

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété Intellectuelle
Bureau international



(43) Date de la publication internationale
7 juillet 2011 (07.07.2011)

PCT

(10) Numéro de publication internationale

WO 2011/080408 A2

(51) Classification internationale des brevets :
G01N 21/64 (2006.01) *G01N 33/02* (2006.01)

(74) Mandataires : CABINET ORES et al.; 36 rue de St Pétersbourg, F-75008 Paris (FR).

(21) Numéro de la demande internationale :
PCT/FR2010/000826

(81) États désignés (sauf indication contraire, pour tout titre de protection nationale disponible) : AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(22) Date de dépôt international :
8 décembre 2010 (08.12.2010)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
0906088 16 décembre 2009 (16.12.2009) FR
1001221 26 mars 2010 (26.03.2010) FR

(71) Déposant (pour tous les États désignés sauf US) :
SPECTRALYS INNOVATION; Biocitech, 102 avenue Gaston Roussel, F-93230 Romainville (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement) :
RIZKALLAH, Jad; Brumana, Beyrouth (LB).
BIRLOUEZ-ARAGON, Inès; 4 rue Hélène Boucher, F-95120 Ermont (FR).

(84) États désignés (sauf indication contraire, pour tout titre de protection régionale disponible) : ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Suite sur la page suivante]

(54) Title : METHOD AND SPECTROSCOPIC ANALYSIS APPLIANCE, ESPECIALLY FOR ANALYSING FOOD, WITH MULTI-CHANNEL TREATMENT OF SPECTRAL DATA

(54) Titre : PROCEDE ET APPAREIL D'ANALYSE SPECTROSCOPIQUE, EN PARTICULIER D'ALIMENTS, UTILISANT UN TRAITEMENT MULTIVOIES DES DONNEES SPECTRALES.

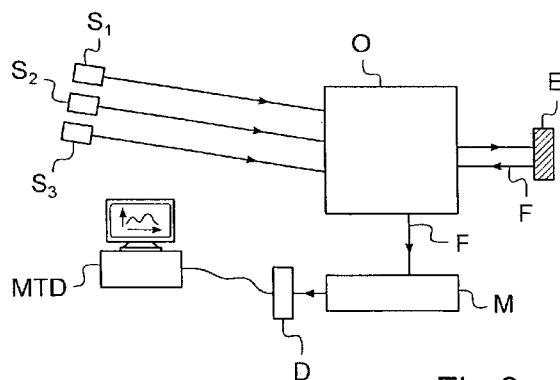


Fig.6

(57) Abstract : The invention relates to a method for the spectroscopic analysis of at least one sample, using a method for analysing spectroscopic data based on a multi-channel statistical model, said method comprising: the illumination of said or each sample to be analysed by a plurality of luminous excitation rays with respective wavelengths; the acquisition and the pre-treatment of frontal fluorescence spectra, each spectrum corresponding to a respective luminous excitation ray; for each sample, the calculation of a score vector by applying said multi-channel statistical model to the pre-treated spectra; and the determination of at least one parameter selected from a quality indicator of said or each sample and a parameter characterising a method that has been applied to said or each sample, from said score vector; the method being characterised in that the average spectral distance between the luminous excitation rays is at least 20 nm, preferably at least 50 nm, over a spectral range of at least 100 nm. The invention also relates to an appliance for implementing such a method.

(57) Abrégé :

[Suite sur la page suivante]

**Publiée :**

- *sans rapport de recherche internationale, sera republiée dès réception de ce rapport (règle 48.2.g)*

Procédé d'analyse spectroscopique d'au moins un échantillon mettant en œuvre une méthode d'analyse des données spectroscopiques basée sur un modèle statistique multivoies, le procédé comportant : l'éclairage dudit ou de chaque échantillon à analyser par une pluralité de rayonnements lumineux d'excitation à des longueurs d'onde respectives; l'acquisition et le prétraitement de spectres de fluorescence frontale correspondant chacun à un rayonnement lumineux d'excitation respectif; pour chaque échantillon, le calcul d'un vecteur de scores par application aux spectres prétraités dudit modèle statistique multivoies; et la détermination d'au moins un paramètre choisi parmi un indicateur de qualité dudit ou de chaque échantillon et un paramètre caractérisant un procédé auquel ledit ou chaque échantillon a été soumis, à partir dudit vecteur des scores; le procédé étant caractérisé en ce que l'écart spectral moyen entre lesdits rayonnements lumineux d'excitation est d'au moins 20 nm, et de préférence d'au moins 50 nm, sur une plage spectrale d'au moins 100 nm. Appareil pour la mise en œuvre d'un tel procédé.

**PROCEDE ET APPAREIL D'ANALYSE SPECTROSCOPIQUE, EN
PARTICULIER D'ALIMENTS, UTILISANT UN TRAITEMENT MULTIVOIES
DES DONNEES SPECTRALES**

L'invention porte sur un procédé et un appareil d'analyse spectroscopique d'au moins un échantillon, en particulier d'un aliment ou médicament, mettant en œuvre une méthode d'analyse des données spectroscopiques basée sur un modèle statistique multivoies. L'invention peut être appliquée en particulier, mais pas uniquement, à l'étude de l'évolution des propriétés nutritionnelles (teneur en vitamines...) et/ou toxicologiques (teneur en contaminants néoformés...) d'un aliment au cours de sa préparation ou conservation. Plus généralement, l'invention peut être appliquée à la détermination de tout indicateur de qualité d'un échantillon, et/ou de tout paramètre caractérisant un procédé auquel ledit échantillon a été soumis.

Les contaminants néoformés (CNF) sont des molécules nouvelles formées dans un aliment à partir des composants propres de cet aliment, du fait de l'action des procédés de transformation, et qui posent des problèmes de santé publique. Peu de laboratoires publics ou privés sont capables aujourd'hui de doser ces contaminants. De plus, si plusieurs méthodes ont été publiées pour le dosage de chacun des contaminants, aucune d'elles n'a réellement fait l'objet d'une standardisation et normalisation. En outre, ces analyses sont couteuses et le délai d'obtention des résultats est long (2 à 3 semaines). L'industrie nécessite des méthodes d'analyses plus rapides et moins couteuses pour assurer un contrôle de la conformité de ses procédés et de la sécurité des aliments qu'elle produit.

Les procédés de transformation alimentaire (cuisson, stérilisation, conservation) peuvent aussi influencer défavorablement les propriétés nutritionnelles des aliments, par exemple en diminuant leur taux de vitamines. La caractérisation de ces procédés et de leur influence sur les propriétés nutritionnelles des aliments pose des problèmes analogues à ceux, précédents, du dosage des contaminants néoformés.

Il est connu de réaliser des analyses, notamment spectroscopiques, des aliments en ayant recours aux méthodes de la chimiométrie, et en particulier à l'analyse multivoies. L'analyse multivoies est

l'extension naturelle de l'analyse multivariée lorsque les données sont arrangées dans des tableaux à trois voies ou plus. Elle se base sur l'utilisation de modèles statistiques telles que « PARAFAC » (« Parallel Factor», c'est à dire modèle à facteurs parallèles) et NPLS (« N-ways Partial Least Squares régression », c'est à dire régression par moindres carrés partiels à n-voies).
5 Ces méthodes, ainsi que leur application à l'analyse de produits alimentaires, sont décrites dans le document [Bro 1998].

Plus précisément, [Rizkallah 2007] décrit une méthode d'analyse d'échantillons non préparés (seulement broyés, si nécessaire) basée
10 sur l'application du modèle PARAFAC à des spectres de fluorescence frontale (« front-face fluorescence »). Cette méthode comporte l'éclairage d'un échantillon par une pluralité de rayonnements lumineux monochromatiques à des longueurs d'ondes respectives, et l'acquisition des spectres de fluorescence correspondants. Les rayonnements d'excitation échantillonnent
15 finement (plusieurs centaines de points) une plage spectrale couvrant le visible et le proche UV ; à leur tour, les spectres d'émission de fluorescence sont échantillonnés spectralement (également sur plusieurs centaines de points). Les données spectrales ainsi recueillies pour chaque échantillon sont agencées en une matrice de grandes dimensions dite « matrice excitation-émission »
20 (MEE), dont une dimension représente les longueurs d'onde d'émission et l'autre dimension représente les longueurs d'onde d'émission.

Les MEE correspondant à plusieurs échantillons d'étalonnage sont analysées par la méthode PARAFAC qui permet d'extraire des informations, appelées « facteurs PARAFAC » qui correspondent aux profils bilinéaires de fluorescence et à leurs intensités relatives. Puis, une régression multiple entre les intensités de fluorescence et les concentrations de contaminants néoformés (ou autres substances d'intérêt) mesurées chimiquement permet de construire un modèle d'étalonnage qui est utilisé ensuite pour prédire les concentrations de contaminants néoformés à partir des
25 MEE recueillies sur de nouveaux échantillons à analyser.

Cette méthode présente deux limites majeures :

- Premièrement, elle nécessite un équipement de coût élevé : un spectrofluorimètre équipé d'une lampe Xénon délivrant l'ensemble

des longueurs d'onde de l'UV proche au visible, et deux monochromateurs capables de faire défiler ces différentes longueurs d'onde à l'excitation comme à l'émission pour les photons émis par l'échantillon.

Deuxièmement, sa mise en œuvre est longue car, même si
5 on évite la préparation de l'échantillon, l'acquisition de chaque EEM prend 15 à 45 min selon la résolution spectrale recherchée. L'analyse de ces matrices de données, comprenant un nombre considérable de variables, est également très lourde malgré l'automatisation des algorithmes de traitement des données.

Une méthode similaire est décrite par [Rizkallah et al., 2008].

10 Le document [Nahorniak 2003] étudie l'influence de la résolution spectrale (et donc de la taille de la matrice EEM) sur l'erreur de prédiction de la concentration d'un fluorophore par application de la méthode PARAFAC. Dans ce document on considère des solutions diluées et de composition simple et connue ; en outre, toutes les mesures sont effectuées
15 avec une résolution spectrale élevée, de 2 nm/pixel, la réduction de résolution étant obtenue en calculant des moyennes spectrales.

