte des écarts types observés sur ces estimations. Par exemple, pour les différences apparentes les plus élevées qui sont observées pour les lots 5 et 12 (les estimations du modèle MS₈ étant les plus faibles), statistiquement il y a plus de 92% de probabilité que ces différences ne soient pas significatives. On peut dire globalement qu'il n'y a *probablement pas de différences*, statistiquement parlant, entre les deux méthodes.

Attention : indépendamment de la remarque précédente, *toute comparaison éventuelle n'est pas une comparaison des performances des étalonnages* internes ou externes car *ils ne sont pas effectués à partir de la même matrice d'étalonnage*, la méthode TQ₃ incluant la forme ϵ pure dans la matrice d'étalonnage.

Examen des lots 1, 5 et 12 : néanmoins, pour les lots 5 et 12, nous avons fait l'hypothèse que la différence observée entre les estimations de la teneur en ϵ des 2 modèles pouvait correspondre à un problème *d'hétérogénéité des pastilles*, conséquence du mélange-broyage qui est effectué manuellement. Nous avons remarqué que plus la granulométrie de l'échantillon est élevée, plus l'homogénéité de la pastille est difficile à obtenir. La non homogénéité des pastilles peut se traduire par une modification des spectres qui, associée au

fait que les zones spectrales choisies pour les 2 modèles ne sont pas identiques, se traduit par une différence des estimations. Pour en revenir à une différence éventuelle de comportement des lots 1 et 12, l'examen du tableau 23 nous permet de remarquer que la granulométrie du lot 12 est particulièrement élevée (34,5 μ m pour le lot 1 et 151 μ m pour le lot 12) alors que ces lots ne se différencient ni par leur pureté chimique, ni par leur teneur en impureté pentanitrique.

Examen du lot 9 : la teneur en forme ϵ du lot 9 ($[\epsilon] \approx 90-91\%$) est beaucoup plus faible que celles des autres lots alors que sa teneur en forme α est la plus élevée ($[\alpha] \approx 9-11\%$). Or la forme α est la forme hydratée du CL20, elle se forme majoritairement dès qu'il y a présence d'eau dans un lot de CL20. Il en résulte que le taux d'eau du lot 9 est de 5 à 6 fois celui d'un lot de CL20 «normal», d'où l'abaissement concomitant de la teneur en forme ϵ .

Examen des lots 13 et 14 : rappelons que ces lots sont obtenus par traitement du lot 9 dans le but d'en diminuer l'hydratation en éliminant un maximum de forme α . *Les résultats obtenus après traitement paraissent aberrants* dans la mesure où la teneur estimée en forme ϵ est $>100\%$ et que celles en impuretés α , β et γ sont nulles. On peut expliquer les résultats obtenus par le fait que *les formes α , β et γ ont été éliminées en partie ou totalement au cours du traitement* (en particulier la forme α). Quoiqu'il en soit, *après traitement*, les concentrations finales en chacune des formes sont *hors domaine d'étalonnage*. Les estimations obtenues résultent de l'extrapolation des modèles, elles traduisent simplement le haut degré de pureté des lots traités en forme ϵ .

Influence de la concentration en impureté pentanitrique : en examinant le tableau 14, on peut constater que *d'un lot à l'autre la teneur en impureté pentanitrique peut varier de 0,2 à 1,3%*. En particulier les lots 13 et 14 précédents ont une teneur de 1,3%, double de celle des étalons qui ont servi à la matrice d'étalonnage. Or rappelons que cette dernière a été conçue *sous l'hypothèse d'une teneur quasi constante* en impureté pentanitrique. *Si cette condition n'est pas respectée, il en découle une insuffisance prévisionnelle des modèles* d'étalonnage (ce qui dans le cas ci-dessus s'ajoute aux effets de la recristallisation).

8.7.3.4 - Discussion des résultats pour l'analyse des formes α , β et γ

Un examen superficiel pourrait suggérer que l'on obtienne en moyenne des estimations des teneurs en forme α plus faibles avec le modèle MS₈ qu'avec le modèle TQ₃, que ce serait l'inverse pour la forme β et qu'il n'y aurait pas de différences pour la forme γ .

Comme pour la forme ϵ , *la comparaison directe des différences observées* entre les estimations des deux méthodes d'analyse *n'est pas possible* «arithmétiquement» car il faut tenir compte des écarts types observés sur ces estimations. Là encore, on peut dire globalement qu'il n'y a pas de différences, statistiquement parlant, entre les deux méthodes.

8.7.4 - *Conclusion*

Ce premier développement d'une méthode d'analyse du polymorphisme du CL20 nous a permis de *valider la méthodologie d'étalonnage basée sur une approche chimométrique* que nous avons utilisée. A partir des résultats de l'étalonnage «externe» (traitement des données avec SIMCA), *il a été possible d'obtenir un étalonnage «interne» acceptable* (utilisation de TURBO QUANT intégré au logiciel du spectromètre). En effet, pour les échantillons appartenant au domaine d'étalonnage, les écarts entre valeurs prédites et valeurs réelles sont de l'ordre de grandeur de la dispersion expérimentale des mesures : *la qualité* de cet étalonnage peut donc être *qualifiée d'acceptable* dans le domaine de concentrations utilisé.

Pour l'ensemble des lots industriels que nous avons analysés à partir de ce premier étalonnage, il s'est avéré que *leur teneur en forme ϵ était beaucoup plus élevée* que nous le pensions et que de plus, certaines formes polymorphes pouvaient être absentes (ou tout au moins non quantifiables). Il en résulte que les domaines d'étalonnage des formes α , β et γ que nous avons fixés entre 1 et 10% ne correspondent pas aux teneurs réelles, ils doivent être recentrés sur ces dernières (de 0,1 à 2% ?), *ce qui implique un nouvel étalonnage*.

Parmi les sources de variabilité affectant la précision du dosage spectrométrique, nous devons également *nous interroger sur l'influence de l'impureté pentanitree*. Nous avons fait l'hypothèse a priori que le procédé industriel conduisait à une teneur approximativement constante de l'ordre de 0,7% (facteur constant de notre matrice d'étalonnage). Si cette hypothèse est invalidée par la réalité industrielle, *cette impureté doit faire partie des éléments à quantifier*, elle doit donc être prise en compte dans la matrice d'étalonnage.

Enfin, il est probable que *la granulométrie de départ de l'échantillon influe sur la dispersion* des caractéristiques des pastilles : plus la granulométrie est élevée, plus la dispersion des caractéristiques est élevée. Pour pallier cette difficulté et minimiser l'influence de la variabilité intrinsèque du geste «manuel», *il faut envisager l'utilisation d'un vibro-broyeur*.

Malgré les limites de cette première tentative, *les résultats sont très encourageants*. Notons surtout que, *pour la première fois, le groupe SNPE dispose d'une méthode d'analyse polymorphique quantitative* de lots industriels de CL20.

IX - DEVELOPPEMENT DE LA METHODE D'ANALYSE QUANTITATIVE : APPROCHE DEFINITIVE

9.1 - Travaux préliminaires

Avant de procéder à un second étalonnage, il nous a paru nécessaire d'acquérir des informations supplémentaires concernant les étalons dont nous disposons (via une étude ATD), les limite de détection en IR et l'influence éventuelle de la teneur en impureté pentanitree sur le comportement du modèle d'étalonnage.

9.1.1 - Analyse thermique différentielle

Toutes les mesures ont été effectuées avec un appareil DSC20 à $8^{\circ}\text{C}\cdot\text{min}^{-1}$, la masse de l'échantillon a varié de 6,4 à 9,3mg selon l'étude. Comme valeurs de référence des températures et des enthalpies de transition, nous avons retenu celles issues des travaux de M.F.FOLTZ [23] (annexe 7).

Forme α (ref A2) : pour cet échantillon le pic de transition est décalé par rapport aux valeurs de la littérature : nous observons une température de transition de $154,2^{\circ}\text{C}$ avec une enthalpie de $35,1\text{J}\cdot\text{g}^{-1}$ pour $172,5^{\circ}\text{C}$ et $12,5\text{J}\cdot\text{g}^{-1}$ reportés dans la littérature. Nous avons examiné un second échantillon de forme α (ref A3) : nous avons observé un thermogramme complexe présentant quatre zones de transition avec deux pics simples et deux massifs

- deux pics vers 86 et 100°C avec respectivement des enthalpies $< 0,4\text{J}\cdot\text{g}^{-1}$,
- un premier massif M_1 dans l'intervalle de température $135-158^{\circ}\text{C}$ constitué de 3 pics à $140,1^{\circ}\text{C}$, $144,6^{\circ}\text{C}$ et $151,7^{\circ}\text{C}$ avec respectivement des enthalpies de $0,7\text{J}\cdot\text{g}^{-1}$, $0,5$ et $1,1\text{J}\cdot\text{g}^{-1}$ (soit une enthalpie totale de $2,3\text{J}\cdot\text{g}^{-1}$),

- un second massif M_2 dans l'intervalle de température 163-203°C constitué de 2 pics à 174,7°C et 184,8°C avec respectivement des enthalpies de 2,0 et 5,8J.g⁻¹ (soit une enthalpie totale de 7,8J.g⁻¹).

Forme β (ref B2) : pour cet échantillon le pic de transition est décalé par rapport aux valeurs de la littérature : nous observons une température de transition de 146,3°C avec une enthalpie de 4,2J.g⁻¹ pour 145,5°C et 1,37J.g⁻¹ reportés dans la littérature. Nous avons examiné un second échantillon de forme β (ref B3) : nous observons alors deux températures de transition à 142,4 (pic principal) et 174,3°C (pic secondaire) avec respectivement des enthalpies de 2,2 et 1,45J.g⁻¹.

Forme γ (ref G3) : nous n'observons aucune différence avec les valeurs de M.F.FOLTZ. Il semblerait que l'échantillon analysé soit pur.

Forme ϵ (ref E3) : nous observons une température de transition à 169,7°C avec une enthalpie de 24,4J.g⁻¹ alors que la littérature signale 168,2°C et 17,4J.g⁻¹. Ces faibles écarts peuvent être probablement attribués à une légère différence en taux d'impureté(s).

Les plus grandes différences entre les caractéristiques thermiques de nos échantillons et celles de la littérature étant observées pour la forme α qui est une forme hydratée du CL20, nous avons pensé que cela pouvait provenir d'une déshydratation dans la mesure où 1% d'eau implique une enthalpie de 23J.g⁻¹. Afin d'étudier cette hypothèse nous avons fait des analyses comparatives ATD et IR de l'étalon α (ref A3) avant et après traitement thermique sous vide à 100°C jusqu'à poids constant. Après traitement thermique de l'échantillon (ref A4) :

- en ATD nous n'observons plus qu'un seul massif contenant 3 pics à 169,2°C, 176,1 et 180,7°C avec respectivement des enthalpies de 0,8J.g⁻¹, 1,52 et 3,34J.g⁻¹ (soit une enthalpie totale de 5,6J.g⁻¹). Un très petit pic observé à 146,8°C avec une enthalpie de 0,1J.g⁻¹ suggère la présence d'une faible quantité de forme β estimée à 0,6%. Le traitement thermique de l'échantillon a entraîné la disparition des deux premiers pics et du massif M_1 et une modification du massif M_2 avec l'apparition d'un troisième pic, un domaine de température qui passe de 163-203°C à 156-194°C et une baisse d'enthalpie totale de 7,8 à 5,7J.g⁻¹.

- en spectroscopie IR nous n'observons aucune modification spectrale.

- le dosage de l'eau des échantillons avant et après ce traitement thermique donne respectivement 1,9 et 1,1%, soit une perte de 0,8% d'eau (le taux théorique d'eau dans la forme α est $\approx 1\%$).

Pour vérifier également l'influence d'une hydratation éventuelle de l'étalon β , nous avons effectué un dosage d'eau et obtenu 0,4%. Cette teneur est de l'ordre du 1/5 de celle de l'étalon α (1,9%). Ces résultats sont cohérents avec les enthalpies totales observées pour nos échantillons (3,7J.g⁻¹ pour β et 10,8J.g⁻¹ pour α).

Nous nous sommes ensuite intéressé à l'influence éventuelle du gradient de température en ATD, en examinant le comportement d'un étalon α avec un gradient de 1°C.min⁻¹ à la place des 8°C.min⁻¹ de l'analyse ATD de départ. Si nous n'avons pas constaté de variation significative de l'enthalpie (32 pour 30,2J.g⁻¹), nous avons observé un abaissement du domaine de température de transition de (119-190°C pour un maximum du pic à 162,2°C pour 8°C.min⁻¹, alors que le domaine est de 111-184°C et le max du pic à 145°C pour 1°C.min⁻¹).

Enfin, nous avons mesuré la répétabilité de l'analyse ATD à partir de 3 analyses de l'étalon ϵ (ref E3) dans des conditions identiques.

Conclusion

Nous confirmons que la déshydratation peut être responsable du comportement de l'étalon α en ATD. Il en résulte que le taux d'eau contenu dans cette forme doit être un facteur de bruit pour l'analyse quantitative infrarouge.

9.1.2 - Sensibilité spectrale de la forme α

Nous avons examiné les variations spectrales de la forme α (ref A3) en y ajoutant respectivement 10 et 20% de forme β (ref B3) : les variations spectrales induites sont détectables visuellement. Les spectres IR obtenus sont reportés figure 61.

