

## ■ Analyse de talcs par spectroscopie proche infrarouge. Mise en œuvre d'un outil de contrôle de matières premières

J.-C. Boulou

*Ingénieur d'applications Infrarouge et Raman, Société Bruker Spectrospin, 34 rue de l'Industrie, 67166 Wissembourg Cedex, France*

La recherche de la qualité est devenue pour beaucoup d'industries un besoin et même un véritable impératif. La maîtrise de la qualité nécessite un certain investissement qui aura une incidence directe sur la sécurité et la productivité de la fabrication et des produits.

La qualité du produit final dépend du contrôle et de la maîtrise de toutes les étapes de la fabrication, mais aussi d'un suivi des produits de base : construire un outil qualité passe obligatoirement par un contrôle des matières premières. Indépendamment d'une contrainte imposée dans certains secteurs industriels (cf. industries pharmaceutiques), le contrôle systématique (toutes les matières, tous les lots, tous les contenants) des matières premières est un élément indispensable de la chaîne de qualité.

Depuis quelque temps la spectroscopie proche infrarouge s'est montrée un outil très intéressant et très performant pour ce type d'analyse de matières premières. Parmi tous les avantages du proche infrarouge on peut citer :

- la facilité et la rapidité de la mesure par une analyse directe sans préparation de l'échantillon et pouvant être faite par un non-spécialiste.
- la possibilité d'utiliser des fibres optiques offrant une délocalisation et une multiplication des points de mesure.
- la performance de la mesure par la sensibilité de la spectroscopie proche infrarouge à la fois aux propriétés chimiques et physiques des produits.

Depuis une dizaine d'années la spectroscopie proche infrarouge est devenue un outil de choix pour le contrôle des matières premières. Elle répond parfaitement aux exigences du contrôle et du gain en productivité en permettant une augmentation du nombre de mesures tout en diminuant la charge de travail du laboratoire.

La fiabilité et la robustesse de la méthode nécessitent une mise au point minutieuse et rigoureuse. Elle requiert des compétences spectroscopiques. Son utilisation en routine reste, par contre, très simple et ne nécessite aucune qualification.

En traitant un double exemple d'analyse de talcs nous nous proposons d'aborder la méthodologie initiale de construction d'une base de données et son utilisation pour le contrôle de matières premières.

### La méthodologie générale

On a l'habitude de dire qu'en spectroscopie proche infrarouge chaque cas est un cas particulier. Sans infirmer totalement cette phrase on peut dégager, dans le cadre de la construction d'un outil analytique de contrôle de matières premières par spectroscopie proche infrarouge, une approche générale et un certain nombre de règles à respecter.

#### Le principe de base de l'analyse

L'objectif d'un contrôle des matières premières est double. Il s'agit dans un premier temps d'identifier, dans le sens de confirmer, la nature du produit puis de le valider par rapport à une propriété qu'elle soit chimique, physique voire par rapport à une origine, un fournisseur ou une qualité. Pour pouvoir mener à bien cette validation comparative il est évident qu'il faut construire un référentiel qui intégrera toutes les matières à contrôler et qui aura défini le pouvoir et les conditions de discrimination de tous ces produits à analyser.

Cette étape nécessite du temps et des échantillons, car pour définir une référence correcte il faut mesurer plusieurs fois plusieurs lots de chaque matière. La Pharmacopée Européenne, par exemple, préconise un minimum de 3 à 5 mesures par lot et un minimum de 10 lots pour définir un spectre référence significatif d'un produit.

La mise au point d'un outil d'analyse efficace et répondant parfaitement aux exigences de la demande nécessite une modélisation mathématique des spectres. L'utilisation de ces programmes mathématiques requiert un minimum de connaissance dans leurs principes.

#### Les outils mathématiques

Nous reprendrons dans cette partie deux des principes mathématiques parmi les plus couramment utilisés et qui ont servi à l'analyse des talcs par l'intermédiaire du logiciel Bruker Opus™ et de son module Ident d'analyse discriminante.

#### La méthode standard : calcul de distances euclidiennes

Dans cette méthode, chaque spectre est représenté par un point dans un espace mathématique à  $N$  dimensions définies

par les longueurs d'onde choisies ( $N =$  nombre d'éléments spectraux). Dans cet espace on peut calculer une distance euclidienne entre deux spectres  $a$  et  $b$  à partir de leurs ordonnées  $a(k)$  et  $b(k)$  par

$$D = \sqrt{\sum_k (a(k) - b(k))^2}$$

Les produits similaires ayant des spectres très voisins seront donc représentés par des points proches et délimiteront ainsi des zones dans l'espace. Comme la variation des spectres n'est pas constante à toutes les longueurs d'onde mais est focalisée sur certaines fréquences, les zones spatiales constituées par les différentes familles de produits tendent vers des ellipsoïdes. C'est la raison pour laquelle on peut être amené à calculer non plus une distance euclidienne simple  $D$  mais une distance non-euclidienne appelée distance de Mahalanobis. Pour le calcul de cette distance de Mahalanobis l'unité de distance vectorielle est normalisée en fonction de l'amplitude de la variation dans la direction considérée.

Ces distances vont être exploitées de la façon suivante : pour une matière donnée dont on aura mesuré  $n$  lots, on définit sa zone d'influence par le point moyen représentatif du spectre moyen dans l'espace multidimensionnel, la déviation standard  $S_0$  des  $n$  lots et la distance moyenne  $D_M$  des points individuels par rapport au point moyen.

La validation globale de la méthode se fera par un test statistique de non-recouvrement des zones spatiales définies par toutes les matières à tester.

La distance de tolérance  $D_t$  est calculée par :

$$D_t = D_{max} + S_0/4$$

où  $D_{max}$  est la distance au point moyen du point le plus éloigné.

On peut aussi calculer  $D_t$  à partir d'un facteur de probabilité défini sur une répartition gaussienne des points.

L'analyse d'un nouveau lot se fait par calcul de sa distance  $D$  au point le plus proche pour l'identification du produit et par la vérification  $D < D_t$  pour la validation en conformité du lot.

Cette méthode standard est simple d'utilisation, de compréhension et de gestion. Elle est peu visuelle mais elle permet par contre de mieux séparer des produits très voisins par rapport à l'analyse en composantes principales décrites ci-après.