Il résulte de cette étude que l'erreur de prédiction est minimisée – et vaut environ 0,5-1% – pour une résolution en excitation de 6 à 10 nm/pixel, puis augmente fortement pour des résolutions inférieures (augmentation d'un facteur 2 à 6 en passant à une résolution de 20 nm/pixel). Il semble donc difficile d'utiliser des résolutions inférieures à 20 nm/pixel.

L'invention vise à procurer une méthode d'analyse plus simple, plus rapide, nécessitant un équipement plus simple et moins couteux, et par conséquent mieux adaptée aux exigences industrielles.

25 Conformément à l'invention un tel but est atteint par un procédé d'analyse spectroscopique d'au moins un échantillon mettant en œuvre une méthode d'analyse des données spectroscopiques basée sur un modèle statistique multivoies, le procédé comportant :

30 a) l'éclairage dudit ou de chaque échantillon à analyser par une pluralité de rayonnements lumineux d'excitation à des longueurs d'onde respectives ;

b) l'acquisition de spectres de fluorescence frontale dudit ou de chaque échantillon, correspondant chacun à un rayonnement lumineux d'excitation respectif ;

c) un prétraitement des spectres de fluorescence acquis ;

5 d) pour chaque échantillon, le calcul d'un vecteur de *scores* (*ou intensité lumineuse*) par application aux spectres prétraités dudit modèle statistique multivoies, identifié par un vecteur de *loadings* (*ou spectre*) d'excitation de la fluorescence et par un vecteur de *loadings* (*ou spectre*) d'émission ; et

10 e) la détermination d'au moins un paramètre choisi parmi un indicateur de qualité dudit ou de chaque échantillon et un paramètre caractérisant un procédé auquel ledit ou chaque échantillon a été soumis, à partir dudit vecteur des *scores* ;

15 le procédé étant caractérisé en ce que l'écart spectral moyen entre lesdits rayonnements lumineux d'excitation est d'au moins 20 nm (de préférence d'au moins 50 nm) sur une plage spectrale d'au moins 100 nm.

20 Avantageusement, le nombre de rayonnements lumineux d'excitation, et de spectres de fluorescence correspondants pour chaque échantillon, peut être compris entre deux et six, et de préférence entre trois et cinq, sur une plage spectrale d'au moins 100 nm, et de préférence d'au moins 150 nm.

25 Cette méthode est innovante du fait qu'elle applique des outils d'analyse multivoies comme le modèle PARAFAC, normalement utilisés pour des MEE de larges dimensions, à un ensemble constitué par un nombre limité de spectres (généralement deux à six, de préférence entre trois et cinq), échantillonnant de manière très grossière la plage spectrale d'excitation des échantillons. Cette plage est relativement large (au moins 100, voire 150 nm, ou plus) parce que la composition physico-chimique desdits échantillons est largement inconnue, donc la position exacte des spectres d'absorption des 30 fluorophores contenus dans le dit échantillon l'est aussi.

Les présents inventeurs ont découvert, de manière surprenante, que les modèles multivoies continuent à être applicables et fournissent des résultats exploitables même en utilisant une résolution

spectrale en excitation aussi grossière, et un nombre aussi restreint de longueurs d'onde d'excitation.

Il importe de souligner que d'un point de vue physique, les « *loadings* » représentent les spectres d'émission ou d'excitation de chaque facteur, et les « *scores* » les intensités desdits facteurs. Un « facteur PARAFAC » représente la contribution d'un fluorophore, ou d'un mélange de fluorophores, commune aux échantillons analysés. Il faut aussi souligner que les spectres de fluorescence sont déformés par l'interaction entre les photons d'excitation et les photons émis avec la matrice alimentaire et que la condition de trilinéarité généralement admise pour l'application du modèle PARAFAC n'est pas satisfaite. Néanmoins, les inventeurs ont montré que ce modèle est utilement appliqué à de telles matrices EEM pour en décrire la composition en fluorophores.

Il faut bien comprendre que le procédé de l'invention ne vise pas (au moins : pas nécessairement) à déterminer la teneur des fluorophores eux-mêmes. Ce procédé exploite simplement des corrélations, en général non-linéaires ou multilinéaires, entre les évolutions des spectres d'excitation et d'émission de fluorescence et certains paramètres de qualité de l'échantillon (teneur en nutriments et/ou contaminants qui, le plus souvent, ne sont pas fluorescents ; charge microbienne...) ou certains paramètres caractérisant un procédé auquel ledit échantillon a été soumis (valeur stérilisatrice, pasteurisatrice ou cuisatrice, ou un produit temps-température, par exemple).

Une comparaison avec le document précité [Nahorniak 2003] est particulièrement intéressante. Ce document se rapporte à un cas particulièrement favorable, à savoir l'analyse de solutions de composition simple et connue. Au contraire, l'invention se rapporte à l'analyse par fluorescence frontale d'échantillons complexes et fortement diffusants, dont la composition est largement inconnue. En outre, dans le document [Nahorniak 2003], les spectres sont acquis à haute résolution spectrale d'excitation, puis moyennés, ce qui est de nature à réduire l'erreur de prédiction. Au contraire, dans le cas de l'invention, seulement deux à six spectres sont acquis, avec une résolution spectrale en excitation extrêmement faible, inférieure à 20 nm/pixel voire à 50 nm/pixel. Malgré cela, un choix opportun des paramètres du modèle

multivoies et/ou un prétraitement opportun des données permettent de limiter l'erreur de prédiction à des valeurs tout à fait acceptables, typiquement inférieures à 15% voire à 10%.

Un procédé selon l'invention peut comporter également une
5 phase préliminaire d'étalonnage comportant :

- i) l'éclairage d'une pluralité d'échantillons d'étalonnage par lesdits rayonnements lumineux d'excitation ;
- ii) l'acquisition des spectres de fluorescence frontale des échantillons, correspondant auxdits rayonnements lumineux d'excitation ;
- 10 iii) un prétraitement des spectres de fluorescence acquis ;
- iv) la détermination, par une méthode itérative, desdits vecteurs de *loadings* d'excitation et d'émission du modèle multivoies, ainsi que d'un vecteur de scores pour chaque échantillon d'étalonnage ; et
- v) la détermination d'une fonction de régression liant lesdits vecteurs de scores aux valeurs, connues, dudit ou de chaque paramètre pour lesdits échantillons d'étalonnage.

Les spectres de fluorescence peuvent être des spectres d'autofluorescence ou, dans certains cas, des spectres de fluorescence induite par un marqueur ajouté à l'échantillon. Par exemple, l'estimation de la charge microbienne d'un échantillon peut être grandement facilitée par l'utilisation d'une sonde fluorescente.

Avantageusement, les spectres de fluorescence peuvent être prétraités par soustraction d'une contribution due à la diffusion Rayleigh du premier ordre du rayonnement lumineux d'excitation, ladite contribution étant calculée au moyen d'un modèle linéaire généralisé. En particulier, ledit modèle linéaire généralisé peut être déterminé à partir d'une première portion du spectre à prétraiter, comprenant seulement ladite contribution due à la diffusion Rayleigh, et est utilisé pour prédire la contribution de diffusion Rayleigh dans une deuxième portion dudit spectre dans laquelle elle se superpose à une contribution de fluorescence. La soustraction de la diffusion Rayleigh est nécessaire aussi dans les procédés d'analyse connus de l'art antérieur, mais devient critique dans le procédé de l'invention. L'utilisation d'un modèle linéaire généralisé, au lieu des techniques connues de remplacement par des zéros ou

des valeurs manquantes, s'avère particulièrement avantageuse lorsqu'un nombre très faible de longueurs d'onde d'excitation est utilisé, comme dans le cas de l'invention.

Ensuite, les spectres de fluorescence peuvent aussi être prétraités en les normalisant. Plusieurs échantillons, d'étalonnage et/ou à analyser, peuvent aussi être prétraités en effectuant une correction multiplicative de dispersion (MSC : « multiplicative scatter correction »).

Selon un mode de réalisation préféré de l'invention, ledit modèle statistique multivoies peut être un modèle PARAFAC, les spectres de fluorescence de chaque échantillon étant représentés sous forme concaténée. Dans ce cas, les vecteurs de *loadings* dudit modèle PARAFAC peuvent être déterminés par une méthode itérative qui est arrêtée lorsque la diminution de la fonction de perte due à la dernière itération descend au-dessous d'une valeur de seuil (« paramètre de convergence ») comprise entre 10^{-4} et 10^{-2} . Cette valeur doit être comparée à celle utilisée généralement de 10^{-6} . En effet, l'utilisation d'un paramètre de convergence plus grand facilite la convergence du modèle PARAFAC lorsqu'un faible nombre de longueurs d'onde d'excitation est utilisé.

En variante, ledit modèle statistique multivoies peut être un modèle NPLS, les spectres de fluorescence de chaque échantillon étant représentés sous forme concaténée. Dans ce cas, les spectres de fluorescence peuvent utilement être prétraités par correction orthogonale du signal.

Avantageusement, lesdits rayonnements lumineux d'excitation peuvent comprendre : un premier rayonnement à une longueur d'onde comprise entre 270 et 300 nm ; un deuxième rayonnement à une longueur d'onde comprise entre 300 et 360 nm ; et un troisième rayonnement à une longueur d'onde comprise entre 400 et 500 nm.

Ledit ou au moins un dit paramètre à déterminer peut être choisi, notamment, parmi : la teneur dudit ou de chaque échantillon en une substance d'intérêt ; une charge microbienne dudit ou de chaque échantillon ; et une valeur stérilisatrice, pasteurisatrice ou cuisatrice, ou un produit temps-température, d'un procédé auquel ledit ou chaque échantillon a été soumis.

