



HAL
open science

Analyse multibloc d'images hyperspectrales multimodales et multirésolutions : cartographie des parois de tiges de maïs par microspectroscopies en rayonnement synchrotron

Fatma Allouche

► **To cite this version:**

Fatma Allouche. Analyse multibloc d'images hyperspectrales multimodales et multirésolutions : cartographie des parois de tiges de maïs par microspectroscopies en rayonnement synchrotron. Traitement du signal et de l'image [eess.SP]. ONIRIS - L'UNIVERSITÉ NANTES ANGERS LE MANS, 2012. Français. NNT: . tel-04410347

HAL Id: tel-04410347

<https://hal.inrae.fr/tel-04410347v1>

Submitted on 22 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de Doctorat

Fatma ALLOUCHE

Mémoire présenté en vue de l'obtention
du grade de Docteur d'Oniris
Sous le label de l'Université Nantes Angers Le Mans

Discipline : Automatique et traitement du signal
Spécialité : Traitement du signal et des images
Laboratoire : Biopolymères Interactions Assemblages- INRA
Soutenu le 26 Octobre 2012

École doctorale 498 Végétal Environnement Nutrition Aliment Mer (VENAM)

Analyse multibloc d'images hyperspectrales multimodales et multirésolutions : cartographie des parois de tiges de maïs par microspectroscopies en rayonnement synchrotron

JURY

Rapporteurs :	M. Duponchel Ludovic M. Rutledge Douglas	-Professeur -Professeur	Université Lille 1 Agroparitech
Examineurs :	Mme. Sevestre-Ghalila Sylvie M. Réfrégiers Matthieu M. Hanafi Mohamed	-Maître de conférences -Responsable d'équipe -Ingénieur de recherche	Paris V / ENIT (Tunisie) Synchrotron SOLEIL ONIRIS
Directeur de Thèse :	Mme. Guillon Fabienne	-Directrice de recherche	INRA Nantes
Co-encadrants de Thèse :	Mme. Devaux Marie-Françoise M. Jamme Frédéric	-Ingénieur de recherche -Ingénieur de recherche	INRA Nantes Synchrotron SOLEIL

À mon père

Remerciement

Je tiens tout d'abord à remercier ma directrice de thèse, Fabienne Guillon, pour m'avoir accueilli au sein de son équipe et pour m'avoir fait confiance malgré les connaissances plutôt légères que j'avais en 2009 en biologie et en spectroscopie.

Un acteur particulièrement incontournable de cette thèse est Marie Françoise Devaux. Je ne la remercierais jamais assez pour s'être présente largement dans le suivi de cette thèse. Ses conseils, encouragements journaliers et sa grande disponibilité ont contribué très amplement à la qualité de ce travail.

Je remercie également Frédéric Jamme pour son aide, ses conseils, sa gentillesse, son enthousiasme, son soutien tout au long de ces trois années.

Je tiens ensuite à remercier Mohamed Hanafi pour toutes les connaissances qu'il m'a transmises dans le domaine des analyses multiblocs. Son recul sur le sujet a beaucoup apporté à ce travail de recherche.

Je remercie cordialement M. Ludovic Duponchel, Professeur de l'université de Lille 1 et M. Douglas Rutledge, Professeur à Agroparitech qui ont accepté d'être rapporteurs de cette thèse. Je souhaite adresser également mes remerciements à Mme Sevestre Ghalila Sylvie et M. Réfrégiers Matthieu d'avoir fait l'honneur d'examiner mon travail.

Je remercie également toute personne qui a contribué à la réalisation de ce travail. Je remercie Paul Robert pour l'interprétation des spectres infrarouge, Cécile Barron pour l'interprétation des spectres Raman, Sylvie Durand et Brigitte Bouchet pour leur préparation des échantillons et leur participation aux manips. Je remercie également les membres de comité de suivi de ma thèse : Achim Kohler, Danielle Nuzillard et Paul Dumas pour leurs critiques judicieuses et constructives de ce travail.

Je remercie également les relectrices attentives, Sylvie, Camille et Brigitte qui m'ont aidé à rendre ce document plus lisible et clair.

Grand merci à tout le personnel de l'équipe PVPP et BIBS, dont certains, qui se reconnaîtront, m'auront plus d'une fois dépanné là où rien ne les y obligeait. Merci aux post-docs, thésards et stagiaires que j'ai côtoyé à l'unité BIA durant ma thèse.

Merci également aux membres des équipes SMIS & DISCO pour leur accueil chaleureux et leur sympathie.

Merci pour toute ma famille et mes amis.

Un grand merci pour mes premières fans : ma mère et ma sœur 'SouSou'. Leur présence et leurs encouragements sont pour moi les piliers fondateurs de ce que je suis et de ce que je fais.

Enfin, je remercie mon cher Bouali 'DabDoubi' pour sa compréhension, son soutien quotidien indéfectible, ses corrections, sa patience et son amour...

Ces remerciements ne peuvent s'achever, sans une pensée pour mon père qui aurait été certainement très fier de moi, s'il était à mes côtés.

Table des matières

Table des matières	i
Nomenclature	v
Liste des Figures	vii
Liste des Tableaux	xii
Introduction Générale	1
Partie A :	6
Introduction à la spectroscopie et à l’analyse d’images hyperspectrales pour la cartographie chimique de tissus végétaux	6
Introduction	9
I Structure et composition chimique des parois cellulaires de tige de maïs et verrous à leur dégradation enzymatique	9
1 Anatomie de la tige de maïs	9
2 La paroi cellulaire.....	11
3 Les constituants pariétaux	12
4 Barrières aux dégradations enzymatiques	18
5 Conclusion.....	21
II Spectroscopies et microspectroscopies	22
Introduction	22
1 Spectroscopie	22
2 Complémentarité entre les méthodes de spectroscopie	32
3 Microspectroscopies	32
4 Pré-traitement	38
Conclusion.....	40
III Chimiométrie et analyse des données spectrales	41
1 Spectres, tableaux de spectres et images hyperspectrales	41
2 Chimiométrie.....	45
Conclusion.....	56
VI Conclusion : Couplage d’image hyperspectrale pour la caractérisation chimique d’échantillons complexes	57
1 Position du problème.....	57
2 Démarche adoptée	58

Partie B :	60
Données...Acquisition & exploration	60
Introduction	62
I Préparation des tiges de maïs	62
1 Echantillons	62
2 Choix de la région	63
3 Préparation de l'échantillon	66
II Acquisition conjointe des images hyperspectrales multimodales sur les lignes SMIS et DISCO du Synchrotron SOLEIL	67
1 Imagerie hyperspectrale	67
2 Protocole d'acquisition conjointe des données	69
3 Description des collections des images hyperspectrales	72
III Pré-traitement	75
1 Déformations spectrales et spatiales des données	75
2 Stratégie	78
3 Méthodes de pré-traitement	79
4 Applications aux images hyperspectrales	82
IV Analyse individuelle de la collection 4f11mb-2	87
1 Exploration des données	87
2 Analyse en composantes principales de la collection de données 4f11mb-2	90
Conclusion	97
Partie C :	99
Appariement des images hyperspectrales multirésolutions pour l'analyse multibloc	99
Introduction	101
I Recalage des images hyperspectrales	101
1 Notions de recalage d'images	101
2 Etapes de recalage	103
3 Transformation géométrique	104
4 Caractéristiques des images hyperspectrales	106
5 Stratégie adoptée	109
6 Méthodes	111
7 Matrices de transformations	117
II Appariement de spectres	118
1. Projection des images de fluorescence et Raman dans l'espace infrarouge	118

2 Recherche des pixels infrarouge ayant des imagerie homologues en fluorescence et en Raman.....	120
3 Appariement des imagerie de fluorescence et Raman	122
III Construction des tableaux de données spectrales pour l’analyse multibloc	122
VI Résultats et discussions.....	124
1 Appariement de la collection 4f11mb-1	124
2 Appariement de la collection 4f11mb-2.....	126
Conclusion.....	128
Partie D :	129
Couplage des images hyperspectrales en utilisant les méthodes multibloc.....	129
Introduction	131
I Couplage de deux images hyperspectrales multirésolutions : extension de l’analyse inter -batterie de Tucker.....	132
Introduction	132
1 Analyse de tableaux multiples : Méthodes.....	132
2 Résultats et discussions	137
Conclusion.....	145
II Couplage de trois images hyperspectrales par l’analyse de co-inertie multiple.....	145
Introduction	145
1 Analyse de co-inertie Multiple : Méthodes et outils d’interprétation	146
2 Résultats et discussions	151
Conclusion.....	162
III. Couplage de trois images hyperspectrales multirésolutions : extension de l’analyse de co-inertie multiple	163
Introduction	163
1 Extension de l'analyse de co-inertie multiple à l'analyse de tableaux ayant un nombre d'entrées hétérogènes.....	163
2 Analyse de l’échantillon 4f11mb-2 en utilisant l’analyse de co-inertie multiple trilineaire	168
Conclusion.....	181
Conclusion.....	181
Discussions	183
1 Acquisition conjointe des données	184
2 Pré-traitement des spectres.....	184

3 Appariement des données.....	185
4 Analyse multibloc	186
5 Analyse et interprétation des données	189
Références Bibliographiques.....	191
Annexe A	205
Contributions scientifiques.....	206

Nomenclature

Dans tout ce qui va suivre, nous allons adopter les conventions de notations suivantes :

n, m, l, \dots scalaires

N, M, L, \dots dimensions des tableaux

$\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$ vecteur

$\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \dots$ matrices ou tableaux à deux entrées

$\underline{\mathbf{X}}, \underline{\mathbf{Y}}, \underline{\mathbf{Z}}, \dots$ tableaux à trois entrées

$[\mathbf{X}, \mathbf{Y}]$ concaténation des variables des matrices \mathbf{X} et \mathbf{Y}

(h) indice de la composante pour la dimension courante

(H) valeur du nombre de composantes

z indice de la tranche courante d'un tableau à trois entrées

Z nombre total de tranches

$Z(j)$ nombre total de tranches du $j^{\text{ème}}$ tableau à trois entrées

K, J nombre total des tableaux à deux, à trois entrées

k, j indice courant des tableaux à deux, à trois entrées

$\mathbf{X}_k^{(h)}$ $k^{\text{ème}}$ tableau à deux entrées à la dimension courante (h)

$\underline{\mathbf{X}}_j^{(h)}$ $j^{\text{ème}}$ tableau à trois entrées à la dimension courante (h)

$\mathbf{X}_{j(z)}^{(h)}$ $z^{\text{ème}}$ tranche du $j^{\text{ème}}$ tableau à trois entrées à la dimension courante (h)

$\overline{\mathbf{X}}_j^{(h)}$ somme pondérée du $j^{\text{ème}}$ tableau à trois entrées à la dimension courante (h)

$\boldsymbol{\alpha}_j^{(h)}$ vecteur de pondération du $j^{\text{ème}}$ tableau à trois entrées à la dimension courante (h)

$\boldsymbol{\alpha}_{j(z)}^{(h)}$ valeur de pondération de chaque tranche du $j^{\text{ème}}$ tableau à trois entrées à la dimension courante (h)

$\tilde{\mathbf{u}}^{(s)}$ profil variable global à la dimension courante (h) avant normalisation

$\tilde{\mathbf{u}}_k^{(s)}$ profil variable du $k^{\text{ème}}$ tableau à la dimension courante (h) avant normalisation

$\mathbf{u}_k^{(s)}$ profil variable du $k^{\text{ème}}$ tableau à la dimension courante (h) normalisé

$\mathbf{c}_g^{(h)}$ composante globale à la dimension courante (h)

$\mathbf{c}_{\mathbf{X}_j}^{(h)}$ composante du $k^{\text{ème}}$ tableau à deux entrées à la dimension courante (h)

$\mathbf{c}_{\underline{\mathbf{X}}_j}^{(h)}$ composante du $j^{\text{ème}}$ tableau à trois entrées à la dimension courante (h)

$\mathbf{c}_{\mathbf{x}_j}^{(h)}$ composante de la somme pondérée du $j^{\text{ème}}$ tableau à trois entrées à la dimension courante

(h)

$\mathbf{R}_{\mathbf{x}_k}^{(h)}$ matrice des résidus du $k^{\text{ème}}$ tableau à deux entrées à la dimension courante (h)

$\mathbf{R}_{\mathbf{x}_{j(z)}}^{(h)}$ matrice des résidus de la $z^{\text{ème}}$ tranche du $j^{\text{ème}}$ tableau à trois entrées à la dimension courante (h)

$\lambda_k^{(h)}$ contribution du $k^{\text{ème}}$ tableau à la détermination de la composante globale à la dimension courante (h)

$\lambda_g^{(h)}$ contribution de tous les tableaux à la détermination de la composante globale à la dimension courante (h)

\mathbf{I} matrice identité

\mathbf{X}' transposée de la matrice \mathbf{X}

$\text{svd}(\mathbf{X})$ décomposition en valeurs singulières de la matrice \mathbf{X}

$\text{trace}(\mathbf{X})$ somme des éléments diagonaux de la matrice \mathbf{X}

$\text{var}(\mathbf{X})$ variance de la matrice \mathbf{X}

$\text{cov}(\mathbf{x}, \mathbf{y})$ covariance entre les vecteurs \mathbf{x} et \mathbf{y}

$\text{corr}(\mathbf{x}, \mathbf{y})$ corrélation entre les vecteurs \mathbf{x} et \mathbf{y}

$\| \cdot \|$ norme euclidienne

Liste des Figures

Figure A1 Anatomie de la tige de maïs	9
Figure A2 Types cellulaires de la tige de maïs.	10
Figure A3 Faisceau d'une tige de maïs en coupe transversale	11
Figure A4 Représentation schématique de la paroi cellulaire des végétaux	12
Figure A5 Molécule de cellulose	13
Figure A6 Structure des microfibrilles de cellulose.....	14
Figure A7 Structure chimique de l'arabinoxylane	15
Figure A8 Structure chimique de l'acide férulique.....	16
Figure A9 Structure chimique de l'acide para-coumarique	16
Figure A10 Unités H, G et S de la macromolécule de lignine	17
Figure A11 Représentation simplifiée a)-d'un mode de liaison inter-unité labile (β -O-4) et b)- d'un mode de liaison	17
Figure A12 Représentation schématique d'une paroi secondaire de poacées.....	18
Figure A13 Etapes de la dégradation enzymatique	19
Figure A14 Spectre électromagnétique	23
Figure A15 Diagramme de Jablonski	24
Figure A16 Vibration de la molécule CO ₂ (changement du moment dipolaire).....	25
Figure A17 Vibration de la molécule CO ₂ (changement de la polarisabilité).....	26
Figure A18 Schéma d'un Synchrotron.....	33
Figure A19 Microspectrophotomètre infrarouge de la ligne SMIS	34
Figure A20 Microspectrophotomètre Raman de la ligne SMIS.....	35
Figure A21 Composantes du microspectrophotomètre de la ligne DISCO	36
Figure A22 Tableau de données de spectres	42
Figure A 23 Tableau de données à trois entrées.....	42
Figure A24 Couplage de trois tableaux de données.....	43
Figure A25 Structure d'une image hyperspectrale	44
Figure A26 Représentation des données issues d'une image hyperspectrale	45
Figure A27 Les étapes principales de l'analyse en composantes principales	47
Figure A28 Illustration de l'analyse multibloc de deux tableaux à deux entrées.....	50
Figure A29 Analyse multibloc de K tableaux de données	53
Figure A30 Dépliement d'une image hyperspectrale.....	55

Figure A31 Analyse en composantes principales d'une image hyperspectrale	56
Figure A32 Démarche adoptée pour le couplage d'images hyperspectrales multirésolutions.	59
Figure B1 Plante du génotype F2 de maïs au stade floraison femelle.	62
Figure B2 Prélèvements des entrenœuds sous épi	63
Figure B3 Observation de section de tige de maïs au stade ensilage.....	64
Figure B4 Section de la tige de maïs avant et après dégradation enzymatique.....	65
Figure B5 Faisceau sélectionné.....	66
Figure B6 Spectres de la lame ZnS	67
Figure B7 Prototype du porte échantillon construit.	69
Figure B8 Exemple de superposition de l'image spectrale et de l'image visible.....	70
Figure B9 Exemple de planification pour l'analyse conjointe d'une collection d'images hyperspectrales	71
Figure B10 Protocole d'acquisition conjointe des images hyperspectrales	72
Figure B11 Acquisition conjointe des images de la collection 4f11mb-1	73
Figure B12 Acquisition conjointe des images de la collection 4f11mb-2	74
Figure B13 Exemples de collections de spectres.	75
Figure B14 Images de fluorescence.	76
Figure B15 Exemples de spectres de l'image de fluorescence	77
Figure B16 Représentation surfacique des intensités de l'image de la figure B-14b..	77
Figure B17 Exemples de profils de l'image de fluorescence	78
Figure B18 Fonctionnement de la transformation de top-hat à une dimension	79
Figure B19 Transformation du facteur de normalisation par "bouchage des trous".....	81
Figure B20 Transformation du facteur de normalisation par lissage	81
Figure B21 Exemples de spectres après normalisation	82
Figure B22 Etapes de pré-traitement des spectres infrarouge de la collection 4f11mb-1.....	83
Figure B23 Etapes de pré-traitement des spectres infrarouge de la collection 4f11mb-2.....	84
Figure B24 Etapes de pré-traitement des spectres de fluorescence.....	85
Figure B25 Etapes de pré-traitement des spectres Raman	86
Figure B26 Spectres sélectionnés dans les types cellulaires	87
Figure B27 Spectres moyens en fonction des types cellulaires.....	88
Figure B28 Spectres de référence de fluorescence.....	90
Figure B29 Analyse en composantes principales de l'image infrarouge : composante 1	91

Figure B30 Analyse en composantes principales de l'image infrarouge : composante 2	92
Figure B31 Analyse en composantes principales de l'image infrarouge : composante 3	92
Figure B32 Analyse en composantes principales des images de fluorescence : composante	193
Figure B33 Analyse en composantes principales des images de fluorescence : composante	294
Figure B34 Analyse en composantes principales des images de fluorescence : composante	394
Figure B35 Analyse en composantes principales des images Raman : composante 1.....	95
Figure B36 Analyse en composantes principales des images Raman : composante 2.....	96
Figure B37 Analyse en composantes principales des images Raman : composante 3.....	97
Figure C1 Région à cartographier	101
Figure C2 Régions cartographiées	102
Figure C3 Images acquises avec différents instruments	102
Figure C4 Projection des images sur l'image de référence	103
Figure C5 Correspondance entre les différents pixels des images.....	103
Figure C6 Transformation basée sur la rotation.....	104
Figure C7 Transformation basée sur le changement d'échelle	105
Figure C8 Transformation géométrique basée sur la translation.	106
Figure C9 Exemples d'images spectrales et d'images visibles.....	107
Figure C10 Représentation des pixels multirésolutions.....	109
Figure C11 Stratégie adoptée pour recalibrer les images spectrales par rapport aux images de référence.....	110
Figure C12 Matrices de passages entre image spectrale et image de référence.....	110
Figure C 13 Matrice de la corrélation croisée entre les deux images a et b.....	112
Figure C14 Rotation de l'image spectrale infrarouge.....	113
Figure C15 Modèle de la pyramide.....	114
Figure C16 Exemple d'une pyramide Gaussienne.....	115
Figure C17 Image visible infrarouge avant et après dégradation.....	116
Figure C18 Matrices de passages entre les différentes images spectrales en passant par l'image de référence	117
Figure C19 Matrices de passages entre les différents images spectrales	118
Figure C20 Images individuelles de fluorescence projetées sur l'image infrarouge.....	119
Figure C 21 Images individuelles Raman projetées sur l'image infrarouge.....	119
Figure C 22 Région cartographiée totale.....	120
Figure C23 Région commune	120

Figure C24 La région commune totale avant érosion et après érosion	121
Figure C25 Les points centraux des surfaces des pixels infrarouge à la résolution de la fluorescence et du Raman.	121
Figure C26 Images Label	122
Figure C27 Dépliage des pixels multirésolutions	122
Figure C28 Empilement de tous les spectres	123
Figure C29 Etablissement de la structure de données.....	124
Figure C30 Fusion de l'image visible de référence avec les images spectrales de la collection 4f11mb-1	125
Figure C31 Structure de données de la collection 4f11mb-1	126
Figure C32 Fusion de l'image visible de référence avec les images spectrales de la collection 4f11mb-2	127
Figure C 33 Structure de données de la collection 4f11mb-2.....	128
Figure D1 Moyenner un tableau à trois entrées selon la troisième entrée.	132
Figure D2 Centrage d'un tableau de données à trois entrées	134
Figure D3 Matricisation d'un tableau à trois entrées en tableaux à deux entrées.....	134
Figure D4 Profils spectraux infrarouge et de fluorescence 1et 2 de l'analyse inter-batterie de Tucker.....	139
Figure D5 Images composantes infrarouge et de fluorescence 1et 2 de l'analyse inter-batterie de Tucker.....	140
Figure D6 Profils spectraux infrarouge et de fluorescence 1et 2 de l'analyse inter-batterie de Tucker à trois entrées	143
Figure D7 Images composantes de fluorescence 1 et 2 de l'analyse inter-batterie de Tucker à trois entrées.	144
Figure D8 Première étape de l'analyse de co-inertie multiple.....	147
Figure D9 Adaptation de la structure de données ayant un nombre d'entrées hétérogènes à l'analyse de co-inertie multiple	151
Figure D10 Profils spectraux 1 de l'analyse de co-inertie multiple.....	155
Figure D11 Images composantes 1 de l'analyse de co-inertie multiple	156
Figure D12 Profils spectraux 2 de l'analyse de co-inertie multiple.....	157
Figure D13 Images composantes 2 de l'analyse de co-inertie multiple	159
Figure D14 Profils spectraux 3 de l'analyse de co-inertie multiple.....	160

Figure D15 Images composantes 3 de l'analyse de co-inertie multiple	161
Figure D16 Les deux premiers profils spectraux globaux obtenus en utilisant l'analyse de co-inertie trilinéaire.	169
Figure D17 Profils spectraux 1 de l'analyse de co-inertie trilinéaire	172
Figure D18 Images composantes 1 de l'analyse de co-inertie trilinéaire	173
Figure D19 Profils spectraux 2 de l'analyse de co-inertie trilinéaire	174
Figure D20 Images composantes 2 de l'analyse de co-inertie trilinéaire	175
Figure D21 Profils spectraux 3 de l'analyse de co-inertie trilinéaire	176
Figure D22 Images composantes 3 de l'analyse de co-inertie trilinéaire	177
Figure D23 Profils spectraux 4 de l'analyse de co-inertie trilinéaire	178
Figure D24 Images composantes 4 de l'analyse de co-inertie trilinéaire	179
Figure D25 Profils variables de la troisième entrées.....	180
Figure D26 Profils variables repliés de la troisième entrées	180
Figure D27 Profil de type Gaussien	181

Liste des Tableaux

Tableau A1 Composition approximative (% poids sec) des parois cellulaires des tiges de maïs mature.....	12
Tableau A2 Vibration infrarouge et Raman de la molécule CO ₂	26
Tableau A3 Nombre d'ondes infrarouge caractéristiques des composantes chimiques des parois cellulaires de poacées.....	27
Tableau A4 Nombre d'ondes Raman caractéristiques des composants chimiques des parois cellulaires de poacées.....	30
Tableau B1 Pourcentages de variance expliquée par les composantes principales de l'image hyperspectrale infrarouge de l'échantillon 4f1 1mb-2.....	91
Tableau B2 Pourcentages de variance expliquée par les composantes principales de fluorescence de l'échantillon 4f1 1mb-2.....	93
Tableau B3 Pourcentages de variance expliquée par les composantes principales des images hyperspectrales Raman de l'échantillon 4f1 1mb-2.....	95
Tableau C1 Caractéristiques des images spectrales	108
Tableau C2 Caractéristiques des images visibles.....	108
Tableau D1 Résultats de l'analyse inter-batterie de Tucker : tableau infrarouge.....	138
Tableau D2 Résultats de l'analyse inter-batterie de Tucker : tableau de fluorescence.	138
Tableau D3 Indicateurs de l'analyse inter-batterie de Tucker.....	139
Tableau D4 Résultats de l'analyse inter-batterie de Tucker à trois entrées : tableau infrarouge.....	142
Tableau D5 Résultats de l'analyse inter-batterie de Tucker à trois entrées : tableau de fluorescence.....	142
Tableau D6 Indicateurs de l'analyse inter-batterie de Tucker à trois entrées.....	142
Tableau D7 Indicateurs de l'analyse de co-inertie multiple.....	152
Tableau D8 Résultats de l'analyse de co-inertie multiple : tableau infrarouge.....	153
Tableau D9 Résultats de l'analyse de co-inertie multiple : tableau de fluorescence.....	154
Tableau D10 Résultats de l'analyse de co-inertie multiple : tableau Raman.....	154

Tableau D11 Contribution du tableau infrarouge, fluorescence et Raman à l'analyse de co-inertie trilinéaire	169
Tableau D12 Indicateurs de la co-inertie trilinéaire	170
Tableau D13 Résultats de l'analyse de co-inertie trilinéaire : tableau infrarouge.....	171
Tableau D14 Résultats de l'analyse de co-inertie trilinéaire : tableau de fluorescence.....	171
Tableau D15 Résultats de l'analyse de co-inertie trilinéaire : tableau Raman	172

Introduction Générale

Contexte socio-économique

L'augmentation du prix de l'énergie et la prise de conscience du réchauffement climatique renforcent la nécessité de trouver des produits de substitution aux carburants fossiles. La biomasse végétale, principale réserve de polymères carbonés renouvelables, peut servir à la production de "bioéthanol" [Schubert, 2006]. Les poacées appelées anciennement graminées, représentent une part importante de la biomasse disponible et n'entrent pas en compétition avec des usages alimentaires. Parmi les poacées, on peut citer la canne à sucre, le blé ou le maïs. Le maïs est cultivé à la fois pour ses grains et en tant que plante fourragère. Sous forme de grains ou plante, il est considéré comme la première céréale cultivée dans le monde avant le riz et le blé.

Les tiges de maïs sont étudiées comme plante modèle et source pour la production de biocarburants issus de produits non alimentaires appelé biocarburant de deuxième génération [Naik *et al.*, 2010].

Les parois cellulaires sont composées d'environ 75% de polysaccharides qui sont théoriquement dégradables sous la forme de monomères ou oligomères utilisables pour la production de bio-plastiques et bio-éthanol. Cette transformation, appelée saccharification, peut être réalisée par voie chimique [Octave, 2009]. Des procédés alternatifs moins polluants sont recherchés impliquant des enzymes de dégradation des polysaccharides. Dans le cas de la production du bioéthanol de deuxième génération, la dégradation est réalisée par l'ajout d'enzymes purifiées ou par des cocktails enzymatiques issus de microorganismes qui dégradent la biomasse [Hansen *et al.*, 2011].

Les parois cellulaires de maïs sont constituées majoritairement par des celluloses et des hémicelluloses. Les parois contiennent également des composés aromatiques : lignines et acides hydroxycinnamiques. A l'état initial, la dégradabilité enzymatique des parois lignifiées est faible et variable. La dégradabilité peut être liée à la teneur en lignine qui est un des verrous majeurs à la conversion biologique de la cellulose en bioéthanol [Berlin *et al.*, 2006]. D'autres paramètres tels que la structure des lignines, la quantité d'acides hydroxycinnamiques [Taboada *et al.* 2009, Zhang *et al.*, 2011-b] ou leurs liaisons avec les polysaccharides (directes ou par l'intermédiaire des acides féruliques dans le cas des poacées) pourraient moduler l'efficacité de la saccharification.

L'amélioration de la saccharification des parois cellulaires représente un point essentiel pour déterminer l'intérêt et mettre au point des filières de bioraffinage de la biomasse végétale.

Le projet Européen Renewall dans lequel s'intègre cette thèse a pour objectif d'identifier et de hiérarchiser les facteurs limitant la dégradation de la biomasse lignocellulosique. Des lignées mutantes contrastées en teneur de composition des parois sont analysées pour leur aptitude à la saccharification. Parallèlement, une analyse exhaustive des mutants avant dégradation doit être réalisée afin d'expliquer les résultats de saccharification. La composition des parois cellulaires varie en fonction des différents tissus constitutifs de la tige. Ainsi l'hétérogénéité des différents tissus sont à prendre en compte pour expliquer les variations de dégradabilité observées entre les différentes lignées.

Dans ce travail, l'analyse *in situ* de la composition chimique des parois cellulaires est considérée en fonction des types cellulaires. Les analyses chimiques classiques sont en général longues. Une méthode spectrale appropriée, rend possible l'analyse en quelques minutes sans altérer la composition de l'échantillon. La microspectrophotométrie permet d'acquérir des spectres de manière spatialement résolue.

Choix des gammes spectrales

L'étude de la composition chimique de la paroi cellulaire nécessite plusieurs techniques pour analyser l'ensemble des polymères qui la constituent. Dans cette thèse, nous proposons d'appliquer différentes microspectroscopies : le moyen-infrarouge, la fluorescence UV et le Raman. Dans la pratique, les polysaccharides et les acides phénoliques (lignines ou acides hydroxycinnamiques) répondent en infrarouge et en Raman [Himmelsbach *et al.*, 2002 ; Robert *et al.*, 2005 ; Gierlinger *et al.*, 2008-a, 2008-b]. La microspectroscopie infrarouge est adaptée particulièrement pour les hémicelluloses, la cellulose et la lignine. La microspectroscopie Raman est plutôt sensible aux acides phénoliques (lignines et acides hydroxycinnamiques) mais révèle également la cellulose et une partie des hémicelluloses tels que les xylanes.

La microspectroscopie de fluorescence UV peut être utilisée également pour cartographier la distribution des composés phénoliques. Les phenylpropanoïdes qui incluent les lignines et les acides hydroxycinnamiques liés aux polysaccharides par des liaisons esters présentent différents profils spectraux de fluorescence dans la région spectrale UV [Saadi *et al.*, 1998 ; Andersson et Akin, 2008]. L'excitation à 280 nm en microspectroscopie de fluorescence UV permet de différencier les acides hydroxycinnamiques (acide férulique et para-coumarique) par des émissions à 310 et 320 nm [Andersson et Akin, 2008].

Imagerie hyperspectrale

Le principe de base de la microspectroscopie repose sur la méthode dite «point par point». Cette méthode consiste à acquérir des spectres d'une manière supervisée. Dans cette thèse, l'hétérogénéité spatiale est explorée à partir d'images hyperspectrales.

La microspectroscopie classique présente un certain nombre de limites : résolution, rapport signal sur bruit, temps d'acquisition. Le rayonnement synchrotron permet d'en améliorer beaucoup. En microspectroscopie classique, la résolution des images infrarouge est limitée par une taille de pixel minimale de $20\ \mu\text{m} \times 20\ \mu\text{m}$. La source synchrotron permet d'acquérir des spectres de bonne qualité avec une résolution pouvant atteindre $5\ \mu\text{m} \times 5\ \mu\text{m}$.

L'équipement mis au point pour la fluorescence UV sur le site synchrotron SOLEIL permet l'acquisition d'images hyperspectrales pour des longueurs d'ondes d'excitation allant de 250 à 350 nm, *a priori* adaptées pour l'analyse des composés phénoliques et des lignines [Andersson et Akin, 2008; Saadi *et al.*, 1998]. Pour toutes ces raisons, deux projets ont été soumis et acceptés pour acquérir des images hyperspectrales au synchrotron SOLEIL dans le moyen infrarouge et en fluorescence UV. La microspectroscopie Raman en utilisant une source de lumière conventionnelle a une résolution de $1\ \mu\text{m} \times 1\ \mu\text{m}$ permettant l'étude des parois végétale [Gierlinger *et al.*, 2012].

Analyse des données spectrales

La chimiométrie permet l'analyse des images hyperspectrales et l'exploitation de la plus grande partie de l'information sans émettre d'hypothèses trop restrictives. En amont, elle permet le pré-traitement des images hyperspectrales. En aval, elle permet la description des données sous forme synthétique.

L'image hyperspectrale se présente sous la forme d'un cube de données à deux dimensions spatiales et une dimension spectrale. En analyse d'images multivariées, les dimensions spatiales sont ignorées. Le cube des données est déplié pour obtenir un tableau à deux dimensions pouvant être analysé avec toute la panoplie des méthodes chimiométriques. La méthode de base de l'analyse d'images hyperspectrales multivariées est l'analyse en composantes principales [Geladi et Grahn, 1996].

Les méthodes multibloc ont été retenues pour analyser d'une manière conjointe les images hyperspectrales [Hanafi et Kiers, 2006 ; Hanafi *et al.*, 2011]. L'analyse conjointe nécessite l'appariement des données issues de chaque image hyperspectrale et l'établissement de la

structure des données reliant les différents blocs. Dans notre cas les images hyperspectrales n'ont pas la même résolution spatiale. Le pixel de l'image infrarouge a une surface de $5 \times 5 \mu\text{m}$. La taille du pixel en fluorescence et en Raman est de $1 \mu\text{m} \times 1 \mu\text{m}$. Chaque spectre infrarouge est apparié avec l'ensemble des spectres qui lui correspondent en fluorescence et en Raman. Cette structure de données ne peut pas être analysée avec les méthodes multibloc existantes. Des approches seront développées en étendant les méthodes développées dans le cadre de l'analyse conjointe des données spectrales [Mazerolles *et al.*, 2002 ; Hanafi *et al.*, 2006 ; Böcker *et al.*, 2007].

Objectif et démarche de la thèse

L'objectif de la thèse est de développer toutes les méthodes et les approches pour cartographier les variations chimiques dans les tiges de maïs à l'aide d'images hyperspectrales acquises avec les trois microspectroscopies : infrarouge, fluorescence UV et Raman. Ce travail devait tenir compte de toutes les étapes de l'analyse depuis l'acquisition de données jusqu' aux développements d'outils nécessaires à l'interprétation des variations spectrales.

La première étape pour réaliser ces objectifs était de développer un protocole d'acquisition conjointe d'images hyperspectrales pour la même section de l'échantillon au synchrotron SOLEIL.

La deuxième étape a été consacrée à l'analyse exploratoire de chaque image hyperspectrale : infrarouge, Raman et de fluorescence. Les données spectrales devaient être corrigées et normalisées. Elles étaient analysées en utilisant une méthode d'analyse multivariée pour déterminer l'information apportée par chaque méthode spectrale.

La troisième étape consistait à appairer les trois ensembles de données et à établir la structure de données. Des techniques de recalage d'images et d'appariement de données spectrales sont indispensables pour réaliser cette troisième étape.

La quatrième étape a concerné le développement des méthodes multibloc pour analyser la structure de données établie. Les méthodes développées sont basées sur des travaux de Hanafi *et al.* [Hanafi et Kiers, 2006 ; Hanafi *et al.*, 2011].

Organisation du manuscrit

Le manuscrit est organisé en quatre parties. La partie A décrit le contexte scientifique de la thèse. Elle est composée de trois grandes sections : la première section présente l'organisation des parois cellulaires dans le maïs et les verrous à leurs dégradabilité.

La deuxième section présente les méthodes de microspectroscopies utilisées dans le cadre de la thèse : l'infrarouge, la fluorescence et le Raman. La troisième section présente les méthodes chimiométriques descriptives utiles pour l'analyse spectrale en commençant par l'analyse en composantes principales jusqu'aux méthodes d'analyse des données structurées en blocs.

La partie B de ce manuscrit s'intéresse à la présentation des données acquises au cours de la thèse. Cette partie commence par la description du protocole d'acquisition conjointe des images hyperspectrales sur un même échantillon et finit par l'exploration des données acquises par analyse en composantes principales. L'exploration des images hyperspectrales d'une manière individuelle nous donne une idée sur l'information apportée par chaque modalité. Des méthodes de pré-traitement des images hyperspectrales ont été développées dans le cadre de cette partie. Ces méthodes ont été appliquées aux données présentées dans cette partie.

La partie C de ce manuscrit décrit l'appariement des images hyperspectrales multimodales et l'établissement de la structure de données. Cette partie inclut la présentation de la méthode de recalage utilisée pour apparier les données et la stratégie d'appariement des spectres issus de chaque image hyperspectrale.

La partie D présente les méthodes développées pour d'analyse conjointe d'images hyperspectrales multirésolutions et multimodalités. Dans une première section, le couplage de deux images hyperspectrales multirésolutions est développé. La deuxième section présente la méthode pour coupler plusieurs images hyperspectrales en dégradant les images à haute résolution spatiale. Cette méthode a été étendue dans la troisième section pour supporter l'analyse des images hyperspectrales multirésolutions sans altérer leurs résolutions spatiales.

Partie A :

Introduction à la spectroscopie et à l'analyse d'images hyperspectrales pour la cartographie chimique de tissus végétaux

Sommaire

Introduction	9
I Structure et composition chimique des parois cellulaires de tige de maïs et verrous à leur dégradation enzymatique.....	9
1 Anatomie de la tige de maïs	9
2 La paroi cellulaire.....	11
3 Les constituants pariétaux	12
3.1 La cellulose	13
3.2 Les hémicelluloses	14
3.3 Les composés phénoliques	15
3.3.1 Acide férulique.....	15
3.3.3 Les lignines.....	16
3.4 Organisation des polymères dans la paroi.....	17
4 Barrières aux dégradations enzymatiques	18
4.1 Notions de dégradation enzymatique	18
4.2 Limites à la dégradation enzymatique.....	19
4.2.1 Accessibilité des polysaccharides aux enzymes	19
4.2.2 Absorption de l'enzyme	21
4.2.3 Dégradation en fonction des types cellulaires.....	21
5 Conclusion.....	21
II Spectroscopies et microspectroscopies	22
Introduction	22
1 Spectroscopie	22
1.1 Notions de spectroscopie.....	22

1.1.1 Onde électromagnétique (modèle ondulatoire)	22
1.1.2 Le Photon (modèle corpusculaire)	23
1.1.3 Interaction onde/matière	24
1.2 Spectroscopies vibrationnelles	24
1.2.1 Vibration des molécules polyatomiques	25
1.2.2 Règles de sélection	25
1.3 La spectroscopie infrarouge	26
1.3.1 Principe	26
1.3.2 Spectroscopie infrarouge des parois cellulaires de poacées	27
1.4 La spectroscopie Raman.....	29
1.4.1 Principe	29
1.4.2 Spectroscopie Raman des composants chimiques présents dans les parois cellulaires de poacées	29
1.5 La spectroscopie de fluorescence	31
1.5.1 Principe	31
1.5.2 Spectroscopie de fluorescence des parois cellulaires	31
2 Complémentarité entre les méthodes de spectroscopie.....	32
3 Microspectroscopies.....	32
3.1 Caractéristiques des microspectrophotomètres de laboratoire	32
3.2 Source synchrotron.....	33
3.2.1 Ligne de lumière SMIS	34
3.2.2 Ligne de lumière DISCO	35
3.3 Microspectroscopie des végétaux.....	36
3.3.1 Microspectroscopie infrarouge des polymères des parois cellulaires.. 36	
3.3.2 Microspectroscopie Raman des polymères des parois cellulaires	37
3.3.3 Microspectroscopie de fluorescence des polymères des parois cellulaires	37
4 Pré-traitement	38
4.1 Pourquoi le pré-traitement?	38
4.2 Etapes de pré-traitement.....	38
4.2.1 Lissage des spectres	38
4.2.2 Correction de la ligne de base	38
4.2.3 Normalisation	39
4.3 Combinaison de méthodes de pré-traitement	40
Conclusion.....	40

III Chimiométrie et analyse des données spectrales	41
1 Spectres, tableaux de spectres et images hyperspectrales	41
1.1 Spectres	41
1.2 Tableaux de spectres	41
1.2.1 Tableau à deux entrées	41
1.2.2 Tableaux à trois entrées	42
1.2.3 Ensemble de tableaux de données.....	43
1.3 Image hyperspectrale.....	43
1.3.1 Structure de données.....	43
1.3.2 Représentation graphique d'images hyperspectrales	44
2 Chimiométrie.....	45
2.1 Notations	45
2.2 Généralités sur les méthodes	46
2.3 Analyse en composantes principales.....	46
2.4 Analyse multibloc	49
2.4.1 Analyse de deux tableaux de données.....	49
2.4.2 Analyse de K tableaux de données	52
2.5 Analyse multivariée d'images hyperspectrales	54
2.5.1 Méthodes d'analyse d'images hyperspectrales.....	54
2.5.2 Analyse en composantes principales d'images hyperspectrales.....	55
Conclusion.....	56
VI Conclusion : Couplage d'image hyperspectrale pour la caractérisation chimique	
d'échantillons complexes	57
1 Position du problème.....	57
2 Démarche adoptée	58

Introduction

Cette partie décrit le contexte scientifique de notre projet. Elle présente dans un premier temps la structure et les composés constitutifs de la biomasse lignocellulosique. Dans un second temps, elle introduit les spectroscopies infrarouge, Raman et de fluorescence et montre leur apport pour la caractérisation du matériel lignocellulosique. Dans un dernier temps, cette partie décrit les outils chimiométriques adaptés à l'analyse des données spectrales d'une manière individuelle et d'une manière conjointe.

I Structure et composition chimique des parois cellulaires de tige de maïs et verrous à leur dégradation enzymatique

1 Anatomie de la tige de maïs

La tige de maïs est constituée de plusieurs segments appelés entrenœuds séparés par des nœuds. Les feuilles d'une plante de maïs apparaissent une à une à l'extrémité d'un entre-nœud (Fig. A1). L'épi fait suite à la fleur femelle et se place à l'aisselle d'une feuille dans la partie médiane de la tige.

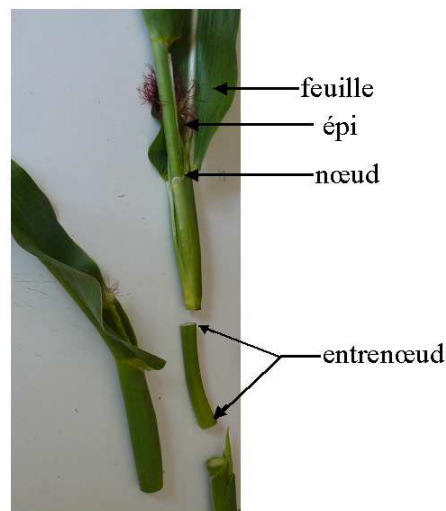


Figure A1 Anatomie de la tige de maïs

La figure A2-a présente une section transversale de tige de maïs. La tige est constituée de différents tissus, principalement le parenchyme et un ensemble de faisceaux vasculaires distribués dans toute la surface de la section. La tige est bornée par un ensemble de petites cellules à parois condensées et de faisceaux vasculaires formant l'écorce. Dans la partie externe de la tige, une assise de petites cellules jointives est observée. Elle constitue l'épiderme (Fig. A2-b).

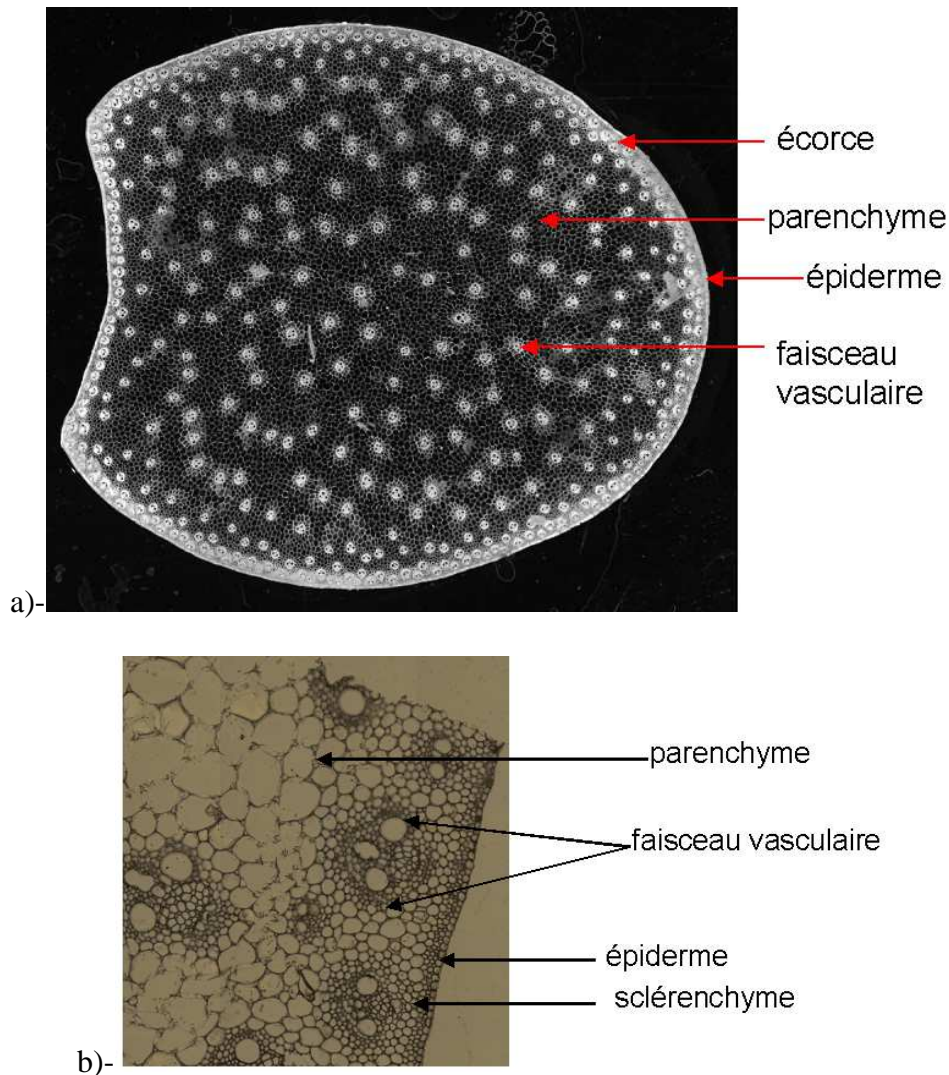


Figure A2 Types cellulaires de la tige de maïs a)- coupe transversale d'une tige de maïs acquise avec le système de macrovision 'BlueBOX' (INRA-Nantes) b)- coupe transversale d'une tige de maïs acquise au microscope infrarouge (INRA-Nantes).

L'écorce est un tissu de protection constituée par une assise continue de cellules imperméables qui permettent à la plante de conserver son eau (Fig. A2) [Taiez et Zeiger, 2002 ; Evert, 2006]. Les parenchymes sont des tissus peu différenciés. Ils assurent, dans la tige, le stockage des réserves. Ces réserves peuvent être notamment des polysaccharides (amidon) ou du saccharose. Les faisceaux vasculaires sont responsables de la circulation de la sève brute et de la sève élaborée dans la plante. La sève brute correspond à la sève ascendante puisée des racines pour être redistribuée aux parties végétatives. La sève élaborée correspond à la sève descendante. Elle provient essentiellement des feuilles, siège de la photosynthèse. Dans la tige, les faisceaux vasculaires sont constitués essentiellement de xylème, de phloème et de sclérenchyme (Fig. A3).

Le xylème et le phloème assurent la circulation de la sève brute et de la sève élaborée respectivement. En plus des éléments des vaisseaux, le tissu xylémien comprend des cellules

parenchymateuses. Le phloème est borné d'un côté par le protophloème qui s'est compressé au cours du développement de la plante et de métaphloème. Le transport de la sève élaborée est gouverné par les complexes cellules compagnes/tubes criblés qui représentent un tissu vivant. Le troisième type cellulaire constituant un faisceau vasculaire est le sclérenchyme. Le sclérenchyme assure le soutien de la plante. Il est caractérisé par des parois épaisses et lignifiées.

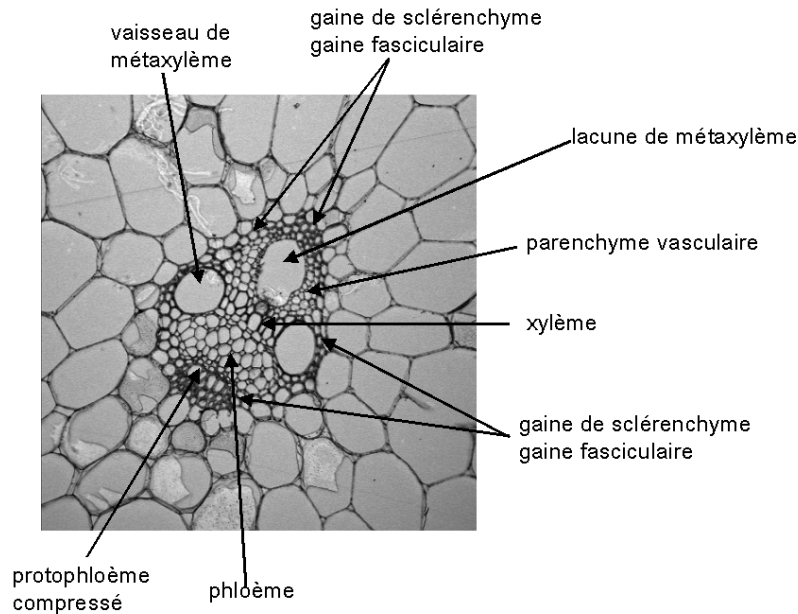


Figure A3 Faisceau d'une tige de maïs en coupe transversale. Image en fond clair acquise au microscope confocal (plateforme BIBS- INRA (Nantes)).

2 La paroi cellulaire

La cellule végétale est entourée d'une paroi qui définit la morphologie de la cellule de la plante. Dans le cas des végétaux lignocellulosiques, elle est constituée principalement de polysaccharides (cellulose, hémicelluloses et en plus faible proportions, de pectines) et de composés de nature phénolique (lignines, acide férulique, acide para-coumarique...) [Taiez et Zeiger, 2002]. Elle contient également des protéines et des composantes inorganiques minérales.

La paroi cellulaire (Fig. A4) est formée de la lamelle moyenne, la paroi primaire et pour certains tissus, de la paroi secondaire [Evert, 2006]. La lamelle moyenne est la couche située entre deux cellules assurant l'adhésion intercellulaire. La paroi primaire est fine et souple (environ 100 nm d'épaisseur), fortement hydrophile (60% de l'eau). Elle est constituée d'une matrice riche en pectines, d'un réseau de cellulose et d'hémicelluloses et de protéines. La paroi

secondaire est présente dans les tissus de soutien et les tissus conducteurs. Elle se dépose sur la paroi primaire après l'arrêt de l'expansion et de la différenciation des cellules. Dans la majorité des cas, elle est constituée de trois strates: S1, la première à partir de la paroi primaire, S2, et enfin S3, la couche la plus interne à la cellule. La couche S2, généralement la plus épaisse, détermine la résistance mécanique des cellules. Cette paroi secondaire est riche en cellulose, rigide et inextensible. La paroi des cellules du sclérenchyme et des vaisseaux de xylème est lignifiée ce qui renforce sa rigidité. La paroi secondaire peut atteindre une épaisseur considérable surtout dans les cellules de soutien [Evert, 2006].

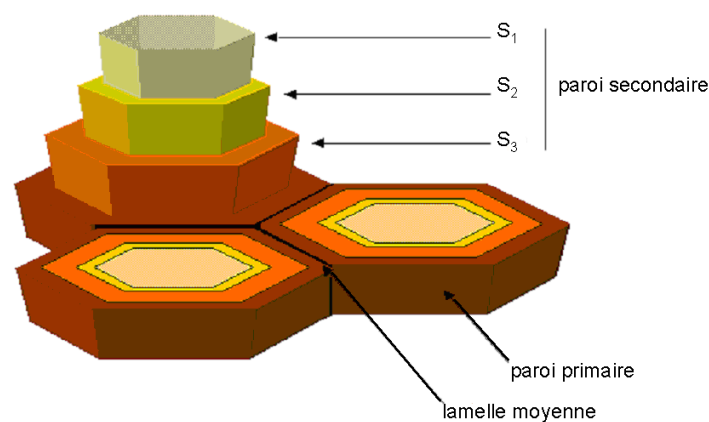


Figure A 4 Représentation schématique de la paroi cellulaire des végétaux

3 Les constituants pariétaux

La composition des parois varie selon les espèces végétales, le stade de développement de la plante et les types cellulaires. Les polysaccharides représentent la partie la plus importante des constituants de la paroi végétale de 60% à 85%. Les lignines sont quantitativement importantes, puisqu'elles constituent entre 10 et 25% de la biomasse végétale terrestre (chiffre variant avec les espèces et le stade de développement) [Pauly et Keegstra, 2008].

La composition chimique des parois cellulaires des tiges de maïs a été estimée par Lynd *et al.*, (1999). Cette composition est présentée dans le tableau ci-dessous :

Tableau A1 Composition approximative (% poids sec) des parois cellulaires des tiges de maïs mature [Lynd *et al.*, 1999]

cellulose	hémicellulose	Lignine	Protéines	Cendre
46%	32%	12%	5%	5%

Le tableau A1 montre que la cellulose est le constituant chimique le plus abondant. La teneur en lignines est relativement faible par rapport aux espèces ligneuses pour lesquels la teneur en lignines peut atteindre jusqu'à 25%. Dans cette étude le pourcentage d'acide phénolique n'a pas été présenté. Une autre étude sur les poacées a montré que le pourcentage des acides phénoliques est de 5% [Vogel, 2008]. Dans cette section, les composants chimiques les plus abondants dans la biomasse lignocellulosique sont présentés.

3.1 La cellulose

La cellulose représente 10-14% de la matière sèche des parois primaires et 40-60% de la matière sèche des parois secondaires [Hématy et Höfte, 2007]. C'est un homopolymère non ramifié constitué de longues chaînes linéaires d'unités D-glucose liées entre elles par une liaison glucosidique β -1-4 (Fig. A5).

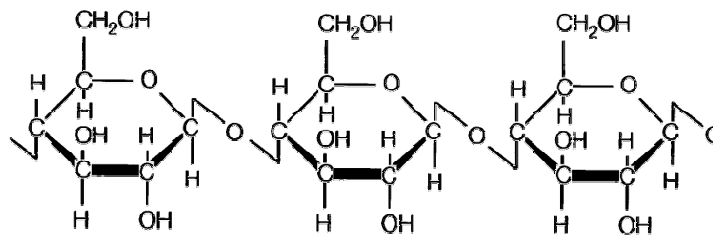


Figure A5 Molécule de cellulose

Les chaînes de cellulose sont organisées parallèlement sous forme de microfibrilles (Fig. A6). La disposition relative des unités de glucose permet l'établissement de liaisons hydrogène intra et inter-chaînes, stabilisant la cellulose. Ces liaisons hydrogènes sont responsables de l'établissement de régions cristallines au sein des microfibrilles.

Dans la paroi primaire, les microfibrilles sont orientées aléatoirement, ce qui permet l'expansion des cellules dans toutes les directions. Une légère orientation parallèle à l'axe de croissance de la plante, déterminé par l'ensemble des contraintes s'exerçant sur la cellule permet une croissance anisotrope de celle-ci [Crowell *et al.*, 2010]. Le diamètre des microfibrilles dans la paroi primaire est inférieur à celui observé dans la paroi secondaire. Dans la paroi secondaire, les microfibrilles présentent une orientation privilégiée variant entre les différentes strates composant cette paroi ce qui a pour effet d'augmenter sa résistance mécanique.

La structure de la cellulose et son organisation en microfibrilles avec une présence de domaines cristallins explique sa résistance marquée aux stress mécaniques et à la dégradation, et sa faible solubilité dans l'eau.

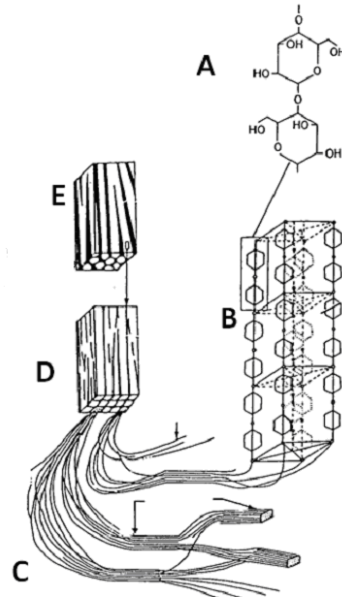


Figure A6 Structure des microfibrilles de cellulose : A : cellobiose unité de base de la cellulose constituée de deux glucose liés en β -(1-4) par une liaison glycosidique ; B : chaîne de cellulose ; C : structure de microfibrilles, assemblage de molécules de cellulose ; D : fragment de macrofibrilles ; E : fragment de la couche pariétale montrant des microfibrilles de cellulose (blanc) et une matrice de hémicellulose/lignine (noir) [Esau, 1977].

3.2 Les hémicelluloses

Les hémicelluloses sont des hétéropolymères d'oses, constitués d'une chaîne principale de monomères liés par des liaisons β -1-4, pouvant porter des chaînes latérales constituées de plusieurs oses [Scheller et Ulvskov, 2010]. Elles peuvent également porter des groupements acétyles. Leur degré de polymérisation est dix à cents fois inférieur à celui de la cellulose [Bledzki et Gassan, 1999]. Les hémicelluloses sont globalement le constituant pariétal le moins résistant : ces polymères se biodégradent rapidement [Zhang et al., 2011-a].

Les hémicelluloses regroupent les xyloglucanes, les xylanes, les mannanes, les beta (1-3)(1-4) glucanes. Les hémicelluloses constituent une matrice autour des microfibrilles des celluloses. Chez les poacées, les hémicelluloses sont majoritairement des xylanes (environ 25% des parois de tiges matures). Ces polysaccharides sont composés d'unités xylose liées en β -1-4, et substitués par des unités d'arabinose en C-2 et C-3 (arabinoxylanes) (Fig. A7) et plus rarement par des unités d'acide α -D-glucuronique ou son dérivé 4-O méthyle en position C-2

du xylose de la chaîne principale (glucurono(arabino)xylanes). Des résidus D-galactose et D-glucose sont également présents en plus faible quantité. Certains arabinoxylanes sont substitués irrégulièrement par des groupements acétyles en position C-2 et C-3 des unités xylose. Les xylanes peuvent être estérifiés à la position C-5 des unités arabinose par l'acide férulique et l'acide para-coumarique.

La nature des substituants, le degré de substitution des xylanes varie selon l'espèce botanique mais aussi en fonction des tissus et le stade de développement de la plante. La substitution des xylanes par l'arabinose est en général plus élevée dans les tissus jeunes comparés aux tissus plus âgés dans les tiges de maïs [Jung et Casler, 2006-a].

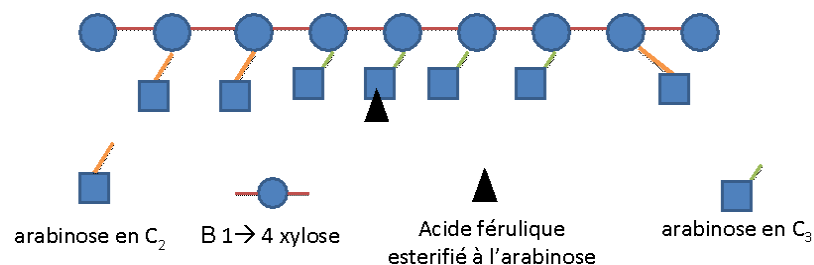


Figure A7 Structure chimique de l'arabinoxylane

3.3 Les composés phénoliques

Les parois des poacées sont caractérisées par la présence d'acides hydroxycinnamiques (acide férulique et acide para-coumarique). Les deux acides se distinguent par le nombre de groupements métoxyles présents sur le cycle aromatique.

3.3.1 Acide férulique

L'acide férulique (Fig. A8) est l'acide hydroxycinnamique majoritaire des parois non lignifiées des graminées. Il est présent estérifié aux unités arabinose des xylanes ou liés aux unités G des lignines par des liaisons éther. L'oxydation de l'acide férulique au sein de la paroi aboutit à la formation de dimères d'acide férulique. Ces dimères permettent la formation de ponts covalents entre les polymères qui les portent. L'acide férulique joue un rôle dans la réticulation des arabinoxylanes des parois cellulaires et l'établissement de liaisons covalentes xylanes-lignines. Il se dépose précocement dans la paroi et servirait de point d'ancrage à la polymérisation et réticulation des lignines [Hatfield et Marita, 2010].

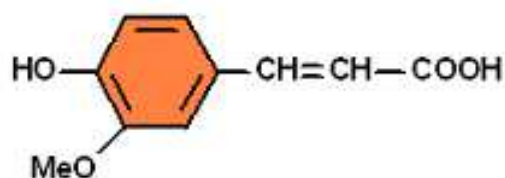


Figure A8 Structure de l'acide férulique

3.3.2 Acide para-coumarique

L'acide para-coumarique (Fig. A9) peut être estérifié à l'arabinose des xylanes mais on le trouve majoritairement lié par une liaison ester aux unités S des lignines. C'est d'ailleurs un indicateur du dépôt de ces unités dans les parois. Il peut représenter jusqu'à 20% du contenu en lignines chez le maïs et le sorgho [Morrison et Archibald, 1998 ; Hatfield, *et al.*, 1999 ; Ralph *et al.*, 2004]. Son rôle n'est pas encore complètement élucidé.

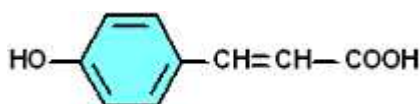


Figure A9 Structure chimique de l'acide para-coumarique

3.3.3 Les lignines

Les lignines sont essentielles pour la plante puisqu'elles permettent sa protection et son port dressé [Siegel et Puerner, 1969]. Leurs trois principales unités sont les unités *p*-hydroxyphényle (H), guaiacyle (G) et syringyle (S) qui sont produites à partir de 3 monolignols, respectivement l'alcool *p*-coumarylique, l'alcool coniférylique et l'alcool sinapylique [Boerjan *et al.*, 2003]. La présence du groupement métoxyle (OCH₃) fait la différence entre les trois unités (Fig. A10). L'unité H ne contient pas de groupement métoxyle. Les unités G et S contiennent respectivement un et deux groupements métoxyyles. Ces monolignols sont oxydés en radicaux phénoliques, grâce à des oxydases (peroxydases et laccases). Les radicaux obtenus se couplent ensuite spontanément. Ces unités sont assemblées via des liaisons labiles β-O-4 et des liaisons résistantes carbone-carbone ou biphényle éther (Fig. A11-a).

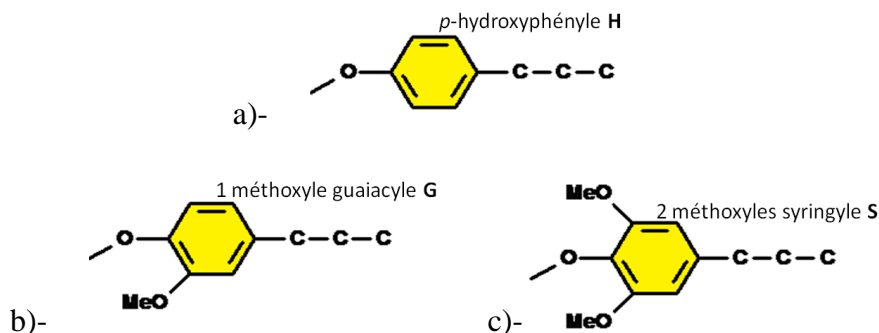


Figure A10 Unités H, G et S de la macromolécule de lignine a)- unité H b)- unité G c)- unité S

La structure des lignines dépend de l'espèce végétale, du stade de développement de la plante et des tissus. Les lignines de poacées (angiospermes monocotylédones) sont constituées des trois unités G, S et H (les unités H restant minoritaires) et se différencient des lignines d'angiospermes dicotylédones par la présence d'acide *p*-coumarique et d'acide férulique. Chez le maïs, les unités H représentent en moyenne 4% des lignines, les unités G, 35% et les unités S, 61% [Barrière *et al.*, 2007]. Les parois des cellules des faisceaux vasculaires contiennent principalement des lignines riches en G alors que les parois des cellules interfasciculaires contiennent des lignines comportant des unités G et S [He et Terashima, 1991]. Au cours de la maturation de la plante, les lignines ont tendance à s'enrichir en unités S [He et Terashima, 1991].

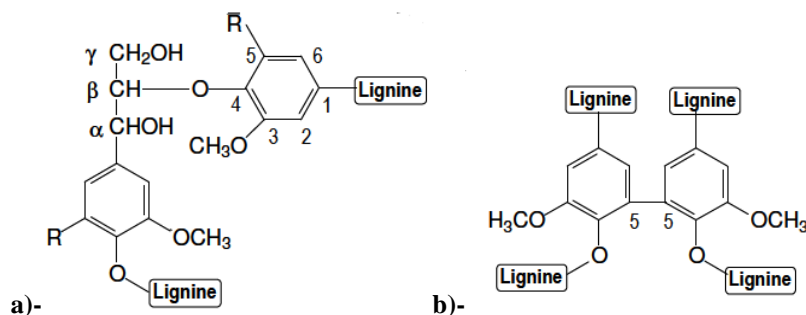


Figure A11 Représentation simplifiée a)-d'un mode de liaison inter-unité labile (β -O-4) et b)-d'un mode de liaison inter-unité biphenyle (exemple de liaison condensée)

3.4 Organisation des polymères dans la paroi

Les polymères constitutifs de la paroi cellulaire s'organisent en un réseau tridimensionnel (Fig. A.12). La cohésion de cette structure est assurée par des liaisons de natures diverses. Chez les poacées, les microfibrilles de cellulose sont enchâssées dans une matrice non cellulosique composée par des xylanes et/ou par des lignines. Plus particulièrement, les

poacées renferment des acides phénoliques qui ont un rôle très important dans l'établissement des liaisons au sein de la paroi. En effet, ils participent à la réticulation des polymères pariétaux en formant des ponts entre les polysaccharides et/ou entre les lignines par des liaisons de type ester qui se fixent généralement sur les lignines [Grabber *et al.*, 2004 ; Buranov et Mazza, 2008].

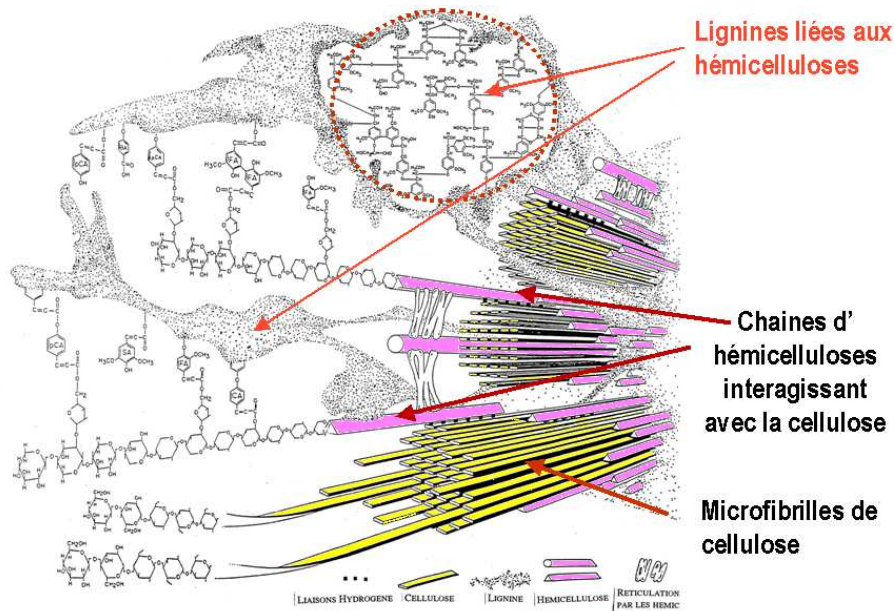


Figure A12 Représentation schématique d'une paroi secondaire de poacées [Bidlack *et al.*, 1992] : les microfibrilles de cellulose sont enchassées dans une matrice d'hémicelluloses et de lignines. Ces polymères peuvent être réticulés par l'acide férulique

Au-delà de des multiples liaisons covalentes au sein de la matrice des parois cellulaires, des interactions non covalentes assurent la cohésion entre cette matrice et l'armature cellulosique. Les régions non substituées par des arabinoxylanes sont susceptibles de s'associer avec la cellulose par des liaisons hydrogènes [Carpita et Gibeaut, 1993]. Ces nombreuses liaisons renforcent les cohésions pariétales et rendent la paroi plus difficilement biodégradable [Hatfield *et al.*, 1999].

4 Barrières aux dégradations enzymatiques

4.1 Notions de dégradation enzymatique

Dans ce paragraphe, nous allons présenter brièvement la dégradation enzymatique. La dégradation d'une molécule se fait en plusieurs étapes : la diffusion des enzymes et l'accessibilité à leur substrat, l'absorption des enzymes et l'hydrolyse proprement dite (Fig. A13). Pour que la dégradation soit efficace, il faut que l'enzyme accède aux liaisons.

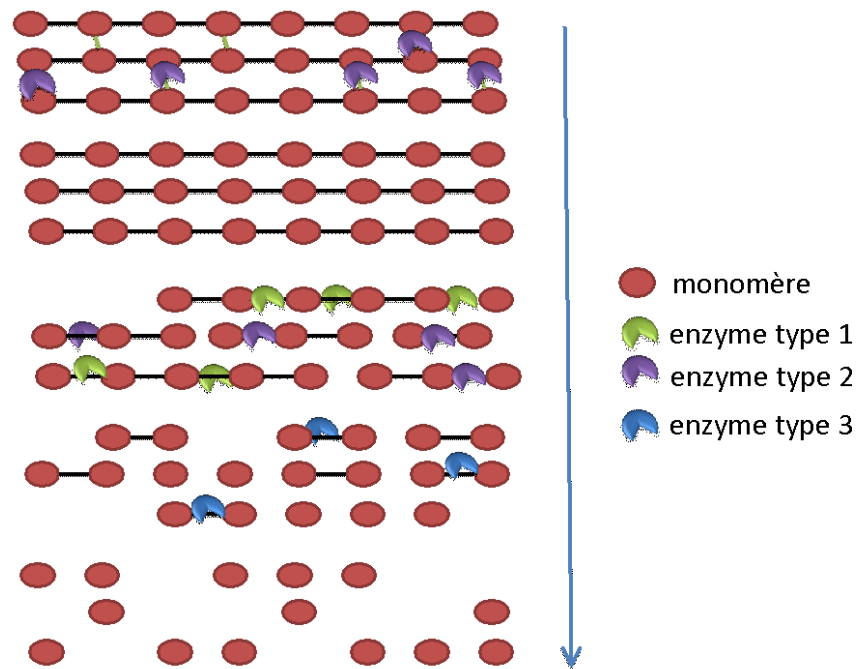


Figure A13 Etapes de la dégradation enzymatique

La dégradation enzymatique d'une molécule dépend fortement de l'accessibilité des liaisons des polymères aux enzymes. Si les liaisons sont accessibles pour un réseau lignocellulosique donné, la dégradation peut se faire avec succès. Par contre, si ce même type de liaisons est caché ou masqué par d'autres types de molécules telles que la lignine, la dégradation ne sera pas possible.

4.2 Limites à la dégradation enzymatique

La dégradation de la biomasse lignocellulosique dépend non seulement de la composition chimique du matériel à dégrader mais également de l'organisation de ses constituants et de leurs interactions au sein de la paroi cellulaire. Les caractéristiques de la biomasse influant sur la dégradation sont listées ci-dessous :

4.2.1 Accessibilité des polysaccharides aux enzymes

L'organisation et en particulier les interactions entre les polymères pariétaux au sein de la paroi jouent un rôle clé dans l'accessibilité des enzymes à leur substrat. Les lignines occupent l'espace libre laissé par les autres constituants et de fait, jouent un rôle de protection physique en limitant le volume accessible aux enzymes au sein de la paroi. Méchin *et al.*, (1998, 2000) ont montré que la teneur en lignines dans la paroi de tige de maïs était le premier facteur explicatif des variations de dégradabilité entre différentes lignées. L'impact de la structure des lignines sur la dégradabilité des parois fait toujours débat. Dans le cas du maïs, une plus faible

dégradabilité des parois a été corrélée à la présence de lignines riches en unité S [Riboulet *et al.*, 2008]. Des travaux de modélisation moléculaire suggèrent que les lignines riches en liaison β -O-4 adoptent une structure linéaire capable de s'absorber à la surface de la cellule [Bessombes et Mazeau, 2005] limitant la dégradabilité de cette dernière. Cette hypothèse semble en accord avec des résultats récents soulignant une corrélation négative entre la teneur en unités S engagées dans des liaisons β -O-4 et la dégradabilité enzymatique de parois de maïs [Zhang *et al.*, 2011-b]. Grabber *et al.*, (2008) ont conclu que la teneur en lignines est plus déterminante de l'aptitude à la dégradation des parois que la composition des lignines en utilisant des parois primaires de culture cellulaire de maïs féruloylées ou non, artificiellement lignifiées avec des monolignols variés. Indépendamment de cet effet barrière physique, les lignines peuvent limiter la dégradation de la paroi à travers leurs propriétés hydrophobes et leur capacité à «fixer» les enzymes limitant leur progression au sein de la paroi [Rémond Zilliox *et al.*, 1997]. Les polysaccharides matriciels, hémicelluloses et pectines représentent également une barrière physique pour atteindre la cellulose. Leur dégradation est souvent un préalable pour une dégradation extensive de la cellulose [Devaux *et al.*, 2006]. Les interactions entre polymères sont également importantes. Ainsi la réticulation des arabinoxylyanes via la dimérisation de l'acide férulique diminue la dégradabilité des xylanes mais aussi des autres polymères de la paroi primaire, cellulose et pectines [Grabber *et al.*, 1997]. Les interactions covalentes lignines – hémicelluloses via l'acide férulique sont souvent mentionnées comme facteurs influençant la dégradation. Jung and Calser (2006-a) ont examiné la dégradabilité des parois dans les entrenœuds de tige de maïs en fonction des tissus et du stade de développement de l'organe. S'ils ont montré clairement que le taux de dégradation des polysaccharides était corrélé négativement avec la teneur en lignines, la relation entre dégradabilité et réticulation lignine-hémicellulose via l'acide férulique apparaît plus complexe. La corrélation est négative seulement pour les entrenœuds collectés durant les phases d'élongation. La fréquence de réticulation lignines-hémicellulose est en effet maximale juste avant la fin de l'élongation puis décroît ensuite. Cette diminution s'explique par une accumulation massive et rapide des lignines et une concentration en acide férulique plus faible dans la paroi secondaire comparativement à la paroi primaire. La plus faible association lignines – hémicelluloses dans les parois secondaires pourrait expliquer qu'en dépit de leur lignification, elles sont partiellement dégradées.

4.2.2 Absorption de l'enzyme

La capacité de l'enzyme de se fixer sur son substrat peut également limiter la dégradation. La présence de nombreux substituants sur le squelette peut gêner la fixation de l'enzyme. Ainsi, l'élimination des groupements acétyles augmente la dégradabilité des xylandes [Grohamann *et al.*, 1989]. La présence de nombreux substituants arabinose et acide glucuronique perturbent la reconnaissance par des endoxylanases de certains sites de coupure le long de la chaîne d'unités xylose [White *et al.*, 1993].