#### L'analyse en composantes principales

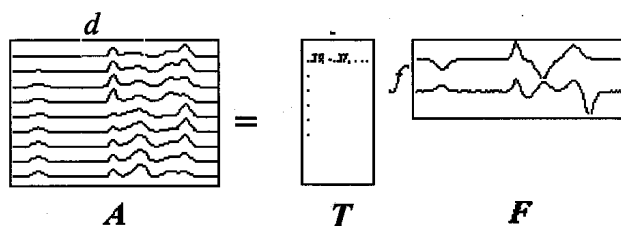
L'analyse en composantes principales (ACP) fait partie des outils mathématiques de compression de données. Elle part du principe que l'information contenue dans un grand nombre de variables (les spectres dans le cas de la spectroscopie proche infrarouge) peut être représentée par quelques données dites latentes (encore appelées Loadings, facteurs ou vecteurs) car non directement visibles.

Dans une première étape dite de factorisation, on va calculer par une série d'opérations matricielles ces variables latentes qui vont permettre, par des combinaisons linéaires, la modélisation de l'ensemble des données initiales. Le but de cette analyse mathématique étant de réduire le plus

possible le nombre de variables, on va dans un deuxième temps déterminer le nombre minimum de facteurs qu'il faut utiliser pour minimiser l'erreur de la modélisation. Les facteurs calculés par la factorisation n'ont pas tous la même importance, on ne s'intéressera et on n'utilisera que les plus significatifs pour le modèle : les composantes principales. Ces composantes représentent en fin de compte des directions privilégiées de variance qui vont nous permettre de définir des familles ou des classes de produits.

$$[a] = T_{1a} \cdot f_1 + T_{2a} \cdot f_2 + T_{3a} \cdot f_3 + \dots$$

avec  $[a]$  : vecteur représentant le spectre  $a$ ,  
 $f$  : les spectres-facteurs (ou facteurs ou vecteurs),  
 $T_{ia}$  : les coefficients multiplicatifs (ou scores) des facteurs  $f$  qui permettent de reconstruire le spectre  $a$ .



Pour le calcul mathématique on fait les opérations matricielles suivantes :

à partir de  $s$  spectres référence dont on utilise  $d$  points  $d_i$  est le vecteur colonne représentant le spectre  $i$

$$D = [d_1 \ d_2 \ d_3 \ \dots \ d_s] : \text{matrice des données}$$

$$D^T = \text{matrice transposée de } D$$

on calcule la matrice de covariance  $Z$

$$Z = D^T \times D.$$

Les vecteurs-propres et les valeurs propres de la matrice  $Z$  sont obtenus par sa diagonalisation et sa transformation orthogonale :

$$L^T \times Z \times L = G.$$

Les vecteurs colonnes de la matrice  $L$  sont les vecteurs propres orthogonaux  $I_i$  de la matrice  $Z$ .

$$I_i \times I_j = 0 \ (i \neq j)$$

$$I_i \times I_i = 1.$$

La matrice  $G$  contient les valeurs propres  $\lambda_i$  de  $Z$  sur sa diagonale principale, les autres éléments de la matrice étant nuls.

D'où l'on peut écrire :

$$Z \times I_i = \lambda_i \times I_i.$$

La matrice des données  $D$  est factorisée par multiplication par la matrice propre  $L$  :

$$F = D \times L$$

$$\text{avec } F^T \times F = (D \times L)^T \times (D \times L) = L^T \times Z \times L = G$$

$L$  étant orthogonale :  $D = D \times 1 = D \times L \times L^T = F \times L^T$ .

Les spectres référence  $d$  sont représentés par combinaisons linéaires des spectres facteurs de  $F$  avec comme coefficients les colonnes de la matrice  $L^T$  :

$$d_i = L^T_{1,i} \times f_1 + L^T_{2,i} \times f_2 + \dots + L^T_{s,i} \times f_s.$$

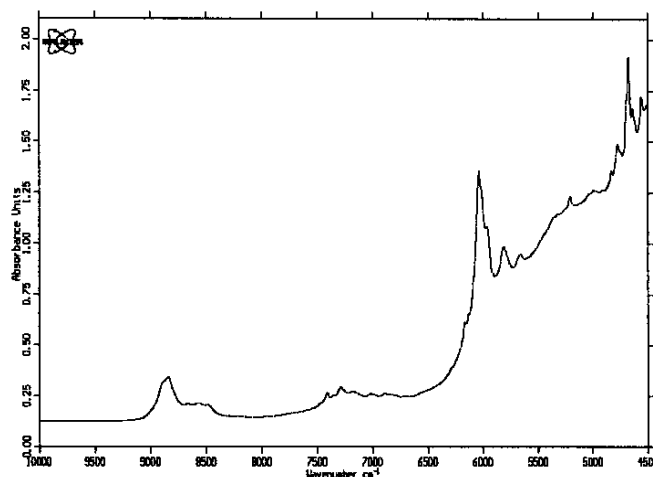


Figure 1. Spectre de référence.

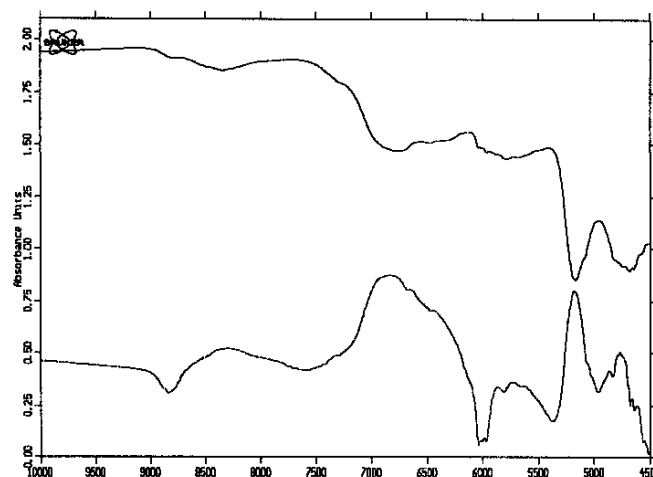


Figure 2. 1<sup>er</sup> facteur (en haut), 2<sup>ème</sup> facteur (en bas).

Les éléments  $L^T_{ij}$  sont les nouvelles coordonnées du point représentant le spectre  $d$  dans la nouvelle base vectorielle constituée par les facteurs  $f$ .