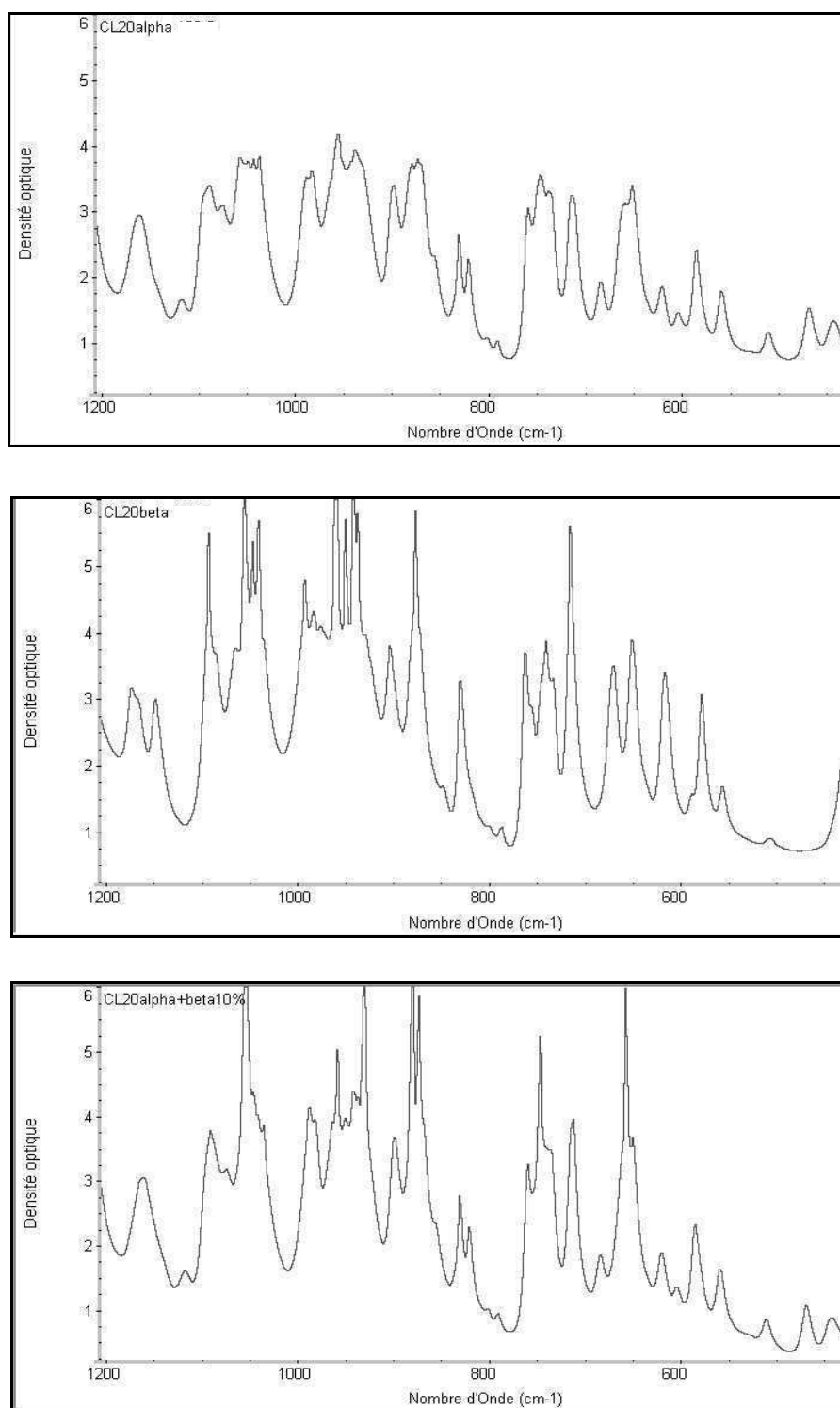


Figure 61 : spectres IR de alpha seul, bêta seul et mélange alpha +10% bêta

9.1.3 - Influence de l'impureté pentanitrée

Pour vérifier l'influence du composé pentanitré sur la méthode de dosage du CL20 dont il est l'impureté majoritaire, nous avons choisi le lot 8 de CL20 qui contient 0,8% de

dérivé pentanitré auquel nous avons successivement ajouté 0,7%, 1,2% et 2,2% de cette impureté. Nous avons analysé ces trois échantillons à l'aide du modèle d'étalonnage TQ₃. Le spectre de l'impureté pentanitrée est reporté figure 62, les résultats des dosages sont reportés dans le tableau 25.

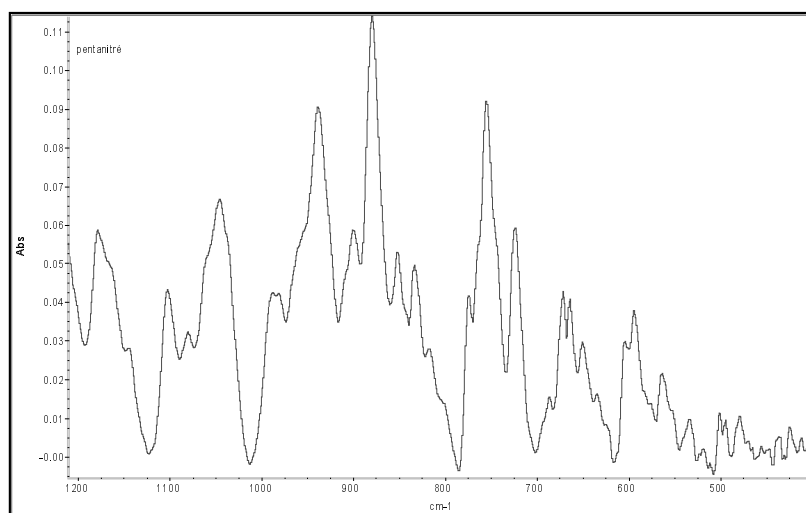


Figure 62 : spectre IR de l'impureté pentanitrée

Tableau 25 : teneurs en formes ϵ , α , β et γ après ajout de l'impureté pentanitrée

Taux total de penta	Analyses du lot 8 en fonction des ajouts de l'impureté pentanitrée							
	Alpha		Bêta		Gamma		Epsilon	
	%	σ	%	σ	%	σ	%	σ
0.8%	2.2	0.4	0.0		<1.0		98.9	1.3
1.5%	2.0	0.2	0.0		<1.0		99.1	0.7
2.0%	<1.0		<1.0		<1.0		98.7	0.7
3.0%	1.3	0.2	0.0		1.2	0.1	98.2	0.5

Malgré les grandes incertitudes sur les très faibles teneurs des formes α , β et γ , ces résultats suggèrent néanmoins *une corrélation négative entre teneur en impureté pentanitrée et estimation de la teneur en forme α* : cette dernière diminue quand la teneur en impureté augmente.

Pour pallier ce phénomène, *il est nécessaire d'intégrer la teneur en impureté pentanitrée* à la matrice d'étalonnage.

9.1.4 - Conclusion

A partir des informations issues de ces travaux préliminaires, nous pouvons définir le domaine d'étalonnage définitif qui va faire l'objet du second volet de nos travaux. Par rapport au premier volet, il apparaît nécessaire, non seulement de modifier les contraintes de concentration sur les formes mésomorphes, mais aussi de tenir compte de la présence du dérivé pentanitré.

L'analyse de nouveaux lots industriels du CL20 montre que les teneurs en formes polymorphes α , β et γ sont extérieures au domaine d'étalonnage préalablement étudié ; *pour développer une méthode d'analyse valable, il est nécessaire de faire un étalonnage dans le domaine d'intérêt réel*. La nouvelle gamme d'étalonnage en chacun des composants a été choisie de manière à correspondre à la variabilité des lots industriels. Les limites retenues sont les suivantes :

pour ϵ	:	$90,0 \leq x_1 \leq 98,8\%$
pour α	:	$0,3 \leq x_2 \leq 4,0\%$
pour β	:	$0,3 \leq x_3 \leq 2,0\%$
pour γ	:	$0,3 \leq x_4 \leq 2,0\%$
pour penta	:	$0,3 \leq x_5 \leq 2,0\%$

9.2 - Mise au point d'une nouvelle technique de préparation des pastilles de KBr

Nous avons décidé tout d'abord de réexaminer notre procédure de préparation de la pastille de KBr afin d'en *minimiser la variabilité expérimentale*. Cet examen nous a permis d'identifier deux sources de variabilité : l'homogénéisation du mélange KBr+CL20 et le rapport entre le diamètre de la pastille et la section du faisceau infrarouge, ce dernier n'analysant que le quart de la surface de la pastille. Pour pallier ces défauts, nous avons décidé d'utiliser :

- un vibro-broyeur à billes afin de *maximiser l'homogénéité* des dispersions dans le KBr tout en libérant l'opérateur de gestes longs et fastidieux ;

- une mini pastilleuse de 5mm de diamètre, de marque SPECAC, pour obtenir des pastilles presque entièrement scrutées par le faisceau infrarouge dont le diamètre est de 4mm. Ces caractéristiques géométriques permettent d'*éviter la saisie de deux spectres avec rotation intermédiaire* de la pastille.

Le vibro-broyeur utilisé (RETSCH type MM200, figure 63) comprend 2 bols de broyage en agate de 5ml contenant chacun 2 billes en agate de 7mm de diamètre.

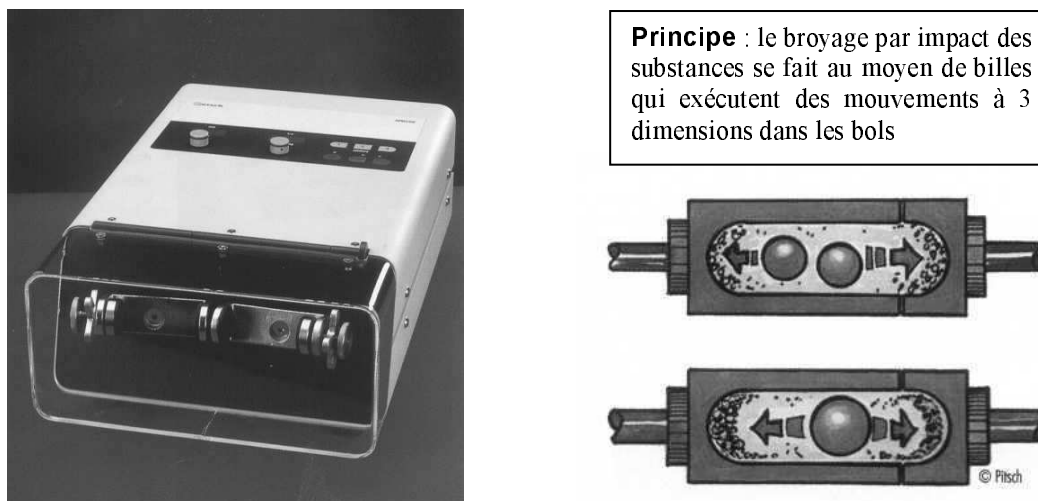


Figure 63 : schéma et principe du vibro-broyeur MM200

Cette modification de la méthode de préparation des pastilles implique une mise au point préliminaire destinée à identifier les conditions minimisant la dispersion des résultats. L'expérimentation correspondant à cette étude a été structurée selon la méthodologie des plans d'expériences.

L'échantillon de CL20 utilisé pour la réalisation des expériences a été synthétisé au laboratoire TCMP du CRB. Ce produit qui va nous servir de référence (G4) est amorphe, il titre 99,56% de CL20 et contient 0,22% d'impureté pentanitree. L'examen de son spectre IR suggère qu'il s'agit d'une forme de type gamma. Comme précédemment, les puretés polymorphique et chimique ont été contrôlées.

9.2.1 - Facteurs expérimentaux et réponses

On désire préparer deux pastilles, l'une contenant le mélange KBr+CL20 à étudier, l'autre ne contenant que du KBr pur destinée à mesurer le bruit de fond de la mesure (background). Afin que les produits servant à les préparer aient « la même histoire », en réalité des granulométries aussi voisines que possible, on introduit le mélange KBr+CL20 dans le premier bol du vibro-broyeur et du KBr pur dans le deuxième.

La concentration en CL20, le taux de remplissage des bols, la fréquence de broyage et la durée de broyage sont quatre facteurs expérimentaux que nous avons choisi d'étudier. Pour atteindre la teneur finale en CL20, ce premier broyat est dilué 10 fois par du KBr, en utilisant le même vibro-broyeur. Comme pour la première phase, durée de mélange et fréquence

des bols sont deux facteurs expérimentaux à étudier. Enfin, pour la réalisation de la pastille, on peut faire varier la pression et la durée de pastillage.

Pour contrôler l'hygroscopie du KBr, chaque mélange est conservé 24h au dessiccateur (sous vide, en présence de silicagel). Pour réaliser une pastille, on prélève 45mg de produit broyé. En résumé, le protocole de préparation de la pastille laisse apparaître trois étapes, chacune pouvant faire l'objet d'une étude par plans d'expériences :

- le broyage
- le mélangeage-dilution
- le pastillage

Les « réponses » mesurées sont les « propriétés » qui décrivent l'évolution des produits broyés et les caractéristiques de la pastille obtenue :

- la granulométrie après le broyage,
- la granulométrie après le mélangeage ;
- l'épaisseur de la pastille
- l'énergie absorbée par la pastille
- l'absorbance lors de la saisie du spectre IR

9.2.2 - Matrices d'expériences et plans d'expérimentations

Pour le broyage : les facteurs retenus sont au nombre de 4 : la fréquence et la durée de broyage, le taux de remplissage des bols et la concentration en CL20. Le domaine de variation choisi pour ces facteurs est décrit dans le tableau 26.

Tableau 26 : facteurs et domaine expérimental pour le broyage

Facteurs		Niveau -1	Niveau 0	Niveau 1
X ₁	Fréquence de broyage (Hz)	10	15	20
X ₂	Durée de broyage (min.)	5	10	15
X ₃	Taux de remplissage des bols (%)	13	15	17
X ₄	[CL20] dans le KBr (%)	6	8	10

Notre *objectif est de réaliser un « criblage » des facteurs* potentiellement influents : pour chacune des réponses étudiées, il s'agit seulement d'identifier les facteurs fortement, moyennement ou faiblement influents. Pour ce faire, on postule un modèle polynomial du

premier degré reliant l'effet des facteurs X_i (exprimés en variables codées $-1, +1$) sur la variation de la réponse Y étudiée :

$$Y = b_0 + \sum b_i X_i$$

Le maximum de précision sur l'estimation des coefficients d'un tel modèle est obtenu en *structurant les essais selon une matrice d'Hadamard* de 8 essais. De plus nous répétons 3 fois un essai au centre du domaine expérimental (valeur codée 0) : la valeur moyenne obtenue pour ce neuvième essai permet de vérifier (ou non) la linéarité du modèle (mise en évidence d'une courbure éventuelle de la fonction de réponse) ; d'autre part la répétition permet d'estimer la variance expérimentale de cette nouvelle procédure. La matrice d'expériences, le plan d'expérimentation associé et la granulométrie obtenue sont reportés tableau 27.