4.2.3 Dégradation en fonction des types cellulaires

Des différences de dégradabilité ont été rapportées entre les différents types cellulaires sur des échantillons de canne à sucre [Siqueira *et al.*, 2011] et de tiges de maïs [Jung et Casler, 2006-b]. Les variations observées de dégradabilité des tissus ont été mises en relation avec la composition des parois. Dans le cas de la canne à sucre, les parois des cellules du parenchyme dans le centre de la tige, peu lignifiées sont rapidement hydrolysées par une cellulase avec 63% de la cellulose dégradée après 72h d'hydrolyse. Dans l'écorce, seulement 20% de la cellulose est hydrolysée et une corrélation négative est observée entre la teneur en lignines des tissus et l'hydrolyse de la cellulose. Dans les entrenœuds de tiges de maïs [Jung et Casler, 2006-a], au stade jeune correspondant à la période d'élongation, tous les tissus à l'exception du protoxylème, qui est le seul tissu lignifié à ce stade, sont dégradés. Après ce stade d'élongation, la dégradabilité des tissus décroît à l'exception de celle du phloème qui n'est pas lignifié. Les tissus ayant des parois secondaires épaisses et lignifiées nécessitent un temps d'incubation en présence des enzymes plus long avant d'observer une dégradation.

5 Conclusion

Les parois cellulaires des tiges de végétaux lignocellulosiques sont constituées en majorité de polysaccharides pouvant être dégradés - saccharifiés - sous la forme de monomères utilisables pour la production de biocarburants de deuxième génération [Naik *et al.*, 2010]. Dans la paroi, les microfibrilles de cellulose forment un réseau inclus dans une matrice d'hémicellulose. Les lignines et les réticulations des polymères au sein de cet ensemble peuvent rendre la paroi très résistante à la dégradation. Afin de mieux comprendre les facteurs de dégradation et d'optimiser les procédés de pré-traitements et de digestion enzymatique, les relations entre les spécificités structurales et histologiques des lignocelluloses et leur dégradabilité enzymatique

doivent être mieux appréhendées. Ces données doivent être recueillies à une échelle locale (au sein des tissus), là où les méthodes globales de biochimie sont inadaptées (étape d'extraction limitée et échelle de caractérisation inappropriée).

II Spectroscopies et microspectroscopies

Introduction

Les méthodes de spectroscopie sont des outils puissants pour la caractérisation des échantillons biologiques. Elles fournissent des informations sur la structure moléculaire des échantillons analysés sous la forme de spectres plus ou moins complexes. L'interprétation de ces spectres permet de révéler les composants chimiques constituant l'échantillon considéré.

Cette partie du manuscrit présente les méthodes de spectroscopie infrarouge, Raman et de fluorescence appliquées en particulier pour la caractérisation des constituants chimiques présents dans les parois cellulaires de poacées [Robert *et al.*, 2005 ; Himmelsbach et Akin, 1998 ; Andersson et Akin, 2008]. Dans une première section, les principes physiques de la spectroscopie sont brièvement introduits. La section suivante est dédiée à la présentation de la spectroscopie infrarouge ainsi que son application pour la caractérisation des polymères présents dans les parois cellulaires de poacées. Dans les sections 1.2 et 1.3, le même schéma d'étude est repris pour les spectroscopies Raman et de fluorescence. La cartographie *in situ* des parois cellulaires nécessite de recueillir l'information d'une manière localisée. Dans ce cas, la microspectroscopie est alors la méthode appropriée. Dans la deuxième section, les techniques de microspectroscopies infrarouge, Raman et de fluorescence ainsi que leurs applications analytiques pour la caractérisation des parois cellulaires sont présentées. Nous finirons ce chapitre, par la présentation des méthodes de pré-traitement des spectres dans le but de les rendre facilement exploitables.

1 Spectroscopie

1.1 Notions de spectroscopie

1.1.1 Onde électromagnétique (modèle ondulatoire)

Le spectre électromagnétique s'étend des ondes radio aux rayons Gamma (Fig. A14). Les radiations électromagnétiques sont définies par leur longueur d'onde λ ou leur fréquence ν . La fréquence est le nombre de longueur d'ondes par unité de temps.

La lumière peut être ainsi caractérisée par une longueur d'onde λ ou une fréquence ν suivant la relation suivante:

$$\nu = \frac{c}{\lambda} \quad \text{Equation A1}$$

où c est la vitesse de la lumière dans le vide $c = 3 \times 10^8 \text{ ms}^{-1}$

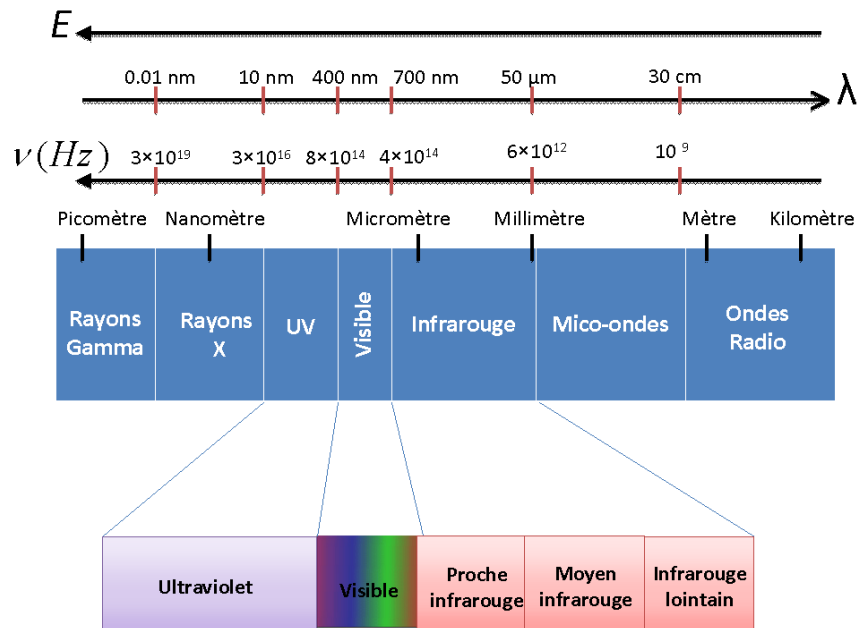


Figure A14 Spectre électromagnétique

1.1.2 Le Photon (modèle corpusculaire)

Selon le modèle quantique, on peut également définir la lumière comme étant composée de petites particules appelées photons. Les photons ont la propriété de transporter une quantité d'énergie E . L'énergie E d'un photon de fréquence ν est donnée par la relation :

$$E = h\nu \quad \text{Equation A2}$$

D'après les équations A1 et A2 :

$$E = h\nu = h\frac{c}{\lambda} \quad \text{Equation A3}$$

où h est la constante de Planck $h = 6.626 \times 10^{-34} \text{ Js}$

L'énergie E est proportionnelle à la fréquence ν et inversement proportionnelle à la longueur d'onde λ . L'énergie E est discrète. Elle est répartie sur plusieurs niveaux, appelées niveaux vibrationnels.

1.1.3 Interaction onde/matière

Toute molécule possède une énergie électronique répartie sur différents niveaux. Chaque niveau électronique comporte plusieurs niveaux vibrationnels (Fig. A15). Lorsqu'une onde électromagnétique rencontre la matière, de l'énergie peut être transférée aux molécules, ainsi elles passent d'un état S_0 à un état S_1 associés à deux niveaux d'énergie différents correspondant respectivement à l'état stable et au premier état excité d'une molécule donnée. Ce changement d'énergie peut se manifester sous forme d'absorption, de diffusion, d'émission d'ondes électromagnétiques. L'analyse de ces différents types d'interactions entre l'onde et la matière donne naissance à différents types de spectroscopies.

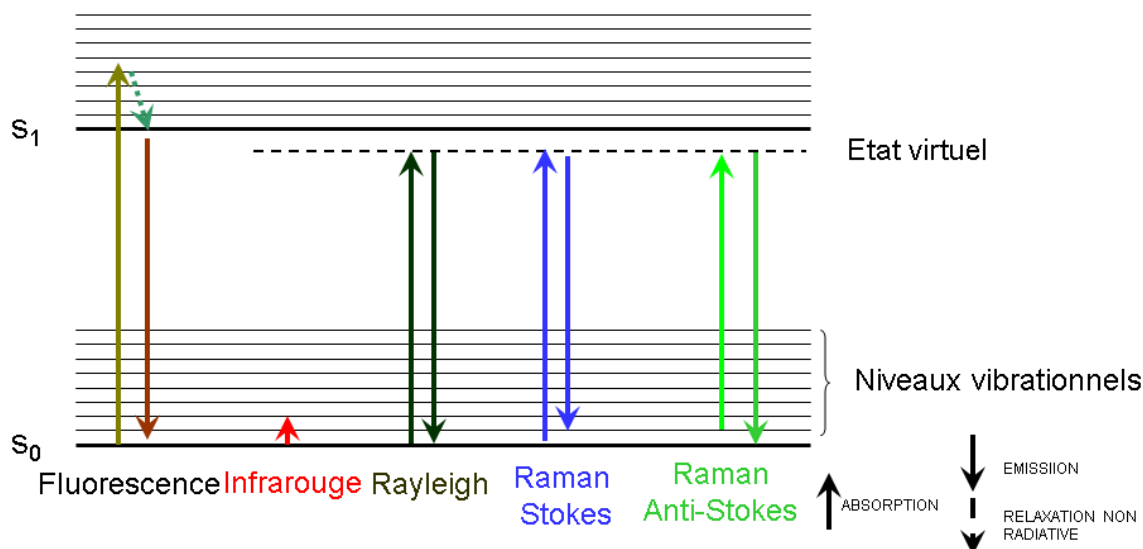


Figure A15 Diagramme de Jablonski : Représentation des transitions électroniques pour une molécule diatomique. S_0 représente l'état fondamental de la molécule, S_1 représente l'état excité.

Trois types de phénomènes résultants de l'interaction de la lumière avec la matière sont à la base de la spectroscopie dont on peut citer l'absorption infrarouge, la diffusion Raman et l'émission de fluorescence.

1.2 Spectroscopies vibrationnelles

L'infrarouge et le Raman sont deux techniques de spectroscopie vibrationnelle. La nature différente des processus d'interaction à l'origine de l'effet Raman et de l'infrarouge font que certaines vibrations des liaisons chimiques de la molécule seront seulement actives en infrarouge et d'autres seulement actives en Raman, et d'autres le seront pour les deux ou ni pour l'une ni pour l'autre.

1.2.1 Vibration des molécules polyatomiques

Les mouvements des molécules, suite à leur vibration, se présentent sous la forme d'élongations ou de déformations. Les élongations (symétriques ou antisymétriques) font intervenir la longueur de liaisons, les angles formant ces liaisons restent constants. Pour les déformations, les liaisons gardent leur longueur, mais les angles qu'elles forment varient.

Il existe différentes règles pour savoir si un mode de vibration est actif ou pas en infrarouge ou en Raman appelées règles de sélection.

1.2.2 Règles de sélection

Un mode de vibration est actif en infrarouge s'il y a un changement du moment dipolaire. Il est actif en Raman, s'il y a un changement de la polarisabilité. Si la molécule a un centre de symétrie, alors aucun mode de vibration n'est actif à la fois en infrarouge et en Raman (Tableau A2). Si la molécule ne possède pas de centre de symétrie alors certaines vibrations seront à la fois actives en infrarouge et en Raman.

Pour illustrer les règles de sélection, nous présentons la molécule CO_2 comme exemple.

La molécule CO_2 , étant linéaire, possède 4 modes de vibration -2 modes d'élongation et 2 modes de déformations- (Fig. A16-17).

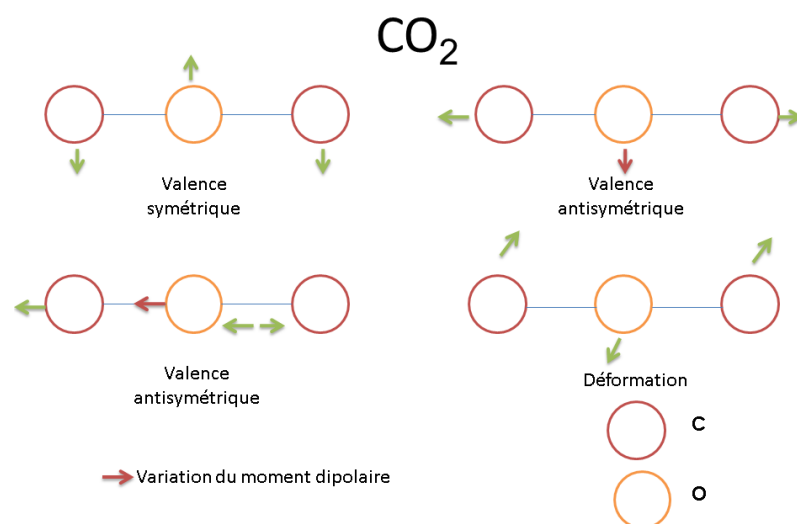
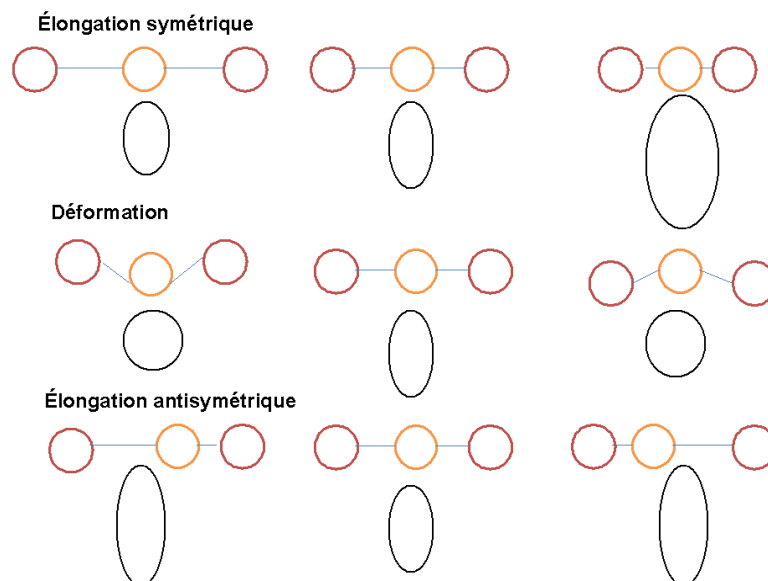


Figure A16 Vibration de la molécule CO_2 (changement du moment dipolaire)

Figure A17 Vibration de la molécule CO₂ (changement de la polarisabilité)Tableau A2 Vibration infrarouge et Raman de la molécule CO₂

Mode de vibration	Infrarouge	Raman
élongation symétrique	Inactive	Active
déformation	Active	Inactive
élongation antisymétrique	Active	Inactive

L'infrarouge est actif lorsqu'il y a une variation du moment dipolaire. Dans le cas de l'élongation symétrique (Fig. A16 et 17), le moment dipolaire ne varie pas. Par conséquent, l'infrarouge est inactif pour une élongation symétrique de la molécule CO₂.

Pour l'élongation antisymétrique et les déformations, le moment dipolaire est variable. Dans ce cas, l'infrarouge est actif.

Le changement de polarisabilité ne se fait que pour l'élongation symétrique. De ce fait, le Raman n'est actif que quand il s'agit d'une élongation symétrique. Il reste inactif pour la déformation et l'élongation antisymétrique (Tableau A2).

1.3 La spectroscopie infrarouge

1.3.1 Principe

En détectant les vibrations caractéristiques des molécules, la spectroscopie infrarouge permet d'effectuer l'analyse des fonctions chimiques présentes dans le matériau. Les spectres infrarouge sont calculés à partir de l'absorbance A en se basant sur la loi de Beer-Lambert :

$$A = \epsilon \cdot c \cdot d = -\log\left(\frac{I}{I_0}\right) \quad \text{Equation A4}$$

où

ε le coefficient d'absorption

c la concentration de l'échantillon

d l'épaisseur de l'échantillon

I intensité transmise

I_0 intensité incidente

1.3.2 Spectroscopie infrarouge des parois cellulaires de poacées

Les différents constituants des parois de poacées ont été étudiés en spectroscopie infrarouge par de nombreux auteurs. Les principaux composés lignocellulosiques étudiés dans la littérature en microspectroscopie infrarouge sont :

- les acides phénoliques
- les lignines
- la cellulose
- les hémicelluloses telles que les xylanes, les arabinoxylanes, les xyloglucanes et les

arabinogalactanes.

Le tableau A3 résume les nombres d'ondes caractéristiques des composants chimiques listés ci-dessous présents dans les parois cellulaires de poacées.

Les spectres infrarouge des composants chimiques des parois cellulaires de poacées sont répartis en trois régions spectrales principales :

- la première région 1800-1200 cm^{-1} révèle les lignines et les composés phénoliques.
- la deuxième région 1200-900 cm^{-1} révèle les polysaccharides tels que la cellulose et

les hémicelluloses.

- la troisième région 900-800 cm^{-1} les lignines et l'anomérisation α ou β de la liaison glycosidique des polysaccharides.

Pour un produit donné, l'attribution des bandes n'est pas forcément la même si ce produit est analysé dans un mélange complexe ou sous la forme de produit pur. Ceci est lié au fait que l'absorbance à une longueur d'onde donnée ne résulte pas d'une composante chimique unique et que certaines bandes peuvent se superposer ou au fait que certaines constituants chimiques sont masqués par d'autres constituants. Par exemple, on ne peut pas distinguer l'acide férulique de l'acide para-coumarique en faisant l'analyse d'une paroi cellulaire en microspectroscopie infrarouge. L'utilisation d'autres techniques de microspectroscopie s'avère indispensable pour valider la présence d'un composant chimique au sein de la paroi.

TableauA3 Nombre d'ondes infrarouge caractéristiques des composantes chimiques des parois cellulaires de poacées

Longueurs d'ondes caractéristiques (cm ⁻¹)	Vibration	Composantes chimiques	Référence bibliographique
1720	élongation C=O	lignine acide para-coumarique	Sun <i>et al.</i> , 2000
1665	élongation conjuguée C=O	lignine	Kacurakova <i>et al.</i> , 2002
1638	élongation conjuguée C=O	lignine	Xiao <i>et al.</i> , 2001
1595	vibration aromatique C=C	lignine	Joseleau et Ruel, 1997
1510	vibration aromatique C=C	lignine	Joseleau et Ruel, 1997
1467	vibration aromatique C=C et déformation C-H	lignine	Sun <i>et al.</i> , 2000
1466	déformation C-H	lignine	Xiao <i>et al.</i> , 2001
1426	élongation C-H	lignine	Xiao <i>et al.</i> , 2001
1420	vibration aromatique C=C	lignine	Sun <i>et al.</i> , 2000
1387	CH ₃	lignine	Sun <i>et al.</i> , 2000
1378	déformation acide hydroxycinnamique C-O	acide férulique	Kacurakova <i>et al.</i> , 1999
1375	phenolique O-H	phénolique	Guo <i>et al.</i> , 2009
1341	respiration unité S	lignine	Sun <i>et al.</i> , 2000
1325	déformation unité S	lignine	Joseleau et Ruel, 1997
1320		cellulose	Mohebbi, 2008
1270	déformation unité G	lignine	Joseleau et Ruel, 1997
1250	déformation acide hydroxycinnamique C-O	ester	Vermerris <i>et al.</i> , 2002
1230	déformation C-O	ester	Mohebbi, 2008
1170	élongation ester C-O	ester	Sun <i>et al.</i> , 2000
1163	C-O-C	cellulose	Yan <i>et al.</i> , 2008
1111	déformation C ₂ -O ₂	cellulose	Yan <i>et al.</i> , 2008
1080	C-O C-C	xyloglucan arabinogalactan	Kacurakova <i>et al.</i> , 2002 Marga <i>et al.</i> , 2003
1065		xylane	Kacurakova <i>et al.</i> , 1999
1061	élongation C ₃ -O ₃	cellulose	Yan <i>et al.</i> , 2008
1045		xylane	Kacurakova <i>et al.</i> , 1999
1042	C-O C-C	xylane	Xiao <i>et al.</i> , 2001
1035	C ₆ -O ₆	cellulose	Yan <i>et al.</i> , 2008
1030	beta(1-3)	xylane	Kacurakova <i>et al.</i> , 1999
944		xyloglucan	Kacurakova <i>et al.</i> , 2002
870	déformation unité S vibration aromatique C-H	lignine	Joseleau et Ruel, 1997
840	déformation unité S vibration aromatique C-H	lignine	Joseleau et Ruel, 1997
837	hydroxyphenylpropane	lignine	Sun <i>et al.</i> , 2000
810	vibration aromatique C-H	lignine	Joseleau et Ruel, 1997

1.4 La spectroscopie Raman

1.4.1 Principe

Le principe de la spectroscopie Raman consiste à exciter de la matière avec une source de lumière monochromatique et analyser la lumière diffusée. Deux types de diffusion peuvent avoir lieu. La première est majoritaire et elle correspond à la diffusion élastique appelée également la diffusion Rayleigh. Dans ce cas, le photon est diffusé sans changement d'énergie. La deuxième est minoritaire et elle correspond à la diffusion inélastique. Dans ce cas le photon est diffusé avec une légère perte d'énergie correspondant à une transition vibrationnelle (diffusion Raman Stokes ou anti-Stokes) (Fig. A15). Comme il s'agit d'un processus intrinsèquement très faible, des sources de lumière intense telles que des lasers sont nécessaires.

1.4.2 Spectroscopie Raman des composants chimiques présents dans les parois cellulaires de poacées

Des spectres Raman de différents constituants des parois des tiges de poacées ont été étudiés en littérature afin d'obtenir une signature spectrale de chacun des polymères selon leurs caractéristiques. Les principaux composés lignocellulosiques étudiés en microspectroscopie Raman sont :

- les acides féruliques
- les acides para-coumariques
- les lignines
- la cellulose
- les hémicelluloses telles que les arabinoxylanes

Le tableau A.4 résume les nombre d'ondes caractéristiques de quelques composants chimiques présents dans les parois cellulaires de poacées

Le spectre Raman des composants chimiques des parois cellulaires de poacées est réparti en quatre régions spectrales principales :

- la première région 1800-1500 cm^{-1} révèle la lignine et les composés phénoliques.
- la deuxième région 1460-1291 cm^{-1} révèle les polysaccharides tels que la cellulose et les hémicelluloses.
- la troisième région 1270-1170 cm^{-1} révèle la lignine.
- la quatrième région 1170-800 cm^{-1} révèle les polysaccharides tels que la cellulose et les hémicelluloses.

Tableau A4 Nombre d'ondes Raman caractéristiques des composants chimiques des parois cellulaires de poacées

Longueurs d'ondes caractéristiques (cm ⁻¹)	Vibration	Composants chimiques	Référence bibliographique
1599-1658	élongation de la double liaison du cycle phényl conjuguée à l'élongation de la liaison C=O	lignine	Himmelsbach et Akin, 1998
1641-1612	élongation O-H	acide p-coumarique	Himmelsbach et Akin, 1998
1599,1630	vibration des doubles liaisons conjuguées du cycle phényl	lignine	Himmelsbach et Akin, 1998
1600-1631		lignine	Himmelsbach et Akin, 1998
1507	élongation de la liaison C=O	lignine G	Pilo-Veloso <i>et al.</i> , 1993
1465	déformation C-O-H	cellulose	Himmelsbach et Akin, 1998
1462	déformation C-H déformation O-H	arabinoxylane	Barron <i>et al.</i> , 2006
1453	déformation C-H ₃ déformation C-H	acide férulique lignine G	Himmelsbach et Akin, 1998
1408	déformation C-H	cellulose	Himmelsbach et Akin, 1998
1378	déformation C-H	cellulose	Himmelsbach et Akin, 1998
1337	déformation C-H	cellulose	Himmelsbach et Akin, 1998
1332	élongation de la liaison C=O	lignine S+G	Pilo-Veloso <i>et al.</i> , 1993
1291	déformation C-H	cellulose	Himmelsbach et Akin, 1998
1272	élongation de la double liaison du cycle phényl conjuguée à l'élongation de la liaison C=O	lignine	Pilo-Veloso <i>et al.</i> , 1993
1287	déformation CH ₃ CH	acide férulique	Himmelsbach et Akin, 1998
1267	déformation CH déformation O-H	arabinoxylane	Barron <i>et al.</i> , 2006
1265	déformation CH ₃ CH ₂ CH	acide férulique	Himmelsbach et Akin, 1998
1219	déformation CH ₃ CH ₂ CH	acide férulique	Himmelsbach et Akin, 1998
1199	élongation C-O et C=O	lignine S+G	Ram <i>et al.</i> , 2003
1176	déformation C-H ₃	acide férulique	Himmelsbach et Akin, 1998
1170	élongation C-O et C=O	lignine S+G	Ram <i>et al.</i> , 2003
1152	déformation de CH, C-C	cellulose arabinoxylane	Barron <i>et al.</i> , 2006
1131	élongation de C=C élongation de C-O-C élongation de C-C	cellulose arabinoxylane	Barron <i>et al.</i> , 2006
1093	étirement de C-C	cellulose arabinoxylane	Himmelsbach et Akin, 1998 Philippe <i>et al.</i> , 2006-a
997	élongation de C-C	cellulose	Himmelsbach et Akin, 1998
970	déformation de CH ₂	cellulose	Himmelsbach et Akin, 1998
896	élongation de C-O-C	arabinoxylane	Himmelsbach et Akin, 1998

1.5 La spectroscopie de fluorescence

1.5.1 Principe

La fluorescence UV se produit quand une molécule éclairée par des radiations de courtes longueur d'onde, l'absorbe et ré-émet des photons de lumière de longueur d'onde plus élevée pour revenir à son état stable. Un rayonnement incident de courte longueur d'onde excite un électron du nuage électronique qui entoure le noyau de l'atome ; l'énergie absorbée par l'électron le propulse dans les couches périphériques du nuage d'où il retombe par sauts quantiques, émettant à chaque fois un rayonnement de fluorescence de plus grande longueur d'onde. La fluorescence est dite primaire, appelée aussi autofluorescence, quand les produits fluorescents naturellement. L'autofluorescence est favorisée par la présence d'électrons π (électrons présents dans les doubles liaisons), qui sont facilement excités. Les composés autofluorescents comportent en général des noyaux aromatiques ou pyrroliques. Toute molécule fluorescente présente deux types de spectres : un spectre d'excitation et un spectre d'émission. Le spectre d'excitation indique les longueurs d'onde pour lesquelles l'intensité de lumière émise à une longueur d'onde donnée est maximale. Celles-ci sont sélectionnées pour exciter les molécules d'intérêt et enregistrer les spectres d'émission. Le spectre d'émission représente l'intensité de fluorescence émise en fonction de la longueur d'onde d'émission pour une longueur d'onde d'excitation donnée.

1.5.2 Spectroscopie de fluorescence des parois cellulaires

Les composés phénoliques (lignines et acides hydroxycinnamiques) sont les principaux composants chimiques de poacées étudiés en spectroscopie de fluorescence UV. Pour caractériser les lignines des travaux ont été effectués sur des molécules modèles [Lundquist *et al.*, 1978] ou principalement sur des lignines de bois [Olmstead et Gray, 1993, 1997]. Pour une excitation à 280 nm et à 340 nm des molécules modèles de lignines, les maxima des spectres d'émission ont été observés dans la région 414 nm et 465 nm [Lundquist *et al.*, 1978]. Les études sur le bois ont montré que pour une excitation à 360 nm, les maxima des spectres d'émission ont été obtenus dans la région 420-460 nm [Olmstead et Gray, 1993, 1997].

Une des caractéristique de la spectroscopie de fluorescence UV est qu'elle permet de différencier l'acide férulique de l'acide para-coumarique à 280 nm [Saadi *et al.*, 1998 ; Andersson et Akin, 2008]. Nous pouvons donc conclure que la longueur d'excitation à 280 nm est la plus adaptée pour révéler à la fois les lignines et les acides hydroxycinnamiques et pour distinguer l'acide férulique de l'acide para-coumarique.

2 Complémentarité entre les méthodes de spectroscopie

Des liaisons symétriques telles que C=C, C-C ou S-S répondent fortement en Raman. Les liaisons antisymétriques telles que les liaisons O-H, C=O, C-O ou N-H répondent en infrarouge. En pratique, les polysaccharides et les lignines sont sensibles aux rayonnements infrarouges et Raman. La spectroscopie Raman révèle les acides hydroxycinnamiques [Himmelsbach et Akin, 1998]. La présence d'un cycle phényl- des lignines et des acides phénoliques en liaison avec des polysaccharides est aussi détectable en fluorescence UV [Andersson et Akin, 2008; Saadi *et al.*, 1998]. Les deux acides phénoliques, à savoir l'acide férulique et l'acide para-coumarique, peuvent être distingués en fluorescence ultraviolet. Ces deux composants chimiques sont très difficilement discriminés en utilisant la spectroscopie Raman ou infrarouge. En conclusion, ces spectroscopies sont complémentaires du point de vue de l'identification des molécules : les composés phénoliques sont autofluorescents et ont une signature caractéristique en Raman, l'infrarouge est particulièrement adapté pour les polysaccharides. Le Raman est complémentaire de l'infrarouge permettant d'observer les liaisons non polaires des polymères.

3 Microspectroscopies

3.1 Caractéristiques des microspectrophotomètres de laboratoire

Un microspectrophotomètre est un système d'analyse spectrale (spectromètre) couplé à un microscope optique. Le microscope optique permet de focaliser la lumière à un endroit donné de l'échantillon. Le spectrophotomètre permet d'acquérir des spectres en "mode" point par point. Il permet également d'acquérir des images.

La résolution spatiale est dépendante entre autres de la source utilisée. Les microspectrophotomètres infrarouge de laboratoire qui utilise une source thermique ne permettent pas d'acquérir des images à haute résolution. Les radiations synchrotron ont des caractéristiques exceptionnelles par rapport aux sources des microspectrophotomètres de laboratoire. L'utilisation d'une source synchrotron permet d'acquérir des images à haute résolution spatiale. Dans le cas de la microspectroscopie Raman et de la microscopie de fluorescence, la résolution latérale est nettement meilleure (1 μm , environ 0,2 μm , respectivement) comparativement à la microspectroscopie infrarouge et le développement du dispositif confocal permet une meilleure résolution en profondeur. La microscopie confocale à fluorescence en mode spectral permet de discriminer plus aisément les molécules sur la base

de leur signature spectrale lorsque leurs spectres se chevauchent partiellement. Cependant, les plages des longueurs d'onde excitation-émission proposées pour les microscopes de laboratoire restent limitées. En particulier le domaine des UV entre 190-350 nm n'est pas couvert. La source synchrotron de fluorescence a l'avantage de couvrir la longueur d'onde 280 nm en excitation.

3.2 Source synchrotron

Dans un synchrotron (Fig. A18), les électrons sont générés dans un canon à électrons. Ils sont ensuite accélérés par un accélérateur linéaire de particules jusqu'à des vitesses proche de la vitesse de la lumière et insérés dans le booster. Lorsque les électrons atteignent leur énergie nominale (2,75 GeV au synchrotron SOLEIL), ils sont dirigés dans l'anneau de stockage. L'anneau de stockage est un polygone avec des aimants de courbure. Ces aimants de courbure (appelé également aimants dipolaires) sont utilisés pour diriger les électrons dans l'anneau de stockage. Quand les électrons passent à travers les aimants, ils perdent de l'énergie sous forme de lumière. La lumière émise par les électrons déviés s'étend des rayons X jusqu'à l'infrarouge.

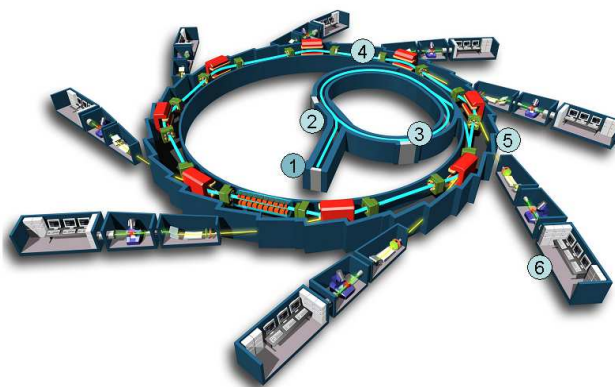


Figure A18 Schéma d'un Synchrotron : 1)-le canon d'électron, 2)- l'accélérateur linéaire, 3)- le booster 4)- l'anneau de stockage, 5)- la ligne de lumière, 6)- la station expérimentale (laboratoire).

Le rayonnement synchrotron est collecté sur des lignes de lumières qui permettent d'étudier de nombreux échantillons dans différents domaines tel que la géologie [Guilhaumou *et al.*, 1998] et la biologie [Miller *et al.*, 1998; Dumas et Miller, 2003] à très hautes résolutions spatiales et spectrales.

Dans les sections suivantes, les lignes de lumière du synchrotron SOLEIL (Source Optimisée de Lumière à Energie Intermédiaire du LURE (Laboratoire d'Utilisation du Rayonnement Electromagnétique)), SMIS (Spectroscopie et microscopie dans l'infrarouge en utilisant le

synchrotron) et DISCO (Dichroïsme, Imagerie et Spectrométrie d'échantillons Chimiques et biologiques) dédiées respectivement aux microspectroscopies infrarouge et de fluorescence sont décrites.

3.2.1 Ligne de lumière SMIS

Deux types de microspectrophotomètres sont disponibles sur la ligne de lumière SMIS :

- 3 microspectrophotomètres infrarouge utilisant la source synchrotron et un microspectrophotomètre utilisant une source interne.
- un microspectrophotomètre Raman utilisant une source laser. Trois lasers sont disponibles sur la ligne SMIS : 532 nm, 633nm, et 785nm.

3.2.1.1 Microspectroscopie infrarouge

Le spectromètre infrarouge à transformé de Fourier (FT-IR) est un Nicolet 5700 couplé à un microscope Continuum XL (Thermo Fisher Scientific, WI, USA) qui permet l'acquisition des spectres infrarouge et des images visibles en fond clair (Fig. A19). Le détecteur du microscope infrarouge est un détecteur (MCT-A) mercure cadmium tellurure, (avec une taille de pixel de 50 μm) refroidi à l'azote liquide. Le microscope fonctionne en mode confocal, à l'aide d'un objectif Schwarzschild (d'ouverture numérique $NA = 0,65$) d'un grossissement 32X et un condenseur d'appariement équivalent [Dumas *et al.*, 2006].



Figure A19 microspectrophotomètre infrarouge de la ligne SMIS

3.2.1.2 Microspectroscopie Raman

Le microspectrophotomètre Raman de la ligne SMIS est un microscope confocal Raman DXR (Thermo Fisher Scientific, WI, USA) (Fig. A20). Il est équipé d'une source laser qui sera focalisée sur un échantillon grâce à un microscope à différents objectifs. Afin d'éviter dans certains cas une dégradation thermique de l'échantillon, le temps d'acquisition ainsi que la

puissance du laser sont réglés de manière à réduire l'intensité du rayonnement incident. Le signal émis est dirigé vers le spectrophotomètre en passant au travers d'un trou confocal ou fente. Puis le signal est focalisé sur la fente d'entrée du spectrophotomètre. Le spectre Raman est ensuite analysé par un détecteur multicanal CC (Coupled Charge Detector) qui permet de mesurer simultanément l'intensité des différentes longueurs d'ondes du spectre. Une platine motorisée est utilisée pour déplacer l'échantillon.



Figure A20 microspectrophotomètre Raman de la ligne SMIS

3.2.2 Ligne de lumière DISCO

La ligne de lumière DISCO est dédiée à la biochimie, la chimie et la biologie cellulaire. Le microscope est construit autour d'un microscope (Olympus IX71) inversé (Fig. A21). Les lentilles intermédiaires en verre ont été remplacées par des lentilles en quartz car elles sont transparentes dans l'UV. L'objectif utilisé est un ultrafluar de grossissement 40X (Zeiss, Allemagne) avec une ouverture numérique de 0,6 et une correction de 0,17 pour la lamelle de quartz. L'excitation de l'échantillon est fournie par le rayonnement synchrotron issue d'un aimant de courbure [Giuliani *et al.*, 2009]. Le faisceau "blanc" de 180 à 600 nm est monochromatisé par un monochromateur, iHR320 (Jobin-Yvon, Fr) équipé d'un réseau de 100 traits par millimètre avec une efficacité minimale à 250 nm (Scientific Spectrum, Inc, Irvine, CA, USA) avec des fentes d'entrée et de sorties fixées à 0,1mm, correspondant à une bande passante de 0,137 nm [Jamme *et al.*, 2010].



Figure A21 Composantes du microspectrophotomètre de la ligne DISCO

3.3 Microspectroscopie des végétaux

3.3.1 Microspectroscopie infrarouge des polymères des parois cellulaires

La microspectroscopie infrarouge a été utilisée pour distinguer des différences de composition de la paroi en fonction des types cellulaires [Stewart, 1996 ; McCann *et al.*, 1997, 2001] et en fonction du stade de développement de la plante [Toole *et al.*, 2007, 2010].

Dans le cadre d'études sur la mise en place des parois végétales dans l'albumen du blé, la microspectroscopie infrarouge s'est avérée performante pour déterminer les compositions relatives en hémicelluloses en fonction des tissus et des stades de développement [Philippe *et al.*, 2006-b, Toole *et al.*, 2007, 2010, Saulnier *et al.*, 2009, Robert *et al.*, 2011] ou en fonction des cultivars [Toole *et al.*, 2011]. Les variations du taux de substitution des arabinoxylyanes ont pu par ailleurs, être mis en évidence en fonction des tissus et des stades de développement. Les cellules centrales de l'albumen ou proches de l'embryon sont plus riches en $\beta(1-3)(1-4)$ glucanes que les cellules du reste de l'albumen. Les arabinoxylyanes dans les cellules de l'albumen amylicé apparaissent plus substitués aux stades jeunes au moment de la différenciation cellulaire qu'aux stades avancés. Par microspectroscopie FT-IR, 50 cultivars de blé pour la panification ont pu être classés en fonction du degré de substitution des arabinoxylyanes dans l'albumen amylicé. Les lignines ont également été étudiées par microspectroscopie FT-IR [Popescu *et al.*, 2007]. Popescu *et al.* (2007) ont démontré que l'infrarouge permettait de discriminer les lignines de différentes sources sur la base de leurs proportions d'unités guaiacyles et syringyles. La microspectroscopie infrarouge a été également utilisée pour classifier différents mutants d'*Arabidopsis* affectés dans la

composition des parois cellulaires [Mouille *et al.*, 2003]. Des mutants d'arabidopsis montrant une altération de la teneur en cellulose ont pu être identifiés.

Le rayonnement synchrotron a été notamment utilisé pour caractériser les parois des cellules à aleurone et des cellules modifiées à aleurone [Jamme *et al.*, 2008]. Dans le cas présent, l'utilisation d'un rayonnement synchrotron a permis d'explorer des variations à l'échelle cellulaire.

3.3.2 Microspectroscopie Raman des polymères des parois cellulaires

La microspectroscopie Raman a été utilisée pour distinguer les variations de la composition et la structure de la paroi au cours du développement du lin [Himmelsbach et Akin, 1998]. La microspectroscopie Raman s'est avérée performante pour la caractérisation des parois cellulaires du blé [Philippe *et al.*, 2006-a ; Robert *et al.*, 2011]. Des études ont été menées en microspectroscopie Raman afin d'évaluer la qualité du blé en terme de dureté et de déterminer la teneur en protéines dans l'albumen amylopectine [Piot *et al.*, 2001 ; Barron *et al.*, 2005].

La microspectroscopie Raman est également utilisée pour l'étude des parois cellulaires du bois [Gierlinger et Schwanninger, 2006].

Les informations obtenues en microspectroscopie infrarouge peuvent être complétées par la microspectroscopie Raman dans la gamme spectrale (1800-800) cm^{-1} pour caractériser les parois cellulaires de poacées [Cael *et al.*, 1973 ; Himmelsbach et Akin, 1998 ; Jarvis et McCann, 2000].

3.3.3 Microspectroscopie de fluorescence des polymères des parois cellulaires

La gamme spectrale UV-visible a été beaucoup utilisée pour étudier les lignines et différencier l'acide férulique de l'acide para-coumarique [Koch et Kleist, 2001 ; Koch, 2004 ; Lybeer et Koch, 2005 ; Kim *et al.*, 2008 ; Rehbein *et al.*, 2010].

La technique principale qui a suscité de l'intérêt pour cartographier la distribution des composés phénoliques dans les parois cellulaires est la microscopie de fluorescence UV [Rudall et Caddick, 1994 ; Harris et Hartley, 1976]. Peu d'études existent en microspectroscopie de fluorescence UV. Une étude a été réalisée sur les grains de blé par Saadi *et al.* (1998) en imagerie spectrale de fluorescence [Saadi *et al.*, 1998]. Le but de cette étude était de cartographier la distribution des composés phénoliques dans la couche à aleurone sur des coupes de grains de blé. Ces auteurs ont montré qu'il était possible de différencier l'acide férulique de l'acide para-coumarique par microspectroscopie UV. Siqueira

et al. (2011) ont utilisé la microspectroscopie de fluorescence pour caractériser la composition chimique des parois cellulaires de la canne à sucre avant et après dégradation enzymatique.

4 Pré-traitement

4.1 Pourquoi le pré-traitement?

Dans de nombreuses situations, les spectres présentent des variations incontrôlées d'intensité globale. Les pré-traitements sont donc nécessaires pour améliorer le signal et indispensables pour une meilleure analyse des données. Ils permettent de réduire l'effet du bruit ou d'une déformation du spectre.

4.2 Etapes de pré-traitement

Le pré-traitement des spectres peut inclure le filtrage du bruit, la correction de la ligne de base et la normalisation de spectres.

4.2.1 Lissage des spectres

Les techniques de filtrage spectral permettent de lisser les spectres en éliminant le bruit. La forme la plus simple de lissage est la "moyenne mobile". La moyenne mobile consiste à remplacer chaque valeur du spectre par la moyenne des valeurs de ses voisins. Les filtres les plus populaires utilisés en spectroscopie sont les filtres binomiaux [Marchand et Marmet, 1983] et les filtres dérivatifs [Savitzky et Golay, 1964].

Le lissage binomial consiste à convoluer le spectre considéré par un filtre Gaussien. La méthode de Savitzky et Golay (1964) permet de définir des poids en se fondant sur la méthode des moindres carrées [Bertrand et Dufour, 2000 ; Lasch, 2012]. L'idée générale de leur méthode et de considérer en plus du nombre des voisins, un polynôme sur lequel le segment du spectre est ajusté au sens des moindres carrées. La méthode de lissage Savitzky-Golay dépend alors de deux paramètres : la taille de la fenêtre utilisée (nombre de voisins) exprimée en nombre de points et le degré du polynôme utilisé pour l'ajustement. Dans la pratique, les auteurs donnent les poids correspondants à des fenêtres de taille de 5 à 25 points avec des polynômes d'ordre de 1 à 5 [Bertrand et Dufour, 2000].

4.2.2 Correction de la ligne de base

La plupart des méthodes de correction de ligne de base font l'hypothèse implicite que le spectre observé est la somme d'un signal pertinent contenant l'information spectrale et d'un

signal de fond correspondant aux variations incontrôlées. La correction consiste à modéliser le fond et à le soustraire du signal. La méthode la plus simple pour corriger la ligne de base consiste à estimer le fond par un ensemble de segments affines passant par les minima locaux des spectres. Le spectre corrigé est simplement la soustraction des valeurs des segments affines du spectre d'origine. Cette méthode repose sur le même principe que la méthode 'rubber band' [Liu et Koenig, 1987]. Dans le cas de la méthode 'rubber band', les spectres sont ajustés en utilisant des fonctions polynomiales.

La méthode de Savitzky et Golay (1964) est utilisée également pour corriger la ligne de base en appliquant une dérivée première ou une dérivée seconde aux spectres. La dérivée seconde a eu plus de succès que la dérivée première parce qu'elle fournit des spectres plus facile à interpréter.

4.2.3 Normalisation

La procédure de la normalisation consiste à ramener les spectres à la même intensité (même échelle) afin de pouvoir les comparer aussi bien sur le plan quantitatif que sur le plan qualitatif. Une correction simple consiste à diviser chaque spectre s par un nombre F_n supposé représentatif de l'intensité générale de ce spectre

$$sn(\lambda) = s(\lambda) / F_n \quad \text{Equation A5}$$

sn représente le spectre normalisé

$s(\lambda)$ est la variable d'indice λ du spectre s

La valeur du facteur F_n peut être calculée suivant différentes manières [Bertrand et Dufour, 2000].

Une des manières les plus simples est de la calculer comme la norme du spectre s , en d'autres termes normer tous les spectres à 1.

$$F_n = \sqrt{\sum_{s=1}^S s(\lambda)^2} \quad \text{Equation A6}$$

où S est le nombre des longueurs d'ondes du spectre s .

Cette méthode est appelée normalisation vectorielle unitaire (unit vector normalisation).

Une autre possibilité est de considérer le coefficient F_n comme la somme des valeurs du spectre s .

$$Fn = \sum_{s=1}^S s(\lambda)$$

Equation A7

Des méthodes de correction plus complexes nécessitent la présence d'une collection de spectres (matrice de spectres) ou d'un spectre de référence. Parmi ces méthodes, on peut citer la correction multiplicative de diffusion (MSC) [Geladi *et al.*, 1985] et l'extension de la correction multiplicative de diffusion (EMSC) [Martens *et al.*, 2003].

4.3 Combinaison de méthodes de pré-traitement

Dans la pratique, la correction des spectres requiert plusieurs étapes de pré-traitement. En effet, dans la majorité des études, le pré-traitement consiste en une combinaison de deux ou plusieurs étapes de pré-traitement exécutées séquentiellement. Une des combinaisons fréquentes pour pré-traiter des spectres infrarouge commence par une dérivée seconde tel que la dérivation par Savitzky Golay suivi par une normalisation vectorielle unitaire [Lasch *et al.*, 2003]

Il y a peu d'études disponibles où l'efficacité des différents modes de pré-traitement a été étudiée [Delwiche et Graybosch, 2003 ; Afseth *et al.*, 2006 ; Heraud *et al.*, 2006]. La conception et la compilation d'un système efficace de pré-traitement et l'optimisation de ses paramètres sont souvent basées sur l'expérience et l'intuition des spectroscopistes. En réalité, l'obtention d'une relation mathématique directe entre les spectres bruts et le spectre idéal est une tâche irréalisable pour des échantillons complexes et hétérogènes. Le choix d'une méthode de pré-traitement reste aussi une tâche assez compliquée.

Conclusion

La microspectroscopie est une méthode puissante pour analyser *in situ* la variabilité de composition chimique des parois en fonction des types cellulaires. Les gammes du moyen infrarouge et du Raman sont adaptées à l'analyse des polysaccharides et des lignines avec des sensibilités différentes. Les principaux acides hydroxycinnamiques sont potentiellement identifiables par leur signature en fluorescence UV. La complémentarité des méthodes spectrales permet de fournir une analyse complète de la variabilité des parois cellulaires. Dans le chapitre suivant, nous présenterons un éventail de méthodes existantes pour analyser les données spectrales en chimiométrie en commençant par l'analyse de données issues d'une seule technique spectrale jusqu'à l'analyse conjointe de plusieurs données spectrales.

III Chimiométrie et analyse des données spectrales

1 Spectres, tableaux de spectres et images hyperspectrales

1.1 Spectres

Un spectre se présente comme une suite de nombre. Chacun de ces nombres représente la valeur de la mesure spectrale à une longueur d'onde donnée. La séquence des nombres associés à un spectre est un vecteur que nous désignerons par une minuscule en gras \mathbf{x} .

1.2 Tableaux de spectres

1.2.1 Tableau à deux entrées

Une étude chimométrique de données spectrales est basée sur l'acquisition de spectres pour une collection d'échantillons. Ces spectres forment une matrice, qu'on appelle également tableau à deux entrées ou à deux modes, notée par une majuscule en gras \mathbf{X} (Fig. A22). La première dimension N du tableau correspond aux spectres ou aux échantillons. La deuxième dimension S correspond aux variables spectrales. Les variables spectrales représentent les longueurs d'ondes dans le cas de la spectroscopie de fluorescence et les nombres d'ondes dans le cas des spectroscopies infrarouge et Raman. Dans la suite, on parlera de variables spectrales pour se référer aux longueurs d'ondes et aux nombres d'ondes.

Les tableaux de données spectrales ont les propriétés suivantes :

- Le tableau de spectres \mathbf{X} comprend généralement beaucoup plus de variables que d'individus c'est à dire beaucoup plus de variables spectrales que de spectres.

- Les variables spectrales sont très corrélées entre elles [Bertrand et Dufour, 2000].

Pour analyser les tableaux de données spectrales en chimiométrie, on utilise l'analyse des données multivariées. L'objectif des méthodes multivariées est de calculer des variables synthétiques de tout l'ensemble de données. Une des méthodes les plus utilisées dans le cadre de l'analyse exploratoire de données est l'analyse en composantes principales [Bertrand et Dufour, 2000 ; Coimbra *et al.*, 1999 ; Dufour *et al.*, 2001].

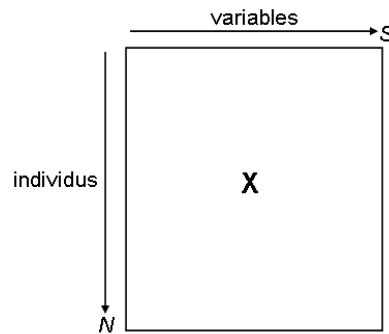


Figure A22 Tableau de données de spectres, N individus et S variables spectrales

1.2.2 Tableaux à trois entrées

Dans certaines problématiques chimiométriques, les mesures spectrales s'effectuent en fonction de deux variables, par exemple, spectrales et temporelles. Le tableau de données obtenu est un tableau à trois entrées $N \times S \times K$ [Smilde *et al.*, 2004 ; Holbrook *et al.*, 2006 ; Rutledge et Jouan-Rimbaud Bouveresse, 2007 ; Lakhal *et al.*, 2012] (Fig. A23). Les N individus et également les S variables sont supposés les mêmes pour l'ensemble des tableaux à deux entrées. Ces tableaux sont dits appariés selon le mode des individus et des variables [Smilde *et al.*, 2004].

Ces tableaux à trois entrées se présentent sous la forme d'un ensemble de K tableaux de données à deux entrées noté $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$. Chaque tableau \mathbf{X}_k contient les mesures de S variables sur N individus. L'un des modèles les plus connus et aussi le plus utilisé pour l'analyse des tableaux à trois entrées est le modèle PARAFAC [Smilde *et al.*, 2004].

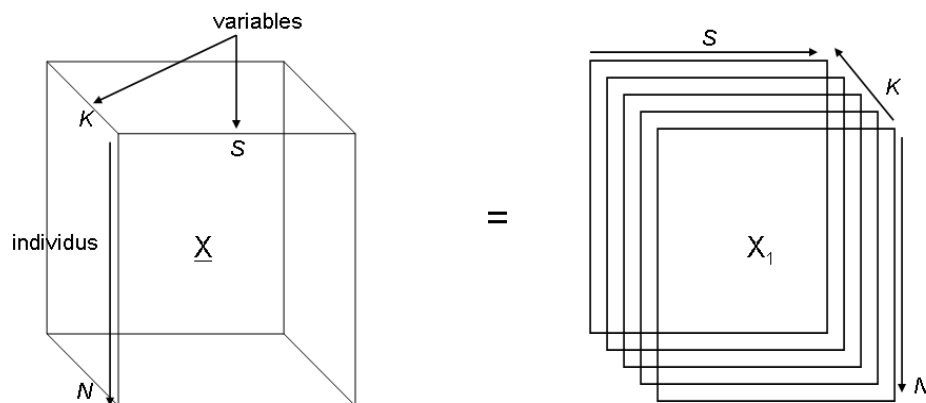


Figure A 23 Tableau de données à trois entrées. Ce tableau contient N individus selon le premier mode, S variables selon le deuxième et K variables selon le troisième mode, respectivement. Un tableau à trois entrées peut être considéré comme K tableaux à deux entrées.

1.2.3 Ensemble de tableaux de données

Parfois, la caractérisation des échantillons complexes nécessite l'utilisation de plusieurs techniques pour pouvoir cerner au mieux les propriétés du produit à analyser. Par exemple, des mesures spectrales sont effectuées sur N échantillons en utilisant des spectroscopies différentes. L'analyse des données mène à considérer ensemble différents tableaux $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$ (Fig. A24). La structure des données résultante consiste en K tableaux ayant chacun N individus et un nombre de variables différent qui dépend de la technique utilisée

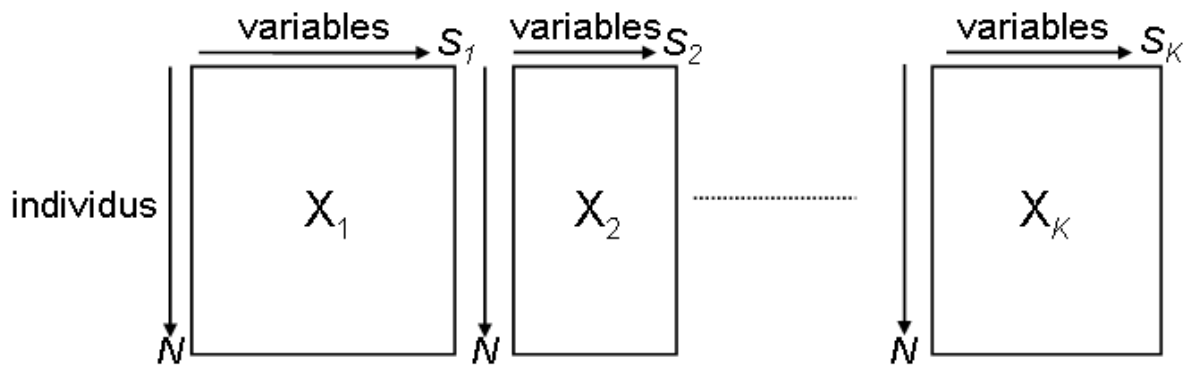


Figure A24 Couplage de K tableaux de données appariés selon le premier mode des individus N . Les variables S_1, S_2 et S_K sont différentes.

Dans ce cas, il est nécessaire de disposer de méthodes d'analyse adaptées au couplage d'un ensemble de K tableaux [Mazerolles *et al.*, 2002; Hanafi *et al.*, 2006 ; Böcker *et al.*, 2007].

1.3 Image hyperspectrale

1.3.1 Structure de données

Une image monochrome peut être représentée par une matrice à deux dimensions (N_b, N_c) où N_l est le nombre de lignes et N_c est le nombre de colonnes.

Une image hyperspectrale est une image pour laquelle, à chaque pixel, un spectre complet est acquis (Fig. A25). Elle peut être considérée comme un tableau de données à trois entrées avec deux entrées correspondant aux dimensions spatiales N_b, N_c et une entrée spectrale S . Les dimensions spatiales correspondent à l'espace scène de l'image. La troisième entrée est relative à la dimension spectrale habituellement considérée comme des variables dans une analyse multivariées de données spectrales.

Une image hyperspectrale peut être considérée comme un empilement d'images observées dans des conditions spectrales différentes ou une collection de spectres. Une représentation

possible de ces données peut être faite sous la forme d'un cube hyperspectral que nous désignerons par une majuscule en gras soulignée **X**.

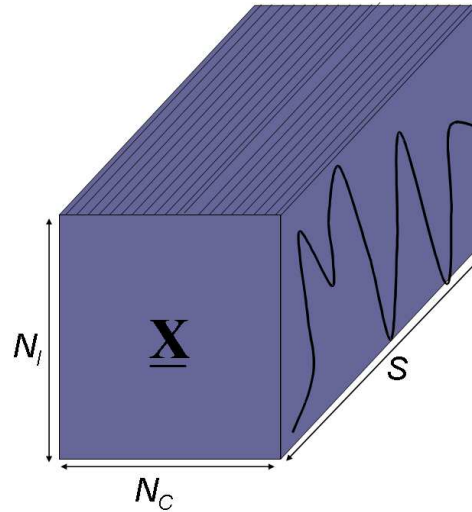


Figure A25 Structure d'une image hyperspectrale

Il est à noter qu'une image hyperspectrale diffère d'un tableau à trois entrées par la nature des entrées : 2 modes individus (N_b , N_c) et un mode variable S .

1.3.2 Représentation graphique d'images hyperspectrales

La représentation des images hyperspectrales n'est pas évidente. En considérant l'image hyperspectrale comme un ensemble de spectres, elle peut être représentée par le graphe de la collection spectrale. Cela revient à ignorer les dimensions spatiales (N_b , N_c) et de les tracer comme un ensemble de $N_l \times N_c$ spectres. Cette approche, ne permet de révéler que la variabilité spectrale au sein de la collection des spectres. Afin de révéler la variabilité spatiale, différentes approches permettant de calculer une image à deux dimensions sont possibles. Par exemple, on peut représenter les variations spatiales en sélectionnant une longueur d'onde caractéristique (Fig. A26). Il est possible également, de calculer l'aire sous un pic ou le rapport de l'aire sous deux pics pour établir une image de la variation chimique de la scène observée. Les logiciels d'acquisition d'images spectrales contiennent généralement des outils de visualisation qui permettent d'explorer les images de manière rapide. L'affichage peut être réalisé après lissage, correction de ligne de base, dérivée, normalisation ou après analyse multivariée de spectres afin de mettre en évidence les différences pertinentes entre les spectres en fonction de leur localisation spatiale.

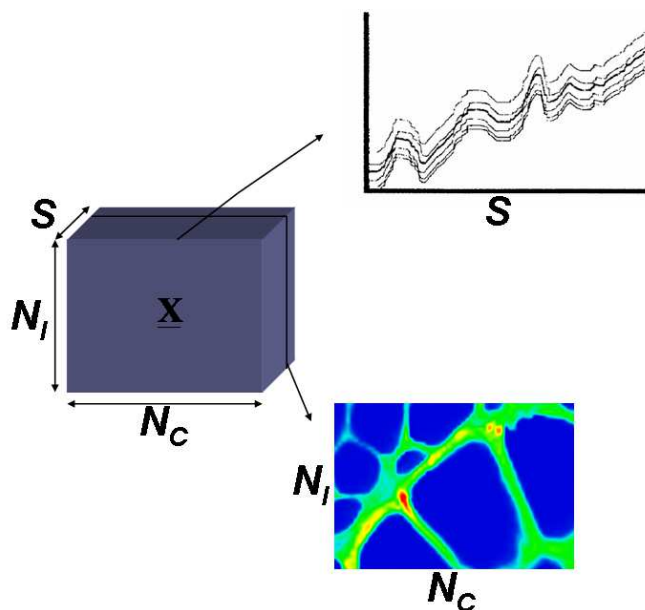


Figure A26 Représentation des données issues d'une image hyperspectrale, (haut) collection de spectres, (bas) affichage de la variabilité spatiale.

2 Chimométrie

La chimométrie peut être définie comme l'ensemble des méthodes statistiques, graphiques ou symboliques permettant d'analyser l'information obtenue dans le domaine de la chimie analytique. La chimométrie s'applique à toutes les étapes de l'analyse, depuis la conception de l'expérience jusqu'à ce que les données soient complètement exploitées.

2.1 Notations

Dans tout ce qui va suivre, nous allons adopter les conventions de notations suivantes.

Les scalaires sont représentés par des lettres minuscules: n, m, l, \dots

Les dimensions des tableaux sont représentées par des lettres majuscules en italique N, M, L, \dots

Les vecteurs sont notés par des lettres minuscules en gras : $\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$

Les matrices appelées également tableaux à deux entrées sont représentées par des lettres majuscules en gras : $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \dots$

La transposée d'une matrice \mathbf{X} sera notée \mathbf{X}'

Les tableaux à trois entrées seront représentés par des lettres majuscules en gras soulignées : $\underline{\mathbf{X}}, \underline{\mathbf{Y}}, \underline{\mathbf{Z}}, \dots$

Il est aussi pratique de définir un tableau à trois entrées comme étant une série de tableaux à deux entrées. Un tableau à trois entrées $\underline{\mathbf{X}}$ ($N \times S \times K$) peut ainsi être noté comme suit : $\underline{\mathbf{X}} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K]$.

2.2 Généralités sur les méthodes

La structure de données la plus simple obtenue dans le cadre de l'analyse spectrale est un tableau à deux entrées. Le tableau est analysé dans le but de comparer le groupe des individus et de synthétiser l'information spectrale. La méthode de base pour réaliser cette tâche est l'analyse en composantes principales. Lorsqu'on dispose de K tableaux à deux entrées appariés selon la dimension des individus mais ayant un nombre de variables différents, l'analyse multibloc est utilisée pour décrire les variations entre les échantillons étudiés (globalement ou au niveau de chacun des blocs) et pour identifier les blocs de variables qui ont des typologies communes et ceux qui ont des typologies spécifiques.

Dans ce qui va suivre, nous allons présenter en détail l'analyse en composantes principales basée sur la diagonalisation, l'algorithme NIPALS, et l'analyse multibloc de K tableaux à deux entrées.

2.3 Analyse en composantes principales

Considérons le tableau \mathbf{X} de données centrées selon la dimension spectrale. L'analyse en composantes principales [Jolliffe, 1986 ; Saporta, 1990 ; Lebart *et al.*, 2000] permet de remplacer les variables d'origine, fortement redondantes par des variables synthétiques, les composantes principales \mathbf{C} non corrélées entre elles et appelées également composantes factorielles ou variables latentes. Les composantes principales $\mathbf{C} = \mathbf{X}\mathbf{U}$, sont des combinaisons linéaires des variables de départ.

Les composantes \mathbf{C} sont calculées de manière à représenter le plus de variance possible en optimisant le critère suivant :

$$\text{Max (Variance } (\mathbf{C}\mathbf{C}')) \text{ sous la contrainte } \mathbf{U}\mathbf{U}' = \mathbf{I} \quad \text{Equation A8}$$

où \mathbf{I} est la matrice identité de taille $S \times S$

En pratique, le calcul des composantes principales est basé sur la diagonalisation de la matrice de variance-covariance

$$\mathbf{V} = \mathbf{X}\mathbf{X}' \quad \text{Equation A9}$$

La diagonalisation de la matrice V revient à la décomposer en une matrice de vecteurs propres U et une matrice de valeurs propres L (Fig. A27). Les vecteurs propres sont les coefficients de la combinaison linéaire appelés en anglais 'loadings'. Ils sont utilisés pour calculer les composantes appelés en anglais 'scores' à partir du tableau original $C= XU$.

Les valeurs propres L représentent les valeurs à partir desquelles on peut calculer le pourcentage de variance décrite par les composantes. Les composantes sont rangées suivant un ordre décroissant. Les premières composantes révèlent l'information pertinente apportée par les données, par contre, les dernières composantes décrivent du bruit.

Dans le cas des données spectrales, les vecteurs propres sont souvent représentés sous la forme de courbes appelées «profils spectraux». Ces courbes présentent des valeurs positives et négatives associées aux différences observées entre les échantillons. Ce sont ces profils qui permettent l'interprétation des composantes en termes de spectroscopies [Robert *et al.*; 1996].

Dans la suite, on appellera les profils spectraux ou les vecteurs propres "les profils variables". Des cartes factorielles peuvent être tracées à partir des valeurs des composantes pour deux composantes principales choisies parmi les premières obtenues. Ces cartes fournissent un outil de comparaison des échantillons.

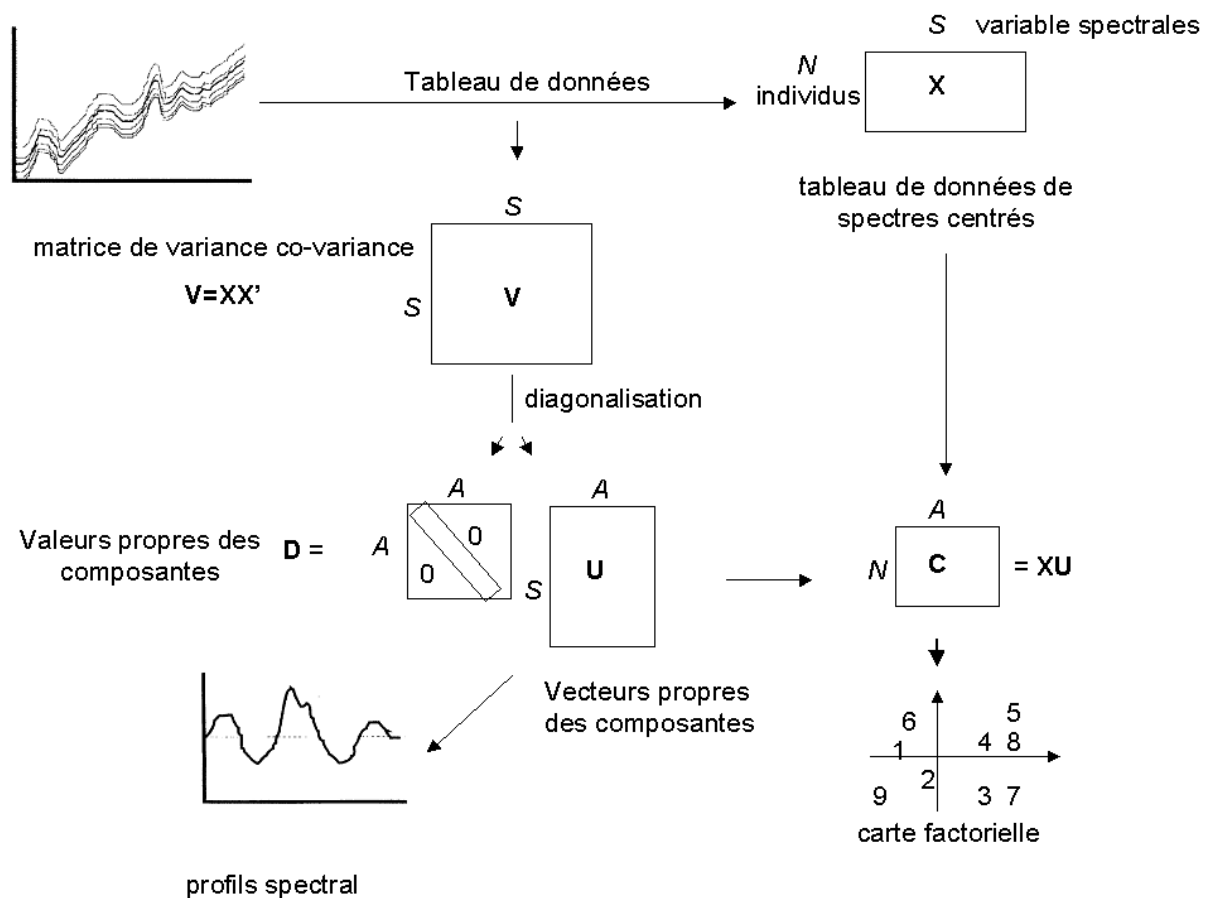


Figure A27 Les étapes principales de l'analyse en composantes principales

La reconstitution des données en analyse en composantes principales s'écrit de la manière suivante :

$$\mathbf{X}_r = \mathbf{C}\mathbf{U} \quad \text{Équation A10}$$

En se reposant sur la formule de reconstitution des données, le critère de l'analyse en composantes principales peut être également présenté comme la minimisation du résidu \mathbf{R} où \mathbf{R} est égal à :

$$\mathbf{R} = \mathbf{X} - \mathbf{X}_r \quad \text{Equation A11}$$

Le critère de minimisation s'écrit de la manière suivante :

$$R = \min \|\mathbf{X} - \mathbf{C}\mathbf{U}\|^2 \quad \text{Équation A12}$$

Algorithme NIPALS

L'algorithme NIPALS proposé par Wold, a pour objectif le calcul des profils variables \mathbf{U} et des composantes \mathbf{C} avec une écriture différente de celle de l'analyse en composantes principales. Cette méthode a été écrite pour minimiser le même critère que celui en analyse en composantes principales en considérant les composantes une par une:

$$R = \min \|\mathbf{X} - \mathbf{c}\mathbf{u}\|^2 \quad \text{Equation A13}$$

où

\mathbf{c} est la composante

\mathbf{u} est le profil variable

Les premières composantes de l'algorithme NIPALS sont obtenues en utilisant l'algorithme itératif suivant :

1. Sélectionner aléatoirement un vecteur colonne \mathbf{x}_i de la matrice \mathbf{X}

2. Projeter la matrice \mathbf{X} sur \mathbf{x}_i pour trouver le profil variable associé $\mathbf{u}^{(1)}$. $\mathbf{u}^{(1)} = \frac{\mathbf{X}' \mathbf{x}_i}{\mathbf{x}_i' \mathbf{x}_i}$

3. Normaliser le profil variable $\mathbf{u}^{(1)}$ à 1. $\mathbf{u}^{(1)} = \frac{\mathbf{u}^{(1)}}{\|\mathbf{u}^{(1)}\|}$

4. Projeter \mathbf{X} sur $\mathbf{u}^{(1)}$ pour calculer la composante $\mathbf{c}_x^{(1)}$, $\mathbf{c}_x^{(1)}$ représente la nouvelle valeur de \mathbf{x}_i . $\mathbf{c}_x^{(1)} = \mathbf{X}\mathbf{u}^{(1)}$

5. Répéter les étapes 2 à 5 jusqu'à convergence du critère $\|\mathbf{c}_x^{(1)} - \mathbf{x}_i\| < \varepsilon$.

$\mathbf{c}_X^{(1)}$ et $\mathbf{u}^{(1)}$ sont respectivement la composante et le profil variable recherchés pour la dimension 1.

Une fois la première dimension ($\mathbf{c}_X^{(1)}$, $\mathbf{u}^{(1)}$) calculée, la part de \mathbf{X} décrite par $\mathbf{c}_X^{(1)}$ et $\mathbf{u}^{(1)}$ obtenue par la formule de reconstitution $\tilde{\mathbf{X}} = \mathbf{c}_1 \mathbf{u}_1'$ est soustraite de la matrice \mathbf{X} . La procédure de soustraction de l'information obtenue est appelée **déflation** [Smilde *et al.*, 2003].

$$\mathbf{X}^{(2)} = \mathbf{X} - \tilde{\mathbf{X}} \quad \text{Équation A14}$$

Sur la matrice $\mathbf{X}^{(2)}$ appelée matrice des résidus, la procédure de recherche des profils variables et des composantes est appliquée. Les profils variables et les composantes sont ainsi calculés séquentiellement.

A l'étape $h > 2$, la déflation s'écrit de la manière suivante :

$$\mathbf{X}^{(h)} = \mathbf{X}^{(h-1)} - \tilde{\mathbf{X}}^{(h-1)} \quad \text{Equation A15}$$

Il a été montré que l'algorithme NIPALS permettait de retrouver les composantes et les profils variables de l'analyse en composantes principales, sans avoir à former explicitement la matrice de variance covariance. Un des intérêts de l'algorithme NIPALS est qu'il autorise le calcul des profils variables et des composantes dans le cas des données manquantes.

2.4 Analyse multibloc

2.4.1 Analyse de deux tableaux de données

Nous disposons de deux tableaux \mathbf{X}_1 , \mathbf{X}_2 constitués respectivement de S_1 et S_2 variables spectrales mesurées sur les mêmes N individus. Par la suite, ces tableaux \mathbf{X}_1 , \mathbf{X}_2 sont supposés centrés. La structure des données est illustrée par la figure A28. L'analyse multibloc de deux tableaux de données a pour but de décomposer les deux tableaux en un ensemble de composantes ($\mathbf{C}_{X1}, \mathbf{C}_{X2}$) et en un ensemble profils variables ($\mathbf{U}_1, \mathbf{U}_2$)

La décomposition est réalisée en maximisant un critère tel que la corrélation ou la covariance entre les composantes. Ce critère maximise la corrélation au carré $corr^2(\mathbf{c}_{X1}, \mathbf{c}_{X2})$ ou la covariance au carré $cov^2(\mathbf{c}_{X1}, \mathbf{c}_{X2})$.

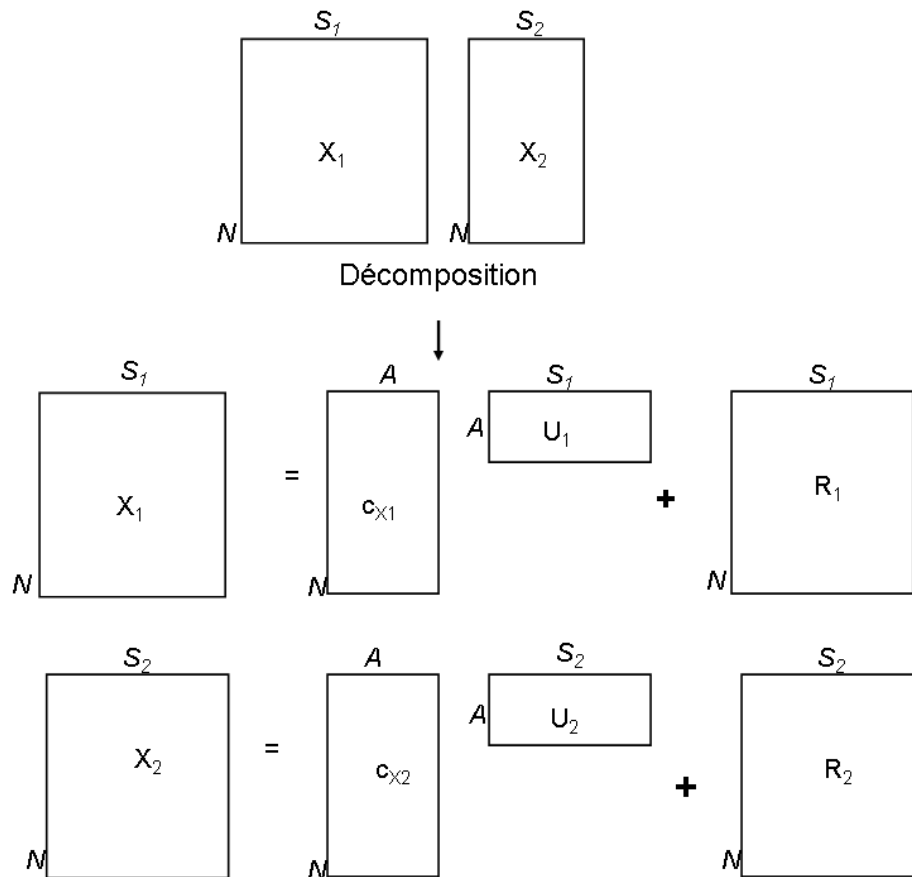


Figure A28 Illustration de l'analyse multibloc de deux tableaux à deux entrées. Chaque tableau X_k est décomposé en A composantes C_k et A profils variables U_k . La reconstitution du tableau X_k est le produit des composantes par les profils variables + le résidu $X_k = C_k \cdot U_k + R_k$

Plusieurs auteurs se sont intéressés à faire la liaison entre deux tableaux de données [Hotelling, 1936 ; Tucker, 1958 ; Chessel et Hanafi, 1996 ; Tenenhaus, 1999]. Parmi les principales méthodes développées dans ce cadre, on peut citer l'analyse canonique des corrélations [Hotelling, 1936], l'analyse inter-batterie de Tucker [Tucker, 1958], et l'analyse PLS2 [Wold *et al.*, 1987].