Un spectre arbitraire  $u$  sera défini par :

$$u = F \times k + e$$

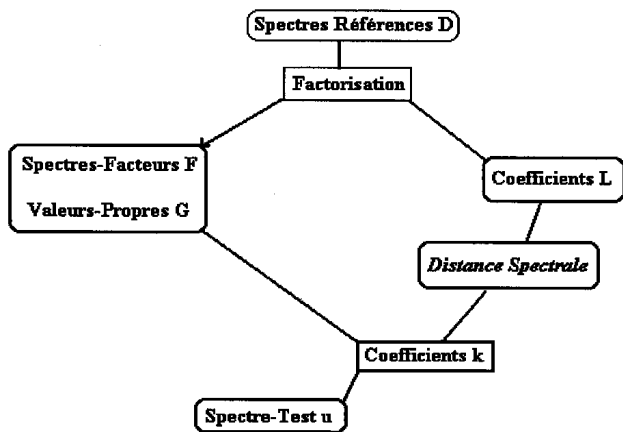
où  $e$  est l'erreur entre le spectre  $u$  et le spectre reconstitué ; erreur que l'on cherche à minimiser d'où :

$$k = (F^T \times F)^{-1} \times F^T \times u = G^{-1} \times F^T \times u.$$

La factorisation ayant pour but de réduire le nombre de données on ne va pas utiliser les  $s$  facteurs, mais uniquement les plus importants, c'est-à-dire ceux ayant les valeurs propres les plus grandes. Si on ne prend que les  $r$  premiers facteurs, le calcul de la distance entre le spectre  $u$  et le spectre référence  $d_a$  se réduit à :

$$D = \sqrt{(k_1 - L^T_{1j})^2 + (k_2 - L^T_{2j})^2 + \dots + (k_r - L^T_{rj})^2}$$

d'où l'on peut remplacer sans problème les éléments  $L^T_{ia}$  par les éléments  $L_{ai}$ .



On retiendra de l'analyse que le calcul des distances spectrales se fait à partir d'un nombre de points beaucoup plus faibles ( $r$ ) et est donc plus rapide, et que l'information intéressante a été condensée sur un nombre plus petit d'éléments.

Les facteurs définissent en fin de compte des directions de variance privilégiées et en représentant un facteur par rapport à un autre, on obtiendra une projection bi ou tridimensionnelle des familles de spectres par rapport à la propriété considérée. L'analyse en composantes principales apporte par ce biais un outil graphique de détermination et de vérification du caractère discriminatoire des produits et facilite la mise en place de la méthode.

Le choix des facteurs et leur nombre à utiliser sont des éléments déterminants de l'analyse en composantes principales. Le nombre de facteurs doit être optimum. Trop petit il ne permet pas de discriminer tous les produits, trop grand, il surestime le modèle et rend la discrimination sensible à des éléments aléatoires.

### Les règles de base

La réalisation d'un outil de contrôle de matières premières n'est pas toujours simple au départ. Cela nécessite beaucoup de produits et de temps. La qualité finale du système, sa robustesse et sa fiabilité dépendent directement du soin et de la méthode apportés à cette construction.

Il convient dès le départ de parfaitement établir le cahier des charges auquel doit répondre le système. Il faut analyser complètement le contexte du contrôle pour définir :

- le mode de mesure : mesure directe ou prélèvement
- le type de mesure : fibres optiques, sphère d'intégration, à travers un contenant d'échantillonnage

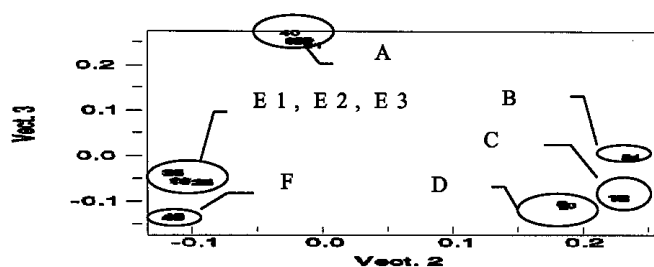


Figure 3. Exemple de répartition par classe de produits par projection de vecteurs propres.

- le lieu de mesure : site de prélèvement, laboratoire
- le ou les opérateurs : préleveur, technicien de laboratoire
- la nature des produits à contrôler : liquide, solide.

Il faut aussi définir les contraintes et les limites de la discrimination : doit-on par exemple analyser une matière globalement ou convient-il d'en discriminer des sous-ensembles vis-à-vis d'une qualité ou d'un fournisseur ? Faut-il effectivement intégrer la totalité des matières ou se limiter aux plus courantes ?

Le cahier des charges impose inévitablement des contraintes, mais il n'est pas rare d'avoir à trouver un compromis entre besoin et faisabilité.

La réalisation d'un système de contrôle par proche infrarouge démarre toujours par l'acquisition (et donc par la recherche de produits) d'un maximum de spectres émanant d'un plus grand nombre possible de produits et de lots. À partir du moment où la quantité de spectres commence à être statistiquement significative, on peut envisager d'aborder petit à petit le traitement mathématique des données disponibles. En général toute démarche logique va du caractère général vers les points particuliers. Cette approche appliquée à notre cas nous amène à priori à tester le caractère discriminatoire global des grandes familles de produits puis à s'intéresser aux sous-ensembles plus pointus. Cette approche logique et globale n'est pas toujours réalisable, car on doit aussi faire face à des contraintes de priorité et de fréquence de certains produits. Très souvent on mène de front en la vérifiant au fur et à mesure du degré d'avancement des opérations, l'analyse globale et l'analyse des cas particuliers. Concrètement, et c'est le cas qui va être détaillé par la suite, on va :

- vérifier le caractère discriminatoire des produits A (ex. Talc) et B (ex. Oxyde de Titane) l'un vis-à-vis de l'autre en définissant une méthode  $M_1$ .
- vérifier le caractère discriminatoire des sous-ensembles A1, A2, A3, ... (ex. talcs de granulométrie différente) entre eux en utilisant  $M_1$  ou en redéfinissant une méthode  $M_A$ .
- vérifier le caractère discriminatoire des sous-ensembles B1, B2, B3, ... entre eux en utilisant  $M_1$  ou en redéfinissant une méthode  $M_B$ .
- vérifier le caractère discriminatoire de l'ensemble A1, A2, A3, ..., B1, B2, B3, ... en utilisant  $M_1$  ou en redéfinissant une méthode globale  $M_2$ .
- rajouter un produit C et réitérer l'opération.