Les notions de valeur stérilisatrice, pasteurisatrice ou cuisatrice sont bien connues par elles-mêmes. On considère une étape d'un procédé thermique au cours de laquelle l'échantillon est maintenu à une température T_i (en °C, à cœur de l'échantillon) pendant un temps t_i . La valeur de 5 pasteurisation de cette étape vaut :

$$VP_i = t_i \times 10^{\frac{T_i - 70}{Z}}$$

sa valeur stérilisatrice vaut :

$$VS_i = t_i \times 10^{\frac{T_i - 121,1}{Z}}$$

et sa valeur cuisatrice vaut :

$$VC_i = t_i \times 10^{\frac{T_i - 100}{Z}}$$

10

où le paramètre Z dépend du microorganisme considéré (pour les valeurs pasteurisatrice et stérilisatrice) ou de la nature de l'échantillon (pour la valeur cuisatrice).

Pour un procédé multi-étapes on prend la somme des valeurs 15 pasteurisatrices ou stérilisatrices des différentes étapes.

Les échantillons à analyser et d'étalonnage peuvent être, notamment, des échantillons d'un produit choisi parmi un aliment et un médicament.

Un autre objet de l'invention est l'utilisation d'un procédé tel que 20 décrit ci-dessus pour mesurer l'évolution des propriétés nutritionnelles, microbiologiques et/ou toxicologiques d'un aliment au cours de sa préparation ou conservation.

Encore un autre objet de l'invention est un appareil d'analyse spectroscopique d'au moins un échantillon comportant : un ensemble de 25 sources lumineuses pour éclairer ledit ou chaque échantillon à analyser par des rayonnements lumineux d'excitation respectifs, présentant des longueurs d'onde différentes avec un écart spectral moyen d'au moins 20 nm sur une plage spectrale d'au moins 100 nm ; un moyen d'acquisition des spectres de fluorescence frontale émis par ledit ou de chaque échantillon lorsqu'il est 30 éclairé par lesdits rayonnements lumineux d'excitation ; et un moyen de

traitement des spectres de fluorescence acquis, adapté pour mettre en œuvre un procédé tel que décrit ci-dessus. De préférence un tel appareil peut comporter entre deux et six (et de préférence entre trois et cinq) desdites sources lumineuses.

5 Avantageusement, un tel appareil peut comporter : une première source lumineuse émettant un rayonnement à une longueur d'onde comprise entre 270 et 300 nm ; une deuxième source lumineuse émettant un rayonnement à une longueur d'onde comprise entre 300 et 360 nm ; et une troisième source lumineuse émettant un rayonnement à une longueur d'onde
10 comprise entre 400 et 500 nm.

D'autres caractéristiques, détails et avantages de l'invention ressortiront à la lecture de la description faite en référence aux dessins annexés donnés à titre d'exemple et qui représentent, respectivement :

- La figure 1, trois spectres « bruts » de fluorescence d'un échantillon de chicorée torréfiée, correspondant à trois rayonnements d'excitation à des longueurs d'ondes différentes ;
- Les figures 2A, 2B, 2C, trois graphiques illustrant l'opération de soustraction de la diffusion de Rayleigh par la méthode utilisant un modèle linéaire généralisé ;
- Les figures 3A, 3B, 3C, trois graphiques comparant ladite méthode de soustraction de la diffusion de Rayleigh (figure 3C) à deux autres méthodes connues de l'art antérieur : le remplacement par des zéros (figure 3A) et le remplacement par des valeurs manquantes (figure 3B) ;
- La figure 4, une illustration schématique de la notion de concaténation des spectres ;
- La figure 5, une illustration schématique de la méthode PARAFAC ;
- La figure 6, le schéma de principe d'un appareil d'analyse selon un mode de réalisation de l'invention ;
- Les figures 7 et 8, des graphiques illustrant l'application d'une méthode selon un premier mode de réalisation de l'invention, basée sur un modèle PARAFAC, à la prédiction de la teneur en acrylamide de différents échantillons de chicorée ;

- Les figures 9A, 9B, 10A et 10B, des graphiques illustrant l'application d'une méthode selon un deuxième mode de réalisation de l'invention, basée sur un modèle NPLS, à la prédition de la teneur en vitamine C de différents échantillons de haricots verts.

5 Le procédé de l'invention utilise le signal de fluorescence émis à la surface de l'aliment après illumination par des faisceaux lumineux à des longueurs d'onde déterminées du domaine UV-visible (approximativement : 250 – 750 nm). Ce signal est analysé par des méthodes chimiométriques qui permettent d'extraire l'information qui se trouve corrélée aux paramètres que
10 l'on cherche à mesurer. L'existence d'une telle corrélation est déduite du fait que, au cours de la production agricole, la conservation et la transformation, la fluorescence intrinsèque des constituants naturels de l'aliment (vitamines, protéines et autres constituants naturels ou ajoutés intentionnellement ou non), ainsi que leur réflectance, évoluent, alors que dans le même temps, de
15 nouveaux signaux apparaissent du fait de la formation de nouvelles molécules (dont les contaminants néoformés). On parle de fluorescence néoformée ou acquise selon que les fluorophores sont formés de novo ou proviennent de l'environnement. L'évolution conjointe des signaux natifs (SN), néoformés (SNF) et nouvellement acquis (SNA) est corrélée de manière robuste aux
20 modifications physiques, physico-chimiques, chimiques ou microbiologiques de l'aliment, en particulier aux changements des paramètres de qualité, induits au cours de la production, la conservation et/ou la transformation. Les facteurs de changement qui influent sur la qualité de l'aliment sont les rayonnements ultraviolets oxydants, la destruction de microorganismes ou, au contraire, le
25 développement de certains d'entre eux pouvant entraîner la synthèse de mycotoxines, l'intervention humaine sur les cultures (engrais, pesticides ...), ou l'application de procédés modifiant la température, la pression ou tout autre paramètre physique au sein de l'aliment et qui induisent par voie de conséquence une modification de la composition physico-chimique et des
30 paramètres de qualité.

Les rayonnements lumineux d'excitation présentent des longueurs d'onde choisies de manière à explorer le spectre UV-visible le plus largement possible. De manière générale, on peut choisir a priori :

- une longueur d'onde comprise entre 270 et 300 nm permettant d'exciter le tryptophane, les phénols comme l'acide chlorogénique ou l'hydroxytyrosol ou encore la vitamine E, molécules émettant dans l'UV ;
- 5 • une longueur d'onde comprise entre 400 et 450 ou 500 nm permettant d'exciter la riboflavine, les porphyrines et la chlorophylle, molécules émettant dans le visible (500-700 nm) ;
- une ou deux longueurs d'onde comprises entre 300 et 360 voire 10 400 nm sont introduites pour exciter les fluorophores néoformés, essentiellement des produits de Maillard et des produits de peroxydation lipidique, mais également les mycotoxines émettant dans l'UV lointain-proche visible (400-500 nm).

Ces longueurs d'onde seront modifiées en cas d'applications autres qu'à l'industrie alimentaire. En tout cas, pour plus de précision, il est 15 possible de choisir les longueurs d'onde les plus proches possibles des maxima des loadings représentant le vecteur excitation obtenu par la décomposition PARAFAC d'une matrice MEE complète obtenue avec un fluorimètre de laboratoire sur d'un lot d'échantillons représentatifs.

Le nombre de longueurs d'onde utilisées est compris entre 2 et 20 6, avantageusement entre 3 et 5, et de préférence égal à 5, permettant d'exciter tour à tour un nombre correspondant de groupes de fluorophores parmi ceux décrits plus haut. L'intensité du rayonnement d'excitation est choisie de sorte à ce que l'énergie d'émission de fluorescence de ces fluorophores soit significativement modifiée au cours des étapes de transformation que l'on 25 cherche à caractériser. Par exemple, l'intensité doit être suffisante pour que l'apparition de mycotoxines liée au développement de champignons sur une graine au cours de sa transformation modifie sensiblement l'émission de fluorescence recueillie sur le détecteur à partir d'un échantillon non contaminé. Elle doit également permettre de mettre en évidence l'application de tel 30 traitement thermique caractérisé par un couple temps-température ; ou encore l'application d'un lavage, blanchiment ou traitement de décontamination.

La figure 6 montre un schéma très simplifié d'un appareil pour la mise en œuvre du procédé de l'invention. Cet appareil comprend trois

sources lumineuses S_1 , S_2 , S_3 émettant chacune un faisceau de rayonnement monochromatique à une longueur d'onde différente dirigée de sorte à éclairer l'échantillon E. Le signal de fluorescence F (en fait, mélange de fluorescence et diffusion Rayleigh du 1^{er} et parfois du 2nd ordre) émis par cet échantillon est 5 transporté par une fibre optique vers un spectromètre qui décompose le rayonnement lumineux émis en spectre. Les spectres acquis sont traités par un moyen de traitement des données MTD (typiquement un ordinateur programmé d'une manière opportune) qui permet d'en extraire l'information chimiométrique recherchée.

10 Les sources lumineuses peuvent être, typiquement, des diodes électroluminescentes, voire des lasers (de préférence à semiconducteur) si des intensités plus importantes sont requises.

L'analyse des données spectroscopiques, effectuée par l'ordinateur MTD, comporte quatre étapes principales :

- 15 - le prétraitement des spectres ;
- la détermination des vecteurs des «loadings» du modèle multivoies à l'aide des spectres des échantillons d'étalonnage ;
- la détermination des vecteurs des «scores» en appliquant ce modèle aux spectres des échantillons à analyser ; et
20 - la détermination d'une information de nature chimique (teneur en une ou plusieurs substances d'intérêt), physique (produit temps-température, valeur stérilisatrice, pasteurisatrice ou cuisatrice), physico-chimique ou microbiologique (charge microbienne) à partir desdits vecteurs des «scores».

25 On considère d'abord le prétraitement des spectres, en référence à un exemple spécifique, dans lequel un échantillon de chicorée torréfiée est éclairé successivement par trois rayonnements d'excitation à 280, 340 et 429 nm. Chacun des trois spectres de fluorescence est constitué de 1515 valeurs d'intensité spectrale pour autant de longueurs d'ondes λ différentes. La résolution spectrale est de 0,25 nm/pixel, mais elle peut être 30 divisée par 5, voire plus, sans qu'une dégradation des résultats du procédé de l'invention ne soit observée.