L'analyse canonique de corrélations permet de calculer les composantes c_{X1} et c_{X2} en maximisant la corrélation entre ces deux composantes. Cependant, sa forte sensibilité à la multicolinéarité des variables la rend peu utilisable dans le cadre de l'analyse spectrale. Pour pallier à ce problème, Devaux *et al.*, (1993) ont appliqué l'analyse canonique des corrélations sur les composantes principales des tableaux d'origine X_1 et X_2 . Il faut noter que la recherche des couples de composantes de corrélation maximale, donne des composantes canoniques bien corrélées entre elles, mais pas nécessairement explicatives de leur tableau d'origine, l'inertie de chacun des tableaux n'étant pas prise en compte par le critère basé sur la corrélation [Gleason, 1976 ; Tenenhaus, 1998].

Tucker [1958] propose une modification du critère de l'analyse canonique en basant le critère de maximiser une covariance et non plus une corrélation. Les composantes c_{X1} et c_{X2} recherchées sont à la fois corrélées entre elles, mais aussi représentatives du tableau initial.

Aujourd'hui, la majorité des méthodes liant deux tableaux de données est basée sur la maximisation du critère de la covariance auquel est associé des contraintes différentes de normes et de déflations [Burnham *et al.*, 1996]. Le choix d'un critère, d'un type de déflation ou d'une contrainte de normalisation oriente le choix d'une méthode. L'approche PLS2 qui est une généralisation de l'algorithme NIPALS, cherche à prédire un tableau Y par un autre tableau X . La méthode PLS2 est l'approche la plus utilisée de nos jours [Blanco et Peguero ; 2008 ; Pedro et Ferreira, 2007]. Son algorithme est le suivant :

Algorithme PLS2

1. Initialiser le vecteur $c_Y^{(0)}$: on prend en général une colonne de Y .
2. Calculer le profil variable de X : $u_1^{(1)} = \frac{X' c_Y^{(0)}}{c_Y^{(0)' c_Y^{(0)}}$
3. Calculer la composante de X $c_X^{(1)} = X u_1^{(1)}$
4. Calculer Le profil variable de Y : $u_2^{(1)} = \frac{Y' c_X^{(1)}}{c_X^{(1)' c_X^{(1)}}$
5. Calculer la nouvelle valeur de la composante de Y $c_Y^{(1)} = \frac{Y' u_2^{(1)}}{u_2^{(1)' u_2^{(1)}}$
6. Répéter les étapes 2 à 5 jusqu'à convergence du critère $d = \|c_Y^{(0)} - c_Y^{(1)}\|$

De la même manière que l'algorithme NIPALS, une fois les composantes et les profils variables à la première dimension obtenus, le calcul des composantes pour les dimensions supérieures se fait au moyen de la déflation en utilisant les équations suivantes :

$$X^{(h)} = X^{(h-1)} - \frac{c_X^{(h-1)} c_X^{(h-1)'}}{\|c_X^{(h-1)}\|^2} X^{(h-1)} \quad \text{Equation A16}$$

$$Y^{(h)} = Y^{(h-1)} - \frac{c_X^{(h-1)} c_X^{(h-1)'}}{\|c_X^{(h-1)}\|^2} Y^{(h-1)} \quad \text{Equation A17}$$

La même procédure sera itérée sur le tableau $Y^{(h)}$ et $X^{(h)}$ pour obtenir les composantes pour les dimensions supérieures

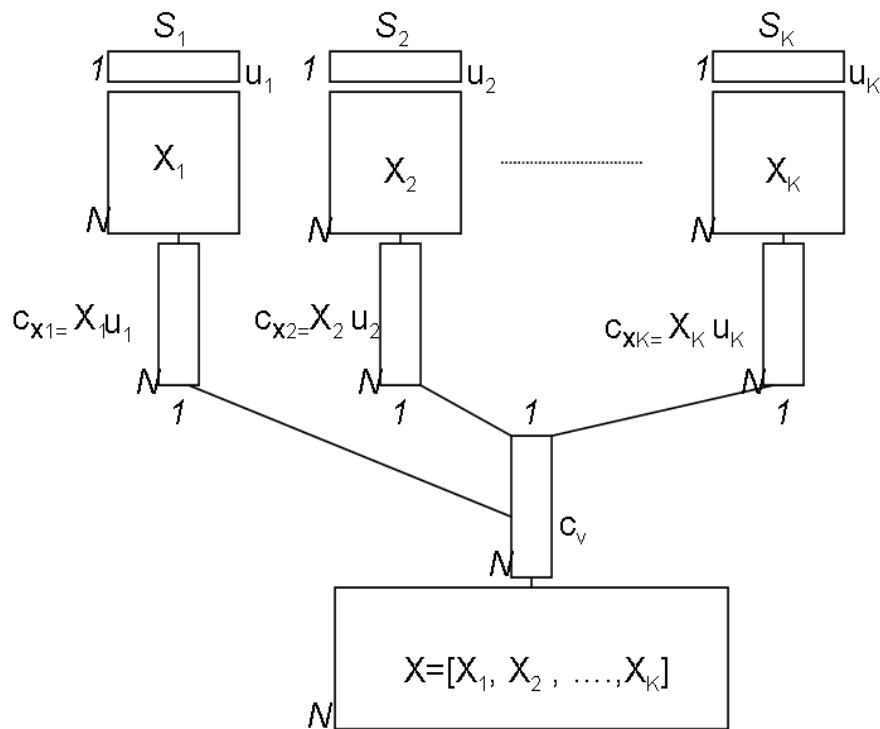
La méthode PLS2 est appliquée dans le cas où un tableau \mathbf{Y} est prédit par un tableau \mathbf{X} . La déflation du tableau \mathbf{X} et \mathbf{Y} en PLS2 est faite par rapport aux composantes du tableau de prédiction \mathbf{X} . La méthode PLS2 est donc une méthode non symétrique. La méthode PLS2 est basée sur une modification, portant sur les déflations de l'analyse inter-batterie de Tucker. Les solutions du premier ordre sont donc équivalentes. Ce sont les solutions à des ordres supérieurs qui sont différentes. La déflation du tableau \mathbf{X} et \mathbf{Y} en analyse inter-batterie de Tucker est faite par rapport à leur composante respective. Ce qui fait de l'analyse inter-batterie de Tucker une méthode symétrique [Tucker, 1958].

2.4.2 Analyse de K tableaux de données

Considérons un ensemble de tableaux $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$ constitués respectivement de S_1, S_2, \dots, S_K variables spectrales (Fig. A29) mesurées sur les mêmes N individus. Par la suite, ces tableaux sont supposés centrés. La structure des données est illustrée par la figure A24. L'analyse multibloc de K tableaux de données a pour but de décomposer les K tableaux en composantes $\mathbf{C}_{X_1}, \mathbf{C}_{X_2}, \dots, \mathbf{C}_{X_K}$ calculées pour être le plus ressemblantes possibles. Les méthodes conçues pour analyser deux tableaux de données ont été étendues pour pouvoir analyser K tableaux de données. Le critère d'optimisation adopté est basé sur des carrés de covariances entre les composantes \mathbf{C}_{X_k} associées à chaque bloc \mathbf{X}_k et une composante globale \mathbf{c}_g . Le critère de maximisation de la $\text{cov}^2(\mathbf{c}_{X_1}, \mathbf{c}_{X_2})$ écrit pour deux tableaux est étendue à plus de deux tableaux en introduisant une variable globale \mathbf{c}_g .

$$\sum_{k=1}^K \text{cov}^2(\mathbf{c}_{X_k}, \mathbf{c}_g) \quad \text{Equation A18}$$

Le calcul de matrice de variance covariance n'est possible que pour deux tableaux. C'est dans ce sens, qu'une composante globale liant les différentes composantes \mathbf{C}_{X_k} est nécessaire. La composante globale est interprétée comme donnant une vision synthétique de tout l'ensemble de l'information apportée par les tableaux.

Figure A29 Analyse multibloc de K tableaux de données

Un certain nombre de méthodes d'analyse de K tableaux à deux entrées ont été proposées : l'analyse canonique Généralisée de Carroll [Carroll, 1968], l'Analyse de Co-inertie Multiple (ACOM) [Chessel et Hanafi, 1996], l'Analyse en Composantes Communes et Poids Spécifiques (ACCPS) [Qannari *et al.*, 2000], l'Analyse en Composantes Principales Consensuelles (CPCA) [Westerhuis *et al.*, 1998] et l'analyse en composante Hiérarchiques (HPCA) [Wold *et al.*, 1996]. De nouveaux outils d'interprétation ont été développés autour de ces méthodes, rendant plus pertinente l'exploitation de données [Mazerolles *et al.*, 2002; Böcker *et al.*, 2007]. Des travaux permettant de faire la liaison entre les méthodes multibloc ont été également développés. Par exemple, Hanafi *et al.* (2011) ont montré qu'il existe une liaison entre l'analyse en composantes principales consensuelles (CPCA) et l'analyse de co-inertie multiple (ACOM) [Hanafi *et al.*, 2011]. En effet, les deux méthodes précédentes optimisent le même critère. La différence réside dans la façon d'appliquer la déflation. La déflation est faite sur les composantes par bloc pour l'analyse de co-inertie multiple, par contre, la déflation est faite sur les composantes globales pour la l'Analyse en Composantes Principales Consensuelles. Les solutions de premier ordre sont similaires mais les solutions d'ordres supérieures diffèrent.

2.5 Analyse multivariées d'images hyperspectrales

2.5.1 Méthodes d'analyse d'images hyperspectrales

L'analyse d'images hyperspectrales a été développée autour des images satellites [Richards et Jia, 1999], astrophysiques [Hege *et al.*, 2003], contrôle de qualité [Liu *et al.*, 2010 ; Qiao *et al.*, 2007] etc... L'application en microspectroscopie est en émergence [Kohler *et al.*, 2007 ; Williams *et al.*, 2009].

L'imagerie hyperspectrale a été utilisée pour détecter et cartographier une grande variété de matériaux ayant des spectres de réflectance caractéristique. Par exemple, les images hyperspectrales ont été utilisées par les géologues pour la cartographie minérale [Clark *et al.*, 1992] et pour la détection des propriétés des sols, y compris l'humidité, la teneur en matière organique, et la salinité [Clark *et al.*, 1992]. Les scientifiques du végétal ont utilisé avec succès l'imagerie hyperspectrale pour identifier les espèces végétales [Bindlish et Barros, 2011] et pour étudier la chimie du végétal [Aber et Martin, 1995]. Le personnel militaire a utilisé l'imagerie hyperspectrale pour détecter les véhicules militaires sous couvert végétal partiel, et beaucoup d'autres cibles militaires étaient l'objectif de détection.

L'analyse d'image multivariée a pour objectif d'extraire des informations à partir des images dans le but de, par exemple, l'analyse exploratoire [Geladi. *et al.*, 1986; De Juan *et al.*, 2005], la segmentation ou la classification en régions [Lin. *et al.*, 2006 ; Piqueras *et al.*, 2011], et la prédiction [Geladi *et al.*, 2004 ; Burger et Geladi, 2006].

L'objectif de l'analyse est généralement de déterminer des régions dans l'image basée sur l'ensemble des variables spectrales. Lorsque la dimension spectrale est considérée comme prépondérante dans la structure de données, la chimiométrie s'impose naturellement comme méthode de traitement des images hyperspectrales. Les outils d'analyse multivariées à savoir les méthodes descriptives [Jaumot *et al.*, 2005; De Juan *et al.*, 2009] et les méthodes prédictives [Geladi et Esbensen, 1991 ; Chevallier, 2006 ; Salzer et Siesler, 2009 ; Prats-Montalbán *et al.*, 2009] s'appliquent. D'autres types de méthodes basées sur les approches non linéaires tel que les champs de Markov et l'estimation Bayésienne [Arngren *et al.*, 2011 ; Moussaoui *et al.*, 2008] sont également très utilisés dans l'analyse d'images hyperspectrales.

2.5.2 Analyse en composantes principales d'images hyperspectrales

Lorsque l'analyse de la variabilité spectrale est considérée comme prépondérante, il est souvent admis d'ignorer les dimensions spatiales N_l et N_c de l'image et de considérer le cube de données comme une collection de spectres (Fig. A30). Cela revient à déplier le cube pour former une matrice bidimensionnelle de dimensions $N_l \times N_c$ et S semblable à celle obtenue en spectroscopie classique [Geladi et Grahn, 1996]. Le procédé utilisé est appelé dépliement ou matricisation [Smilde *et al*, 2004]. La transformation inverse est appelée repliement. Dans ce paragraphe, l'analyse en composantes principales est décrite comme étant une méthode de base de l'analyse multivariée d'image hyperspectrale en chimométrie.

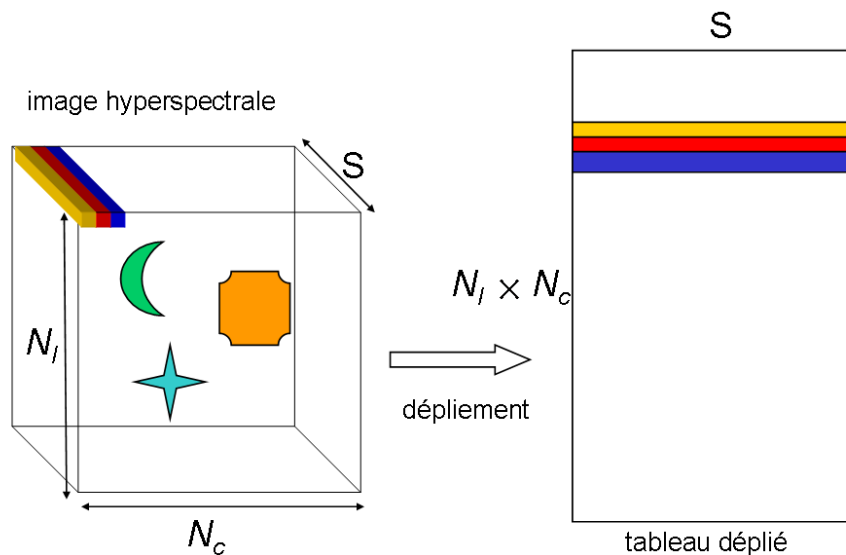


Figure A30 Dépliement d'une image hyperspectrale

Une fois l'image hyperspectrale dépliée, le tableau obtenu peut être analysé en utilisant l'analyse en composantes principales (Fig. A31). Les A profils variables sont interprétés pour révéler les variations spectrales, A étant le nombre de composantes calculées. Les A valeurs propres représentent comme en analyse classique le pourcentage de variance décrite par chaque composante principale calculée. Les composantes se présentent sous la forme d'un tableau de dimension $N_l \times N_c$ lignes et A colonnes. Les A composantes sont par la suite repliées pour obtenir les images composantes synthétiques individuelles. Elles peuvent être affichées soit en niveau de gris où la valeur la plus faible de la composante est représentée en noir et la valeur la plus élevée est représentée en blanc ou sous la forme d'une image couleur en superposant 3 images composantes et en l'affichant comme une image RGB [Pedersen, 1994 ; Geladi et Grahn, 1996].

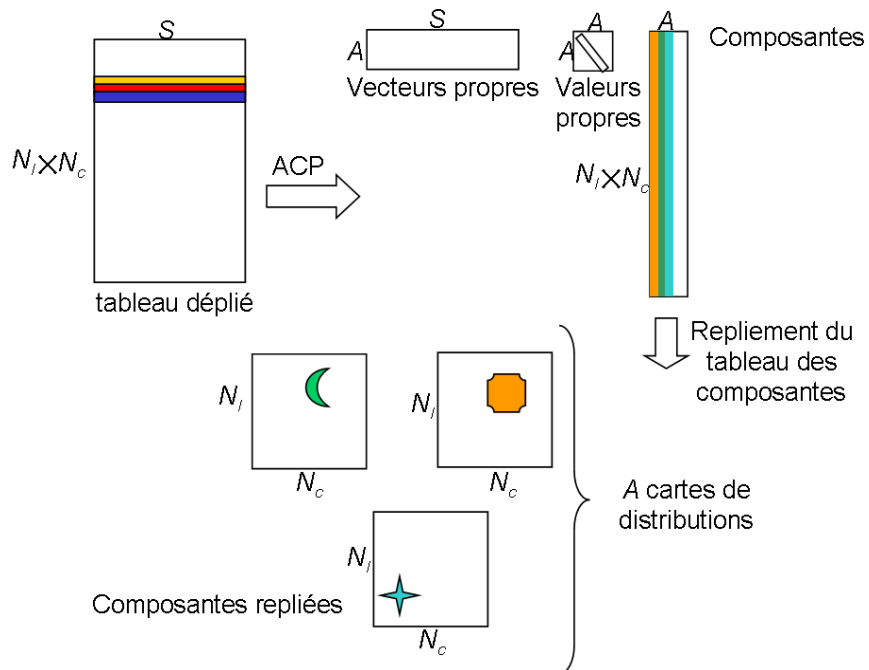


Figure A31 Analyse en composantes principales d'une image hyperspectrale

Conclusion

Dans cette partie nous avons présenté les structures de données que peuvent prendre les données spectrales à savoir :

- un tableau de spectres : les données sont issues d'une seule technique spectrale.
- plusieurs tableaux de données : les données sont issues de plusieurs techniques spectrales.
- un cube de données : les données issues de l'imagerie hyperspectrale

Pour analyser de telles structures de données, des méthodes d'analyse multivariées ont été présentées. L'analyse en composantes principales représente la méthode la plus utilisée pour analyser un tableau de spectres à deux entrées. L'analyse d'images hyperspectrales a été présentée autour de l'analyse d'un tableau à deux entrées en ignorant les dimensions spatiales afin de constituer un tableau de données à deux entrées. Le couplage de plusieurs tableaux de données a donné naissance aux méthodes multibloc. Les propriétés des composantes et des profils variables des méthodes multibloc sont fortement liées au critère d'optimisation, à la déflation ainsi qu'aux contraintes de normalisation imposées. Le choix d'une méthode d'analyse par rapport à une autre dépend des propriétés des composantes et des profils variables recherchés.

VI Conclusion : Couplage d'image hyperspectrale pour la caractérisation chimique d'échantillons complexes

1 Position du problème

Notre projet a pour but final de mettre en évidence les relations entre la composition des parois cellulaires et leur dégradabilité. La saccharification par voie enzymatique est proposée dans le cadre du développement durable. Des recherches sont menées pour déterminer les facteurs limitant la dégradation. Ces études ont montré que cette dégradation ne dépend pas que de la composition globale des parois cellulaires mais aussi de la spécificité de la composition chimique des tissus de la plante [Jung et Casler, 2006-a] Dans ce travail, nous proposons d'explorer la variation chimique au sein des types cellulaires, le maïs étant choisi comme une plante modèle et d'application.

L'investigation de la variabilité de composition des parois cellulaires est analysée par microspectroscopie en acquérant des images hyperspectrales. Une seule technique spectrale ne permet pas de révéler toutes les composantes chimiques, nous avons proposé d'appliquer trois méthodes spectroscopiques complémentaires et différentes : l'infrarouge, la fluorescence UV et le Raman, afin de distinguer les différentes composantes chimiques présentes dans la paroi cellulaire.

Classiquement, la microspectroscopie est utilisée en mode point par point. Les spectres sont acquis de manière supervisée à des positions sélectionnées par l'utilisateur. Nous proposons d'acquérir des images hyperspectrales afin d'explorer l'hétérogénéité spatiale sans *a priori*.

L'objectif de la thèse est de développer les méthodes et les approches pour cartographier la composition chimique des tiges de maïs en analysant de manière conjointe des images hyperspectrales acquises avec trois techniques spectrales. Les développements ont concerné toute l'analyse de l'acquisition de données jusqu'au développement des outils et leur interprétation.

L'acquisition des images hyperspectrales pour un même échantillon avec différents appareils dégage plusieurs contraintes :

- acquisition non simultanée des images avec des contraintes de temps d'acquisition,
- résolutions spatiale et spectrale spécifiques pour chaque méthode,
- présence de déformations spectrales, variations de ligne de base, phénomènes de diffusion de la lumière, etc...

Nous avons retenu la chimiométrie pour analyser d'une manière conjointe des images acquises avec les trois méthodes spectroscopiques. La réalisation de cet objectif nécessite le développement de méthodes pour mettre en forme les données multirésolutions et les analyser avec les méthodes appropriées. Ces méthodes comprennent :

- Le recalage spatial des images des trois modalités.
- L'appariement de pixels de différentes résolutions.
- L'analyse de données conjointes basée sur des méthodes multibloc.
- La détermination de la complémentarité et de la redondance entre les gammes spectrales.

2 Démarche adoptée

La première étape de ce travail consiste en la mise au point d'un protocole d'acquisition conjointe d'images hyperspectrales de la même région d'un échantillon. La deuxième étape est dédiée à l'analyse individuelle de chaque image hyperspectrale. Une étape primordiale, avant de passer à l'analyse, est le pré-traitement de données. Les images hyperspectrales infrarouge, fluorescence et Raman doivent alors être corrigées et normalisées avant de les soumettre aux méthodes d'analyse d'image multivariées. Le but de l'analyse individuelle des images hyperspectrales est de déterminer l'information fournie par chacune des techniques utilisées. La troisième étape consiste en l'appariement de données multirésolutions, l'établissement de la structure de données et le développement de l'analyse conjointe de données. L'appariement des images hyperspectrales se fait par le moyen des techniques de recalage d'images. Chaque image hyperspectrale est une structure à trois dimensions : deux dimensions spatiales et une dimension spectrale. Ces images doivent être combinées pour former un ensemble de données pouvant être analysées en utilisant les techniques multibloc. Pour établir la structure de données, les différences des résolutions spatiales et spectrales des trois techniques doivent être prises en compte. Nous avons utilisé l'environnement Matlab pour développer tous les outils relatifs à l'analyse et à l'interprétation de données. La figure A32 résume les différentes étapes de ce travail.

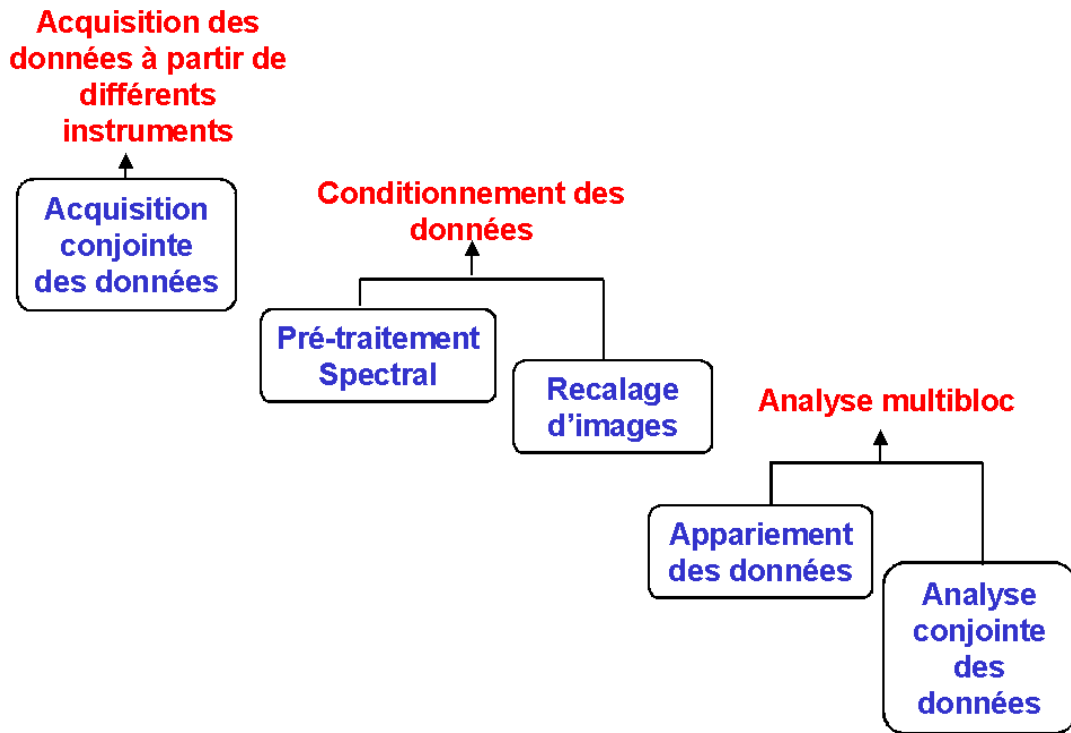


Figure A32 Démarche adoptée pour le couplage d'images hyperspectrales multirésolutions

Partie B :

Données...Acquisition & exploration

Sommaire

Introduction	62
I Préparation des tiges de maïs	62
1 Echantillons	62
2 Choix de la région	63
3 Préparation de l'échantillon	66
II Acquisition conjointe des images hyperspectrales multimodales sur les lignes SMIS et DISCO du Synchrotron SOLEIL	67
1 Imagerie hyperspectrale	67
1.1 Imagerie infrarouge	68
1.2 Imagerie Raman	68
1.3 Imagerie de fluorescence.....	69
2 Protocole d'acquisition conjointe des données	69
3 Description des collections des images hyperspectrales	72
3.1 Collection des images hyperspectrales 4f1 1mb-1	73
3.2 Collection des images hyperspectrales 4f1 1mb-2	74
III Pré-traitement	75
1 Déformations spectrales et spatiales des données	75
1.1 Exemples de spectres	75
1.2 Variation spatiale.....	76
2 Stratégie.....	78
3 Méthodes de pré-traitement.....	79
3.1 Elimination des spikes.....	79
3.2 Filtrage du bruit	80
3.3 Normalisation spatiale et spectrale.....	80
3.3.1 Prise en compte des régions sans signal	80

3.3.2 Prise en compte des variations locales d'intensité	81
4 Applications aux images hyperspectrales.....	82
4.1 Pré-traitement des spectres infrarouge	82
4.2 Pré-traitement des spectres de fluorescence.....	84
4.3 Pré-traitement des spectres Raman.....	85
IV Analyse individuelle de la collection 4f11mb-2	87
1 Exploration des données.....	87
2 Analyse en composantes principales de la collection de données 4f11mb-2.....	90
2.1 Analyse en composantes principale de l'image infrarouge.....	90
2.2 Analyse en composantes principales des images hyperspectrales de fluorescence	92
2.3 Analyse en composantes principales des images hyperspectrales Raman.....	95
Conclusion	97

Introduction

Ce chapitre est dédié à la présentation des données acquises au cours de la thèse depuis la préparation des échantillons jusqu'à l'exploration des données. La première partie décrit le protocole d'acquisition des images hyperspectrales. La deuxième partie est consacrée aux pré-traitements des données acquises. La troisième et dernière partie, présente un exemple d'analyse individuelle au sein de chaque spectroscopie. Les analyses individuelles ont été faites pour présenter l'information apportée par chaque méthode.

I Préparation des tiges de maïs

1 Echantillons

L'échantillon modèle retenu pour réaliser les développements méthodologiques de la thèse est le génotype de maïs "F2", génotype de référence pour l'INRA, souvent utilisé pour réaliser des hybrides d'intérêt agronomique [Tovar-Gómez *et al.*, 1997 ; Machinet *et al.*, 2011]. Trois stades de développement étaient initialement prévus dans l'étude pour lesquels les taux de croissance des lignines sont observés [Jung et Casler, 2006-a] : 9 feuilles, floraison femelle et ensilage. Le stade 9 feuilles est celui où 9 feuilles ligulées sont visibles. Le maïs atteint la floraison femelle quand les épis présentent leurs soies et toutes les feuilles sont différenciées (Fig. B1). Le stade ensilage est le stade de maturité physiologique de la plante.



Figure B1 Plante du génotype F2 de maïs au stade floraison femelle.

Les tiges de maïs ont été récoltées et les entrenœuds sous épi ont été prélevés et conservés dans un mélange à 70% (v/v) d'éthanol/eau (Fig.B2).



Figure B2 Prélèvements des entrenœuds sous épi

2 Choix de la région

La figure B3 montre les propriétés d'autofluorescence des parois cellulaires, en macroscopie et microscopie confocale.

La figure B3-a montre une image acquise sur une coupe de tige de maïs au stade ensilage au microscope de fluorescence (Nikon AZ100) de la plateforme IMAGIF de Gif sur Yvette avec une excitation à trois longueurs d'ondes 340, 380 et 488 nm. Dans la figure B3-b, un faisceau vasculaire de tige de maïs au stade ensilage est observé en microscopie confocale (plateforme BIBS INRA-Nantes) avec une excitation simultanée à 375, 488 et 561 nm. Les images révèlent des variations chimiques en fonction des types cellulaires. Dans l'image acquise en macroscopie de fluorescence (Fig. B3-a), l'écorce et la gaine de sclérenchyme apparaissent en vert. La couleur verte correspond à une réponse après excitation dans le bleu. Cette réponse révèle la plus grande concentration en lignine au sein de cet échantillon. La couleur rose apparaissant au niveau du parenchyme avec des couleurs différentes au milieu de la coupe, près de l'écorce et à l'intérieur du faisceau correspond à une réponse après excitation dans l'UV. Dans l'image confocale du faisceau (Fig. B3-b), différentes couleurs apparaissent (orange, jaune, vert, bleu...) montrant la grande variabilité de composition des parois des cellules. Sur la figure B3-b, on distingue le sclérenchyme, le phloème, le xylème, le parenchyme vasculaire ainsi que le parenchyme entourant les faisceaux par différentes

couleurs. Ce ci montre que les types cellulaires ont des teneurs différentes en lignines [Koch et Kleist, 2001].

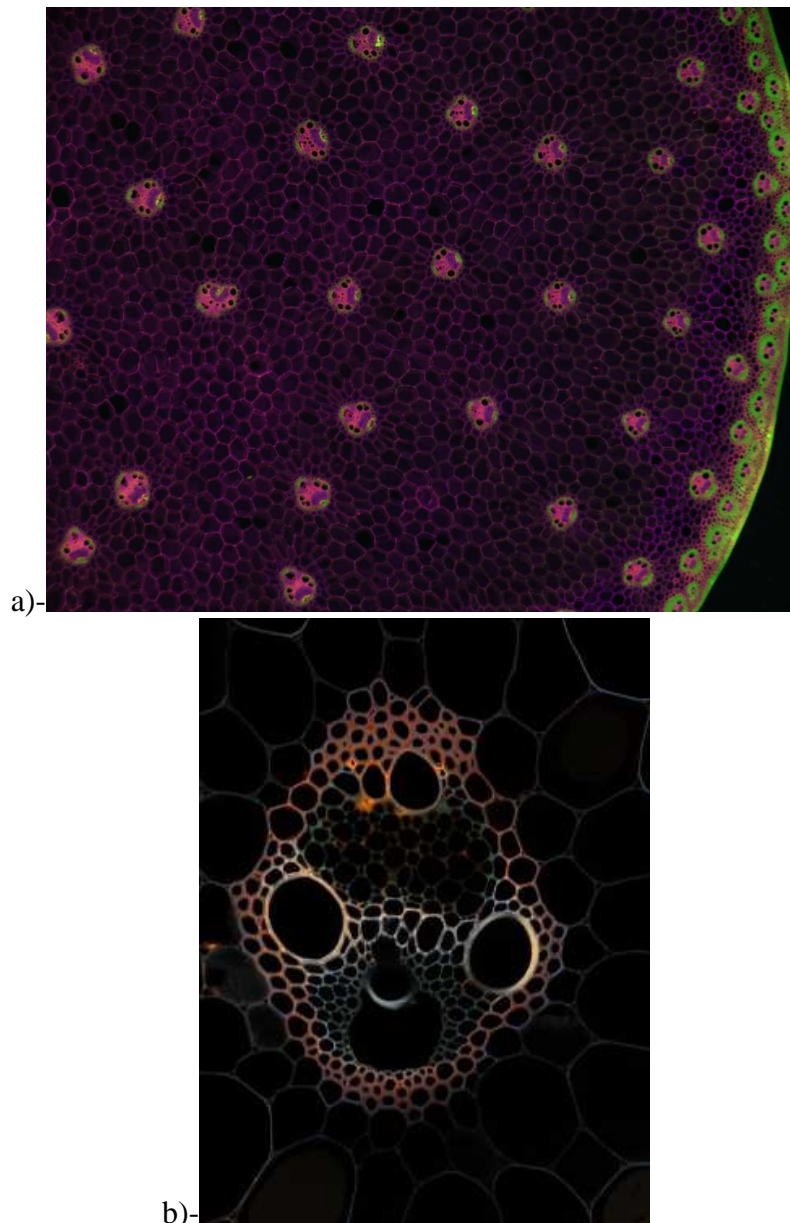


Figure B3 Observation de section de tige de maïs au stade ensilage en utilisant : a)- la fluorescence en macrovision, image en fausse couleur : canal rouge (émission dans le bleu après excitation à 380 nm (UV)), canal vert (émission dans le vert après excitation à 488 nm (bleu)), canal bleu émission dans le bleu après excitation à 340 nm) b)- la microscopie confocale de fluorescence, rendu d'image couleur après excitation simultanée aux longueurs d'ondes (375, 488, 561) et récupération sur le détecteur multicanal du microscope confocal (Nikon A1).

L'impact de la variation de composition sur la dégradation enzymatique a été mis en évidence en réalisant des dégradations enzymatiques sur coupe (Fig. B4). Des enzymes de type cellulase et xylanase ont été utilisées en excès.

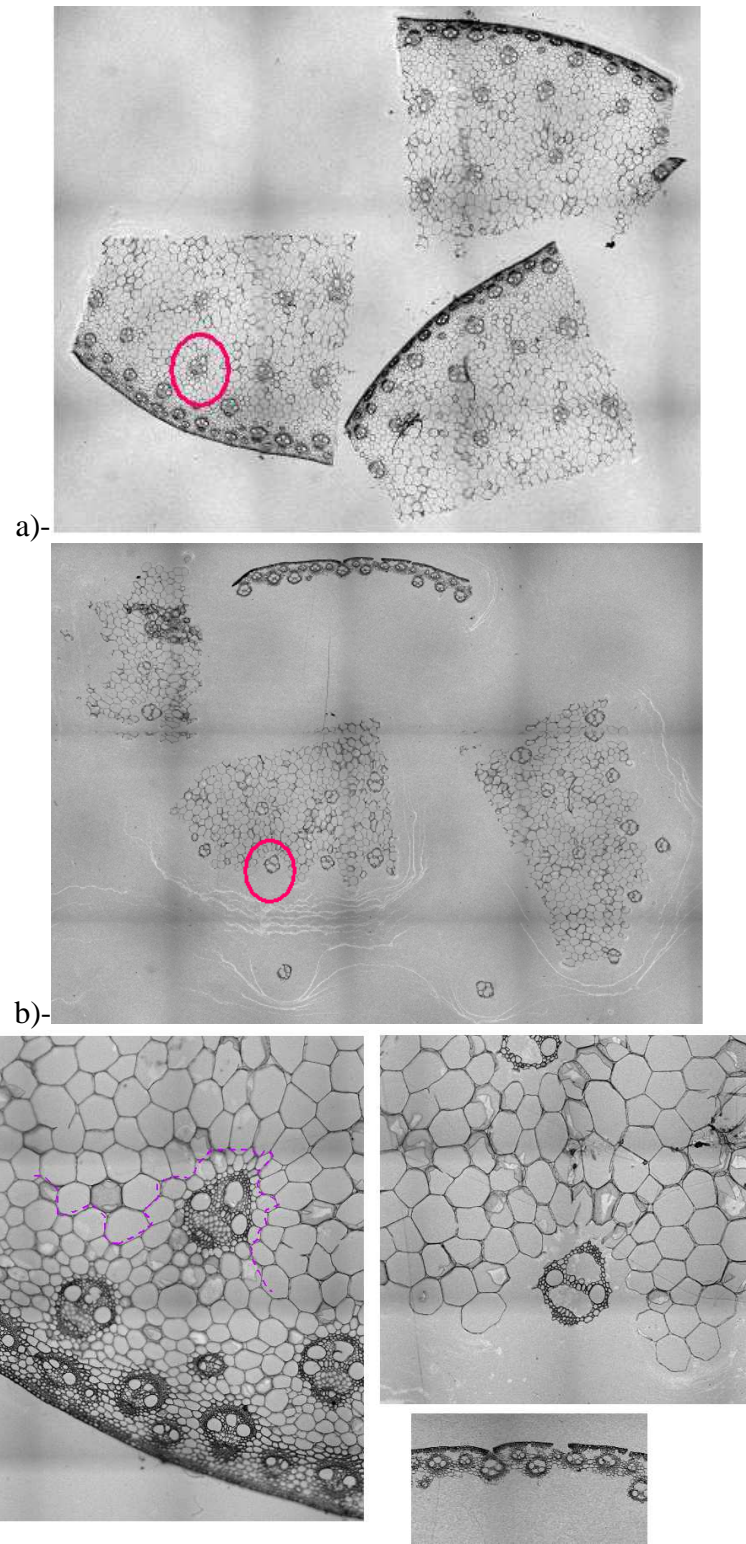


Figure B4 Section de tige de maïs a)-avant dégradation enzymatique b)- après dégradation enzymatique c)-Dégradation du faisceau du bord. Dans l'image de gauche, la frontière des cellules dégradées/ non dégradées a été tracée.

Après dégradation, l'écorce est séparée du reste du tissu conduisant à la libération de faisceaux dans la coupe. Les cellules de parenchyme près du bord, du sclérenchyme et celles

du parenchyme entourant les faisceaux ont été dégradées. A l'intérieur du faisceau, les cellules du phloème ainsi que celles du parenchyme vasculaire ont été également dégradées (Fig. B4). Au vu de l'ensemble de ces données, une région dans un faisceau près du bord de l'écorce a servi de support au développement des méthodes chimiométriques. Les résultats sont interprétés sur la base de la comparaison de composition de différents types cellulaires : du xylème, phloème, sclérenchyme et parenchyme (Fig. B5).

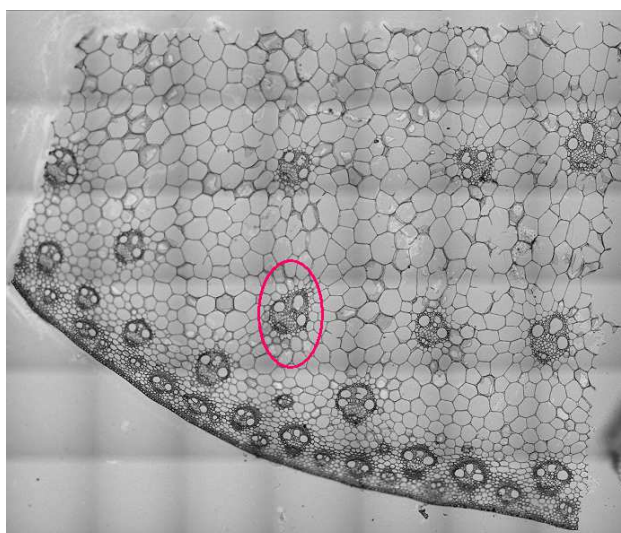


Figure B5 Faisceau sélectionné

3 Préparation de l'échantillon

Les échantillons ont été prélevés à partir du milieu des entre-nœuds de tige de maïs, stockés dans 70% (v/v) d'éthanol/eau, inclus dans de la paraffine et découpés à l'aide d'un microtome. Certaines cellules contenaient des grains d'amidon. Les coupes ont été désamidonnées en utilisant une alpha-amylase. Elles ont été déparaffinées en suivant le protocole décrit dans Jamme *et al.*, (2008). En outre, pour ne pas masquer les signaux des lignines par ceux des protéines, les protéines ont été dégradées à l'aide d'une protéase [Prosky *et al.*, 1984].

Nous avons testé plusieurs épaisseurs de coupe : 10, 15, 20, 30 et 50 μm . Malgré les essais de tests des épaisseurs de coupes et les expériences préliminaires faites sur le microspectrophotomètre infrarouge de notre laboratoire, il n'a pas été possible d'obtenir des signaux suffisamment intenses pour le stade 9 feuilles et des signaux non saturés au stade ensilage pour tous les types cellulaires. Donc, la suite du travail a été réalisée uniquement à partir de coupes au stade floraison femelle.

Les résultats préliminaires ont montré qu'au stade floraison femelle, une épaisseur de coupe de 10 μm était adaptée pour l'acquisition spectrale dans le moyen infrarouge, la fluorescence et le Raman. Les coupes ont été recueillies sur des lames de ZnS, transparentes dans le moyen

infrarouge. Il a été validé que ces lames sont adaptées pour les deux autres spectroscopies. En effet le ZnS n'est pas fluorescent dans le proche UV et il présente un signal en Raman à des nombres d'ondes n'appartenant pas à la région de $1800-800\text{ cm}^{-1}$ -la région qui nous intéressera du spectre Raman- (Fig B6).

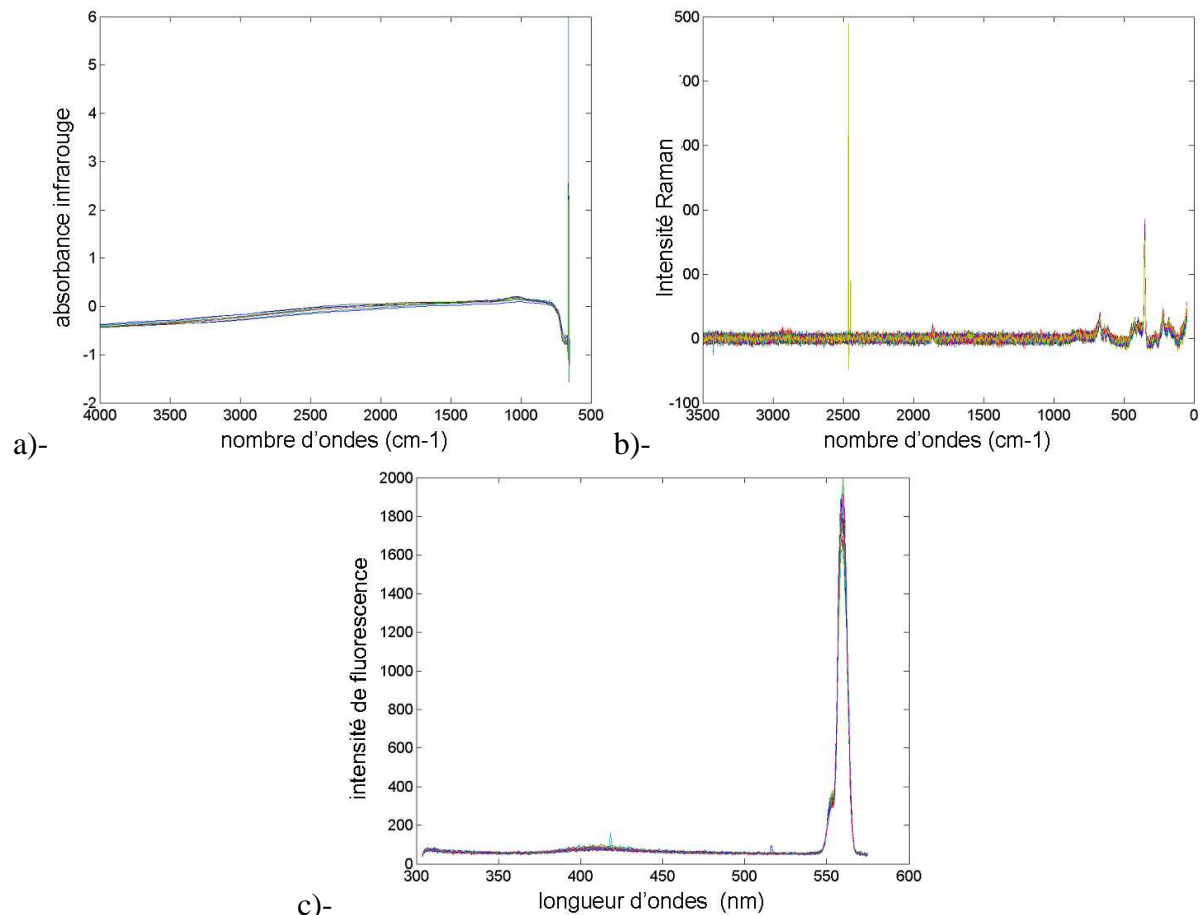


Figure B6 Spectres de la lame ZnS a)-infrarouge b)- Raman c)- fluorescence

II Acquisition conjointe des images hyperspectrales multimodales sur les lignes SMIS et DISCO du Synchrotron SOLEIL

1 Imagerie hyperspectrale

Deux projets ont été soumis pour obtenir du temps de faisceau au Synchrotron SOLEIL. Le premier projet a été accepté durant la première année de thèse et le deuxième durant la deuxième année. L'objectif du premier projet était de comparer les stades de développement et mettre au point l'acquisition conjointe. Le second projet avait pour but l'amélioration de la qualité des données acquises lors de la première année et des nouveaux jeux de données équivalents à ceux de la première année ont été acquis. Un autre objectif du deuxième projet

était la comparaison des parois de cellules avant et après dégradation enzymatique. Toutes les images hyperspectrales ont été acquises au synchrotron SOLEIL. Les images infrarouge ont été acquise sur la ligne de lumière SMIS avec la source synchrotron [Dumas *et al.*, 2006], les images Raman en utilisant un laser et les images de fluorescence sur la ligne de lumière DISCO en utilisant la partie UV de la source de rayonnement synchrotron [Giuliani *et al.*, 2009]. En raison du caractère non destructif de la spectroscopie infrarouge, les images infrarouge ont été acquises en premier lieu. Par la suite, le même échantillon a été cartographié en microspectroscopie de fluorescence. Le Raman a été la dernière technique utilisée en raison de la possibilité de destruction des échantillons liée à l'échauffement dû au laser d'excitation.

1.1 Imagerie infrarouge

Le microscope continuum XL permet l'acquisition des spectres infrarouge et des images visibles en fond clair en utilisant le logiciel OMNIC (Thermo Fisher, USA). Le microscope a été utilisé en mode transmission en utilisant l'objectif à 32 X. L'objectif 32 X permet d'avoir des champs de vue de l'ordre de $250 \times 250 \mu\text{m}^2$.

La taille de l'ouverture a été fixée à $10 \times 10 \mu\text{m}^2$ lors de la première session d'acquisition. Les images acquises durant la deuxième session, ont été obtenues en utilisant une taille d'ouverture de $12 \times 12 \mu\text{m}^2$ pour améliorer le rapport signal / bruit. Les spectres ont été acquis dans la gamme infrarouge de $4000\text{-}800 \text{ cm}^{-1}$ avec une résolution spectrale de 8 cm^{-1} et 128 spectres accumulés. Le temps d'acquisition pour chaque spectre était égal à 50 s. Le pas de déplacement a été fixée à $5 \mu\text{m}$.

1.2 Imagerie Raman

Les images hyperspectrales ont été acquises en utilisant un microscope confocal Raman DXR (Thermo Fisher Scientific, WI, USA) avec un laser d'excitation de 532 nm d'une puissance de 10 mW. Pour se repérer dans l'échantillon, le microscope Raman permet d'acquérir des images visibles en fond noir. Les images visibles ont été acquises avec un objectif longue distance 100X (LMPlan FL, N,A=0.8) (Olympus, USA). L'ordre de grandeur du champ de vue est de $60 \times 60 \mu\text{m}^2$.

Les spectres Raman ont été acquis entre 50 et 3500 cm^{-1} avec une résolution de 8 cm^{-1} . La collecte des données et le contrôle de l'image ont été effectués à l'aide du logiciel OMNIC (Thermo Fisher Scientific, WI, USA). Lors de l'acquisition, les spectres ont été corrigés

automatiquement à l'aide d'un polynôme de degrés 5 pour corriger la ligne de base due à une éventuelle fluorescence de l'échantillon.

Le temps d'acquisition pour chaque spectre a été fixé à 5 s. Le pas de déplacement était de 1 μm .

1.3 Imagerie de fluorescence

Le microscope Olympus IX71 permet l'acquisition des spectres de fluorescence UV et des images visibles en fond clair en utilisant le logiciel Labspec (HORIBA JOBIN YVON - Raman). L'objectif utilisé était l'objectif ultrafluar 40 X (Zeiss, Allemagne) permettant d'acquérir un champ de vue maximal de $100 \times 100 \mu\text{m}^2$. Pour cette étude, la longueur d'onde d'excitation a été fixée à 280 nm. Cette longueur d'onde a été choisie comme pouvant différencier particulièrement l'acide férulique de l'acide para-coumariques [Saadi *et al.*, 1998 ; Andersson et Akin, 2008]. Les spectres ont été enregistrés entre 294 et 565 nm par pas de 0,265 nm. Le temps d'acquisition a été fixé à 7 s par spectre. La taille du pas de déplacement a été fixée à $1 \times 1 \mu\text{m}^2$.

La lame infrarouge utilisée étant trop petite pour le porte échantillon du microscope de fluorescence, un dispositif faisant l'intermédiaire entre le porte échantillon du microscope de fluorescence de DISCO et la lame infrarouge a été construit à l'atelier CRAIS de l'unité BIA de l'INRA de Nantes. La figure B7 montre le prototype du porte échantillon construit. Dans la figure B7-a, le porte échantillon est représenté seul. Dans la figure B7-b, il est représenté en mode d'utilisation : la lame infrarouge est représentée en rectangle rouge, la lame de quartz en rectangle vert et le porte échantillon du microscope en gris.

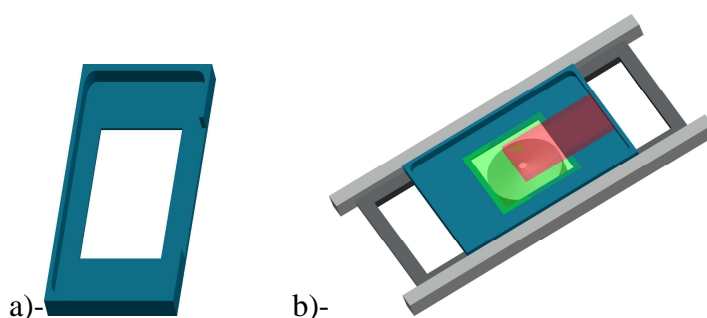


Figure B7 Porte échantillon a)- seul b)- déposé sur le porte échantillon du microscope de fluorescence.

2. Protocole d'acquisition conjointe des données

L'analyse conjointe des données nécessite d'acquérir des spectres pour exactement les mêmes régions pour les trois spectroscopies. Pour chaque spectroscopie, deux types d'images sont

acquises : les images visibles et les images hyperspectrales. Les images visibles servent à se repérer dans l'échantillon et à sélectionner précisément la région à cartographier. La procédure de sauvegarde de l'image visible par les deux logiciels OMNIC et LABSPEC est basée sur des copies d'écran, ce qui engendre des résolutions variables pour la même scène et une qualité médiocre des images visibles. Ces images visibles associées à l'acquisition sont de tailles comparables aux images spectrales (Fig. B8). Elles ne permettent pas de repérer la localisation exacte de la région observée dans le cas de la fluorescence et du Raman.

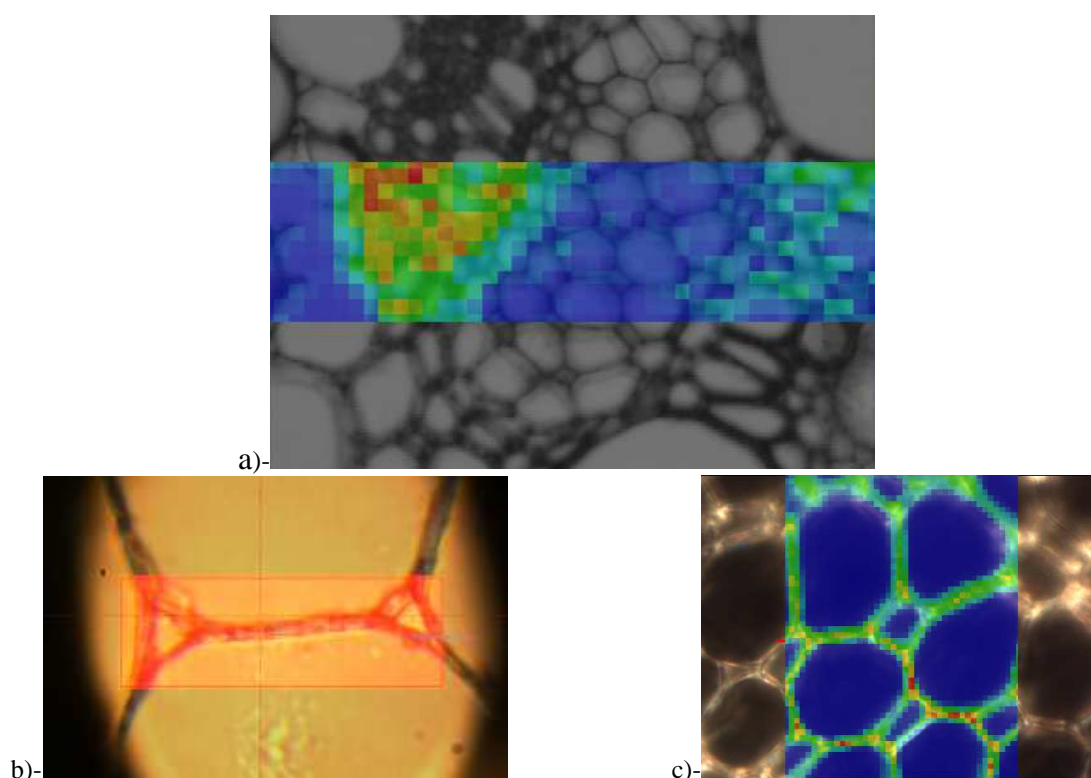


Figure B8 Exemple de superposition de l'image spectrale et de l'image visible a)- infrarouge b)- fluorescence c)- Raman. Les images ont été superposées en utilisant les outils fournis par le logiciel OMNIC pour les images infrarouge et Raman et le logiciel Labspec pour l'image de fluorescence.

Nous avons donc travaillé à partir d'une image visible de référence ayant un plus grand champ de vue et une meilleure qualité pour les trois spectroscopies.

Lors de la première année de thèse, l'image de référence a été acquise en utilisant la deuxième caméra du microscope de fluorescence de la ligne DISCO qui permet d'acquérir et sauver des images de haute qualité. Celle de la deuxième année a été acquise en utilisant le microscope confocal de fluorescence de L'INRA de Nantes. La solution de la deuxième année a été retenue parce qu'elle permet de préparer de manière très précise le protocole d'acquisition avant de partir au Synchrotron. Les régions à cartographier ont été sélectionnées *a priori* à partir des images de référence. Par exemple, sur la figure B9, les régions délimitées par le

rectangle sur l'image d'un faisceau vasculaire représentent les régions à cartographier telles qu'elles ont été définies *a priori*.

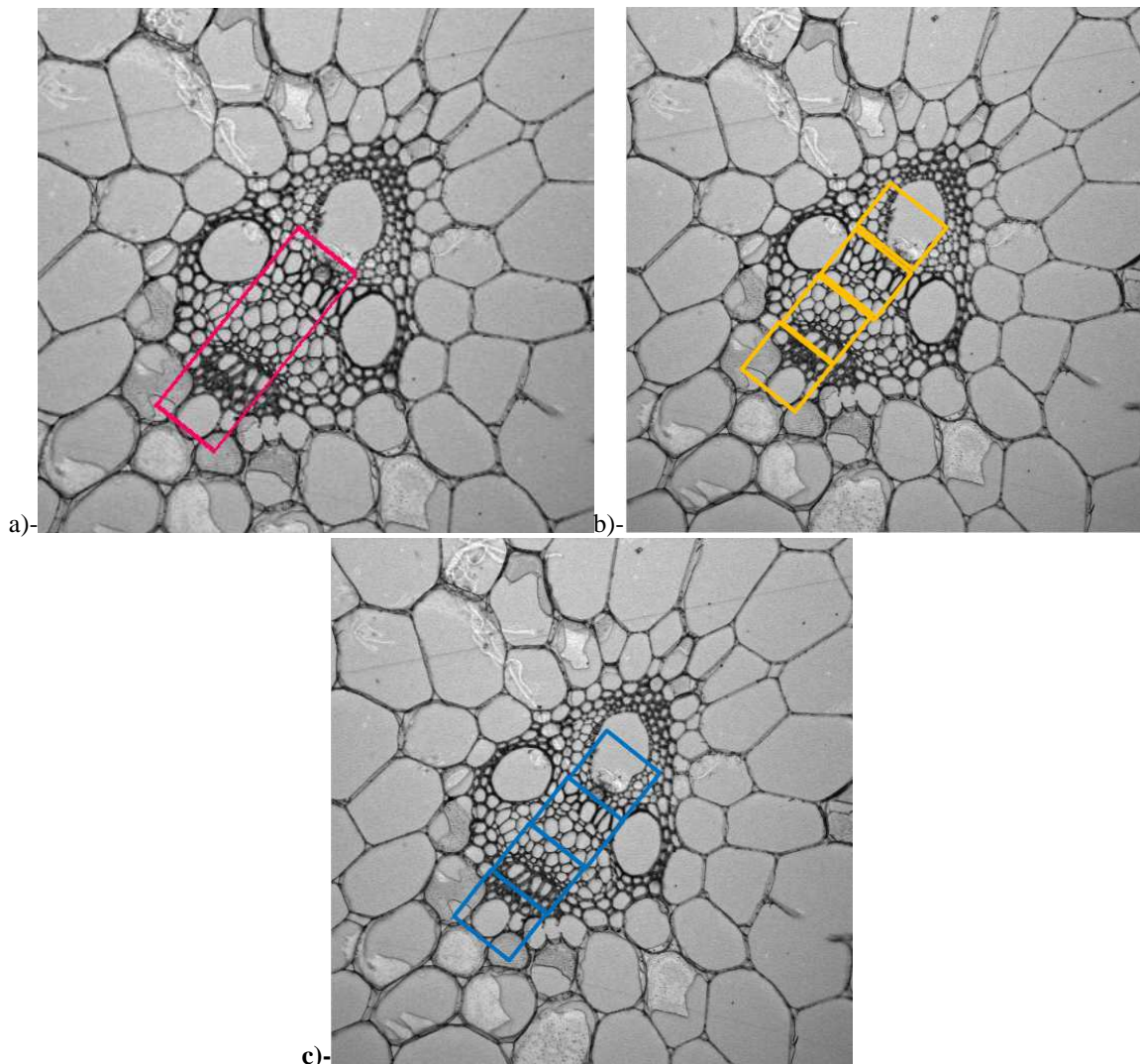


Figure B9 Exemple de planification d'acquisition pour l'analyse conjointe d'une collection d'images hyperspectrales. a)-région de l'image infrarouge (rouge) b)-régions des images Raman (jaune) c)- régions des images de fluorescence (bleu)

Les images hyperspectrales infrarouge ont les plus basses résolutions et les plus grands champs de vue par rapport aux images hyperspectrales de fluorescence et Raman. La stratégie pour pouvoir cartographier les mêmes régions pour les trois spectroscopies est de fixer, en première étape, la région à cartographier pour la spectroscopie ayant le plus grand champ de vue, dans notre cas la microspectroscopie infrarouge. La région à cartographier en infrarouge sera par la suite décomposée en un ensemble de régions ayant de plus petits champs de vue pouvant être cartographiées en fluorescence et Raman.

La figure B10 montre la comparaison entre le schéma prévisionnel et l'acquisition réellement effectuée. A l'acquisition des problèmes ont été rencontrés :

- La présence d'un angle de rotation entre les différentes images due au fait qu'il était très difficile de mettre l'échantillon selon le même angle en le déplaçant d'un microscope à un autre.

- Les images ne sont pas tout à fait superposées à cause du champ de vue et du grandissement du microscope, les petits champs de vue ne permettant pas de se repérer facilement dans l'échantillon.

- Les images spectrales et visibles sont inversées en fluorescence (effet miroir). Les inversions sont liées au chemin optique.

- Des images ont été cartographiées en petites surfaces à cause des contraintes liées à la limitation en temps d'acquisition et à la perte de faisceau lors de l'acquisition. Dans la figure B10-b, une partie de l'image de fluorescence est manquante à cause de la perte de faisceau lors de l'acquisition.

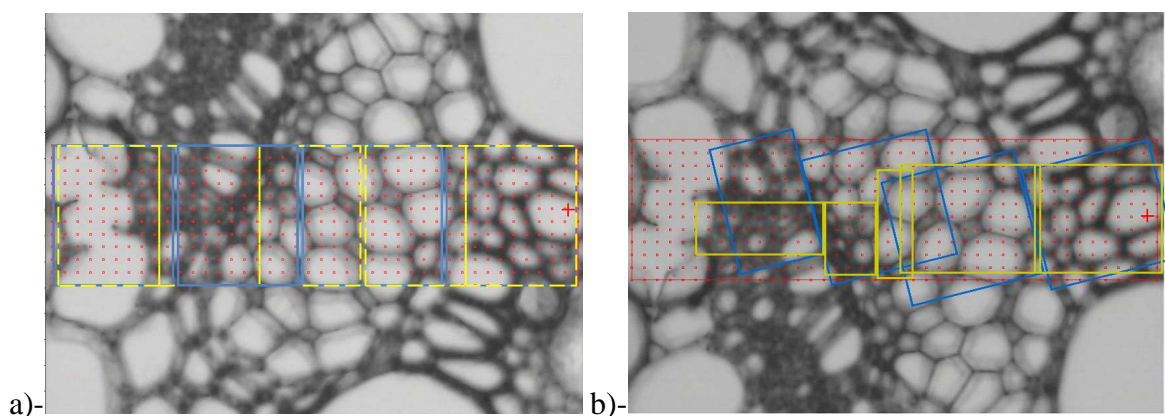


Figure B10 Protocole d'acquisition conjointe des images hyperspectrales infrarouge (rouge), fluorescence (bleu) et Raman (jaune) a)- régions préparées avec le protocole d'acquisition conjointe b)- régions cartographiées.

3 Description des collections des images hyperspectrales

Toutes les collections d'images hyperspectrales acquises pendant la thèse n'ont pas servi aux développements méthodologiques. Nous nous limitons dans cette section à la description des jeux de données qui vont être utiles pour ce qui va suivre dans le document. Les images ont été acquises pour le génotype F2 au stade floraison femelle. Ce génotype est nommé '**4f11m**'. Un faisceau près du bord de l'épiderme désignée par la lettre '**b**'. Ainsi notre échantillon à analyser est nommé **4f11mb**. Pour les deux sessions d'acquisition, des jeux de données homologues ont été acquis. Les spectres acquis dans le faisceau ont été comparés en fonction des différents types cellulaires présents.

3.1 Collection des images hyperspectrales 4f11mb-1

La collection d'images hyperspectrales 4f11mb-1 a été choisie parmi celles acquises pendant la première année de thèse '1', d'où son appellation. L'image infrarouge correspondait à un champ de vue de taille $270 \times 50 \mu\text{m}^2$. Il a été nécessaire d'acquérir dix images de fluorescence et 8 images Raman pour cartographier cette même région. Au cours de cette session d'acquisition, le laser Raman a endommagé la coupe. De ce fait, les images Raman n'ont pas été prises en compte dans cette collection d'images et seulement trois images de fluorescence ont été considérées pour les développements méthodologiques. La région spectrale commune des images infrarouge et de fluorescence couvre une région dans le faisceau vasculaire qui contient trois types cellulaires : parenchyme, phloème et sclérenchyme (Fig. B11).

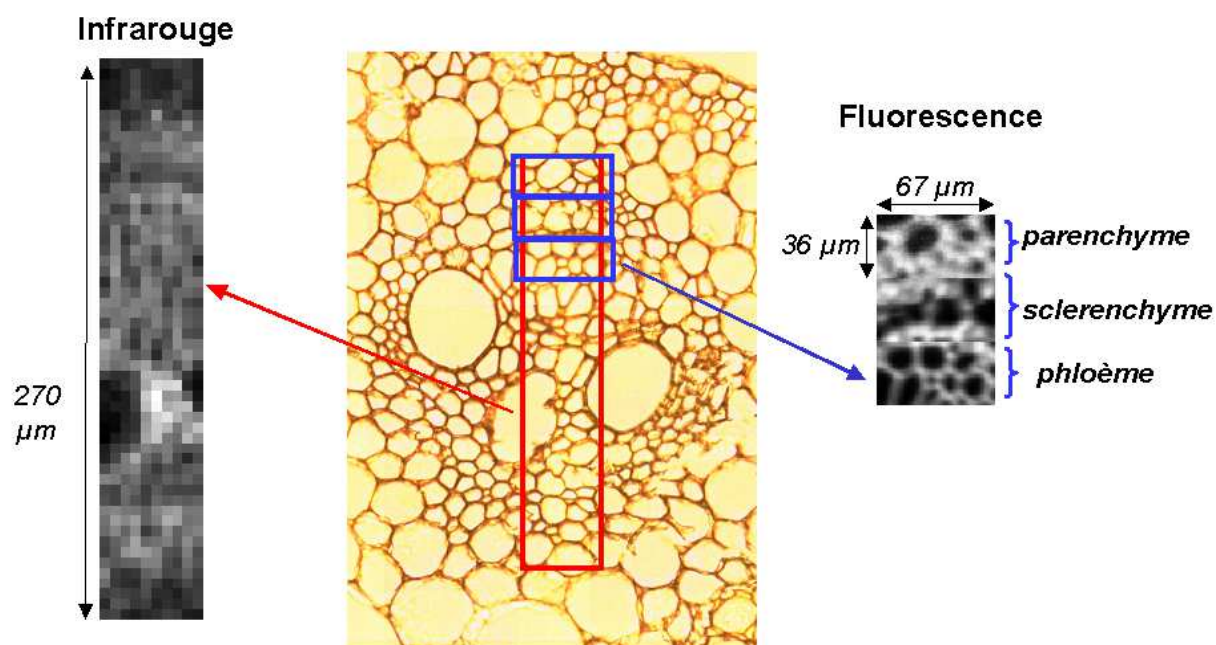


Figure B11 Acquisition conjointe des images de la collection 4f11mb-1 : Milieu : l'image en fond clair du faisceau vasculaire près de l'épiderme. Gauche: image spectrale infrarouge ayant un champ de vue de $270 \mu\text{m} \times 50 \mu\text{m}$. Les intensités de niveau de gris ont été calculées pour chaque pixel comme étant la somme des intensités spectrales de la région des sucres entre 1200 et 950 cm^{-1} . A droite : trois images de fluorescence adjacentes contenant trois types cellulaires : parenchyme, sclérenchyme et phloème. Les intensités de niveaux de gris ont été calculées pour chaque pixel comme la somme des intensités spectrales entre 360 - 540 nm .

3.2 Collection des images hyperspectrales 4f11mb-2

Suite aux dommages causés par la spectroscopie Raman sur l'échantillon 4f11mb en première année, un échantillon équivalent a été cartographié la deuxième année de thèse. Sa collection d'images associée est nommée 4f11mb-2.

Cette collection contient principalement trois types cellulaires: sclérenchyme, phloème et xylème. Une région de $210 \times 60 \mu\text{m}^2$ a été cartographiée en microspectroscopie infrarouge au milieu du faisceau vasculaire (Fig. B12-c). Quatre images adjacentes de fluorescence ont été considérées : trois ayant une taille de $47 \times 51 \mu\text{m}^2$ et une ayant une taille de $24 \times 51 \mu\text{m}^2$ (Fig. B12-d). Suite à un problème de perte de faisceau, la quatrième image de fluorescence était plus petite en taille que les trois autres images. Cinq images Raman ont été nécessaires pour cartographier la même région cartographiée qu'en infrarouge et en fluorescence. Les images Raman sont de taille 41×49 , 42×54 , 15×54 , 27×35 et $24 \times 43 \mu\text{m}^2$ (Fig. B12-e). La résolution des images de fluorescence et Raman permet de distinguer la structure des parois cellulaires.

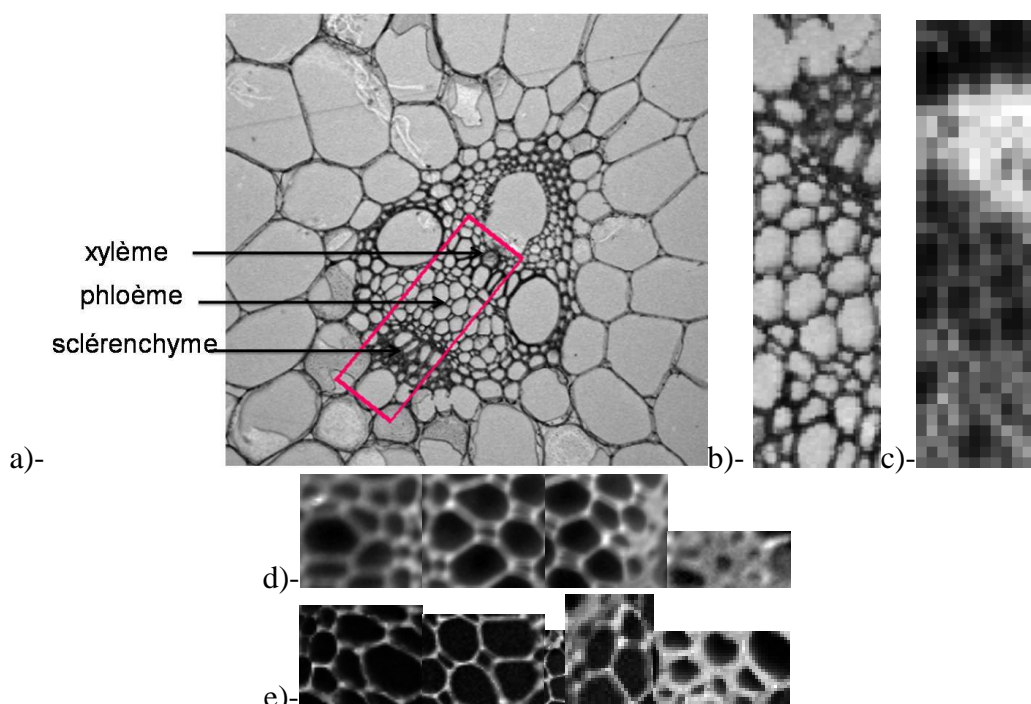


Figure B12 a)-Image en fond clair d'un faisceau vasculaire proche de l'épiderme dans une section transversale d'un entrenœud de tige de maïs. La région analysée est délimitée par le rectangle. **b)-** La région choisie contient principalement du xylème, du phloème et des cellules du sclérenchyme **c)-** Image spectrale infrarouge. Les intensités de niveaux de gris ont été calculées pour chaque pixel comme étant la somme des intensités spectrales de la région entre 1800 et 800 cm^{-1} . **d)-** Images de fluorescence. Les intensités de niveaux de gris ont été calculées pour chaque pixel comme la somme des intensités spectrales entre 360 - 540 nm . **e)-** Images Raman. Les intensités de niveaux de gris ont été calculées pour chaque pixel comme la somme des intensités spectrales entre 1800 - 800 cm^{-1} .

III Pré-traitement

En imagerie hyperspectrale, les spectres sont acquis par un balayage de points adjacents pour former une matrice à trois entrées avec deux dimensions spatiales et une dimension spectrale. Les spectres contiennent à la fois des informations chimiques et physiques. Dans le cas des sections de tige de maïs, des spectres peuvent être acquis à l'intérieur des cellules dans les régions où il n'y a pas de matière. Lorsque les spectres sont acquis sur le bord d'une paroi, l'intensité globale du spectre est intermédiaire.

D'un point de vue "image", ces variations d'intensités dépendent de la morphologie de l'échantillon. De plus, des variations peuvent être liées à la variation d'épaisseur de l'échantillon et dans notre cas des structures cellulaires. C'est pour ces raisons qu'il est nécessaire d'utiliser des méthodes de pré-traitement avant toute analyse chimiométrique, afin de permettre l'étude de la quantité relative de composés chimiques.

1 Déformations spectrales et spatiales des données

1.1 Exemples de spectres

La figure B13 montre des spectres bruts acquis pour les trois spectroscopies.

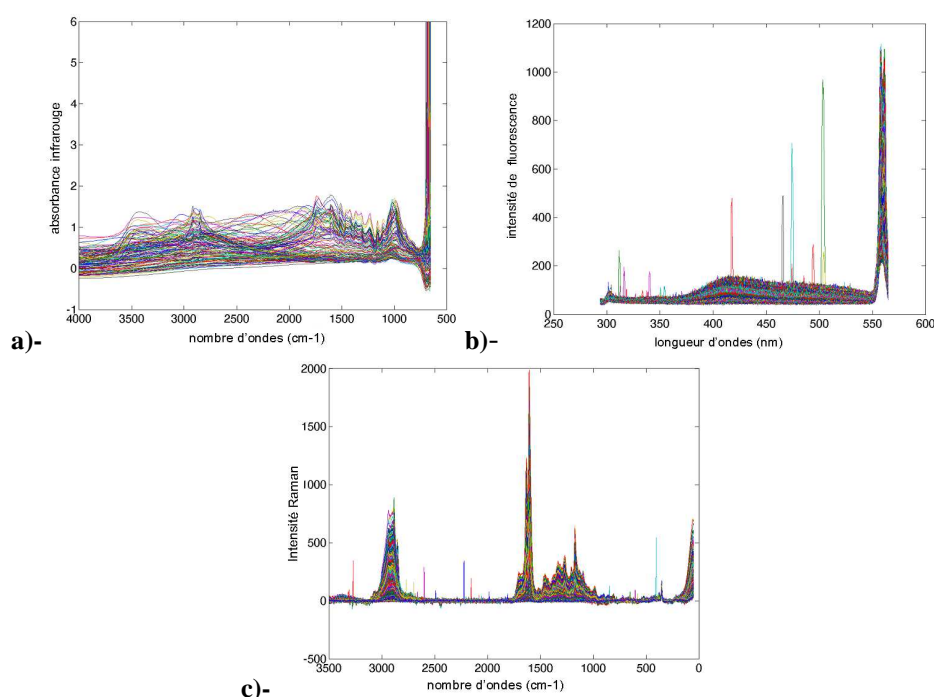


Figure B13 Exemples de collections de spectres : a) -infrarouge b)- fluorescence c)-Raman.

En fonction des spectroscopies, différentes déformations spectrales sont observées :

- des variations de lignes de base en infrarouge.