Tableau 27 : matrice d'expériences, plan d'expérimentation et résultats du broyage

N°Exp	X_1	X_2	X_3	X_4	Fréq. broyage (Hz)	Durée broyage (min)	Rempliss. des bols (%)	[CL20] (%)	Granulo. obtenue (μm)
1	1	1	1	-1	20	15	17	6	92
2	-1	1	1	1	10	15	17	10	275
3	-1	-1	1	1	10	5	17	10	343
4	1	-1	-1	1	20	5	13	10	154
5	-1	1	-1	-1	10	15	13	6	321
6	1	-1	1	-1	20	5	17	6	142
7	1	1	-1	1	20	15	13	10	72
8	-1	-1	-1	-1	10	5	13	6	365
9	0	0	0	0	15	10	15	8	343
10	0	0	0	0	15	10	15	8	290
11	0	0	0	0	15	10	15	8	329

Pour le mélangeage : les facteurs retenus sont la fréquence, la durée de mélangeage et la granulométrie du broyat à diluer. Ce dernier facteur est imposé par la première phase du protocole le broyage, les valeurs de la granulométrie étant converties en variables codées de -1 à 1 . Le domaine de variation choisi pour ces facteurs est décrit dans le tableau 28.

Tableau 28 : facteurs et domaine expérimental pour le mélangeage

Facteurs		Niveau -1	Niveau 0	Niveau 1
X_5	Fréquence de «mélangeage» (Hz)	10	15	20
X_6	Durée du «mélangeage» (min.)	5	10	15
G_1	Granulométrie du broyat (μm)	72	218.5	365

La matrice d'expériences et le plan d'expérimentation sont constitués de 11 essais. A noter que pour l'étude des facteurs X_5 et X_6 l'étude est effectuée par un plan factoriel répété. La matrice d'expériences, le plan d'expérimentation et les résultats sont reportés tableau 29.

Tableau 29 : matrice d'expériences, plan d'expérimentation et résultats du mélangeage

N°Exp.	X_5	X_6	G_1	Fréq. mélang. (Hz)	Durée mélang. (min)	Granulo. broyat (μm)	Granulo. obtenue (μm)
1	1	1	-0.86	20	15	92	103
2	-1	-1	0.39	10	5	275	315
3	-1	1	0.85	10	15	343	242
4	1	1	-0.44	20	15	154	85
5	-1	1	0.70	10	15	321	327
6	1	-1	-0.52	20	5	142	147
7	1	-1	-1.00	20	5	72	127
8	-1	-1	1.00	10	5	365	332
9	0	0	0.85	15	10	343	288
10	0	0	0.49	15	10	290	255
11	0	0	0.75	15	10	329	248

La variable codée sans dimension X_j (codage entre -1 et 1) et la variable naturelle U_j correspondante exprimée dans son système d'unité sont liées par la relation : $X_{ij} = (U_{ij} - U_j^0) / \Delta U_j$ avec U_j^0 = valeur de la variable U_j au centre de son domaine de variation [c'est à dire $(U_{j \max} + U_{j \min}) / 2$] et ΔU_j = demi-étendue de la variation de la variable U_j [c'est à dire $(U_{j \max} - U_{j \min}) / 2$].

Pour le pastillage : pour cette étape les facteurs retenus sont la granulométrie après mélangeage (facteur imposé par la phase de mélangeage-dilution du protocole) notée G_2 , la pression de pastillage et la durée d'application de la pression. Le domaine de variation choisi pour ces facteurs est décrit dans le tableau 30.

Tableau 30 : facteurs et domaine expérimental pour le pastillage

Facteurs		Niveau -1	Niveau 0	Niveau 1
G_2	Granulométrie après mélangeage (μm)	85	208.5	332
X_7	Pression de pastillage (tonne)	1	1.5	2
X_8	Durée d'application de la pression (s)	30	60	90

Ces trois facteurs sont étudiés en 11 essais avec la même stratégie que les trois précédents. La matrice d'expériences, le plan d'expérimentation et les résultats sont reportés tableau 31.

Tableau 31 : matrice d'expériences, plan d'expérimentation et résultats du pastillage

N°Exp.	G ₂	X ₇	X ₈	Granulo. mélang. (µm)	Press. Past. (tonne)	Durée. Past. (s)	Epais. Past. (mm)	Energie absorbée
1	-0.85	-1	-1	103	1.0	30	0.83	9.58
2	0.86	1	-1	315	2.0	30	0.83	9.64
3	0.27	-1	1	242	1.0	90	0.83	9.61
4	-1.00	1	-1	85	2.0	30	0.84	9.65
5	0.96	1	1	327	2.0	90	0.83	8.92
6	-0.50	1	1	147	2.0	90	0.82	9.58
7	-0.66	-1	1	127	1.0	90	0.83	9.65
8	1.00	-1	-1	332	1.0	30	0.82	9.55
9	0.64	0	0	288	1.5	60	0.82	9.25
10	0.38	0	0	255	1.5	60	0.82	9.32
11	0.32	0	0	248	1.5	60	0.83	9.39

9.2.3 - Mise en œuvre de l'expérimentation

La quantité de CL20 dans le KBr dépend à la fois du pourcentage de CL20 et du taux de remplissage : en associant ces deux facteurs on obtient au total 5 concentrations différentes ; dans une première étape nous avons donc réalisé ces 5 « pré-mélanges » (tableau 32). Pour un pré-mélange donné, la fréquence de broyage est constante, seule la durée du broyage varie.

Tableau 32 : préparation de la pastille : première étape (broyage)

N°Exp.	« Pré-mélange »	Remplissage (%)	[CL20] (%)	Masse CL20 (mg)	Masse KBr (mg)	Durée broy (min)	Fréquence Broy (Hz)
1	1	17	6	68.85	1078.65	15	20
6		17	6			5	20
2	2	17	10	114.75	1032.75	15	10
3		17	10			5	10
7	4	13	10	87.75	789.75	15	20
4		13	10			5	20
5	3	13	6	52.65	824.85	15	10
8		13	6			5	10
9-11	5	15	8	80.80	929.20	10	15

La seconde étape consiste à diluer 10 fois chaque pré-mélange dans le KBr (100mg de pré-mélange+900mg de KBr), toujours à l'aide du vibro-broyeur (tableau 33).

Tableau 33 : préparation de la pastille : deuxième étape (dilution-mélangeage)

N°Exp.	« Pré-mélange »	[CL20] (%)	Masse (1) (mg)	Masse KBr (mg)	Durée mélang. (min)	Fréquence mélang. (Hz)
1	1	0.6	100	900	15	20
6		0.6			5	20
2	2	1.0	100	900	5	10
3		1.0			15	10
7	4	1.0	100	900	5	20
4		1.0			15	20
5	3	0.6	100	900	15	10
8		0.6			5	10
9-11	5	0.8	100	900	10	15

9.2.3.1 - Saisie des spectres

Pour obtenir le maximum de précision sur les absorbances dues aux seuls analytes présents dans le mélange, le « background » (bruit de fond) doit être le plus représentatif possible des conditions de préparation ; on l'obtient à partir de la pastille préparée avec le KBr pur ayant subi les mêmes cycles de broyage que l'échantillon analysé.

Compte tenu de la taille de la pastille, il suffit de saisir un seul spectre par pastille. Par contre *il est toujours nécessaire de répéter la mesure d'un échantillon à analyser* : nous avons choisi d'effectuer *5 répétitions* (5 pastilles). Ces 5 prélèvements étant issus du même bol, le background correspondant est *commun* aux 5 mesures.

Les pastilles sont réalisées à partir de 45mg de la dispersion finale dans le KBr. Les spectres ont été saisis en transmission de 4000 à 400cm^{-1} à l'aide d'un porte pastille de 5mm, avec les paramètres de réglage du spectromètre suivants : gain = 1 ; résolution = 4cm^{-1} ; nombre de scans = 100.

9.2.3.2 - Résultats expérimentaux

Les résultats obtenus ont été regroupés dans différents tableaux (tableaux 34, 35 et annexe 8). Les granulométries sont mesurées une heure après le broyage. La granulométrie du KBr de départ varie de *366 à 286 μm* . La granulométrie du CL20 utilisé (lot G4) pour la réalisation de ce plan est de *96 μm* .

Tableau 34 : préparation de la pastille : première étape (broyage)

N°Exp.	« Pré-mélange »	[CL20] théo (%)	[CL20] exp (%)	Bol 1		Bol 2	
				Masse CL20 exp (mg)	Masse KBr exp (mg)	Masse KBr théo pour BG* (mg)	Masse KBr exp pour BG* (mg)
1	1	6	6.0	68.9	1078.6	1147.5	1147.5
6		6	6.0	68.9	1078.6	1147.5	1147.6
2	2	10	10.0	114.8	1032.7	1147.5	1147.5
3		10	10.0	114.7	1032.7	1147.5	1147.5
7	4	10	10.0	87.8	789.8	877.5	877.5
4		10	10.0	87.8	789.7	877.5	877.5
5	3	6	6.0	52.6	824.9	877.5	877.5
8		6	6.0	52.7	824.8	877.5	877.4
9	5	8	8.0	80.8	929.3	1010.0	1010.0
10		8	8.0	80.8	929.3	1010.0	1009.9
11		8	8.0	80.8	929.3	1010.0	1009.9

BG = background

Tableau 35 : préparation de la pastille : deuxième étape (mélangeage)

N°Exp.	« Pré-mélange »	[CL20] théo (%)	[CL20] exp (%)	Bol 1		Bol 2	
				Masse (1) exp (mg)	Masse KBr exp (mg)	Masse KBr (1) exp pour BG (mg)	Masse KBr ajout exp pour BG (mg)
1	1	0.6	0.6	99.9	900.1	100.0	900.0
6		0.6	0.6	100.0	900.0	100.0	900.0
2	2	1.0	1.0	99.9	900.0	99.9	900.0
3		1.0	1.0	99.9	900.0	100.1	900.0
7	4	1.0	1.0	100.0	900.0	100.0	900.0
4		1.0	1.0	100.0	900.0	100.1	900.1
5	3	0.6	0.6	99.9	900.0	100.1	900.1
8		0.6	0.6	100.0	899.9	100.1	900.0
9	5	0.8	0.8	100.0	900.0	100.1	900.1
10		0.8	0.8	100.0	899.9	99.9	900.0
11		0.8	0.8	99.9	900.1	99.9	900.1

9.2.4 - Exploitation des Résultats

9.2.4.1 - Granulométrie après broyage (Y_1)

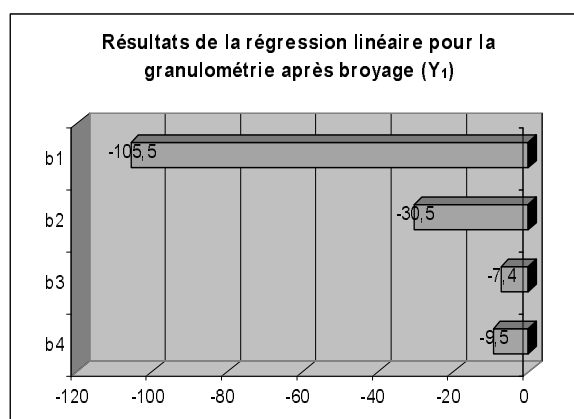
Rappelons que pour un plan de criblage, le modèle postulé est un polynôme du premier degré. Cette granulométrie n'étant fonction que des 4 premiers facteurs, ce polynôme s'écrit :

$$Y_1 = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$

Les coefficients de ce modèle sont estimés par régression linéaire à partir des variables codées et des seuls essais de la matrice d'Hadamard (essais 1 à 8), ils sont reportés numériquement et sous forme graphique dans le tableau 36. Les variables codées étant sans dimension, les effets des facteurs sont alors indépendants des unités choisies et *directement comparables* entre eux.

Tableau 36 : résultats de la régression linéaire pour la granulométrie après broyage (Y_1)

Coefficients	Valeur
b_0 (constante)	220.1
b_1 (fréquence broy.)	-105.5
b_2 (durée broy.)	-30.5
b_3 (taux remplissage)	-7.4
b_4 ([CL20])	-9.5



L'examen de ces résultats montre que la « granulométrie après broyage » dépend essentiellement de la fréquence de broyage (facteur X_1) et dans une moindre mesure de la durée du broyage (X_2). L'effet de la fréquence de broyage $b_1 = -105,5$ signifie qu'en moyenne, la granulométrie du broyat diminue de $2 \times 105,5 = 211 \mu\text{m}$ quand la fréquence de broyage passe de 10 à 20Hz. De la même manière, la granulométrie du broyat diminue de $2 \times 30,5 = 61 \mu\text{m}$ quand la durée de broyage passe de 5 à 15min. Le taux de remplissage du bol (X_3) et la concentration en CL20 (X_4) ont des effets négligeables dans leur domaine de variation.

L'objectif recherché étant de minimiser la granulométrie (ce qui va de pair avec une maximisation de l'homogénéité du mélange) il faut donc choisir une **fréquence de broyage de 20Hz** et une **durée de broyage de 15min**.

Information apportée par les points au centre (essais 9-11)

Valeur moyenne de la réponse au centre du domaine expérimental = $320,7 \mu\text{m}$;

Variance expérimentale estimée à partir des répétitions = 754,3 soit un écart-type de $27,5 \mu\text{m}$.

Courbure de la surface de réponse = $320,7 - 220,1 = 100,6$. L'écart entre la valeur prédite par le modèle au centre du domaine expérimental (b_0) et la valeur expérimentale étant de l'ordre de 4 écarts-types, il existe une courbure de la surface de réponse.

Le but de l'étude n'étant pas d'obtenir un modèle prédictif mais simplement de « peser » les facteurs, il n'est pas nécessaire de poursuivre l'étude en postulant un modèle quadratique.

9.2.4.2 - Granulométrie après mélangeage (Y_2)

Pour cette réponse qui concerne le deuxième stade de la préparation de la pastille, *la granulométrie G_1 obtenue à la fin du premier stade de broyage devient un facteur d'entrée*.