Les spectres « bruts » (figure 1) sont dominés par la diffusion de Rayleigh du premier ordre, à la longueur d'onde des rayonnements d'excitation (280, 340 et 429 nm). Le spectre de chaque signal de fluorescence, en effet, est partiellement superposé au rayonnement d'excitation diffusé, qui 5 est beaucoup plus intense. Ce problème est connu, et est généralement résolu en remplaçant les valeurs d'intensité correspondant à la région spectrale de superposition par des zéros (figure 3A) ou des valeurs manquantes (figure 3B). Sur ces figures les références F_{280} , F_{340} et F_{429} représentent les spectres de fluorescence relatifs aux excitations à 280, 340 et 429 nm respectivement, 10 dans lesquels les intensités des régions spectrales de chevauchement avec la diffusion de Rayleigh ont été remplacées par des zéros ou des valeurs manquantes.

Ces méthodes présentent des inconvénients, car le remplacement par des zéros engendre des artefacts et donc fausse l'analyse 15 en créant des variances artificielles. Le remplacement par valeurs manquantes engendre des difficultés au niveau de la convergence des modèles multivoies utilisés pour l'analyse des données. De plus, la présence des valeurs manquantes empêche l'utilisation des prétraitements de standardisation (voir ci-après), qui ne peuvent être appliqués qu'à des valeurs réelles.

20 Ces problèmes sont acceptables dans le cas des techniques conventionnelles, basées sur l'utilisation de MEE de grandes dimensions, mais deviennent facilement rédhibitoires dans un procédé tel que celui de l'invention, basé sur l'analyse de données très faiblement résolues en longueur d'onde d'excitation.

25 Pour cette raison, l'invention propose de supprimer la contribution de la diffusion de Rayleigh du premier ordre par une technique innovante qui se base sur la prédiction de la région de diffusion qui chevauche la fluorescence via un modèle linéaire généralisé (GLZ) avec une fonction de liaison log [Davidson 1998 ; Rizkallah 2007]. Ce modèle est calibré sur une 30 région du spectre dans laquelle la contribution de la fluorescence est négligeable, et l'intensité spectrale est attribuée exclusivement à la diffusion (référence RD sur les figures 2A et 2B).

Le modèle linéaire généralisé est de la forme:

$$f(\mu_y) = b_0 + b_1 x$$

Dans l'équation, $f(\mu_y)$ est la fonction de liaison de μ_y , valeur attendue de y , y étant le vecteur des intensités de diffusion de Rayleigh qui ne se chevauchent pas avec la fluorescence, tandis que x est un vecteur d'indices 5 (1,2,3, etc.) de même taille que y .

Grace au modèle généralisé, des relations non linéaires (entre x et y) peuvent être modélisées via la fonction de liaison. Le modèle généralisé peut être utilisé pour modéliser des variables dépendantes ayant des distributions appartenant à la famille exponentielle (Normale, Gamma, Poisson, 10 etc.). La régression linéaire multiple est un cas spécial du modèle GLZ qui correspond à une fonction de liaison égale à la fonction identité et à une variable dépendante (y) présentant une distribution normale.

Les paramètres b_i du modèle GLZ sont estimés par la méthode statistique de maximisation de la vraisemblance (L) :

$$15 \quad L = F(Y, \text{modèle}) = \prod_{i=1}^n p[y_i, b_i].$$

$p[y_i, b_i]$ étant la probabilité de y_i conditionnée par b_i .

Le but est trouver les paramètres qui donnent la plus grande probabilité (densité jointe) de réalisation de y pour toutes les observations. Une estimation itérative (algorithme de Fisher, qui est une méthode quasi-20 Newtonienne) est utilisée pour trouver les paramètres b_i en maximisant L :

$$\frac{\partial \log(L)}{\partial b_i} = 0$$

Une fois les b_i estimés, le modèle GLZ est appliqué sur les indices de la diffusion correspondant à la région spectrale superposée avec la fluorescence (fig2B – région de prédiction RP) pour prédire les intensités de la 25 diffusion « pure ». Après cette prédiction, les spectres de diffusion complets (partie réelle et prédictive) sont soustraits des MEE pour obtenir le spectre de fluorescence pure SF (figure 2C – à comparer avec la figure 2A montrant le spectre de fluorescence partiellement superposé au spectre de la lumière diffusée).

30 Les opérations successives ne portent plus sur les trois spectres de fluorescence considérés individuellement, mais sur les spectres

concaténés, c'est à dire disposés l'un à la suite de l'autre dans une colonne unique (voir la figure 4).

Les spectres concaténés peuvent être soumis selon les cas à :

- Une simple normalisation par division de chaque valeur d'émission de chaque spectre concaténé, par la norme de ce même vecteur X_i .
5 Cette normalisation est appliquée à chaque échantillon ($i=1$ à I échantillons), soit :

$$x_{ij}(\text{normalisé}) = \frac{x_{ij}}{\|x_i\|}$$

où j est le nombre de longueurs d'ondes du spectre d'émission concaténés (j= 1 à 4545) et $\|x_i\| = \sqrt{\sum_j |x_{ij}|^2}$; ou

- Une correction par « *Multiplicative scatter correction* » (correction multiplicative de la dispersion) qui consiste à réaliser une régression entre le vecteur x_i (le spectre d'émission concaténé de chaque échantillon i) et le vecteur concaténé moyen de tous les échantillons ($i=1$ à I échantillons), puis à retrancher l'ordonnée à l'origine et à diviser par la pente respective chaque intensité du spectre concaténé d'émission pour chaque échantillon, comme décrit ci-dessous:

$$x_i = a_i + b_i \bar{x}_i$$

$$x_{ij}(\text{MSC}) = \frac{x_{ij} - a_i}{b_i}$$

Voir à ce propos : [Bro 1998 ; Dhanoa 1994]

Après avoir acquis et prétraité les spectres de fluorescence d'un certain nombre d'échantillons d'étalonnage, il est possible de déterminer le modèle statistique multivoies qui sera utilisé pour l'analyse d'autres échantillons, à caractériser. Cela comporte le calcul des vecteurs de « *loadings* » de ce modèle.

On considère d'abord le cas d'un modèle multilinéaire de type « PARAFAC » [Bro 1998]. Le principe, illustré par la figure 5, est de décomposer une structure à trois voies (cube de données) X en une somme de

produits externes de trois vecteurs (« triades ») a_i , b_j , c_k , plus un résidu E, lui aussi sous la forme d'un « cube de données ». Les trois voies constituant le « cube » X sont : les échantillons, les rayonnements d'excitation, les longueurs d'onde d'émission. On peut donc écrire :

$$5 \quad x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk},$$

Où : « i » est l'index des échantillons, « j » celui des rayonnements d'excitation, « k » celui des longueurs d'onde d'émission, « f » celui des F facteurs de décomposition PARAFAC.

10 La concaténation des spectres permet d'écrire la décomposition PARAFAC sous forme matricielle :

$$X_{IJK} = A(C \otimes B)^T;$$

où :

- I, J et K sont les nombres d'échantillons, de longueurs 15 d'onde d'excitation et de longueurs d'onde d'émission respectivement ;

- X_{IJK} est la matrice des fluorescences concaténées de tous les échantillons ;

- B_{JF} et C_{KF} sont les vecteurs (à JF et KF éléments, respectivement) « loadings » d'excitation et d'émission (profils bilinéaires de fluorescence) ;

- A_{IF} est le vecteur (à IF éléments) des « scores » (ou intensités des loadings spectraux) ;

- le symbole \otimes représente le produit tensoriel de Khatri-Rao ;

et

25 - l'exposant T indique la transposée de la matrice colonne.

Les vecteurs des loadings sont calculés à partir d'échantillons d'étalonnage choisis de manière à être représentatifs de la variabilité totale attendue au sein du groupe des échantillons à prédire subséquemment. En d'autres termes, le modèle est empirique ; il n'est donc valide que pour des 30 échantillons similaires à ceux qui ont servi à l'étalonnage.

Les paramètres A, B et C du modèle PARAFAC peuvent être calculés par la méthode d'alternance des moindres carrés (méthode itérative

non linéaire). Dans cette méthode une première estimation du vecteur A est calculée conditionnellement sur des valeurs initiales aléatoires attribuées à B et C pour minimiser la somme des carrés des résidus. Le paramètre B est ensuite actualisé en utilisant l'estimation de A, puis le paramètre C est actualisé en 5 utilisant la nouvelle valeur de B ; et ainsi de suite. Chaque actualisation itérative de A, B et C améliore donc la solution (baisse de la surface d'erreur). L'algorithme converge lorsque l'amélioration de la solution au niveau d'une itération devient très petite (par défaut, ce critère est de 10^{-6}).

La convergence du modèle PARAFAC pour des données 10 faiblement résolues et qui, de plus, ne sont pas véritablement trilinéaires, peut être problématique [Bro 1998 ; Harshman 1984 ; Rizkallah 2007]. En effet, la surface d'erreur du modèle peut contenir des minima locaux (points plus bas que leur voisinage mais plus élevés que le vrai minimum de la surface), des points col, des terrains plats, et des vallées très étroites qui peuvent ralentir ou 15 bien empêcher la convergence de l'algorithme.

Les présents inventeurs ont remarqué qu'en imposant une limitation sur le nombre d'itérations (par exemple 30 itérations maximum) et/ou en augmentant le critère de convergence (10^{-2} ou 10^{-3} au lieu de 10^{-6}) on améliore significativement le modèle au niveau des paramètres (*loadings* et 20 *scores*), comme cela sera montré ci-après à l'aide de la figure 7. En effet, un nombre excessif d'itérations peut dégrader la pertinence du modèle.

La convergence peut être facilitée en imposant une contrainte de non négativité puisqu'on sait a priori que les spectres de fluorescence ainsi que leurs intensités relatives ne peuvent pas prendre des valeurs négatives.

25 Enfin, il a été remarqué que le prétraitement de suppression de la diffusion de Rayleigh par application d'un modèle GLZ facilite la convergence du procédé itératif de calcul des vecteurs de *loadings* et de *scores* du modèle PARAFAC.

Dans le cas de données parfaitement multilinéaires, le modèle 30 PARAFAC n'admet en théorie qu'une solution unique. Cependant, dans le cas d'une MEE limitée à 3 excitations, et lorsque la fluorescence est recueillie en surface, la trilinéarité de la matrice MEE est fortement perturbée.