- la présence de "spikes" en fluorescence et Raman. Les spikes sont des pics intenses et étroits causés par la présence de rayons cosmiques dans l'atmosphère qui sont détectés par les appareils.

- des signaux plus ou moins bruités

L'objectif du pré-traitement est de se débarrasser de ces déformations. De plus, toute la région spectrale n'est pas exploitable ou intéressante pour analyser les parois cellulaires du maïs. En infrarouge et Raman, seule la région entre 1800 et 800 cm^{-1} a été retenue. En fluorescence, le pic intense à 560 nm correspond à la première harmonique de la bande de Rayleigh et la région avant 350 nm ne contient pas de signal. Les régions spectrales avant 365 nm et après 540 nm sont donc éliminées de l'analyse.

1.2 Variation spatiale

Les coupes de tiges de maïs ont une structure particulière. Les cellules ayant été vidées de leur contenu lors de la préparation des échantillons, un réseau de paroi est observé. Dans l'image spectrale résultante, on s'attend donc à observer des régions sans signal à l'intérieur des cellules et des spectres avec des intensités variables sur les parois. La figure B14-b montre la somme des intensités spectrales après élimination du bruit et des spikes (voir paragraphe 3.1) dans la région 365-540 nm pour une image de fluorescence.

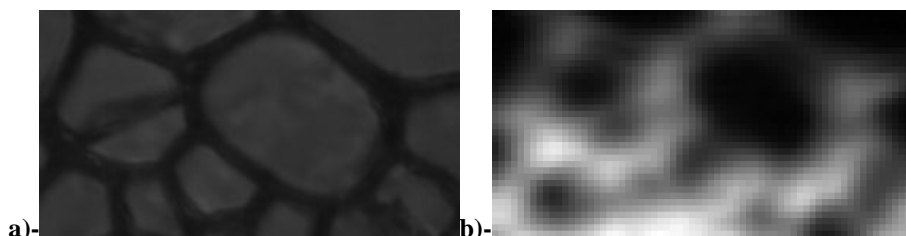


Figure B14 images de fluorescence : a)- Image visible en fond clair de cellule de parenchyme proche faisceau. b)- Image de la somme des intensités spectrales entre 365 et 540 nm après élimination des spikes et du bruit.

Les intensités contiennent plusieurs types d'informations. Elles représentent la structure cellulaire de l'échantillon. Les spectres sont caractérisés par des intensités plus importantes dans les zones de jonctions cellulaires et des intensités plus faibles sur le bord des parois ou pour certaines parois (Fig. B15).

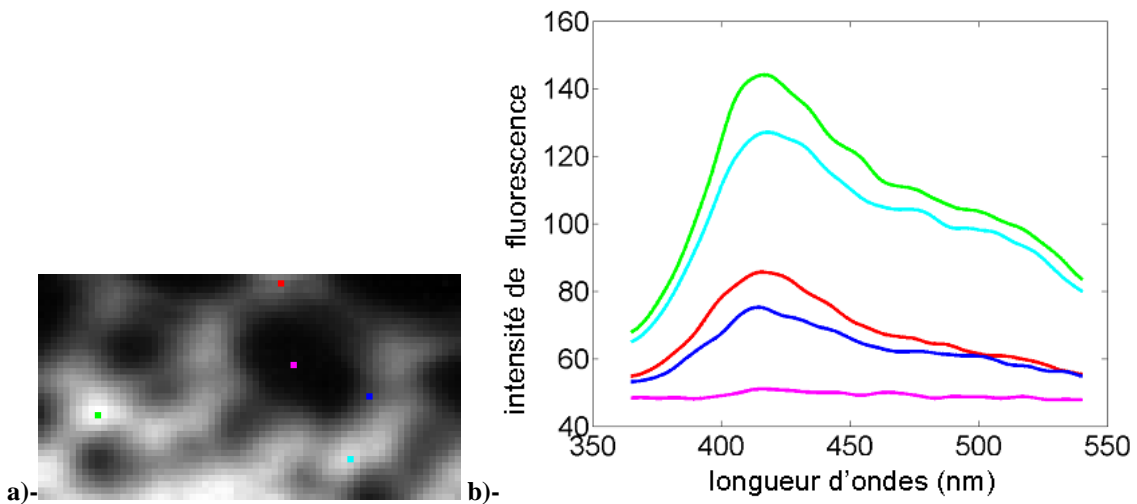


Figure B15: Fluorescence : exemples de spectres. Rose : intérieur de cellule, bleu : bord de paroi, rouge : zone de jonction en haut de l'image, cyan : zone de jonction en bas de l'image, vert : zone de jonction à gauche de l'image.

Des variations d'intensités liées à la position dans l'image apparaissent également : les spectres dans la partie haute de l'image sont moins intenses que ceux dans la partie basse de l'image. Ces variations apparaissent clairement dans la représentation surfacique des intensités de l'image exemple présentée dans la figure B14-b (Fig. B16). Ces variations lentes sont appelées "inhomogénéités d'intensité" ou "shading" en traitement d'images [Tomazevic *et al.*, 2002 ; Vovk *et al.*, 2007]. Elles peuvent être dues à du photoblanchiment en fluorescence ou à des hétérogénéités de coupes conduisant à une acquisition spectrale à différents plans focaux par rapport à la surface de la coupe.

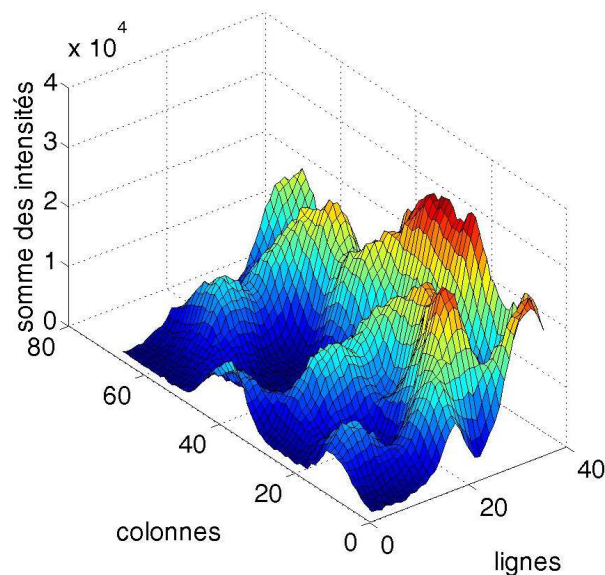


Figure B16 Fluorescence : représentation surfacique des intensités de l'image de la figure 14b. Les intensités les plus faibles correspondent aux pixels en haut de l'image.

2. Stratégie

De manière classique, en spectroscopie, les spectres sont prétraités individuellement en donnant à chacun la même intensité globale : par exemple en divisant chaque spectre par la somme des intensités totale ou la somme des intensités sous un pic ou bien par leur norme [Bertrand et Dufour, 2000]. Dans notre cas, cela reviendrait à donner la même importance aux spectres de l'intérieur des cellules qu'aux spectres de parois et donc à introduire des spectres de bruit dans l'analyse. Une idée naturelle est d'éliminer les spectres correspondant à une absence de signal en définissant un seuil d'intensité au-dessus duquel un spectre serait considéré comme valide [Gorzsás *et al.*, 2011]. Cette méthode présente plusieurs inconvénients :

- la valeur du seuil est difficile à définir lorsque des variations lentes d'intensité sont observées comme dans la figure B16. Doit-on appliquer le même seuil dans des régions de fort et faible contraste ? La détermination d'un seuil relatif n'est pas non plus évidente.

- des spectres de faible intensité mais contenant une information pertinente seront éliminés sur le bord des parois.

- la structure cellulaire spatiale liée à l'acquisition d'image hyperspectrale est perdue : notamment, la variation d'intensité qui décrit la présence d'une paroi dans l'image passe par un maximum au milieu de la paroi et diminue pour aller vers l'intérieur des cellules.

Ces différentes notions peuvent être visualisées sur un "profil" image. Dans la figure B17-a, deux lignes ont été tracées dans l'image de la somme des intensités (Fig. B14-b). Les intensités des pixels sur les deux lignes sont représentées comme des courbes (Fig. B17-b).

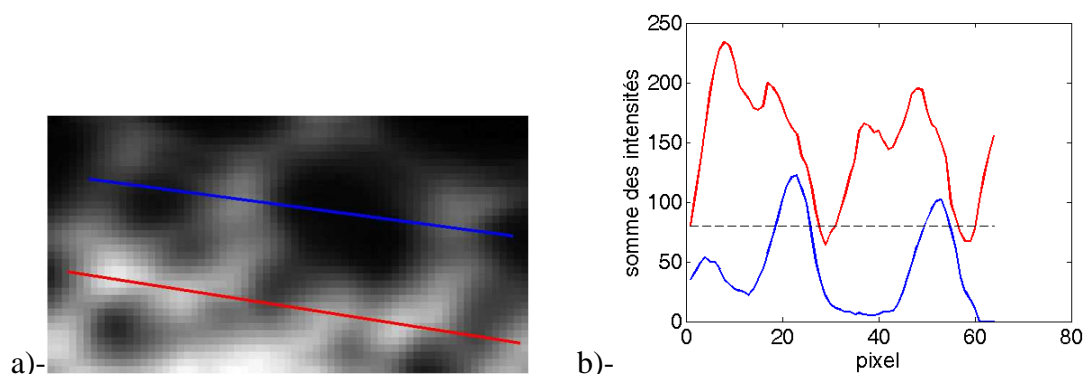


Figure. B17 Fluorescence : exemple de profil. Bleu : haut de l'image, rouge : bas de l'image. Pointillé noir : seuil d'intensité à 80.

Dans cette thèse, nous proposons de combiner un pré-traitement à la fois spectral et spatial dont l'objectif est de préserver la structure cellulaire observée tout en éliminant les

inhomogénéités d'intensité. Les corrections de ligne de base, l'élimination des spikes et du bruit ont été appliquées aux spectres considérés individuellement. La normalisation des intensités a été réalisée à partir d'une méthode tenant compte du voisinage spatial des spectres. La méthode a été mise au point avec l'hypothèse que les variations d'intensité spectrale dans un petit voisinage pouvaient être considérées comme pertinentes. Ainsi, les plus grandes intensités dans les zones de jonctions cellulaires ont été considérées comme une information réelle et non pas comme un artefact de mesure.

Les sections suivantes sont dédiées à la description des méthodes de pré-traitement développées et à leur application aux images hyperspectrales infrarouge, fluorescence et Raman acquises pendant la thèse.

3 Méthodes de pré-traitement

3.1 Elimination des spikes

Classiquement, les spikes sont identifiés par leur intensité et éliminés manuellement après l'acquisition des images spectrales. Dans cette thèse, une méthode automatique d'élimination des spikes a été mise au point. Cette méthode utilise le fait que les spikes sont des pics étroits et intenses. La transformée en chapeau haut de forme est une transformation classique en traitement d'images dont l'objectif est de sélectionner les objets plus petits qu'un masque d'une taille et d'une forme données. Ce masque doit être plus intense que le fond de l'image [Soille, 1999]. La transformation s'applique également à des données à 1 dimension. Le principe consiste à identifier des pics plus étroits qu'une fenêtre d'une taille donnée et d'une intensité supérieure à un seuil donné. La figure. B18 montre les étapes de la transformation pour une courbe exemple.

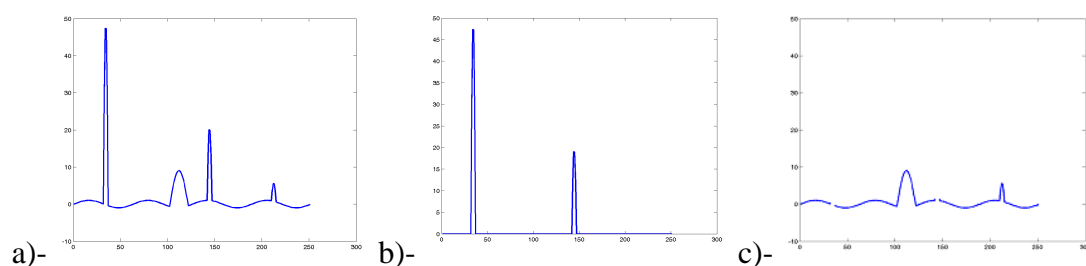


Figure B18 Fonctionnement de la transformation de top-hat à une dimension. Gauche : courbe avec 3 spikes d'intensité variable et un pic plus large. Milieu : spikes détectés pour une fenêtre de taille 5 points et une intensité de 5. Droite : courbe après soustraction des spikes.

3.2 Filtrage du bruit

Le bruit a été éliminé par filtrage par la transformée de Fourier selon la méthode décrite dans [Gonzalez *et al.*, 2009]. Pour chaque spectre, la transformée de Fourier est calculée et un filtre gaussien est appliqué pour ne retenir que les composantes de plus basse fréquence du signal. La transformée de Fourier inverse permet d'obtenir le spectre lissé.

3.3 Normalisation spatiale et spectrale

Classiquement, la normalisation des spectres est effectuée spectre par spectre selon l'équation :

$$\text{sn}(\lambda) = s(\lambda) / Fn \quad \text{Equation B1}$$

où s est le spectre à normaliser, Fn est le facteur de normalisation, sn le spectre normalisé et λ une longueur d'onde considérée. Le facteur de normalisation peut être calculé comme étant l'intensité à une longueur donnée, l'aire sous un pic, la somme ou la norme des intensités du spectre complet [Bertrand et Dufour, 2000].

Pour prendre en compte à la fois l'information spatiale et spectrale, nous proposons de calculer le facteur de normalisation en tenant compte des pixels voisins. Ce facteur sera appliqué par la suite à chaque spectre de l'image hyperspectrale.

Notons $\mathbf{Fn}(V(x,y))$ ce facteur de normalisation où $V(x,y)$ représente le voisinage du spectre localisé à la coordonnée spatiale (x,y) . Nous avons choisi de baser le calcul du facteur de normalisation sur la somme des intensités du spectre à normaliser. Considérons l'image de la somme des intensités dont une représentation est donnée sur la figure B14-b. Cette image représente également l'image des facteurs de normalisation correspondant à l'équation B1. Nous proposons ici de modifier les valeurs de ce facteur de normalisation en transformant l'image par des opérations de traitement d'images.

3.3.1 Prise en compte des régions sans signal

Afin de ne pas normaliser les spectres de l'intérieur des cellules et les spectres des parois de la même manière, il est nécessaire que le facteur de normalisation soit très supérieur à la somme des intensités spectrales observée pour ces pixels. L'opération géodésique "bouchage de trous" (fillholes) du traitement d'image [Soille, 1999] permet d'affecter aux pixels d'un "trou" la plus petite valeur observée sur son contour, un "trou" étant défini comme un minimum local entouré par des valeurs plus élevées. La figure B19 montre le résultat de l'opération

"bouchage des trous" sur un des profils de la figure B14-b. Au niveau de l'intérieur des cellules l'intensité du facteur de normalisation a été rehaussée.

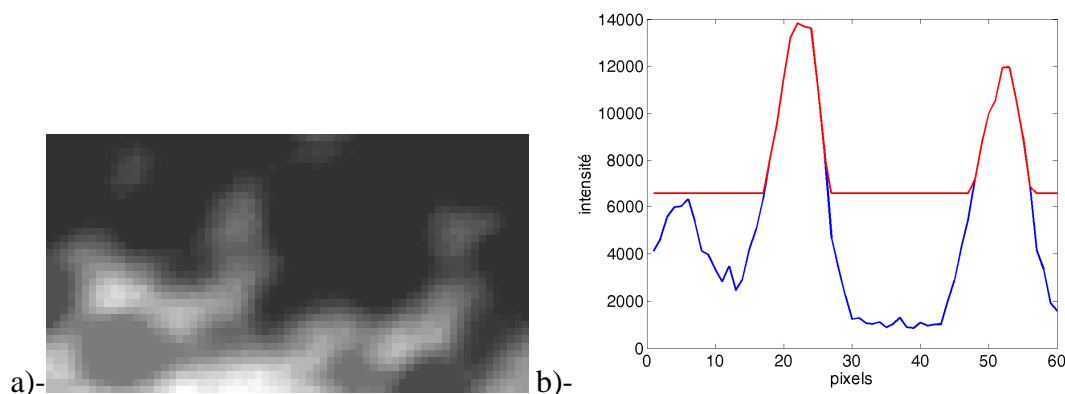


Figure B19 Transformation du facteur de normalisation par "bouchage des trous". a)- Image de la somme des intensités après "bouchage des trous." b)- Effet sur le profil bleu de la figure B14-b. Bleu : profil de départ, rouge : profil après "bouchage des trous".

3.3.2 Prise en compte des variations locales d'intensité

L'utilisation des facteurs de normalisation après bouchage des trous va permettre de préserver la différence d'intensité entre l'intérieur des cellules et les parois mais ne permettra pas de conserver les différences au niveau des parois ou des jonctions cellulaires. En utilisant un lissage avec une moyenne mouvante de taille $\nu \times \nu$, ν étant la taille du voisinage considéré, les détails plus petits que la fenêtre sont supprimés (Fig. B20). La taille du voisinage est choisie en fonction de la structure observée et de la méthode d'acquisition des données. Dans l'exemple, 11 pixels correspondent à un voisinage de $11 \times 11 \mu\text{m}^2$, taille supérieure à l'épaisseur des parois végétales. Cette taille permet de supprimer les variations au niveau des jonctions cellulaires.

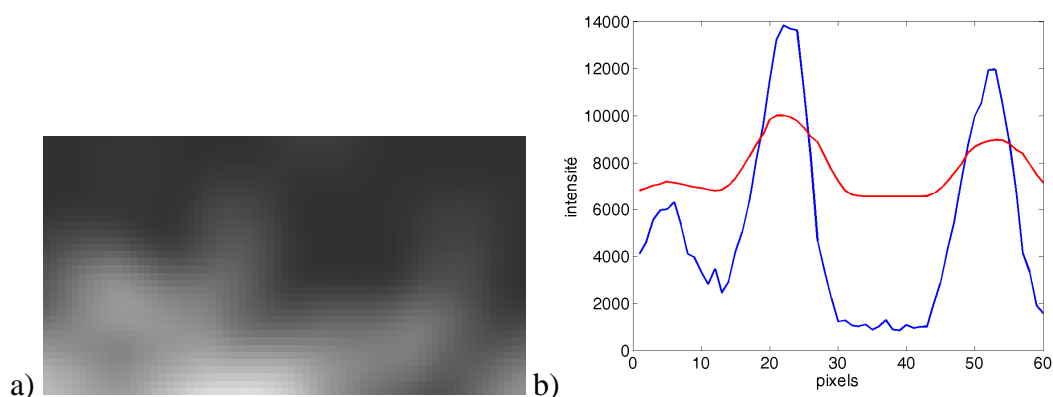


Figure B20 Transformation du facteur de normalisation par lissage a) image de la somme des intensités après lissage avec une moyenne mouvante de taille 11 x 11 pixels. b) effet sur le profil bleu de la Figure B14-b. Bleu : profil de départ, rouge : profil après lissage.

3.3.3 Normalisation

L'image de la figure B21-a représente l'ensemble des facteurs de normalisation $\mathbf{Fn}(V(x,y))$ dépendant d'un voisinage du spectre.

L'équation B1 écrite pour l'ensemble des spectres de l'image devient alors :

$$\mathbf{Sn}(x,y,\lambda) = \mathbf{S}(x,y,\lambda) / \mathbf{Fn}(V(x,y)) \quad \text{Equation B2}$$

où $\mathbf{S}(x,y,\lambda)$ représente l'intensité du spectre à normaliser localisé à la coordonnée spatiale (x,y) pour la longueur d'onde λ et $\mathbf{Sn}(x,y,\lambda)$ l'intensité normalisée. Le résultat pour l'image exemple est donné figure B21.

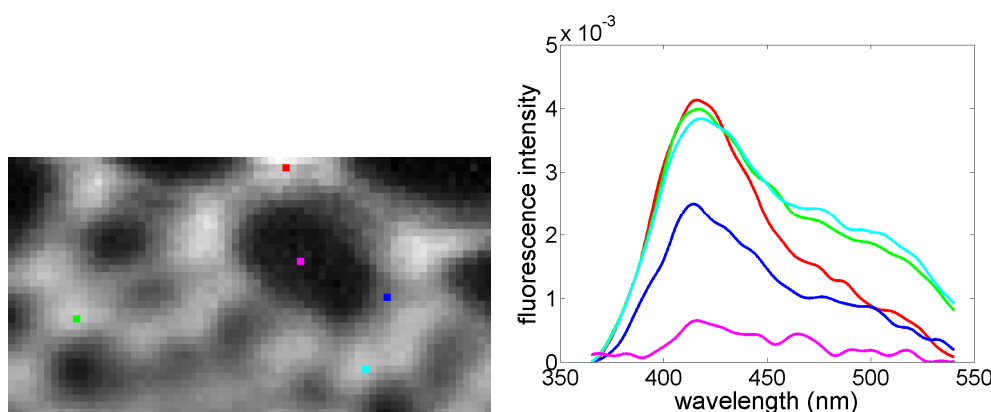


Figure B21 a) image des sommes des intensités après normalisation. b) exemples de spectres. Rose : intérieur de cellule, bleu : bord de paroi, rouge : zone de jonction en haut de l'image, cyan : zone de jonction en bas de l'image, vert : zone de jonction à gauche de l'image.

Les spectres des zones de jonctions vert et cyan sont maintenant superposés et la différence avec le spectre rouge de la zone de jonction en haut de l'image apparaît clairement dans la région entre 460 et 540 nm. La moindre intensité sur le bord des parois ainsi qu'à l'intérieur de la cellule est préservée.

4 Applications aux images hyperspectrales

Pour chaque spectroscopie, le pré-traitement a été réalisé en deux étapes. La première étape se composait de corrections spécifiques à chaque spectroscopie. La seconde est une étape de normalisation qui a pris en compte à la fois l'information spectrale et spatiale.

4.1 Pré-traitement des spectres infrarouge

La région $1800-800 \text{ cm}^{-1}$ a été sélectionnée. Les spectres infrarouge des collections 4f11mb-1 et 4f11mb-2 ont été pré-traités de deux manières différentes. Concernant la collection 4f11mb-1, les spectres infrarouge ont été corrigés en utilisant une dérivée seconde de

Savitzky-Golay en utilisant un polynôme de degré 3 et une fenêtre de 9 points. Les spectres ont été normalisés en utilisant la norme vectorielle unitaire (Fig. B22).

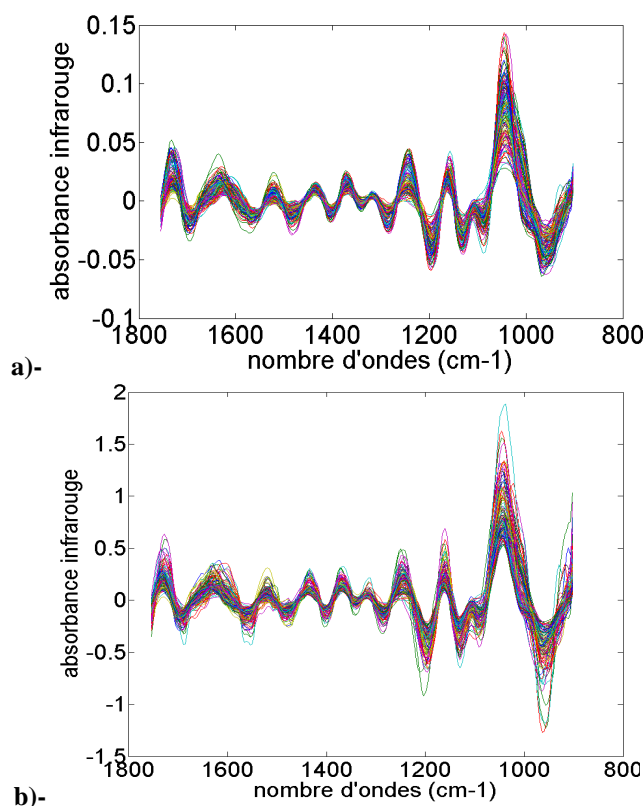


Figure B22 Etapes de pré-traitement des spectres infrarouge de la collection 4f11mb-1: a)-sélection des longueurs d'ondes + dérivée de Savitzky Golay b)- normalisation en utilisant la norme unitaire

Pour la collection 4f11mb-2, la ligne de base a été corrigée en utilisant des segments linéaires liant des points de référence choisis pour la collection spectrale entière. Sur la figure 23, les étapes de pré-traitement des spectres infrarouge de la collection 4f11mb-2 sont présentées. Pour corriger la ligne de base, les points de référence ont été sélectionnés à partir du spectre moyen à 800, 1184, 1280, 1384, 1480, 1528 et 1800 cm^{-1} .

Pour chaque spectre, les segments linéaires ont été ajustés en déplaçant les points sélectionnés vers le plus proche point minimal. En dépit de l'ajustement de chaque segment, certaines valeurs négatives ont été observées après la correction de la ligne de base, principalement dans la région 1800-1500 cm^{-1} (Fig. B23-b).

Les spectres ayant une somme des intensités négatives supérieure à un seuil de 0,2 dans le 1800-1500 cm^{-1} ont été éliminés. Les spectres éliminés présentaient des déformations pouvant être liés à la diffusion de mie. L'informations associée à ces spectres est non pertinente. Pour le reste de la collection, la valeur minimale a été mise à 0 par une correction de type "offset"

(Fig. B23-c). Finalement, les spectres ont été normalisés, en utilisant la normalisation spatiale spectrale avec un voisinage de taille 5×5 pixels (Fig. 23-d).

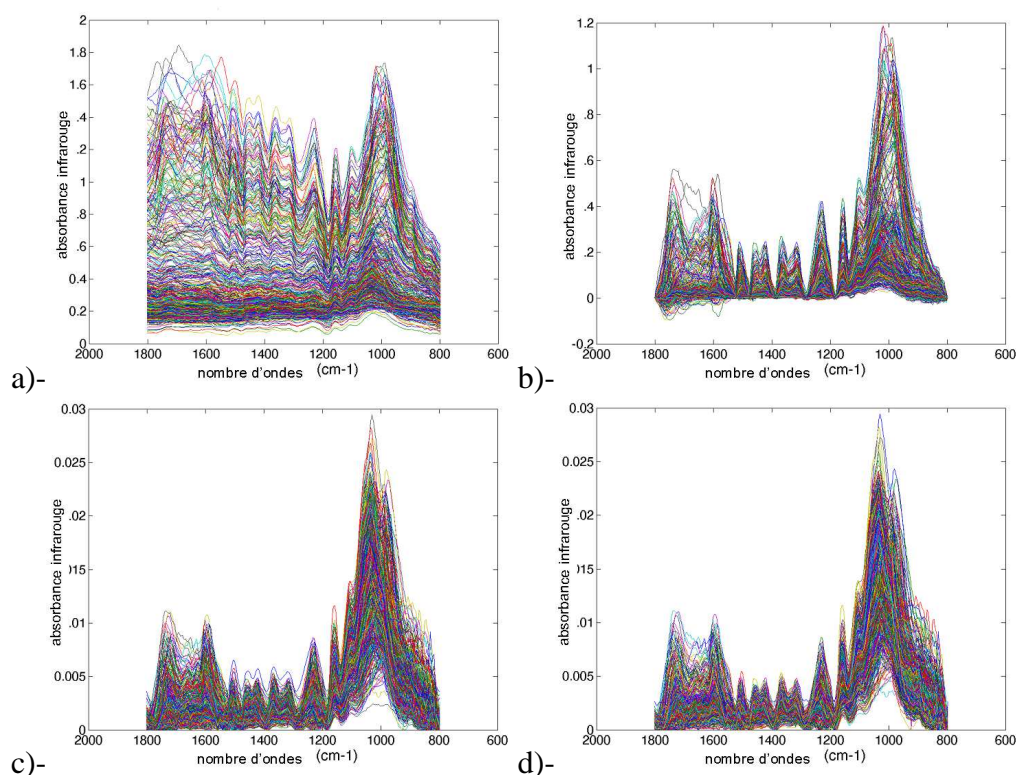


Figure B23 Etapes de pré-traitement des spectres infrarouge : a)- spectres de la région ($1800-800 \text{ cm}^{-1}$) b)- correction de la ligne de base c)- normalisation des spectres d)- élimination des spectres ayant des valeurs négatives.

4.2 Pré-traitement des spectres de fluorescence

Les pics fin appelés “spikes” ont été éliminés en utilisant la procédure décrite dans le paragraphe 3.1 (Fig. B24-b). La taille du chapeau haut de forme a été fixée à 21 points = 5.57 nm et le seuil d’intensité a été fixé à 40. En fluorescence cette intensité correspond à la valeur du bruit de fond. Un filtre gaussien à transformée de Fourier de taille 21 a été appliqué (Fig. B24-c).

La gamme spectrale comprise entre 365 et 540 nm a été choisie comme étant la région où les lignines et les acides phénoliques fluorescent (Fig. B24-d).

La normalisation spectrale et spatiale a été appliquée en considérant un voisinage de taille 11×11 pixels (Fig. 24-e).

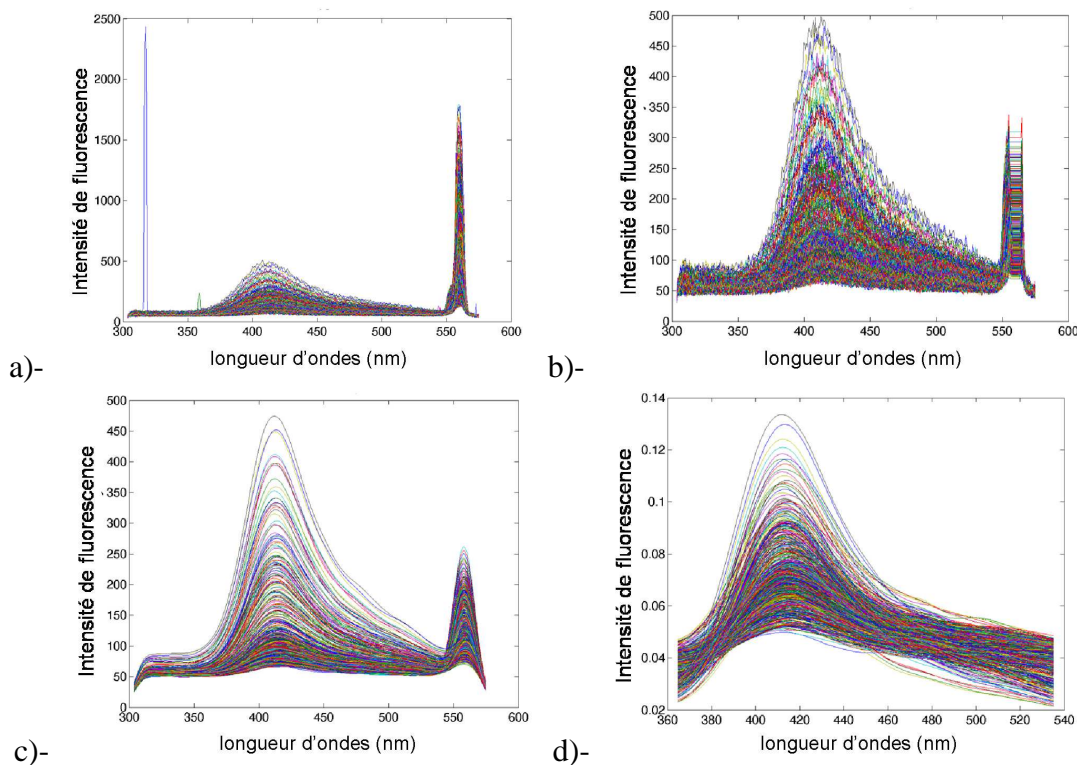


Figure B24 Etapes de pré-traitement des spectres de fluorescence : a) spectres bruts b) élimination des spikes c) filtrage en utilisant la transformée de Fourier d) Sélection de la plage de longueur d'ondes (365-540 nm) + normalisation spatiale/spectrale.

4.3 Pré-traitement des spectres Raman

Des spikes ont été également observés dans les spectres Raman (Fig. B25-a). Dans le cas des spectres Raman, les pics sont souvent fins et le pic des phénols à 1600 cm^{-1} est aussi intense qu'un spike. Les spikes ont été éliminés en appliquant la procédure décrite en 3.1 avec deux passages et des paramètres différents. D'abord, les spikes les plus intenses de largeur inférieure à 11 cm^{-1} et d'intensité supérieure à 200 en intensité ont été éliminés, ensuite, les spikes moins intenses de largeur inférieure à 7 cm^{-1} et de d'intensité supérieure à 50 ont été également éliminés (Fig. B25-b et c). Le bruit a été supprimé par transformée de Fourier en utilisant un filtre gaussien de taille 125 (Fig. B25-d). Comme pour le moyen infrarouge, le domaine spectral entre 800 et 1800 cm^{-1} a été sélectionné. La normalisation spatiale et spectrale a été appliquée en considérant un voisinage de taille 11×11 pixels (Fig. B25- e). Certains spectres Raman présentent des valeurs négatives. Dans cette étude, nous avons considéré tous les spectres de la collection et nous n'avons pas éliminés des spectres en Raman.

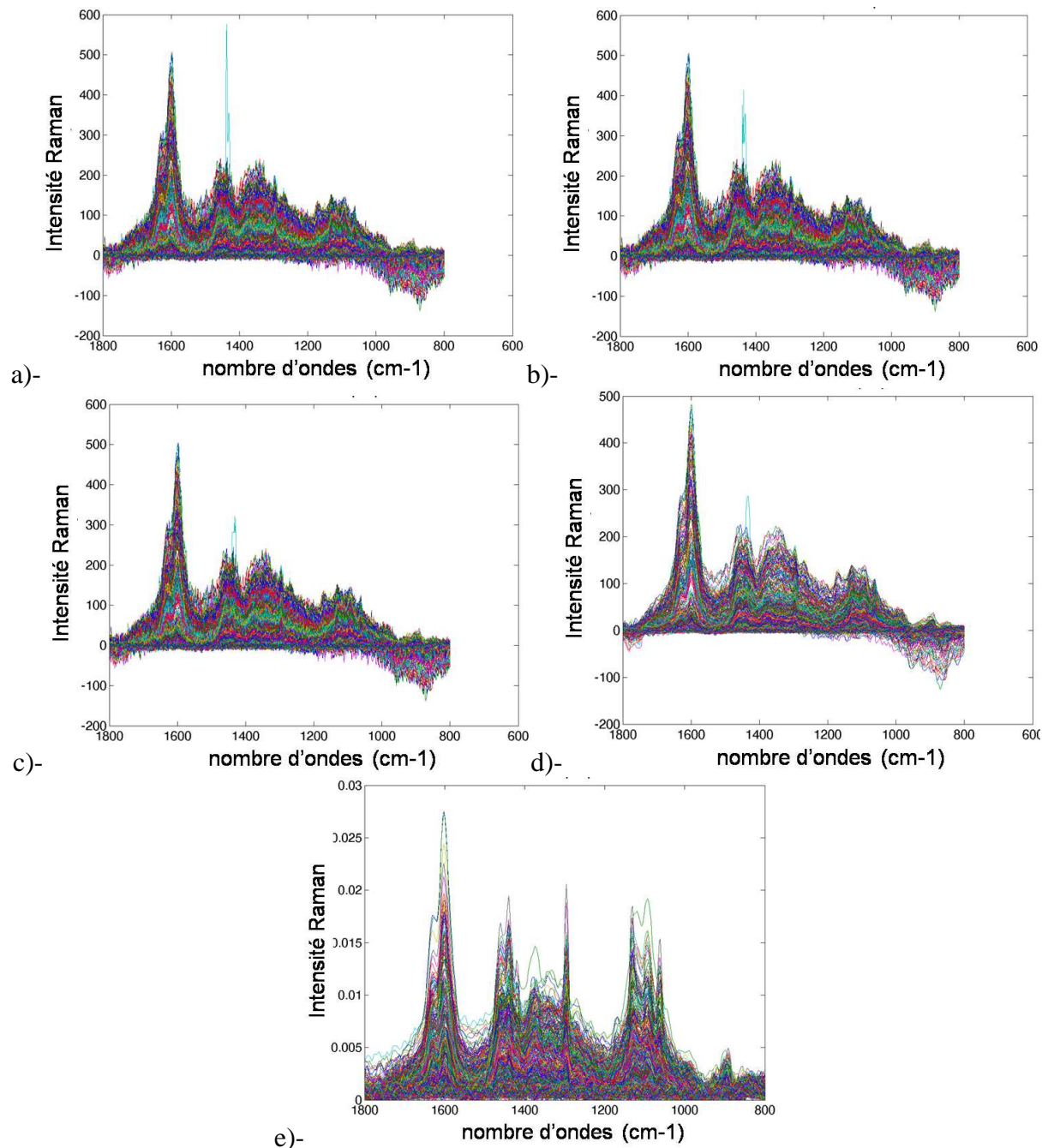


Figure B25 Etapes de pré-traitement des spectres Raman : a)- sélection la plage de longueur d'ondes 1800-800 cm^{-1} b)- 1^{ère} élimination des spikes c)- 2^{ème} élimination des spikes d)- filtrage en utilisant la transformée de Fourier e)-normalisation spatiale/spectrale.

IV Analyse individuelle de la collection 4f11mb-2

Comme les deux collections d'images 4f11mb-1 et 4f11mb-2 sont équivalentes, et que la deuxième comprend des images issues des trois spectroscopies, dans cette section, nous présenterons la collection 4f11mb-2.

1 Exploration des données

Afin d'explorer la variabilité chimique au sein de la collection 4f11mb-2, des spectres pré-traités de chaque type cellulaire ont été sélectionnés. Les spectres ont été sélectionnés à partir des images "somme des intensités". Les spectres des trois types cellulaires, sclérenchyme, phloème et xylème, ont été examinés. Dans la figure B26, un exemple des profils tracés sur des images Raman est présenté.

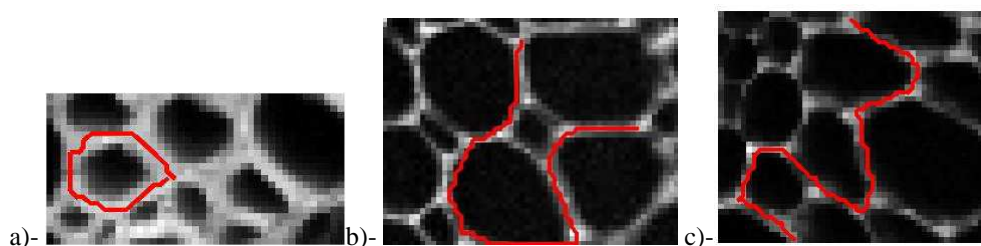


Figure B26 Spectres sélectionnés dans le type cellulaire a)-sclérenchyme b)-phloème c)-xylème. Les images en niveaux de gris sont des images Raman calculées à partir de la somme des intensités des spectres dans la région $1800-800\text{ cm}^{-1}$.

Les spectres sont présentés dans la figure B27. L'interprétation des spectres moyens infrarouge a été faite avec précaution. En effet, la majorité des composants chimiques donnent naissance à des bandes qui se superposent. De plus, plusieurs groupements chimiques sont presque identiques dans la majorité des polymères appartenant aux parois cellulaires. Par conséquent, une seule bande ne permet pas d'interpréter la présence d'un composant chimique. Des interprétations pertinentes doivent tenir compte de plusieurs bandes, qui indiquent tous à la fois la présence d'un même composant. Dans notre interprétation, nous avons utilisé la littérature pour attribuer nos bandes.

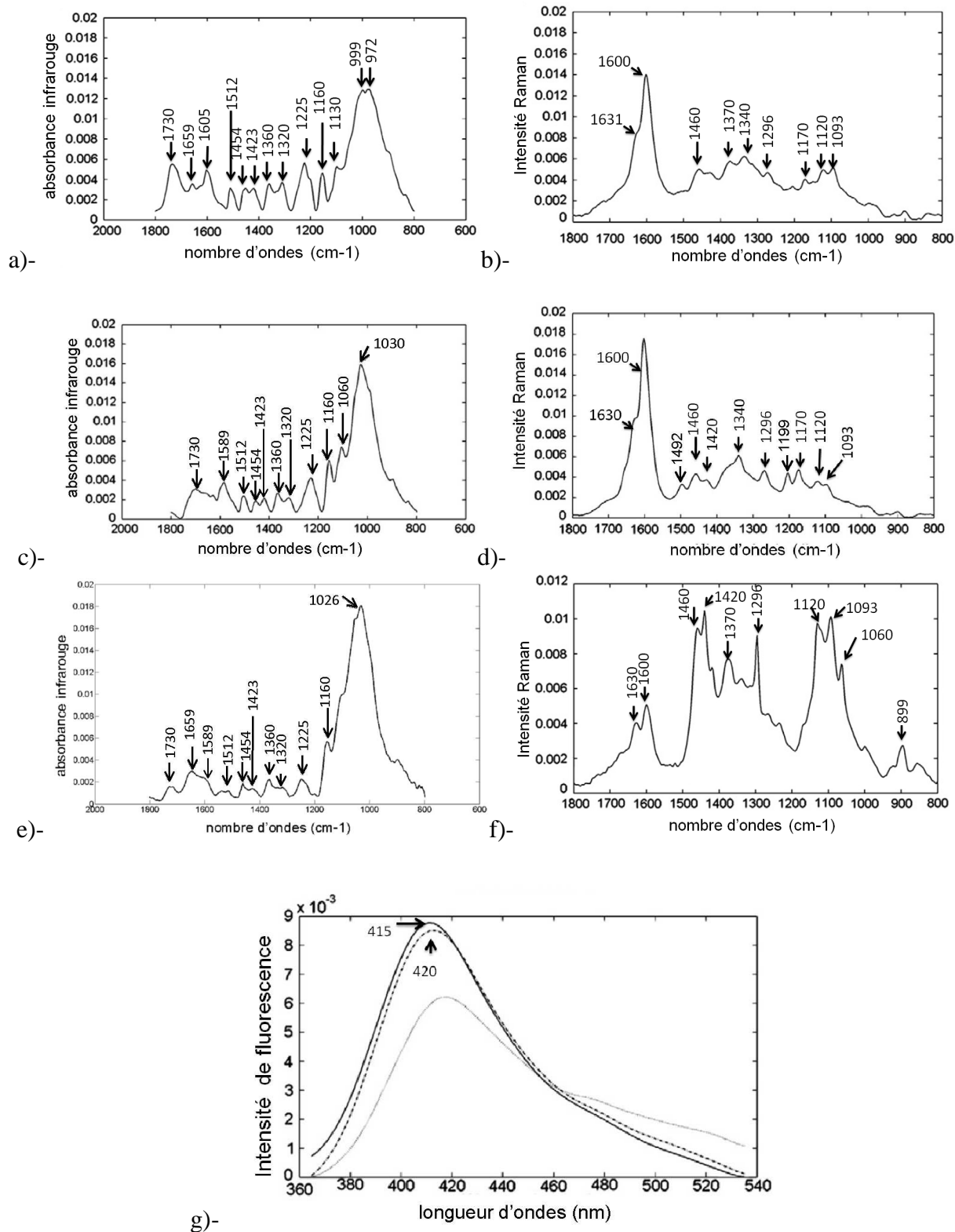


Figure B27 Spectres moyens en fonction des types cellulaires: spectres moyen infrarouge des cellules a)- du sclérenchyme c)- du xylème e)- du phloème, spectres moyen Raman des cellules b)- du sclérenchyme d)- du xylème f)- du phloème g)- spectres moyens de fluorescence des cellules du sclérenchyme (ligne pointillée), des cellules du phloème (ligne continue) et des cellules du xylème (ligne en tirets).

Un pic large et intense est observé dans la région 1200-800 cm^{-1} dans chacun des spectres infrarouge (Fig. B27-a, c et e). Ce pic est observé dans le spectre du sclérenchyme, du xylème et du phloème, avec des maxima respectifs à (999 - 972) cm^{-1} , 1030 cm^{-1} et 1026 cm^{-1} .

Le pic à (999 - 972) cm^{-1} est non attribué dans la littérature.

Les pics à 1030 cm^{-1} et 1026 cm^{-1} peuvent correspondre à une signature de cellulose ou de xylanes.

L'allure générale des spectres des cellules du sclérenchyme et du phloème est assez similaire dans la région 1800 - 1200 cm^{-1} .

Plusieurs pics sont observés dans les cellules du sclérenchyme et du xylème :

- 1160, 1320 cm^{-1} attribués à la cellulose et aux xylanes.
- 1423, 1454, 1512, 1589 cm^{-1} attribués aux lignines.
- 1225 cm^{-1} attribué aux liaisons esters.

La bande à 1130 cm^{-1} apparaissant dans le spectre de sclérenchyme pourrait être attribuée aux lignines.

Dans la région 1600-1800 cm^{-1} , le spectre moyen du sclérenchyme diffère du spectre moyen du xylème.

Sur le spectre moyen du phloème, le pic des polysaccharides est le plus intense par rapport aux spectres moyens de sclérenchyme et du xylème.

Les spectres Raman du xylème et du sclérenchyme sont assez semblables (Fig. B29-b et d).

Les pics sont observés à:

- 1093-1120, 1296, 1340, 1370, 1460 cm^{-1} qui peuvent être attribués aux polysaccharides : cellulose et xylane.
- 1170, 1600-1631 cm^{-1} attribués à la structure aromatique de la lignine.

Le pic à 1199 cm^{-1} apparaissant seulement dans le spectre moyen du xylème est attribué à la lignine.

Le pic à 1060 et à 1420 cm^{-1} sont non attribués dans la littérature. Ils n'apparaissent que dans les cellules du phloème.

Le spectre moyen Raman du phloème diffère à la fois du xylème et du sclérenchyme (Fig. B27-b, d et f). Les pics du spectre du phloème sont beaucoup plus intenses que ceux du sclérenchyme et du xylème et ils sont attribués à des polysaccharides. Toutefois, les polysaccharides réagissent fortement dans le cas des cellules du phloème. Les résultats obtenus en infrarouge et en Raman ont clairement confirmés que les parois cellulaires du sclérenchyme et du xylème contenaient des lignines alors que les parois cellulaires du phloème sont enrichies en polysaccharides et xylanes.

Dans le cas du spectre moyen d'émission de fluorescence du phloème, le pic est observé à 415 nm alors que dans le cas du spectre du sclérenchyme, le maximum du pic est déplacé à 420 nm, de la fluorescence est également observée entre 440 nm et 540 nm (Fig. 27-g).

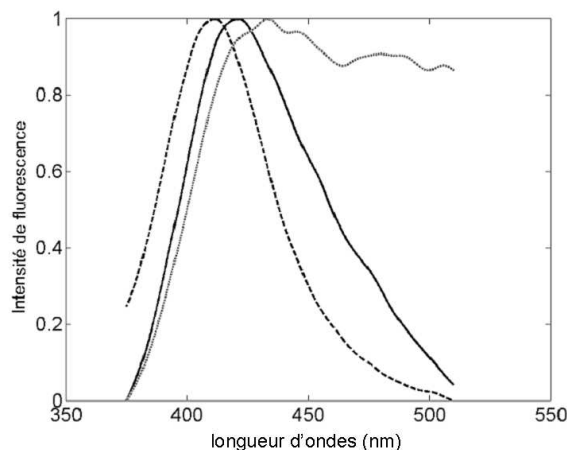


Figure B28 Spectres de référence de fluorescence de l'acide férulique (ligne continue), le para-coumarique (ligne en tirets) et de lignine (ligne en pointillés).

La figure B28 montre trois spectres acquis à l'aide du micro-spectrophotomètre de fluorescence de la ligne DISCO pour trois composés de référence: l'acide férulique, l'acide para-coumarique et la lignine. L'acide férulique et para-coumarique ont montré un pic centré autour de 420 et 410 nm, respectivement, tandis que la lignine a montré une fluorescence continue à partir de 420 nm. Cette interprétation confirme que les parois cellulaires du sclérenchyme contiennent de la lignine.

2 Analyse en composantes principales de la collection de données 4f11mb-2

Afin d'interpréter l'information fournie par les images hyperspectrales infrarouge, fluorescence et Raman de la collection de données 4f11mb-2. Trois composantes principales pour chaque spectroscopie ont été examinées.

2.1 Analyse en composante principale de l'image infrarouge

Le tableau B1 montre les pourcentages de variance obtenus pour chaque composante principale infrarouge. Les trois premières composantes infrarouge décrivent 87% de la variabilité totale (Tableau B1).

Tableau B1 Pourcentages de variance expliquée par les composantes principales de l'image hyperspectrale infrarouge de l'échantillon 4f11mb-2

composante 1	composante 2	composante 3	pourcentage cumulé
54.17	23.47	9.87	87.51

Le premier profil spectral infrarouge (Fig. B29-a) est un spectre d'intensité. Il décrit la variation d'intensité paroi/intérieur de la cellule. L'obtention de ce profil spectral est liée à la méthode de pré-traitement utilisée qui conserve les informations d'intensité de la structure cellulaire.

Dans l'image de la première composante infrarouge (Fig. B29-c), les pixels apparaissant en noir sont les pixels appartenant aux parois. Les pixels clairs sont les pixels appartenant à l'intérieur de cellules. Les pixels rouges sont les pixels éliminés de l'analyse.

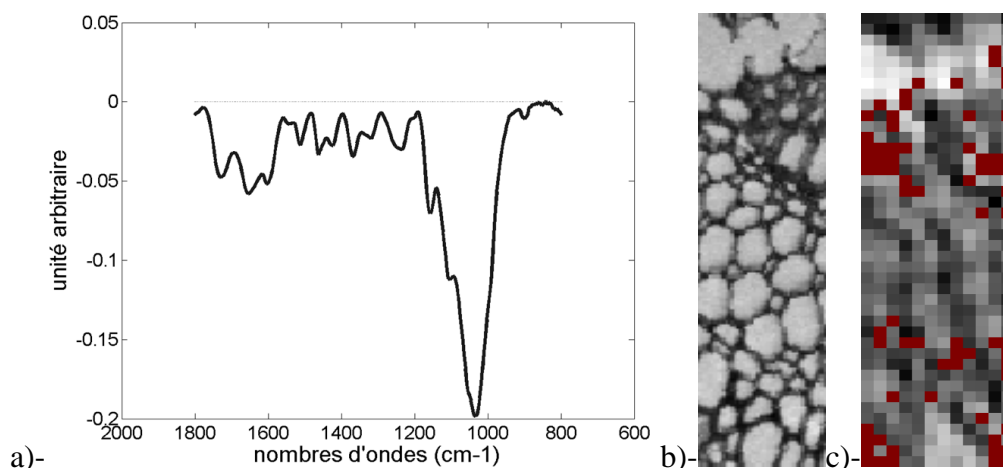


Figure B29 a)- profil spectral infrarouge 1 b)- image fond clair de la région analysée c)- image composante infrarouge 1.

Le profil spectral 2 (Fig. B30-a) révèle des bandes d'absorption marquées. Les plus importantes sont les suivantes :

- En négatif : 1068, 1110 cm^{-1} qui peuvent être attribuées à la cellulose et aux xyloxygènes.
- En positif : 1230, 1330, 1419, 1504, 1589 cm^{-1} sont des bandes qui peuvent être attribuées aux lignines et aux liaisons esters.

Les pixels clairs de la deuxième composante image révèlent les lignines et les esters (Fig. B30-c). Les pixels noirs révèlent la cellulose et les xyloxygènes. Avec cette interprétation, les cellules de sclérenchyme et de xylème sont riches en lignines et en esters. Les cellules du phloème contiennent de la cellulose et des xyloxygènes. Les cellules du xylème révèlent des pixels moins intenses, ils sont donc moins riches en lignines.

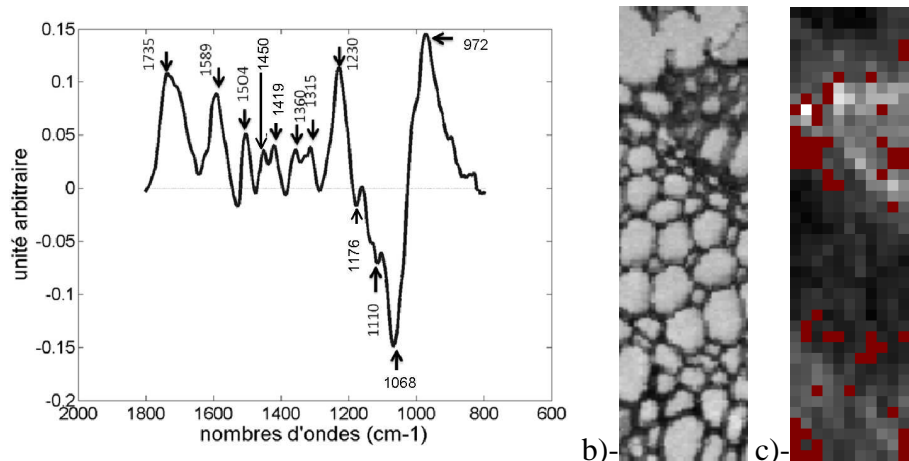


Figure B30 a)- profil spectral infrarouge 2 b)- image fond clair de la région analysée c)- image composante infrarouge 2.

Le profil spectral 3 (Fig. B31-a) possède principalement deux bandes d'absorption en négatif à 945 et 1072 cm^{-1} attribuées aux xylanes opposées à des bandes en positif à 1235, 1320, 1512, 1589 cm^{-1} attribuées aux lignines et aux liaisons esters. L'image composante 3 infrarouge est assez compliquée et ne permet pas de mettre en évidence les variations spatiales exceptés pour quelques pixels contrastés (Fig. B31-c).

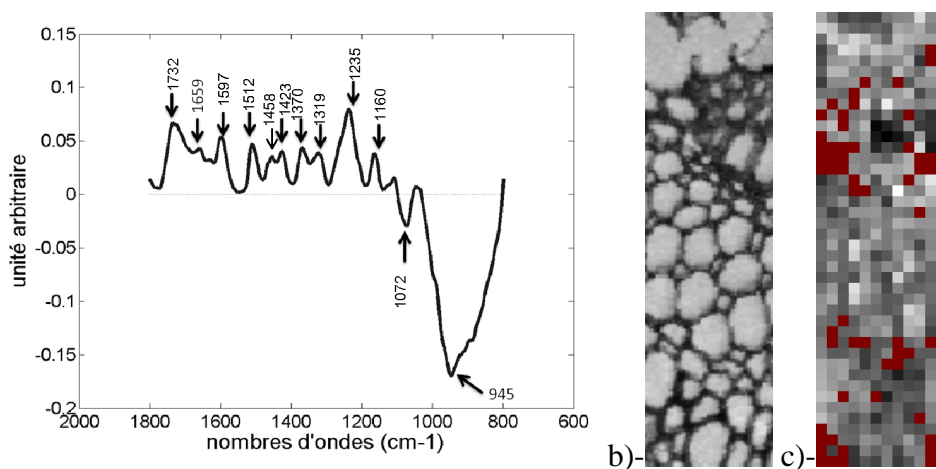


Figure B31 a)- profil spectral infrarouge 3 b)- image fond clair de la région analysée c)- image composante infrarouge 3.

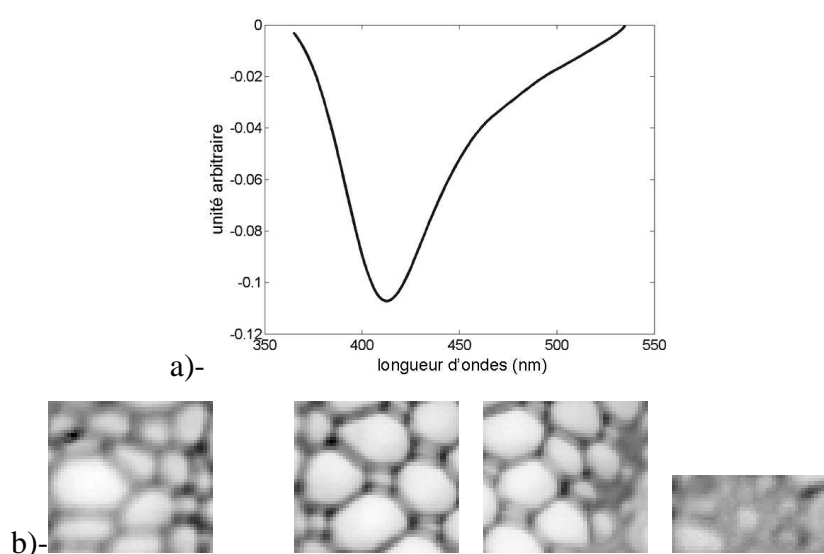
2.2 Analyse en composantes principales des images hyperspectrales de fluorescence

Les images hyperspectrales de fluorescence de la collection 4f11mb-2 ont été dépliées et concaténées dans un même tableau de données à deux entrées. Le tableau de données obtenu est soumis à l'analyse en composantes principales. Trois composantes décrivant 99.68% de la variabilité totale ont été examinées.

Tableau B 2 Pourcentages de variance expliquée par les composantes principales de fluorescence de la collection 4f11mb-2

composante 1	composante 2	composante 3	pourcentage cumulé
96.18	3.17	0.33	99.68

Le premier profil spectral (Fig. B32-a) est un profil révélant la différence d'intensité entre les parois et les fonds de cellules. Dans la première composante image de fluorescence (Fig. B32-b), les pixels apparaissant en clair sont les pixels appartenant aux fonds de cellules. Les pixels noirs sont les pixels appartenant aux parois.

**Figure B32 a)- profil spectral de fluorescence 1 b)- images composantes de fluorescence 1.**

Le second profil spectral oppose un pic négatif à 390 nm à la région entre 450 nm et 504 nm en positif (Fig. B33-a). En comparant ce profil spectral aux spectres de produits purs (Fig. B28), on peut conclure qu'il oppose des pixels riches en acides hydroxycinnamiques à des pixels riches en lignines. Les pixels apparaissant sur la composante 2 en noir sont donc riches en acides hydroxycinnamiques et les pixels clairs sont riches en lignine (Fig. B33-b). Avec cette interprétation, nous pouvons interpréter que les parois cellulaires de sclérenchyme sont riches en lignine et les parois cellulaires du phloème contiennent des acides hydroxycinnamiques.

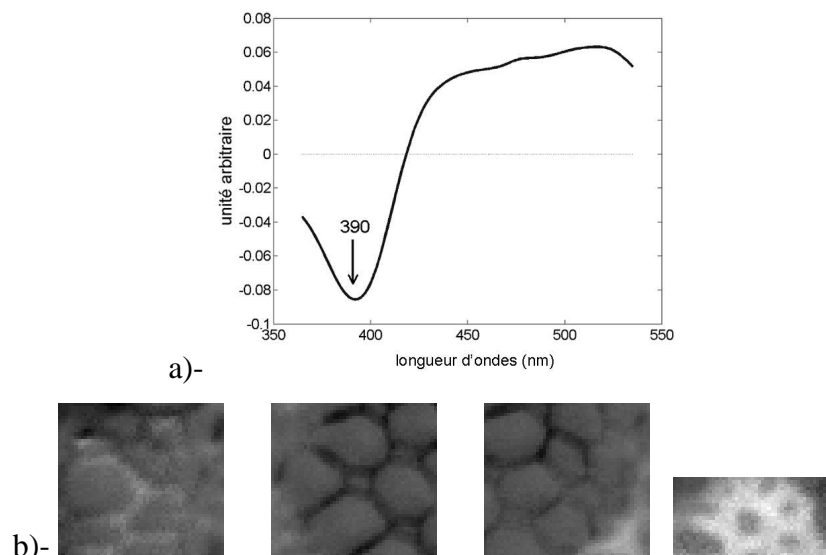


Figure B33 a)- profil spectral de fluorescence 2 b)- images composantes de fluorescence 2.

Le troisième profil spectral, oppose un pic positif à 420 nm à deux pics négatifs, le premier à 390 nm et le deuxième dans la région des lignines à partir de 509 nm (Fig. B34-a). En comparant le troisième profil spectral aux spectres de produits purs, nous pouvons interpréter qu'il oppose des pixels riches en acide férulique aux pixels qui contiennent des lignines. Sur la troisième composante (Fig. B34-b), les cellules du xylème et les cellules de sclérenchyme apparaissent en pixels noirs. Les pixels appartenant à la frontière séparant le phloème des autres types cellulaires apparaissent en clairs. Ils sont donc riches en acides féruliques.

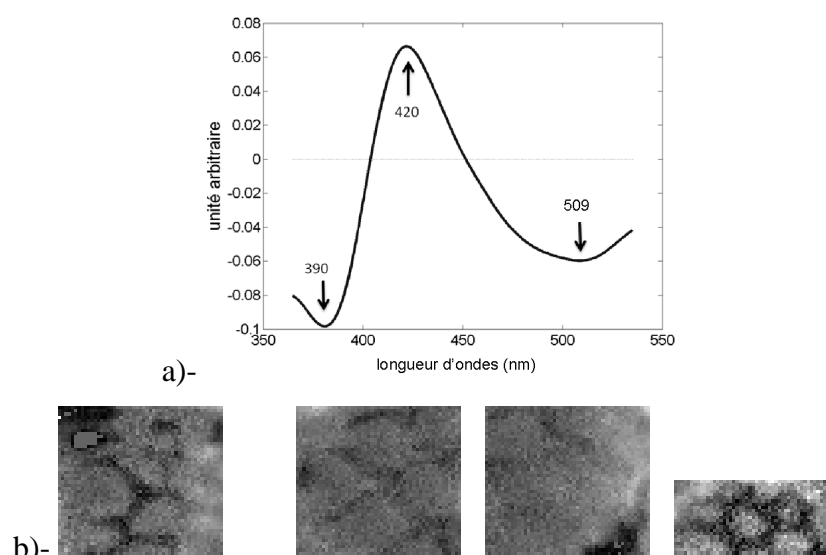


Figure B34 a)- profil spectral de fluorescence 3 b)- images composantes de fluorescence 3.

2.3 Analyse en composantes principales des images hyperspectrales Raman

Les images hyperspectrales Raman de la collection 4f11mb-2 ont été dépliées et concaténées dans un même tableau de données à deux entrées. Le tableau de données obtenu est soumis à une analyse en composantes principales. Trois composantes décrivant 97.81% de la variabilité totale ont été examinées.

Tableau B 3 Pourcentages de variance expliquée par les composantes principales des images hyperspectrales Raman de l'échantillon 4f11mb-2

composante 1	composante 2	composante 3	pourcentage cumulé
83.08	13.50	1.23	97.81

Le premier profil spectral Raman, comme en fluorescence et en infrarouge, révèle la variation d'intensité entre les spectres (Fig. B35-a). Dans la première image composante Raman (Fig. B35-b), les pixels apparaissant en noir sont les pixels appartenant aux intérieurs de cellules. Les pixels clairs sont les pixels appartenant aux parois.

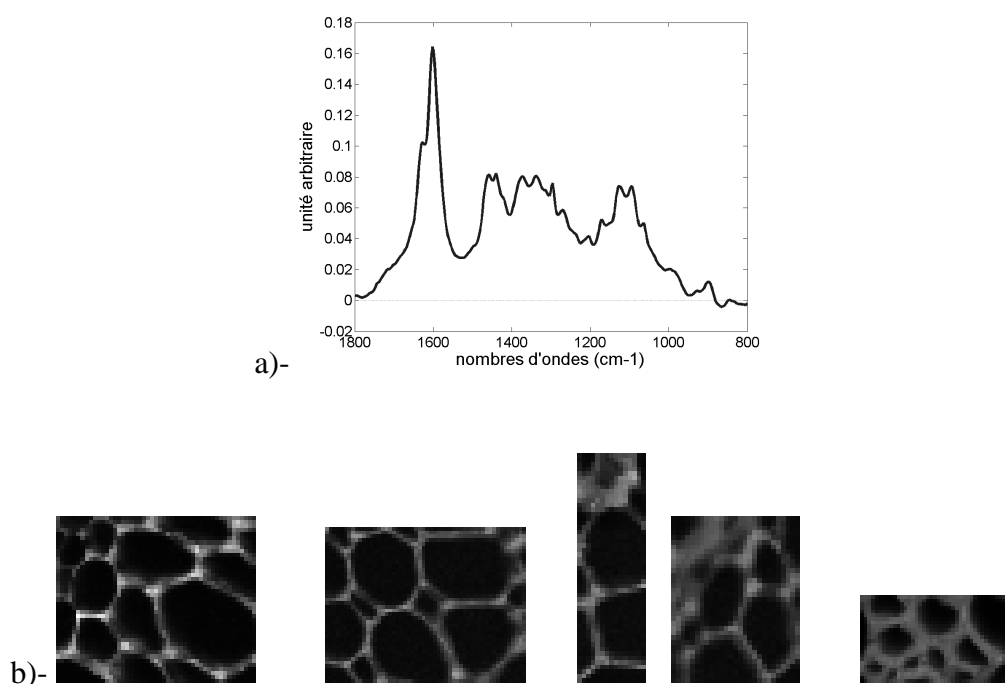


Figure B35 a)- profil spectral Raman 1 b)- images composantes Raman 1.

Le deuxième profil spectral oppose un pic positif à $(1600-1630) \text{ cm}^{-1}$ qui est un pic caractéristique des phénoliques à plusieurs pics négatifs : à $1093-1131$, 1296 , 1340 , 1370 , 1463 cm^{-1} sont caractéristiques des polysaccharides (Fig. B36-a). La deuxième composante

image oppose les parois cellulaires du phloème apparaissant en noir aux parois cellulaires de sclérenchyme apparaissant en blanc. Certaines parois cellulaires de xylème apparaissent également en pixel noir et d'autres en pixel blanc mais peu intense par rapport à celle du sclérenchyme sauf pour les jonctions cellulaires du xylème (Fig. B36-b). D'après cette interprétation, les cellules de sclérenchyme sont riches en lignines. Les cellules du phloème contiennent des polysaccharides. Les cellules du xylème contiennent un mélange de polysaccharides et de lignines avec une teneur plus élevée dans les jonctions des parois.

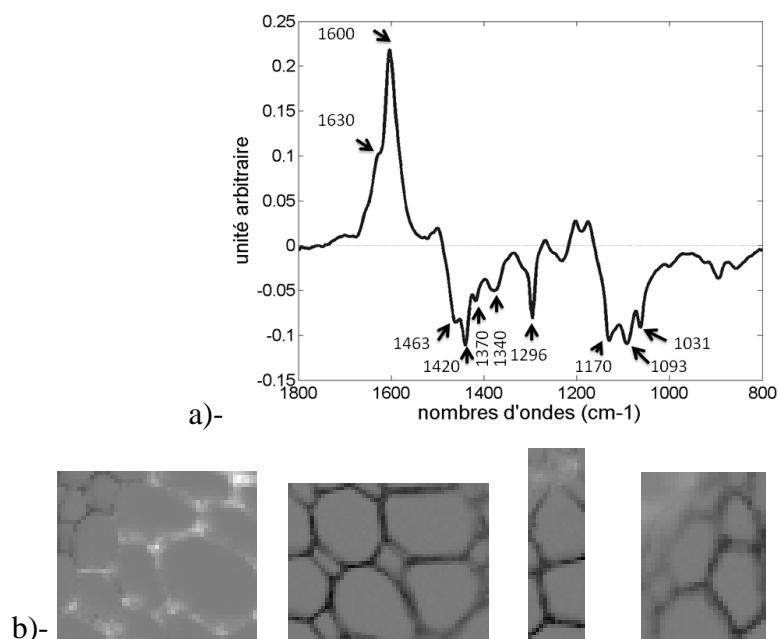


Figure B36 a)- profil spectral Raman 2 b)-images composantes Raman 2.

Les pics apparaissant dans le deuxième profil spectral en positif et en négatif, apparaissent tous en positif sur le troisième profil spectral Raman (Fig. B37-a), la partie négative ne contient pas de bande intéressante ou susceptible d'être attribuée en spectroscopie Raman. Le troisième profil spectral Raman peut être dû à des déformations de la ligne de base des spectres. Les cellules du xylème apparaissent en pixels noirs, les déformations de ligne de base sont alors observées pour les spectres du xylème (Fig. B37-b). La déformation de la ligne de base peut être liée à la différence d'épaisseur de l'échantillon ou à la différence du plan focal au moment de l'acquisition des images.

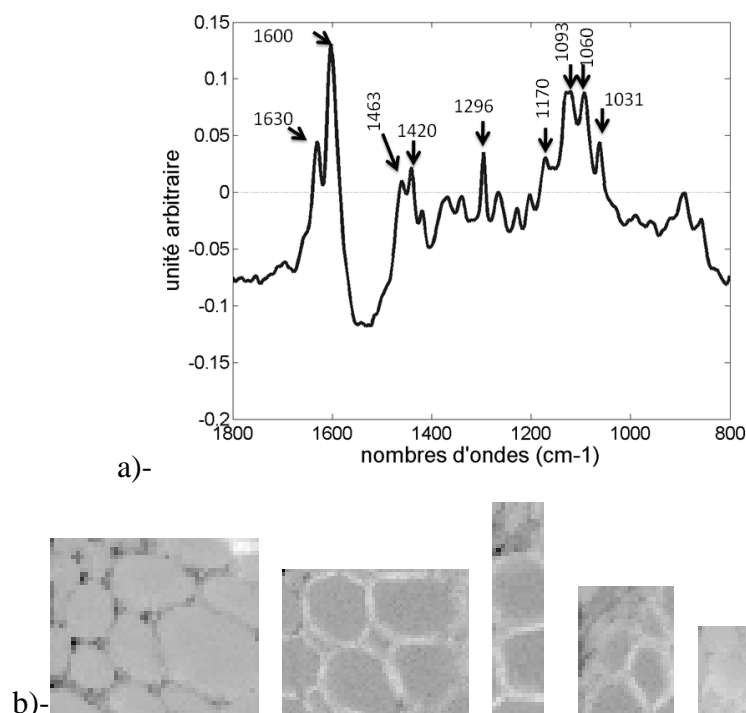


Figure B37 a)- profil spectral Raman b)- images composantes Raman 3.

Conclusion

Deux collections d'images hyperspectrales ont été sélectionnées parmi celles acquises au Synchrotron SOLEIL :

- 4f11mb-1 : c'est une collection contenant une image infrarouge et trois images de fluorescence. Cette collection a été acquise au cours de la première année de thèse dans un faisceau près de l'épiderme.

- 4f11mb-2 : c'est une collection contenant une image infrarouge, 4 images de fluorescence et 5 images Raman. Cette collection a été également acquise au cours de la deuxième année de thèse dans un faisceau près de l'épiderme.

Chaque collection a été acquise à la même section d'un faisceau vasculaire d'une tige de maïs recueillie au stade de la floraison femelle. Les trois microspectroscopies ont été utilisées afin d'analyser à la fois les polysaccharides, les lignines et les composés phénoliques présents dans les parois cellulaires. Les analyses et les explorations individuelles des données des images hyperspectrales au sein de chaque modalité ont été effectuées après pré-traitement des données. Le pré-traitement des images hyperspectrales consiste en la correction et la normalisation de spectres. La correction des spectres s'est basée sur la correction de la ligne de base (à cause des effets de diffusion et artefacts), l'élimination du bruit et le lissage. La normalisation de spectres a été faite avec une méthode qui prend en compte la nature spectrale

et spatiale des données. Les données ont été analysées individuellement en utilisant l'analyse en composantes principales afin de montrer l'information fournie par chacune des modalités. Les pics obtenus en spectroscopie Raman sont cohérents avec ce qui a été présenté en littérature. Certains pics en infrarouge correspondent aux pics présentés dans la littérature, d'autres présentent des différences d'un delta. Ces différences peuvent être dues à des bandes qui se chevauchent d'autant que nous sommes en train d'analyser des spectres de mélange assez complexes.

Partie C : Appariement des images hyperspectrales multirésolutions pour l'analyse multibloc

Sommaire

Introduction	101
I Recalage des images hyperspectrales	101
1 Notions de recalage d'images	101
2 Etapes de recalage	103
3 Transformation géométrique	104
3.1 La Rotation	104
3.2 Le changement d'échelle.....	105
3.3 La Translation	105
3.4 Transformation affine.....	106
4 Caractéristiques des images hyperspectrales.....	106
5 Stratégie adoptée	109
6 Méthodes	111
6.1 Choix des éléments caractéristiques à apparier.....	111
6.2 Mesure de ressemblance entre les éléments de l'image de référence et l'image à recaler	111
6.3 Recherche des facteurs d'échelle et des angles de rotation.....	112
6.3.1 Algorithme	112
6.3.2 Contraintes de calcul et choix des bornes du facteur d'échelle et de la rotation	113
6.4 Les pyramides d'images	114
6.5 Algorithme de recalage	116
7 Matrices de transformations	117
II Appariement de spectres	118
1 Projection des images de fluorescence et Raman dans l'espace infrarouge.....	118

2 Recherche des pixels infrarouge ayant des imagerie homologues en fluorescence et en Raman.....	120
3 Appariement des imagerie de fluorescence et Raman	122
III Construction des tableaux de données spectrales pour l'analyse multibloc	122
VI Résultats et discussions.....	124
1. Appariement de la collection 4f11mb-1	124
1.1 Recalage d'images	124
1.2 Structure de données	125
2 Appariement de la collection 4f11mb-2.....	126
2.1 Recalage d'images	126
2.2 Structure de données	127
Conclusion.....	128

Introduction

L'analyse conjointe des données infrarouge, fluorescence et Raman acquises sur une même coupe nécessite d'apparier les spectres acquis. La première étape du processus est d'identifier les spectres acquis à un endroit donné pour chaque image spectrale. Cette étape est réalisée au moyen de techniques de recalage d'images. La deuxième étape consiste à construire les tableaux de données pour l'analyse conjointe à partir des images hyperspectrales.

Dans cette partie, la méthode de recalage d'images hyperspectrales utilisée pour apparier des données issues des différentes microspectroscopies est présentée. Après recalage, les spectres infrarouge, de fluorescence et Raman homologues peuvent être extraits des images. La première partie présente les stratégies ainsi que les méthodes adoptées pour recaler les images hyperspectrales infrarouge, de fluorescence et Raman. La deuxième partie décrit les développements réalisés afin d'apparier les spectres et de les organiser dans une structure de données susceptible d'être analysée par les méthodes multibloc. La dernière partie est réservée à la présentation des résultats obtenus pour les collections des images 4f11mb-1 et 4f11mb-2.

I Recalage des images hyperspectrales

1 Notions de recalage d'images

Le recalage est le calcul d'une transformation permettant l'alignement de deux ou de plusieurs images de la même scène afin de comparer ou de fusionner leurs informations respectives. Le recalage des images a été le sujet de nombreuses études [Zitova et Flusser, 2003 ; Petrou *et al.*, 2004 ; Burger et Burge, 2008]. Les méthodes de recalage ont été développées pour aligner des images obtenues à partir de coupes sériées, pour réaliser des images panoramiques [Gledhill *et al.*, 2003], pour l'analyse du mouvement [Perperidis *et al.*, 2005 ; Zhu *et al.*, 2007] et pour fusionner les informations provenant de différents capteurs etc... Notre problématique se positionne dans cette dernière catégorie.