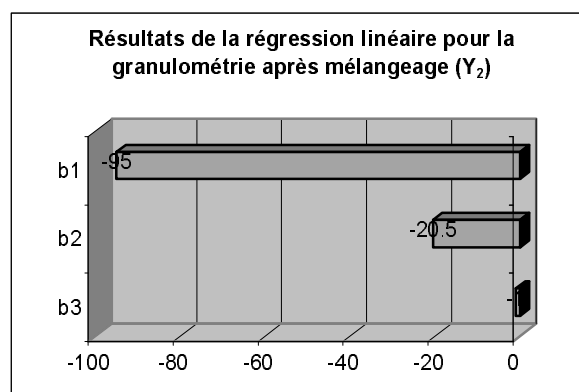
Le modèle postulé devient :

$$Y_2 = b_0 + b_5 X_5 + b_6 X_6 + b_7 G_1$$

Comme pour la réponse Y_1 , les coefficients de ce modèle sont estimés à partir des seuls essais de la matrice d'Hadamard (essais 1 à 8), ils sont reportés numériquement et sous forme graphique dans le tableau 37.

Tableau 37 : résultats de la régression linéaire pour la granulométrie après mélangeage (Y_2)

Coefficients	Valeur
b_0 (constante)	224.5
b_1 (fréquence mélangeage)	-95
b_2 (durée mélangeage)	-20.5
b_3 (granulo. broyat)	-1



Comme pour la granulométrie après broyage, la granulométrie après mélangeage dépend essentiellement de la fréquence (facteur X_5) et de la durée (X_6). En moyenne, la granulométrie finale diminue de $2 \times 95 = 190 \mu\text{m}$ quand la fréquence passe de 10 à 20Hz et de $2 \times 20,5 = 41 \mu\text{m}$ quand la durée passe de 5 à 15min. La granulométrie d'entrée n'a pas d'influence sur la granulométrie finale. Ce résultat montre que, du point de vue granulométrique, cette phase de mélangeage n'est en réalité qu'une phase de broyage du KBr neuf qui représente 90% de la charge du bol du vibro-broyeur, le broyat issu de la première phase n'en représentant que 10%.

L'objectif recherché étant le même que celui de l'étape précédente, il faut choisir une **fréquence de mélangeage de 20Hz** et une **durée de mélangeage de 15min**.

Information apportée par les points au centre (essais 9-11)

Valeur moyenne de la réponse au centre du domaine expérimental = $263,7 \mu\text{m}$;

Variance expérimentale estimée à partir des répétitions = 456,3 soit un écart-type de $21,4 \mu\text{m}$.

Courbure de la surface de réponse = $263,7 - 224,5 = 39,2$. L'écart entre la valeur prédite par le modèle au centre du domaine expérimental (b_0) et la valeur expérimentale étant de l'ordre de 1,5 écart-type, il n'y a pas de courbure de la surface de réponse.

9.2.4.3 - Epaisseur de la pastille et énergie du faisceau IR absorbée (Y_3 et Y_4)

Pour ces réponses qui concernent les caractéristiques de la pastille obtenue in fine, l'examen du tableau 31 et de l'annexe 8 montre qu'elles ne varient pas de manière significative d'un essai à l'autre.

Que ce soit pour l'épaisseur ou pour l'énergie absorbée, la dispersion des 8 essais de la matrice d'expérience est identique à celle obtenue par répétition du point au centre : *les réponses sont insensibles aux variations des facteurs* dans le domaine étudié.

Ce résultat signifie que *la méthode d'échantillonnage utilisant le vibro-broyeur* dans les conditions définies ci-dessus, *est une méthode robuste*.

9.2.5 - Variabilité spectrale

Nous avons réalisé les 11 spectres obtenus à partir des pastilles issues de l'expérimentation précédente. Pour une concentration donnée en CL20, on devrait observer des spectres rigoureusement superposables, or comme c'était le cas pour les spectres de notre première étude d'analyse quantitative, nous n'observons pas un bon retour à la ligne de base. Ce phénomène ne peut être expliqué que par une *stabilité photométrique insuffisante de notre spectromètre*.

Un exemple de spectre obtenu est reporté figure 64 (mélange 4). Les densités optiques obtenues pour les 11 mélanges, pour les trois fréquences les plus intenses du spectre, sont rassemblées dans le tableau 38.

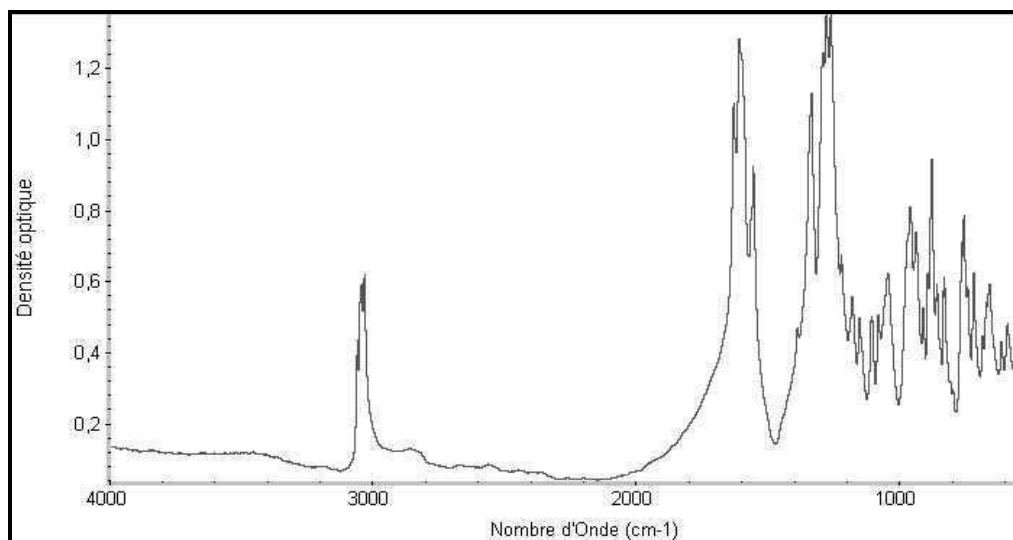


Figure 64 : spectre IR du mélange 4

Toutes les densités optiques étant $<1,5$, nous pouvons appliquer la loi de BEER LAMBERT quel que soit le mélange retenu pour la réalisation de la nouvelle matrice d'étalonnage.

Tableau 38 : résultats des calculs de densités optiques pour les trois bandes principales observées sur le spectre

Spectres du mélange N°	bandes		
	3039.26 cm ⁻¹	1608.34 cm ⁻¹	1278.57 cm ⁻¹
1	0.36	0.82	0.82
2	0.14	0.17	0.16
3	0.18	0.23	0.21
4	0.53	1.21	1.20
5	0.18	0.26	0.24
6	0.24	0.45	0.44
7	0.55	1.24	1.28
8	0.10	0.13	0.12
9	0.19	0.31	0.29
10	0.20	0.32	0.30
11	0.22	0.36	0.35

9.2.6 - Conclusion

Les fréquences de broyage et de mélangeage sont fortement influentes sur les granulométries obtenues, les durées de broyage et de mélangeage l'étant plus faiblement. Le taux de remplissage, la concentration en CL20, la pression de pastillage et la durée de pression n'ont que peu d'influence dans les domaines de variation choisis pour notre étude.

Pour la réalisation des expériences du nouveau plan d'étalonnage, pour chacune des étapes, il faudra utiliser le vibro-broyeur à 20Hz pendant 15min. Pour les niveaux des facteurs peu influents, l'opérateur les fixera en fonction de leur facilité de mise en œuvre pourvu que les valeurs choisies appartiennent au domaine expérimental étudié et qu'elles soient fixées une fois pour toute. Le protocole retenu est schématisé par la figure 65.

9.3 - Matrice d'étalonnage et plan de mélanges avec contraintes

Rappelons que, les contraintes sur les concentrations des constituants du mélange des formes polymorphes du CL20 se matérialisent par un ensemble de troncatures du domaine expérimental qui conduisent à un polyèdre nommé «polyèdre des contraintes». Dans l'espace des mélanges nos contraintes définissent un polyèdre caractérisé par 16 sommets, 32 arêtes, 24 faces de dimension 2, 8 faces de dimension 3 et un barycentre, soit un ensemble de 81 points informatifs. Pour 5 composants il n'est plus possible de représenter le polyèdre des contraintes. Nous disposons également de 16 points tests internes potentiel proposés par le logiciel utilisé NEMROD (centres de gravité des polyèdres délimités par chaque sommet et

les milieux des arêtes qui en sont issues), ce qui avec les 81 points précédemment définis constitue un ensemble de 97 points informatifs. C'est parmi ces points informatifs que l'on choisira les points d'étalonnage

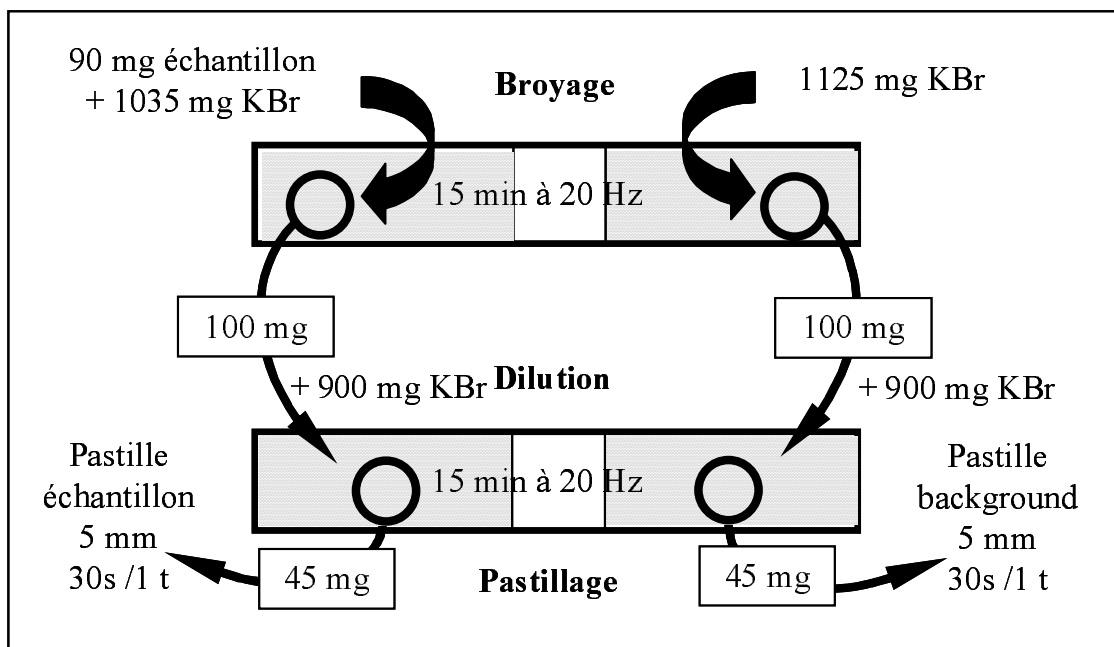


Figure 65 : protocole d'échantillonnage

9.4 - Construction de la Matrice d'étalonnage à partir des points informatifs

Comme nous disposons d'un ensemble de 97 mélanges informatifs candidats à la construction de la matrice d'étalonnage, nous devons sélectionner parmi eux les meilleurs points. Compte tenu de la loi de BEER LAMBERT, le modèle d'étalonnage est un modèle linéaire. Pour un tel modèle, les points les plus informatifs sont les sommets du polyèdre des contraintes, ici les 16 sommets. Ces points étant par nature à la périphérie du domaine expérimental étudié, il est nécessaire de choisir des points à l'intérieur du domaine pour valider le modèle (et/ou pour l'affiner) : nous avons retenu le barycentre et les 16 points tests internes définis § 9.3. Tous calculs faits, *il en est résulté une matrice d'étalonnage constituée de 33 mélanges étalons* informatifs dont les coordonnées sont rassemblées dans le tableau 39.

Nous avons également retenu les points 34 à 41 (centre des faces de dimension 3) qui ne sont pas indispensables, mais qui serviront pour la validation du modèle.

Tableau 39 : coordonnées des points informatifs d'étalonnage et de validation

		N° Exp	Alpha	Bêta	Gamma	Pentanitré	Epsilon
Sommets	C A L I B R A T I O N	1	0.003	0.003	0.003	0.003	0.988
		2	0.003	0.02	0.003	0.003	0.971
		3	0.003	0.003	0.02	0.003	0.971
		4	0.003	0.02	0.02	0.003	0.954
		5	0.003	0.003	0.003	0.02	0.971
		6	0.003	0.02	0.003	0.02	0.954
		7	0.003	0.003	0.02	0.02	0.954
		8	0.003	0.02	0.02	0.02	0.937
		9	0.04	0.003	0.003	0.003	0.951
		10	0.04	0.02	0.003	0.003	0.934
		11	0.04	0.003	0.02	0.003	0.934
		12	0.04	0.02	0.02	0.003	0.917
		13	0.04	0.003	0.003	0.02	0.934
		14	0.04	0.02	0.003	0.02	0.917
		15	0.04	0.003	0.02	0.02	0.917
		16	0.04	0.02	0.02	0.02	0.9
Centre de gravité	C	17	0.0215	0.0115	0.0115	0.0115	0.944
Points tests internes	A L I B R A T I O N	18	0.0123	0.0073	0.0073	0.0073	0.966
		19	0.0123	0.0158	0.0073	0.0073	0.9575
		20	0.0123	0.0073	0.0158	0.0073	0.9575
		21	0.0123	0.0158	0.0158	0.0073	0.949
		22	0.0123	0.0073	0.0073	0.0158	0.9575
		23	0.0123	0.0158	0.0073	0.0158	0.949
		24	0.0123	0.0073	0.0158	0.0158	0.949
		25	0.0308	0.0158	0.0158	0.0158	0.9405
		26	0.0308	0.0073	0.0073	0.0073	0.9475
		27	0.0308	0.0158	0.0073	0.0073	0.939
		28	0.0308	0.0073	0.0158	0.0073	0.939
		29	0.0308	0.0158	0.0158	0.0073	0.9305
		30	0.0308	0.0073	0.0073	0.0158	0.939
		31	0.0308	0.0158	0.0073	0.0158	0.9305
		32	0.0308	0.0073	0.0158	0.0158	0.9305
		33	0.0308	0.0158	0.0158	0.0158	0.922
Faces de dimension 3	V A L I D A T I O N	34	0.003	0.0115	0.0115	0.0115	0.9625
		35	0.0215	0.003	0.0115	0.0115	0.9525
		36	0.0215	0.0115	0.003	0.0115	0.9525
		37	0.0215	0.0115	0.0115	0.003	0.9525
		38	0.04	0.0115	0.0115	0.0115	0.9255
		39	0.0215	0.02	0.0115	0.0115	0.9355
		40	0.0215	0.0115	0.02	0.0115	0.9355
		41	0.0215	0.0115	0.0115	0.02	0.9355

9.5 - Réponses expérimentales

Comme pour le premier plan de mélange, les réponses expérimentales sont les absorbances du spectre infrarouge. Il y a autant de réponses individuelles que de fréquences spectrales étudiées.