L'idéal serait d'utiliser une seule bibliothèque générale pour tous les produits et sous-familles, mais en pratique il faut souvent faire appel à des sous-bibliothèques et des méthodes spécifiques.

## L'analyse de Talcs

### L'objectif

Deux types de contrôle sont demandés sur des talcs :

- 3 talcs de fournisseurs différents dont la différence essentielle est la granulométrie (10, 20 et 50 microns)

que l'on notera par la suite par Talc10, Talc20 et Talc50. Ces produits se retrouveront entre autres avec des lactoses de 3 types différents.

- 3 catégories de talcs, un talc brut et deux talcs traités différemment que l'on notera respectivement par Talc\_0, Talc\_1 et Talc\_2. Ces trois produits seront très souvent associés à leurs équivalents en oxyde de titane.

Pour l'instant aucune analyse n'est effectuée sur ces produits par manque d'outils spécifiques adaptés, alors que les quantités réceptionnées représentent un pourcentage relativement important.

**Le choix de la mesure**

Dans les deux cas une mesure par fibres optiques à travers un flacon de prélèvement a été choisie pour répondre aux contraintes suivantes :

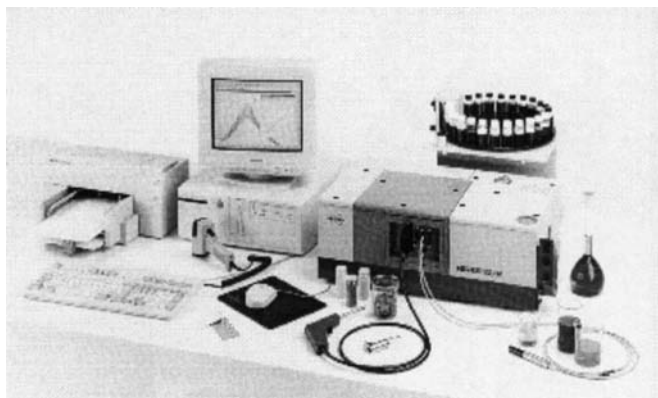
- élimination de tout risque de contamination liée à l'immersion directe de la sonde dans le contenant.
- Maintien du prélèvement pour l'échantillothèque.
- Présence de liquides parmi les autres produits à analyser imposant les fibres optiques plus universelles qu'une sphère d'intégration.

Les mesures ont été réalisées sur un système proche infrarouge **Bruker**, le **VectorN-F**, équipé d'un système de fibres optiques avec un passeur automatique adapté au nombre et au type de flacon de prélèvement. Le spectromètre travaille en standard avec une résolution spectrale de 8 cm<sup>-1</sup> (1,2 nm à 1250 nm) pour une gamme spectrale de 15 500 à 4000 cm<sup>-1</sup> (650 à 2500 nm) en utilisant une séparatrice Quartz et un détecteur InAs à refroidissement Peltier.

**Étude des talcs traités**

- Les produits et les spectres.

Trois types de talcs sont à contrôler : le talc brut (Talc\_0), et deux talcs traités (Talc\_1 et Talc\_2). Il est à noter qu'un troisième talc traité n'a pas été pris en compte, car il est progressivement remplacé par le Talc\_2. Chaque produit a été mesuré trois fois et moyenné avec en tout :



**Spectromètre Bruker VECTOR22N**

- 3 lots de Talc\_0
- 18 lots de Talc\_1
- 10 lots de Talc\_2

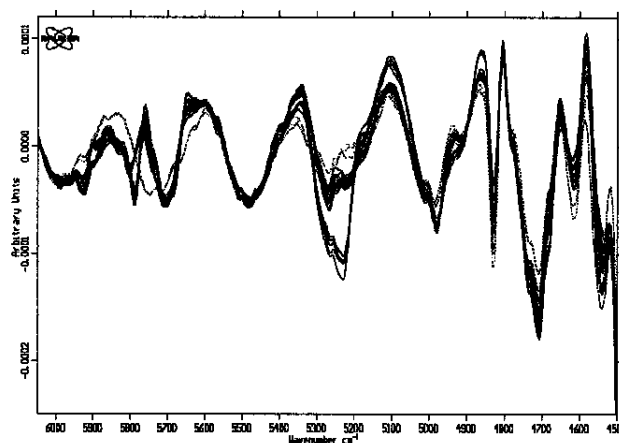
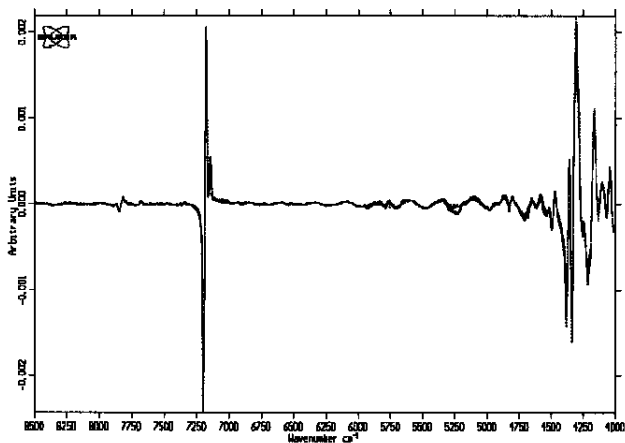
(la différence du nombre de lots est directement liée à la fréquence des produits) soit en tout 31 spectres pour lesquels on attend 3 classes.

- L'analyse en composantes principales.

Le caractère discriminatoire des 3 familles dans la zone spectrale 6050 – 4500 cm<sup>-1</sup> visible sur la figure 4b est confirmée par l'analyse en composantes principales par rapport à l'utilisation d'une gamme spectrale plus étendue. (Fig. 5)

Les résultats de l'ACP sur la gamme réduite sont présentés par les projections en figure 6 où l'on a tracé un vecteur *n* par rapport au vecteur *m* (successivement Vecteur1/Vecteur2, Vecteur1/Vecteur3, Vecteur2/Vecteur3 et Vecteur2/Vecteur4) (Fig. 6).