Supposons que l'objet à cartographier est celui présenté dans la figure C1.

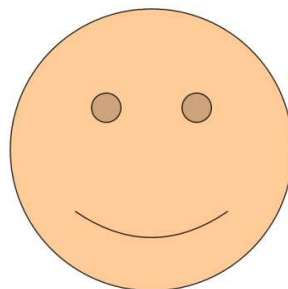


Figure C1 Région à cartographier

Dans la figure C2, les régions cartographiées par les trois instruments sont délimitées par des rectangles dans chaque région.

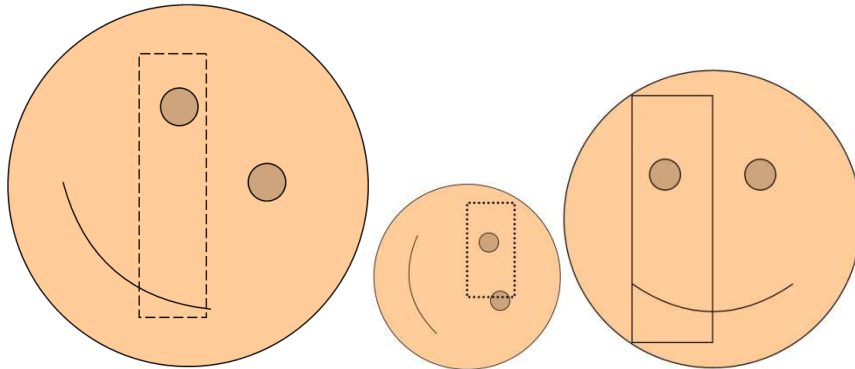


Figure C2 Régions cartographiées

Dans l'exemple présenté, les trois instruments ont des résolutions spatiales différentes conduisant à voir l'objet à différentes échelles. Des rotations sont également observées entre les images. De plus, les images acquises ne couvrent pas exactement la même région (Fig. C3). L'objectif du recalage est de trouver les pixels acquis en commun pour les trois images.

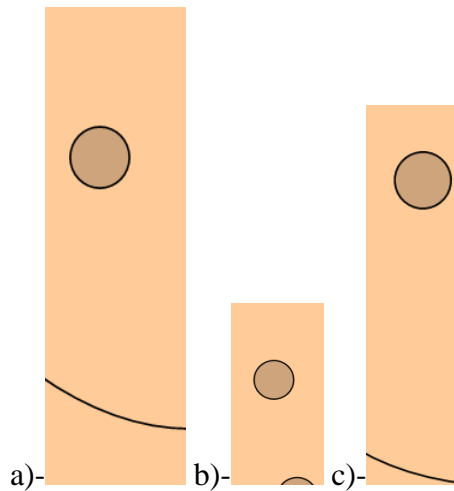


Figure C3 Images acquises avec l'instrument a)-1 b)-2 c)-3

L'idée naturelle est de projeter les trois images sur un même espace de référence et de calculer l'intersection entre elles (Fig. C4). Les pixels sont identifiés par leurs coordonnées en x et y dans chaque image. Les pixels communs sont identifiés par leurs coordonnées dans l'image de référence.

Afin de trouver les coordonnées des pixels dans la région commune, nous devons déterminer tout d'abord la transformation qui permet de passer du système des coordonnées des images individuelles vers le système de coordonnées de l'image de référence (Fig. C5).

Les méthodes de recalage cherchent à déterminer la meilleure transformation au sens d'un critère de similarité. Il existe différents types de transformations que Maintz et Viergever (1998) énumèrent :

- Affine : transformations conservant le parallélisme.
- Projective : met en correspondance des lignes non parallèles.
- Déformable : transforme des lignes droites en courbes.

Ces modèles de transformation peuvent être appliqués à toute ou à une partie de l'image. On parle alors respectivement de modèles globaux ou locaux.

Lorsque des images sont acquises sur une même coupe *a priori* sans déformations spatiales, un modèle de transformation affine incluant des rotations, des translations et des changements d'échelle est suffisant.

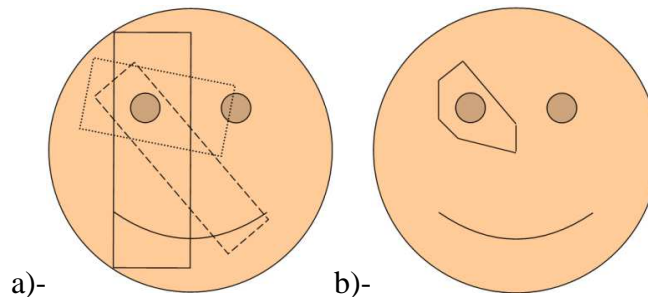


Figure C4 a)- Projection des images sur l'image de référence b)- Calcul de l'intersection des images dans l'image de référence

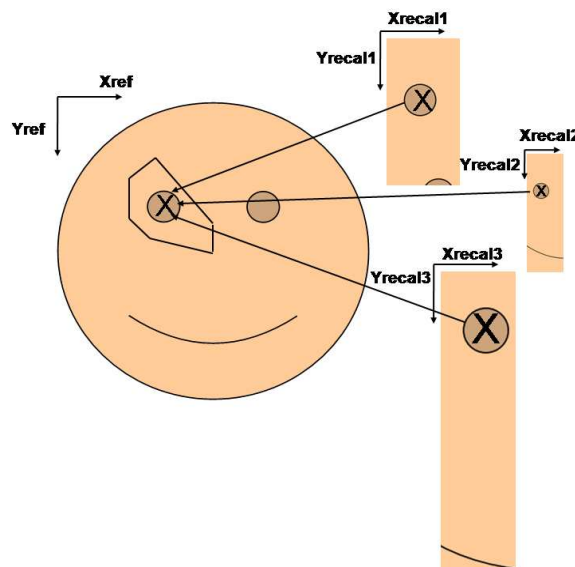


Figure C5 Correspondance entre les différents pixels des images

2 Etapes de recalage

La majorité des méthodes de recalage d'images est basée sur quatre étapes. La première est la détection d'éléments caractéristiques (contours, points, régions). La deuxième est la mise en

correspondance des éléments caractéristiques. La troisième est l'estimation des paramètres de la transformation permettant d'aligner l'image à recaler Im_{recal} avec l'image de référence Im_{ref} . La quatrième est le calcul de l'image transformée.

La transformation permettant de calculer les coordonnées du pixel (x_{recal}, y_{recal}) de l'image à recaler Im_{recal} dans l'image de référence Im_{ref} est exprimée comme suit :

$$(x_{ref}, y_{ref}) = \mathbf{T} (x_{recal}, y_{recal}) \quad \text{Equation C1}$$

Dans le cas du recalage affine, la transformation \mathbf{T} est obtenue à partir des transformations géométriques usuelles du plan.

3 Transformation géométrique

3.1 La Rotation

Les coordonnées (x_{ref}, y_{ref}) dans l'image de référence d'un point de coordonnées (x_{recal}, y_{recal}) dans l'image à recaler ayant subit une rotation d'angle θ autour d'un axe z peuvent être obtenues avec les formules trigonométriques élémentaires [Burger et Burge, 2008] (Fig. C6) :

$$x_{ref} = x_{recal} \cos \theta - y_{recal} \sin \theta \quad \text{Equation C2}$$

$$y_{ref} = x_{recal} \sin \theta + y_{recal} \cos \theta \quad \text{Equation C3}$$

Cette écriture peut se mettre également sous la forme matricielle suivante :

$$\begin{pmatrix} x_{ref} \\ y_{ref} \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_{recal} \\ y_{recal} \end{pmatrix}$$

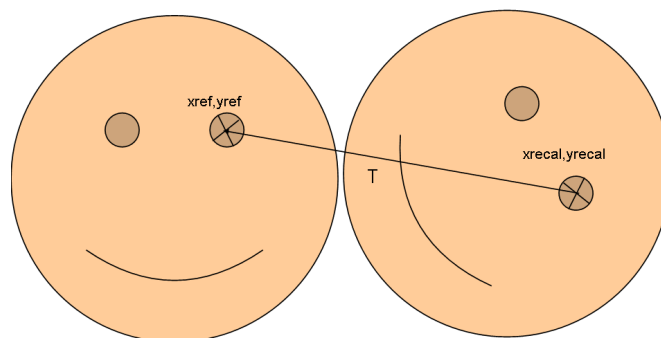


Figure C6 Transformation d'un angle $\theta=60^\circ$ d'un point (x_{recal}, y_{recal}) vers (x_{ref}, y_{ref})

3.2 Le changement d'échelle

Le changement d'échelle revient à multiplier les coordonnées (x_{recal}, y_{recal}) par un facteur s appelé facteur d'échelle. Dans certains cas, on peut avoir un facteur d'échelle en x et en y $\mathbf{s}=(s_x, s_y)$.

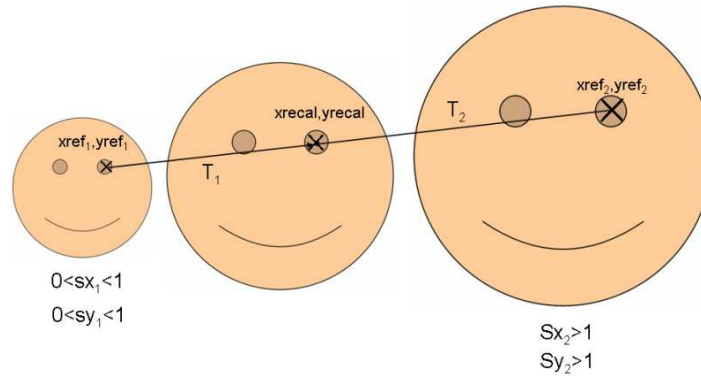


Figure C7 Changement d'échelle de l'image au milieu. La transformation T_1 permet de calculer les coordonnées (x_{ref1}, y_{ref1}) à partir des coordonnées (x_{recal}, y_{recal}) . La transformation T_2 permet de calculer les coordonnées (x_{ref2}, y_{ref2}) à partir des coordonnées (x_{recal}, y_{recal}) .

La transformation T s'écrit de la manière suivante :

$$x_{ref} = s_x x_{recal} \quad \text{Equation C4}$$

$$y_{ref} = s_y y_{recal} \quad \text{Equation C5}$$

Cette écriture peut se mettre également sous la forme matricielle suivante :

$$\begin{pmatrix} x_{ref} \\ y_{ref} \end{pmatrix} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \begin{pmatrix} x_{recal} \\ y_{recal} \end{pmatrix} \quad \text{Equation C6}$$

3.3 La Translation

La translation est définie par un vecteur $\mathbf{t}=(tx, ty)$ (Fig. C8). Les coordonnées du point (x_{recal}, y_{recal}) par la translation de vecteur \mathbf{t} est le point (x_{ref}, y_{ref}) calculé comme suit :

$$x_{ref} = x_{recal} + tx \quad \text{Equation C7}$$

$$y_{ref} = y_{recal} + ty \quad \text{Equation C8}$$

La forme matricielle s'écrit de la manière suivante :

$$\begin{pmatrix} x_{ref} \\ y_{ref} \end{pmatrix} = \begin{pmatrix} x_{recal} \\ y_{recal} \end{pmatrix} + \begin{pmatrix} tx \\ ty \end{pmatrix} \quad \text{Equation C9}$$

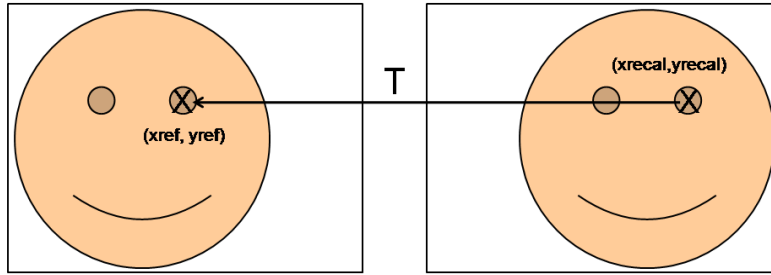


Figure C8 Transformation géométrique T basée sur la translation.

3.4 Transformation affine

La transformation affine est une transformation linéaire qui est exprimée par une partie multiplicative relative à la rotation et au facteur d'échelle, et une partie additive relative à la translation.

On considère ici que $s_x=s_y=s$.

Dans un espace bidimensionnel, une transformation affine s'écrit de la manière suivante.

$$\begin{pmatrix} x_{ref} \\ y_{ref} \end{pmatrix} = s \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x_{recal} \\ y_{recal} \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad \text{Equation C10}$$

Pour des raisons pratiques de calcul, August Ferdinand Möbius [Burger et Burge, 2008], a proposé un système de coordonnées qui permet d'inclure la translation dans la matrice de transformation par l'ajout d'une dimension supplémentaire. Ce nouveau système de coordonnées est appelé alors système de coordonnées homogènes.

Une transformation affine incluant une rotation d'angle θ , une translation $\mathbf{t}=(t_x,t_y)$ et un facteur d'échelle s est donc modélisée par la matrice suivante :

$$\mathbf{T} = \begin{pmatrix} s \cdot \cos(\theta) & -s \cdot \sin(\theta) & t_x \\ s \cdot \sin(\theta) & s \cdot \cos(\theta) & t_y \\ 0 & 0 & 1 \end{pmatrix} \quad \text{Equation C11}$$

Avec cette écriture, il est possible de calculer la transformation inverse \mathbf{T}^{-1} .

4 Caractéristiques des images hyperspectrales

Les images infrarouge, de fluorescence et Raman ont été acquises avec des systèmes instrumentaux dont les résolutions spatiales et spectrales sont différentes. La Figure C9-a, c et e, montre respectivement un exemple d'image spectrale infrarouge, de fluorescence et Raman

après pré-traitement des spectres avec leurs image visible associée. La Figure C9-g montre l'image de référence acquise au microscope confocal de Nantes.

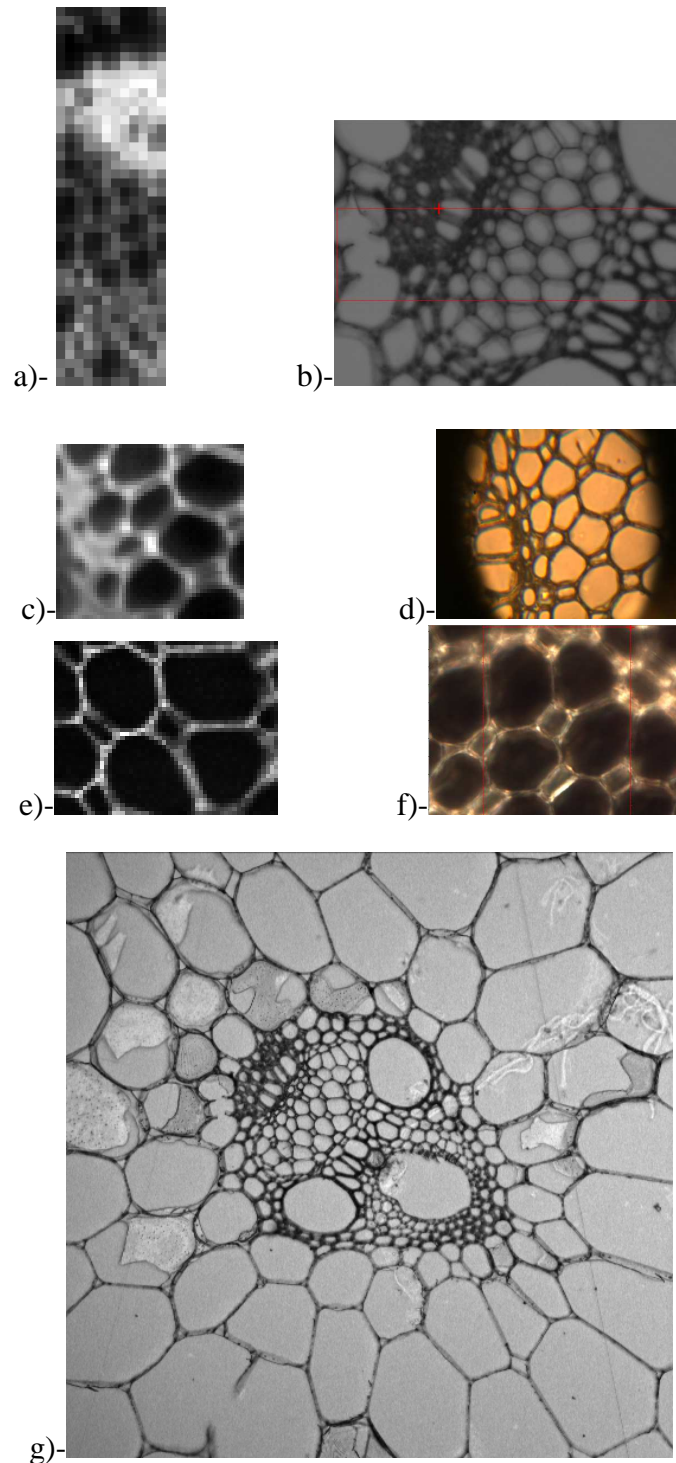


Figure C9 a)- Image spectrale infrarouge. Les intensités en niveau de gris ont été calculées pour chaque pixel comme étant la somme des intensités spectrales de la région entre 1800 et 800 cm^{-1} b) Image visible en fond clair acquise au microscope infrarouge c)- Image spectrale de fluorescence. Les intensités en niveau de gris ont été calculées pour chaque pixel comme étant la somme des intensités spectrales de la région entre 365 et 540 nm d)- Image visible en fond clair acquise au microscope de fluorescence e)- Image spectrale Raman. Les intensités en niveau de gris ont été calculées pour chaque pixel comme étant la somme des intensités spectrales de la région entre 1800 et 800 cm^{-1} f)- Image visible en fond noir acquise au microscope Raman g)- Image visible en fond clair acquise au microscope confocal.

Le tableau C1 synthétise les différences de résolutions entre les images spectrales ainsi que la taille maximale du champ de vue pour chaque microspectroscopie, le champ de vue étant la surface maximale pouvant être cartographiée.

Tableau C1 Caractéristiques des images spectrales

	Résolution (μm^2)	Champ de vue (μm^2)	Dimension (pixels) des images de la collection 4f11mb-1	Dimension (pixels) des images de la collection 4f11mb-2	couleur de la paroi	couleur du fond de cellule
Infrarouge	5×5	1350×250	270×50	210×60	blanc	noir
Fluorescence	1×1	100×100	36×67 36×67 36×67	47×51 47×51 47×51 24×51	blanc	noir
Raman	1×1	60×60	pas d'images Raman	41×49 42×54 15×54 27×35 24×43	blanc	noir

La sauvegarde des images visibles infrarouge, de fluorescence et Raman est basée sur une procédure d'impression d'écran. Les images visibles sont de qualité médiocre et de résolution variable. Le tableau C2 synthétise les propriétés des images visibles ainsi que la taille maximale du champ de vue pour chaque microspectroscopie.

Tableau C2 Caractéristiques des images visibles

	Champ de vue (μm^2)	Dimension (en pixel) des images de la collection 4f11mb-1	Dimension (en pixel) des images de la collection 4f11mb-2	couleur de la paroi	couleur du fond de cellule
Infrarouge	250×250	520×1093	697×925	noir	gris
Fluorescence	100×100	480×416 480×381 480×381	480 × 381 480 × 381 480 × 381 445 × 381	noir	orange
Raman	60×60	-	385× 509 385× 508 383× 508 383× 509 384 × 511	jaune	noir
référence mosaïque	-	1024 × 1280	-	noir	orange
référence confocale	650×650	-	512×512	noir	gris

Les images acquises ont les propriétés suivantes :

- Pour chaque modalité spectrale, des petits décalages ont été observés entre l'image visible et l'image spectrale.
- La taille du champ de vue des images visibles est très petite et dépasse légèrement celle des images spectrales.
- Les images de référence ont des résolutions et des tailles très supérieures aux images spectrales. Un grand contraste entre les parois et les fonds des cellules est observé pour les images de référence et un faible contraste est observé pour les images spectrales.

Les images spectrales n'ont pas la même résolution spatiale. Un pixel infrarouge représente une surface de $5 \times 5 \mu\text{m}^2$ par contre la taille d'un pixel de fluorescence ou de Raman a une surface de $1 \times 1 \mu\text{m}^2$. Un pixel en infrarouge correspond alors à un voisinage de $25=5 \times 5$ pixels en fluorescence et à un voisinage de 25 pixels en Raman. En d'autres termes, chaque pixel infrarouge correspond à une imagerie en fluorescence et une imagerie en Raman.

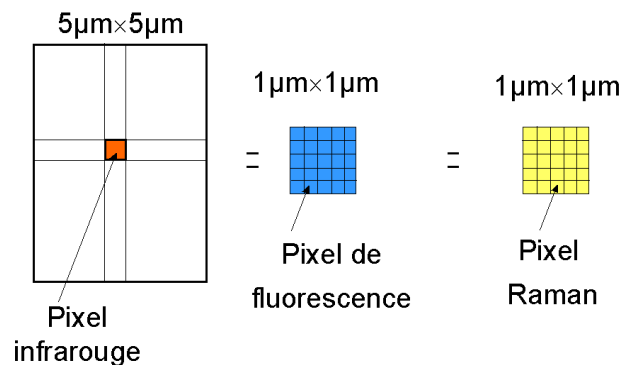


Figure C10 Représentation des pixels multirésolutions

5 Stratégie adoptée

Au vu des caractéristiques des images spectrales et des images visible, le recalage de l'image spectrale directement sur l'image de référence est une tâche très difficile.

Le recalage des images spectrales sur l'image de référence est réalisé en deux étapes (Fig. C11) :

- chaque image spectrale est recalée par rapport à son image visible associée. Les deux images ont des champs de vue de tailles comparables. Le recalage entre l'image spectrale et l'image visible a pour but de déterminer les petits décalages.
- chaque image visible est recalée par rapport à l'image de référence. Ces deux types d'images ont des caractéristiques de contraste paroi/fond de cellules comparables.

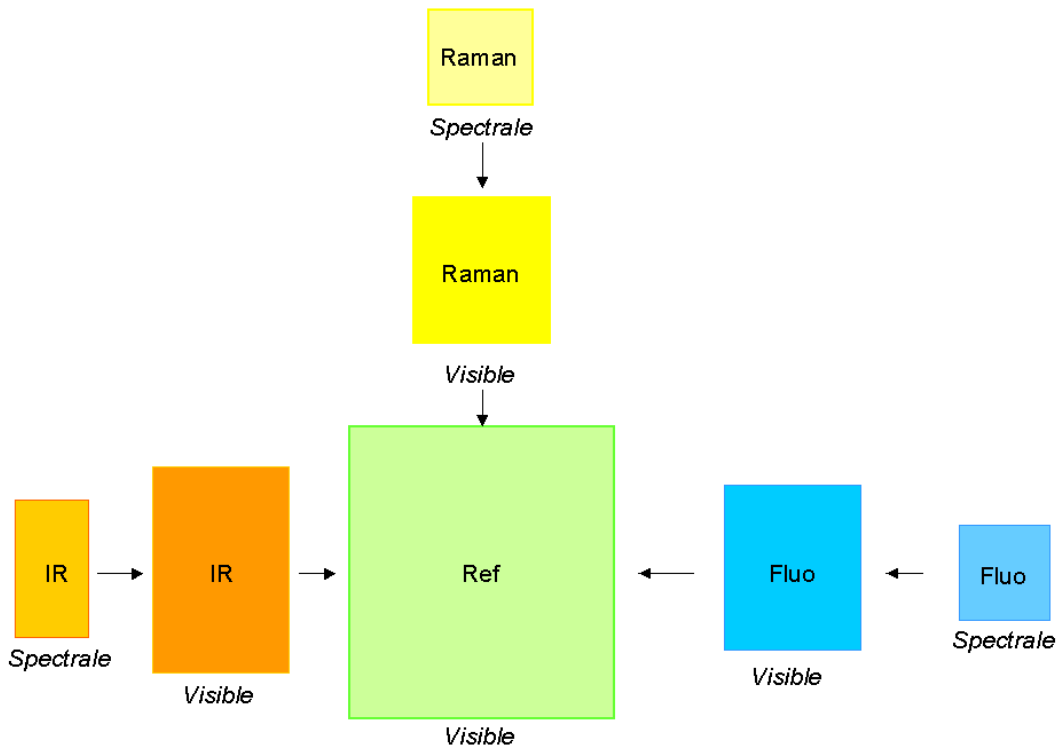


Figure C11 Stratégie adoptée pour recaler les images spectrales par rapport aux images de référence

Calcul de la matrice de passage de l'image spectrale vers l'image de référence

Le calcul de la transformation affine T permettant de passer directement de l'espace de l'image spectrale vers l'espace de l'image de référence est obtenu en multipliant deux matrices de transformations affines :

- T_s permettant de passer de l'image spectrale vers l'image visible.
- T_v permettant de passer de l'image visible vers l'image de référence.

$$T = T_v T_s$$

Equation C12

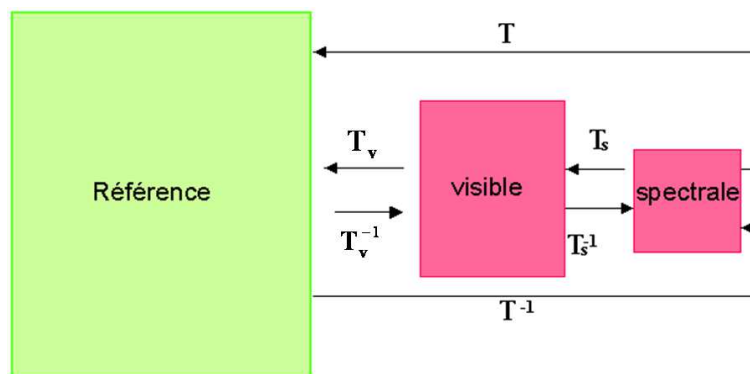


Figure C12 Matrices de passages entre image spectrale et image de référence

6 Méthodes

6.1 Choix des éléments caractéristiques à apparier

Les images spectrales et les images visibles sont beaucoup plus petites en dimension et en résolution que l'image de référence. Une approche basée région a été retenue. Dans les approches basées régions, la plus petite image est intégralement recherchée dans l'image de référence.

6.2 Mesure de ressemblance entre les éléments de l'image de référence et l'image à recaler

Les intensités de gris sont comparées à travers une mesure de similarité comme les corrélations, les mesures d'entropies [Neemuchwala, 2003 ; Lu, 2008] ou la somme des différences des niveaux de gris. Dans le cas d'une translation simple, la méthode la plus classique est le calcul de la corrélation croisée [Burger et Burge, 2008]. Le recalage est effectué en déplaçant l'image à recaler Im_{recai} dans l'image de référence Im_{ref} en mesurant la corrélation entre les pixels correspondants entre les deux images.

Le coefficient de corrélation croisée appliquée aux coordonnées x et y de l'image de référence est donné par :

$$C_N(x,y) = \frac{\sum_{s,t} [w(s,t) - \bar{w}] [f(x+s, y+t) - \bar{f}_{xy}]}{\sqrt{\sum_{s,t} [w(s,t) - \bar{w}]^2 \sum_{s,t} [f(x+s, y+t) - \bar{f}_{xy}]^2}} \quad \text{Equation C13}$$

Où :

w : les intensités de gris de l'image à recaler

\bar{w} : valeurs moyennes des intensités de gris de l'image à recaler

f : les intensités de gris de l'image de référence

\bar{f}_{xy} : la moyenne des intensités de gris de la région considérée dans l'image de référence.

s et x : indices des lignes de l'image à recaler et de référence

t et y : indices des colonnes de l'image à recaler et de référence

Un exemple de corrélation croisée entre une image spectrale et une image visible en fluorescence est présenté dans la figure C13. Le point clair brillant apparaissant au milieu de l'image indique que la valeur de corrélation croisée maximale est trouvée à cet endroit de l'image visible.

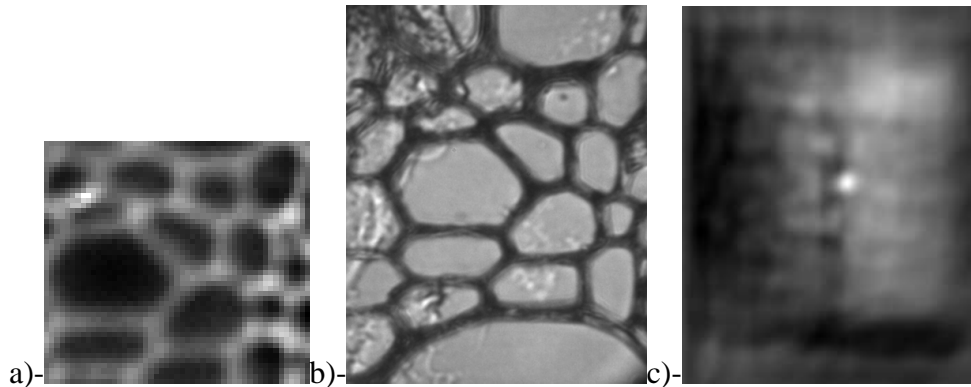


Figure C 13 a)-Image spectrale de fluorescence b)- image associé au canal vert de l'image visible acquise au microscope de fluorescence c)- Matrice de la corrélation croisée entre les deux images a et b.

6.3 Recherche des facteurs d'échelle et des angles de rotation

La méthode de corrélation croisée ne peut aligner que les images ayant un décalage en translation. Dans notre cas, la rotation et le facteur d'échelle doivent être déterminés. La solution est de tester différentes valeurs de facteur d'échelle et d'angle de rotation, puis de calculer les translations par corrélation croisée. Les valeurs maximales des corrélations permettent de sélectionner à la fois le facteur d'échelle, l'angle de rotation et le vecteur de translation. Une approche par dichotomie a été retenue pour tester les différentes combinaisons du facteur d'échelle et d'angle de rotation.

6.3.1 Algorithme

L'algorithme de la recherche est le suivant :

Etape 1: sélectionner un intervalle $[s_{min}, s_{max}]$ de recherche du facteur d'échelle avec un pas

$$\text{initial : } pass = \frac{s_{max} - s_{min}}{10} .$$

Etape 2: sélectionner un intervalle de recherche de l'angle de rotation $[\theta_{min}, \theta_{max}]$ avec un

$$\text{pas initial } pas_{\theta} = \frac{\theta_{max} - \theta_{min}}{10}$$

Etape 3: créer l'ensemble des images à recaler Im_{recal} par transformation en utilisant toutes les combinaisons {facteur d'échelle + rotation}.

Etape 4 : application de la corrélation normalisée entre chaque image à recaler et l'image de référence.

Etape 5 : sélectionner l'angle θ_j et le facteur d'échelle associé s_i correspondant à la valeur maximale de corrélation croisée.

Etape 6 : sélectionner un nouvel intervalle de facteurs d'échelles $[s_{min}, s_{max}] = [s_i - pass, s_i + pass]$

Etape 7 : sélectionner un nouvel intervalle d'angles $[\theta_{min}, \theta_{max}] = [\theta_i - pas \theta, \theta_i + pas \theta]$

Etape 8 : réduire le pas de recherche $pass = pass/10$

Etape 9 : réduire le pas de recherche $\theta = pas \theta/10$

Etape 10 : Répéter Etape1 → Etape9 jusqu'à ce qu'on obtienne une précision de 1 pour l'angle θ_j et de 0.01 pour le facteur d'échelle s_i .

6.3.2 Contraintes de calcul et choix des bornes du facteur d'échelle et de la rotation

Les bornes de l'intervalle des facteurs d'échelles $[s_{min}, s_{max}]$ sont déterminées en fonction de la taille des images à recaler Im_{recal} et celle de référence Im_{ref} . Dans notre cas, l'image à recaler Im_{recal} est beaucoup plus petite que l'image de référence Im_{ref} . Dans l'optique de recaler la petite image sur la grande, l'intervalle de recherche du facteur d'échelle est $[1; +\infty[$. Pour borner notre recherche nous allons chercher plutôt $\frac{1}{s}$ dans l'intervalle $]0.1]$.

Les bornes théoriques de recherche de l'angle de rotation $[\theta_{min}, \theta_{max}]$ sont $[0^\circ, 360^\circ]$. La rotation d'une image conduit à l'augmentation de sa taille et altère les pixels du bord (Fig. C14). Les déformations sont beaucoup plus apparentes pour les images de petites tailles (dans notre cas l'image spectrale). A cause de ce problème, notre recherche consiste à chercher l'angle $-\theta$ en appliquant la rotation à l'image de référence plutôt qu'à l'image spectrale.

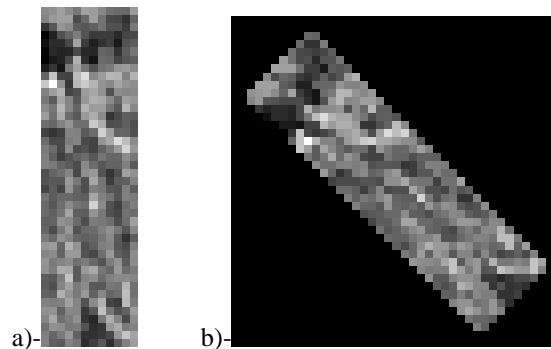


Figure C14 Rotation de l'image spectrale infrarouge avec un angle $\theta=45^\circ$ a)- image avant rotation b)- image après rotation

Les résolutions ainsi que le nombre des pixels des images de référence sont très différents de ceux des images à recaler. Nous proposons dans la suite de rechercher les paramètres de transformations en appliquant une décomposition pyramidale aux images considérées.

6.4 Les pyramides d'images

La notion de décomposition en pyramide a été initiée par Burt et Adelson [Burt et Adelson, 1983] pour la compression d'images. Elle a été utilisée avec d'autres méthodes tel que la segmentation, la recherche de motifs [Bonmassar et Shwartz, 1998] ou la corrélation [Adelson *et al.*, 1984]. Cette méthode est utilisée par exemple pour la détection des contours : faire la détection à basse résolution pour affiner leur localisation entre deux niveaux successifs en revenant à l'image d'origine. Cette méthode est la plus courante pour réaliser des mosaïques d'images [Dani et Chaudhuri, 1995].

Une image en pyramide se présente comme un empilement d'images de résolutions décroissantes (Fig. C15). Le niveau de base est l'image originale et les $N - 1$ niveaux successifs sont de résolution de plus en plus réduite. En montant dans la pyramide d'image, à chaque niveau la taille de l'image diminue de moitié.

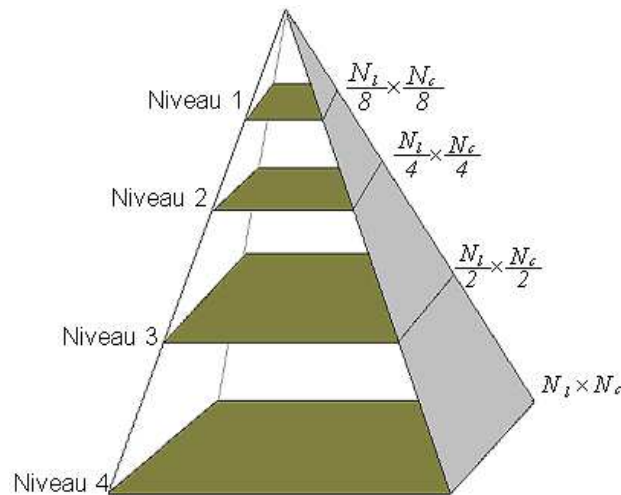


Figure C15 Modèle de la pyramide

Le principe de la méthode de construction de pyramide sans perte d'information, définie par Burt et Adelson [Burt et Adelson, 1983], [Adelson *et al.*, 1984] est présentée ci-dessous.

Une pyramide est générée par une opération de réduction de résolution de l'image de départ. Pour un niveau N donnée cette opération peut s'écrire de la manière suivante :

$$g_N(i, j) = \sum_{m=-2}^2 \sum_{n=-2}^2 w(m, n) g_{N-1}(2i + m, 2j + n) \quad \text{Equation C14}$$

Où $g_N(i, j)$ est la valeur du pixel (i, j) au niveau N et $w(m, n)$ est le poids affecté à chaque pixel pour la réduction. Un certain nombre d'opérateurs peuvent être utilisés, le filtre moyen au voisinage qui produit une pyramide dite "moyenne", le filtre passe-bas Gaussien, qui réalise une pyramide Gaussienne, des opérateurs de morphologie mathématique pour créer une

pyramide morphologique ou sans filtre qui donne une pyramide d'échantillonnage. La qualité de l'image sortante dépend de la sélection du filtre.

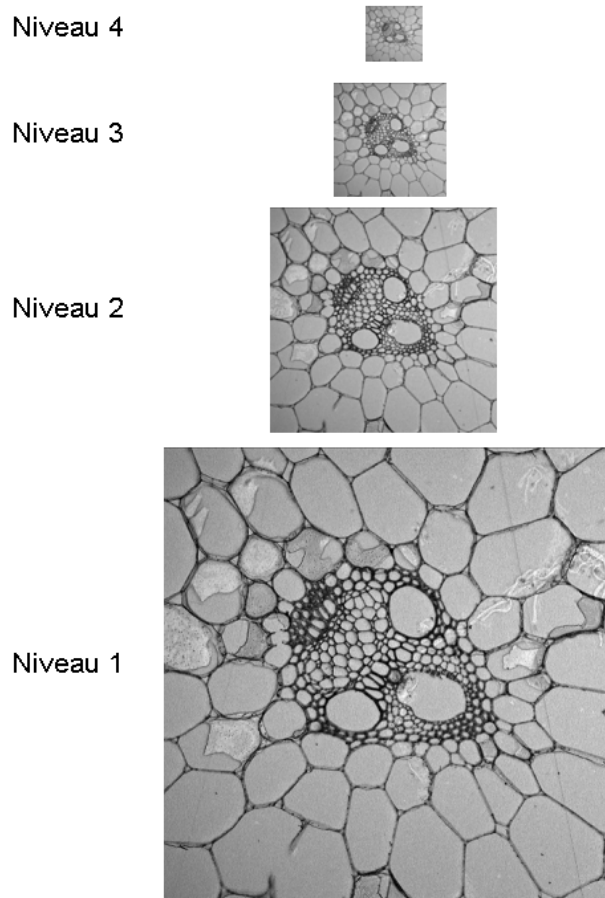


Figure C16 Exemple d'une pyramide Gaussienne obtenue à partir de l'image acquise au microscope confocal.

La figure C16 montre l'application d'une pyramide Gaussienne à l'image de référence de la collection d'images 4f11mb-2. L'application de la pyramide Gaussienne diminue la taille des images de moitié à chaque niveau. Nous avons retenu cette approche pour ramener les images de référence et les images à recaler à des tailles et des résolutions comparables.

Sur la figure C17, nous présentons un exemple de dégradation en pyramide de l'image visible infrarouge. Il est clair qu'à faible résolution, l'image visible infrarouge (Fig. C17-b) a des caractéristiques de contraste comparable à l'image spectrale (Fig. C17-C) ce qui n'était pas le cas pour l'image d'origine. L'image spectrale infrarouge et l'image de référence n'ont pas le même teinte paroi/fond de cellule (Fig. C17-b et c). Il était nécessaire d'inverser les niveaux de gris de l'image visible pour avoir le même contraste que l'image spectrale (Fig. C17-d).

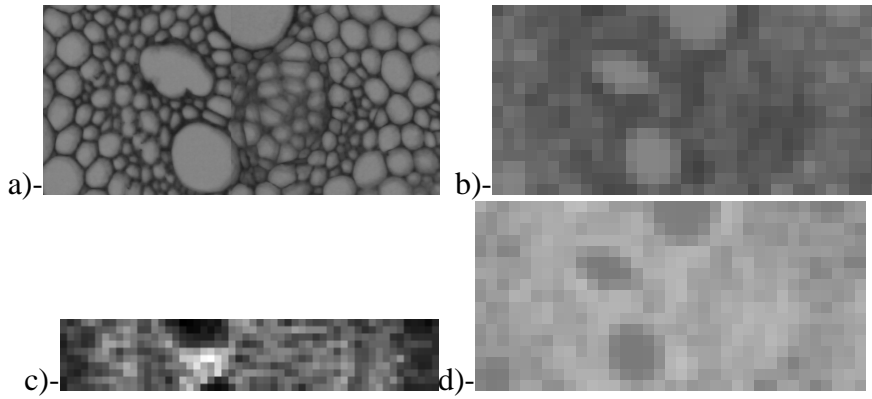


Figure C17 a)-Image visible infrarouge b)-Image visible infrarouge obtenue après dégradation avec application de la pyramide Gaussienne avec un niveau $N_{Ref}=4$ c)-Image spectrale infrarouge d)- image infrarouge C17-b dégradé avec des niveaux de gris inversés.

Contraintes de calcul

Dans cette approche multirésolutions, le niveau de la pyramide doit être choisi pour préserver la structure cellulaire au sein des images. L'algorithme développé propose toutes les solutions possibles et l'expert doit choisir un niveau adapté. Le niveau de l'image à recalcr dans la pyramide est choisi pour assurer que l'image à recalcr reste plus petite en nombre de ligne et nombre de colonne que l'image de référence.

6.5 Algorithme de recalage

Etape 1 : choisir le niveau de la pyramide N_{Ref} à appliquer à l'image de référence

Etape 2 : calculer le niveau de la pyramide N_{Recal} à appliquer à l'image à recalcr

Etape 3 : sélectionner l'intervalle $[s_{min}, s_{max}]$ de la recherche du facteur d'échelle avec un pas

$$\text{initial : } pass = \frac{s_{max} - s_{min}}{10}$$

Etape 4 : sélectionner l'intervalle de la recherche de la rotation $[\theta_{min}, \theta_{max}]$ avec un pas

$$\text{initial } pas\theta = \frac{\theta_{max} - \theta_{min}}{10}$$

Etape 5 : créer l'ensemble des images à recalcr Im_{recal} par transformation en utilisant toutes les combinaisons {facteur d'échelle + rotation}.

Etape 6 : appliquer la corrélation normalisée entre les images à recalcr et l'image de référence une par une.

Etape 7 : sélectionner l'angle θ_j et le facteur d'échelle s_i relatif à la valeur maximale de la corrélation croisée.

Etape 8 : remonter à un niveau plus haut dans la pyramide $N_{Ref} = N_{Ref} - 1$ et $N_{Recal} = N_{Recal} - 1$

Etape 9 : sélectionner un nouvel intervalle de recherche des facteurs d'échelles

$$[s_{min}, s_{max}] = [(s_i - \text{pass}) * 2^{(N_{Ref} - N_{Recal})}, (s_i + \text{pass}) * 2^{(N_{Ref} - N_{Recal})}]$$

Etape 10 : sélectionner un nouvel intervalle d'angles $[\theta_{min}, \theta_{max}] = [\theta_i - \text{pas } \theta, \theta_i + \text{pas } \theta]$

Etape 11 : $\text{pass} = \text{pass} / 10$

Etape 12 : $\text{pas } \theta = \text{pas } \theta / 10$

Etape 13 : Répéter Etape 3 → Etape 9 jusqu'à obtenir $N_{Ref} = 0, N_{Recal} = 0, \theta = 1$ et $s = 0.01$

7 Matrices de transformations

A la fin de la procédure de recalage nous disposons des matrices de transformation T_{IR} , T_{fluo} et T_{Raman} reliant, respectivement, les différentes images spectrales infrarouge, fluorescence et Raman à l'image de référence selon l'équation C12. Ces matrices sont nécessaires pour trouver les spectres homologues pour les trois spectroscopies.

Les matrices inverses respectives T_{IR}^{-1} , T_{fluo}^{-1} et T_{Raman}^{-1} permettent le passage de l'espace de l'image de référence vers l'espace des images spectrales (Fig. C18).

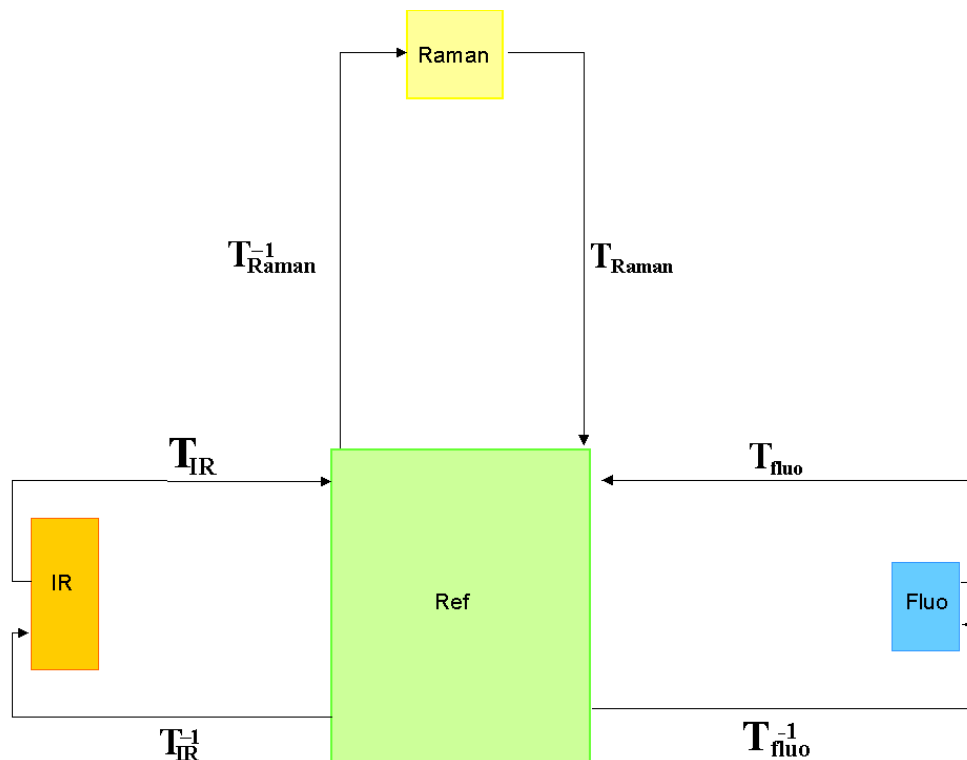


Figure C 18 Matrices de passages entre les différentes images spectrales en passant par l'image de référence.

A partir de ce schéma, il est possible de passer directement de l'espace infrarouge vers l'espace de fluorescence et Raman (Fig. C19). Les matrices de passages de l'espace infrarouge vers l'espace de fluorescence et à l'espace Raman sont obtenues respectivement comme suit :

$$\mathbf{T}_{\text{IR-fluo}} = \mathbf{T}_{\text{fluo}} \mathbf{T}_{\text{IR}}^{-1} \quad \text{Equation C15}$$

$$\mathbf{T}_{\text{IR-Raman}} = \mathbf{T}_{\text{Raman}} \mathbf{T}_{\text{IR}}^{-1} \quad \text{Equation C16}$$

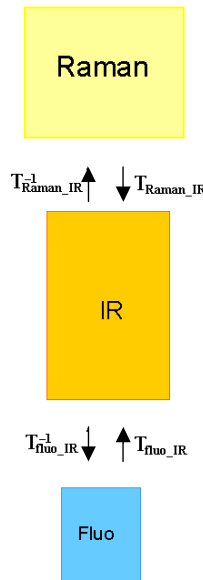


Figure C19 Matrices de passages entre les différentes images spectrales

II Appariement de spectres

L'intersection entre les images spectrales a été calculée de manière à minimiser leur nombre de rotations (Fig. C4). L'image spectrale infrarouge, qui est l'image la moins résolue et qui a le plus grand champ de vue a été retenue comme un nouvel espace de référence. L'espace de référence étant choisi, la suite du processus consiste à projeter les images spectrales de fluorescence et Raman sur l'image spectrale infrarouge. L'étape suivante est d'identifier les spectres infrarouge ayant des imageries homologues en fluorescence et en Raman. Dans ce qui va suivre, nous allons illustrer notre travail à partir de la collection 4f11mb-2.

1 Projection des images de fluorescence et Raman dans l'espace infrarouge

Dans un premier temps, la région cartographiée en infrarouge est considérée à la même résolution que la fluorescence et le Raman.

La projection se fait en calculant les coordonnées des pixels extrêmes de chaque image de fluorescence et Raman dans l'espace infrarouge. Pour une image ayant N_l lignes et N_c

colonnes, on appelle les pixels extrêmes les pixels de coordonnées $(1,1)$, $(1,N_c)$, $(N_l,1)$ et (N_l, N_c) .

Les coordonnées des pixels homologues des points extrêmes d'une image de fluorescence dans l'espace infrarouge sont calculées comme suit :

$$\begin{pmatrix} x_{IR} \\ y_{IR} \\ I \end{pmatrix} = \mathbf{T}_{fluor-IR} \begin{pmatrix} N_l \\ N_c \\ I \end{pmatrix} \quad \text{Equation C17}$$

Des images binaires à la résolution la plus haute (résolution des images de fluorescence et Raman) ont été obtenues. Les régions cartographiées en fluorescence et Raman projetées dans l'image infrarouge sont données dans les figures C20 et C21.

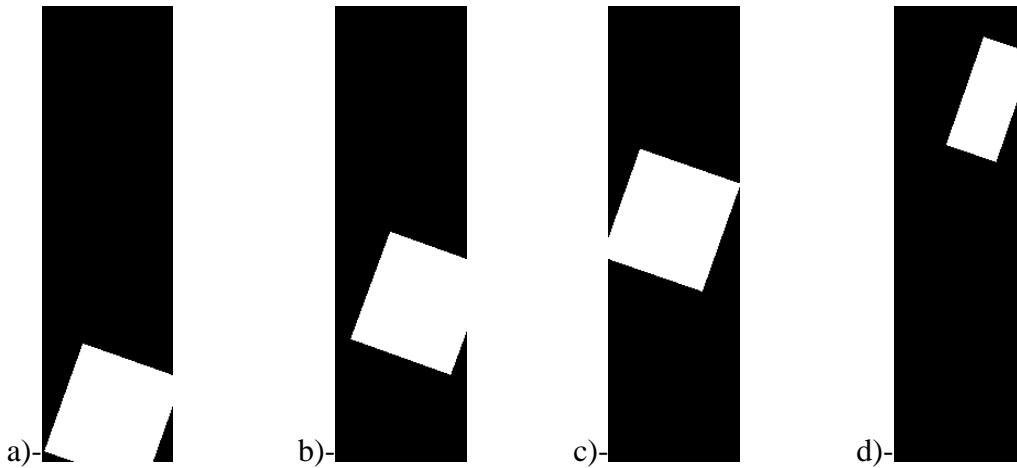


Figure C20 images individuelles de fluorescence projetées sur l'image infrarouge

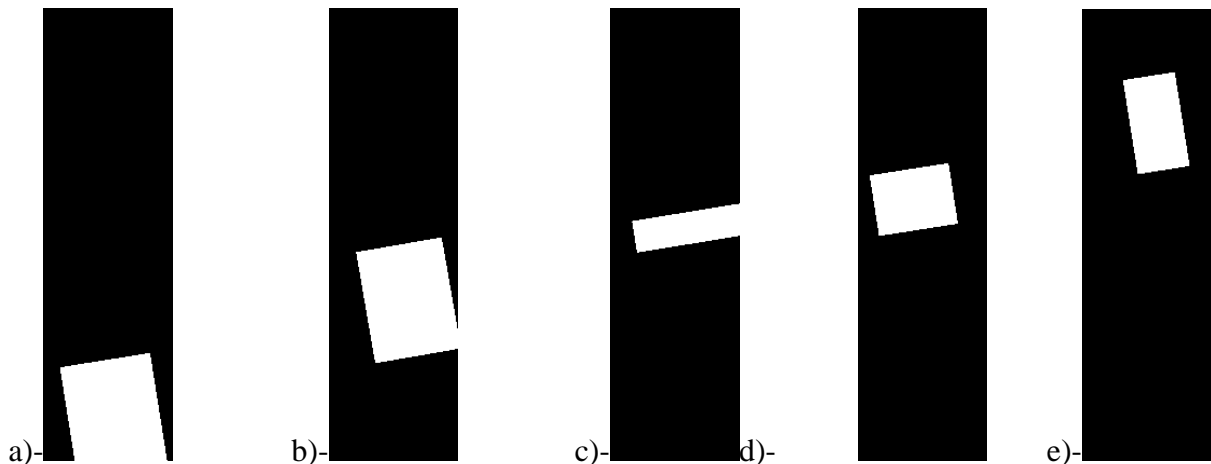


Figure C 21 images individuelles Raman projetées sur l'image infrarouge

- La région totale cartographiée est obtenue par union des images individuelles projetées (Fig. C21)

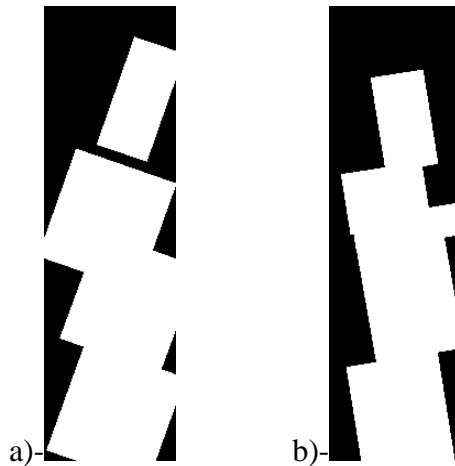


Figure C 22 Région cartographiée totale a)- en fluorescence b)- en Raman

- La région commune est obtenue par intersection des images totales projetées (Fig. C23).



Figure C23 Région commune

2 Recherche des pixels infrarouge ayant des imagerie homologues en fluorescence et en Raman

L'image de la région commune est obtenue est à la résolution des images de fluorescence et Raman. La simple réduction en résolution de l'image commune n'est pas suffisante pour trouver les pixels infrarouge ayant des imagerie homologues en fluorescence et en Raman. En effet, à un pixel infrarouge, doit correspondre une imagerie entière, i.e. un carré $p \times p$ pixels. La réduction de la résolution conduit à des effets de bord très important. Les pixels infrarouge de la région commune, peuvent être obtenus en érodant l'image région commune avec un élément structurant carré ayant la même surface que l'imagerie (Fig. C24). L'érosion est une opération élémentaire de morphologie mathématique [Serra, 1982] qui permet de ne

conserver que les pixels d'une région que lorsque l'élément structurant centré sur eux est complètement inclus dans la région.



Figure C 24 La région commune totale a)- avant érosion b)- après érosion

Les coordonnées des pixels infrarouge dans l'image spectrale infrarouge (x_{IR}, y_{IR}) sont calculées en passant l'image érodée à la résolution de l'image infrarouge.

Les coordonnées des points centraux (x_c, y_c) de l'imagette de fluorescence sont obtenues en transformant les coordonnées de chaque pixel infrarouge appartenant à la région commune vers l'espace de l'image de fluorescence et Raman.

$$\begin{pmatrix} x_{fluo} \\ y_{fluo} \\ I \end{pmatrix} = \mathbf{T}_{IR_fluo} \begin{pmatrix} x_{IR} \\ y_{IR} \\ I \end{pmatrix} \quad \text{Equation C18}$$

$$\begin{pmatrix} x_{Raman} \\ y_{Raman} \\ I \end{pmatrix} = \mathbf{T}_{IR_Raman} \begin{pmatrix} x_{IR} \\ y_{IR} \\ I \end{pmatrix} \quad \text{Equation C19}$$

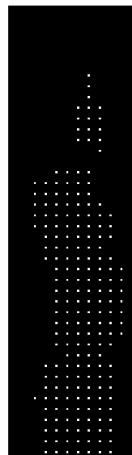


Figure C 25 Les points centraux des surfaces des pixels infrarouge à la résolution de la fluorescence et du Raman.

A la fin de cette étape, les pixels infrarouge ayant des imagerie homologues en fluorescence et en Raman sont trouvés.

3 Appariement des imagerie de fluorescence et Raman

Cette étape a été réalisée en utilisant des images Label (Fig. C26). Une image Label est une image indexée où les pixels prennent des valeurs indexées en fonction de leur région d'origine. Dans la figure C26, les pixels des régions cartographiées totales en fluorescence et en Raman sont numérotés en fonction de leur image d'origine. L'image d'origine du pixel central des imagerie de fluorescence et Raman est retrouvée à partir de l'image Label. L'ensemble des pixels des imagerie sont sélectionnés. Lorsque l'imagerie appartient à une zone de recouvrement entre les deux images d'origine, les pixels de la première image de la collection ont été retenus.

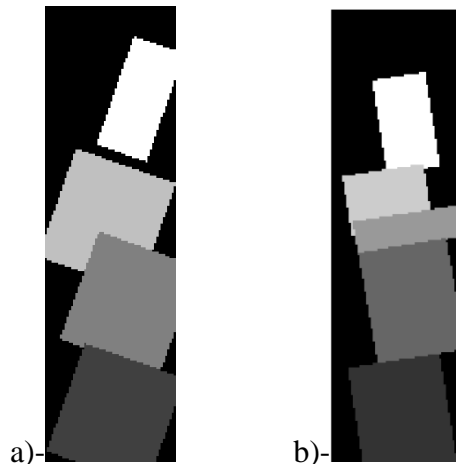


Figure C 26 Les images Label a)- de fluorescence b)- Raman

III Construction des tableaux de données spectrales pour l'analyse multibloc

Classiquement, en analyse des données, les images hyperspectrales sont dépliées en un tableau de spectres à deux dimensions – une dimension spatiale relative aux pixels et une dimension de variables spectrales. Dans notre cas, chaque pixel infrarouge correspond à une imagerie de taille $P \times P$ en fluorescence et en Raman.

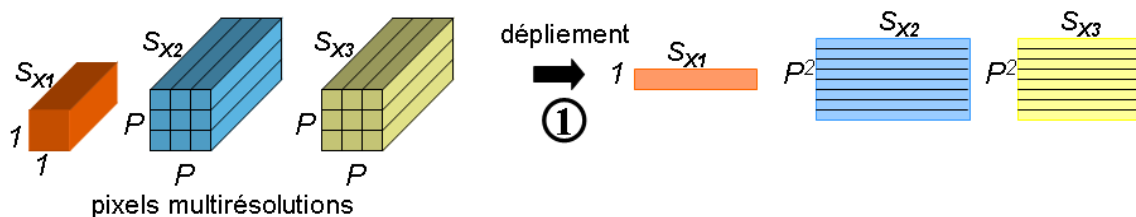


Figure C 27 Dépliage des pixels multirésolutions

Pour respecter les différences de résolutions, notre structure de données a été construite de la manière suivante (Fig. C27 et C28):

- considérer l'ensemble des spectres infrarouge ayant des imagettes homologues en fluorescence et en Raman.

- déplier chaque imagette de fluorescence et Raman associée aux spectres infrarouge (Fig. C27).

- associer à chaque spectre infrarouge une matrice de P^2 spectres de l'imagette dépliée (Fig. C27 et 28).

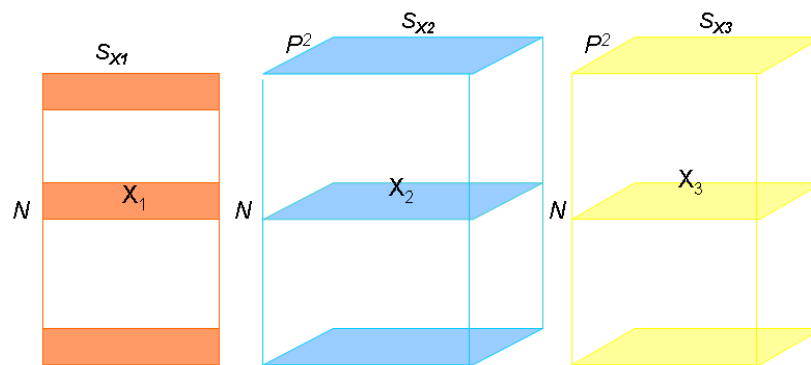


Figure C28 Empilement de tous les spectres

La structure finale de données est un ensemble de trois tableaux (Fig. C28) :

- Un tableau à deux entrées des données infrarouges X_1 avec une dimension spatiale et une dimension spectrale : premier mode: N pixels, second mode S_{x_1} nombre d'onde, N étant le nombre de spectres infrarouge pour lesquels des spectres de fluorescence et Raman ont également été acquis.

- Un tableau à trois entrées des données de fluorescence X_2 ; premier mode: N pixels, second mode S_{x_2} et troisième mode: P^2 pixels.

- Un tableau à trois entrées des données Raman X_3 (premier mode: N pixels, second mode S_{x_3} et troisième mode: P^2 pixels).

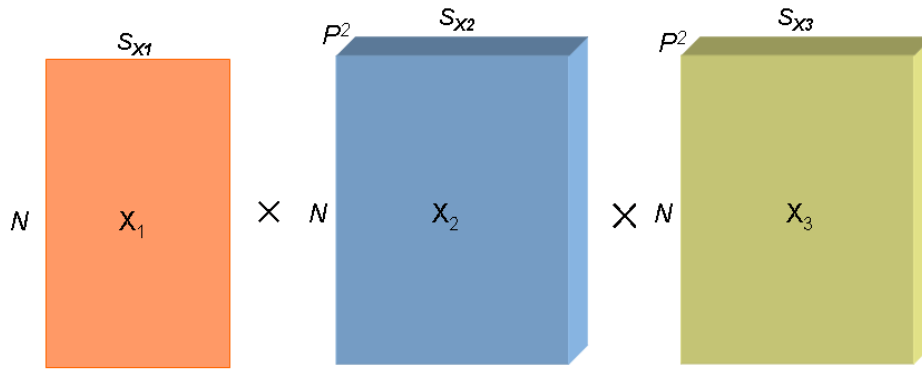


Figure C 29 Etablissement de la structure de données

VI Résultats et discussions

1 Appariement de la collection 4f11mb-1

1.1 Recalage d'images

Les images spectrales ont été recalées par rapport à l'image de référence qui est dans ce cas l'image mosaïque prise au microscope de fluorescence de la ligne DISCO. Les facteurs d'échelles et les angles de rotation de l'image infrarouge étaient de 42,7 et 88°, respectivement. Les facteurs d'échelle ont été estimés à 6,3, 6,2 et 6,3 pour les trois images de fluorescence et les angles de rotation étaient de 1°, 0° et 0°. La résolution théorique des images infrarouge est de 5 μm celle de la fluorescence est de 1 μm . En pratique, le rapport des facteurs d'échelle obtenu est de $\frac{42}{6} = 7$, indiquant qu'un pixel infrarouge correspondait à 7 \times 7 pixels en fluorescence. Après vérification de l'étalonnage du microscope de fluorescence, la résolution spatiale lors de l'acquisition était de 0.8 μm au lieu de 1 μm . Ainsi, le rapport obtenu après recalage est proche du rapport entre les deux résolutions des systèmes d'acquisition.

Pour cette analyse, chaque pixel infrarouge dans la région d'intérêt a été couplé à ses 49 spectres homologues de fluorescence pour construire le tableau à trois entrées des données de fluorescence. La figure C30 montre la fusion des images spectrales 4f11mb-1 avec l'image de référence après recalage.

Pour fusionner les images spectrales avec les images de référence (Fig. C30), une interpolation bilinéaire est utilisée [Burger et Burge, 2008].

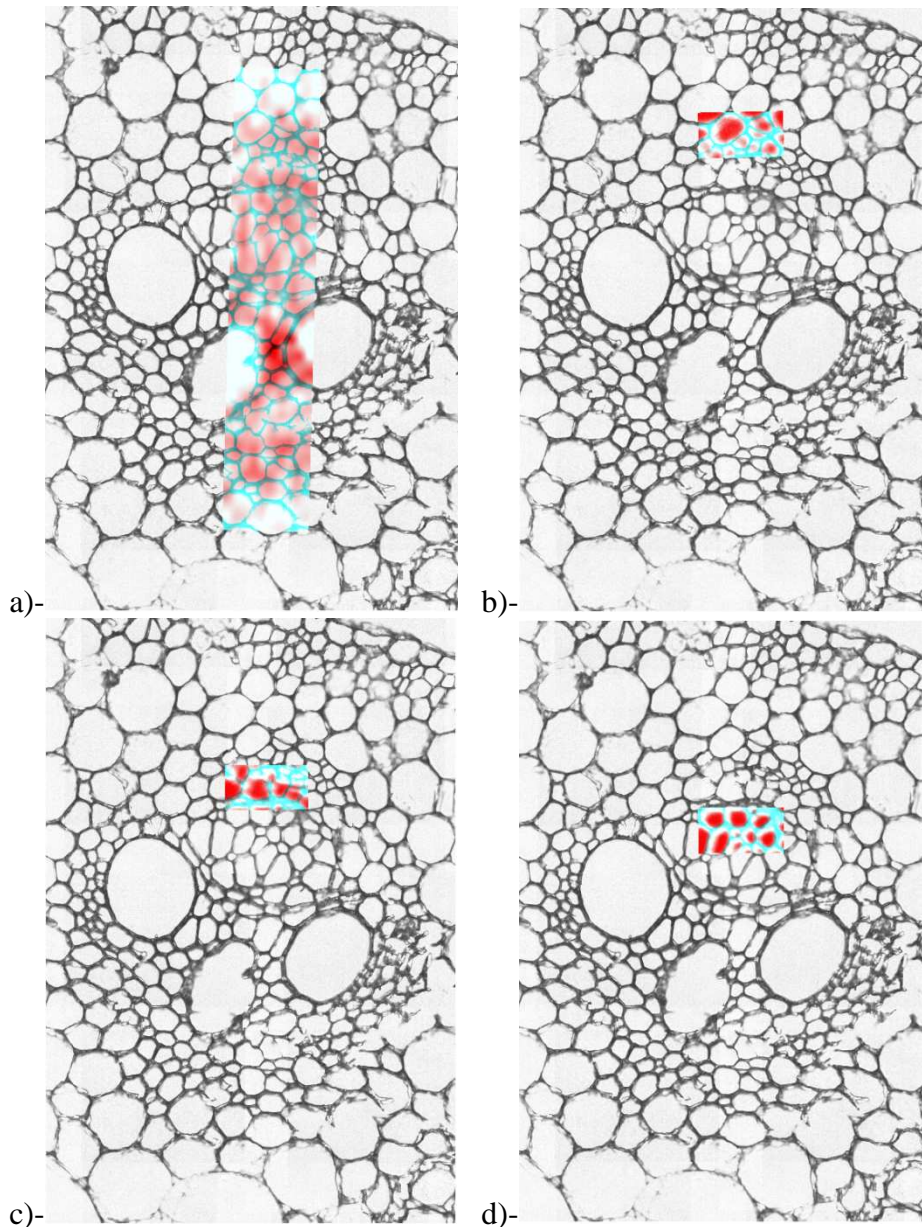


Figure C 30 Fusion de l'image visible de référence avec: a) l'image spectrale infrarouge (intensité de la de la région entre 1800-800 cm⁻¹), b, c, d) l'image spectrale de fluorescence (somme des intensités spectrales entre 360-540 nm)

1.2 Structure de données

La structure des données finale obtenue pour la collection des images 4f11mb-1 est un ensemble de deux tableaux de données : un tableau de données à deux entrées avec une dimension spatiale et une dimension spectrale (première entrée: 117 pixels et deuxième entrée : 245 nombres d'ondes) couplé à un tableau de données à trois entrées à deux dimensions spatiales et une dimension spectrale (première entrée : 117 pixels, deuxième entrée : 662 longueurs d'ondes et troisième entrée : 49 pixels) (Fig. C31).

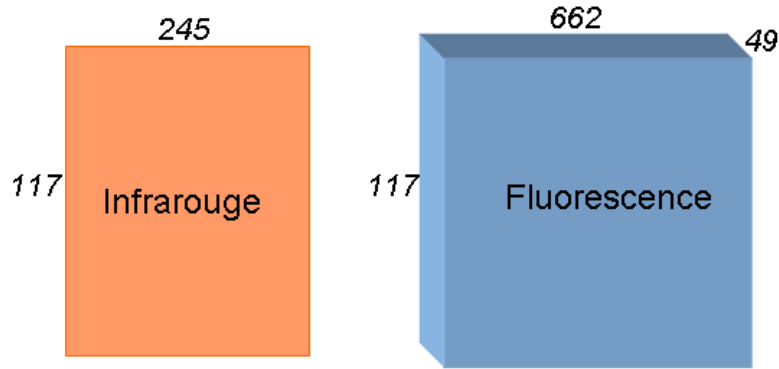


Figure C31 Structure de données de la collection 4f11mb-1

2 Appariement de la collection 4f11mb-2

2.1 Recalage d'images

Dans le cas de la collection 4f11mb-2, les images spectrales ont été recalées par rapport à l'image de référence acquise au microscope confocal de l'INRA de Nantes. Les facteurs d'échelle et les angles de rotation de l'image infrarouge étaient de 3,8 et 8° , respectivement. Les facteurs d'échelle pour les quatre images de fluorescence étaient de 0,8 et les angles de rotation étaient de -28° . Pour les images Raman, les facteurs d'échelles étaient de 0,8, 0,75, 0,8, 0,8, 0,8 et facteurs de rotation étaient de 1° , $1,8^\circ$, $1,2^\circ$, $1,2^\circ$, 1° . La quatrième image Raman a un facteur d'échelle et un angle de rotation plus petit que les autres images. Ceci peut être dû à sa petite taille, comparée aux autres images de la collection. Le rapport entre les facteurs d'échelles obtenus est égal à 5 en arrondi. Ce facteur correspond exactement au rapport de résolution théorique entre l'image spectrale infrarouge d'une part et de fluorescence et Raman d'autre part. La figure C32 présente la fusion de l'image de référence avec chacune des images spectrales de la collection 4f11mb-2 après recalage. Les images de fusion ont été calculées comme des images RGB en superposant l'image visible dans le canal rouge et l'image spectrale dans le canal vert et bleu.

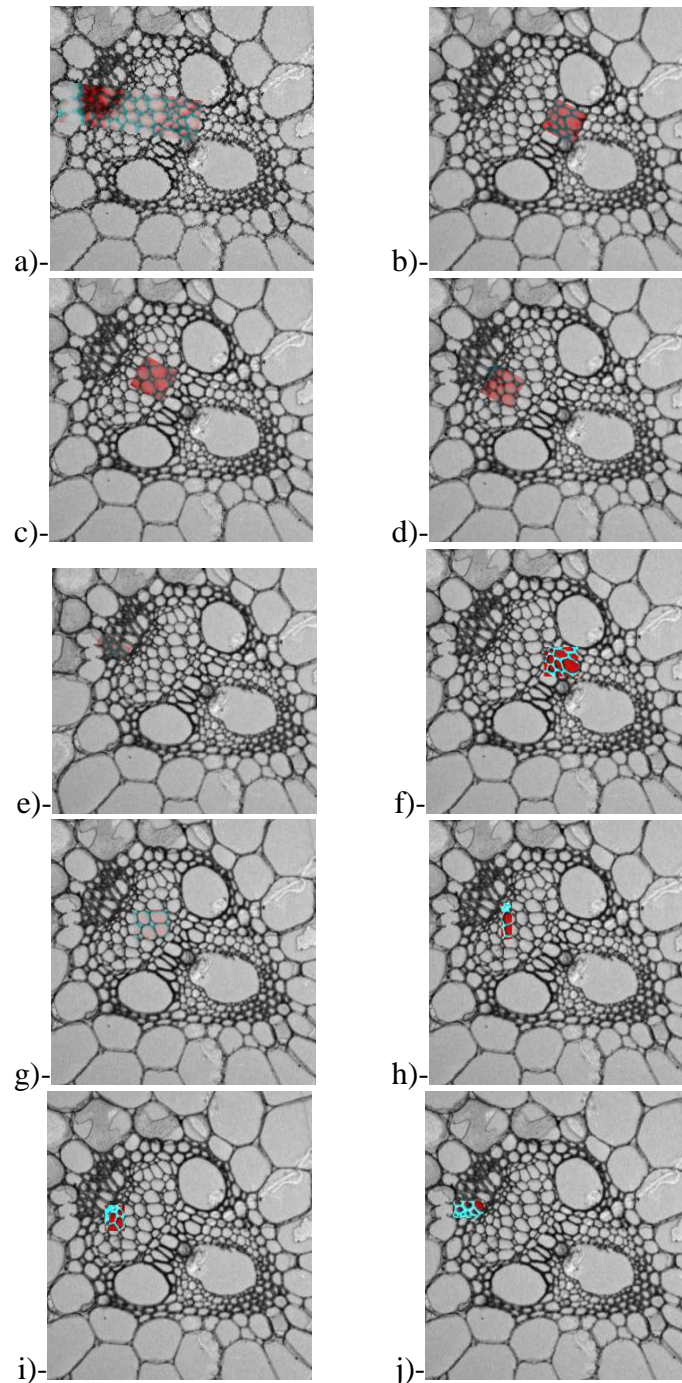


Figure C32 Fusion de l'image visible de référence avec : a)- l'image spectrale infrarouge (intensité de la de la région entre $1800-800\text{ cm}^{-1}$), b, c, d et e)- l'image spectrale de fluorescence (somme des intensités spectrales entre $360-540\text{ nm}$), f, g, h, i, j) l'image spectrale Raman (intensité de la de la région entre $1800-800\text{ cm}^{-1}$). Les images de fusion ont été calculées comme des images RGB avec le canal rouge = image visible, le canal vert et bleu = canaux intensités spectrales.

2.2 Structure de données

La structure finale des données est un ensemble de trois tableaux (Fig. C32) :
 - Un tableau à deux entrées des données infrarouge avec une entrée spatiale et une entrée spectrale (première entrée: 194 pixels et la deuxième entrée: 245 nombre d'ondes, respectivement).

- Un tableau de données à trois entrées de fluorescence (première entrée: 194 pixels, deuxième entrée: 322 longueurs d'ondes et troisième entrée: 25 pixels).
- Un tableau à trois entrées des données Raman (première entrée: 194 pixels, deuxième entrée: 346 nombres d'onde et troisième entrée: 25 pixels).