De ce fait, la solution n'est plus unique et il est nécessaire de développer des critères de choix du nombre de facteurs ainsi que du modèle final.

Ce choix est guidé par plusieurs critères connus en soi [Bro 5 1998 ; Harshman1984]:

- La vérification de la conformité entre les paramètres spectraux obtenus (vecteurs B et C) et la connaissance a priori des fluorophores présents dans l'échantillon analysé.

- Un critère connu sous le nom de CORCONDIA [Bro 1998] 10 qui sert à vérifier le pourcentage de déviation du modèle par rapport à un modèle multi-linéarité parfait.

- L'étude du pourcentage de variance des données de fluorescence expliqué par le modèle.

- L'étude de la structure des résidus, qui doivent être 15 aléatoires.

- L'étude de la structure des scores (vecteur A) en termes de cohérence de l'évolution des scores par rapport à ce qui est attendu de l'application de la transformation sur le lot d'échantillons d'étalonnage.

- La répétabilité et la reproductibilité des scores, etc.

20 Les scores ainsi obtenus par la décomposition PARAFAC des MEE des échantillons d'étalonnage sont ensuite utilisés pour construire une régression multiple (linéaire ou généralisée) sur les concentrations du ou des paramètres de qualité à mesurer. Les coefficients de la régression ainsi que les vecteurs B et C sont stockés dans une base de données du modèle.

25 Le modèle est validé par application à de nouveaux échantillons connus pour lesquels les valeurs prédictes sont comparées aux valeurs mesurées par les méthodes de référence.

Les MEE obtenues sur de nouveaux échantillons sont 30 prétraités (élimination de Rayleigh et standardisation) de la même façon que les MEE des échantillons d'étalonnage. Le modèle PARAFAC et le modèle de régression sont appliqués sur les nouveaux échantillons en utilisant les coefficients stockés (B, C et les coefficients de régression) obtenus à partir des échantillons d'étalonnage. Dans le cas de l'utilisation d'une normalisation de

type MSC, le spectre moyen des échantillons d'étalonnage fournit la référence pour corriger les spectres des nouveaux échantillons.

A partir du cube des matrices prétraitées des nouveaux échantillons à analyser, les nouveaux scores sont calculés sur la base des 5 vecteurs « *loadings* » B et C comme suit:

$$\mathbf{A}_{\text{nouveau}} = (\mathbf{B} \otimes \mathbf{C})^+ * \mathbf{X}_{\text{nouveau}};$$

où l'exposant + indique l'inverse généralisé du produit tensoriel et $\mathbf{X}_{\text{nouveau}}$ les données spectrales concaténées et prétraitées des nouveaux échantillons. Les nouveaux scores sont enfin multipliés par les coefficients de 10 régression (obtenus sur les échantillons d'étalonnage) pour obtenir les paramètres de qualité des nouveaux échantillons.

Ainsi les opérations nécessaires pour prédire les concentrations des paramètres de qualité dans les échantillons nouveaux à analyser consistent à :

- 15 - prétraiter les nouveaux spectres (éliminer la diffusion et standardiser) ;
- calculer les scores des échantillons à analyser à partir des vecteurs B et C en mémoire ; et
- appliquer l'équation de régression en mémoire pour obtenir 20 les concentrations des paramètres.

La méthode décrite ci-dessus a été appliquée pour prédire la teneur en acrylamide dans des échantillons de chicorée ($n=68$).

Pour chaque échantillon, trois spectres ont été acquis avec des excitations à 280, 340 et 429 nm, prétraités par suppression de la diffusion et 25 standardisation, et concaténés ; puis les MEE ainsi obtenues ont été décomposées en quatre facteurs PARAFAC.

Sur la figure 7 :

- 30 - les graphiques de la première colonne, libellés « A », représentent les « scores » des quatre facteurs PARAFAC (respectivement : points noirs, cercles gris, carrés, étoiles) pour 100 échantillons ;
- les graphiques de la deuxième colonne, libellés « B » représentent les « *loadings* » d'excitation des quatre facteurs PARAFAC (courbes $F_1^B, F_2^B, F_3^B, F_4^B$) pour ces mêmes 100 échantillons ;

- les graphiques de la troisième colonne, libellés « C » représentent les « *loadings* » d'émission des quatre facteurs PARAFAC (courbes F^C_1 , F^C_2 , F^C_3 , F^C_4) pour ces mêmes 100 échantillons.

Les graphiques de la première ligne se rapportent au cas où la diffusion de Rayleigh a été supprimée par introduction de valeurs manquantes ; ceux de la troisième ligne, au cas où elle a été supprimée par introduction de zéros ; ceux de la troisième et quatrième ligne, au cas où la suppression de la diffusion de Rayleigh a été effectuée en utilisant un modèle linéaire généralisé.

À l'exception des données qui incluent des valeurs manquantes, toutes les autres données ont été aussi standardisées par MSC.

Les résultats montrent que les valeurs manquantes au niveau du chevauchement de la diffusion et de la fluorescence (fig. 7C ligne 1) sont mal prédites tandis que les zéros produisent des artefacts (fig. 7C ligne 2) constitués par des variations brusques d'intensité. Les modèles prétraités avec GLZ produisent des paramètres spectraux acceptables, avec des variations d'intensité graduelles (fig. 7C lignes 3 et 4).

Le critère de convergence du modèle de la figure 7 lignes 1, 2 et 3 est de 10^{-6} (critère communément utilisé) ; celui du modèle de la figure 7 ligne 4 est de 10^{-2} . La restriction imposée sur la convergence du modèle soit par limitation du nombre d'itérations permissible et/ou en augmentant le critère de convergence améliore très significativement les résultats modèle en termes de structure des scores (fig. 7A lignes 3 et 4) et de séparation des facteurs (fig. 7B et C lignes 3 et 4). En particulier, on peut remarquer l'évolution de la courbe F^B_2 qui, dans la troisième ligne de la figure, montre une contribution des trois longueurs d'onde d'excitation ; au contraire, dans la quatrième ligne, la séparation des excitations est nettement meilleure. De même, l'épaulement de la courbe F^C_1 est réduit considérablement lorsque le facteur de convergence passe de 10^{-6} à 10^{-2} .

Les scores issus du modèle présenté dans la figure 7, ligne 4 ont été utilisés pour calibrer la teneur en acrylamide chimiquement mesurée dans les échantillons. Les teneurs en acrylamide chimiquement mesurées et prédites par le modèle basé sur les scores PARAFAC de la fluorescence mesurée sur le capteur sont présentés dans la figure 8. Sur cette figure, les

teneurs en acrylamide mesurées conventionnellement sont reportées sur l'axe des abscisses, celles prédictes par le modèle le sont sur l'axe des ordonnées. N est le nombre d'échantillons, R est le coefficient de corrélation et le RMSEC est la racine carré de la moyenne des carrés des erreurs. On remarquera en 5 particulier la valeur élevée du coefficient de corrélation (presque 95%) ; la RMSEC vaut 465,47 µg/kg, c'est à dire environ 15% de la valeur moyenne de la teneur en acrylamide mesurée.

Quand la régression issue des scores PARAFAC de fluorescence n'aboutit pas à des résultats convenables (par exemple coefficient 10 de corrélation $R < 0,85$), une autre stratégie qui vise à améliorer la prédiction du modèle de régression est appliquée. Cette méthode se base sur la régression multivoies NPLS, de préférence couplée à un prétraitement par Correction Orthogonale du Signal (OSC) [Wold 1998].

Le prétraitement OSC vise à réduire ou éliminer la variance 15 contenue dans le signal de fluorescence qui est orthogonale (non corrélée) à la variable dépendante (y) avant d'exécuter la régression. L'OSC est appliquée sur les données concaténées après élimination de la diffusion par la méthode GLZ et prétraitement de standardisation par MSC.

L'algorithme le plus fréquemment utilisé pour la correction OSC 20 est le NIPALS (« non linear iterative partial least squares », c'est à dire algorithme non linéaire et itérative par moindres carrés partiels). La première étape consiste à extraire le vecteur des scores t de la première composante principale (décomposition en valeurs singulières) et l'orthogonaliser sur le vecteur y : $t_{\text{nouveau}} = (1 - y(y^T y)^{-1} y^T) t$. Ensuite, une régression type MLR ou PLS 25 (multivariée) est appliquée pour prédire le vecteur orthogonalisé : $w = X^+ t_{\text{nouveau}}$ (la convergence est testée sur le nouveau score $t^* = Xw$). Lorsque t^* converge, le vecteur propre correspondant est calculé $p = X^T t^*$ et l'effet orthogonal est ensuite éliminé des données de fluorescence ($X_{\text{nouveau}} = X - t^* p^T$). Dans l'application à l'invention, un seul facteur OSC est calculé puis éliminé avant la 30 réorganisation des données de fluorescence en une structure trois voies. Dans le cas d'une seule variable y , le modèle converge en une seul itération.

Contrairement au modèle de décomposition PARAFAC, le modèle de régression NPLS [Bro 1998] cherche des facteurs explicatifs à la fois

des données de fluorescence \underline{X} (trois voies) et de y (la variable dépendante). La régression NPLS cherche donc une décomposition des données \underline{X} qui aboutit à des scores qui ont une covariance maximale avec y ; une régression interne entre les scores NPLS de \underline{X} et y aboutit ensuite à l'étalonnage :

$$5 \quad \underline{X}^{(I \times JK)} = \mathbf{T}(\mathbf{W}^K \otimes \mathbf{W}^J)^T + \mathbf{E}_1 ;$$

$$\mathbf{y} = \mathbf{T}\mathbf{b} + \mathbf{E}_2 ;$$

où \mathbf{T} est le vecteur des scores de \underline{X} (ayant la meilleure covariance avec y) et les vecteurs *loadings* respectives pour les excitations et les émissions sont \mathbf{W}^K et \mathbf{W}^J .