On constate facilement que l'introduction du 4ème facteur perturbe la séparation des produits. On en déduit par contre qu'en utilisant les trois premiers facteurs on obtient une bonne discrimination des familles. Cependant, il ne faut pas négliger le fait que ce nombre de 3 est peut-être



Figures 4a et 4b. Spectres des trois types de talc : a) entre 8500 et 4000 cm<sup>-1</sup> (1175 – 2500 nm) ; b) entre 6050 et 4500 cm<sup>-1</sup> (1650 – 2220 nm).

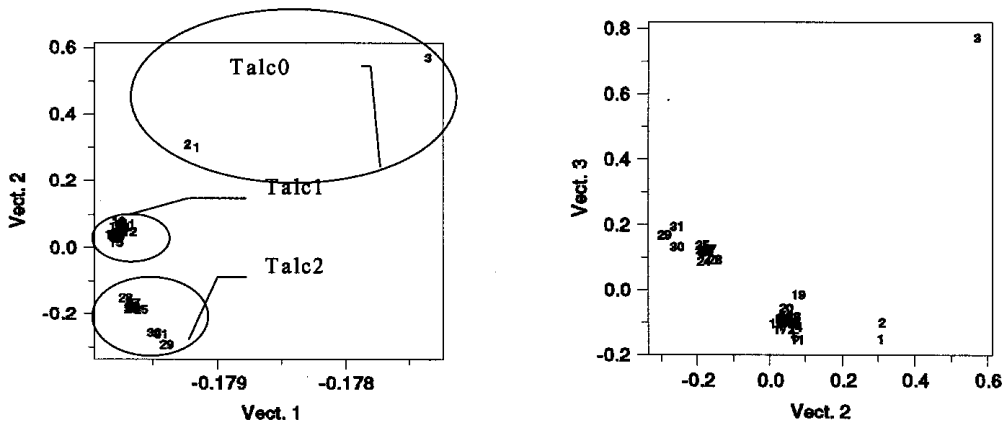


Figure 5. ACP sur la gamme spectrale 8500 – 4000 cm<sup>-1</sup> (1175 – 2500 nm).

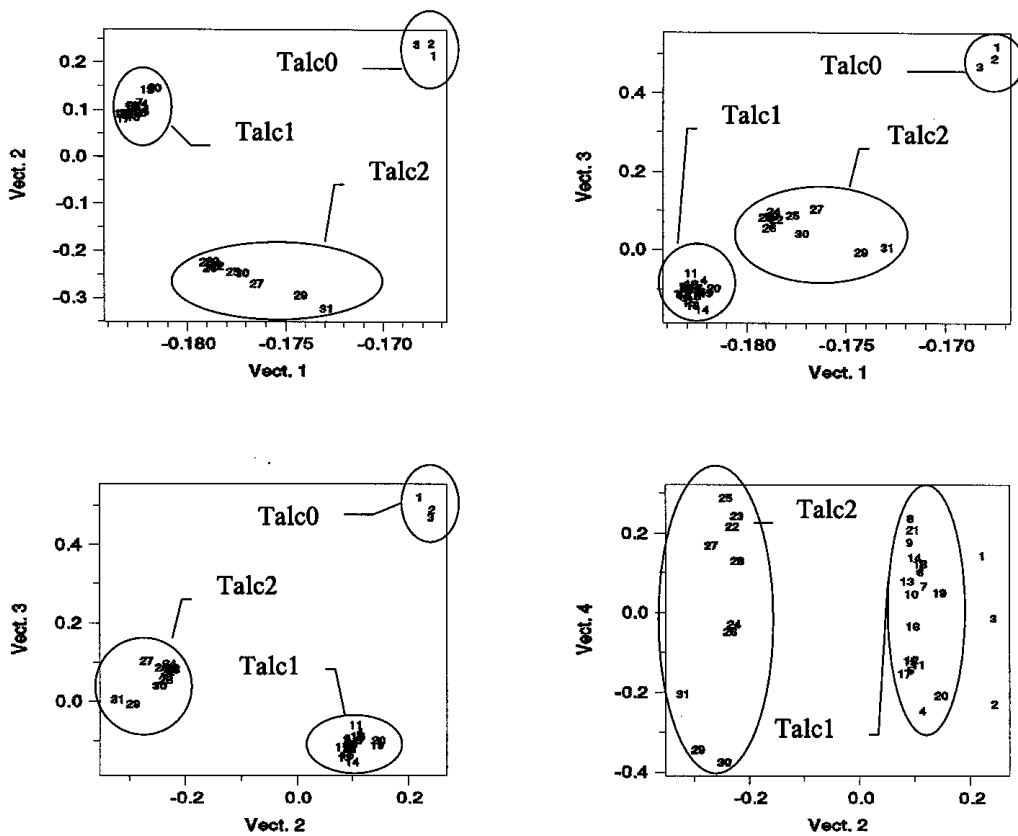


Figure 6. ACP entre 6050 et 4500 cm<sup>-1</sup> (1650 – 2220 nm).

Nombre de Facteurs	Facteur(s) utilisé(s)	Spectres confondus	Spectres inconnus	Spectres identifiés
1	Vecteur 1	10	0	21
1	<b>Vecteur 2</b>	<b>0</b>	<b>0</b>	<b>31</b>
1	Vecteur 3	8	1	22
2	Vecteurs 1 et 2	0	0	31
2	Vecteurs 1 et 3	8	1	22
2	Vecteurs 2 et 3	0	0	31
3	Vecteurs 1, 2 et 3	0	0	31

suresstimé par rapport au nombre de classes à séparer (3) et que le but par la suite est de compléter notre première série par de nouveaux produits.

Par un test de validation on va tester individuellement les 31 spectres originaux et vérifier combien sont parfaitement reconnus en fonction du nombre et des facteurs utilisés.

On déduit de cette validation qu'à priori l'utilisation d'un facteur suffit à discriminer les trois familles de talcs.

- Conclusion sur l'analyse des talcs traités.