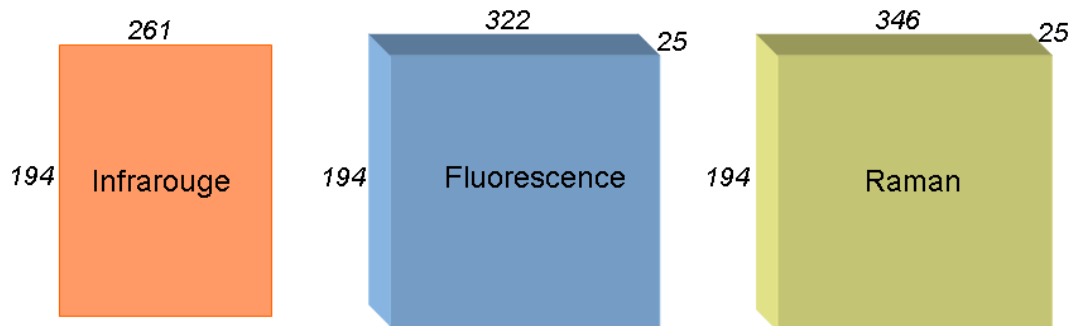


Figure C 33 Structure de données de la collection 4f11mb-2

Conclusion

Les outils indispensables pour l'appariement pertinent des collections d'images acquises ont été développés afin d'assurer l'analyse conjointe des données. Dans ce cadre, nous avons mis en place une méthode générique pour le recalage affine d'images hyperspectrales adaptée à nos données. Cette méthode est invariante aux translations, rotations et aux changements d'échelles. Elle est basée sur la combinaison d'une technique multirésolution 'les pyramides', une recherche dichotomique et la corrélation croisée. Une fois les images recalées sur un espace de référence, la région commune pour laquelle des spectres ont été acquis pour les trois méthodes de microspectroscopie est calculée en utilisant des techniques d'analyse d'images. L'étape suivante était d'organiser la structure de données des spectres appariés, tout en préservant l'ensemble des informations acquises. Les pixels infrarouge, fluorescence et Raman sont hétérogènes en terme de taille et de résolution. Chaque spectre infrarouge a été apparié à un ensemble de spectres de fluorescence et Raman. La structure de données obtenue contient un tableau de données infrarouge à deux entrées couplé à deux tableaux à trois entrées de fluorescence et Raman. Une telle structure de données ne peut être analysée directement avec un modèle de décomposition multibloc classique. De nouvelles approches ont été proposées et développées dans le chapitre suivant.

Partie D :

Couplage des images hyperspectrales en utilisant les méthodes multibloc

Sommaire

Introduction	131
I. Couplage de deux images hyperspectrales multirésolutions : extension de l'analyse inter-batterie de Tucker.....	132
Introduction	132
1 Analyse de tableaux multiples : Méthodes.....	132
1.1 Analyse inter-batterie de Tucker	133
1.2 Extension de l'analyse inter-batterie de Tucker pour l'analyse des tableaux hétérogènes.....	134
1.3 Propriétés.....	136
1.4 Outils d'interprétation.....	137
2 Résultats et discussions	137
2.1 Données analysés	137
2.2 Analyse inter-batterie de Tucker	138
2.2 Analyse inter-batterie de Tucker à trois entrées.....	142
Conclusion.....	145
II Couplage de trois images hyperspectrales par l'analyse de co-inertie multiple.....	145
Introduction	145
1 Analyse de co-inertie Multiple : Méthodes et outils d'interprétation	146
1.1 Brève présentation de l'analyse de co-inertie multiple	146
1.2 Outils d'interprétation.....	149
1.2.1 Contribution des tableaux de données à la structure commune.....	149
1.2.2 Reconstitution de tableaux de données à partir des composantes.....	149
1.2.3 Covariances et corrélations entre les composantes par bloc et les composantes communes.....	149
1.2.4 Corrélation entre la composante globale et les composantes par bloc	150
1.2.5 Outils d'interprétation spatiaux.....	150

2 Résultats et discussions	151
2.1 Données analysées.....	151
2.2 Pourcentage de variances et contributions des tableaux de données	152
2.3 Interprétation spectrale et spatiale.....	154
2.3.1 Composante 1.....	154
2.3.2 Composante 2.....	157
2.3.3 Composante 3.....	160
Conclusion.....	162
III. Couplage de trois images hyperspectrales multirésolutions : extension de l’analyse de co-inertie multiple	163
Introduction	163
1 Extension de l'analyse de co-inertie multiple à l'analyse de tableaux ayant un nombre d'entrées hétérogènes.....	163
1.1 Algorithme	163
1.2 Propriétés.....	165
1.3 Outils d'interprétation.....	167
1.3.1 Contribution des tableaux de données à la structure commune.....	167
1.3.2 Covariances entre composantes par bloc et composantes communes	167
1.3.3 Corrélation entre les composantes.....	168
2 Analyse de l'échantillon 4f1 1mb-2 en utilisant l'analyse de co-inertie multiple trinéaire	168
2.1 Pondération des tableaux de données.....	168
2.2 Analyse de la collection de données 4f1 1mb-2.....	170
2.3 Interprétation spectrale et spatiale.....	171
2.3.1 Composante 1.....	172
2.3.2 Composante 2.....	174
2.3.3 Composante 3.....	176
2.3.4 Composante 4.....	178
2.4 Interprétation de la troisième entrée des blocs à trois entrées.....	180
Conclusion.....	181
Conclusion.....	181

Introduction

Le couplage des images hyperspectrales nécessite le développement de méthodes depuis l'acquisition jusqu'à l'analyse et l'interprétation en passant par une étape de conditionnement de données. L'étape de conditionnement comporte le pré-traitement et l'appariement des spectres. Le protocole d'acquisition conjointe ainsi que le pré-traitement ont été décrits dans la partie B. La procédure d'appariement a été établie dans la partie C. La structure de données obtenue comporte un ensemble de tableaux de données hétérogènes à deux et à trois entrées. Elle ne peut pas être analysée avec des méthodes d'analyse classiques. La différence entre le nombre d'entrées des tableaux de données est due à la différence de taille des pixels des images hyperspectrales. L'idée naturelle pour coupler des images hyperspectrales multirésolutions est de les considérer à une même résolution -la plus basse-. Dans ce cas, la structure de données ne contient que des tableaux à deux entrées qui peuvent être analysés avec des méthodes comme PLS2 [Wold *et al.*, 2001]. Sur la base de ces méthodes, les approches tenant en compte la différence de résolution ont été développées. Dans une première étape, la collection d'images 4f11mb-1 a été analysée en utilisant l'analyse inter-batterie de Tucker [Tucker, 1958]. Outre sa robustesse, cette méthode symétrique et orthogonale permet une analyse simultanée et une interprétation assez simple d'un couple de tableaux à deux entrées. Le modèle de décomposition inter-batterie de Tucker peut être étendu si l'un des deux tableaux est à trois entrées. L'approche a été développée en considérant le tableau à trois entrées comme étant un ensemble de tableaux à deux entrées. Pour analyser plus de deux tableaux de données, l'analyse de co-inertie multiple a été sélectionnée parmi les méthodes existantes [Chessel et Hanafi 1996 ; Hanafi *et al.*, 2011]. Comme l'analyse inter-batterie de Tucker, l'analyse de co-inertie multiple présente l'avantage d'être symétrique et orthogonale. L'analyse de co-inertie multiple a également été étendue pour s'adapter à la structure de données. Cette méthode a été illustrée en utilisant le jeu de données 4f11mb-2.

Dans cette partie, les approches retenues et développées pour coupler des images hyperspectrales multirésolutions sont présentées. Des outils d'interprétation et de visualisation sont proposés. Cette partie est organisée comme suit : le chapitre 1 est dédié au couplage de deux images hyperspectrales en utilisant l'extension de l'analyse inter-batterie de Tucker. Ce chapitre contient toutes les analyses relatives au jeu de données 4f11mb-1. Le deuxième chapitre a concerné le couplage de plusieurs images hyperspectrales en utilisant l'analyse de co-inertie multiple en considérant toutes les images à la plus basse résolution. Cette méthode a servi pour analyser le jeu de données 4f11mb-2. Dans le troisième chapitre, l'extension de

l'analyse de co-inertie multiple a été développée pour pouvoir analyser la structure de données sans altération de résolution. Dans ce cas, tous les tableaux à trois entrées sont considérés comme un ensemble de données à deux entrées. Cette méthode a été appliquée aux données de l'échantillon 4f11mb-2.

I Couplage de deux images hyperspectrales multirésolutions : extension de l'analyse inter-batterie de Tucker

Introduction

L'objectif de ce chapitre est de présenter l'analyse conjointe de deux images hyperspectrales de résolutions différentes. Après appariement et organisation des données, nous devons analyser deux tableaux de données ayant un nombre hétérogène d'entrées, dans notre cas deux et trois entrées respectivement. L'approche naturelle pour analyser une telle structure de données, est de moyennner le tableau à trois entrées selon la dimension spectrale pour obtenir un tableau à deux entrées (Fig. D1).

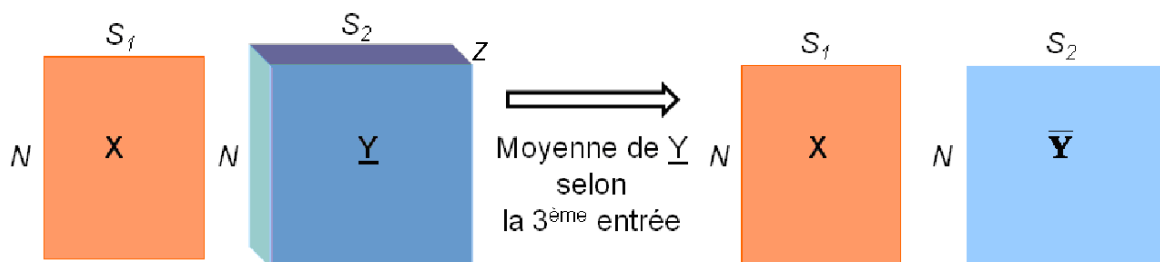


Figure D1 Moyennner un tableau à trois entrées selon la troisième entrée. La structure de données obtenue consiste en deux tableaux à deux entrées.

Dans ce cas, deux méthodes peuvent être appliquées : la PLS2 ou l'analyse inter batterie de Tucker. L'analyse inter-batterie de Tucker a été retenue préférablement à PLS2 afin de préserver le rôle symétrique de chaque domaine spectral. La méthode a été étendue pour analyser la structure de données en conservant le nombre hétérogène de modes.

1 Analyse de tableaux multiples : Méthodes

Dans cette section, l'analyse inter-batterie et son extension sont présentées en détails. Elles ont été respectivement appliquées à la collection 4f11mb-1 en faisant la moyenne selon la dimension spectrale et en gardant la structure de données sans altération.

Ce travail a fait l'objet d'une publication [Allouche *et al.*, 2012] et d'une conférence orale au congrès "Les Journées de Statistique".

1.1 Analyse inter-batterie de Tucker

L'analyse inter-batterie de Tucker [Tucker, 1958], également appelée l'analyse de co-inertie en littérature française [Chessel et Hanafi, 1996], est une méthode multivariée qui vise à relier deux tableaux de données à deux entrées appariés selon le mode des individus. Notons \mathbf{X} et \mathbf{Y} les tableaux de données à deux entrées centrés selon les variables spectrales.

Les valeurs des composantes $\mathbf{c}_X = \mathbf{X}\mathbf{u}$ ($\|\mathbf{u}\|=1$) et $\mathbf{c}_Y = \mathbf{Y}\mathbf{v}$ ($\|\mathbf{v}\|=1$) sont calculées en maximisant la covariance au carré :

$$\text{cov}^2(\mathbf{c}_X, \mathbf{c}_Y) \quad \text{Equation D1}$$

Les premières composantes $\mathbf{c}_X^{(1)}$ et $\mathbf{c}_Y^{(1)}$ sont similaires à celles obtenues à l'aide de la méthode régression aux moindres carrés partiels PLS2. Dans le cas de la méthode PLS2, à une étape donnée h ($1 < h \leq H$), les composantes $\mathbf{c}_X^{(h)}$ et $\mathbf{c}_Y^{(h)}$ sont obtenues en faisant une déflation sur les valeurs des composantes $\mathbf{c}_X^{(h)}$ du tableau des variables explicatives \mathbf{X} .

$$\mathbf{X}^{(h)} = \mathbf{X}^{(h-1)} - \frac{\mathbf{c}_X^{(h-1)} \mathbf{c}_X^{(h-1)'}}{\|\mathbf{c}_X^{(h-1)}\|^2} \mathbf{X}^{(h-1)} \quad \text{Equation D2}$$

$$\mathbf{Y}^{(h)} = \mathbf{Y}^{(h-1)} - \frac{\mathbf{c}_X^{(h-1)} \mathbf{c}_X^{(h-1)'}}{\|\mathbf{c}_X^{(h-1)}\|^2} \mathbf{Y}^{(h-1)} \quad \text{Equation D3}$$

Dans l'analyse inter-batterie de Tucker, \mathbf{X} et \mathbf{Y} sont déflatés par rapport à leurs profils variables respectifs $\mathbf{u}^{(h)}$ et $\mathbf{v}^{(h)}$.

$$\mathbf{X}^{(h)} = \mathbf{X}^{(h-1)} - \mathbf{c}_X^{(h-1)} \mathbf{u}^{(h-1)'} \quad \text{Equation D4}$$

$$\mathbf{Y}^{(h)} = \mathbf{Y}^{(h-1)} - \mathbf{c}_Y^{(h-1)} \mathbf{v}^{(h-1)'} \quad \text{Equation D5}$$

Ce qui fait de l'analyse inter-batterie de Tucker une méthode symétrique par rapport à PLS2. Dans la pratique, les matrices \mathbf{U} et \mathbf{V} sont calculées en utilisant la décomposition en valeurs singulières de la matrice de variance co-variance $\mathbf{X}'\mathbf{Y}$ entre les deux tableaux \mathbf{X} et \mathbf{Y} , tous les vecteurs propres étant calculés simultanément :

$$\text{svd}(\mathbf{X}'\mathbf{Y}) = \mathbf{U}\mathbf{L}\mathbf{V}' \quad (\mathbf{U}\mathbf{U}'=\mathbf{I}, \mathbf{V}\mathbf{V}'=\mathbf{I}) \quad \text{Equation D6}$$

où \mathbf{L} est une matrice diagonale. Le $h^{(\text{ème})}$ élément diagonal de \mathbf{L} ou la $h^{(\text{ème})}$ valeur singulière est égale à la covariance entre les composantes calculées $\text{cov}(\mathbf{c}_X^{(h)}, \mathbf{c}_Y^{(h)})$, \mathbf{I} est la matrice

identité. Dans le cas de PLS2, les profils variables ne peuvent pas être calculés simultanément. Une approche itérative est donc nécessaire [Wold *et al.*, 1987].

1.2 Extension de l'analyse inter-batterie de Tucker pour l'analyse des tableaux hétérogènes

Dans cette section, l'analyse inter-batterie de Tucker a été étendue dans le cas où \underline{Y} est un tableau de données à trois entrées.

Notons \underline{X} et \underline{Y} les deux tableaux de données à deux et à trois entrées centrés selon la première entrée, respectivement, et Z la troisième entrée du tableau de données \underline{Y} .

Le centrage d'un tableau à trois entrées \underline{Y} se fait selon la première entrée. Pour chaque colonne du tableau de données de départ, la valeur moyenne est calculée et soustraite de chaque élément de cette colonne. Un cube de valeurs moyennes \underline{Y}_c est alors obtenue (Fig. D2).

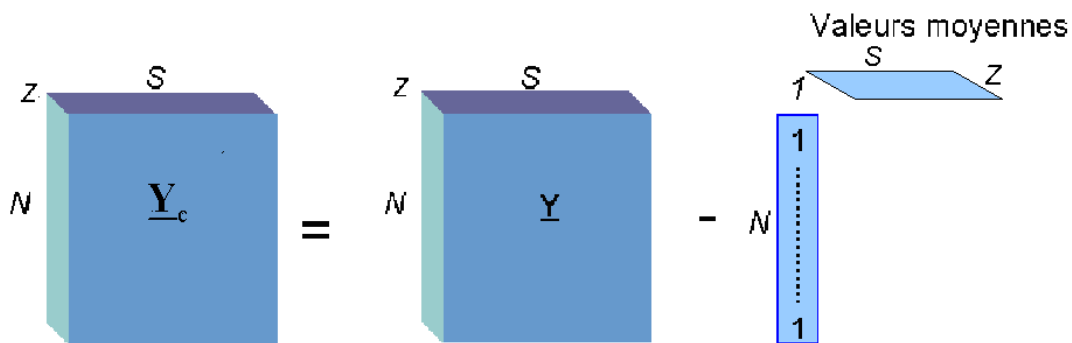


Figure D2 Centrage d'un tableau de données à trois entrées.

Un tableau de données à trois entrées peut être considéré comme un ensemble de Z de tableaux de données à deux entrées $\underline{Y}_{(z)}$ appelés tranches (Fig. D3).

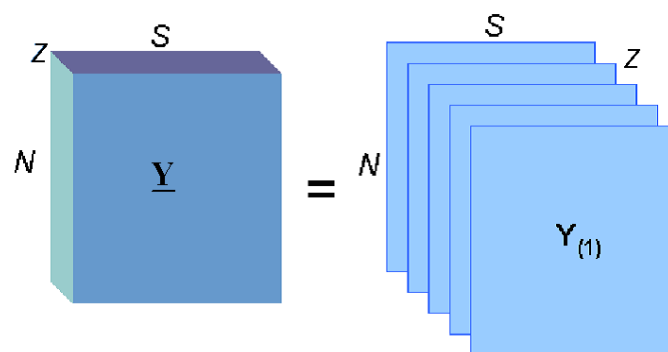


Figure D3 Représentation d'un tableau à trois entrées en un ensemble de tableaux à deux entrées

Comme dans l'analyse inter-batterie de Tucker, les profils variables \mathbf{u} et \mathbf{v} sont recherchés afin de maximiser la somme des covariances au carré entre les composantes $\mathbf{c}_x = \mathbf{X}\mathbf{u}$ ($\|\mathbf{u}\|=1$) et $\mathbf{c}_{Y(z)} = \mathbf{Y}_{(z)}\mathbf{v}$ ($\|\mathbf{v}\|=1$) ; ($z = 1, 2, \dots, Z$) pour chaque tableau ce qui revient à maximiser le critère suivant :

$$\text{Max} \sum_{z=1}^Z \text{cov}^2(\mathbf{c}_x, \mathbf{c}_{Y(z)}) \text{ sous les contraintes } \|\mathbf{u}\|=1 \text{ et } \|\mathbf{v}\|=1 \quad \text{Equation D7}$$

Notons que pour tous les $\mathbf{Y}_{(z)}$ il y a un seul profil variable \mathbf{v} .

La procédure commence en tirant au sort un couple arbitraire de vecteurs ($\mathbf{u}^{(0)}, \mathbf{v}^{(0)}$). Les covariances entre \mathbf{c}_x et chaque $\mathbf{c}_{Y(z)}$ sont calculées. La somme pondérée des matrices $\mathbf{Y}_{(z)}$ est calculée en considérant les covariances entre \mathbf{c}_x et $\mathbf{c}_{Y(z)}$ comme coefficient de pondération. Appelons $\bar{\mathbf{Y}}$ cette somme pondérée. La décomposition en valeurs singulières est effectuée sur la matrice $\mathbf{X}'\bar{\mathbf{Y}}$. La première paire de vecteurs propres est extraite. Ces étapes sont répétées jusqu'à ce que la convergence soit atteinte. L'annexe A établit les propriétés de monotonie de cette procédure itérative qui montre la convergence de l'algorithme. La procédure se résume comme suit :

1. Tirer au sort un couple de vecteurs ($\mathbf{u}=\mathbf{u}^{(0)}, \mathbf{v}=\mathbf{v}^{(0)}$) (avec $\|\mathbf{u}^{(0)}\|=1$ et $\|\mathbf{v}^{(0)}\|=1$)
2. Calculer les composantes $\mathbf{c}_x = \mathbf{X}\mathbf{u}$ et $\mathbf{c}_{Y(z)} = \mathbf{Y}_{(z)}\mathbf{v}$ avec $Z = 1, 2, \dots, Z$
3. Calculer les covariances $\alpha_{(z)} = \text{cov}^2(\mathbf{c}_x, \mathbf{c}_{Y(z)})$ avec $z = 1, 2, \dots, Z$
4. Calculer la somme pondérée du tableau à trois entrées $\bar{\mathbf{Y}} = \sum_{z=1}^Z \alpha_{(z)} \mathbf{Y}_{(z)}$
5. Réaliser une décomposition en valeurs singulières $\mathbf{X}'\bar{\mathbf{Y}} = \mathbf{U}\mathbf{L}\mathbf{V}'$
6. Considérer $\mathbf{u}=\mathbf{u}^{(1)}$ = la première colonne de \mathbf{U}

$\mathbf{v}=\mathbf{v}^{(1)}$ = la première colonne de \mathbf{V}

7. Itérer la procédure depuis l'étape 2 jusqu'à $\|\mathbf{u}^{(1)} - \mathbf{u}^{(0)}\| + \|\mathbf{v}^{(1)} - \mathbf{v}^{(0)}\| < \text{epsilon}$

$\mathbf{u}^{(1)}$ et $\mathbf{v}^{(1)}$ sont conservés comme étant les premiers profils variables. Les composantes des tableaux \mathbf{X} et $\underline{\mathbf{Y}}$ sont calculées par :

$$\mathbf{c}_x^{(1)} = \mathbf{X}\mathbf{u}^{(1)} \quad \text{Equation D8}$$

$$\mathbf{c}_{\mathbf{Y}_{(z)}}^{(l)} = \mathbf{Y}_{(z)} \mathbf{v}^{(l)} \quad \text{Equation D9}$$

Les profils variables \mathbf{u} et \mathbf{v} et des composantes \mathbf{c}_x et $\mathbf{c}_{\mathbf{Y}_{(z)}}$ pour les dimensions supérieures sont obtenus après déflation des tableaux \mathbf{X} et $\mathbf{Y}_{(z)}$ par rapport à leurs profils variables respectifs.

$$\mathbf{X}^{(h)} = \mathbf{X}^{(h-1)} - \mathbf{c}_x^{(h-1)} \mathbf{u}^{(h-1)} \quad \text{Equation D10}$$

$$\mathbf{Y}_{(z)}^{(h)} = \mathbf{Y}_{(z)}^{(h-1)} - \mathbf{c}_{\mathbf{Y}_{(z)}}^{(h-1)} \mathbf{v}^{(h-1)} \quad \text{avec } z = 1, 2 \dots Z \quad \text{Equation D11}$$

1.3 Propriétés

Propriété 1

Les profils variables $\mathbf{u}^{(h)}$ et $\mathbf{v}^{(h)}$ sont orthogonaux [Chessel et Hanafi, 1996].

Propriété 2

Les composantes $\mathbf{c}_x^{(h)}$ et $\mathbf{c}_{\mathbf{Y}_{(z)}}^{(h)}$ ne sont pas orthogonales.

Propriété 3

Les composantes $\mathbf{c}_x^{(h)}$ et les profils variables $\mathbf{u}^{(h)}$ ($1 \leq h \leq H ; 1 \leq z \leq Z$) ; permettent la reconstitution du tableau de données \mathbf{X} comme suit :

$$\mathbf{X} = \sum_{h=1}^H \mathbf{c}_x^{(h)} \mathbf{u}^{(h)} + \mathbf{R}_x^{(h)} \quad \text{Equation D12}$$

où $\mathbf{R}_x^{(h)}$ est le résidu du tableau \mathbf{X} .

Les composantes $\mathbf{c}_{\mathbf{Y}_{(z)}}^{(h)}$ et les profils variables $\mathbf{v}^{(h)}$ ($1 \leq h \leq H ; 1 \leq z \leq Z$) ; permettent la reconstitution de chaque tranche $\mathbf{Y}_{(z)}$ du tableau de données \mathbf{Y} comme suit :

$$\mathbf{Y}_{(z)} = \sum_{h=1}^H \mathbf{c}_{\mathbf{Y}_{(z)}}^{(h)} \mathbf{v}^{(h)} + \mathbf{R}_{\mathbf{Y}_{(z)}}^{(h)} \quad \text{Equation D13}$$

où $\mathbf{R}_{\mathbf{Y}_{(z)}}^{(h)}$ est le résidu du tableau $\mathbf{Y}_{(z)}$.

Propriété 4

La variance du tableau de données à deux entrées \mathbf{X} décrite par les composantes $\mathbf{c}_x^{(h)}$ est égale à :

$$\sum_{h=1}^H \mathbf{c}_X^{(h)} \mathbf{c}_X^{(h)} \quad \text{Equation D14}$$

La variance du tableau de données à trois entrées \mathbf{Y} décrite par les composantes $\mathbf{c}_{Y(z)}^{(h)}$ est la somme des variances de chaque composante $\mathbf{c}_{Y(z)}^{(h)}$ des tranches $\mathbf{Y}_{(z)}$:

$$\sum_{z=1}^Z \sum_{h=1}^H \mathbf{c}_{Y(z)}^{(h)} \mathbf{c}_{Y(z)}^{(h)} \quad \text{Equation D15}$$

1.4 Outils d'interprétation

L'analyse inter-batterie de Tucker et son extension ont été conçues pour décrire les covariances entre deux tableaux de données. Les profils variables \mathbf{u} et \mathbf{v} des deux tableaux sont interprétés pour révéler les variations spectrales. Les composantes \mathbf{c}_X et $\mathbf{c}_{Y(z)}$ sont repliées en utilisant l'image de la région commune totale (Fig. C23 et C25) et affichées comme image composante pour révéler les variations spatiales. Des indicateurs sont nécessaires pour étudier l'information expliquée par chaque tableau de données et pour évaluer l'information commune apportée par l'analyse conjointe. Les variances expliquées par les composantes permettent de révéler l'information fournie par chaque tableau dans l'analyse. Les pourcentages de covariance entre les composantes à une dimension h donnée $\text{cov}^2(\mathbf{c}_X^{(h)}, \mathbf{c}_Y^{(h)})$ ainsi que les coefficients de corrélation $\text{corr}(\mathbf{c}_X^{(h)}, \mathbf{c}_Y^{(h)})$ entre les composantes sont calculées pour évaluer les relations entre les composantes de chaque tableau.

2 Résultats et discussions

2.1 Données analysées

La collection d'images 4f11mb-1 est constituée d'une image infrarouge et 3 images de fluorescence. Ces images contiennent trois types cellulaires : sclérenchyme, phloème et parenchyme. Les spectres infrarouge étaient prétraités en utilisant la dérivée seconde de Savitzky Golay et normalisé à 1. Les spectres de fluorescence étaient normalisés en utilisant la normalisation spatiale et spectrale avec des tailles de fenêtres de $11 \times 11 \mu\text{m}^2$. Après appariement de spectres, la structure de données établie contient un tableau à deux entrées et un tableau à trois entrées (Fig. C31) : un tableau à deux entrées \mathbf{X}_1 de spectres infrarouge de dimensions 194×261 et un tableau à trois entrées \mathbf{X}_2 de spectres de fluorescence de dimensions $194 \times 322 \times 49$.

La structure de données obtenue ne peut pas être soumise directement à l'analyse inter-batterie de Tucker. Dans une première étape, le tableau à trois entrées a été moyenné selon la troisième entrée pour obtenir deux tableaux à deux entrées et l'analyse inter-batterie de Tucker a été appliquée de manière classique. Dans une seconde étape, l'analyse inter-batterie a été étendue en considérant le tableau à trois entrées comme un ensemble de tableaux à deux entrées.

2.2 Analyse inter-batterie de Tucker

Les 49 tranches du tableau de données à trois entrées de fluorescence ont été moyennées. La matrice de covariance entre le tableau de données infrarouge et de fluorescence a été soumise à une décomposition en valeurs singulières. Le pourcentage de variance expliquée des composantes, le pourcentage de covariance totale des composantes ainsi que les coefficients de corrélations ont été calculés (Tableau D1, D2 et D3).

Tableau D1 Résultats du tableau des données infrarouge. La première ligne représente les pourcentages de variance de chaque composante principale. La deuxième ligne représente les pourcentages de variance de chaque composante de l'inter-batterie de Tucker.

Composantes	1	2	3	4	5
pourcentage de variance de chaque composante principale	56.07	19.97	4.38	3.81	3.57
pourcentage de variance de chaque composante de l'analyse inter-batterie de Tucker	45.45	24.97	7.06	3.88	2.68

Tableau D2 Résultats du tableau des données de fluorescence. La première ligne représente les pourcentages de variance de chaque composante principale. La deuxième ligne représente les pourcentages de variance de chaque composante de l'inter-batterie de Tucker.

Composantes	1	2	3	4	5
pourcentage de variance de chaque composante principale	88.82	11.07	0.07	0.02	0.01
pourcentage de variance de chaque composante	88.43	11.42	0.09	0.04	0.01

Deux composantes correspondant à 90 et 10% de la covariance totale ont été examinées. Les profils variables et les images composantes ont été calculés pour chacun des domaines spectraux (Fig. D4 et D5). Les composantes obtenues en analyse inter-batterie de Tucker ont permis de calculer le pourcentage de variance décrite par le modèle. Les composantes de fluorescence correspondaient à 88 et 11% de la variance totale et les composantes infrarouge correspondaient à 45 et 25%. Dans le cas de la fluorescence, 99% de la variabilité a été décrite par les deux composantes. Les composantes de fluorescence sont très semblables à celles obtenues par analyse en composantes principales. Dans le cas des données infrarouge, 30% de

la variance n'a pas été décrite par le modèle qui pourrait correspondre à des variations spectrales non corrélées avec les données de fluorescence. Les coefficients de corrélation entre les composantes infrarouge et de fluorescence sont relativement faibles montrant que les composantes de l'analyse inter-batterie décrivent la variance des tableaux plutôt que leur ressemblance.

Tableau D3 Indicateurs de couplage entre le tableau de données infrarouge et de fluorescence. La première ligne représente la covariance entre les composantes de l'analyse inter-batterie de Tucker. La deuxième ligne représente la corrélation entre les composantes de l'analyse inter-batterie de Tucker.

Composantes	1	2	3	4	5
pourcentage de la covariance	89.67	10.30	0.02	0.00	0.00
coefficient de corrélation	0.39	0.50	0.48	0.32	0.25

Notre démarche d'interprétation va suivre le schéma suivant :

- interpréter les profils variables les plus "simples", dans notre cas les profils variables de fluorescence.
- rechercher les liens entre les profils variables infrarouge et de fluorescence homologues.
- utiliser les critères calculés tel que la variance et la corrélation comme indicateurs de la qualité de lecture.

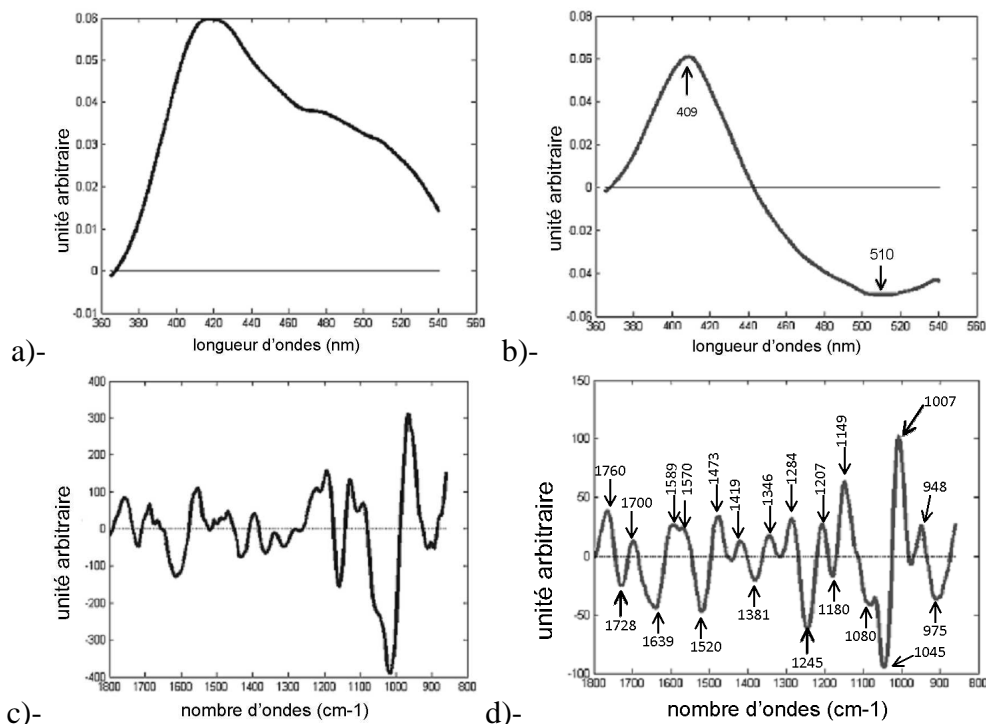


Figure D4 Profils spectraux 1et 2 : a-b)- de fluorescence, c-d)- infrarouge.

Les images composantes (Fig. D5) ont été obtenues en combinant l'image visible de référence avec les composantes repliées tout en préservant leur résolution spatiale réelle. L'image composante a été affichée en fausse couleur: les pixels en bleu correspondent aux valeurs les plus basses des composantes. Les pixels en rouge correspondent aux valeurs les plus grandes des composantes. Cette représentation permet l'interprétation des variations d'intensité en fonction de leur localisation spatiale et donc en fonction des types cellulaires, phloème, sclérenchyme ou parenchyme (Fig. D5).

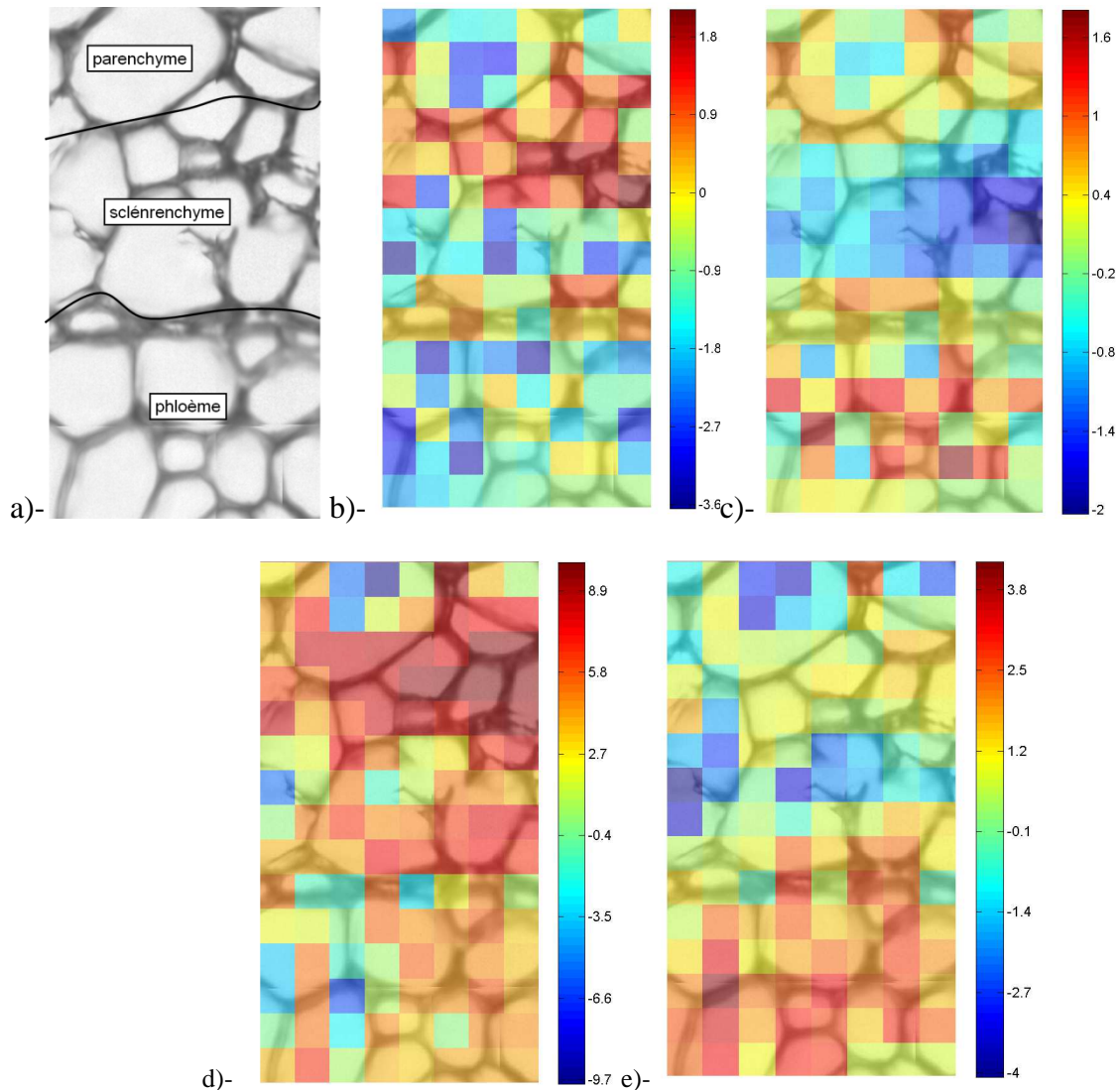


Figure D5 Images composantes obtenues en utilisant l'analyse inter-batterie de Tucker : a)- image fond clair de la région sélectionnée b-c)- images composantes 1 et 2 de fluorescence, d-e) images composantes infrarouge 1 et 2. Ces images ont été obtenues en superposant l'image fond clair avec l'image composante repliée avec un coefficient de transparence égal à 0,5. La carte de couleur 'Jet' de Matlab a été appliquée pour les composantes après ajustement des valeurs entre 0 et 255 pour la visualisation. Les barres d'échelles indiquent les valeurs minimales et maximales réelles des valeurs des composantes.

Le premier profil spectral de fluorescence $\mathbf{v}^{(1)}$ (Fig. D4-a) décrit les variations d'intensité entre les spectres. L'image composante associée $\mathbf{c}_Y^{(1)}$ (Fig. D5-a) a révélé la position où l'intensité maximale de fluorescence a été observée. Malgré la faible résolution, la composante de

fluorescence $\mathbf{c}_Y^{(1)}$ oppose des pixels avec des signaux faibles à l'intérieur des cellules à des pixels appartenant aux parois cellulaires, où la fluorescence a eu lieu. Le profil spectral infrarouge $\mathbf{u}^{(1)}$ associé (Fig. D4-c) ne correspond pas à un profil d'intensité. En effet, les spectres de fluorescence et infrarouge ont été prétraités différemment. Dans le cas de la fluorescence, la normalisation spectrale a été appliquée en tenant compte du voisinage, afin de préserver la structure des parois cellulaires et d'éviter la normalisation des régions où il n'y a aucun signal. Dans le cas des spectres infrarouge, nous avons considéré qu'il n'y avait pas de régions ne comportant aucun signal en raison de la taille du pixel. Les spectres infrarouge ont été prétraités pour avoir tous la même intensité globale. Les principales variations du premier profil spectral infrarouge $\mathbf{u}^{(1)}$ (Fig. D4-c) ont été observées dans la région 1200-800 cm^{-1} correspondant à la région des sucres. Les pixels intenses ont été localisés dans la région des petites cellules de sclérenchyme. Le coefficient de corrélation entre les composantes $\mathbf{c}_X^{(1)}$ et $\mathbf{c}_Y^{(1)}$ est égal à 0.39 montrant que cette première composante décrit la variance des tableaux plutôt que leur ressemblance. En effet, l'analyse inter-batterie de Tucker est basée sur la covariance. La première composante de fluorescence décrit une variance de 88%. Cette composante est alors due à un effet d'intensité de fluorescence.

Le second profil spectral de fluorescence $\mathbf{v}^{(2)}$ (Fig. D 4-b) oppose un pic positif à 409 nm à des longueurs d'ondes négatives entre 450 nm et 550 nm. Le profil spectral a été comparé aux spectres des trois composés de référence - l'acide férulique, para-coumarique et la lignine (Fig. B28). La deuxième composante $\mathbf{c}_Y^{(2)}$ a été interprétée comme une opposition de pixels avec une teneur forte en acides hydroxycinnamiques (acide férulique et acide para-coumarique) contre les pixels ayant un contenu plus élevé en lignine. L'image composante correspondante (Fig. D5-b) oppose des pixels dans les régions des cellules de sclérenchyme à des pixels dans la région du phloème et du parenchyme. Dans le deuxième profil spectral infrarouge $\mathbf{u}^{(2)}$ (Fig. D4-c). Des pics négatifs ont été observés à :

- 1045 et 1080 cm^{-1} qui ont été attribués aux xylanes.
- 1381,1520,1639 cm^{-1} qui ont été attribués aux lignines.
- 1245 cm^{-1} qui correspondent à des liaisons esters.

L'infrarouge est sensible aux sucres qui ne sont pas fluorescents. Les acides phénoliques sont en faible proportion par rapport à d'autres polymères [Vogel, 2008]. Ceci rend leur étude difficile en spectroscopie infrarouge. En assemblant les deux types d'informations, la conclusion suivante a pu être tirée : les xylanes sont co-localisés avec les lignines. Les acides hydroxycinnamiques étaient négativement corrélés aux lignines et aux xylanes.

Les images composantes de fluorescence $c_Y^{(2)}$ et infrarouge $c_X^{(2)}$ (Fig. D5-b et d) révèlent le phloème et la région du parenchyme en couleur rouge et la région du sclérenchyme en bleu. Avec cette interprétation, les acides hydroxycinnamiques ont été identifiés principalement dans le phloème et le parenchyme tandis que les cellules du sclérenchyme sont enrichies en lignine et en xylanes.

2.3 Analyse inter-batterie de Tucker à trois entrées

L'analyse a été faite en prenant en compte les pixels des imagerie c'est à dire la plus haute résolution. La covariance au carré entre le tableau de données infrarouge et chacune des 49 tranches du tableau de données à trois entrées de fluorescence est maximisée. Le tableau de fluorescence dépliée a été soumis à l'analyse en composantes principales (Tableau D5).

Tableau D4 Résultats du tableau de données infrarouge : pourcentages de variance des composantes de l'inter-batterie de Tucker.

Composantes	1	2	3	4	5
pourcentage de variance de chaque composante de l'analyse inter-batterie de Tucker	49.73	22.29	6.50	3.27	3.00

Tableau D5 Résultats du tableau de données de fluorescence. La première ligne représente les pourcentages de variance de chaque composante principale. Les pourcentages sont donnés afin de les comparer avec les pourcentages obtenus en analyse inter-batterie de Tucker.

Composantes	1	2	3	4	5
pourcentage de variance de chaque composante principale	97.66	2.22	0.03	0.01	0.01
pourcentage de variance de chaque composante de l'analyse inter-batterie de Tucker	88.83	8.80	0.01	0.07	0.05

Tableau D6 Indicateurs de couplage entre le tableau de données infrarouge et de fluorescence. La première ligne représente les covariances entre les composantes obtenues en analyse inter-batterie de Tucker.

Composantes	1	2	3	4	5
pourcentage de la covariance entre les deux tableaux	87.48	12.35	0.09	0.03	0.02

Deux composantes correspondant à 87% et 12% (Tableau D6) de la covariance totale ont été examinées. Les images composantes ont été calculées pour chaque domaine spectral, ce qui permet de calculer le pourcentage de variance décrit par le modèle.

Le calcul de coefficients de corrélation entre une composante à trois entrées et une composante à deux entrées n'est pas trivial. Ce travail a été sujet de développements ultérieurs.

Les composantes de fluorescence correspondaient à 89% et 9% de variance totale du tableau de fluorescence et les composantes infrarouges correspondaient à 49% et 22% (Tableau D5 et D4). Les variances des composantes de fluorescence ont été calculées comme la somme des variances obtenues pour chaque tranche du tableau de fluorescence. Les profils spectraux ont été calculés pour chaque domaine spectral (Fig. D6).

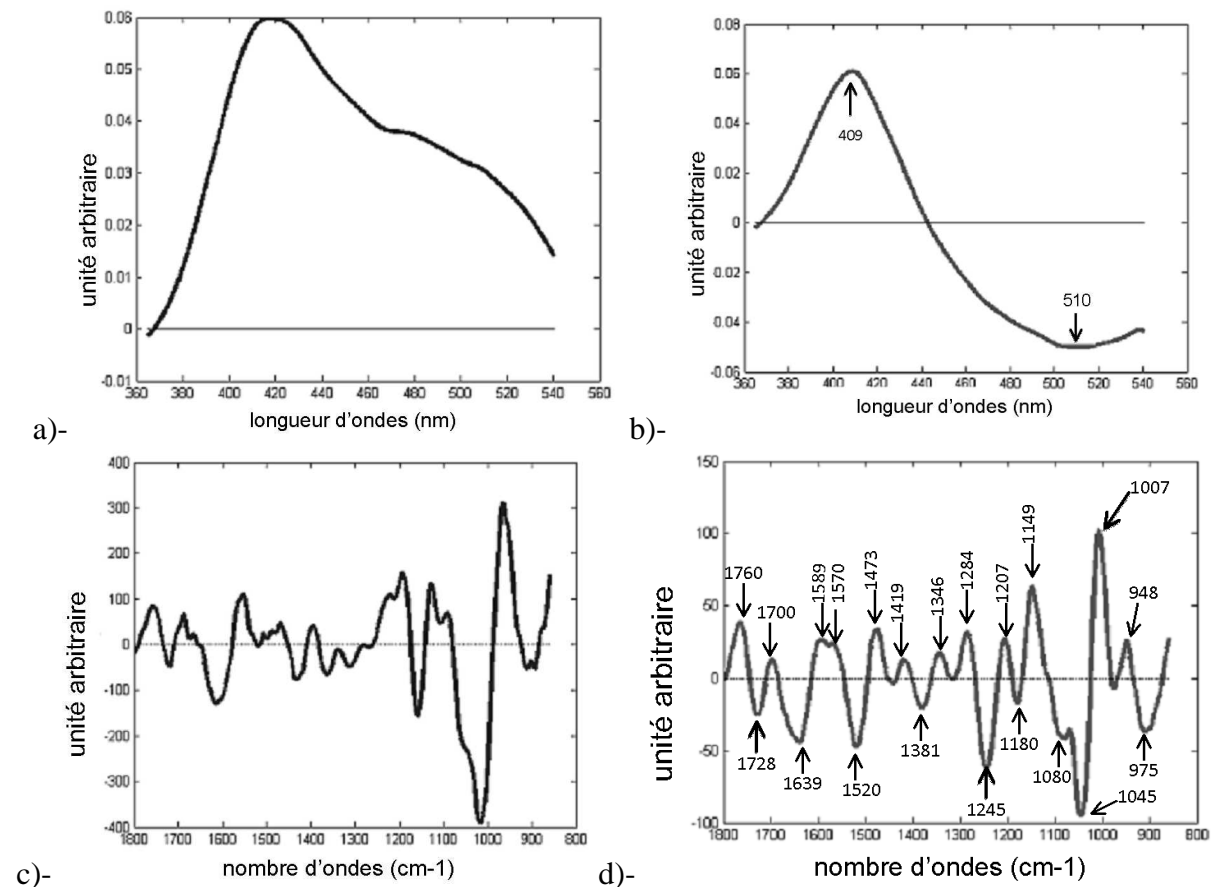


Figure D6 Profils spectraux 1et 2 : a-b)- de fluorescence, c-d)- infrarouge.

Les profils spectraux obtenus étaient identiques à ceux obtenus avec l'analyse inter-batterie de Tucker. Le premier profil spectral de fluorescence révèle des variations d'intensité dans les spectres et le deuxième oppose un pic positif révélant les acides hydroxycinnamiques à un pic négatif révélant la lignine. Par ailleurs, des pics négatifs ont été observés à 1639, 1520, 1045, 1080, et 1245 cm^{-1} dans le second profil spectral infrarouge (Fig. D6-d) comme cela a été déjà observé avec l'analyse inter-batterie de Tucker. Ainsi, la même interprétation spectrale a été

obtenue en utilisant les deux méthodes. En particulier, les images composantes infrarouge ont été identiques à celles obtenues avec l'analyse inter-batterie de Tucker (Fig. D7-c et d).

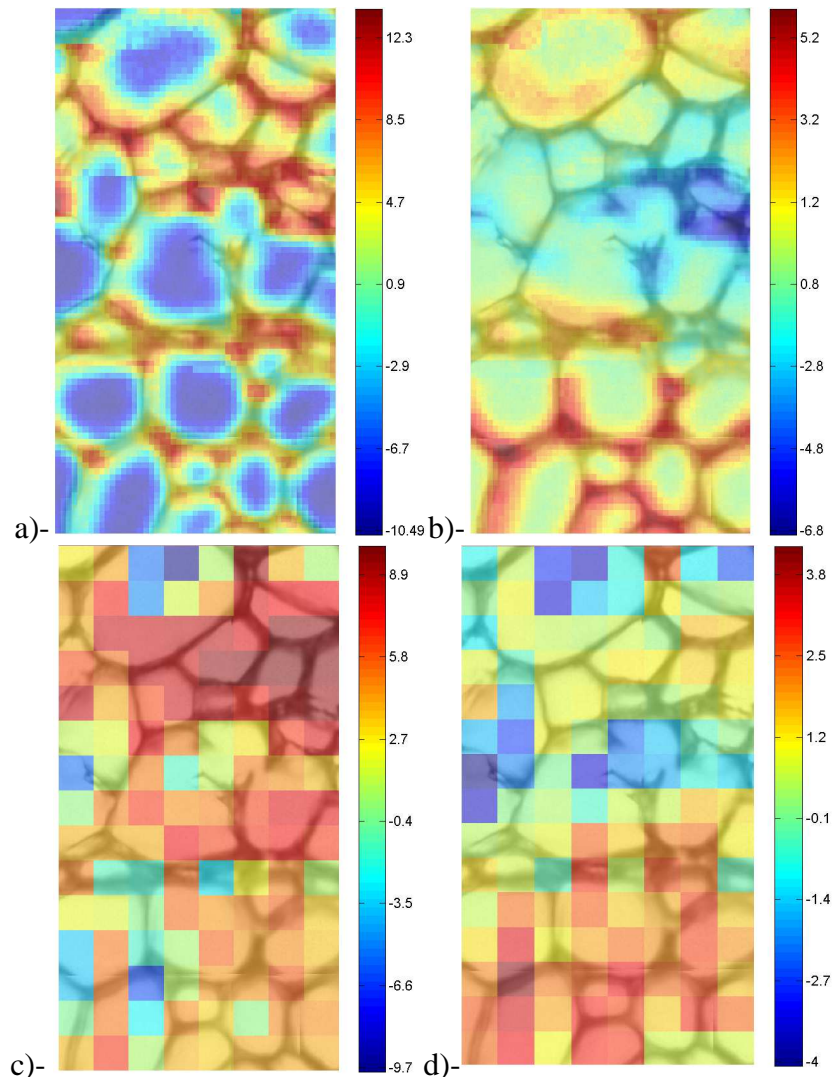


Figure D7 Images composantes obtenues en utilisant l'analyse inter-batterie de Tucker à trois entrées : a-b)- images composantes 1 et 2 de fluorescence c-d)- images composantes 1 et 2 de l'infrarouge. Ces images ont été obtenues en superposant l'image fond clair avec l'image composante repliée avec un coefficient de transparence égal à 0,5. La carte 'Jet' de Matlab a été appliquée pour les composantes après ajustement des valeurs entre 0 et 255 pour la visualisation. Les barres d'échelles indiquent les valeurs minimales et maximales réelles des valeurs des composantes.

Les valeurs des composantes de fluorescence ont été calculées, repliées et affichées en tant qu'images composites avec l'image visible de référence pour révéler les variations spatiales. La résolution des données de fluorescence étant conservée, la structure cellulaire de la région analysée est clairement distinguée. L'image composante 1 (Fig. D7-a) révèle la position où l'intensité de fluorescence maximale a été observée. Les pixels positifs sont localisés en rouge sur les parois des cellules avec des maxima sur les jonctions cellulaires. Les pixels négatifs en bleu correspondent à l'intérieur de cellules. L'image composante 2 de fluorescence (Fig. D7-b) a révélé en couleur rouge les parois cellulaires du phloème avec une plus grande intensité

au niveau des jonctions cellulaires et avec une intensité moindre, sur les parois cellulaires du parenchyme. Les pixels bleus correspondent à une région de la gaine du faisceau du sclérenchyme. Considérant l'interprétation des profils spectraux, les acides hydroxycinnamiques ont été localisés principalement dans le phloème et le parenchyme et en quantité plus élevée dans les jonctions cellulaires. Les cellules du sclérenchyme sont enrichies en lignine co-localisée avec des liaisons esters et des xylanes.

Conclusion

Dans ce chapitre, les données des images hyperspectrales infrarouge et de fluorescence ont été analysées conjointement en utilisant une méthode multivariée reliant deux tableaux de données. Le critère de maximisation retenu pour l'analyse est la covariance au carré. Nous avons choisi d'utiliser l'analyse inter-batterie de Tucker comme étant une méthode symétrique et orthogonale. Cette méthode a été étendue pour préserver la résolution spatiale des images hyperspectrales et par conséquent d'analyser une structure de données ayant un nombre d'entrées hétérogènes : un tableau à deux entrées et un tableau à trois entrées.

Les profils variables obtenus en faisant la moyenne selon la dimension spectrale sont identiques à ceux obtenus avec l'extension de l'analyse inter-batterie de Tucker. L'intérêt de calcul avec le tableau de fluorescence à trois entrées est de prendre en compte l'ensemble des données de fluorescence sans réduire leur résolution.

II Couplage de trois images hyperspectrales par l'analyse de co-inertie multiple

Introduction

Les méthodes les plus populaires pour analyser des tableaux multiples sont basées sur des approches séquentielles qui reposent sur une étape de déflation. La déflation est le processus qui consiste à soustraire l'information décrite par une composante du tableau de données initial avant de calculer la composante suivante. De telles approches séquentielles sont couramment utilisées en chimiométrie [Smilde *et al.*, 2003 ; Hanafi et Kiers, 2006].

L'analyse conjointe de deux tableaux \mathbf{X} et \mathbf{Y} est basée sur la maximisation de la covariance entre deux composantes \mathbf{c}_X et \mathbf{c}_Y .

Pour l'analyse de plus de deux tableaux, le critère ne peut pas être étendu de manière immédiate. Beaucoup de méthodes basées sur la notion de composante commune ou compromis sont développées pour étudier la ressemblance entre les tableaux [Qannari *et al.*, 2000 ; Hanafi *et al.*, 2011]. Par exemple, l'analyse en composantes principales consensuelles et l'analyse de co-inertie multiple optimisent le même critère, i.e., la somme des covariances

entre les composantes par bloc et une composante globale. La différence principale entre les deux méthodes réside dans la déflation. La déflation en analyse de co-inertie multiple est faite sur les composantes par bloc alors qu'en analyse en composantes principales consensuelles, la déflation est faite sur les composantes communes. Une des conséquences est que l'analyse de co-inertie multiple présente l'avantage d'avoir des profils variables orthogonaux par bloc. Etant symétrique et orthogonale, l'analyse de co-inertie multiple a été retenue préférentiellement aux autres méthodes.

Le but de l'analyse conjointe de trois tableaux de données est d'étudier la complémentarité et la redondance entre les trois tableaux. La première section de ce chapitre est dédiée à la présentation de l'analyse de co-inertie multiple. La deuxième section est consacrée aux développements des indicateurs statistiques permettant d'interpréter la participation de chaque tableau de données pour déterminer la composante globale. La troisième section concerne l'analyse et l'interprétation de la collection d'images 4f11mb-2.

1 Analyse de co-inertie Multiple : Méthodes et outils d'interprétation

1.1 Brève présentation de l'analyse de co-inertie multiple

L'objectif de l'analyse de co-inertie multiple est de décrire les similarités entre plusieurs tableaux de données centrées \mathbf{X}_k ($1 \leq k \leq K$), appelées aussi blocs, observés sur les mêmes n échantillons en décrivant le maximum de variance pour chacun d'eux. Une structure commune est nécessaire pour étudier les covariances entre plus de deux tableaux de données. Elle est obtenue en calculant une composante synthétique commune à tous les blocs synthétisant l'ensemble de l'information apportée par les différentes méthodes. Des composantes par bloc sont également calculées. La première étape de la méthode consiste à déterminer une composante globale \mathbf{c}_g et des composantes par bloc $\mathbf{c}_{\mathbf{X}_k} = \mathbf{X}_k \mathbf{u}_k$ ($k=1, \dots, K=\text{nombre de bloc}$) de manière à maximiser la somme des carrés des covariances

$$\sum_{k=1}^K \text{cov}^2(\mathbf{c}_{\mathbf{X}_k}, \mathbf{c}_g) \text{ sous les contraintes : } \|\mathbf{u}_k\| = 1 \text{ et } \|\mathbf{c}_g\| = 1 \quad \text{Equation D16}$$

L'interprétation de ce critère est très simple : les composantes par bloc $\mathbf{c}_{\mathbf{X}_k} = \mathbf{X}_k \mathbf{u}_k$ sont calculées pour être liées autant que possible à la composante globale \mathbf{c}_g , qui elle-même synthétise l'ensemble de l'information de tous les tableaux.

Une solution du problème de maximisation est obtenue en effectuant une analyse en composantes principales du tableau concaténé $\mathbf{X}=[\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K]$. La première composante principale de \mathbf{X} , après standardisation, représente la composante globale $\mathbf{c}_g^{(1)}$ recherchée. Le vecteur propre associé $\tilde{\mathbf{u}}^{(1)}$ porte sur l'ensemble des variables concaténées pour tous les tableaux $\tilde{\mathbf{u}}^{(1)} = [\tilde{\mathbf{u}}_1^{(1)} \tilde{\mathbf{u}}_2^{(1)} \dots \tilde{\mathbf{u}}_K^{(1)}]$. Le vecteur $\tilde{\mathbf{u}}^{(1)}$ peut être fragmenté en sous-vecteurs. Chaque sous-vecteur associé à un tableau de données est normalisé pour obtenir les profils variables relatifs à chacun des blocs :

$$\mathbf{u}_k^{(1)} = \frac{\tilde{\mathbf{u}}_k^{(1)}}{\|\tilde{\mathbf{u}}_k^{(1)}\|} \quad \text{Equation D17}$$

Les composantes par bloc sont obtenues comme suit:

$$\mathbf{c}_{\mathbf{X}_k}^{(1)} = \mathbf{X}_k \mathbf{u}_k^{(1)} \quad \text{Equation D18}$$

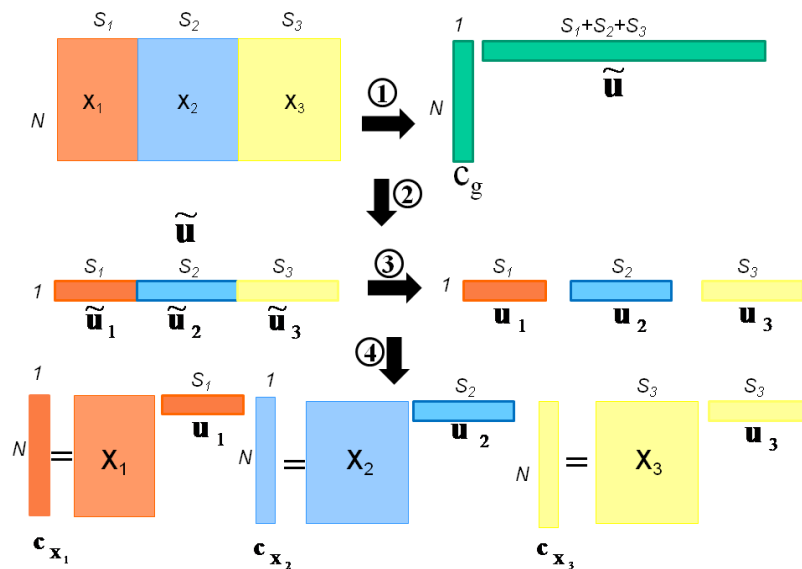


Figure D8 Première étape de l'analyse de co-inertie multiple 1)- Décomposition du tableau concaténé [$\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$] 2)- Découpage du profil variable global en des ensembles de profils variables 3)- Normalisation de chaque profil variable par bloc 4)- Calcul des composantes par bloc à partir des profils variables obtenus à la troisième étape.

La figure D8 résume la première étape de l'analyse de co-inertie multiple.

La deuxième étape de l'analyse de co-inertie multiple a pour but de calculer les profils variables et les composantes d'ordre supérieur à 1. Pour cela, chaque tableau de données est déflaté par rapport à son profil variable.

$$\mathbf{X}_k^{(2)} = \mathbf{X}_k - \mathbf{c}_{\mathbf{X}_k}^{(1)} \mathbf{u}_k^{(1)} \quad \text{Equation D19}$$

La procédure est itérée en fusionnant les blocs $\mathbf{X}_k^{(2)}$ et en calculant la deuxième composante globale $\mathbf{c}_g^{(2)}$, les nouveaux profils variables et les nouvelles composantes sont calculés en utilisant les équations D16, D17 et D18. À l'étape h , le bloc déflaté est obtenu par :

$$\mathbf{X}_k^{(h)} = \mathbf{X}_k^{(h-1)} - \mathbf{c}_{\mathbf{X}_k}^{(h-1)} \mathbf{u}_k^{(h-1)} \quad \text{Equation D20}$$

Propriété 1

Comme il a été montré dans [Smilde *et al.*, 2003 ; Hanafi *et al.*, 2011], les profils variables par bloc ainsi que les composantes globales sont mutuellement orthogonaux.

Propriété 2

Les composantes $\mathbf{c}_{\mathbf{X}_k}^{(h)}$ et les profils variables $\mathbf{u}_k^{(h)}$ ($1 \leq h \leq H ; 1 \leq k \leq K$); permettent la reconstitution du tableau de données k comme suit :

$$\mathbf{X}_k = \sum_{h=1}^H \mathbf{c}_{\mathbf{X}_k}^{(h)} \mathbf{u}_k^{(h)} + \mathbf{R}_k^{(h)} \quad \text{Equation D21}$$

où $\mathbf{R}_k^{(h)}$ est le tableau des résidus après décomposition du tableau \mathbf{X}_k .

Propriété 3

La variance de chaque tableau de données k décrite par les variances des composantes par bloc $\mathbf{c}_{\mathbf{X}_k}^{(h)}$:

$$\sum_{h=1}^H \mathbf{c}_{\mathbf{X}_k}^{(h)} \mathbf{c}_{\mathbf{X}_k}^{(h)} \quad \text{Equation D22}$$

Propriété 4

La composante globale $\mathbf{c}_g^{(h)}$ de l'analyse de co-inertie multiple est la première composante principale standardisée des deux tableaux de données $\mathbf{X}^{(h)}$ ou $\mathbf{c}_{\mathbf{X}}^{(h)}$:

$$\mathbf{X}^{(h)} = [\mathbf{X}_1^{(h)} \mathbf{X}_2^{(h)} \dots \mathbf{X}_K^{(h)}] \quad \text{Equation D23}$$

$$\mathbf{c}_{\mathbf{X}}^{(h)} = [\mathbf{c}_{\mathbf{X}_1}^{(h)} \mathbf{c}_{\mathbf{X}_2}^{(h)} \dots \mathbf{c}_{\mathbf{X}_K}^{(h)}] \quad \text{Equation D24}$$

En effet, la première composante principale de $\mathbf{c}_{\mathbf{X}_k}^{(1)}$ est calculée en maximisant

$$\sum_{k=1}^K \text{cov}^2(\mathbf{c}_{\mathbf{X}_k}, \mathbf{c}_g)$$

sous la contrainte $\|\mathbf{c}_g\| = 1$.

1.2 Outils d'interprétation

Des indicateurs sont proposés pour étudier à la fois l'information commune et l'information spécifique apportée par chaque tableau de données.

1.2.1 Contribution des tableaux de données à la structure commune

Etant une composante principale du tableau des composantes de chaque bloc $\mathbf{c}_{\mathbf{X}_k}^{(h)}$, $\mathbf{c}_g^{(h)}$ peut être écrite comme suit :

$$\mathbf{c}_g^{(h)} = \sum_{k=1}^K \alpha_k \mathbf{c}_{\mathbf{X}_k}^{(h)} \quad \text{Equation D25}$$

Les coefficients α_k (ou, de préférence, leurs carrés) peuvent être utilisés pour évaluer la contribution de chaque bloc pour la détermination des composantes globales. Tous les détails mathématiques et les démonstrations sont disponibles dans [Chessel et Hanafi, 1996 ; Hanafi et Kiers, 2006 ; Hanafi *et al.*, 2011].

Le niveau de similarités entre les tableaux de données pour une composante peut être estimé par le calcul de la contribution de chaque tableau de données, soit à la composante globale, soit au profil variable global. Les coefficients α_k de l'analyse en composante principale du tableau $\mathbf{c}_X^{(h)} = [\mathbf{c}_{\mathbf{X}_1}^{(h)} \mathbf{c}_{\mathbf{X}_2}^{(h)} \dots \mathbf{c}_{\mathbf{X}_k}^{(h)}]$ (Équation D24) et la norme $\|\tilde{\mathbf{u}}_k^{(h)}\|$ où $\tilde{\mathbf{u}}_k^{(h)}$ est le profil variable du tableau de données \mathbf{X}_k avant normalisation, mesure cette contribution. La norme de $\tilde{\mathbf{u}}_k^{(h)}$ a été considérée dans ce travail.

1.2.2 Reconstitution de tableaux de données à partir des composantes

La part de variance de chaque tableau de données \mathbf{X}_k décrite par les composantes par bloc $\mathbf{c}_{\mathbf{X}_k}^{(h)}$ correspond à la variance du tableau impliqué dans le calcul des composantes globales et donc dans l'information commune. Cette information permet de montrer quelles méthodes contribuent le plus à la création de la composante globale.

1.2.3 Covariances entre les composantes par bloc et les composantes communes

Le pourcentage de variance expliquée par la composante globale $\mathbf{c}_g^{(h)}$ également composante principale du tableau de données $\mathbf{c}_X^{(h)} = [\mathbf{c}_{\mathbf{X}_1}^{(h)} \mathbf{c}_{\mathbf{X}_2}^{(h)} \dots \mathbf{c}_{\mathbf{X}_k}^{(h)}]$ est égal à :

$$\lambda_g^{(h)} = \frac{\sum_{k=1}^K \text{cov}^2(\mathbf{c}_{\mathbf{X}_k}^{(h)}, \mathbf{c}_g^{(h)})}{\sum_{k=1}^K \text{var}(\mathbf{c}_{\mathbf{X}_k}^{(h)})^2} \quad \text{Equation D26}$$

Le coefficient $\lambda_k^{(h)} = \frac{\text{cov}^2(\mathbf{c}_{\mathbf{X}_k}^{(h)}, \mathbf{c}_g^{(h)})}{\sum_{k=1}^K \text{var}(\mathbf{c}_{\mathbf{X}_k}^{(h)})^2}$ mesure la part de covariance introduite par chaque

tableau de données $\mathbf{X}_k^{(h)}$ dans la composante globale et permet de mesurer plus précisément la contribution du tableau k pour le calcul de l'information commune entre les tableaux.

Une valeur élevée du pourcentage $\lambda_g^{(h)}$ indique que la composante globale décrit une grande partie de la covariance. Cette covariance peut mesurer le lien entre les tableaux ou décrire un tableau qui aurait une grande variance. Une valeur faible indique qu'il n'a pas été possible de trouver une information commune. Au moins certains tableaux contiennent une information spécifique.

1.2.4 Corrélation entre la composante globale et les composantes par bloc

Les coefficients de corrélation entre les composantes par bloc $\mathbf{c}_{\mathbf{X}_k}^{(h)}$ et les composantes globales $\mathbf{c}_g^{(h)}$ peuvent être proposés pour évaluer le degré de ressemblance entre les composantes par bloc et la composante globale à une dimension donnée h .

1.2.5 Outils d'interprétation spatiaux

Des images composantes communes et par bloc sont obtenues en dépliant respectivement les valeurs des composantes communes et par bloc. Ces images sont par la suite ajustées et affichées sous la forme d'image monochrome ou en niveau de gris.

Les images composantes, telles qu'elles sont calculées en analyse de co-inertie multiple, sont affichées à la résolution des images infrarouge –la résolution la plus basse–.

Afin de calculer les images composantes à la résolution la plus élevée, tous les spectres ont été considérés comme des observations supplémentaires. Ils sont par la suite projetés sur les profils variables de chaque bloc.

$$\mathbf{c}_{\mathbf{X}_k}^{(h)}(i, z) = \mathbf{X}_k(i, \cdot, z) \mathbf{u}_k^{(h)} \quad \text{Equation D27}$$

où i est la $i^{\text{ème}}$ ligne du tableau de données à trois entrées \mathbf{X}_k , z est la $z^{\text{ème}}$ tranche.

2 Résultats et discussions

2.1 Données analysées

La collection d'images 4f11mb-2 est composée d'une image infrarouge, de 4 images de fluorescence et de 5 images Raman qui contiennent trois types cellulaires : sclérenchyme, phloème et xylème. Les spectres infrarouge, de fluorescence et Raman ont été normalisés en utilisant la normalisation spatiale et spectrale mais avec des tailles de fenêtres différentes : 3×3 pixels correspondant à $15 \times 15 \mu\text{m}^2$ pour l'infrarouge et 11×11 pixels correspondant à $11 \times 11 \mu\text{m}^2$ pour la fluorescence et le Raman. Après appariement des spectres, la structure de données établie contient un tableau à deux entrées et deux tableaux à trois entrées. :

- un tableau à deux entrées \mathbf{X}_1 de spectres infrarouge de dimensions 194×262 .
- un tableau à trois entrées \mathbf{X}_2 de spectres de fluorescence de dimensions $194 \times 322 \times 25$.
- un tableau à trois entrées \mathbf{X}_3 de spectres Raman de dimensions $194 \times 346 \times 25$.

Cette structure de données obtenue ne peut pas être soumise directement à l'analyse de co-inertie multiple. Dans une première étape, les deux tableaux à trois entrées sont moyennés selon la troisième entrée pour obtenir trois tableaux à deux entrées pouvant être soumis à l'analyse de co-inertie multiple (Fig. D9).

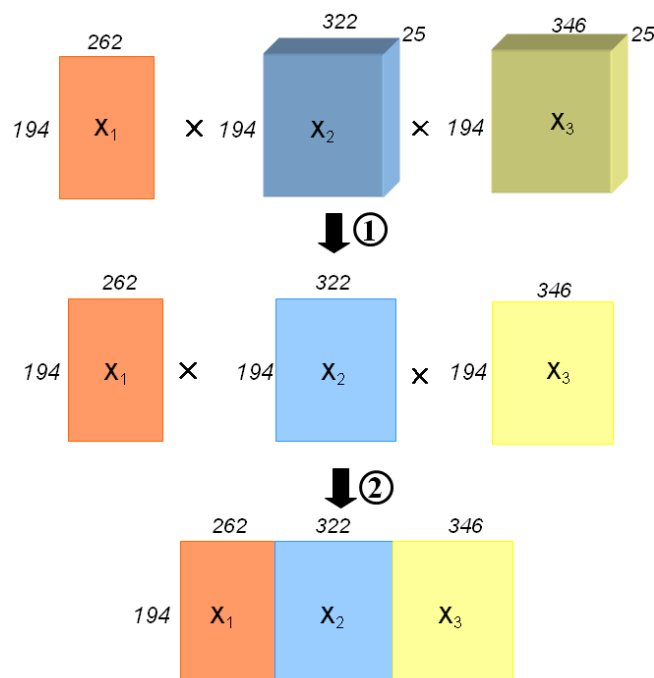


Figure D9 Adaptation de la structure de données ayant un nombre d'entrées hétérogènes à l'analyse de co-inertie multiple 1)- Moyenner les tableaux à trois entrées selon la troisième entrée 2)- Concaténer les trois tableaux à deux entrées.

Les tableaux de fluorescence et Raman ont été moyennés selon la troisième entrée. Les trois tableaux à deux entrées ont été soumis à l'analyse de co-inertie multiple. Les profils variables, les composantes ainsi que les indicateurs ont été calculés. Pour interpréter les résultats obtenus nous allons adopter la démarche suivante :

- interpréter la complémentarité et la redondance des méthodes à l'aide des indicateurs développés dans la section 1.2.
- interpréter les profils variables pour révéler la variabilité spectrale
- interpréter la variabilité spatiale en utilisant les images composantes.

2.2 Pourcentage de variances et contributions des tableaux de données

Les indicateurs développés dans la section 1.2 sont présentés pour les 5 premières composantes dans le tableau D7.

Tableau D7 Indicateurs de l'analyse de co-inertie multiple

Composantes		1	2	3	4	5
$\lambda_g^{(h)}$		72.81	86.94	92.72	89.12	87.24
$\lambda_k^{(h)}$	infrarouge	13.04	50.10	92.24	82.73	85.84
	fluorescence	32.05	3.15	0.03	1.78	0.20
	Raman	27.72	33.69	0.44	4.61	1.20
$\ \tilde{\mathbf{u}}_k^{(h)}\ $	infrarouge	17.90	57.63	99.49	92.82	93.63
	fluorescence	44.00	3.60	0.01	0.21	5.00
	Raman	38.10	31.77	0.50	6.97	1.37
corrélation avec la composante globale	infrarouge	0.75	0.96	0.99	0.99	0.99
	fluorescence	0.90	0.79	0.20	0.83	0.25
	Raman	0.86	0.91	0.25	0.57	0.33

Le tableau D7 résume la contribution de chaque tableau de données à l'analyse en utilisant les différents indicateurs. Le premier indicateur $\lambda_g^{(h)}$ est la contribution des valeurs de toutes les composantes par bloc à la composante globale. Le deuxième indicateur est la contribution de chaque composante par bloc à la détermination de la composante globale. Le troisième indicateur est la norme $\|\tilde{\mathbf{u}}_k^{(h)}\|$ qui mesure la contribution des profils variables par bloc pour les profils variables globaux. Il est à noter que la somme de $\lambda_k^{(h)}$ est égale à $\lambda_g^{(h)}$.

$$\lambda_g^{(h)} = \sum_{k=1}^K \lambda_k^{(h)} \quad \text{Equation D28}$$

D'autres part, $\|\tilde{\mathbf{u}}_k^{(h)}\|$ mesure la part relative de la participation du tableau \mathbf{X}_k dans l'information globale et $\|\tilde{\mathbf{u}}_k^{(h)}\| = \frac{\lambda_k^{(h)}}{\lambda_g^{(h)}}$.

La somme des contributions de chacun des profils variables par bloc pour déterminer le premier vecteur propre global est de 100%, 17,9% sont dus au tableau de données infrarouge, 44% sont dus au tableau de données de fluorescence et 38,1% sont dus au tableau de données Raman.

La somme des $\lambda_k^{(h)}$ peut être interprétée comme une mesure de l'information commune apportée par toutes les spectroscopies. Par exemple, à la dimension $h = 1$ l'information commune décrite est égale à 72,81% et le reste 27.19% est spécifique à chacune des méthodes spectroscopiques. Ce pourcentage est inégalement réparti entre les spectroscopies. La complémentarité et la redondance des tableaux de données peuvent être interprétées à partir des coefficients de corrélation entre les composantes par bloc et la composante globale. Ces coefficients permettent de mesurer le degré de ressemblance entre la composante globale et les composantes par bloc. Par exemple, à la dimension $h=1$, nous remarquons que les composantes de fluorescence et Raman ressemblent fortement à la composante globale. La composante infrarouge est beaucoup moins proche de la composante globale.