10 L'algorithme pour la NPLS avec f facteurs correspondant à des données \underline{X} (trois voies) et une seule variable dépendante est décrit comme suit :

1. Centrer les données \underline{X} et y et on pose $\mathbf{y}_0 = \mathbf{y}$
2. Commencer l'itération $f=1$
- 15 3. Calculer $\mathbf{Z}=\mathbf{X}^T\mathbf{y}$
4. Déterminer \mathbf{w}^J et \mathbf{w}^K en décomposant \mathbf{z} en valeurs singulières
5. Calculer les vecteurs \mathbf{t} et les arranger dans une matrice $\mathbf{T}=[\mathbf{t}_1 \mathbf{t}_2 \mathbf{t}_3 \mathbf{t}_f]$
- 20 6. Calculer les coefficients de régression $\mathbf{b}= (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{y}_0$
7. Chaque échantillon \mathbf{X}_i est remplacé par $\mathbf{X}_i - \mathbf{t}_i\mathbf{w}^J(\mathbf{w}^K)^T$ et $\mathbf{y}=\mathbf{y}_0 - \mathbf{T}\mathbf{b}$
8. $f = f + 1$. Redémarrer de la phase 1 jusqu'à explication suffisante de y .

25 (Dans ces équations, l'indice du facteur f a été omis de \mathbf{t} , \mathbf{w}^J , \mathbf{w}^K , et \mathbf{b})

Etant donné que la NPLS est appliquée sur les données après prétraitement OSC (donc la majeure variance perpendiculaire à y a été éliminée) un seul facteur NPLS est utilisé pour la régression.

Après concaténation des MEE des nouveaux échantillons, le prétraitement MSC est appliqué en utilisant le spectre moyen des échantillons d'étalonnage comme référence. Les facteurs OSC (p et w) calculés sur les échantillons d'étalonnage sont ensuite éliminés des nouveaux échantillons :

$$5 \quad X_c = X_n - X_n * w^* (p^T * w)^{-1} * p^T;$$

X_n est la matrice des MEE concaténées et prétraitées par MSC et X_c la matrice des MEE concaténées corrigées par OSC.

Les loadings (W^J , W^K) et les coefficients de régression b de la NPLS obtenus à partir des échantillons d'étalonnage sont appliqués sur la 10 matrice X_c pour prédire les scores des nouveaux échantillons et leur variable dépendante.

$$T_n = (W^J \otimes W^K)^+ * X_c$$

$$\hat{y} = T_n b + e.$$

La méthode NPLS a été appliquée à la prédiction de la teneur 15 en vitamine C dans des échantillons de haricots verts.

Les figures 9A, 9B, 10A et 10B, des graphiques illustrant l'application d'une méthode selon un deuxième mode de réalisation de l'invention, basée sur un modèle N-PLS, à la prédiction de la teneur en vitamine C de différents échantillons de haricots verts.

20 Après acquisition des matrices MEE de n=60 échantillons d'haricots verts (figure 9A : spectres bruts), élimination de la diffusion (modèle GLZ) et standardisation par MSC, un facteur OSC a été appliqué pour éliminer des données fluorescence la majeure partie de la variance orthogonale (non explicative) à la variable dépendante (teneur en vitamine C chimiquement mesurée) : voir figure 9B, spectres prétraités.

Ensuite, un modèle NPLS avec un facteur a été appliqué pour calibrer le « cube » des MEE prétraités sur la teneur en vitamine C chimiquement mesurée.

30 La figure 10A montre les coefficients de régression NPLS pour les trois spectres d'émission (excitation à 280nm : ligne noire ; excitation à 340nm : ligne grise ; excitation à 429 nm : ligne discontinue). La figure 10B montre un graphique des teneurs en vitamine C mesurées conventionnellement (axe des abscisses) et prédictes par application du modèle NPLS à trois

spectres d'autofluorescence (axe des ordonnées). N est le nombre d'échantillons, R est le coefficient de corrélation et le RMSEC est la racine carré de la moyenne des carrés des erreurs. Là encore on remarquera la valeur très élevée du coefficient de corrélation (supérieure à 95%) et le fait que 5 l'erreur quadratique moyenne RMSEC est de l'ordre de 15%, acceptable pour les applications considérées ici.

L'invention a été décrite en référence à deux exemples particuliers dans lesquels elle a été appliquée à la détermination de la teneur d'échantillons alimentaires en une substance d'intérêt. L'invention n'est 10 cependant pas limitée à ce type d'applications. Plus généralement, elle peut permettre la détermination de différents indicateurs de qualité d'un échantillon ; ces indicateurs peuvent être de type chimique, comme dans les exemples précités, mais également de type microbiologique (charge microbienne), physico-chimique ou chimique. L'invention permet aussi de caractériser les 15 procédés auxquels des échantillons sont soumis, en procurant des estimations de paramètres tels que la valeur stérilisatrice ou pasteurisatrice, le produit temps-température ou la valeur cuisatrice d'un procédé de traitement thermique.

Références :

20 [Bro 1998] Bro, R. « Multi-way Analysis in the Food Industry Models, Algorithms, and Applications », PhD thesis, Universiteit van Amsterdam, 1998.

[Dhanoa 1994] Dhanoa, M.S.; Lister, S.J.; Sanderson, R.; Barnes, R.J. « The link between multiplicative scatter correction (MSC) and 25 standard normal variate (SNV) transformations of NIR spectra. » *J. Near Infrared Spectrosc.* 1994, 2, 43-47.

[Harshman 1984] Harshman, R. A. « « How can I know if it's real? » A catalog of diagnostics to use with three-mode factor analysis and multidimensional scaling. » In: *Research methods for multimode data analysis*. 30 Law, H.G.; Snyder, C.W.; Hattie, J.A.; McDonald, R.P., eds. pp. 566-571, Praeger, New York, 1984.

- [Rizkallah 2007] Rizkallah, J. « Chemometric Analysis of Front Face Fluorescence Data – Estimation, of Neoformed Contaminants in Processed Foods. » PhD thesis, Agro Paris Tech France, 2007.
- [Wold 1998] Wold, S.; Antti, H.; Lindgren, F.; Öhman. 5 « Orthogonal signal correction of near-infrared spectra. » *Chemometr. Intell. Lab. Syst.* 1998, 44, 175-185.
- [Nahorniak 2003] Nahorniak Michelle L., Booksh Karl S. « Optimizing the implementation of the PARAFAC method for near-real time calibration of excitation-emission fluorescence analysis » *J. Chemometrics* 10 2003; 17; 608-617.
- [Rizkallah et al., 2008] Rizkallah J., Lakhal L., Birlouez-Aragon I.: « Front face fluorescence to monitor food processing and neoformed contamination » In « Optical Methods for Monitoring Fresh and Processed Food - Basics and Applications for a better Understanding of Non-Destructive Sensing » Chapter 4. Editor Zude M. Publisher, CRC Press, USA (2008)
- [Davidson 1998] Davidson, R. ; MacKinnon J.G. « Econometric theory and methods. Generalized least squares and related methods » in « Econometric Theory and Methods », Oxford University Press 2004, pp 255-261.

REVENDICATIONS

1. Procédé d'analyse spectroscopique d'au moins un échantillon mettant en œuvre une méthode d'analyse des données spectroscopiques basée sur un modèle statistique multivoies, le procédé 5 comportant :
 - a) l'éclairage dudit ou de chaque échantillon à analyser (E) par une pluralité de rayonnements lumineux d'excitation à des longueurs d'onde respectives ;
 - b) l'acquisition de spectres de fluorescence frontale (F_{280} , F_{340} , 10 F_{429}) dudit ou de chaque échantillon, correspondant chacun à un rayonnement lumineux d'excitation respectif ;
 - c) un prétraitement des spectres de fluorescence acquis ;
 - d) pour chaque échantillon, le calcul d'un vecteur de scores par application aux spectres prétraités dudit modèle statistique multivoies, 15 identifié par un vecteur de *loadings* d'excitation de la fluorescence et par un vecteur de *loadings* d'émission ; et
 - e) la détermination d'au moins un paramètre choisi parmi un indicateur de qualité dudit ou de chaque échantillon et un paramètre caractérisant un procédé auquel ledit ou chaque échantillon a été soumis, à 20 partir dudit vecteur des scores ;
le procédé étant caractérisé en ce que l'écart spectral moyen entre lesdits rayonnements lumineux d'excitation est d'au moins 50 nm, sur une plage spectrale d'au moins 100 nm.
2. Procédé selon la revendication 1 dans lequel le nombre de 25 rayonnements lumineux d'excitation, et de spectres de fluorescence correspondants pour chaque échantillon, est compris entre deux et six, et de préférence entre 3 et 5, sur une plage spectrale d'au moins 100 nm, et de préférence d'au moins 150 nm.
3. Procédé selon l'une des revendications précédentes, 30 comprenant également une phase préliminaire d'étalonnage comportant :
 - i) l'éclairage d'une pluralité d'échantillons d'étalonnage par lesdits rayonnements lumineux d'excitation ;

- ii) l'acquisition des spectres de fluorescence frontale des échantillons, correspondant auxdits rayonnements lumineux d'excitation ;
 - iii) un prétraitement des spectres de fluorescence acquis ;
 - iv) la détermination, par une méthode itérative, desdits vecteurs des *loadings* d'excitation et d'émission du modèle multivoies, ainsi que d'un vecteur de scores pour chaque échantillon d'étalonnage ; et
 - v) la détermination d'une fonction de régression liant lesdits vecteurs de scores aux valeurs, connues, dudit ou de chaque paramètre pour lesdits échantillons d'étalonnage.
- 10 4. Procédé selon l'une des revendications précédentes, dans lequel les spectres de fluorescence sont prétraités par soustraction d'une contribution due à la diffusion Rayleigh du premier ordre du rayonnement lumineux d'excitation, ladite contribution étant calculée au moyen d'un modèle linéaire généralisé.
- 15 5. Procédé selon la revendication 4, dans lequel ledit modèle linéaire généralisé est déterminé à partir d'une première portion (RD) du spectre à prétraiter, comprenant seulement ladite contribution due à la diffusion Rayleigh, et est utilisé pour prédire la contribution de diffusion Rayleigh dans une deuxième portion (RP) dudit spectre dans laquelle elle se superpose à une 20 contribution de fluorescence.
6. Procédé selon l'une des revendications précédentes, dans lequel les spectres de fluorescence sont prétraités en les normalisant.
7. Procédé selon l'une des revendications précédentes, dans lequel les spectres de fluorescence de plusieurs échantillon, d'étalonnage et/ou 25 à analyser, sont prétraités en effectuant une correction multiplicative de dispersion.
8. Procédé selon l'une des revendications précédentes, dans lequel ledit modèle statistique multivoies est un modèle PARAFAC, les spectres de fluorescence de chaque échantillon étant représentés sous forme 30 concaténée.
9. Procédé selon la revendication 8 dans lequel les vecteurs de *loadings* dudit modèle PARAFAC sont déterminés par une méthode itérative

qui est arrêtée lorsque la diminution de la fonction de perte due à la dernière itération descend au-dessous d'une valeur de seuil comprise entre 10^{-4} et 10^{-2} .