L'étude précédente a permis de montrer qu'il est parfaitement possible d'analyser et de contrôler sélectivement les

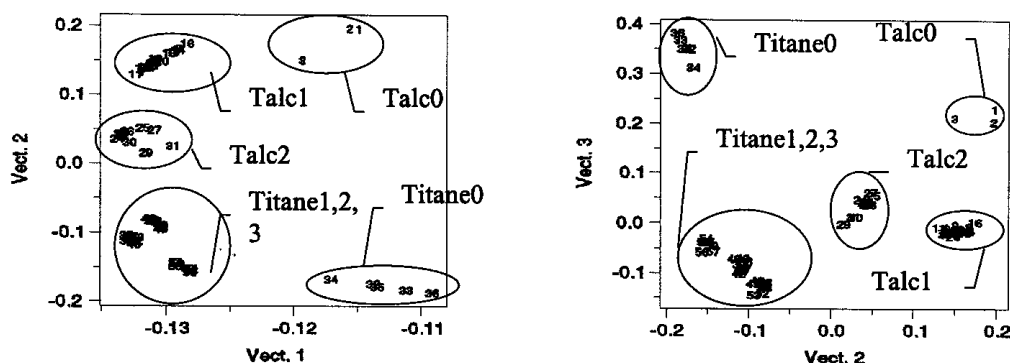


Figure 7. ACP Talcs + Oxydes de Titane : représentation des facteurs 1, 2 et 3.

trois types de talcs envisagés. On retiendra de façon complémentaire qu’il est très important pour la suite du travail de ne pas se limiter à la seule constatation du caractère discriminatoire, mais de compléter l’étude par des essais de validation pour déterminer le nombre optimum de facteurs à utiliser pour obtenir la discrimination.

- Ajout d’une nouvelle matière.

Nous allons maintenant rajouter à nos trois types de talcs, quatre types d’oxydes de titane (un brut et trois traités différemment). L’ensemble représentera 7 classes à discriminer pour un total de 60 spectres.

Par des essais préliminaires on a vérifié que les oxydes de titane se discriminaient comme les talcs dans une même zone spectrale étendue.

- Les produits :

3 talcs : 10, 20 et 50 microns (6 à 7 lots par produit)  
3 lactoses : D10, D80 et EPD (6 à 7 lots par produit).

Aucun lot n’a été particulièrement sélectionné ou spécifiquement éliminé en amont ou en aval du calcul (Fig. 8).

- L’analyse spectrale.

Bien que le talc ne présente pas beaucoup de bandes dans le domaine du proche infrarouge, une analyse spectrale visuelle simple permet de constater qu’il existe des différences entre les produits. Elles sont faibles mais néanmoins significatives pour le proche infrarouge. Cette analyse spectrale permet aussi de prévoir la ou les gammes spectrales qu’il faudra utiliser pour l’ACP.

- Résultats de l’analyse en composantes principales.

Les deux graphes ci-contre (Fig. 9) illustrent parfaitement à la fois les possibilités et les limitations d’une exploitation graphique trop rapide de l’ACP. Si on ne représente que le 1<sup>er</sup> facteur en fonction du 2<sup>ème</sup> on ne constate pas de discrimination entre deux des trois produits alors qu’il y a effectivement séparation totale si on prend les 2<sup>ème</sup> et 3<sup>ème</sup> facteurs. En réduisant la gamme spectrale on arrive à trouver une zone réduite pour laquelle les facteurs 1 et 2 offrent une bonne séparation alors que l’utilisation des facteurs 2 et 3 est pratiquement indépendante de la gamme spectrale. Pouvoir utiliser une gamme spectrale suffisamment étendue est toujours beaucoup plus favorable pour la suite quand il s’agira d’introduire d’autres matières dans l’analyse. Par validation on constate que l’on pourrait même n’utiliser qu’un facteur, le deuxième.

Nombre de Facteurs	Facteur(s) utilisé(s)	Spectres confondus	Spectres inconnus	Spectres identifiés
1	Vecteur 1	47	0	13
1	Vecteur 2	9	0	51
1	Vecteur 3	50	0	10
2	Vecteurs 1 et 2	3	0	57
2	Vecteurs 1 et 3	49	1	10
2	Vecteurs 2 et 3	1	0	59
3	Vecteurs 1, 2 et 3	1	0	59
3	<b>Vecteurs 2, 3 et 4</b>	<b>0</b>	<b>0</b>	<b>60</b>

On constate que l’on peut discriminer globalement les deux classes (talc et oxyde de titane) ainsi que les sous-classes (bruts et traités) en utilisant une même méthode avec trois facteurs. Il serait intéressant d’affiner les résultats, par un nombre de lots plus important, au niveau du spectre confondu avec les vecteurs 2 et 3 et avec les vecteurs 1, 2 et 3 (voir tableau) pour vérifier s’il s’agit bien d’une limite de la méthode ou d’un lot présentant une anomalie quelconque. On peut aussi noter que le vecteur 1 n’est *a priori* pas le plus significatif pour la discrimination.

**Analyse de talcs de granulométrie différente**

Dans cet exemple nous allons, en utilisant la même démarche que précédemment, envisager l’analyse de trois catégories de talcs qui seront par la suite analysés avec des lactoses.

Nombre de Facteurs	Facteur(s) utilisé(s)	Spectres identifiés
1	Vecteur 1	1 / 20
1	Vecteur 2	19 / 20
1	Vecteur 3	18 / 20
2	Vecteurs 1 et 2	20 / 20
2	Vecteurs 1 et 3	19 / 20
2	Vecteurs 2 et 3	20 / 20
3	Vecteurs 1, 2 et 3	20 / 20

L’information permettant de discriminer les trois talcs est résumée dans les facteurs 2 et 3.