L'analyse de co-inertie multiple et l'analyse en composantes principales se basent sur le même critère d'optimisation et le même genre de décomposition. Les spectres infrarouge, de fluorescence et Raman ont été normalisés en tenant compte de la nature spatiale et spectrale des images hyperspectrales. Ces raisons conduisent à ce que les composantes de co-inertie soient proches des composantes obtenues par analyse en composantes principales (Tableau D8, D9 et D10).

La première composante décrit 52% de variance totale pour l'infrarouge, 97 et 80% pour la fluorescence et le Raman, respectivement. La composante par bloc de fluorescence décrit quasiment toute l'information du tableau de fluorescence comme en analyse en composantes principales. Cependant à partir de la troisième composante, la composante globale est créée par le tableau infrarouge avec une contribution supérieure à 92%.

Tableau D8 Résultats du tableau de données infrarouge. La première ligne représente les pourcentages de variance de chaque composante principale. La deuxième ligne du tableau représente les pourcentages de variance des composantes de la co-inertie.

Composantes		1	2	3	4	5
ACP	pourcentage de variance de chaque composante principale	54.73	23.47	9.95	2.98	2.63
ACOM	pourcentage de variance de chaque composante par bloc	52.36	25.47	10.04	3.03	2.65

Tableau D9 Résultats du tableau de données de la fluorescence. La première ligne représente les pourcentages de variance de chaque composante principale. La deuxième ligne du tableau représente le pourcentage de variance des composantes de la co-inertie.

Composantes		1	2	3	4	5
ACP	pourcentage de variance de chaque composante principale	97.15	2.55	0.15	0.10	0.03
ACOM	pourcentage de variance de chaque composante par bloc	97.14	2.54	0.09	0.10	0.11

Tableau D10 Résultats du tableau de données Raman. La première ligne représente les pourcentages de variance de chaque composante principale. La deuxième ligne du tableau représente le pourcentage de variance des composantes de la co-inertie.

Composantes		1	2	3	4	5
ACP	pourcentage de variance de chaque composante principale	80.36	17.59	0.98	0.50	0.19
ACOM	pourcentage de variance de chaque composante par bloc	80.07	17.81	0.71	0.49	0.32

Dans la section suivante, seulement, les trois premières composantes sont interprétées et discutées. Ces composantes décrivent 88% de la variance totale pour l'infrarouge, 99% pour la fluorescence et 98% pour le Raman.

2.3 Interprétation spectrale et spatiale

2.3.1 Composante 1

Les profils spectraux ont été calculés pour chaque spectroscopie (Fig. D10). Comme il était attendu à partir de la normalisation spatiale et spectrale, les premiers profils spectraux $\mathbf{u}_1^{(1)}$, $\mathbf{u}_2^{(1)}$ et $\mathbf{u}_3^{(1)}$, décrivent les variations d'intensité entre les spectres pour les trois spectroscopies (Fig. D10-a, b et c).

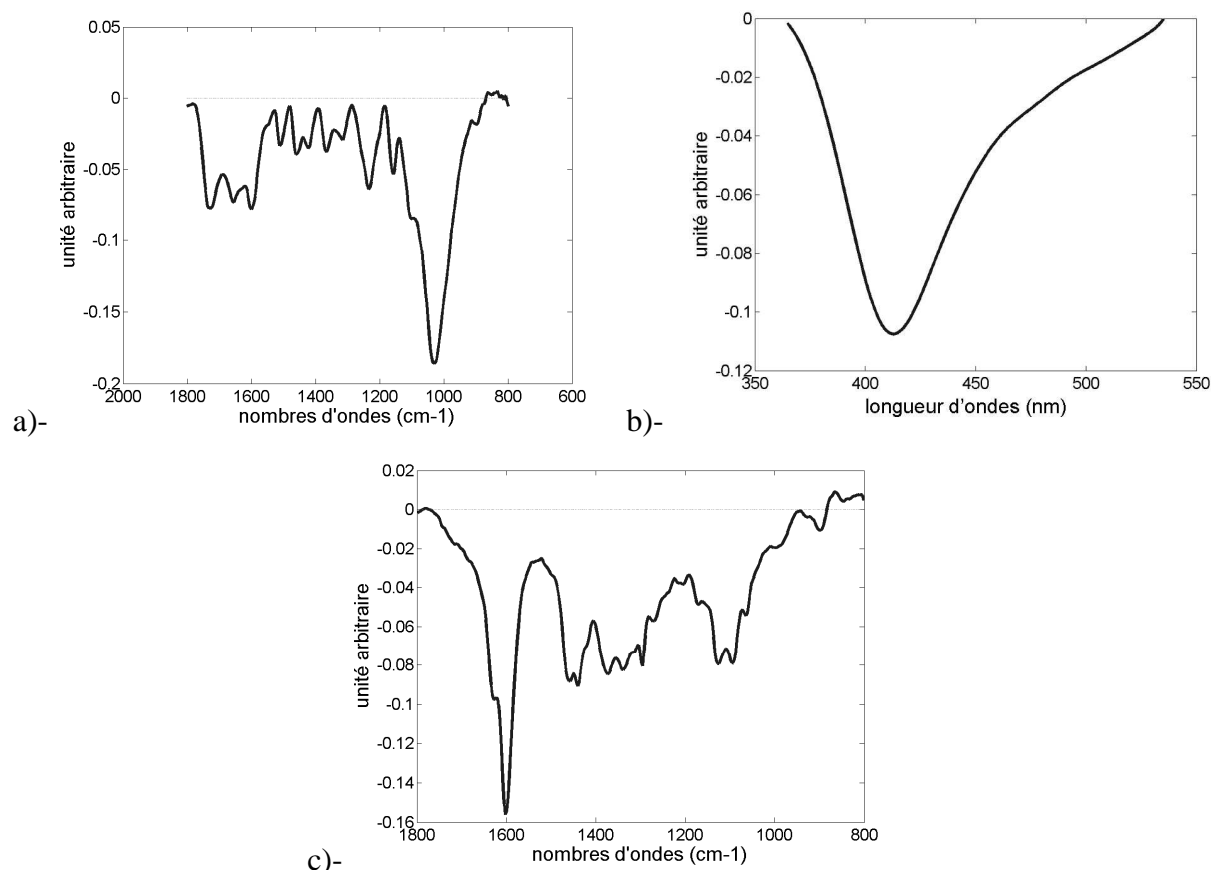


Figure D10 Profils spectraux 1de l'analyse de co-inertie multiple: a)- infrarouge, b)- de fluorescence, c)- Raman.

Les images composantes $c_1^{(1)}$, $c_2^{(1)}$ and $c_3^{(1)}$ (Fig. D11-a, b et c), montrent la position où les intensités minimales ont été observées pour les trois microspectroscopies. Les pixels bleus ont été trouvés principalement dans les jonctions cellulaires et les pixels rouges dans les fonds de cellules. Sur la composante de fluorescence, les intensités minimales ont été observées sur les jonctions des parois cellulaires du phloème (Fig. D11-b). L'image composante Raman a mis en évidence les cellules du xylème avec une forte intensité en négatif. Les intensités les plus fortes en négatif sont observées sur les jonctions cellulaires du xylème (Fig. D11-c). L'image composante infrarouge a révélé tous les pixels appartenant au sclérenchyme, au phloème et au xylème en bleu (Fig. D11-a). Les microspectroscopies de fluorescence et Raman ont révélé les fonds de cellules d'une manière homogène.

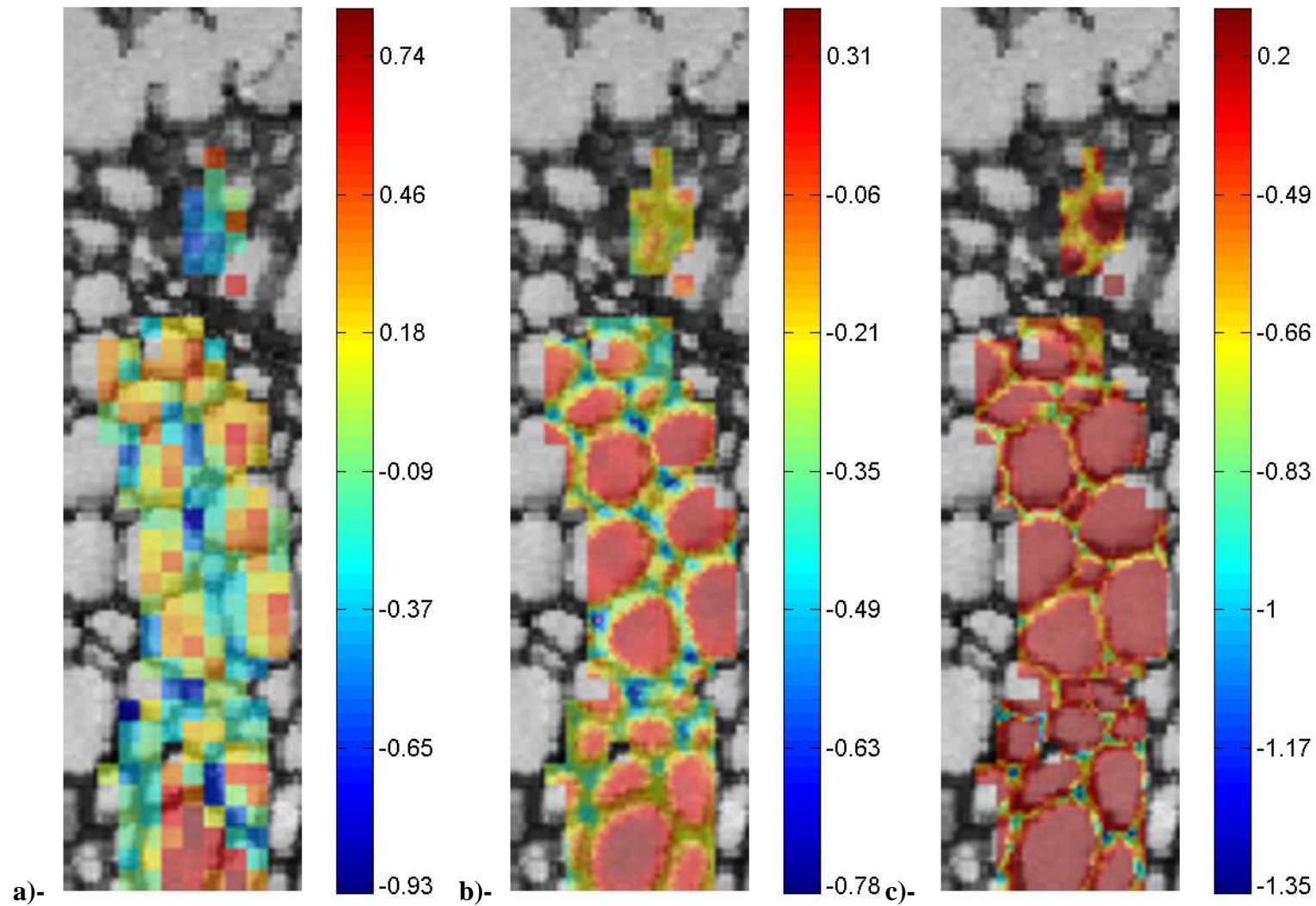


Figure D11 Images composantes 1 de l'analyse de co-inertie multiple a)- infrarouge b)- de fluorescence c)- Raman.

2.3.2 Composante 2

Le second profil spectral infrarouge $u_1^{(2)}$ présente des pic positifs qui correspondent aux polysaccharides (cellulose et xylanes) avec un maximum à 1068 et 1110 cm^{-1} . Des pics négatifs ont été observés à:

- 1420, 1454, 1504 et 1593 cm^{-1} qui ont été attribués aux lignines.
- 1230 cm^{-1} qui correspondent aux liaisons esters.

Le second profil spectral de fluorescence $u_2^{(2)}$ (Fig. D12-b) oppose un pic positif à 390 nm contre la région entre 450 nm et 550 nm en négatif. Ceci a été interprété comme une opposition entre des signaux révélant la lignine à des signaux d'acides hydroxycinnamiques : acide férulique et acide para-coumarique.

Le profil spectral Raman $u_3^{(2)}$ (Fig. D12-c) présente un pic intense négatif à 1600-1631 cm^{-1} . Ce pic correspond à la lignine. Des pics négatifs ont été observés à : - 1093-1131, 1296, 1340, 1370, 1463 cm^{-1} qui peuvent être attribués aux polysaccharides (cellulose et aux xylanes).

Les trois spectroscopies donnent des résultats cohérents en démontrant des profils opposant la lignine aux polysaccharides.

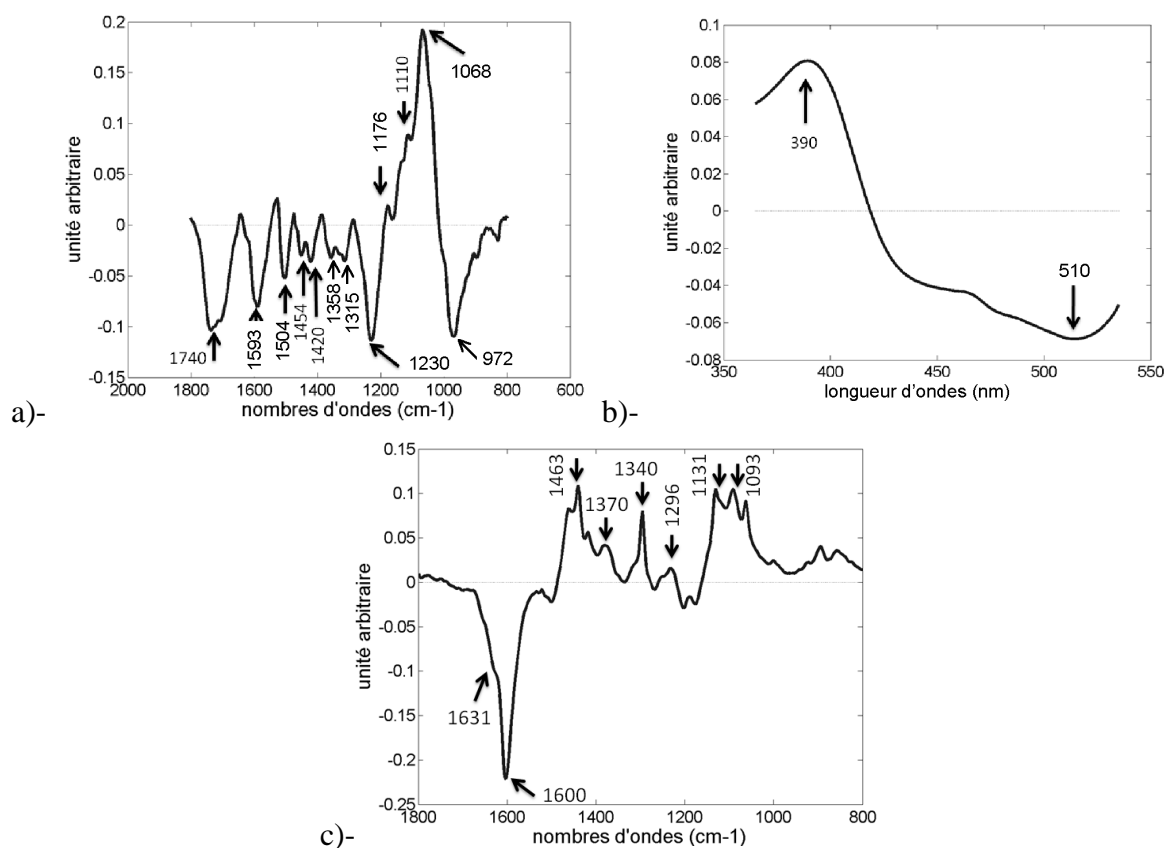


Figure D12 Profils spectraux 2 de l'analyse de co-inertie multiple: a)- infrarouge, b)- de fluorescence, c)- Raman.

La seconde composante oppose des pixels positifs des parois cellulaires du phloème avec des pixels négatifs des parois cellulaires du sclérenchyme et du xylème. Cette observation est cohérente avec les profils spectraux, le sclérenchyme est connu par contenir des lignines et les cellules du phloème sont plus riches en polysaccharides.

Toutefois, le contraste entre les trois tissus diffère d'une spectroscopie à l'autre. Le plus fort contraste est observé en spectroscopie Raman. Les spectroscopies de fluorescence et infrarouge, ont mis en évidence les parois cellulaires de sclérenchyme beaucoup plus que les parois cellulaires du xylème. Cela peut être interprété comme une composition biochimique différente entre les parois cellulaires du sclérenchyme et du xylème. En ce qui concerne la composante de fluorescence (Fig. D13-b), les lignines réagissent fortement dans le cas du sclérenchyme. Ceci peut être interprété par la présence d'une teneur plus élevée de lignines dans le sclérenchyme que dans le xylème. En corrélant l'interprétation spatiale et spectrale, les cellules de xylème sont enrichies en polysaccharides par rapport aux cellules du sclérenchyme. La spectroscopie Raman n'a pas montré ce résultat peut-être en raison de la forte réponse des groupes phénoliques autour $(1603-1630) \text{ cm}^{-1}$ qui pourraient avoir dominé le signal. Les zones de jonction de cellules du xylème sont particulièrement intenses sur la composante image Raman.

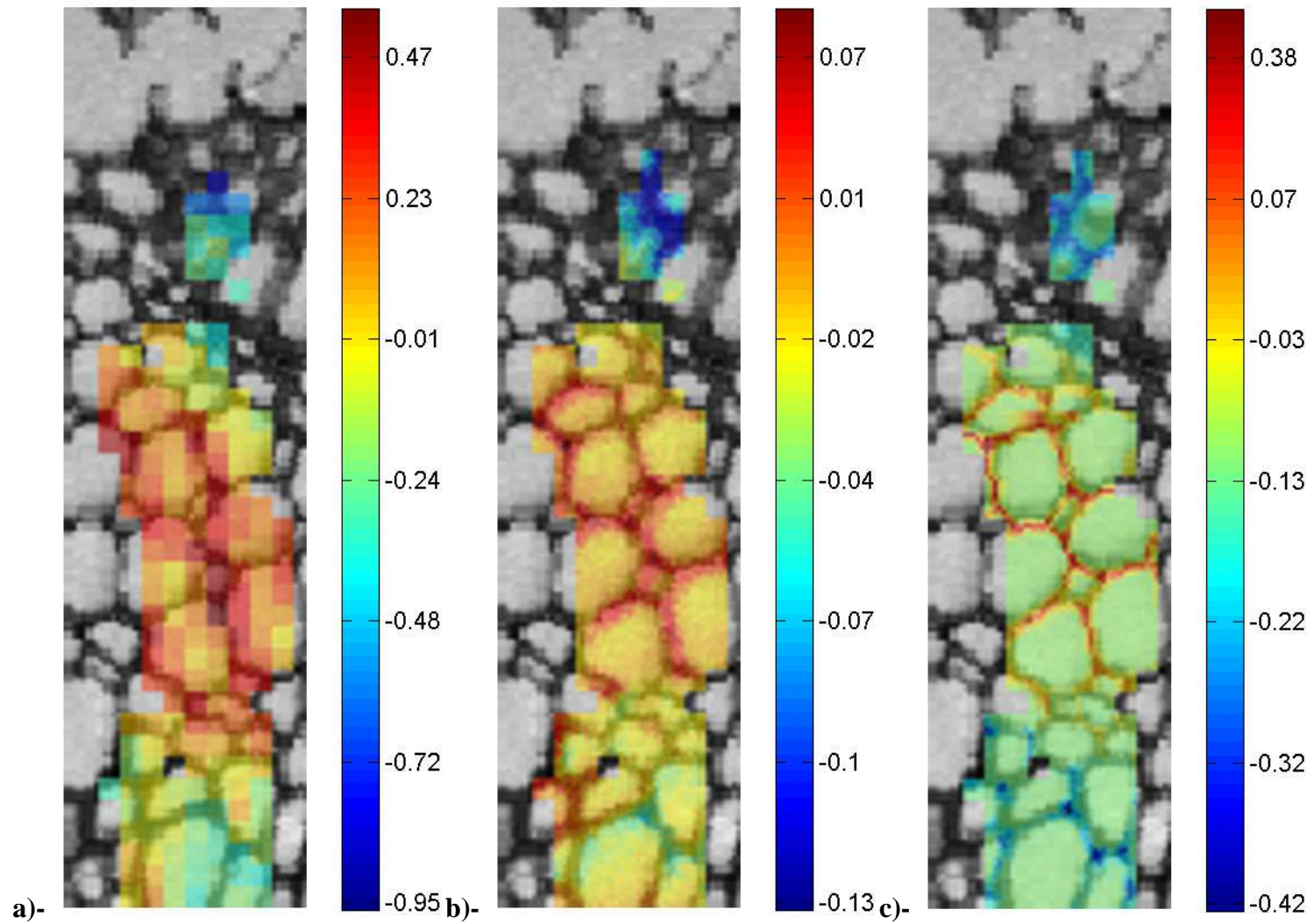


Figure D13 Images composantes 2 de l'analyse de co-inertie multiple a)- infrarouge b)-fluorescence c)-Raman.

2.3.3 Composante 3

Le profil spectral infrarouge $\mathbf{u}_1^{(3)}$ (Fig. D14-a) présente des pics positifs révélant la lignine et les liaisons esters en opposition à un pic négatif à 945 cm^{-1} . Le profil spectral de fluorescence $\mathbf{u}_2^{(3)}$ (Fig. D14-b) a également présenté un pic positif dans la région de lignine (447-456 nm). Le profil spectral Raman $\mathbf{u}_3^{(3)}$ était bruité montrant qu'aucune information pertinente n'a été décrite par cette composante.

Les valeurs des composantes par bloc pour chaque tableau de données ont été calculées, dépliées et affichées en tant qu'images composites avec l'image visible de référence pour révéler les variations spatiales des composantes chimiques. Une carte en couleur a été appliquée pour l'affichage des valeurs des composantes : les valeurs basses sont affichées en bleu et les hautes en rouge.

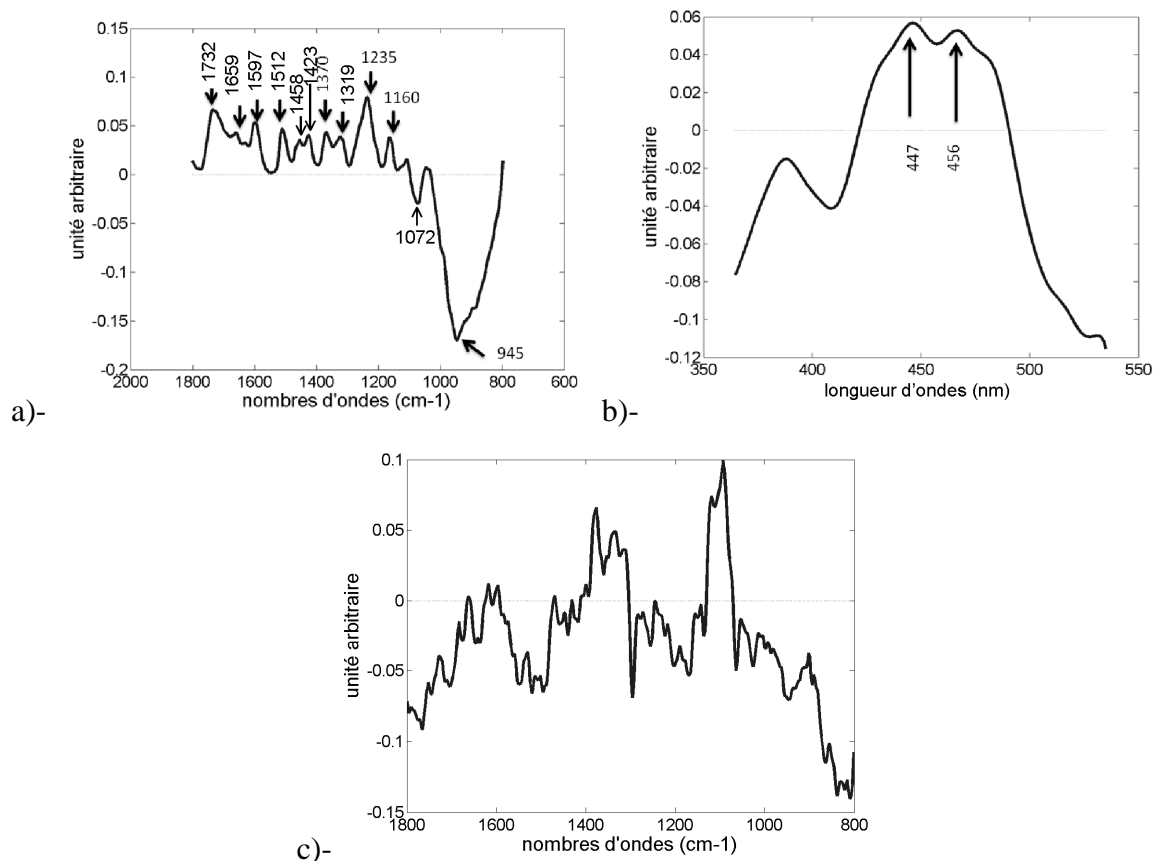


Figure D14 Profils spectraux 3 de l'analyse de co-inertie multiple a)- infrarouge b)-fluorescence c)- Raman.

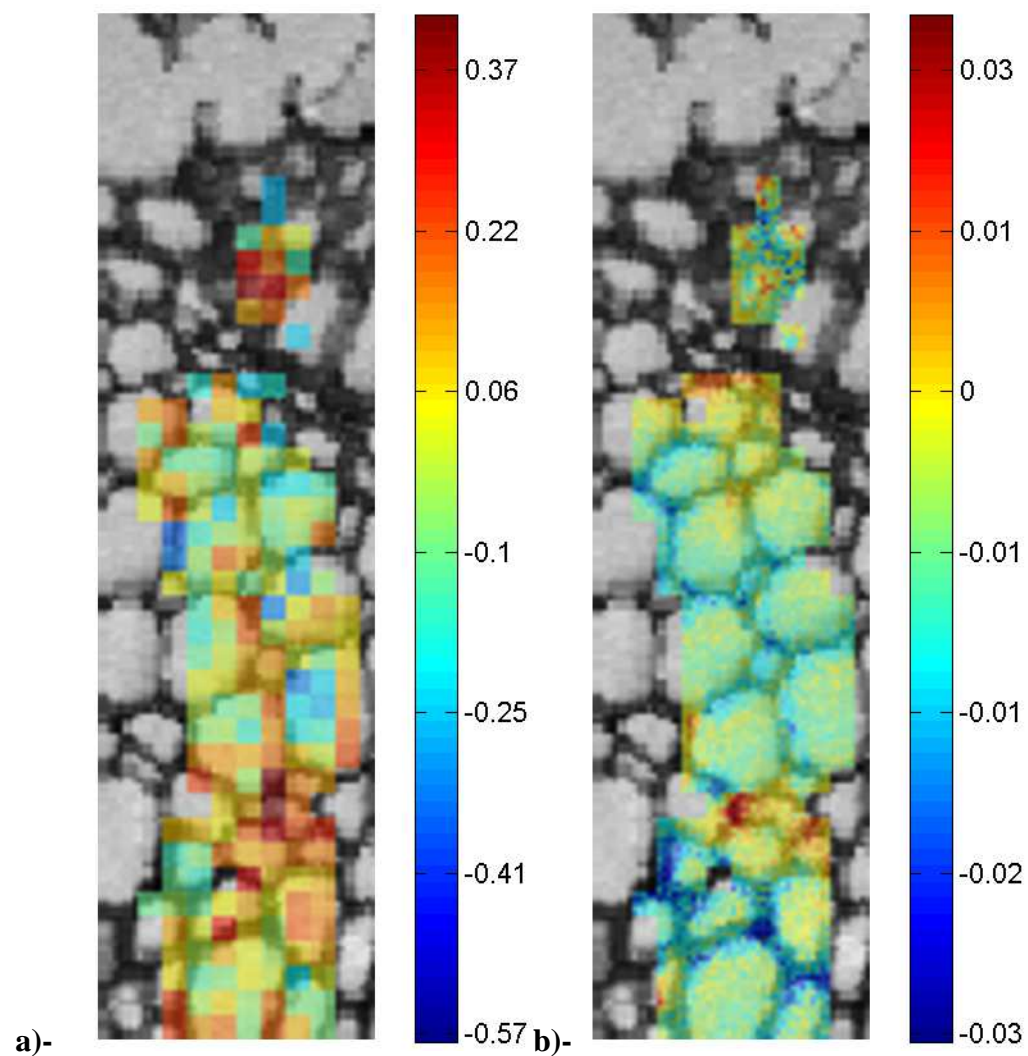


Figure D15 Images composantes 3 de l'analyse de co-inertie multiple a)- infrarouge b)- de fluorescence

La composante 3 n'a pas été interprétée pour le Raman parce que le profil spectral était bruité. Par conséquent, les variations spatiales n'étaient pas pertinentes. Les pixels infrarouge positifs ont été trouvés sur les parois des cellules avec une intensité plus élevée dans les cellules du sclérenchyme et les cellules du xylème. Le profil spectral associé présentait des pics positifs observés pour la lignine et les liaisons esters. Deux petites régions sont révélées en positif dans l'image de fluorescence correspondant aux parois cellulaires de la frontière séparant le phloème du xylème et du sclérenchyme. Son profil spectral associé révèle les lignines.

Conclusion

L'objectif de ce travail était d'appliquer l'analyse de co-inertie multiple pour analyser conjointement les données d'images hyperspectrales et de chercher l'information spécifique et commune portée par chacune des modalités. La méthode est conçue pour calculer les profils variables globaux, les profils variables par bloc, les composantes globales et les composantes par bloc. Plusieurs indicateurs ont été proposés afin d'étudier les liaisons entre les composantes globales et les composantes par bloc. Le résultat obtenu a montré que, dans certains cas, une contribution de faible variance peut révéler des informations pertinentes. Ceci pourrait s'expliquer par le fait qu'un petit nombre de pixels est impliqué dans les covariances par rapport aux autres tableaux de données. Par exemple la troisième composante de fluorescence ayant un écart de 0,1%, révèle des différences dans les parois séparant le phloème du sclérenchyme et du xylème (Fig. D13-b).

Dans le présent travail, les trois premières composantes ont été examinées. En utilisant les outils d'interprétation, nous avons démontré que chacun des tableaux de données a contribué d'une manière non proportionnée pour former les profils variables globaux. Le premier profil variable global décrit les variations d'intensité causées par la présence de pixels sans signal et des pixels avec un signal dans les parois. Les tableaux de données de fluorescence et Raman ont contribué davantage à cette information. Le deuxième profil variable a opposé des parois lignifiées à celles non ou moins lignifiées. Principalement l'infrarouge et Raman ont contribué à cette composante. Pour les autres composants, infrarouge a contribué à plus de 90%.

La pleine résolution des images de fluorescence et Raman a été obtenue dans les images composantes, en considérant les pixels des imagerie comme des individus supplémentaires en les projetant sur le profil variable obtenu.

Dans le chapitre suivant, l'analyse de co-inertie multiple sera étendue dans le but d'analyser un mélange de tableaux de données à deux et à trois entrées en s'appuyant à une approche similaire à celle développée en analyse inter-batterie de Tucker.

III. Couplage de trois images hyperspectrales multirésolutions : extension de l'analyse de co-inertie multiple

Introduction

L'objectif de ce chapitre est d'étendre l'analyse de co-inertie multiple de manière à considérer les tableaux à trois entrées avec leur pleine résolution. Le principe est similaire à celui développé en analyse inter-batterie de Tucker à trois entrées. Pour illustrer cette méthode, nous avons analysé la collection d'images hyperspectrales 4f11mb-2. Dans la première section de ce chapitre, la méthode développée qu'on a appelé analyse de co-inertie trilineaire a été présentée. Dans la deuxième section, la collection 4f11mb-2 est analysée avec l'analyse de co-inertie trilineaire. Les résultats sont comparés avec ceux obtenus en analyse de co-inertie multiple.

1 Extension de l'analyse de co-inertie multiple à l'analyse de tableaux ayant un nombre d'entrées hétérogènes

1.1 Algorithme

Dans cette section, l'analyse de co-inertie multiple a été étendue dans le cas où les tableaux de données à analyser sont un mélange de tableaux à deux et à trois entrées. La méthode est présentée pour un nombre quelconque de tableaux à 2 ou à 3 entrées.

Notons \mathbf{X}_k les K tableaux de données à deux entrées et \mathbf{Y}_j les J tableaux de données à trois entrées, centrés selon la dimension spectrale. Un tableau à trois entrées peut être considéré comme un ensemble de $Z(j)$ tableaux à deux entrées $\mathbf{Y}_{j(z)}$ où $Z(j)$ est le nombre de tranches de chaque tableau. Comme en analyse de co-inertie multiple, une structure commune est nécessaire pour étudier les covariances entre tous les tableaux de données. Un ensemble de composantes par tableau est calculé de manière à maximiser les covariances entre la structure commune et les composantes par tableau. La première étape de la méthode consiste à déterminer une composante globale \mathbf{c}_g et des composantes par tableau, $\mathbf{c}_{\mathbf{X}_k} = \mathbf{X}_k \mathbf{u}_k$ ($\|\mathbf{u}_k\|=1$) ($k=1, \dots, K$) et $\mathbf{c}_{\mathbf{Y}_{j(z)}} = \mathbf{Y}_{j(z)} \mathbf{v}_j$ ($\|\mathbf{v}_j\|=1$) ; ($j=1, \dots, J$ et $z=1, \dots, Z(j)$), qui maximise la somme des carrés des covariances

$$\sum_{k=1}^K \text{cov}^2(\mathbf{c}_{\mathbf{X}_k}, \mathbf{c}_g) + \sum_{j=1}^J \sum_{z=1}^{Z(j)} \text{cov}^2(\mathbf{c}_{\mathbf{Y}_{j(z)}}, \mathbf{c}_g) \quad \text{Equation D29}$$

sous les contraintes : $\|\mathbf{u}_k\|=1$, $\|\mathbf{v}_j\|=1$ et $\|\mathbf{c}_g\|=1$

Les composantes par bloc $\mathbf{c}_{\mathbf{X}_k} = \mathbf{X}_k \mathbf{u}_k$ et par tranche de tableau à trois entrées $\mathbf{c}_{\mathbf{Y}_{j(z)}} = \mathbf{Y}_{j(z)} \mathbf{v}_j$ sont calculées pour être liées autant que possible à la composante globale \mathbf{c}_g .

Comme en analyse inter-batterie de Tucker, la procédure est basée sur le calcul d'une matrice somme pondérée des tableaux à trois entrées.

L'algorithme commence par le choix de J vecteurs tirés au sort $\alpha_j^{(0)}$ avec $\|\alpha_j^{(0)}\|=1$. Pour chaque tableau à trois entrées $\underline{\mathbf{Y}}_j$, toutes les tranches $\mathbf{Y}_{j(z)}$ sont combinées pour calculer une somme pondérée des matrices $\underline{\mathbf{Y}}_j$ dans le but de former une matrice $\overline{\mathbf{Y}}_j$. L'analyse de co-inertie multiple est ensuite appliquée au tableau fusionné $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K, \overline{\mathbf{Y}}_1, \overline{\mathbf{Y}}_2, \dots, \overline{\mathbf{Y}}_J]$. Une fois l'analyse de co-inertie multiple appliquée, on obtient un ensemble de profils variables relatifs à chacun des tableaux de données (\mathbf{u}_k pour \mathbf{X}_k et \mathbf{v}_j pour $\underline{\mathbf{Y}}_j$) et la composante globale \mathbf{c}_g . Les composantes $\mathbf{c}_{\mathbf{Y}_{j(z)}}$ sont calculées en utilisant l'équation $\mathbf{c}_{\mathbf{Y}_{j(z)}} = \mathbf{Y}_{j(z)} \times \mathbf{v}_j$

$\mathbf{c}_{\mathbf{Y}_j}$ est le tableau à trois entrées résultant de la concaténation de tous les vecteurs $\mathbf{c}_{\mathbf{Y}_{j(z)}}$.

Les $\alpha_j^{(l)}$ sont ajustés en projetant chaque composante par tranche $\mathbf{c}_{\mathbf{Y}_{j(z)}}$ sur la composante globale \mathbf{c}_g . Le calcul du vecteur poids $\alpha_j^{(l)}$ se fait tableau par tableau. Au début un scalaire $\beta_{j(z)}^{(l)}$ est calculé en projetant chaque composante par bloc et par tranche sur la composante globale $\beta_{j(z)}^{(l)} = \mathbf{c}_{\mathbf{Y}_{j(z)}}' \times \mathbf{c}_g$. Le scalaire $\beta_{j(z)}^{(l)}$ représente la covariance basée sur la ressemblance entre chaque tranche z de chaque bloc j et la composante globale \mathbf{c}_g .

L'ensemble des $Z(j)$ scalaires $\beta_{j(z)}^{(l)}$ obtenu forme le vecteur $\beta_j^{(l)}$. Le vecteur poids $\alpha_j^{(l)}$ est égal

au vecteur $\beta_j^{(l)}$ normalisé $\alpha_j^{(l)} = \frac{\beta_j^{(l)}}{\|\beta_j^{(l)}\|}$

Ces étapes sont répétées jusqu'à ce que la convergence des $\alpha_j^{(l)}$ soit atteinte. La procédure se résume comme suit :

1. Choisir un ensemble de vecteurs arbitraires et considérer $\alpha_j = \alpha_j^{(0)}$ avec $\|\alpha_j^{(0)}\|=1$
2. Calculer pour chaque tableau $\underline{\mathbf{Y}}_j$ la matrice de la somme pondérée $\overline{\mathbf{Y}}_j = \sum_{z=1}^{Z(j)} \alpha_{j(z)} \times \mathbf{Y}_{j(z)}$

3. Réaliser une analyse de co-inertie multiple sur les K tableaux \mathbf{X}_k ($k=1, \dots, K$) et les J tableaux $\bar{\mathbf{Y}}_j$ ($j=1, \dots, J$).

4. Récupérer la composante globale \mathbf{c}_g et les profils variables : \mathbf{u}_k pour \mathbf{X}_k et \mathbf{v}_j pour $\underline{\mathbf{Y}}_j$

5. Calculer les composantes par tranche $\mathbf{c}_{\mathbf{Y}_{j(z)}} = \mathbf{Y}_{j(z)} \times \mathbf{v}_j$

6. Calculer $\beta_{j(z)}^{(l)} = \mathbf{c}_{\mathbf{Y}_{j(z)}} \times \mathbf{c}_g$ et former le vecteur $\beta_j^{(l)}$

7. Calculer le vecteur poids $\alpha_j^{(l)} = \frac{\beta_j^{(l)}}{\|\beta_j^{(l)}\|}$ et considérer $\alpha_j = \alpha_j^{(l)}$

8. Itérer depuis l'étape 2 jusqu'à $\sum_{j=1}^J \|\alpha_j^{(l)} - \alpha_j^{(0)}\| < \text{epsilon}$

A l'étape $h=2$, les blocs $\mathbf{X}_k^{(2)}$ et $\underline{\mathbf{Y}}_j^{(2)}$ sont calculés en utilisant la déflation :

$$\mathbf{X}_k^{(2)} = \mathbf{X}_k - \mathbf{c}_{\mathbf{X}_k} \mathbf{u}_k', \quad k=1, \dots, K \quad \text{Equation D30}$$

$$\mathbf{Y}_{j(z)}^{(2)} = \mathbf{Y}_{j(z)} - \mathbf{c}_{\mathbf{Y}_{j(z)}} \mathbf{v}_j, \quad j=1, \dots, J \text{ et } z=1, \dots, Z(j) \quad \text{Equation D31}$$

La procédure est réitérée depuis l'étape 2 de l'algorithme en considérant les blocs $\mathbf{X}_k^{(2)}$ et $\underline{\mathbf{Y}}_j^{(2)}$ et en calculant les nouveaux profils variables et les nouvelles composantes.

À l'étape h , le bloc déflaté est obtenu par :

$$\mathbf{X}_k^{(h)} = \mathbf{X}_k^{(h-1)} - \mathbf{c}_{\mathbf{X}_k}^{(h-1)} \mathbf{u}_k^{(h-1)}, \quad k=1, \dots, K \quad \text{Equation D32}$$

$$\mathbf{Y}_{j(z)}^{(h)} = \mathbf{Y}_{j(z)}^{(h-1)} - \mathbf{c}_{\mathbf{Y}_{j(z)}}^{(h-1)} \mathbf{v}_j^{(h-1)}, \quad j=1, \dots, J \text{ et } z=1, \dots, Z(j) \quad \text{Equation D33}$$

1.2 Propriétés

Propriété 1

Comme il a été montré dans [Chessel et Hanafi, 1996 ; Hanafi et Kiers, 2006 ; Hanafi *et al.*, 2011], les profils variables par bloc ainsi que les composantes globales sont mutuellement orthogonaux.

Propriété 2

Les composantes par bloc $\mathbf{c}_{\mathbf{X}_k}^{(h)}$ et les profils variables $\mathbf{u}_k^{(h)}$ ($1 \leq h \leq H$; $1 \leq k \leq K$) ; permettent la reconstitution du tableau de données \mathbf{X}_k comme suit :

$$\mathbf{X}_k = \sum_{h=1}^H \mathbf{c}_{\mathbf{X}_k}^{(h)} \mathbf{u}_k^{(h)} + \mathbf{R}_{\mathbf{X}_k}^{(h)} \quad \text{Equation D34}$$

$\mathbf{R}_{X_k}^{(h)}$ est le tableau des résidus après décomposition du tableau \mathbf{X}_k

Les composantes $\mathbf{c}_{Y_{j(z)}}^{(h)}$ et les profils variables $\mathbf{v}^{(h)}$ ($1 \leq h \leq H ; 1 \leq z \leq Z(j)$) ; permettent la reconstitution de chaque tranche $\mathbf{Y}_{j(z)}$ du tableau de données \mathbf{Y}_j comme suit :

$$\mathbf{Y}_{j(z)} = \sum_{h=1}^H \mathbf{c}_{Y_{j(z)}}^{(h)} \mathbf{v}_j^{(h)} + \mathbf{R}_{Y_{j(z)}}^{(h)} \quad \text{Equation D35}$$

$\mathbf{R}_{Y_{j(z)}}^{(h)}$ est le tableau des résidus après décomposition de la tranche z du tableau à trois entrées \mathbf{Y}_j .

Propriété 3

La variance de chaque tableau de données à deux entrées \mathbf{X}_k décrite par les composantes $\mathbf{c}_{X_k}^{(h)}$ est égale à :

$$\sum_{h=1}^H \mathbf{c}_{X_k}^{(h)} \mathbf{c}_{X_k}^{(h)} \quad \text{Equation D36}$$

où \mathbf{c}_{X_k} est la matrice des composantes, \mathbf{c}_{X_k}

La variance de chaque tableau de données à trois entrées \mathbf{Y}_j décrite par les composantes $\mathbf{c}_{Y_{j(z)}}^{(h)}$ des tranches $\mathbf{Y}_{j(z)}$ est égale à :

$$\sum_{z=1}^{Z(j)} \sum_{h=1}^H \mathbf{c}_{Y_{j(z)}}^{(h)} \mathbf{c}_{Y_{j(z)}}^{(h)} \quad \text{Equation D37}$$

Propriété 4

La composante globale $\mathbf{c}_g^{(h)}$ de l'analyse de co-inertie multiple est la première composante principale standardisée des deux tableaux de données $\mathbf{W}^{(h)}$ ou $\mathbf{C}^{(h)}$:

$$\mathbf{W}^{(h)} = [\mathbf{X}_1^{(h)} \mathbf{X}_2^{(h)} \dots \mathbf{X}_K^{(h)} \bar{\mathbf{Y}}_1^{(h)} \bar{\mathbf{Y}}_2^{(h)} \dots \bar{\mathbf{Y}}_J^{(h)}] \quad \text{Equation D38}$$

$$\mathbf{C}^{(h)} = [\mathbf{c}_{X_1}^{(h)} \mathbf{c}_{X_2}^{(h)} \dots \mathbf{c}_{X_K}^{(h)} \mathbf{c}_{Y_1}^{(h)} \mathbf{c}_{Y_2}^{(h)} \dots \mathbf{c}_{Y_J}^{(h)}] \quad \text{Equation D39}$$

où

$\mathbf{c}_{\bar{\mathbf{Y}}_j}^{(h)}$ est la composante par bloc calculée à partir du tableau pondérée $\bar{\mathbf{Y}}_j^{(h)}$.

$\mathbf{c}_{\bar{\mathbf{Y}}_j}^{(h)} = \sum_{z=1}^Z \alpha_{j(z)}^{(h)} \mathbf{c}_{Y_{j(z)}}^{(h)}$ est la somme pondérée par le vecteur poids $\alpha_j^{(h)}$ de l'ensemble des composantes $\mathbf{c}_{Y_{j(z)}}^{(h)}$.

En effet,

$$\begin{aligned} \mathbf{c}_{\bar{\mathbf{Y}}_j}^{(h)} &= \bar{\mathbf{Y}}_j^{(h)} \mathbf{v}_j^{(h)} = \left(\sum_{z=1}^{Z(j)} \alpha_{j(z)}^{(h)} \mathbf{Y}_{j(z)}^{(h)} \right) \mathbf{v}_j^{(h)} = (\alpha_{j1}^{(h)} \bar{\mathbf{Y}}_{j1}^{(h)} + \alpha_{j2}^{(h)} \bar{\mathbf{Y}}_{j2}^{(h)} + \dots + \alpha_{j(Z)}^{(h)} \bar{\mathbf{Y}}_{j(Z)}^{(h)}) \mathbf{v}_j^{(h)} = \\ &= \alpha_{j1}^{(h)} (\bar{\mathbf{Y}}_{j1}^{(h)} \mathbf{v}_j^{(h)}) + \alpha_{j2}^{(h)} (\bar{\mathbf{Y}}_{j2}^{(h)} \mathbf{v}_j^{(h)}) + \dots + \alpha_{j(Z)}^{(h)} (\bar{\mathbf{Y}}_{j1}^{(h)} \mathbf{v}_j^{(h)}) = \sum_{z=1}^{Z(j)} \alpha_{j(z)}^{(h)} \mathbf{c}_{j(z)}^{(h)} \end{aligned}$$

1.3 Outils d'interprétation

Des indicateurs sont proposés pour étudier à la fois l'information commune et l'information spécifique apportée par chaque tableau de données.

1.3.1 Contribution des tableaux de données à la structure commune

Le niveau de similarités entre les tableaux de données pour une composante peut être estimé par le calcul de la contribution de chaque tableau de données, soit à la composante globale, soit au profil variable global. Les normes $\|\tilde{\mathbf{u}}_k^{(h)}\|$ et $\|\tilde{\mathbf{v}}_j^{(h)}\|$ où $\tilde{\mathbf{u}}_k^{(h)}$ et $\tilde{\mathbf{v}}_j^{(h)}$ sont les profils variables globaux respectifs des tableaux de données \mathbf{X}_k et \mathbf{Y}_j , avant normalisation, mesurent cette contribution.

1.3.2 Reconstitution de tableaux de données à partir des composantes

La part de variance de chaque tableau de données \mathbf{X}_k et \mathbf{Y}_j décrite par les composantes par bloc $\mathbf{c}_{\mathbf{X}_k}^{(h)}$ et par tranche $\mathbf{c}_{\mathbf{Y}_{j(z)}}^{(h)}$ correspondent à l'information impliquée dans le calcul des composantes globales et donc dans l'information commune. Ce pourcentage permet de montrer quelles méthodes contribuent le plus à la création de la composante globale.

1.3.3 Covariances entre composantes par bloc et composantes communes

Le pourcentage de variance expliquée par la composante globale $\mathbf{c}_g^{(h)}$ également composante principale du tableau de données $\mathbf{C}^{(h)} = [\mathbf{c}_{\mathbf{X}_1}^{(h)} \mathbf{c}_{\mathbf{X}_2}^{(h)} \dots \mathbf{c}_{\mathbf{X}_K}^{(h)} \mathbf{c}_{\mathbf{Y}_1}^{(h)} \mathbf{c}_{\mathbf{Y}_2}^{(h)} \dots \mathbf{c}_{\mathbf{Y}_J}^{(h)}]$ est égal à :

$$\lambda_g^{(h)} = \frac{\sum_{k=1}^K \text{cov}^2(\mathbf{c}_{\mathbf{X}_k}^{(h)}, \mathbf{c}_g^{(h)}) + \sum_{j=1}^J \text{cov}^2(\mathbf{c}_{\mathbf{Y}_j}^{(h)}, \mathbf{c}_g^{(h)})}{\sum_{k=1}^K \text{var}(\mathbf{c}_{\mathbf{X}_k}^{(h)})^2 + \sum_{j=1}^J \sum_{z=1}^{Z(j)} \text{var}(\mathbf{c}_{\mathbf{Y}_{j(z)}}^{(h)})^2}$$

Equation D40

Le coefficient $\lambda_k^{(h)} = \frac{\text{cov}^2(\mathbf{c}_{\mathbf{X}_k}^{(h)}, \mathbf{c}_{\mathbf{g}}^{(h)})}{\sum_{k=1}^K \text{var}(\mathbf{c}_{\mathbf{X}_k}^{(h)})^2 + \sum_{j=1}^J \sum_{z=1}^Z \text{var}(\mathbf{c}_{\mathbf{Y}_{j(z)}}^{(h)})^2}$ mesure la part de covariance introduite par

chaque tableau de données $\mathbf{X}_k^{(h)}$ dans la composante globale et permet de mesurer plus précisément la contribution du tableau \mathbf{X}_k pour le calcul de l'information commune entre les tableaux.

Le coefficient $\lambda_j^{(h)} = \frac{\text{cov}^2(\mathbf{c}_{\mathbf{Y}_j}^{(h)}, \mathbf{c}_{\mathbf{g}}^{(h)})}{\sum_{k=1}^K \text{var}(\mathbf{c}_{\mathbf{X}_k}^{(h)})^2 + \sum_{j=1}^J \sum_{z=1}^Z \text{var}(\mathbf{c}_{\mathbf{Y}_{j(z)}}^{(h)})^2}$ mesure la part de covariance introduite par

chaque tableau de données $\mathbf{Y}_j^{(h)}$ dans la composante globale et permet de mesurer plus précisément la contribution du tableau à trois entrées \mathbf{Y}_j pour le calcul de l'information commune entre les tableaux.

1.3.3 Corrélacion entre les composantes

Les coefficients de corrélation entre les composantes par bloc et les composantes globales peuvent être proposés pour évaluer le degré de ressemblance entre les composantes par bloc et la composante globale à une dimension donnée h , indépendamment de la variance.

La corrélation du tableau à trois entrées avec la composante globale a été calculée comme la corrélation entre la composante pondérée de l'ensemble des composantes $\mathbf{c}_{\mathbf{Y}_j}^{(h)}$ avec la composante globale $\mathbf{c}_{\mathbf{g}}^{(h)}$.

$$\text{corr}(\mathbf{c}_{\mathbf{Y}_j}^{(h)}, \mathbf{c}_{\mathbf{g}}^{(h)}) \tag{Equation D41}$$

2 Analyse de l'échantillon 4f11mb-2 en utilisant l'analyse de co-inertie multiple trilineaire

2.1 Pondération des tableaux de données

La collection d'images 4f11mb-2 a été analysée en utilisant l'analyse de co-inertie multiple trilineaire. La structure de données analysée contenait un tableau infrarouge à deux entrées de taille 194×261 couplé à deux tableaux de données à trois entrées de fluorescence et de Raman de taille 194×322×25 et 194×346×25, respectivement. La figure D16 montre les deux profils variables globaux obtenus en analyse de co-inertie multiple trilineaire avant décomposition en sous vecteur et normalisation. Ces deux profils variables montrent que le poids du tableau

infrarouge est très petit par rapport à ceux des tableaux Raman et de fluorescence. Pour la première dimension par exemple, la valeur de la contribution du tableau infrarouge à la détermination du profil variable global est de 0.56%. Par contre la contribution des tableaux de fluorescence et Raman est de 51.20% et de 48.24% respectivement (Tableau D11). Les tableaux à trois entrées contiennent 25 fois plus de spectres ce qui leur donne une plus grande variance totale, bien que les spectres aient été normalisés de manière identique pour les trois spectroscopies.

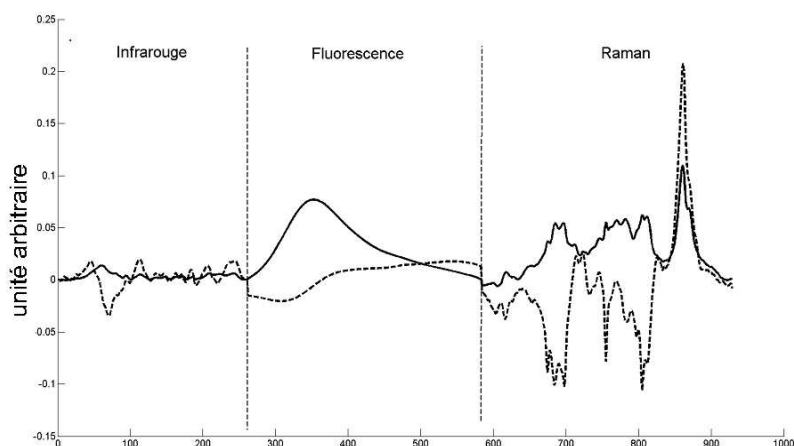


Figure D16 Les deux premiers profils spectraux globaux obtenus en utilisant l'analyse de co-inertie trilineaire.

Nous avons choisi d'équilibrer l'influence de chacun des tableaux *a priori* dans l'analyse. Par exemple, chaque tableau peut être pondéré par la racine carrée de sa première valeur propre ou par la racine carrée de sa trace [Lebart *et al.*, 2000].

La pondération des tableaux de données permet d'équilibrer le rôle de chacun des tableaux et fournit une représentation des composantes et des profils variables qui s'interprètent selon les règles usuelles de l'analyse de données. Deux possibilités de pondération ont été testées sur notre structure de données (Tableau D11). La pondération avec la racine carrée de la trace fait augmenter la contribution du tableau infrarouge à la détermination du profil variable global de 0.56% à 31.99% pour la première dimension alors que la pondération avec la racine carrée de la première valeur propre du tableau infrarouge ne l'a fait augmenter que de 0.56% à 5.28%. Cette augmentation est valable également pour les autres composantes. La pondération avec la racine carrée de la trace de chaque tableau a été retenue par rapport à la pondération avec la première valeur propre.

Tableau D11 Contribution du tableau infrarouge, fluorescence et Raman à la formation du profil variable global. La première ligne représente les valeurs de $\|\tilde{\mathbf{u}}_k^{(h)}\|$ et $\|\tilde{\mathbf{v}}_j^{(h)}\|$ sans pondération du tableau infrarouge. La deuxième ligne représente les valeurs de $\|\tilde{\mathbf{u}}_k^{(h)}\|$ et $\|\tilde{\mathbf{v}}_j^{(h)}\|$ en pondérant le tableau infrarouge par la racine carrée de la plus grande valeur propre. La troisième ligne représente les valeurs de $\|\tilde{\mathbf{u}}_k^{(h)}\|$ et $\|\tilde{\mathbf{v}}_j^{(h)}\|$ en pondérant le tableau infrarouge par la racine carrée de la trace.

Composantes		1	2	3	4	5
sans pondération	infrarouge	0.56	3.03	1.93	5.85	32.35
	fluorescence	51.20	6.57	0.30	4.65	2.90
	Raman	48.24	90.40	97.77	89.50	64.75
pondération avec la première valeur propre	infrarouge	5.28	28.13	98.05	76.95	92.80
	fluorescence	57.22	6.74	0.15	5.84	1.30
	Raman	37.50	65.13	1.80	17.21	5.90
pondération avec la trace	infrarouge	31.99	77.28	99.72	96.34	99.29
	fluorescence	49.45	3.47	0.05	1.50	0.30
	Raman	18.56	19.25	0.23	2.16	0.41

2.2 Analyse de la collection de données 4f11mb-2

Les tableaux pondérés ont été soumis à l'analyse de co-inertie multiple trilinéaire. Le tableau D12 résume les indicateurs présentés dans la section II.2.3 du chapitre précédent.

Les valeurs $\lambda_g^{(h)}$ ont été calculées pour les 5 premières composantes. Les valeurs de $\lambda_k^{(h)}$ sont comprises entre 74% et 97%.

Tableau D12 Indicateurs de la co-inertie trilinéaire

Composantes		1	2	3	4	5
$\lambda_g^{(h)}$		73.70	90.48	96.67	95.37	93.52
$\lambda_k^{(h)}$	infrarouge	23.58	69.93	96.42	91.90	92.86
	fluorescence	36.45	3.15	0.03	1.40	0.13
	Raman	13.68	17.40	0.22	2.07	0.53
corrélation avec la composante globale	infrarouge	0.82	0.99	0.99	0.99	0.99
	fluorescence	0.92	0.79	0.22	0.82	0.25
	Raman	0.78	0.86	0.27	0.59	0.32

Pour la composante 1 la somme des contributions des trois tableaux à la détermination de la composante globale est égale à 73.70%, 23.58% sont dus au tableau infrarouge, 36.45% sont dus au tableau de fluorescence et 13.68% sont dues au tableau Raman. Le tableau de fluorescence contribue le plus à former la composante globale de la première dimension. Par conséquent, cette composante ressemble plus à une composante de fluorescence. Le coefficient de corrélation entre la composante globale et la première composante infrarouge, de fluorescence et Raman est respectivement de 0.82, 0.92 et 0.78. Les trois premières composantes sont donc corrélées entre elles.

Les tableaux infrarouge et Raman ont contribué le plus à former la deuxième composante globale. Le tableau de fluorescence a contribué seulement à 3,15%. Les deuxièmes composantes infrarouge, fluorescence et Raman sont corrélées entre elles.

Pour les autres composantes, l'infrarouge a contribué le plus à former la composante globale. Les indicateurs démontrent que les composantes 3, 4 et 5 sont principalement créées par le tableau infrarouge. En effet le coefficient de corrélation entre la composante infrarouge et la composante globale est de 0.99 pour ces trois composantes.

La pondération des tableaux de données par la racine carrée de leur trace a ramené à des indicateurs similaires à ceux obtenus en analyse de co-inertie multiple.

2.3 Interprétation spectrale et spatiale

Les profils spectraux obtenus en analyse de co-inertie trinéaire sont identiques à ceux obtenus avec l'analyse de co-inertie multiple. Dans cette section, les quatre premières composantes sont interprétées et discutées. Ces composantes décrivent 91% de la variance totale pour l'infrarouge, 99% pour la fluorescence et 98% pour le Raman.

La première composante décrit 54% de la variance totale pour l'infrarouge, 98 et 84% pour la fluorescence et le Raman, respectivement. Ces pourcentages de variance par bloc restent proches des pourcentages de variance de l'analyse en composantes principales pour les trois tableaux. Ceci est valable pour toutes les autres composantes (Tableau D13,14 et 15).

Tableau D13 Résultats du tableau de données infrarouge. La première ligne représente les pourcentages de variance de chaque composante principale. La deuxième ligne représente les pourcentages de variance de chaque composante par bloc.

Composantes		1	2	3	4	5
ACP	pourcentage de variance de chaque composante principale	54.72	23.47	9.94	2.97	2.62
ACOMT	pourcentage de variance de chaque composante par bloc	53.62	24.40	9.98	3.00	2.64

Tableau D14 Résultats du tableau de données de fluorescence. La première ligne représente les pourcentages de variance de chaque composante principale. La deuxième ligne représente les pourcentages de variance de chaque composante par bloc.

Composantes		1	2	3	4	5
ACP	pourcentage de variance de chaque composante principale	97.64	1.91	0.17	0.10	0.04
ACOMT	pourcentage de variance de chaque composante par bloc	97.63	1.91	0.11	0.09	0.14

Tableau D15 Résultats du tableau de données Raman. La première ligne représente les pourcentages de variance de chaque composante principale. La deuxième ligne représente les pourcentages de variance de chaque composante par bloc.

Composantes		1	2	3	4	5
ACP	pourcentage de variance de chaque composante principale	84.68	12.16	0.94	0.37	0.20
ACOMT	pourcentage de variance de chaque composante par bloc	84.53	11.95	0.90	0.40	0.32

2.3.1 Composante 1

Comme attendu, les premiers profils spectraux révèlent des variations d'intensité entre les spectres. Les mêmes composantes que l'analyse de co-inertie multiple ont été obtenues en analyse de co-inertie trilineaire (Fig. D17-a, D17-b et D17-c). Ainsi, la même interprétation spectrale a été obtenue en utilisant l'analyse de co-inertie multiple et l'analyse de co-inertie multiple trilineaire. Les valeurs des composantes images infrarouge, de fluorescence et Raman ont été calculées, repliées et affichées en tant qu'images composites avec l'image visible de référence pour révéler les variations spatiales. La résolution des données de fluorescence et Raman étant conservée, la structure cellulaire de la région analysée est clairement distinguée sans avoir recours aux projections sur les individus supplémentaires tel que pour le cas de l'analyse de co-inertie multiple classique. Les images composantes 1 (Fig. D18-a, D18-b et D18-c) révèlent une intensité maximale sur les parois et une intensité minimale dans les fonds de cellule. Les pixels négatifs sont localisés en rouge sur les parois des cellules. Les pixels négatifs en bleu correspondent à l'intérieur de cellules.

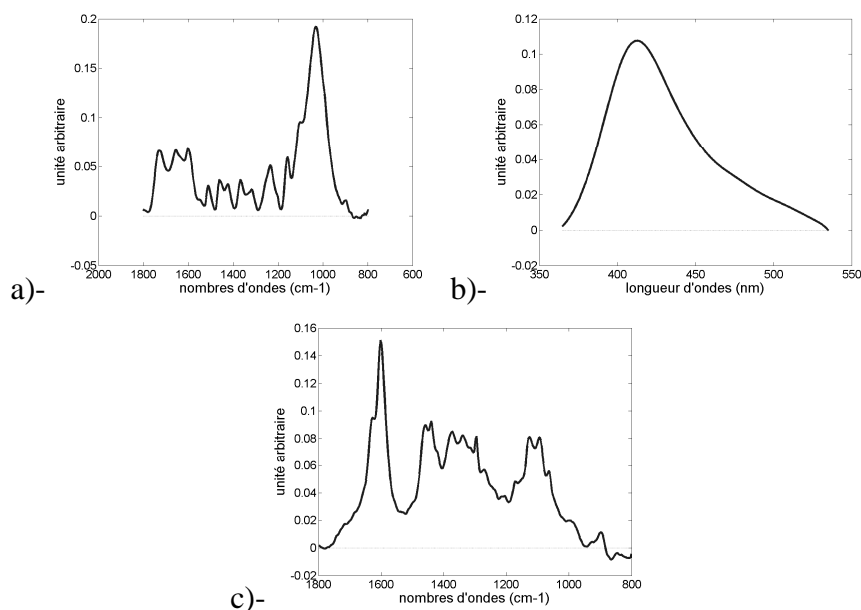


Figure D17 Profils spectraux 1 de l'analyse de co-inertie trilineaire a)- infrarouge b)- de fluorescence c)- Raman

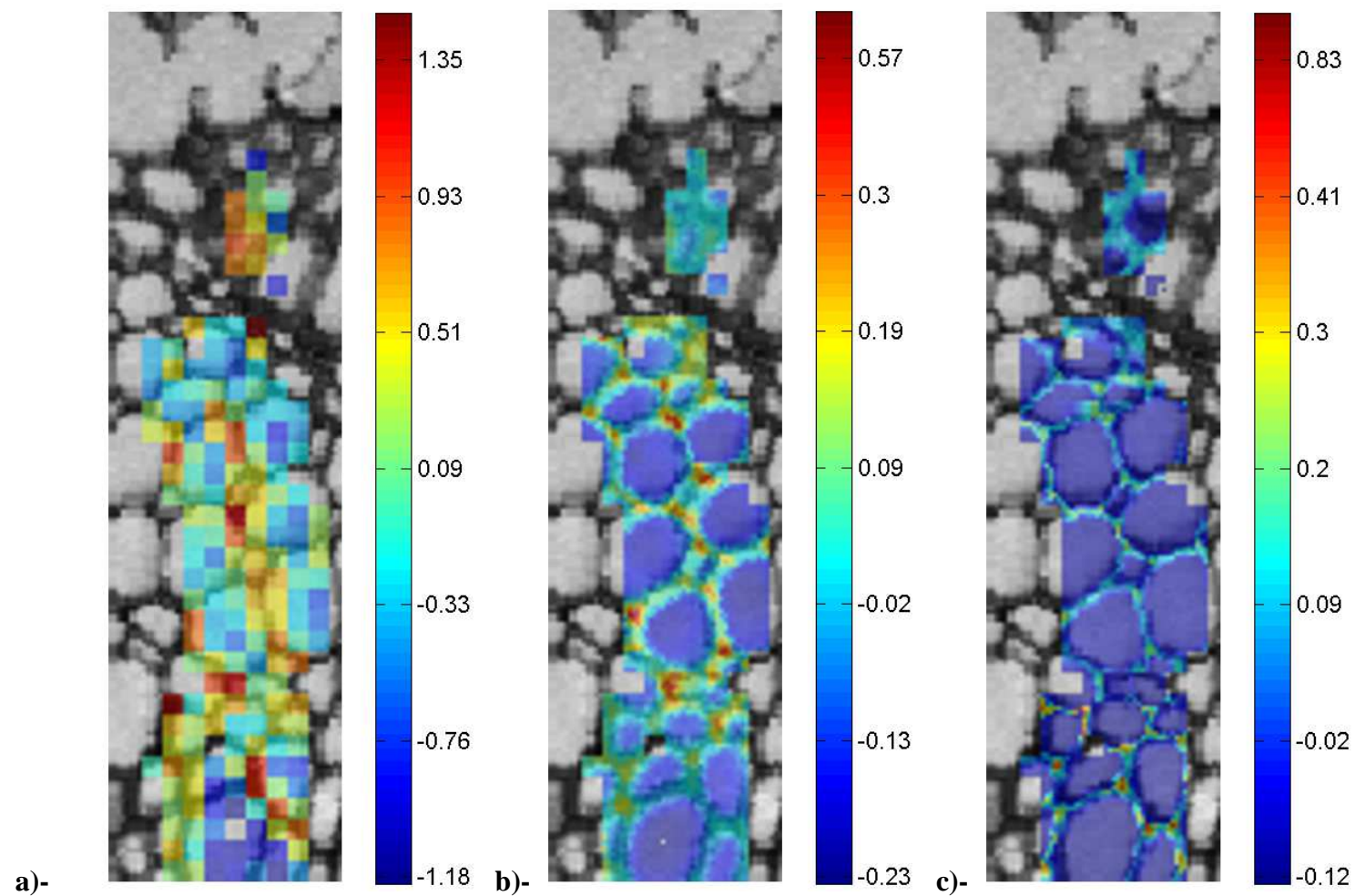


Figure D18 Images composantes 1 de l'analyse de co-inertie trinéaire a)- infrarouge b)- de fluorescence c)- Raman .

2.3.2 Composante 2

Le deuxième profil spectral infrarouge $u_1^{(2)}$ oppose des pics en positif à 1068, 1110 cm^{-1} révélant les polysaccharides (cellulose et xylane) à des pics négatifs à 1230, 1419, 1450, 1504, 1593 cm^{-1} révélant les lignines et les liaisons esters. Le deuxième profil spectral de fluorescence $v_1^{(2)}$ oppose un pic positif à 390 nm à un pic négatif dans la région des lignines. Le profil spectral Raman $v_2^{(2)}$ oppose un pic négatif (1602-1631) révélant la lignine à un ensemble de pics positifs révélant les polysaccharides.

Les images composantes 2 (Fig. D20-a, 20-b et 20-c) ont révélé en couleur rouge les parois cellulaires du phloème avec la plus grande intensité au niveau des jonctions cellulaires. Les pixels bleus correspondent à la région du sclérenchyme. Les pixels correspondant au cellules du xylème sont également bleus mais avec une intensité plus faible que le sclérenchyme. Considérant l'interprétation des profils spectraux, les acides hydroxycinnamiques (acide férulique et acide para-coumarique) ont été localisés principalement dans le phloème. Les cellules du sclérenchyme et du xylème sont enrichies en lignine co-localisée avec des liaisons esters mais avec une teneur moindre dans les cellules du xylème.

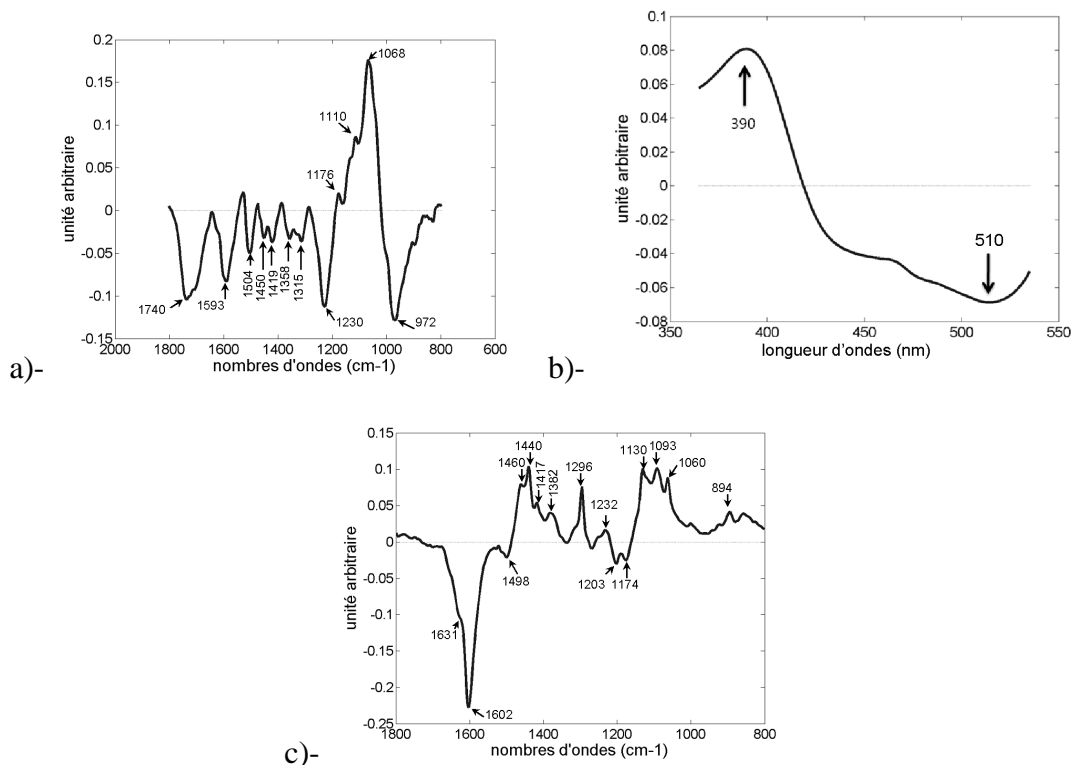


Figure D19 Profils spectraux 2 de l'analyse de co-inertie trilineaire a)- infrarouge, b)- de fluorescence, c)- Raman.

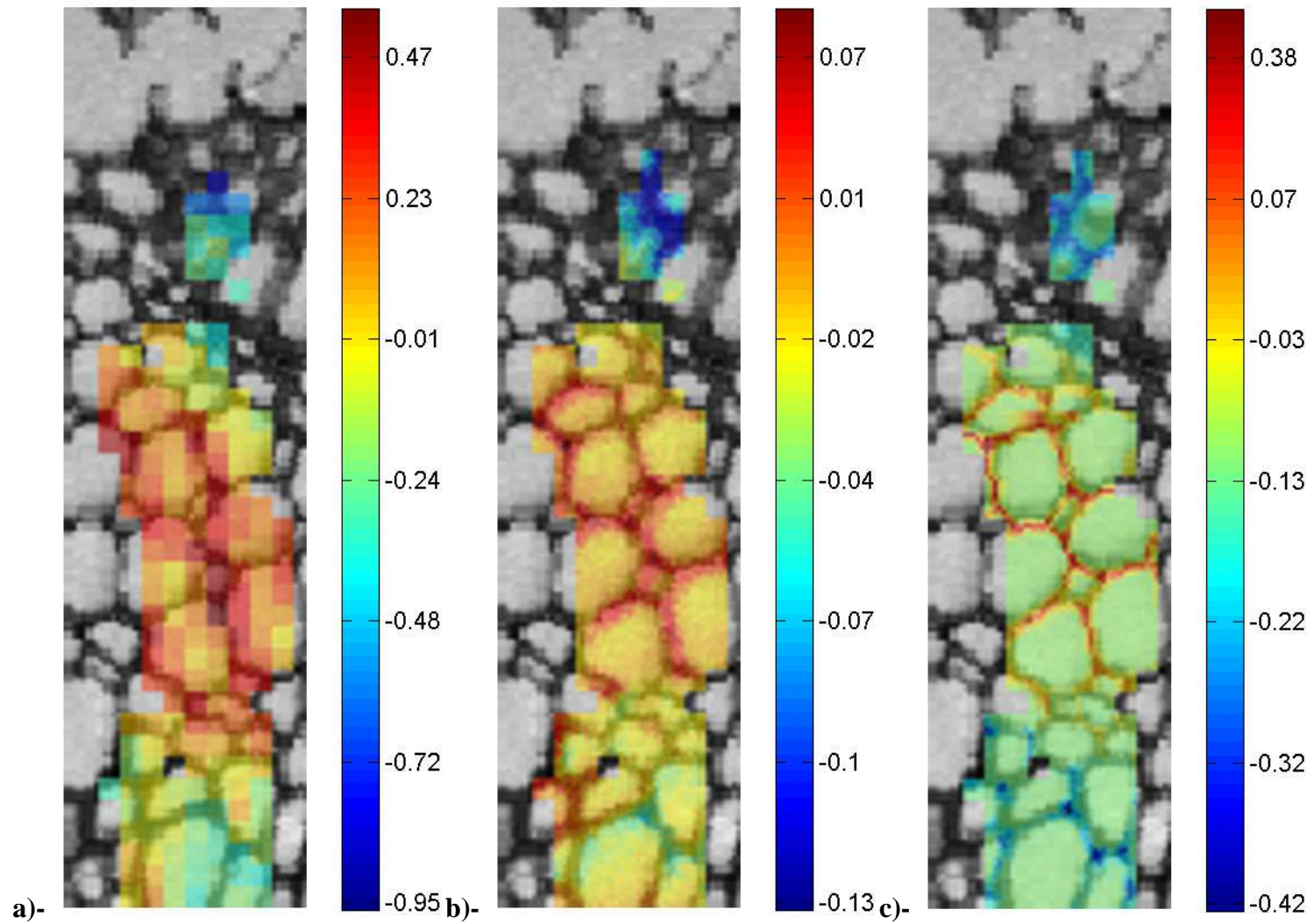


Figure D20 Images composantes 2 de l'analyse de co-inertie trilineaire a)- infrarouge b)- de fluorescence c)- Raman.