10. Procédé selon l'une des revendications 1 à 7, dans lequel ledit modèle statistique multivoies est un modèle NPLS, les spectres de fluorescence de chaque échantillon étant représentés sous forme concaténée.

5 11. Procédé selon la revendication 10, dans lequel les spectres de fluorescence sont prétraités par correction orthogonale du signal.

12. Procédé selon l'une des revendications précédentes dans lequel lesdits rayonnements lumineux d'excitation comprennent :

10 - un premier rayonnement à une longueur d'onde comprise entre 270 et 300 nm ;

- un deuxième rayonnement à une longueur d'onde comprise entre 300 et 360 nm ; et

- un troisième rayonnement à une longueur d'onde comprise entre 400 et 500 nm.

15 13. Procédé selon l'une des revendications précédentes dans lequel ledit ou au moins un dit paramètre à déterminer est choisi parmi :

- la teneur dudit ou de chaque échantillon en une substance d'intérêt ;

20 - une charge microbienne dudit ou de chaque échantillon ; et

- une valeur stérilisatrice, pasteurisatrice ou cuisatrice, ou un produit temps-température d'un procédé auquel ledit ou chaque échantillon a été soumis.

25 14. Procédé selon l'une des revendications précédentes dans lequel les échantillons à analyser et d'étalonnage sont des échantillons d'un produit choisi parmi : un aliment et un médicament.

15. Utilisation d'un procédé selon la revendication 14 pour mesurer l'évolution des propriétés nutritionnelles, microbiologiques et/ou toxicologiques d'un aliment au cours de sa préparation ou conservation.

30 16. Appareil d'analyse spectroscopique d'au moins un échantillon comportant :

- un ensemble de sources lumineuses (S_1 , S_2 , S_3) pour éclairer ledit ou chaque échantillon à analyser (E) par des rayonnements

lumineux d'excitation respectifs, présentant des longueurs d'onde différentes avec un écart spectral moyen d'au moins 50 nm sur une plage spectrale d'au moins 100 nm ;

- un moyen d'acquisition (M, D) des spectres de fluorescence frontale émis par ledit ou de chaque échantillon lorsqu'il est éclairé par lesdits rayonnements lumineux d'excitation ; et
 - un moyen de traitement (MTD) des spectres de fluorescence acquis, adapté pour mettre en œuvre un procédé selon l'une des revendications 1 à 14.

10 17. Appareil d'analyse spectroscopique selon la revendication 16 comportant entre deux et six, et de préférence entre trois et cinq, desdites sources lumineuses.

18. Appareil selon la revendication 17 comportant :

- une première source lumineuse (S_1) émettant un rayonnement à une longueur d'onde comprise entre 270 et 300 nm ;
- une deuxième source lumineuse (S_2) émettant un rayonnement à une longueur d'onde comprise entre 300 et 360 nm ; et
- une troisième source lumineuse (S_3) émettant un rayonnement à une longueur d'onde comprise entre 400 et 500 nm.

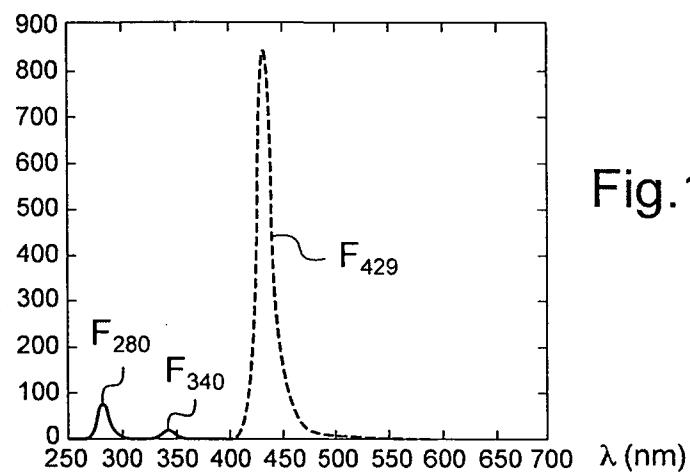


Fig.1

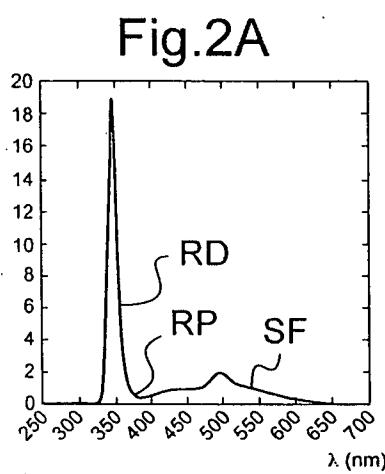


Fig.2B

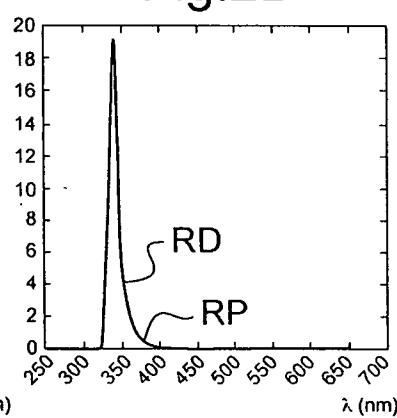


Fig.2C

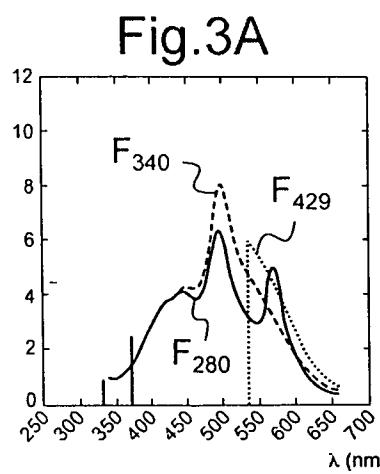
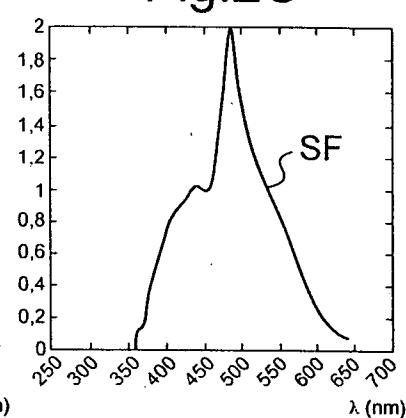


Fig.3B

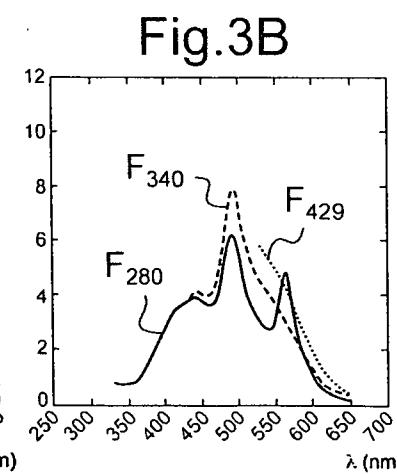
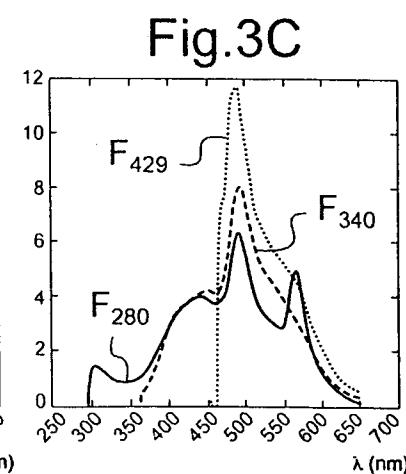