Là encore, l'exploitation numérique peut nécessiter un traitement mathématique préliminaire du spectre brut enregistré par le spectromètre : correction de la dérive de la ligne de base et/ou utilisation de la dérivée première du spectre.

9.6 - Expérimentation du second plan d'expériences

9.6.1 - Réalisation des mélanges étalons

Nous avons réalisé 41 mélanges informatifs : les 33 points de la matrice d'étalonnage et les 8 points tests de validation (centre de faces de dimension 3). Comme pour le plan n°1, ce sont des mélanges synthétiques obtenus à partir d'échantillons des formes polymorphes pures du CL20, en ajoutant cette fois l'impureté pentanitree. Ces formes polymorphes ont été synthétisées au laboratoire de synthèse du CRB à partir du lot de CL20 réf G1 titrant 99,48 % (avec en particulier 0,3% d'impureté pentanitree). Les méthodes de synthèse des formes polymorphes appartiennent au savoir-faire du groupe SNPE.

Les références et les caractéristiques des échantillons des formes polymorphes du CL20 utilisés pour effectuer les mélanges étalons sont rassemblées dans le tableau 40. Leur pureté chimique a été déterminée par HPLC (méthode de normalisation interne corrigée) ; les teneurs observées étant très proche de 100%, les différentes formes polymorphes sont considérées comme «pures». Cette pureté monomorphique a été contrôlée par IR et diffraction X (travaux effectués en collaboration avec Mr TOUPET [85]) (annexe 9) (contrôle de conformité du spectre IR de chaque forme synthétisée par rapport aux spectres de référence issus de travaux antérieurs). Le spectre IR de chacune des formes polymorphes est donné figure 66.

Tableau 40 : références et caractéristiques des substances étalons

Références	GRANULO -METRIE (μm)	TAUX		DENSITE ($\text{g}\cdot\text{cm}^{-3}$) (référence)	PURETE Chimique	CHN		
		Eau (%)	Solvant (%)			Azote (%)	Carbone (%)	Hydro. (%)
E1 (epsilon)	53.0	0.03	AE : 0.08 T : 0.23	2.03 (2.04)	99.48 (PN : 0.35)	39.3 (38.4)	17.0 (16.5)	1.51 (1.38)
G1 (gamma)	77.0	0.07		1.91 (1.92)	99.48 (PN : 0.31)	39.2 (38.4)	16.7 (16.5)	1.47 (1.38)
A1 (alpha)	152.0	1.19	A : 0.92	1.88-1.87 (1.97)	99.48 (PN : 0.31)	38.5 (38.4)	17.2 (16.5)	1.65 (1.38)
B1 (bêta)	5.6	0.07	C : 0.43 THF : 0.33	1.99 (1.98)	99.49 (PN : 0.31)	38.9 (38.4)	17.3 (16.5)	1.58 (1.38)

AE : Acétate d'éthyle T : Toluène A : Acétone THF : Tétrahydrofurane C : Chloroforme

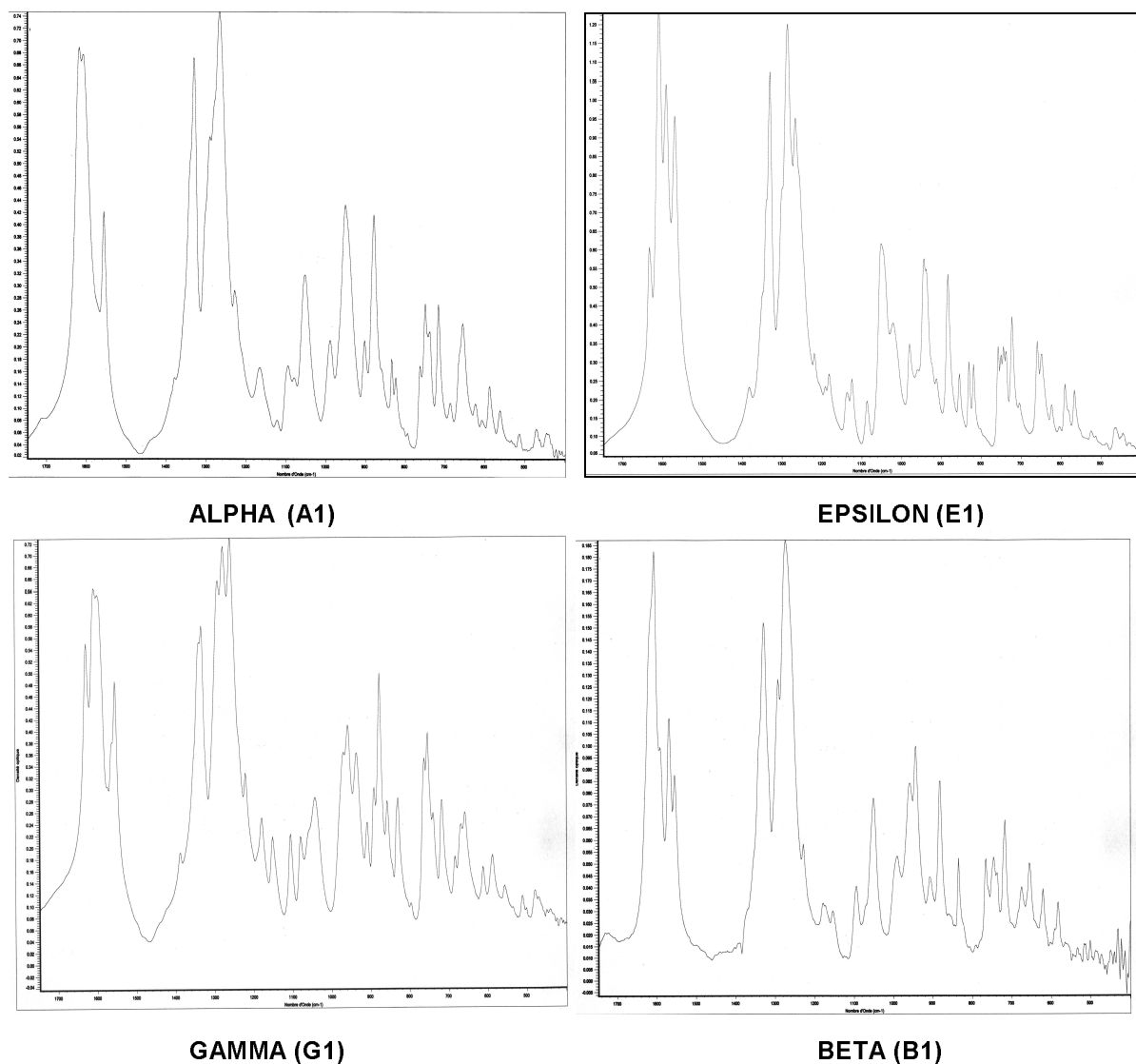


Figure 66 : spectres IR des 4 formes polymorphes du CL20 utilisés pour la réalisation du plan n°2
Zone spectrale 1750-400cm⁻¹

9.6.2 - Techniques d'échantillonnage et de saisie des spectres IR

9.6.2.1 - Appareillage et Logiciels d'exploitation des données

Seule quelques différences sont à noter par rapport à l'appareillage utilisé pour le premier plan de mélanges.

Pesées : balance de précision METTLER TOLEDO modèle AG245 et microbalance METTLER TOLEDO modèle M3.

Matériel pour la réalisation des pastilles : utilisation d'un vibro-broyeur de marque RETSCH type MM200 (2 bols en agate de 5 ml contenant chacun 2 billes en agate de 7 mm de diamètre) et d'une pastilleuse 5mm de marque SPECAC.

9.6.2.2 - Echantillonnage

Les mélanges synthétiques sont obtenus par pesée des formes polymorphes pures et de l'impureté pentanitrée conformément aux compositions données par la matrice d'étalonnage et par dispersion dans du KBr cristallisé anhydre (sachet individuel de 500mg, conservé à l'étuve, de granulométrie et de qualité adaptées à la spectrométrie IR). Afin de réduire la variabilité lors de la fabrication des pastilles, chaque mélange est réalisé selon le protocole de préparation de la pastille déterminé à partir du plan de « criblage » précédent (§9.2.6) :

- à chaque étape 2 pastilles sont préparées dans les mêmes conditions : l'une contenant le mélange CL20+PN+KBr, l'autre ne contenant que du KBr pur pour la réalisation du Background. Le mélange et le KBr pur auront ainsi des caractéristiques de granulométrie et d'hygroscopie aussi proches que possible et les pastilles obtenues à partir de chacun d'eux auront des pouvoirs diffusants voisins.
- introduire dans le premier bol du vibro-broyeur le mélange suivant : 90mg de CL20 ($\alpha+\beta+\gamma+\epsilon$)+PN et 1035mg de KBr. Dans le deuxième bol mettre 1125mg de KBr pur. Régler la fréquence sur 20Hz et la durée sur 15min. Une fois le broyage terminé, et pour pallier l'hygroscopie du KBr, mettre les produits broyés sous vide au dessiccateur (sous P₂O₅) pendant 24 heures.
- prendre 100mg de produits broyés (mélange et KBr pur), ajouter 900mg de KBr, puis réitérer le processus précédent (fréquence 20Hz et 15min de broyage-mélangeage, 24h de dessiccation sous vide et P₂O₅).
- les pastilles proprement dites sont réalisées à partir de 45mg de cette dispersion finale dans le KBr (sous vide, à l'aide d'une pastilleuse et d'une presse hydraulique (pression : 1 tonne pendant 30s)).
- *pour chaque mélange étalon, en raison de la dispersion spectrale observée, on fabrique 5 pastilles* selon ce protocole. Comme ces 5 prélèvements sont issus du même bol, le background correspondant est *commun* aux 5 mesures. Une pastille de KBr pur est effectuée par mélange.

9.6.2.3 - Saisie des spectres

Que ce soit pour le Background ou pour les échantillons, les spectres ont été saisis en transmission de 4000 à 400cm⁻¹ à l'aide d'un porte pastille de 5mm, avec les paramètres de réglage suivants du spectromètre : gain = 1 ; résolution = 4cm⁻¹ ; nombre de scans = 100.

Nous disposons d'un ensemble de **5 spectres par étalon** qui peuvent être considérés comme une **répétition des points d'étalonnage**.

9.7 - Exploitation des résultats

La méthode de travail appliquée est identique à celle utilisée pour le premier plan d'expérience c'est à dire : l'information spectrale sera **exportée** pour pouvoir être exploitée de manière « explicite » avec tous les outils nécessaires en particulier pour identifier les zones spectrales les plus informatives et les modèles les plus performants. Les paramètres optimaux seront ensuite « **importés** » dans le logiciel TURBO QUANT pour **optimiser** la méthode d'analyse qui sera utilisée **en routine**, directement à partir du spectromètre servant à saisir les spectres.

9.7.1 - Méthodes d'analyse développées en « Externe »

9.7.1.1 - Exportation et mise en forme des données spectrales

Les conditions opératoires n'ayant pas changées, nous nous sommes également retrouvé avec un tableau de données similaire à celui du premier plan. Pour l'obtenir nous avons appliqué les mêmes étapes que pour le premier plan (mise en forme préliminaire, réduction du nombre de fréquences informatives). La seule différence est le nombre de spectre qui n'est plus de 212 mais de 208, soit un tableau de 239 descripteurs (colonnes) et 208 individus (lignes).

Contrairement au premier plan, les spectres obtenus présentent tous une forte dérive de la ligne de base. C'est pourquoi nous avons effectué plusieurs essais de traitements, d'où la réalisation de 4 tableaux EXCEL construit comme précédemment :

- un tableau contenant les absorbances brutes ;
- un tableau ne contenant que les données spectrales ayant subi une correction de ligne de base automatique dans la zone de fréquences allant de 1800 à 600cm⁻¹ ;
- un tableau contenant les absorbances transformées en variables centrées ;
- un dernier tableau contenant les dérivées premières des spectres.

Chaque tableau est ensuite transféré dans le logiciel SIMCA et sert de base de données pour élaborer les modèles à l'aide de la régression PLS (algorithmes PLS 1 et 2) [51].

9.7.1.2 - Exploitation des données spectrales

Identification des zones spectrales informatives

La démarche suivie pour identifier les zones réellement informatives est la même que celle du premier plan d'expériences (utilisation de la variance maximale et/ou de la corrélation maximale) avec suppression des zones de fréquences jugées comme étant non informatives à partir des données du premier plan (suppression des mêmes zones que pour le modèle MS₉ du premier plan).

Choix des mélanges de calibration et des mélanges de validation

La matrice de calibration (d'étalonnage) est constituée normalement des mélanges de 1 à 33, les points 34 à 41 représentant les points de validation. Les points aberrants (outliers) s'écartant des modèles créés à l'aide de la régression PLS seront mentionnés au cas par cas.

Les concentrations théoriques n'étant pas toujours réalisables, la matrice des concentrations réelles est connue à partir des données reportées dans le cahier de l'opérateur. Ce sont en toute rigueur les concentrations réelles qui doivent être employées dans les calculs.

Modèles de calibration

Les essais de modèles effectués à partir des variables centrées ou de la correction de la ligne de base entre 1800 et 600cm⁻¹ n'ont pas donné de résultats satisfaisants (valeur de R² (X et Y) cum et de Q² cum faibles quand elles existent, les prédictions obtenues sont assez éloignées de la réalité).

D'autres modèles furent élaborés, soit à partir des données brutes sans aucune suppression, soit en prenant en compte simultanément 2, 3 ou 4 composés, sans que nous obtenions de résultats acceptables.