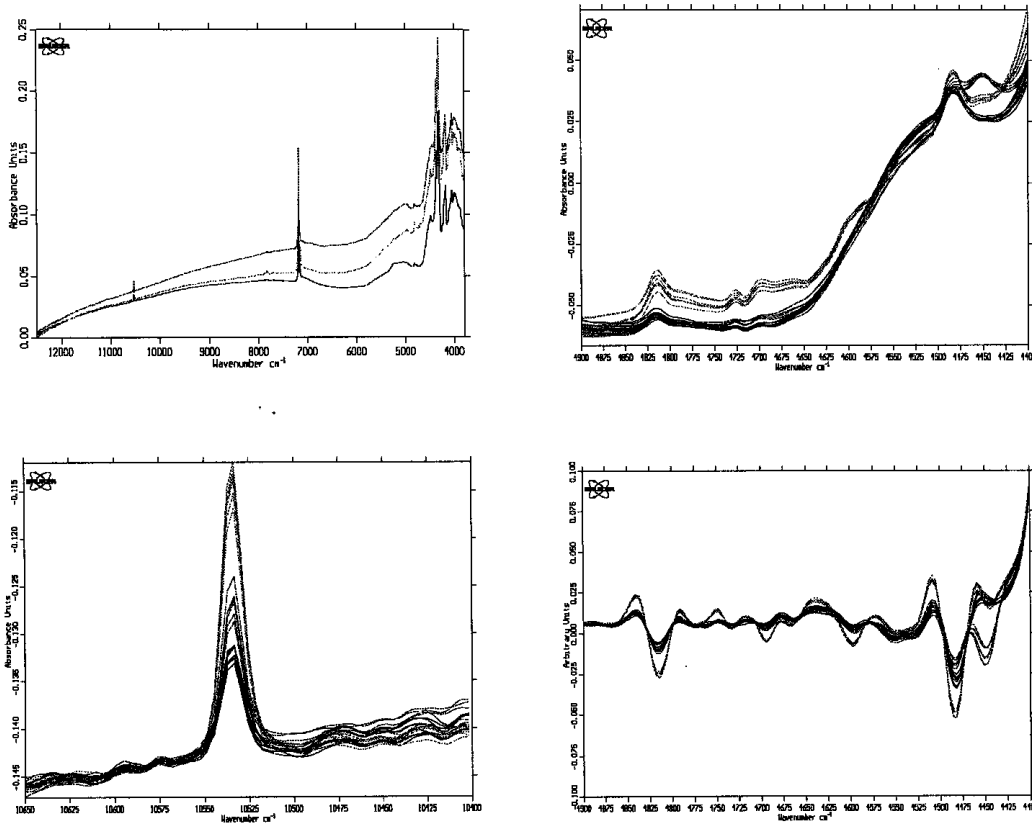


Figure 8. a) Spectres des 3 Talcs entre 12500 et 3800  $\text{cm}^{-1}$ ; b et c) agrandissement des zones 4900 – 4400  $\text{cm}^{-1}$  et 10650 – 10400  $\text{cm}^{-1}$ ; d) dérivée seconde entre 4900 et 4400  $\text{cm}^{-1}$ .

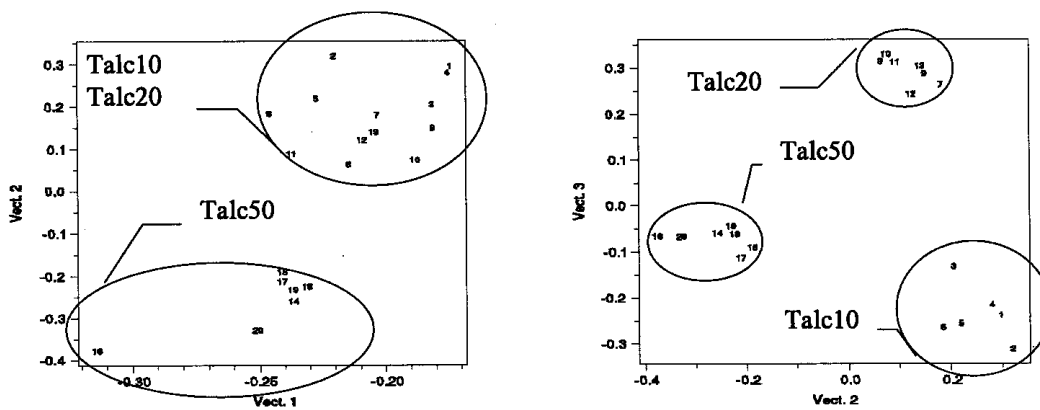


Figure 9. Représentation graphique des résultats de l'ACP : Vecteur1/Vecteur2 (gauche) ; Vecteur2/Vecteur3 (droite).

- Ajout des trois types de lactose.

Intuitivement on peut penser que pour séparer les 6 classes de produits, il faudra 1 facteur pour discriminer les talcs des lactoses, un facteur pour les talcs entre eux et un autre pour les lactoses. Un nombre de 3 facteurs pour l'ACP semble être prévisible et réaliste. Cette hypothèse sera confirmée par la validation effectuée sur les 41 spectres originaux. (Fig. 10).

Trois facteurs (2, 4 et 5) permettent de modéliser parfaitement l'ensemble des spectres et offrent une parfaite discrimination des trois types de talc et des trois lactoses.

Même si cette étude systématique semble fastidieuse, elle permet de mettre l'accent sur un certain nombre de points qui peuvent expliquer les problèmes que l'on peut rencontrer dans la mise en place d'un outil d'analyse discriminante.