Fig.3C



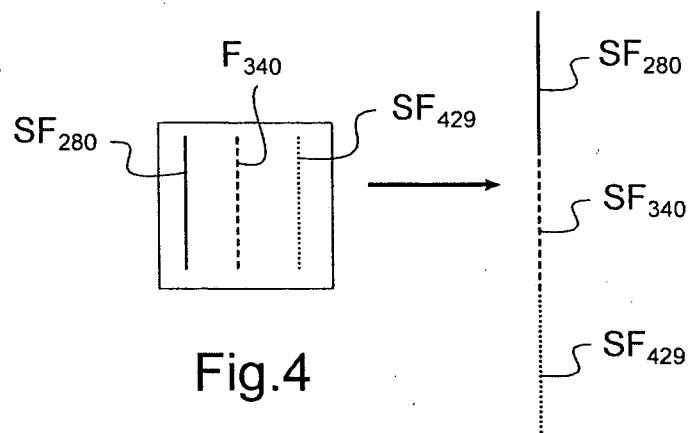


Fig.4

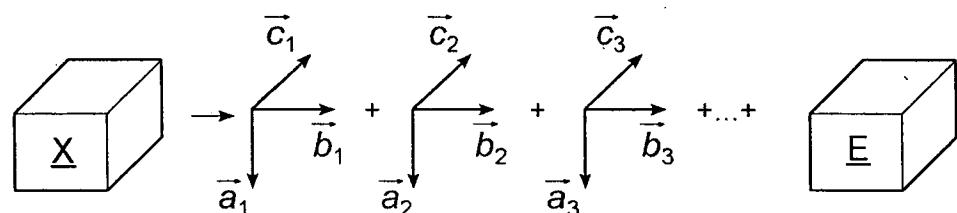


Fig.5

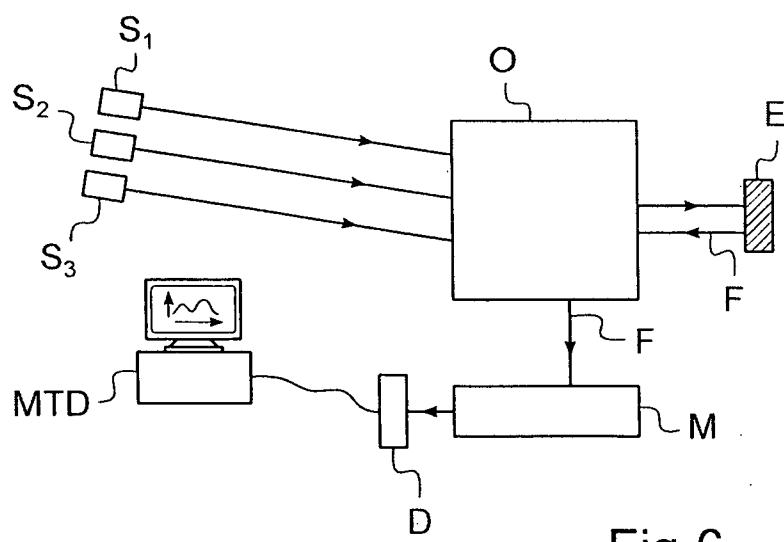
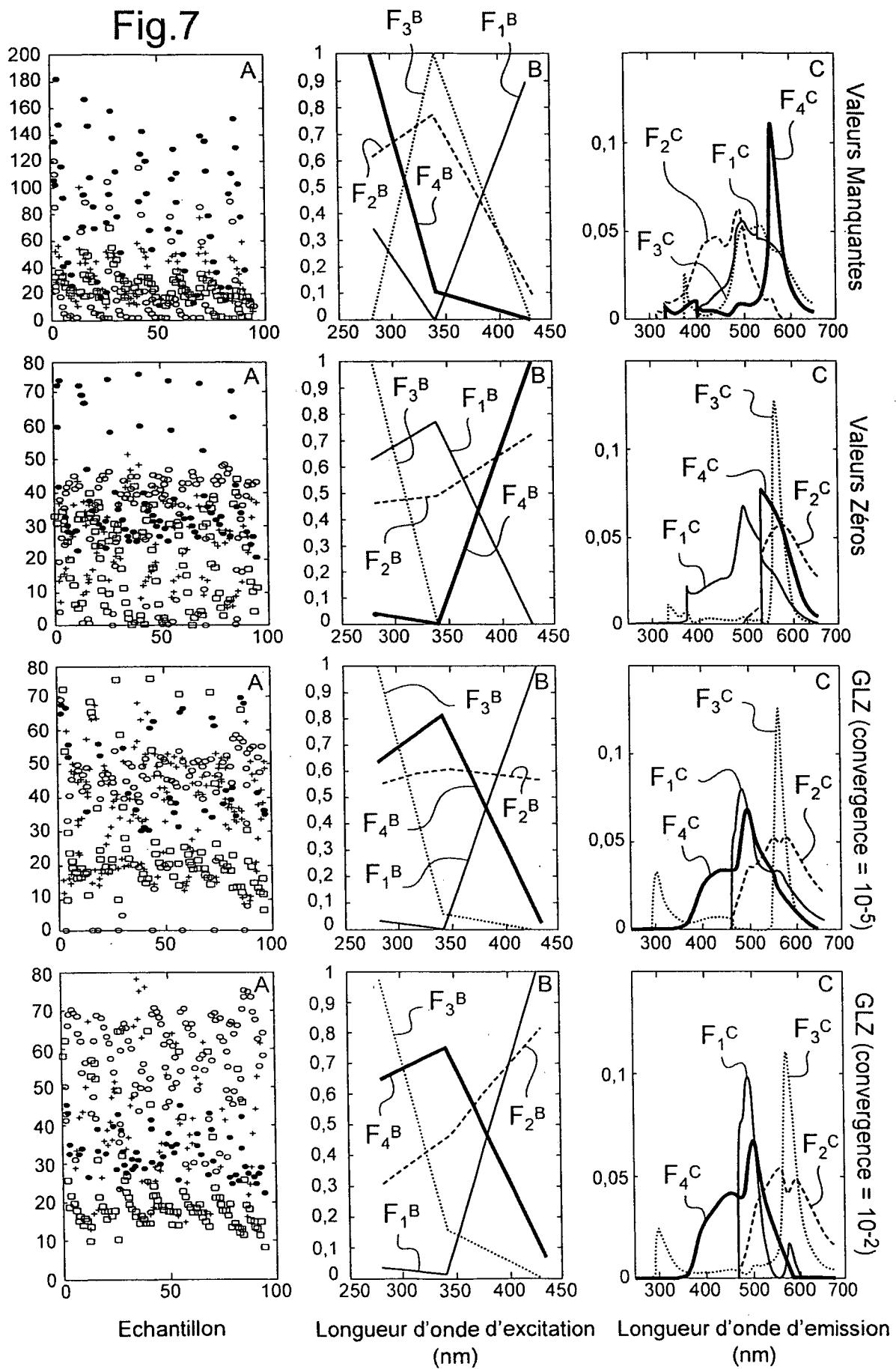


Fig.6

Fig.7



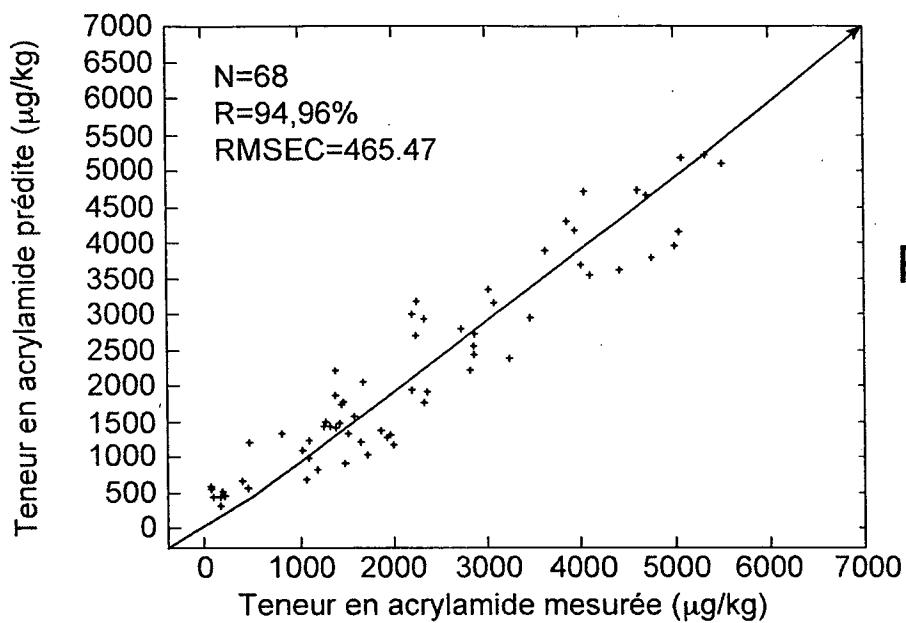


Fig.8

Fig.9A

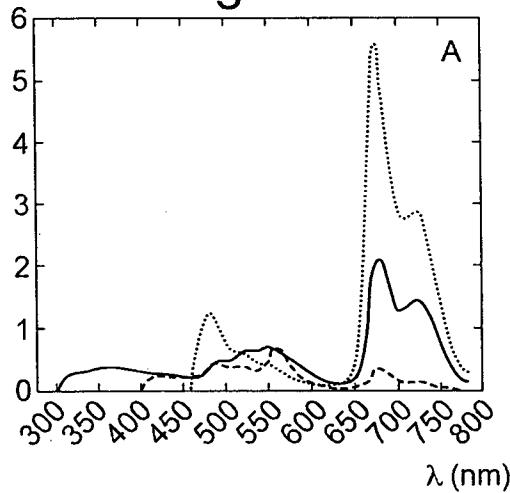


Fig.9B

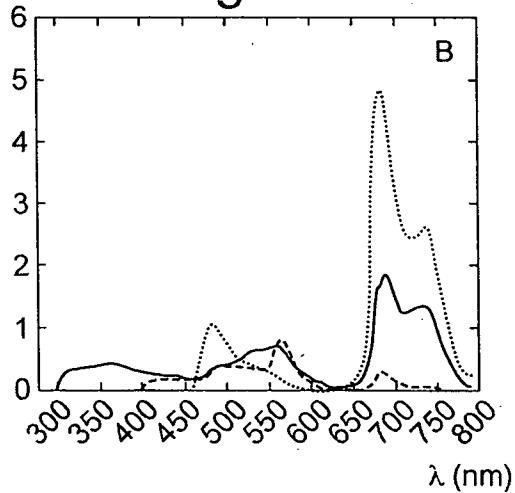


Fig.10A

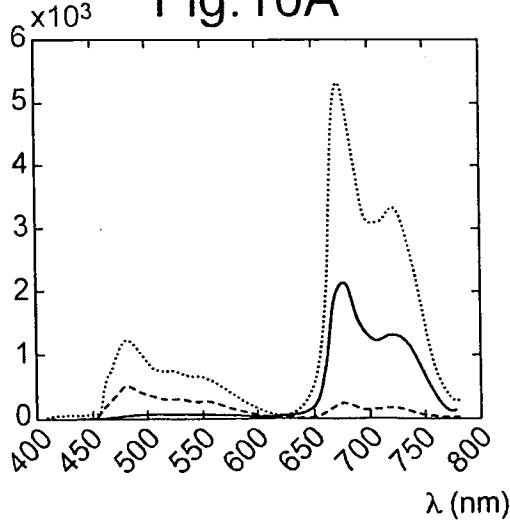


Fig.10B