Finalement les modèles retenus sont élaborés à partir des absorbances brutes et des dérivées premières de tous les spectres des mélanges 1 à 33. En fonction des fréquences et des mélanges retenus pour le calcul, 10 modèles ont été définis (6 pour les absorbances et 4 pour les dérivées premières). Ces modèles sont référencés de MS₁₀ à MS₁₅ pour les absorbances et de MS₁₆ à MS₁₉ pour les dérivées premières. Les modèles MS_{11a15} et MS_{17a19} sont construits à partir de l'algorithme PLS1 (un seul composé intervient dans la régression). Les 2 modèles restant (MS₁₀ et 16) sont élaborés à partir de l'algorithme PLS2 (l'ensemble des composés interviennent dans la régression).

Modèles élaborés à partir des absorbances brutes :

Zones de fréquences supprimées : 399.19 à 553.47cm⁻¹, 576.61cm⁻¹, 599.75 à 615.18cm⁻¹, 669.18 à 692.32cm⁻¹, 761.74 à 808.03cm⁻¹, 838.88cm⁻¹, 892.88cm⁻¹, 1070.30 à 1108.87cm⁻¹, 1147.44cm⁻¹, 1155.15cm⁻¹, 1394.28 à 1509.99cm⁻¹, 1695.12 à 1795.40cm⁻¹.

Remarque : la réunion des zones de corrélation maximale entre absorbances et concentrations, n'a pas permis d'améliorer le modèle par rapport à seule la variance maximale entre les spectres.

Modèles	Réf. fichiers	Conditions de modélisation (concentrations théoriques)
MS ₁₀	M ₃ : M ₂ variance maximale entre les spectres, avec demande de composantes suppl.	Suppression des fréquences ayant les plus petites variances <0.005
MS ₁₁	M ₅ : M ₂ avec Y=epsilon	Suppression des fréquences ayant les plus petites variances <0.005
MS ₁₂	M ₆ : M ₂ avec Y=alpha	Suppression des fréquences ayant les plus petites variances <0.005
MS ₁₃	M ₁₀ : M ₂ avec Y=bêta	Suppression des fréquences ayant les plus petites variances <0.005
MS ₁₄	M ₁₁ : M ₂ avec Y=gamma	Suppression des fréquences ayant les plus petites variances <0.005
MS ₁₅	M ₁₂ : M ₂ avec Y=PN	Suppression des fréquences ayant les plus petites variances <0.005

Modèles élaborés à partir des dérivées premières :

Zones de fréquences supprimées : 476.33 à 553.47cm⁻¹, 584.33cm⁻¹, 599.75 à 630.61cm⁻¹, 653.75cm⁻¹, 676.89 à 707.75cm⁻¹, 730.89cm⁻¹, 738.60cm⁻¹, 754.03cm⁻¹, 769.46 à 808.03cm⁻¹, 854.31cm⁻¹, 862.02cm⁻¹, 900.59cm⁻¹, 916.02cm⁻¹, 923.74cm⁻¹, 954.59 à 970.02cm⁻¹, 993.16cm⁻¹, 1016.30 à 1031.73cm⁻¹, 1070.30 à 1085.73cm⁻¹, 1101.15cm⁻¹, 1108.87cm⁻¹, 1124.30cm⁻¹, 1132.01cm⁻¹, 1147.44 à 1170.58cm⁻¹, 1186.01 à 1232.29cm⁻¹, 1371.14 à 1540.85cm⁻¹, 1648.84 à 3004.55cm⁻¹, 3058.55 à 3197.40cm⁻¹.

Modèles	Réf. fichiers	Conditions de modélisation (concentrations théoriques)
MS ₁₆	M ₆ : corrélations - variance zone commune	Suppression des fréquences communes
MS ₁₇	M ₁₅ : M ₆ avec Y = alpha	Suppression des fréquences communes
MS ₁₈	M ₁₆ : M ₆ avec Y = bêta	Suppression des fréquences communes
MS ₁₉	M ₁₇ : M ₆ avec Y = epsilon	Suppression des fréquences communes

Nombre de composantes retenues pour chaque modèle

Le critère de détermination du nombre optimal de composantes à utiliser dans le modèle d'étalonnage correspond à la minimisation de la somme des carrés des résidus par validation croisée du modèle testé (PRESS, Q² ou test équivalent) (tableaux 41 et 42).

Tableau 41 : modèles élaborés à partir des absorbances brutes

Modèles	Nb de composantes	Modèles	Nb de composantes
MS ₁₀	8	MS ₁₃	9
MS ₁₁	12	MS ₁₄	8
MS ₁₂	8	MS ₁₅	8

Tableau 42 : modèles élaborés à partir des dérivées premières

Modèles	Nb de composantes	Modèles	Nb de composantes
MS ₁₆	12	MS ₁₈	4
MS ₁₇	8	MS ₁₉	4

Examen des résidus

Parmi les outils d'analyse de la qualité de l'étalonnage que nous avons utilisé l'examen de la distribution des résidus via la droite de Henry présente l'avantage de la simplicité (méthode purement graphique) [84].

9.7.1.3 - Analyse des modèles retenus

A partir des absorbances : les tableaux 43 et 44 ci dessous rassemblent les résultats de la validation à partir des points 34 à 41. Comme nous disposons de 5 spectres pour un mélange donné, nous pouvons rendre la valeur moyenne de sa concentration estimée et un écart type représentant la dispersion de ces estimations individuelles.

Tableau 43 : validation des modèles à partir des absorbances

Méthodes	Pts valid.	Compositions réelles					Compositions estimées par les Méthodes de dosage									
		ϵ (%)	α (%)	β (%)	γ (%)	PN (%)	ϵ (%)	σ	α (%)	σ	β (%)	σ	γ (%)	σ	PN (%)	σ
MS ₁₀	34	96.3	0.3	1.1	1.1	1.2	96.2	0.3	1.1	0.4	0.8	0.1	<1.0	0.4	2.0	0.2
	35	95.3	2.1	0.3	1.1	1.1	95.3	0.3	2.6	0.3	0.3	0.1	0.3	0.1	1.4	0.1
	36	95.3	2.1	1.1	0.3	1.2	94.6	0.4	2.6	0.3	1.2	0.1	0.2	0.2	1.4	0.1
	37	95.2	2.1	1.1	1.2	0.3	94.6	0.2	2.4	0.2	1.1	0.2	1.1	0.1	0.8	0.2
	38	92.6	4.0	1.1	1.1	1.2	93.1	0.4	4.0	0.3	1.3	0.2	0.6	0.2	1.2	0.1
	39	93.6	2.1	2.0	1.1	1.2	93.5	0.5	2.3	0.4	1.6	0.2	1.4	0.1	1.2	0.1
	40	93.6	2.1	1.1	2.0	1.2	94.2	0.6	2.6	0.2	1.0	0.2	0.7	0.4	1.5	0.2
	41	93.6	2.1	1.1	1.1	2.0	94.1	0.3	2.4	0.3	1.1	0.2	0.8	0.2	1.7	0.1
MS _{11a14}	34	96.3	0.3	1.1	1.1	1.2	96.1	0.2	0.6	0.0	1.0	0.2	0.3	0.2	2.0	0.2
	35	95.3	2.1	0.3	1.1	1.1	95.1	0.2	2.4	0.2	0.5	0.2	0.7	0.1	1.4	0.1
	36	95.3	2.1	1.1	0.3	1.2	94.5	0.4	2.4	0.3	1.3	0.1	0.4	0.1	1.4	0.1
	37	95.2	2.1	1.1	1.2	0.3	94.5	0.2	2.5	0.1	1.1	0.2	1.2	0.1	0.9	0.2
	38	92.6	4.0	1.1	1.1	1.2	92.8	0.4	4.0	0.3	1.4	0.2	0.9	0.1	1.2	0.1
	39	93.6	2.1	2.0	1.1	1.2	93.5	0.5	2.4	0.3	1.5	0.2	1.5	0.2	1.3	0.1
	40	93.6	2.1	1.1	2.0	1.2	94.0	0.5	2.6	0.3	0.9	0.2	1.0	0.3	1.7	0.2
	41	93.6	2.1	1.1	1.1	2.0	94.0	0.4	2.2	0.4	1.1	0.2	0.9	0.2	1.7	0.1

Rappelons que compte tenu qu'une régression PLS passe par des détours mathématiques complexes, il n'est pas possible de donner ici, de manière accessible, un véritable intervalle de confiance sur une teneur prédite. Néanmoins, nous pouvons examiner deux indicateurs statistiques simples : le coefficient de variation qui rapporte l'écart type d'une mesure à sa valeur moyenne et l'erreur relative sur une teneur en rapportant le résidu à la valeur théorique de la teneur. Ces indicateurs sont rassemblés dans le tableau 44. Une première analyse montre que ces deux modèles peuvent servir de base à un développement de méthodes avec TURBO QUANT.

Tableau 44 : coefficients de variation et erreurs relatives à partir des absorbances

Méthodes	Pts. valid	Teneurs théoriques					Forme ϵ		Forme α		Forme β		Forme γ		Forme PN	
		ϵ (%)	α (%)	β (%)	γ (%)	PN (%)	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %
MS ₁₀	34	96.3	0.3	1.1	1.1	1.2	0.3	0.2	34.3	276.3	16.5	27.9	<1.0	108.1	10.1	71.2
	35	95.3	2.1	0.3	1.1	1.1	0.3	0.0	10.7	23.7	19.5	11.0	39.6	72.8	5.0	21.6
	36	95.3	2.1	1.1	0.3	1.2	0.5	0.8	10.4	24.0	7.5	5.2	85.0	34.8	8.1	22.8
	37	95.2	2.1	1.1	1.2	0.3	0.2	0.6	10.3	11.5	13.4	2.2	12.5	13.2	28.2	137.9
	38	92.6	4.0	1.1	1.1	1.2	0.4	0.5	8.2	1.0	13.1	12.8	30.6	50.5	9.4	1.5
	39	93.6	2.1	2.0	1.1	1.2	0.5	0.2	18.9	10.4	11.4	18.0	9.7	23.8	7.6	2.8
	40	93.6	2.1	1.1	2.0	1.2	0.6	0.7	6.1	20.7	17.5	8.8	50.3	63.5	11.3	26.0
	41	93.6	2.1	1.1	1.1	2.0	0.3	0.5	12.1	12.6	14.6	2.4	28.7	29.8	6.9	17.0
MS _{11à14}	34	96.3	0.3	1.1	1.1	1.2	0.2	0.2	5.3	115.3	15.9	12.3	78.5	74.3	10.8	73.6
	35	95.3	2.1	0.3	1.1	1.1	0.2	0.3	7.5	15.1	41.3	61.7	21.2	39.4	9.3	24.7
	36	95.3	2.1	1.1	0.3	1.2	0.4	0.9	11.4	12.8	9.4	21.3	19.3	30.8	8.4	22.1
	37	95.2	2.1	1.1	1.2	0.3	0.2	0.8	4.1	19.4	16.2	3.6	11.0	2.0	21.5	164.4
	38	92.6	4.0	1.1	1.1	1.2	0.4	0.2	7.7	0.2	12.9	22.6	14.9	22.2	8.3	7.8
	39	93.6	2.1	2.0	1.1	1.2	0.5	0.2	13.1	11.9	10.1	24.1	15.6	30.4	8.3	15.9
	40	93.6	2.1	1.1	2.0	1.2	0.5	0.4	11.0	22.3	24.5	14.7	28.8	49.4	14.7	43.7
	41	93.6	2.1	1.1	1.1	2.0	0.4	0.4	16.4	6.1	16.1	2.8	22.0	18.7	6.5	14.8

L'examen de ce tableau montre que les modèles retenus (MS₁₀ et MS_{11à14}) présentent des performances variables avec la forme polymorphe considérée :

- la prédiction sur ϵ est très fiable pour les 2 modèles (erreur relative moyenne de 0,4% avec MS₁₀ et MS_{11à14}) ;
- α et β sont assez bien estimés à l'exception des très faibles teneurs (0,3%), l'erreur relative moyenne pour α étant de 14,9% avec MS₁₀ et 12,6% avec MS_{11à14} ; pour β respectivement de 11 et 14,5%) ;
- β est le mieux prédit pour une teneur de 0,3% : valeur moyenne des 5 spectres de 0,3% avec une erreur relative de 11% avec MS₁₀. Dans l'ensemble β est mieux estimé que α .
- γ est le plus mal prédit des 4 formes polymorphes, même en excluant la teneur de 0,3% : erreur relative moyenne de 42,3 avec MS₁₀ et de 27% avec MS_{11à14}. Pour ce

dernier composé les coefficients de variation et les erreurs relatives sont particulièrement élevés quelle qu'en soit la teneur ;

- comme les autres composés, l'impureté pentanitree est mal prédite pour une teneur de 0,3% en présentant, en excluant cette teneur, une erreur relative moyenne de 23,3% avec MS₁₀ et de 28,9% avec MS_{11à14}.

A partir des dérivées premières : les tableaux 45 et 46 ci dessous rassemblent les résultats obtenus pour la validation (mélange 34 à 41). Comme précédemment nous disposons de 5 spectres par mélange d'où le calcul de la valeur moyenne de la concentration estimée et d'un écart type représentant la dispersion. Le tableau 46 donne les résultats des calculs du coefficient de variation et de l'erreur relative sur une teneur.