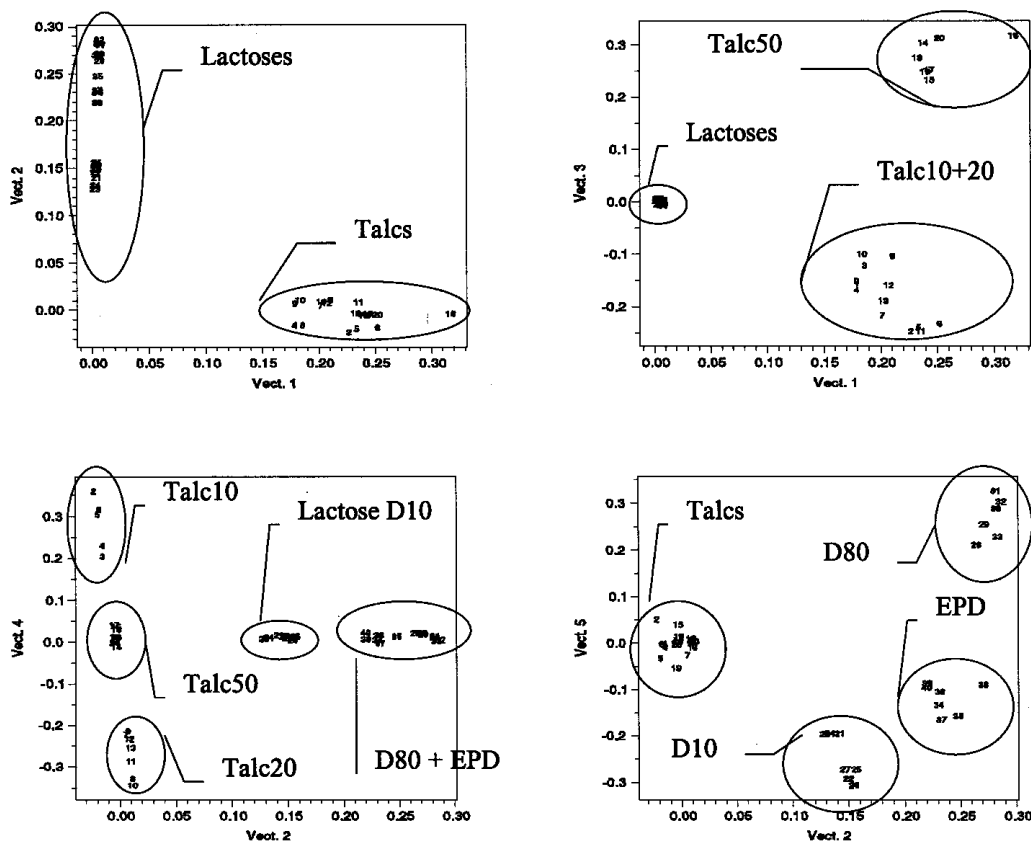


Figure 10. Résultats de l'ACP : a) Vecteur1/Vecteur2 ; b) Vecteur2/Vecteur3 ; c) Vecteur2/Vecteur4 ; d) Vecteur2/5.

Facteur(s) utilisé(s)	Résultat de validation	Produits identifiés
1	21	D80/D10/Talcs
2	38	D80/D10/T10/T20/T50
3	8	T50
1 + 2	19	D10/D80/Talcs
1 + 3	9	T50
1 + 4	20	Lactoses/T10/T20/T50
1 + 5	18	D80
2 + 3	26	D10/D80/T50
2 + 4	38	D10/D80/T10/T20/T50
2 + 5	22	D10/D80/EPD/Talcs
3 + 4	20	Lactoses/T10/T20/T50
3 + 5	25	D80/T50
4 + 5	37	D80/T10/T20/T50

Facteur(s) utilisé(s)	Résultat de validation	Produits identifiés
1+2+3	28	D10/D80/T50
1+2+4	38	D10/D80/T10/T20/T50
1+2+5	22	D10/D80/EPD/Talcs
1+3+4	20	Lactoses/T10/T20/T50
1+3+5	25	D80/T50
1+4+5	37	D80/T10/T20/T50
2+3+4	38	D10/D80/T10/T20/T50
2+3+5	29	D10/D80/EPD/T50
<b>2+4+5</b>	<b>41</b>	<b>D10/D80/EPD/T10/T20/T50</b>
3+4+5	37	D80/T10/T20/T50
2+4+6	2	aucun
1+2+4+5	41	D10/D80/EPD/T10/T20/T50
1+2+3+4	24	D10/T10

Alors que les facteurs sont classés par ordre mathématique décroissant, on peut facilement constater sur cet exemple que le premier facteur, pourtant mathématiquement le plus important, n'est pas le plus discriminant pour séparer les six produits ci-dessus. Le vecteur ayant le pouvoir discriminant le plus important vis-à-vis des propriétés qui nous intéressent est le deuxième. Le vecteur 3 ne semble rien apporter alors que le 4 sépare bien les talcs et le 5 les

lactoses. La validation associant les trois facteurs 2, 4 et 5 est la seule à être parfaite. Dans celle utilisant les quatre vecteurs 1, 2, 4 et 5 le premier n'apporte rien, il est superflu.

Deux remarques sont importantes à noter :

- le nombre, la forme et surtout l'ordre des facteurs sont systématiquement changés dès que l'on rajoute un nouveau produit. Ce n'est pas parce qu'avec ces six

produits initiaux il faut utiliser les vecteurs 2, 4 et 5 qu'il faudra les reprendre si on ajoute une nouvelle matière.

- les résultats obtenus en utilisant la méthode standard par calcul des distances sont nettement plus longs en temps de calcul et surtout moins bons. Il faut fortement réduire la gamme spectrale pour obtenir une discrimination correcte et la séparation des lactoses D80 et EPD reste plus délicate.
- la méthode par factorisation permet d'utiliser un nombre de spectres beaucoup plus important et donc de construire de très grandes bibliothèques.

### Conclusions

Les deux exemples détaillés ci-dessus ont permis de montrer qu'il est parfaitement possible de discriminer différents types de talcs (bruts, traités et de granulométrie différente) et que cette discrimination reste effective après adjonction de produits complémentaires à la base de données.

### Conclusions générales

Par ces deux exemples simples, nous avons voulu montrer les possibilités de la spectroscopie proche infrarouge pour le contrôle en routine de matières premières sur la base de produits qui ne sont pas actuellement analysés. En détaillant l'approche systématique et l'utilisation mathématique des outils actuellement disponibles, nous avons voulu montrer l'importance d'une grande rigueur méthodologique dans la

construction de l'outil analytique de contrôle par proche infrarouge.

La définition d'un cahier des charges complet, de la rigueur et de la méthode, une connaissance des outils mathématiques utilisés sont indispensables à la construction d'un outil de contrôle efficace. Comprendre et valider ses résultats sont une garantie pour la robustesse et la fiabilité de l'outil final. Contrairement à la spectroscopie dans le moyen infrarouge, le proche infrarouge demande un investissement complémentaire dans la réalisation de la méthode. Même si chaque environnement industriel nécessite sa propre méthode, cette méthode est parfaitement transférable d'un site à l'autre si le contexte d'utilisation est similaire.

Les potentialités du proche infrarouge sont très grandes. Il permet non seulement une systématisation des contrôles mais aussi une uniformisation de ce système de contrôle pour des sites industriels identiques.

### Références

- Brown, S. D. *Appl. Spectrosc.* **1995**, 49(12).  
 Burns, D. A.; Ciurczak, E. W. *Pratic. Spectrosc. Ser. 13.*  
 Bruker Opus/Ident Référence Manuel Vs 2.2.  
 Martens, H.; Naes, T. *Multivariate Calibration* Wiley Editions, 1989.  
 Serane, G. *Math. Phys. Appl.* **1965**.  
 Pharmacopée Européenne, 1997; p 42.  
 Lanteri, P.; Longerey, R. *Analisis Mag.* **1996**, 24(2).  
 Hailey, P.A. *European Pharm. Rev.* **1996**.  
 Boulou, J. -C. *Spectra Anal.* **1994**, 176.