Tableau 45 : validation des modèles à partir des dérivées premières

Méthodes	Pts valid.	Compositions réelles					Compositions estimées par les Méthodes de dosage									
		ϵ (%)	α (%)	β (%)	γ (%)	PN (%)	ϵ (%) σ		α (%) σ		β (%) σ		γ (%) σ		PN(%) σ	
MS ₁₆	34	96.3	0.3	1.1	1.1	1.2	95.0	1.0	1.3	1.3	1.0	0.3	1.0	0.2	1.7	0.6
	35	95.3	2.1	0.3	1.1	1.1	95.0	0.6	2.1	0.4	0.4	0.2	1.1	0.4	1.5	0.4
	36	95.3	2.1	1.1	0.3	1.2	93.6	0.4	2.8	0.2	1.2	0.1	1.1	0.3	1.2	0.1
	37	95.2	2.1	1.1	1.2	0.3	94.0	0.5	2.4	0.5	1.3	0.2	1.1	0.3	1.3	0.3
	38	92.6	4.0	1.1	1.1	1.2	93.2	0.6	2.8	0.4	1.6	0.2	0.9	0.3	1.5	0.3
	39	93.6	2.1	2.0	1.1	1.2	93.0	0.7	2.8	0.4	1.5	0.2	1.1	0.3	1.6	0.4
	40	93.6	2.1	1.1	2.0	1.2	93.3	0.7	2.6	0.9	1.1	0.2	1.2	0.2	1.7	0.5
	41	93.6	2.1	1.1	1.1	2.0	94.1	0.8	2.0	0.5	1.1	0.3	1.3	0.5	1.6	0.3
MS _{17à19}	34	96.3	0.3	1.1			95.0	0.7	1.3	1.2	0.9	0.3				
	35	95.3	2.1	0.3			95.1	0.6	2.2	0.5	0.4	0.3				
	36	95.3	2.1	1.1			93.6	0.5	2.7	0.2	1.2	0.1				
	37	95.2	2.1	1.1			94.1	0.5	2.3	0.4	1.3	0.2				
	38	92.6	4.0	1.1			93.2	0.7	2.9	0.5	1.5	0.2				
	39	93.6	2.1	2.0			93.1	0.9	2.6	0.4	1.5	0.2				
	40	93.6	2.1	1.1			93.4	0.8	2.7	0.7	1.1	0.2				
	41	93.6	2.1	1.1			94.1	0.4	2.1	0.5	1.1	0.3				

L'examen de ces 2 tableaux (tableaux 45-46) suggère que les erreurs relatives obtenues à partir des dérivées premières sont plus importantes que celles observées à partir des absorbances brutes. Ceci nous incite à ne retenir que les modèles élaborés avec les absorbances brutes comme base de travail pour le développement d'une méthode avec Turbo Quant, en particulier le modèle MS₁₀.

Tableau 46 : coefficients de variation et erreurs relatives à partir des dérivées premières

Méthodes	Pts. valid	Teneurs théoriques					Forme ϵ		Forme α		Forme β		Forme γ		Forme PN	
		ϵ (%)	α (%)	β (%)	γ (%)	PN (%)	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %
MS ₁₆	34	96.3	0.3	1.1	1.1	1.2	1.0	1.3	99.3	341.3	32.2	14.4	18.9	10.3	36.1	45.9
	35	95.3	2.1	0.3	1.1	1.1	0.6	0.4	19.5	2.41	56.4	34.7	34.4	3.2	27.4	30.1
	36	95.3	2.1	1.1	0.3	1.2	0.4	1.8	7.1	34.7	9.4	8.2	26.6	269.2	10.8	6.4
	37	95.2	2.1	1.1	1.2	0.3	0.6	1.3	22.1	14.4	19.6	14.0	25.5	13.6	26.4	264.4
	38	92.6	4.0	1.1	1.1	1.2	0.7	0.6	15.9	30.2	12.2	40.1	28.4	15.6	17.9	30.6
	39	93.6	2.1	2.0	1.1	1.2	0.8	0.7	15.4	31.9	14.2	24.8	22.5	3.3	27.6	35.7
	40	93.6	2.1	1.1	2.0	1.2	0.7	0.3	34.2	25.0	14.9	2.5	14.7	38.9	28.2	46.9
	41	93.6	2.1	1.1	1.1	2.0	0.8	0.5	25.5	6.4	25.6	2.4	41.2	16.0	20.0	22.9
MS _{17a19}	34	96.3	0.3	1.1			0.8	1.4	94.4	318.7	32.0	15.0				
	35	95.3	2.1	0.3			0.7	0.2	21.7	1.7	56.5	45.1				
	36	95.3	2.1	1.1			0.5	1.8	6.8	26.0	10.4	10.0				
	37	95.2	2.1	1.1			0.6	1.2	18.9	9.9	17.9	17.6				
	38	92.6	4.0	1.1			0.7	0.7	15.8	28.6	16.5	35.6				
	39	93.6	2.1	2.0			0.9	0.5	14.6	23.6	16.0	23.7				
	40	93.6	2.1	1.1			0.8	0.2	27.2	26.8	19.3	0.9				
	41	93.6	2.1	1.1			0.5	0.5	22.2	0.6	26.0	2.4				

9.7.2 - Méthodes d'analyse développées en « Interne » avec Turbo Quant

La régression PLS est également effectuée à partir des 33 spectres (répété chacun 5 fois) des échantillons de référence à l'aide du logiciel TURBO QUANT version 5.0, en utilisant l'ensemble des concentrations des différents constituants.

Huit modèles ont été testés, mais deux seulement ont été retenus. Pour le premier modèle retenu, référencé TQ₄, le choix des fenêtres spectrales informatives a été effectué en ne sélectionnant que les zones de plus grande variance entre les spectres et de plus grande corrélation entre les concentrations et les absorbances. Le critère additionnel de choix retenu est la minimisation du PRESS. Le deuxième modèle référencé TQ₅ a été élaboré à partir du modèle MS₁₀ (absorbances pures) obtenu avec SIMCA, en faisant évoluer le paramètre « fenêtres spectrales » pas à pas jusqu'à obtenir les zones les plus appropriées au logiciel TURBO QUANT. Comme avec SIMCA nous avons essayé :

- de dissocier les réponses (régression type PLS1), les résultats étaient identiques que l'on ait Y=5 ou Y=3 réponses. Ce qui nous a permis de constater que le logiciel TURBO QUANT travaillait avec l'algorithme PLS1.
- de « jongler » entre absorbances pures et dérivées premières. Six modèles ont été construits pour finalement n'en retenir qu'un. Le choix a été effectué selon les critères statistiques habituels : meilleur coefficient de corrélation, meilleurs résultats de calibration, validation croisée et externe, PRESS.

Le tableau 47 résume les fenêtres spectrales retenues et la correspondance des modèles TQ₄ et TQ₅ avec les fichiers de travail.

Tableau 47 : fenêtres spectrales et correspondances

Méthode	Référence interne	Fenêtres spectrales
TQ ₄	PLS02c1 Y=5	4 fenêtres spectr. : 415- 1232; 1359- 1550 ; 1634-1908 ; 2669-3112cm ⁻¹ .
TQ ₅	PLS02c1 Y=5 SIMCA4	7 fenêtres spectr. : 553-599 ; 615-669 ; 692-761 ; 808-1070 ; 1108-1394 ; 1510-1695 ; 2800-3200cm ⁻¹ .

Pour chaque modèle la calibration fut réalisée à partir des mélanges 1 à 33, la validation interne sur les mélanges 37-38 et 41, et la validation externe sur les mélanges 34 à 41.

Pour le modèle TQ₄, à partir des 4 fenêtres spectrales informatives retenues (figure 67) et la possibilité de travailler soit à partir des absorbances brutes, soit des dérivées premières, on a 16 possibilités (algorithme de Yates). Une seule possibilité a été retenue (tableau 48).

Pour le modèle TQ₅, le choix s'est fait avec « l'œil du spectroscopiste » en regardant l'allure du spectre (dérive de la ligne de base ...). Le tableau 48 résume les combinaisons retenues.

Tableau 48 : correspondances entre fenêtres spectrales - absorbances ou dérivées premières

Méthode	Fenêtres spectrales	Absorbances	Dérivées premières
TQ ₄	1	X	
	2, 3 et 4		X
TQ ₅	1, 2, 4 et 7	X	
	3, 5 et 6		X

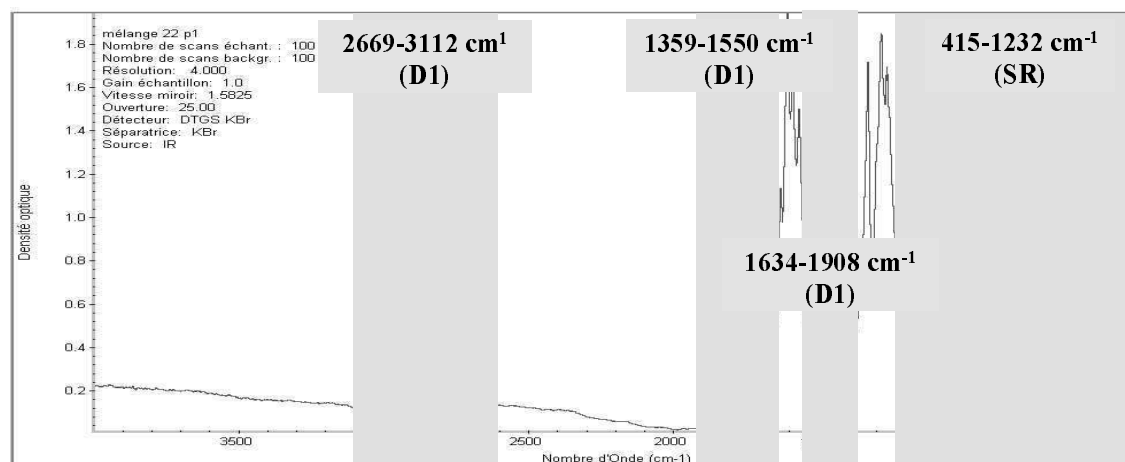


Figure 67 : fenêtres informatives retenues pour la méthode TQ₄

Les graphiques des PRESS nous ont permis de déterminer le nombre optimal de facteurs pour chaque réponse (tableau 49). Une première indication sur la validité des modèles est donnée par le coefficient de régression R^2 . Le tableau 50 rassemble ces coefficients. A l'issue de ce travail nous avons retenu le modèle TQ₄ toujours à partir des mêmes critères statistiques. La comparaison entre valeurs théoriques et valeurs estimées par le modèle TQ₄ est représentée graphiquement figure 68.

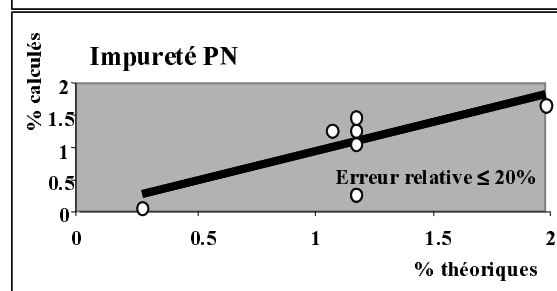
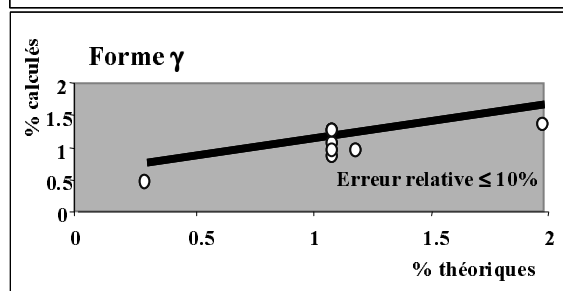
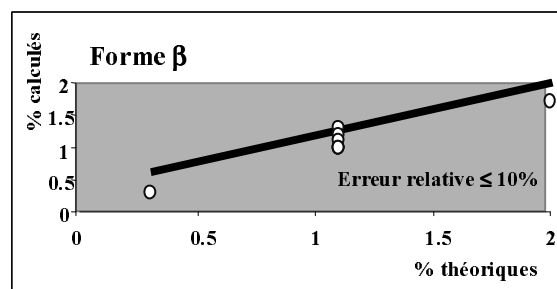
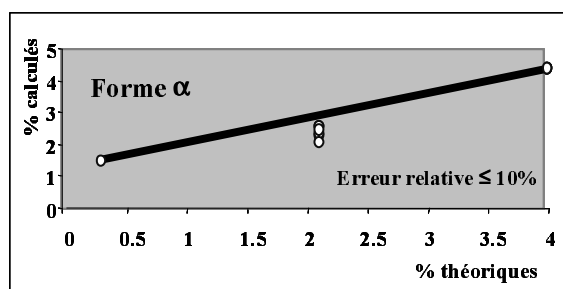
Tableau 49 : nombre de facteurs retenus

	Nombre de facteurs TQ ₄	Nombre de facteurs TQ ₅
α	9	9
β	10	10
γ	10	10
ϵ	10	10
PN	9	10

Tableau 50 : coefficients de régression R^2 obtenus pour les modèles retenus

	TQ ₄	TQ ₅
α	0.972	0.971
β	0.979	0.976
γ	0.956	0.937
ϵ	0.979	0.985
PN	0.935	0.942

Remarque : ces coefficients sont relativement plus faibles que ceux obtenus avec le premier plan d'étalonnage. On peut expliquer ce résultat par l'extension des limites des domaines de concentrations vers les basses teneurs et l'existence d'un cinquième composé (l'impureté pentanitree).



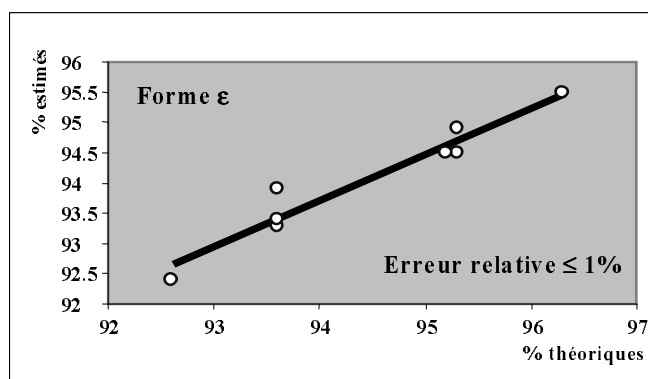


Figure 68 : validation du modèle TQ4

Comme avec SIMCA, nous pouvons rendre une valeur moyenne et un écart type. Le tableau 51 rassemble les résultats des dosages obtenus pour les points tests, dans le tableau 52 nous avons rassemblé coefficients de variation et erreurs relatives.

Tableau 51 : validation des modèles

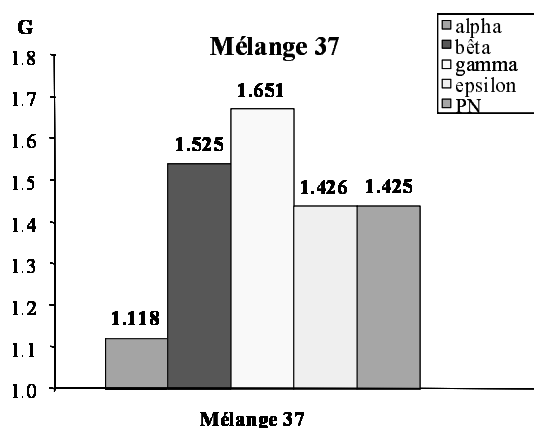
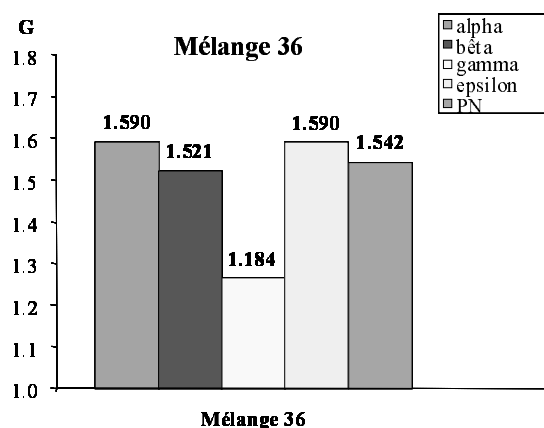
Méthodes	Pts valid.	Compositions réelles					Compositions estimées par les Méthodes de dosage									
		ϵ (%)	α (%)	β (%)	γ (%)	PN (%)	ϵ (%) σ		α (%) σ		β (%) σ		γ (%)	σ	PN (%) σ	
TQ ₄	34	96.3	0.3	1.1	1.1	1.2	95.5	0.1	1.4	0.4	1.0	0.1	0.9	0.2	1.5	0.1
	35	95.3	2.1	0.3	1.1	1.1	94.9	0.2	2.2	0.3	0.3	0.2	1.1	0.4	1.6	0.1
	36	95.3	2.1	1.1	0.3	1.2	94.5	0.3	2.5	0.3	1.3	0.1	0.5	0.1	1.1	0.3
	37	95.2	2.1	1.1	1.2	0.3	94.5	0.3	2.3	0.2	1.2	0.2	1.0	0.1	1.0	0.2
	38	92.6	4.0	1.1	1.1	1.2	92.4	0.5	4.3	0.4	1.1	0.2	1.3	0.2	1.1	0.1
	39	93.6	2.1	2.0	1.1	1.2	93.3	0.4	2.0	0.4	1.7	0.1	1.3	0.3	1.6	0.2
	40	93.6	2.1	1.1	2.0	1.2	93.4	0.3	2.5	0.3	1.0	0.2	1.4	0.2	1.7	0.1
	41	93.6	2.1	1.1	1.1	2.0	93.9	0.5	2.4	0.3	1.0	0.2	1.0	0.2	1.8	0.2
TQ ₅	34	96.3	0.3	1.1	1.1	1.2	96.0	0.2	0.5	0.2	0.9	0.2	0.8	0.2	1.8	0.2
	35	95.3	2.1	0.3	1.1	1.1	94.8	0.3	2.2	0.2	0.4	0.3	1.0	0.4	1.7	0.1
	36	95.3	2.1	1.1	0.3	1.2	94.2	0.6	2.3	0.3	1.2	0.2	0.6	0.4	1.6	0.2
	37	95.2	2.1	1.1	1.2	0.3	94.4	0.3	2.5	0.1	1.1	0.2	1.0	0.3	1.0	0.2
	38	92.6	4.0	1.1	1.1	1.2	92.3	0.4	4.3	0.3	1.0	0.2	0.8	0.1	1.5	0.1
	39	93.6	2.1	2.0	1.1	1.2	93.0	0.5	2.3	0.2	1.6	0.1	1.6	0.4	1.5	0.2
	40	93.6	2.1	1.1	2.0	1.2	93.1	0.4	2.7	0.3	1.0	0.2	1.5	0.4	1.8	0.1
	41	93.6	2.1	1.1	1.1	2.0	93.8	0.4	2.2	0.4	0.8	0.1	1.0	0.1	2.0	0.2

A partir du modèle retenu TQ₄ et à l'aide du test de GRUBBS [86] dont nous résumons la théorie dans l'annexe 10, nous avons recherché les « points aberrants » ou pouvant être considérés comme tels. Il s'agit en fait de détecter une(des) pastille(s) éventuellement « défectueuse(s) ». Les résultats numériques du test sont donnés en annexe 10, nous n'en reportons ici que la représentation graphique (figure 69). Si aucune pastille n'est jugée aberrante, 2 ont été jugées douteuses. N'étant pas aberrantes, ces 2 pastilles ont été prises en compte dans les calculs.

Tableau 52 : coefficients de variation et erreurs relatives

Méthodes	Pts. valid	Teneurs théoriques					Forme ϵ		Forme α		Forme β		Forme γ		Forme PN	
		ϵ (%)	α (%)	β (%)	γ (%)	PN (%)	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %	C.var %	Er.rel %
TQ ₄	34	96.3	0.3	1.1	1.1	1.2	0.1	0.9	25.9	35.1	12.9	13.4	25.4	17.9	6.8	32.2
	35	95.3	2.1	0.3	1.1	1.1	0.2	0.5	12.1	3.4	53.9	12.9	32.8	1.4	5.9	38.5
	36	95.3	2.1	1.1	0.3	1.2	0.3	0.8	11.0	20.6	11.2	20.0	17.8	59.1	28.9	4.1
	37	95.2	2.1	1.1	1.2	0.3	0.3	0.7	7.6	9.9	15.0	6.4	14.0	14.9	18.2	148.6
	38	92.6	4.0	1.1	1.1	1.2	0.6	0.2	9.7	7.2	15.1	0.8	16.4	14.1	8.4	7.2
	39	93.6	2.1	2.0	1.1	1.2	0.4	0.3	18.3	6.3	7.2	15.1	22.0	14.6	11.5	34.3
	40	93.6	2.1	1.1	2.0	1.2	0.3	0.2	10.3	16.3	20.5	10.8	17.3	28.9	4.0	42.6
	41	93.6	2.1	1.1	1.1	2.0	0.5	0.3	14.3	13.2	17.9	13.1	17.1	5.1	10.5	11.1
TQ ₅	34	96.3	0.3	1.1	1.1	1.2	0.2	0.3	37.4	60.9	21.3	17.2	22.4	29.4	8.6	57.3
	35	95.3	2.1	0.3	1.1	1.1	0.3	0.6	7.3	3.8	59.1	37.7	46.4	13.5	8.5	48.4
	36	95.3	2.1	1.1	0.3	1.2	0.7	1.2	13.2	10.6	13.8	7.7	64.3	100	11.9	34.6
	37	95.2	2.1	1.1	1.2	0.3	0.3	0.8	4.5	20.4	14.4	0.2	26.9	18.5	19.8	150.9
	38	92.6	4.0	1.1	1.1	1.2	0.5	0.3	6.9	7.6	15.7	10.0	11.9	31.4	6.7	24.9
	39	93.6	2.1	2.0	1.1	1.2	0.6	0.7	9.5	10.3	9.2	20.0	24.7	40.0	10.7	26.0
	40	93.6	2.1	1.1	2.0	1.2	0.4	0.5	11.2	27.2	18.3	7.7	27.2	24.7	7.8	56.6
	41	93.6	2.1	1.1	1.1	2.0	0.4	0.2	18.3	6.7	14.8	23.0	6.7	6.1	9.6	1.0

Notions importantes : les graphiques reportés figure 69 représentent la valeur de la variable de GRUBBS (Ga ou Gb) la plus forte : $G_a = [x_{\max} - \text{moy}(x)]/s$ et $G_b = [\text{moy}(x) - x_{\min}]/s$ avec x la variable et s l'écart type. On compare les valeurs expérimentales de Ga et Gb aux valeurs théoriques de la statistique de GRUBBS avec dans notre cas : si $G_i > 1,764$ alors la donnée est *aberrante*, si $1,715 < G_i < 1,764$ alors la donnée est *douteuse*.



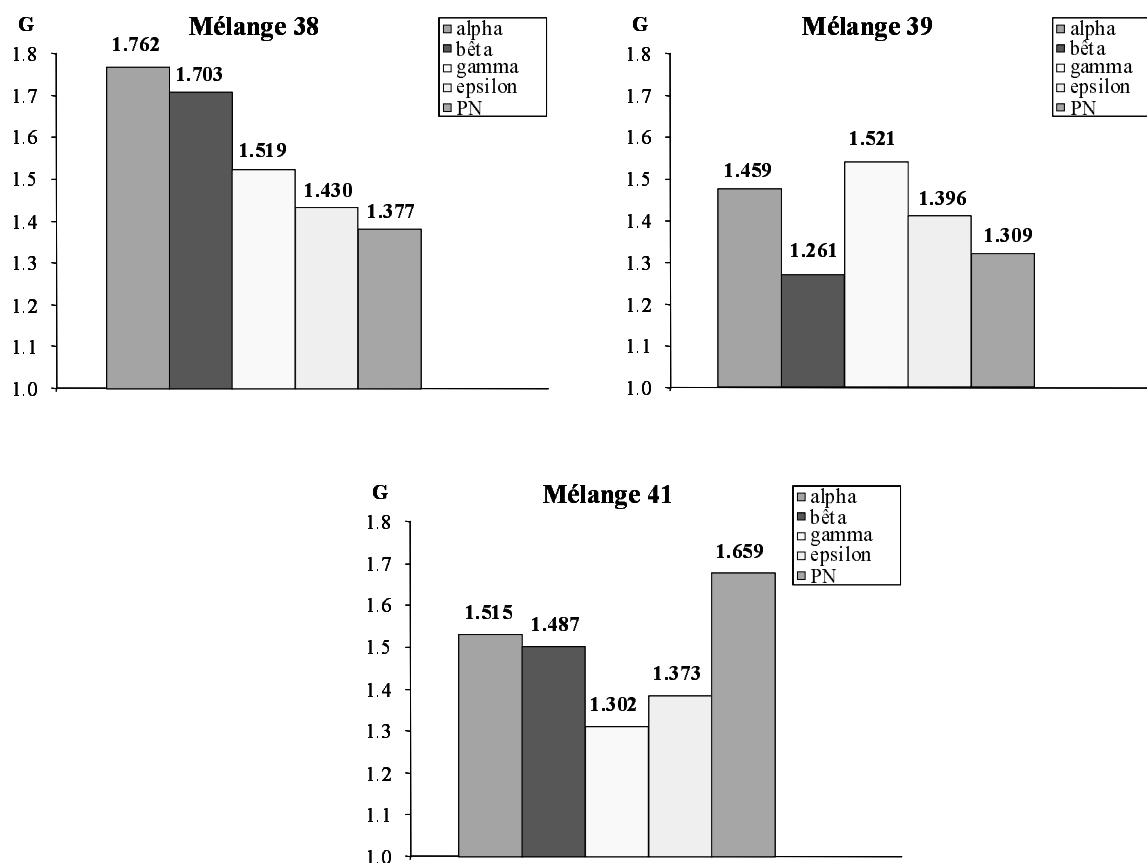


Figure 69 : test de GRUBBS appliqué aux résultats de la méthode TQ₄

L'examen des tableaux 51 et 52 montre un comportement similaire du modèle TQ₄ à celui des modèles développés avec SIMCA :

- dans tous les cas on obtient une qualité de prédiction très satisfaisante pour la teneur en forme ϵ (valeur moyenne des 5 spectres très proche de la valeur réelle, erreur relative moyenne de 0,6% et coefficient de variation moyen de 0,3%).
- pour α et β , d'une manière générale (excepté la teneur 0,3%) la prédiction est assez satisfaisante (les valeurs moyennes des 5 spectres sont assez proches de la valeur réelle, sans tenir compte de la teneur 0,3% l'erreur relative moyenne est de 11,7% pour α et de 14,8% pour β , les coefficients de variation moyen sont de 13% pour les deux formes polymorphes).
- par contre la forme γ est assez bien estimée contrairement aux résultats de SIMCA (les valeurs moyennes des spectres sont proches de la réalité et la précision moyenne est de 15,7%), mais avec une mauvaise répétabilité (coefficient de variation moyen de 24,4% le plus élevé des 4 formes polymorphes).

Les performances obtenues avec les modèles TQ_i et SIMCA sont identiques. Nous obtenons une très bonne prévision sur le composé epsilon. Pour les autres formes polymorphes et l'impureté pentanitree, la prédiction reste acceptable, excepté pour les très faibles teneurs (seuil de quantification IRTF pour ces composés). Globalement la méthode de dosage des formes polymorphes et du pentanitree TQ₄ élaborée à l'aide du moyen infrarouge et de la régression PLS est acceptable. Nous l'avons donc appliquée à l'analyse de lots industriels. Un résumé des descripteurs statistiques de la méthode TQ₄ retenue est donné tableau 53.

Tableau 53 : descripteurs statistiques de la méthode retenue

Composés	R ²	Nb de composantes	PRESS ($\sum r_i^2$)	σ résid.
Forme α	0.972	9	41.9	0.5
Forme β	0.979	10	9.1	0.2
Forme γ	0.956	10	31.4	0.4
Forme ϵ	0.979	10	80.5	0.7
Impureté PN	0.935	9	23.5	0.4

CONCLUSION

Nous avons montré la nécessité d'étudier la technique d'échantillonnage (pastille KBr) afin d'en assurer la fiabilité et la robustesse. Pour ce faire, nous avons effectué un plan de criblage pour déterminer les facteurs influents sur la méthode de préparation de la pastille.

Afin de mettre au point une technique de dosage de la pureté polymorphique dans l'infrarouge moyen (IRTF – MIR) par régression PLS, nous avons réalisé plusieurs étalonnages avec différents domaines de contraintes pour chacun des composés présents. Outre les 4 formes polymorphes du CL20, la prise en compte de l'impureté pentanitree s'est avérée être un facteur important pour le dosage.

La méthode TQ₄ retenue a été optimisée, elle est fiable et robuste pour l'estimation de la forme polymorphe ϵ qui est notre composé d'intérêt, elle est moins précise mais néanmoins acceptable pour la prédiction des 3 autres formes et de l'impureté pentanitree. En répondant au cahier des charges fixé par le groupe SNPE : doser avec précision la forme ϵ dans le domaine 95-98,8% et une estimation du taux des 3 autres formes polymorphe et de l'impureté pentanitree avec la précision demandée par les impératifs techniques et commerciaux, ce modèle peut servir de méthode de dosage de référence pour tous les lots de CL20.

Nous en concluons que notre objectif a été atteint.