2.3.3 Composante 3

Le troisième profil spectral infrarouge $\mathbf{u}_1^{(3)}$ révèle les lignines et les liaisons esters en positif (Fig. D21-a). Le profil spectral de fluorescence $\mathbf{v}_1^{(3)}$ présente deux pics positifs qui sont caractéristiques des lignines (Fig. D21-b). Le profil spectral Raman $\mathbf{v}_2^{(3)}$ est bruité et non interprétable (Fig. D21-c).

L'image composante infrarouge 3 (Fig. D22-a) révèle en positif les lignines. Les pixels rouges apparaissent sur toute la paroi y compris les parois du phloème mais avec un profil spectral différent de celui de la deuxième composante. Les pixels rouges sont plus intenses dans les cellules de sclérenchyme et la zone frontière séparant le phloème du xylème et du sclérenchyme. L'image composante de fluorescence (Fig. D22-b) distingue deux petites régions apparaissant en rouge correspondant aux parois cellulaires de la frontière séparant le phloème des autres types cellulaires. Les parois cellulaires du sclérenchyme apparaissent également en rouge. L'image composante 3 du Raman n'a pas été interprétée parce que le profil variable Raman est bruité et non interprétable.

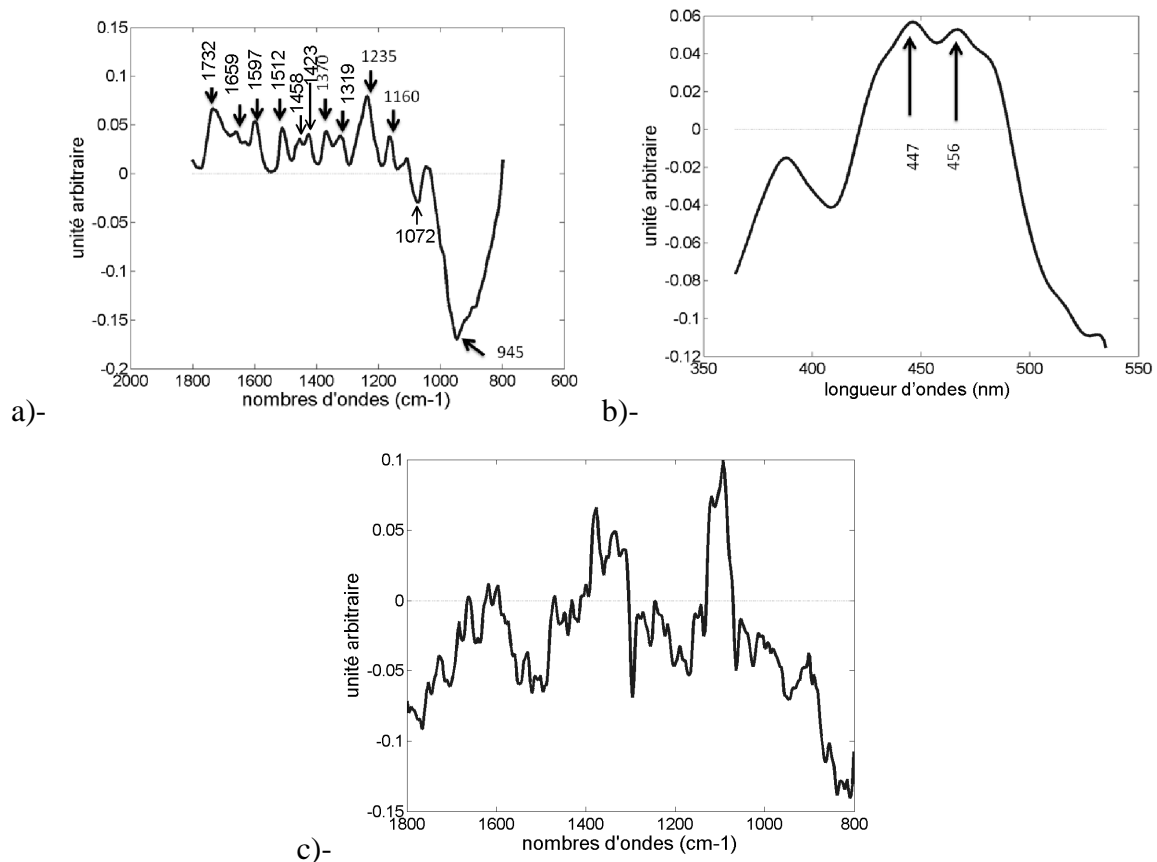


Figure D21 Profils spectraux 3 de l'analyse de co-inertie multiple a)- infrarouge b)- de fluorescence c)- Raman

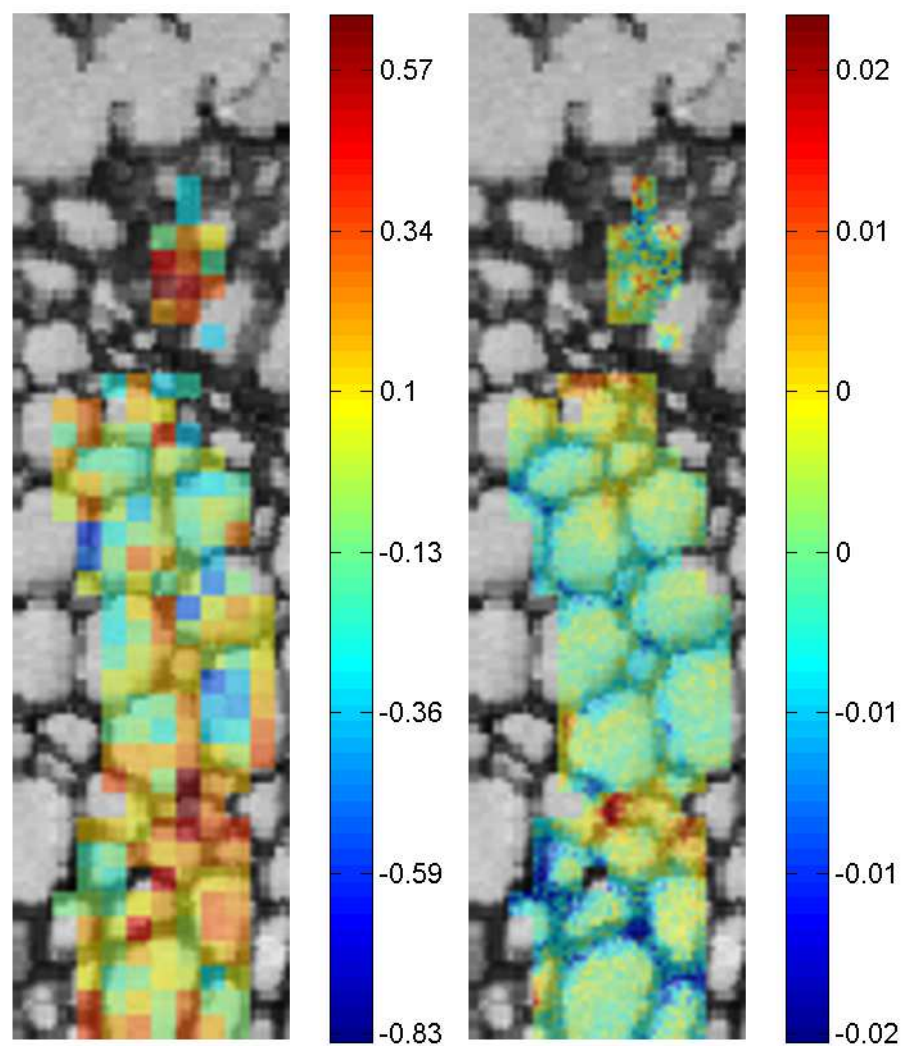


Figure D22 Images composantes 3 de l'analyse de co-inertie trilineaire a)- infrarouge b)- de fluorescence.

2.3.4 Composante 4

Dans l'analyse de co-inertie trilinéaire, nous avons poussé l'interprétation jusqu'à la quatrième composante. Le quatrième profil spectral infrarouge $u_1^{(4)}$ révèle beaucoup de pics en positif et négatif (Fig. D23-a). Son interprétation est assez délicate. Le profil spectral de fluorescence $v_1^{(4)}$ présente trois pics négatifs, un dans la région des acides hydroxycinnamiques et les deux autres sont caractéristiques des lignines [Radotic *et al.*, 2006] (Fig. D23-b). Le profil spectral Raman $v_2^{(3)}$ était bruité et non interprétable, il n'a pas été présenté dans cette analyse.

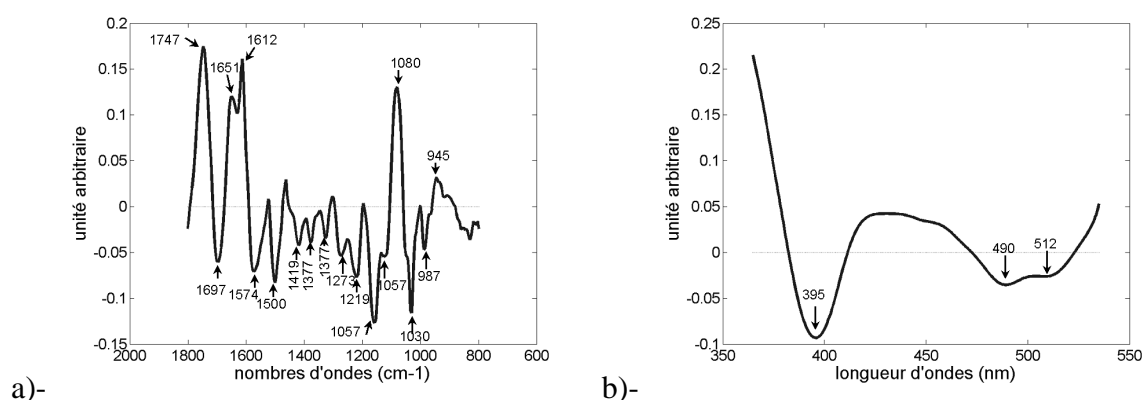


Figure D23 Profils spectraux 4 de l'analyse de co-inertie trilinéaire a)- infrarouge b)- de fluorescence.

Malgré la difficulté de l'interprétation des profil spectraux, les composantes images montrent que les pixels appartenant aux cellules du phloème n'ont pas participé à cette composante (Fig D24 a-b). En effet, elles apparaissent en bleu clair sur la composante infrarouge et en vert sur la composante de fluorescence. La couleur bleu clair et verte sont associées respectivement, à une valeur de composante nulle en infrarouge et en fluorescence. Les composantes opposent des cellules du xylème aux cellules de sclérenchyme. Les cellules du xylème apparaissent en bleu et les cellules de sclérenchyme apparaissent en rouge.

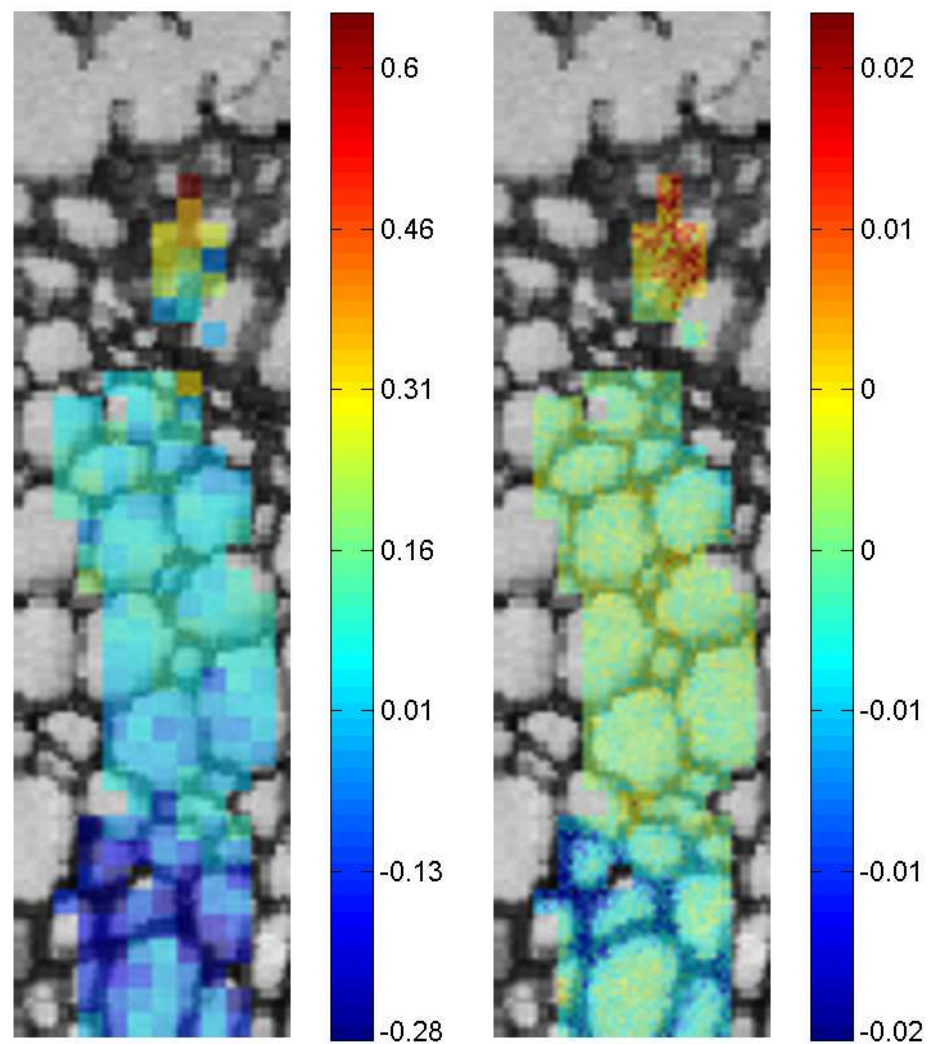


Figure D24 Images composantes 4 de l'analyse de co-inertie trinéaire a)- infrarouge b)- de fluorescence.

2.4 Interprétation de la troisième entrée des blocs à trois entrées

La décomposition en utilisant l'analyse de co-inertie multiple trilineaire permet de calculer des profils variables relatifs à la troisième entrée spatiale de chacun des tableaux de fluorescence et Raman. Les profils variables ont été repliés et affichés sous la forme d'une image en niveaux de gris. La dimension de l'image correspond à la dimension de l'imagette en fluorescence et en Raman. Trois profils variables pour le tableau de fluorescence ainsi que pour le tableau Raman ont été calculés et affichés. Les profils variables 1, 2 et 3 sont affichés en bleu, vert et rouge. Dans notre cas, la dimension de l'imagette est 5×5 pixels. Dans les figures D-25 et D-26, nous avons présenté respectivement l'ensemble des profils variables avant et après repliement.

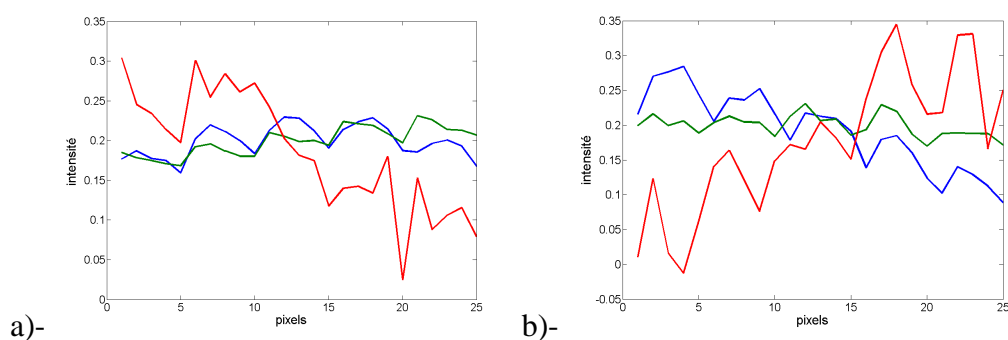


Figure D25 Profils variables de la troisième entrées a)- de fluorescence b)- Raman.

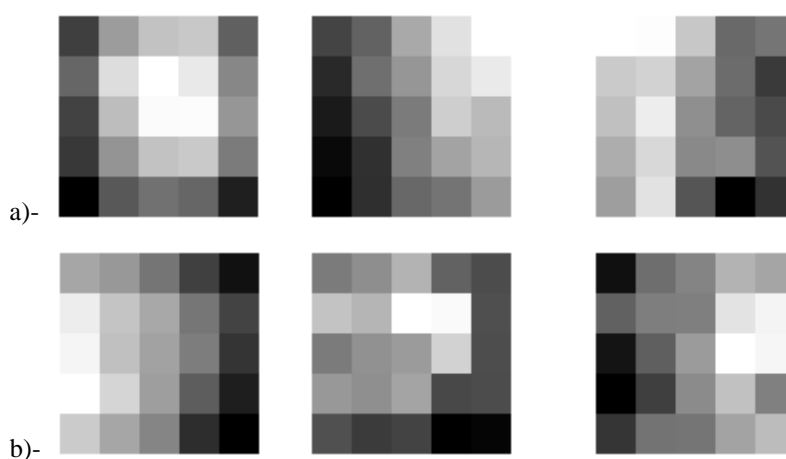


Figure D26 Profils variables repliés de la troisième entrées a)- de fluorescence b)- Raman.

Les profils variables de la troisième entrée sont tous positifs ce qui est attendu puisqu'ils représentent la covariance entre la composante globale et les composantes de fluorescence et Raman. Les profils montrent une structure ayant un effet dents de scie donnant une plus

grande importance aux pixels du centre. Cet effet est voisin de celui observé pour un profil un profil type Gaussien (Fig. D-27).

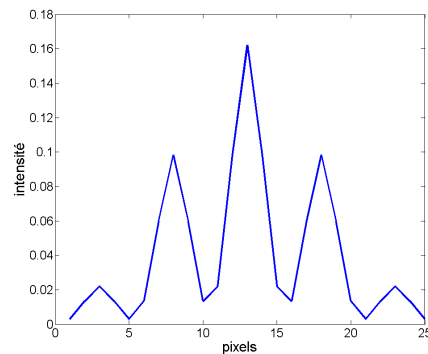


Figure D27 Profil de type Gaussien

Les profils 1 et 2 sont très semblables. Les profils 3 de fluorescence montre des pixels beaucoup plus intense que d'autres, le profils Raman le montre également mais dans le sens contraire. Ceci peut être dû au fait que les pixels fluorescence et Raman ne soient pas appariés un à un. En effet, la procédure d'appariement a été développée pour faire correspondre des imagerie à des pixels infrarouge et non pas les pixels de fluorescence avec les pixels Raman.

Conclusion

Dans ce chapitre, l'analyse de co-inertie multiple a été étendue pour analyser des images hyperspectrales multirésolutions sans altérer leurs qualités. Le développement de cette méthode s'est basé sur le même principe utilisé dans l'analyse inter-batterie de Tucker. Une pondération des différents tableaux est apparue nécessaire pour tenir compte de la variance globale apportée par les tableaux. Les résultats sont similaires avec ceux obtenus en analyse de co-inertie multiple développée dans le deuxième chapitre.

Conclusion

Dans cette partie, plusieurs images hyperspectrales issues de différentes spectroscopies ont été combinées afin de caractériser les parois cellulaires dans les sections de tige de maïs. L'imagerie hyperspectrale infrarouge, de fluorescence en utilisant une source synchrotron et Raman en utilisant une source laser ont été utilisées pour analyser les polysaccharides, les lignines et les composés phénoliques présents dans les parois cellulaires. L'une des caractéristiques des données est que les pixels infrarouges, de fluorescence et Raman ont des résolutions différentes. Les tableaux de données appariés ont été construits en préservant cette

différence de résolution. La structure de données obtenue consiste en un tableau de données à deux entrées couplé à deux tableaux de données à trois entrées. Deux méthodes multivariées ont été étudiées et étendues pour analyser une telle structure de données. La première méthode est l'analyse inter-batterie de Tucker appliquée à deux tableaux à deux entrées. Le tableau de données à trois entrées a été moyenné selon la troisième entrée pour obtenir un tableau de données classique à deux entrées. La méthode a été jugée pertinente pour le couplage d'une manière symétrique deux différentes techniques de spectroscopie. La méthode a été étendue pour le couplage d'un tableau de données à trois entrées avec un tableau de données à deux entrées. Le point fort de cette méthode est de permettre l'analyse conjointe de ces deux tableaux de données sans altérer leurs qualités. Une deuxième méthode, l'analyse de co-inertie multiple, a été par la suite utilisée pour coupler plusieurs images : infrarouge, de fluorescence et Raman. Outre sa robustesse, l'analyse de co-inertie multiple permet de calculer des profils variables globaux et par bloc, des composantes synthétisant l'information de tous les tableaux et des composantes par bloc permettant d'identifier la contribution de chaque spectroscopie à l'analyse conjointe des deux tableaux. L'analyse de co-inertie multiple a été également étendue pour l'analyse d'un mélange de tableaux à deux et à trois entrées.

Discussions

L'objectif de la thèse était de réaliser les développements méthodologiques nécessaires pour le couplage d'images hyperspectrales multirésolutions acquises avec le rayonnement synchrotron dans le moyen infrarouge, en fluorescence UV et avec une source conventionnelle en Raman. Le contexte scientifique était la caractérisation de l'hétérogénéité de composition chimique des parois cellulaires des tiges de maïs, plante considérée comme modèle et d'application pour la production de bio-carburant. L'analyse chimique fine des polymères dans les parois devrait permettre d'identifier les verrous à la saccharification enzymatique de ces parois. La microspectroscopie est un outil puissant et rapide pour analyser *in situ* la distribution des polymères dans les parois cellulaires à l'échelle de quelques micromètres.

Les parois cellulaires des tiges de maïs sont majoritairement composées de polysaccharides, de lignines et d'acides hydroxycinnamiques. L'infrarouge est particulièrement adapté à l'analyse des polysaccharides et des lignines. Le Raman répond fortement en présence des groupements phénoliques contenus dans les lignines et les acides hydroxycinnamiques. Il répond également aux polysaccharides avec une sensibilité différente de celle de l'infrarouge. La fluorescence à 280 nm est sensible aux groupements phénoliques. Les trois techniques sont à la fois complémentaires et redondantes.

Cette thèse porte sur l'étude le couplage d'images hyperspectrales comme moyen de caractérisation *in situ* de la composition chimique des parois cellulaires. Pour relever ce défi, nous avons adopté une démarche à la fois méthodologique et appliquée. Cette démarche a été construite autour des cinq étapes suivantes :

- Etablissement d'un protocole d'acquisition conjointe de données.
- Pré-traitement des données acquises.
- Appariement des données issues des trois microspectroscopies infrarouge, fluorescence et Raman et établissement de la structure de données.
- Développement des méthodes multibloc pour l'analyse de la structure de données établie.
- Analyse et interprétation des résultats obtenus en appliquant les méthodes multibloc.

1 Acquisition conjointe des données

L'acquisition des images avec les trois microspectroscopies pour un même échantillon a nécessité de définir un protocole de préparation des coupes et de choisir un support d'échantillon compatible pour les trois microspectroscopies. En fonction des contraintes d'acquisition de chaque technique, nous avons convenu d'effectuer des coupes sériées d'épaisseur de 10 μm . Le choix de cette épaisseur a été guidée par la qualité des spectres obtenus pour les trois génotypes prévus d'être analysés. Il n'a pas été possible d'obtenir des signaux suffisamment intenses pour le stade 9 feuilles et des signaux non saturés au stade ensilage pour tous les types cellulaires. Le travail a été réalisé uniquement à partir de coupes au stade floraison femelle. L'obtention de spectres de bonne qualité en infrarouge au stade ensilage et au stade 9 feuilles nécessiterait d'adapter les conditions d'acquisitions (taille d'ouverture, temps d'acquisition...) en fonction de la morphologie des parois cellulaires (parois fines, parois épaisses). Une possibilité serait d'acquérir des images ne contenant qu'un seul type cellulaire, par exemple le phloème présentant des parois fines ou le sclérenchyme caractérisé par des parois épaisses, et de reconstituer l'ensemble de l'image spectrale par assemblage d'images. Dans notre étude, nous avons acquis des images infrarouges contenant différents types cellulaires avec les mêmes conditions d'acquisitions.

Le protocole d'acquisition conjointe consiste à acquérir des spectres de la même région pour les trois techniques. Une image de référence a facilité le recalage des images en fluorescence et Raman. Le repérage des régions à cartographier est assez long et l'accès au synchrotron est limité. Pendant ma deuxième année thèse, le protocole d'acquisition a été préparé au préalable à partir d'images acquises au microscope confocal de notre laboratoire (INRA-Nantes).

Des collections d'images ont été acquises au bord et au centre de la tige avant et après dégradation enzymatique. Nous avons pu analyser, au cours de la thèse, les collections d'images acquises dans un faisceau du bord de la tige avant dégradation enzymatique. Les autres collections n'ont pas été totalement exploitées. Il sera intéressant d'analyser toutes les données collectées afin de comparer la composition chimique des parois cellulaires avant et après dégradation enzymatique et à différents endroits de la tige.

2 Pré-traitement des spectres

Avant d'envisager l'analyse conjointe des données acquises selon les trois modalités, il était nécessaire d'explorer séparément chacune des images hyperspectrales. Le but de cette étape

était de déterminer les informations apportées par chacune des techniques et la nature des pré-traitements éventuels des images ou des spectres. Les pré-traitements ont pour objectif la correction des déformations spectrales et spatiales. Une panoplie de méthodes existe en littérature pour pré-traiter les spectres infrarouges, fluorescence et Raman [Lasch, 2012]. L'absence de méthodes de validation fait que les méthodes de pré-traitement sont basées essentiellement sur l'expérience des spectroscopistes.

Le processus de pré-traitement est une exécution séquentielle d'au moins deux étapes de correction de spectres. Ces étapes sont la correction de la ligne de base, l'élimination du bruit et la normalisation des spectres.

De manière classique, en spectroscopie, les spectres sont normalisés individuellement en donnant à chacun la même intensité globale : par exemple en divisant chaque spectre par la somme des intensités totale ou la somme des intensités sous un pic ou bien par leur norme. Dans notre cas, cela reviendrait à donner la même importance aux spectres de l'intérieur des cellules qu'aux spectres de parois et donc à introduire des spectres de bruit dans l'analyse. Une idée naturelle est de fixer un seuil permettant d'éliminer les spectres correspondant à une absence de signal. L'application d'un seuil relatif n'est pas évidente et présente plusieurs inconvénients : des spectres pertinents de faible intensité seront éliminés, la structure cellulaire spatiale liée à l'acquisition d'image hyperspectrale est perdue. Il était donc nécessaire, de développer une méthode de pré-traitement qui tienne compte de la nature spatiale et spectrale des données. Le principe général de la normalisation spatiale et spectrale était de calculer une carte des coefficients de normalisation qui tiennent compte des pixels voisins. Une autre idée serait d'utiliser ces coefficients au moment de l'analyse. Par exemple les coefficients pouvaient servir de pondération des individus en analyse en composantes principales.

La méthode développée a été appliquée aux spectres infrarouge, fluorescence et Raman. Les spectres infrarouge présentaient beaucoup plus de déformations que les spectres Raman et de fluorescence. Quelques spectres présentant des valeurs négatives ont été éliminés après pré-traitement. Une méthode de classification, tel que la Support Vector Machine 'SVM' pourrait être utile pour trier et éliminer automatiquement les spectres aberrants avant pré-traitement [Widjaja *et al.*, 2008].

3 Appariement des données

L'analyse conjointe des images acquises par les trois méthodes de microspectroscopies nécessite d'apparier les spectres issus de chaque modalité. La première étape du processus

était d'identifier les spectres acquis à un endroit donné pour chaque image spectrale. Cette étape a été réalisée au moyen de techniques de recalage d'images. Les images hyperspectrales se présentent sous la forme d'un cube de données. Une représentation sous la forme d'une image monochrome est nécessaire pour pouvoir appliquer le recalage. Différentes approches sont possibles : sélection d'une longueur d'onde caractéristique, calcul de l'aire sous un pic, le rapport de l'aire sous deux pics ou de la somme sous le spectre. L'image la plus contrastée est choisie afin de distinguer les structures cellulaires.

Sur ces bases, nous avons développé une méthode de recalage rigide semi-automatique qui prend en compte la rotation, la translation et le changement d'échelle. L'algorithme développé nécessite l'intervention de l'utilisateur pour pré-traiter les images à recaler et les préparer à la procédure de recalage. Cette intervention nécessite le calcul d'une image monochrome à partir de l'image hyperspectrale ou la transformation éventuelle d'une image dans le cas un effet miroir. Par exemple, un effet miroir est observé entre l'image visible et l'image spectrale en fluorescence. L'utilisateur doit également intervenir pour définir l'intervalle de recherche de l'angle de rotation ce qui permet de minimiser le temps de calcul.

En développant notre procédure de recalage, nous avons supposé qu'il n'y avait pas de déformations au sein des images puisque la même coupe était cartographiée par les trois techniques. Bien que cette hypothèse se soit révélée juste dans la plus part des cas, quelques déformations locales ont été observées notamment au sein de quelques images Raman. Les déformations étaient liées à des différences de plan focal ou au vieillissement de la coupe. Ce problème peut être résolu en utilisant une méthode de recalage d'image qui tient en compte les déformations locales [Ruzena et stane, 1989].

4 Analyse multibloc

4.1 Choix des méthodes

Les méthodes d'analyse structurées en bloc permettent de calculer des composantes synthétiques globales et des composantes spécifiques en se basent sur un critère de covariance entre la composante globale et les composantes par bloc [Hanafi et Kiers, 2006 ; Hanafi *et al.*, 2011]. Les propriétés des composantes et des profils variables résultant sont fortement liées au critère d'optimisation, à la déflation ainsi qu'aux contraintes de normalisation imposées. Le choix d'une méthode d'analyse par rapport à une autre dépend des propriétés des composantes et des profils variables recherchés. Dans notre cas nous avons choisi d'analyser nos données avec des méthodes symétriques fournissant des profils variables orthogonaux.

La structure de données obtenue après appariement de spectres est constituée d'un mélange de blocs à deux et à trois entrées. L'analyse des images hyperspectrales multirésolutions par les méthodes multibloc n'est pas immédiate. Les méthodes existantes, pouvant coupler des images hyperspectrales à une même résolution, doivent être adaptées pour pouvoir supporter la structure de données établie.

4.2 Critère d'optimisation

La covariance a été choisie pour chercher les liens entre les différents blocs en décrivant le maximum de variance pour chaque bloc. Le critère de covariance a été choisi pour décrire la ressemblance entre les tableaux en tenant en compte des informations de plus grande variance. Le pourcentage de variance décrit par la composante 1 dans l'analyse de la collection d'images 4f11mb-2 était de 97% pour la fluorescence, 80% pour le Raman et 53% pour l'infrarouge. Le pourcentage de variance expliqué est lié à la structure et à la nature de l'information fournie par chaque bloc. Cette première composante était liée à l'intensité des spectres. La variance résiduelle pour chaque tableau donne un poids beaucoup plus important au tableau de l'infrarouge.

Dans notre cas, nous avons constaté que la différence de résolution jouait un rôle dans le cas de la co-inertie trinéaire ce qui a conduit à normaliser les tableaux par rapport à leur trace.

Une autre alternative est d'appliquer une pondération à chaque étape de la décomposition pour donner aux tableaux la même importance après chaque étape de déflation.

4.3 Déflation

Plusieurs choix de déflation sont possibles dans le cadre des méthodes multibloc. Afin de faciliter l'interprétation, nous avons choisi une méthode de déflation fournissant des profils variables orthogonaux. Nous avons choisi d'appliquer une déflation sur les profils variables par bloc. Cette déflation fournit dans notre cas des résultats proches de ceux obtenus en faisant une décomposition de chaque bloc par une analyse en composantes principales. Dans le cas de l'analyse en composantes principales consensuelles, la déflation est appliquée sur les composantes globales. Dans ce cas, il n'y a plus d'orthogonalité entre les profils variables et la tâche de l'interprétation devient beaucoup plus compliquée.

4.4 Complémentarité et redondance des méthodes

Pour interpréter la complémentarité et la redondance des méthodes, des indicateurs se basant sur la covariance entre la composante globale et les composantes par bloc ont été développés. Ces indicateurs ont permis de situer la composante globale par rapport à tout l'ensemble des données et spécifiquement par rapport à chaque bloc. La covariance étant très liée à la variance, la prenant comme critère pour évaluer la complémentarité et la redondance entre les méthodes n'est pas toujours évident lorsque les variances diffèrent d'un bloc à l'autre.

D'autres critères peuvent être utiles tel que la corrélation entre la composante globale et les composantes par bloc. La corrélation entre les composantes par bloc et la composante globale montre le degré de ressemblance entre les composantes et permet ainsi de savoir quelles sont les composantes par bloc qui ressemblent le plus à la composante globale, indépendamment de la variance décrite dans les tableaux.

4.5 Profils variables et composantes

Les méthodes multibloc fournissent à la fin de la décomposition des profils variables pour l'analyse spectrale, des composantes communes et spécifiques à chaque bloc qui peuvent être repliées et affichées sous la forme d'images en niveaux de gris ou d'images composites en couleurs pour l'analyse spatiale. Des cartes factorielles permettant de distinguer les individus 'les pixels' en fonction de deux composantes, par bloc ou composantes communes, pourraient être également établis.

4.6 Profils variables de la troisième entrée

Nous avons ébauché une analyse des profils variables de la troisième entrée relative aux pixels imagettes. Dans notre travail, les imagettes ont été dépliées de manière naturelle pour constituer la structure des données appariées. Les imagettes constituent 25 pixels relatives à $5 \times 5 \mu\text{m}^2$ représentant soit un morceau de paroi pouvant être orienté de manière aléatoire, soit une paroi, soit un espace à l'intérieur d'une cellule. Cela a par conséquent que les tranches des tableaux n'ont aucun ordre.

Une idée pour obtenir des profils variables interprétables serait d'adopter une autre méthode de dépliement qui prendrait en compte par exemple un ordre donné, comme par exemple l'intensité en niveaux de gris des pixels, pour étudier la conséquence sur la décomposition multibloc.

4.7 Apport des méthodes multibloc

Un aspect intéressant des méthodes multibloc développées dans ce travail de recherche est leur généralité. Excepté leur apport dans le domaine de couplage d'images hyperspectrales, ces méthodes sont également applicables aux données de même structure provenant d'autres domaines ayant des problématiques de traitement similaires, tels que la sensométrie, le marketing ou l'écologie. Par exemple, les méthodes développées peuvent être appliquées à des données physico-chimiques associées à des données d'une analyse biochimique du même produit. Les données organisées en K tableaux ayant un nombre d'entrées hétérogènes sont de plus en plus rencontrées dans la pratique du fait de l'augmentation du nombre de variables recueillies et de la complexité des questions posées, notamment dans les domaines biologiques.

5 Analyse et interprétation des données

Dans cette thèse deux collections d'images hyperspectrales ont été analysées. La première collection '4f11mb-1' contenait une image infrarouge et trois images de fluorescence acquises dans un faisceau près du bord et contenant trois types cellulaires : parenchyme, sclérenchyme et phloème. La deuxième collection d'images hyperspectrales '4f11mb-2' était également acquise dans un faisceau près du bord. Elle contenait une image infrarouge, quatre images de fluorescence et cinq images Raman contenant trois types cellulaires : sclérenchyme, phloème et xylème.

La collection d'images 4f11mb-1 a été analysée en utilisant l'analyse inter-batterie de Tucker et son extension. Deux composantes ont été retenues et examinées pour ce jeu de données. La deuxième composante a opposé des pixels riches en lignines aux pixels riches en acides hydroxycinnamiques. Les pixels riches en lignines ont été localisés dans les parois des cellules de sclérenchyme par contre les pixels riches en acides hydroxycinnamiques ont été trouvés dans les cellules du phloème avec une plus grande intensité sur les jonctions de cellules et dans les cellules de parenchyme avec une intensité moindre. Pour ce jeu de données, le tableau de fluorescence a contribué à la détermination de la première composante qui a été guidée par l'intensité des spectres de fluorescence puisque les spectres infrarouge n'étaient pas pré-traités de la même manière que les spectres de fluorescence.

Le deuxième collection d'images 4f11mb-2 a été analysée en utilisant l'analyse de co-inertie multiple et son extension. Quatre composantes ont été examinées. La méthode de pré-

traitement tenant compte de l'aspect spectral et spatial des images a pour conséquence d'obtenir une première composante décrivant la somme des intensités en mettant en évidence le contraste paroi/intérieur de cellule. La deuxième composante a opposé les pixels riches en lignines aux pixels riches en polysaccharides. Les polysaccharides ont été localisés principalement dans le sclérenchyme et en quantité plus élevée dans le xylème. Les cellules du sclérenchyme sont particulièrement enrichies en lignine.

En utilisant les outils d'interprétation, nous avons démontré que chacun des tableaux de données a contribué d'une manière non proportionnée pour former les composantes communes. Les tableaux de données de fluorescence et Raman ont contribué davantage à former la première composante commune. L'infrarouge et le Raman ont contribué à la deuxième composante. Pour les autres composantes, l'infrarouge a contribué à plus de 90 %. Les résultats obtenus en analyse de co-inertie trilineaire sont cohérents avec ceux obtenus avec l'analyse de co-inertie multiple. Des différences ont été observées pour les indicateurs statistiques calculées pour évaluer la complémentarité et la spécificité de chaque méthode. Ces différences peuvent être liées soit à la pondération appliquée aux trois tableaux, soit à la différence de résolutions entre les différentes images analysées.

5. Conclusion

Ma thèse est un travail pluridisciplinaire marqué par la présence de l'expérimentation acquisition *in situ* des images hyperspectrales infrarouge, fluorescence UV et Raman, de la chimiométrie et de l'analyse d'image et du signal appliquée à la caractérisation des tissus des végétaux. Ce travail a nécessité une implication au niveau de toutes les étapes du projet depuis l'acquisition des données jusqu'à l'analyse et l'interprétation en passant par l'étape de conditionnement de données. Etant familière avec le domaine du traitement de l'information, j'ai développé au cours de mes travaux de thèse un panel d'outils informatiques et statistiques pour apparier, prétraiter, gérer, intégrer et analyser des images multimodales issues de différentes microspectroscopies. Outre l'enrichissement de mes compétences de base en analyse d'images et des données, j'ai pu également découvrir deux mondes complexes : la spectroscopie et la biologie du végétal.

Références Bibliographiques

- Aber J.D., M.E. Martin, High Spectral Remote Sensing of Canopy Chemistry, Summaries of the Fifth Annual JPL Airborne Earth Science Workshop, 1-4, 1995.
- Adelson E.H., C.H. Anderson, J.R. Bergen, P.J. Burt, J.M. Ogden, Pyramid methods in image processing, *RCA Engineer.* 29 (1984) 33–41.
- Afseth N.K., V.H. Segtnan, J.P. Wold, Raman spectra of biological samples: a study of preprocessing methods, *J. App. Spectrosc.* 60 (2006) 1358–1367.
- Allouche F., M. Hanafi, F. Jamme, P. Robert, F. Guillon, M.F. Devaux, Coupling hyperspectral image data having different spatial resolutions by extending multivariate inter-battery Tucker analysis, *J. Chemometer. Intell. Lab.* 113 (2012) 43-51.
- Andersson W.F., D.E. Akin, Structural and chemical properties of grass lignocelluloses related to conversion for biofuels, *J. Ind Microbial Biotechnol.* 35 (2008) 355-366.
- Arngren M., M.N. Schmidt, J. Larsen, Unmixing of hyperspectral images using bayesian non negative matrix factorization with volume prior, *J. Signal Process. Syst.* 65 (2011) 479-496.
- Barnes R.J., M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra, *App. Spectrosc.* 43 (1989) 772–777.
- Barrière Y., C. Riboulet, V. Méchin, M. Pichon, A.J. Cardinal, T. Lubberstedt, C. Lapiere, Genetics and genomics of lignification in grass cell walls based on maize as a model system, *Genes Genomes Genomics.* 1 (2007) 133-156.
- Barron C., M.L. Parker, E.N.C. Mills, X. Rouau, R.H. Wilson, FTIR imaging of wheat endosperm cell walls in situ reveals compositional and architectural heterogeneity related to grain hardness, *Planta.* 220 (2005) 667–677.
- Barron C., P. Robert, F.Guillon, L. Saulnier, X. Rouau, Structural heterogeneity of wheat arabinoxylans revealed by Raman spectroscopy, *Carbohydr. Res.* 341 (2006) 1186-1191.
- Berlin A., M. Balakshin, N. Gilkes, J. Kadla, V. Maximenko, Inhibition of cellulase, xylanase and β -glucosidase activities by softwood lignin preparations, *J. Biotechnol.* 125 (2006) 198–209.
- Bertrand D., E. Dufour, *La spectroscopie infrarouge et ses applications analytiques*, édition Tec& Doc, collection sciences et techniques agroalimentaires, Paris, 2000.
- Bessombes S., K. Mazeau, The cellulose/lignin assembly assessed by molecular modeling. Part 2: Seeking for evidence of organization of lignin molecules at the interface with cellulose, *Plant Phys. Biochem.* 43 (2005) 277-286.

- Bidlack J.E., M. Malone, R. Benson, Molecular structure and component integration of secondary cell walls in plants, *Proc. Okla. Acad. Sci.* 72 (1992) 51-56.
- Bindlish R., A.P. Barros, Parameterization of vegetation backscatter in radar-based, soil moisture estimation, *Remote Sens. Environ.* 76 (2001) 1-130.
- Blanco M., A. Peguero, An expeditious method for determining particle size distribution by near infrared spectroscopy: Comparison of PLS2 and ANN models, *Talanta* 77 (2008) 647-651.
- Bledzki A.K., J. Gassan, Composites reinforced with cellulose based fibres, *Prog. Polym. Sci.* 24 (1999) 221-274.
- Böcker U., R. Ofstad, Z. Wu, H.C. Bertram, G.D. Sockalingum, B. Egelanddal, A. Kohler, Revealing co-variance structures in FTIR and Raman microspectroscopy spectra - A study on pork muscle fiber tissue subjected to different processing parameters. *J. App. Spectrosc.* 61 (2007) 1032-1039.
- Boerjan W., J. Ralph, M. Baucher, Lignin biosynthesis, *Ann. Rev. Plant. Biol.* 54 (2003) 519-546.
- Bonmassar G., E.L. Schwartz, Representation is Space-Variant, *Behav. Brain Sci.* 21 (1998) 469-470.
- Buranov A.U., G. Mazza, Lignin in straw of herbaceous crop, *Ind. Crop. Prod.* 28 (2008) 237-259.
- Burger J., P. Geladi, Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples, *The Analyst.* 131 (2006) 1152-1160.
- Burger W., M. Burge, *Digital Image Processing*, 1ed., Springer, New York, 2008.
- Burnham A.J., R. Viveros, J.F. Macgregor, Framework for latent variable multivariate regression, *J. Chemometer.* 10 (1996) 31-45.
- Burt P.J., E.H. Adelson, The Laplacian Pyramid as a compact Image Code (1983). *IEEE Trans. Communic.* 31 (1983) 532-540.
- Cael J., J.L. Koenig, J. Blackwell, Infrared and Raman spectroscopy of carbohydrates. *Carbohydr. Res.* 29 (1973) 123-134.
- Carroll J.D., A generalization of canonical correlation analysis to three or more sets of variables, 76th annual convention of the American psychological association, 227-228, 1968.
- Carpita N.C., D.M. Gibeaut, Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth, *Plant J.* 3(1993) 1-30.
- Chessel D., M. Hanafi, Analyses de la co-inertie de K nuages de points, *Rev. Stat. Appl.* 44 (1996) 35-60.

- Chevallier S., D. Bertrand, A. Kohler, Ph. Courcoux, Application of PLS-DA in multivariate image analysis, *J. Chemometr.* 20 (2006) 221–229.
- Clark R.N., G.A. Swayze, A. Gallagher, Mapping the mineralogy and lithology of Canyonlands, Utah with imaging spectrometer data and the multiple spectral feature mapping algorithm, *Summaries of the Third Annual JPL Airborne Geoscience Workshop*, 11-13,1992.
- Coimbra M. A., A. Barros, D.N. Rutledge, I. Delgadillo, FTIR spectroscopy as a tool for the analysis of olive pulp cell-wall polysaccharide extracts, *Carbohydr. Res.* 317 (1999) 145-154.
- Crowell E.F., M. Gonneau, S. Vernhettes, H. Höfte, Regulation of anisotropic cell expansion in plants, *Comptes Rendus Biologies.* 333 (2010) 320-324.
- Dani P., S. Chaudhuri, Automated assembling of images: Image montage preparation, *Pattern Recogn.* 28 (1995) 431–445.
- De Juan A., M. Maeder, T. Hancewicz, R. Tauler, Local rank analysis for exploratory spectroscopic image analysis. Fixed size image window-evolving factor analysis, *J. Chemometer. Intell. Lab.* 77 (2005) 64–74.
- De Juan A., M. Maeder, T. Hancewicz, L. Duponchel, R. Tauler, *Chemometric tools for image analysis*, R. Salzer, H.W. Siesler editors, *Infrared and Raman Spectroscopic Imaging*, Wiley-VCH, 2009.
- Delwiche S.R., R.A. Graybosch, Examination of spectral pretreatments for partial least-squares calibrations for chemical and physical properties of wheat, *J. App. Spectrosc.* 57 (2003) 1517–1527.
- Devaux M.F., P. Robert, M. Safar, E. Vigneau, Canonical correlation analysis of mid and near-infrared oil spectra, *J. Appl. Spectrosc.* 47 (1993) 1024–1029.
- Devaux M.F., I. Taralova, J. Levy-Vehel, E. Bonnin, J.F. Thibault, F. Guillon, Contribution of image analysis to the description of enzymatic degradation kinetics for particulate material, *J. Food Eng.* 77 (2006) 1096-1107.
- Dufour E., M.F. Devaux, P. Fortier, S. Herbert, Delineation of the structure of soft cheeses at the molecular level by fluorescence spectroscopy-relationship with texture, *Int. Dairy J.* 11 (2001) 465-473.
- Dumas P., F. Polack, B. Lagarde, O. Chubar, J.L. Giorgetta, S. Lefrançois, Synchrotron Infrared Microscopy at the French Synchrotron facility SOLEIL, *J. Infrared Phys. Techn.* 49 (2006) 152-160.
- Dumas P., L. Miller, The use of synchrotron infrared microspectroscopy in biological and biomedical investigations, *Vib. Spectrosc.* 32 (2003) 3–21.
- Esau K., *Anatomy of seed plants*. 2nd ed, Wiley, New York, 1977.
- Evert R.F., *Esau's Plant anatomy : meristems, cells, and tissues of the plant body : their structure, function and development*, third ed Wiley-Interscience, New York, 2006.

- Geladi P., D. MacDougall, H. Martens, Linerization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Appl. Spectrosc.* 39 (1985) 491-500.
- Geladi P., H. Grahn, *Multivariate image analysis*, 1ed., Wiley, Chichester, 1996.
- Geladi P., J. Burger, T. Lestander, Hyperspectral imaging: calibration problems and solutions, *J. Chemometer Intell Lab.* 72 (2004) 209-217.
- Geladi P., K. Esbensen, Regression on multivariate images: principal component regression for modeling, prediction and visual diagnostic tools, *J. Chemometr.* 5 (1991) 97-111.
- Geladi P., S. Wold, K.H. Esbensen, Image analysis and chemical information in images, *Anal. Chim. Acta.* 191 (1986) 473-480.
- Gierlinger N., L. Goswami, M. Schmidt, I. Burgert, C. Coutand, T. Rogge, M. Schwanninger, In situ FT-IR microscopic study on enzymatic treatment of Poplar wood cross-sections, *J. Biomacromolecules.* 9 (2008-a) 2194-2201.
- Gierlinger N., L. Sapei, O. Paris, Insights into the chemical composition of Equisetum hyemale by high resolution Raman imaging, *Planta.* 227 (2008-b) 969-980.
- Gierlinger N., M. Schwanninger, Chemical imaging of poplar wood cell walls by confocal Raman microscopy, *Plant Physio.* 140 (2006) 1246-1254.
- Gierlinger N., T. Keplinger, M. Harrington, Imaging of plant cell walls by confocal Raman microscopy, *Nat. Protoc.* 7 (2012) 1694 -1708.
- Giuliani A., F. Jamme, V. Rouam, F. Wien, J.L. Giorgetta, B. Lagarde, O. Chubar, S. Bac, I. Yao, S. Rey, C. Herbeaux, J.L. Marlats, D. Zerbib, F. Polack, M. Réfregiers, DISCO: a low-energy multipurpose beamline at synchrotron SOLEIL, *J. Synchrotron Radiat.* 16 (2009) 835-841.
- Gleason T.C., On redundancy in canonical analysis, *Physiological bulletin.* 83 (1976) 1004-1006.
- Gledhill D., G. Yun Tian, D. Taylor, D. Clarke, Panoramic imaging: a review, *Comput. Graph.* 27 (2003) 435-445.
- Gonzalez R.C., R.E. Woods, S.L. Eddins, *Digital Image Processing using Matlab*, 2 ed., Gatesmark Publishing, New York, 2009.
- Gorzsás A, H. Stenlund, P. Persson, J. Trygg, B. Sundberg, Cell-specific chemotyping and multivariate imaging by combined FT-IR microspectroscopy and orthogonal projections to latent structures (OPLS) analysis reveals the chemical landscape of secondary xylem, *Plant J.* 66 (2011) 903-914.
- Grabber J.H., J. Ralph, C. Lapierre, Y. Barrière, Genetic and molecular basis of grass cell-wall degradability. I. Lignin-cell wall matrix interactions, *R. Biologies.* 327 (2004) 455-465.
- Grabber J.H., J. Ralph, R.D. Hatfield, S.P. Quideau, *p*-Hydroxyphenyl, Guaiacyl, and Syringyl lignins have similar inhibitory effects on wall degradability, *J. Agric. Food Chem.* 45 (1997) 2530-2532.

- Grabber J.H., R.D. Hatfield, F. Lu, J. Ralph, Coniferyl ferulate incorporation into lignin enhances the alkaline delignification and enzymatic degradation of cell walls. *Biomacromolecules*. 9 (2008) 2510–2516.
- Grohamann K., D.J. Mitchell, M.E. Himmel, B.E. Dale, H.A. Schroeder, The role of ester groups in resistance of plant cell wall polysaccharides to enzymatic hydrolysis, *Appl. Biochem. Biotechnol.* 20/21 (1989) 45-61.
- Guilhaumou N., P. Dumas, G.L. Carr, G.P. Williams, Synchrotron infrared microspectrometry applied to petrography in micrometer-scale range: Fluid chemical analysis and mapping, *J. Appl. Spectrosc.* 52 (1998) 1029-1034.
- Guo G.L., D.C. Hsu, W.H. Chen, W.S. Hwang, Characterization of enzymatic saccharification for acid-pretreated lignocellulosic materials with different lignin composition, *Enzyme Microb. Technol.* 45 (2009) 80-87.
- Hanafi M., A. Kohler, E. Quannari, Connections between multiple co-inertia analysis and consensus principal component analysis, *J. Chemometer Intell Lab.* 106 (2011) 37-40.
- Hanafi M., G. Mazerolles, E. Dufour, E.M. Qannari, Common components and specific weight analysis and multiple co-inertia analysis applied to the coupling of several measurement techniques, *J. Chemometer*. 20 (2006) 1-12.
- Hanafi M., H.A.L. Kiers, Analysis of K sets of data with differential emphasis on agreement between and within sets, *Comput. Stat. Data An.* 51 (2006) 1491-1508.
- Hansen M.A.T., J.B. Kristensen, C. Felby, H. Jorgensen, Pre-treatment and enzymatic hydrolysis of wheat straw (*Triticum aestivum* L.) – The impact of lignin relocation and plant tissues on enzymatic accessibility, *J. Bioresource Technol.* 102 (2011) 2804–2811.
- Harris P.J., R.D. Hartley, Detection of bound ferulic acid in the cell walls of the Gramineae by ultraviolet fluorescence microscopy. *Nature*. 259 (1976) 508–510.
- Hatfield R.D., J.M. Marita, Enzymatic processes involved in the incorporation of hydroxycinnamates into grass cell walls, *Phytochem. Rev.* 9 (2010) 35-45.
- Hatfield R.D., J. Ralph, J.H. Grabber, Cell wall cross-linking by ferulates and diferulates in grasses, *J. Sci. Food Agric.* 79 (1999) 403–407.
- He L., N. Terashima, Formation and structure of Lignin in monocotyledons IV. Deposition process and structural diversity of the lignin in the cell wall of sugar cane and rice plant Studied by ultraviolet microscopic spectroscopy, *Holzforschung*. 45 (1991) 191-198.
- Hege E.K., D. O'Connell, W. Johnson, S. Basty, E.L. Dereniak, Hyperspectral imaging for astronomy and space surveillance, S.S. Shen, P.E. Lewis editors, *Imaging Spectrometry IX*, 2003.
- Hématy K., H. Höfte, Cellulose and cell elongation, in the Expanding cell, J.P. Verbelen, K. Vissenberg editors, *Plant Cell. Monographs*, Springer –Verlag, 2007.

- Heraud P., B.R. Wood, J. Beardall, D. McNaughton, Effects of pre-processing of Raman spectra on in vivo classification of nutrient status of microalgal cells, *J. Chemometr.* 20 (2006) 193–197.
- Himmelsbach D.S., D.E. Akin, Near-Infrared Fourier-Transform Raman Spectroscopy of Flax (*Linum usitatissimum* L.) Stems. *J. Agric. Food Chem.* 46 (1998) 991-998.
- Himmelsbach D.S., S. Khalili, D.E. Akin, The use of FT-IR microspectroscopic mapping to study the effects of enzymatic retting of flax (*Linum usitatissimum* L) stems , *J. Sci. Food Agric.* 82 (2002) 685-696.
- Holbrook R.D., J.H. Yen, T.J. Grizzard, Characterizing natural organic material from the Occoquan Watershed (Northern Virginia, US) using fluorescence spectroscopy and PARAFAC, *Sci. Total Environ.* 361 (2006) 249-266.
- Hotelling H., Relations between two sets of variates, *Biometrika.* 28 (1936) 321–377.
- Jamme F., P. Robert, B. Bouchet, L. Saulnier, P. Dumas, F. Guillon, Aleurone cell walls of wheat grain: high spatial resolution investigation using synchrotron infrared microspectroscopy, *J. Appl. Spectrosc.* 62 (2008) 895-900.
- Jamme F., S. Villette, A. Giuliani, V. Rouam, F. Wien, B. Lagarde, M. Réfrégiers, Synchrotron UV fluorescence microscopy uncovers new probes in cells and tissues, *Microsc. Microanal.* 16 (2010) 507–514.
- Jarvis M.C., M.C. MacCann, Macromolecular biophysics of the plant cell wall: concepts and methodology, *Plant Physio. Biochem.* 38 (2000) 1–13.
- Jolliffe I.T., *Principal Component Analysis*, Springer-Verlag, New York, 1986.
- Jaumot J., R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, *J. Chemometer Intell Lab.* 76 (2005) 101–110.
- Joseleau J.P., K. Ruel, Study of lignification by non invasive techniques in growing maize internodes—an investigation by Fourier transform infrared, cross-polarisation-magic angle spinning ¹³C-nuclear magnetic resonance spectroscopy, *Plant Physiol.* 114 (1997) 1123–1133.
- Jung H.G., M.D. Casler, Maize stem tissues: cell wall concentration and composition during development, *Crop Sci.* 46 (2006-a) 1793-1800.
- Jung H.G., M.D. Casler, Maize stem tissues: impact of development on cell wall degradability, *Crop Sci.* 46 (2006-b) 1801-1809.
- Kacuráková M., A.C. Smith, M.J. Gidley, R.H. Wilson, Molecular interactions in bacterial cellulose composites studied by 1D FT-IR and dynamic 2D FT-IR spectroscopy, *Carbohydr. Res.* 337 (2002) 1145-1153.
- Kacuráková M., N. Wellner, A. Ebringerova, Z. Hromadkova, R.H. Wilson, P.S. Belton, Characterization of xylan-type polysaccharides and associated cell wall components by FT-IR and FT-Raman spectroscopies, *Food Hydrocolloid.* 13 (1999) 35-41.

- Kim J.S., K.H. Lee, C.H. Cho, G. Koch, Y.S. Kim, Micromorphological characteristics and lignin distribution in bamboo (*Phyllostachys pubescens*) degraded by the white rot fungus *Lentinus edodes*, *Holzforschung*. 62 (2008) 481-487.
- Koch G., G. Kleist, Application of scanning UV microspectrophotometry to localise lignins and phenolic extractives in plant cell walls, *Holzforschung*. 55 (2001) 1-5.
- Koch G., Topochemical characterization of lignins and phenolic extractives in wood cell walls, *Lenzinger Berichte*. 83 (2004) 6-12.
- Kohler A., D. Bertrand, H. Martens, K. Hannesson, C. Kirschner, R. Ofstad, MIA of a set of FTIR microspectroscopy images of aged bovine muscle tissue combining image and design information, *Anal. Bioanal. Chem.* 389 (2007) 1143-1153.
- Lakhal L., V. Acha, T. Aussenac, PARAFAC analysis of front-face fluorescence data: absorption and scattering effects assessed by means of Monte Carlo simulations, *Chemometer. Intell. Lab.* 116 (2012) 112-122.
- Lasch P., Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging, *J. Chemometer. Intell. Lab.* in press.
- Lasch P., J. Schmitt, M. Beekes, T. Udelhoven, M. Eiden, H. Fabian, W. Petrich, D. Naumann, Antemortem identification of bovine spongiform encephalopathy from serum using infrared spectroscopy, *Anal. Chem.* 75 (2003) 6673-6678.
- Lebart L., A. Morineau, M. Piron, *Statistique exploratoire multidimensionnelle*, 3 ed., Dunod, Paris, 2000.
- Liu J., J.L. Koenig, A new baseline correction algorithm using objective criteria, *J. Appl. Spectrosc.* 41 (1987) 447-449.
- Liu Y., Y. Zhang, L. Guo, Towards realistic assessment of cultivated land quality in an ecologically fragile environment: A satellite imagery-based approach, *Appl. Geogr.* 30 (2010) 271-281.
- Lu X., S. Zhang, H. Su, Y. Chen, Mutual information-based multimodal image registration using a novel joint histogram estimation, *Comput Med Imag Grap.* 32 (2008) 202-209.
- Lundquist K., B. Josefsson, G. Nyquist, Analysis of lignin products by fluorescence spectroscopy, *Holzforschung*. 32 (1978) 27-32.
- Lybeer B., G. Koch, Lignin distribution in the tropical bamboo species *Gigantochloa levis*, *Iawa. J.* 26 (2005) 443-456.
- Lynd L.R., C.E. Wyman, T.U. Gerngross, Biocommodity engineering, *Biotechnol. Prog.* 15 (1999) 777-793.
- Machinet G.E., I. Bertrand, Y. Barrière, B. Chabbert, S. Recous, Impact of plant cell wall network on biodegradation in soil: role of lignin composition and phenolic acids in roots from 16 maize genotypes, *Soil Biol. Biochem.* 43 (2011) 1544-1552.
- Maintz J.A., Viergever M.A., A review of medical image registration, *Med. Image Anal.* 2 (1998) 1-36.

- Marchand P., L. Marmet, Binomial smoothing filter: A way to avoid some pitfalls of least square polynomial smoothing, *Rev. Sci. Instrum.* 54 (1983) 1034-1041.
- Marga F., A. Gallo, K.H. Hasenstein, Cell wall components affects mechanical properties: Studies with thistle flowers, *Plant Physiol. Biochem.* 41 (2003) 792-797.
- Martens H., J.P. Nielsen, S.B. Engelsen, Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures, *Anal. Chem.* 75 (2003) 394-404.
- Mazerolles G., M.F. Devaux, E. Dufour, E.M. Qannari, Ph. Courcoux, Chemometric methods for the coupling of spectroscopic techniques and for the extraction of the relevant information contained in the spectral data tables, *J. Chemometer. Intell. Lab.* 63 (2002) 57-68.
- McCann M.C., L. Chen, K. Roberts, E.K. Kemsley, C. Sene, N.C. Carpita, N.J. Stacey, R.H. Wilson, Infrared microspectroscopy: Sampling heterogeneity in plant cell wall composition and architecture, *Physio. Planta.* 100 (1997) 729-738.
- McCann M.C., M. Bush, D. Milioni, P. Sado, N.J. Stacey, G. Catchpole, M. Defernez, N.C. Carpita, H. Hofte, P. Ulvskov, R.H. Wilson, K. Roberts, Approaches to understanding the functional architecture of the plant cell wall, *Phytochem.* 57 (2001) 811-812.
- Méchin V., O. Argillier, Y. Barrière, V. Menanteau, Genetic variation in stems of normal and brown-midrib 3 maize inbred lines, Towards similarity for in vitro digestibility and cell wall composition, *Maydica.* 43 (1998) 205-210.
- Méchin V., O. Argillier, O. Ménanteau, Y. Barrière, I. Mila, B. Pollet, C. Lapiere, Relationship of cell wall composition to in vitro cell wall digestibility of maize inbred line stems, *J. Sci Food Agric.* 80 (2000) 574-580.
- Miller L.M., C.S. Carlson, G.L. Carr, M.R. Chance, A method for examining the chemical basis for bone disease: synchrotron infrared microspectroscopy, *Cell. Mol. Biol.* 44 (1998) 117-27.
- Mohebbi B., Application of ATR infrared spectroscopy in wood acetylation. *J. Agric. Sci. Technol.* 10 (2008) 253-259.
- Morrison W., D. Archibald, Analysis of graded flax fiber and yarn by pyrolysis mass spectrometry and pyrolysis gas chromatography mass spectrometry, *J. Agric. Food Chem.* 46 (1998) 1870-1876.
- Mouille G., S. Robin, M. Lecomte, S. Pagant, H. Höfte, Classification and identification of arabidopsis cell wall mutants using Fourier transform infrared (FT-IR) microspectroscopy, *Plant J.* 35 (2003) 393-404.
- Moussaoui S., H. Hauksdóttir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Douté, J.A. Benediksson, On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation, *Neurocomput.* 71 (2008) 2194-2208.
- Naik S.N., V.V. Goud, P.K. Rout, A.K. Dalai, Production of first and second generation biofuels: A comprehensive review, *J. Renew Sust Energ Rev.* 14 (2010) 578-597.

- Neemuchwala H., A. Hero, P. Carson, Image matching using alpha-entropy measures and entropic graphs, *Signal Process.* 85, 2003.
- Octave S.T., Biorefinery: Toward an industrial metabolism, *J. Biochimie.* 91 (2009) 659–664.
- Olmstead J.A., D.G. Gray, Fluorescence emission from mechanical pulp sheets, *J. Photochem. Photobiol.* 23 (1993) 59-65.
- Olmstead J.A., D.G. Gray, Fluorescence spectroscopy of cellulose, lignin and mechanical pulps : *J. Pulp Paper Sci.* 23 (1997) 1571-1581.
- Pauly M., K. Keegstra, Cell-wall carbohydrates and their modification as a resource for biofuels, *Plant J.* 54 (2008) 559–568.
- Pedersen F., Interactive explorative analysis of multivariate images using principal components, 1994.
- Pedro A.M.K., M.M.C. Ferreira, Simultaneously calibrating solids, sugars and acidity of tomato products using PLS2 and NIR spectroscopy, *Anal. Chim. Acta.* 595 (2007) 221–227.
- Perperidis D., R.H. Mohiaddin, D. Rueckert, Spatio-temporal free-form registration of cardiac MR image sequences, *Med. Image Anal.* 9 (2005) 441–456.
- Petrou M., Image Registration: An Overview, *Adv. Imag. Elect. Phys.* 130 (2004) 243–291.
- Piló-Veloso D., E.A. Nascimento, S.A.L. Morais, *Quim. Nova.* 16 (1993) 435.
- Piqueras S., L. Duponchel, R. Tauler, A. de Juan, Resolution and segmentation of hyperspectral biomedical images by multivariate Curve Resolution-Alternating Least Squares, *Anal. Chim. Acta.* 705 (2011) 182-192.
- Piot O., J.C. Autran, M. Manfait, Investigation by confocal Raman microspectroscopy of the molecular factors responsible for grain cohesion in the *Triticum aestivum* bread wheat, *J. cereal Sci.* 34 (2001) 191-205.
- Philippe S., C. Barron, P. Robert, M.F. Devaux, L. Saulnier, F. Guillon, Characterization using Raman microspectroscopy of arabinoxylans in the walls of different cell types during the development of wheat endosperm. *J Agr. Food Chem.* 54 (2006-a) 5113–5119.
- Philippe S., P. Robert, C. Barron, M.F. Devaux, L. Saulnier, F. Guillon, Deposition of Cell Wall Polysaccharides in Wheat Endosperm during Grain Development: Fourier Transform-Infrared Microspectroscopy Study, *J. Agric. Food Chem.* 54 (2006-b) 2303-2308.
- Popescu C.M., C. Vasile, M.C. Popescu, G.H. Singurel, V.I. Popa, B.S. Munteanu, Analytical methods for lignin characterization. II. Spectroscopic studies, *Cell Chem. Technol.* 40 (2006) 597-621.

- Prats-Montalbán J.M., A. Ferrer, R. Bro, T. Hancewicz, Prediction of skin quality properties by different multivariate image analysis methodologies, *J. Chemometer. Intell. Lab.* 96 (2009) 6–13.
- Prosky L., N.G. Asp, I. Furda, J.W. DeVries, T.F. Schweizer, B.F. Harland, Determination of total dietary fiber in foods, food products, and total diets: Interlaboratory study. *Journal of the Association of Official Analytical Chemists.* 67 (1984) 1044–1052.
- Qannari E.M., I. Wakeling, Ph. Coucoux, M.F. Macfie, Defining the underlying sensory dimensions, *Food Qual. Prefer.* 11 (2000) 151-154.
- Qiao J., M.O. Ngadi, N. Wang, C. Gariépy, S.O. Prasher, Pork quality and marbling level assessment using a hyperspectral imaging system, *J. Food Eng.* 83 (2007) 10–16.
- Radotic K., A. Kalauzi, D. Djikanovic, M. Jeremic, R.M. Leblanc, Z.G. Cerovic, Component analysis of the fluorescence spectra of a lignin model compound, *J. Photochem. Photobiol.* 83 (2006) 1–10.
- Ralph J., K. Lundquist, G. Brunow, F. Lu, H. Kim, P.F. Schatz, J.M. Marita, R.D. Hatfield, S.A. Ralph, J.H. Christensen, W. Boerjan, Lignins: natural polymers from oxidative coupling of 4- hydroxyphenyl-propanoids, *Phytoch. Rev.* 3 (2004) 29-60.
- Ram M.S., F.E. Dowell, L.M. Seitz, FT-Raman of unsoaked and NaOH-soaked wheat kernels, bran, and ferulic acid, *Cereal Chem.* 80 (2003) 188-192.
- Rehbein M., M. Pereira, G. Koch, O. Kordsachia, Topochemical investigation on the delignification of Eucalyptus globulus chips during neutral sulfite (NS) pulping, *Wood Sci. Technol.* 44 (2010) 435-449.
- Rémond-Zilliox C., P. Debeire, D. reis, B. Vian, Immunolocalization of a purified xylanase during hydrolysis of wheat straw stem, *Int. J. Plant Sci.* 158 (1997) 769-777.
- Riboulet C., D. Denoue, B. Lefevre, Y. Barriere, Genetic variation in maize cell wall for lignin content, lignin composition, p-hydroxycinnamic acid content, and digestibility in set of 19 lines at silage harvest maturity, *Maydica.* 53 (2008) 11–19.
- Richards J.A., X. Jia, *Remote Sensing Digital Image Analysis, an Introduction.* 3 ed., Springer-Verlag, Berlin, 1999.
- Robert P., F. Jamme, C. Barron, B. Bouchet, L. Saulnier, P. Dumas, F. Guillon, Change in wall composition of transfer and aleurone cells during wheat grain development, *Planta.* 406 (2011) 233-393.
- Robert P., M.F. Devaux, D. Bertrand, Beyond prediction: Extracting relevant information from near infrared spectra, *J. Near Infrared Spec.* 4 (1996) 75-84.
- Robert P., M. Marquis, C. Barron, F. Guillon, L.Saulnier, Ft-ir investigation of cell wall polysaccharides from cereal grains, Arabinoxylan infrared assignment, *J. Agr. Food. Chem.* 53 (2005) 7014-7018.
- Rudall P.J., L.R. Caddick, Investigation of the presence of phenolic compounds in monocotyledonous cell walls, using UV Fluorescence Microscopy, *Ann. Bot.* 74 (1994) 483-491.

- Rutledge D.N., D. Jouan-Rimbaud Bouveresse, Multi-way analysis of outer product arrays using PARAFAC, *J. Chemometer. Intell. Lab.* 85 (2007) 170-178.
- Ruzena B.,K. Stane, Multiresolution elastic matching, *Comput. Vision Graph. Image Process.* 46 (1989) 1-21.
- Saadi A., I. Lempereur, S. Sharonov, J.C. Autran, M. Manfait, Spatial distribution of phenolic materials in durum wheat grain as probed by confocal fluorescence spectral imaging, *J. Cereal Sci.* 28 (1998) 107-114.
- Saporta G., *Probabilités, Analyse des données et Statistique*, Ed. Technip, 1990.
- Salzer R., H.W. Siesler (Eds.), *Infrared and Raman spectroscopic imaging*, Wiley-VCH, Weinheim (Germany), 2009.
- Saulnier L., P. Robert, M. Grintchenko, F. Jamme, B. Bouchet, F. Guillon, Wheat endosperm cell walls: Spatial heterogeneity of polysaccharide structure and composition using micro-scale enzymatic fingerprinting and FT-IR microspectroscopy, *J. Cereal Sci.* 50 (2009) 312-317.
- Savitzky A., M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* 36 (1964) 1627-1639.
- Schubert C., Can biofuels finally take center stage? *Nat. Biotechnol.* 24 (2006) 777-784.
- Serra J., *Image Analysis and Mathematical Morphology*, Ac. Press, London, 1982.
- Siegel S.M., N.J. Puerner, Role of gravitational stress in land plant evolution: The gravitational factor in lignification [semi-annual report] [Honolulu]: University of Hawaii, 1969.
- Siqueira G., A. MF Milagres, W. Carvalho, G. Koch, A. Ferraz, Topochemical distribution of lignin and hydroxycinnamic acids in sugar-cane cell walls and its correlation with the enzymatic hydrolysis of polysaccharides, *Biotechnology for Biofuels.* 1 (2011) 4-7.
- Scheller H.V., P. Ulvskov, Hemicelluloses, *Annu. Rev. Plant Biol.* 61 (2010) 263-89.
- Smilde A., B. Ramsus, P. Geladi, *Multi-way Analysis*, 1 ed.,Wiley, Chichester, 2004.
- Smilde A.K., J.A. Westerhuis, S.D. Jong, A framework for sequential multiblock component methods, *J. Chemometer.* 17 (2003) 323-337.
- Soille P., *Morphological Image Analysis*, Springer-Verlag, Berlin-Heidelberg, 1999.
- Stewart D., Fourier Transform infrared microscopy of plant tissues, *App. Spectrosc.* 50 (1996) 357-365.
- Sun R.C., J. Tomkinson, W. Zhu, S.Q. Wang, Delignification of maize stems by peroxymonosulfuric acid, peroxyformic acid, peracetic acid, and hydrogen peroxide. Physicochemical and structural characterization of the solubilized lignins, *J. Agric. Food Chem.* 48 (2000) 1253-1262.

- Taboada A., E. Novo-Uzal, G. Flores, M. Loureda, A.R. Barcelo, A. Masa, F. Pomar, Digestibility of silages in relation to their hydroxycinnamic acid content and lignin composition, *J. Sci. Food Agric.* 90 (2009) 1155-1162.
- Taiez L., E. Zeiger, *Plant physiology*, third ed Sinauer Associates, Sunderland, MA, USA 2002.
- Tenenhaus M., *La régression PLS. Théorie et pratique*, Technip, Paris, 1998.
- Tenenhaus M., *L'approche PLS. Rev. Stat. Appl.* 4 (1999) 75-40.
- Tomazevic D., B. Likar, F. Pernus, Comparative evaluation of retrospective shading correction methods, *J. Microsc. Oxford.* 208 (2002) 212-223.
- Toole G.A., G. Le Gall, I. J. Colquhoun, C. Nemeth, L. Saulnier, A. Lovegrove, T. Pellny, M.D. Wilkinson, J. Freeman, R.A.C. Mitchell, E.N.C. Mills, P.R. Shewry, Temporal and spatial changes in cell wall composition in developing grains of wheat cv. Hereward, *Planta.* 232 (2010) 677-689.
- Toole G.A., R.H. Wilson, M.L. Parker, N.K. Wellner, T.R. Wheeler, P.R. Shewry, E.N.C. Mills, The effect of environment on endosperm cell-wall development in *Triticum aestivum* during grain Wlling: an infrared spectroscopic imaging study, *Planta.* 225 (2007) 1393-1403.
- Toole G.A., G. Le Gall, I.J. Colquhoun, Ph. Johnson, Z. Bedö, L. Saulnier, P.R. Shewry, E.N.C. Mills, Spectroscopic Analysis of Diversity of Arabinoxylan Structures in Endosperm Cell Walls of Wheat Cultivars (*Triticum aestivum*) in the HEALTHGRAIN Diversity Collection, *J. Agric. Food Chem.* 59 (2011) 7075-7082.
- Tovar-Gomez M.R., J.C. Emile, B. Michalet-Doreau, Y. Barriere, In situ degradation kinetics of maize hybrid stalks, *Anim. Feed. Sci. Tech.* 68 (1997) 77-88.
- Tucker L.R., An inter-battery method of factor analysis, *J. Psychometrica.* 23 (1958) 11-136.
- Umesh P.A., S.A. Ralph, FT-Raman spectroscopy of wood: Identifying contributions of lignin and carbohydrate polymers in the spectrum of black spruce (*Picea mariana*). *App. spectrosc.* 51 (1997) 11.
- Vermerris W., K.J. Thompson, L.M. McIntyre, The maize brown midrib1 locus affects cell wall composition and plant development in a dose dependent manner, *Heredity.* 88 (2002) 450-457.
- Vogel J., Unique aspects of the grass cell wall, *Curr. Opin. Plant. Biol.*, 11 (2008) 301-307.
- Vovk U., F. Pernus, B. Likar, A review of methods for correction of intensity inhomogeneity in MRI, *IEEE Trans. Med. Imaging.* 26 (2007) 405-421.
- Westerhuis J.A., T. Kourti, J.F. MaxGregor, Analysis of multiblock and hierarchical PCA and PLS models, *J. Chemometr.* 12 (1998) 463-482.
- White B.A., R.I. Mackie, K.C. Doerner, Enzymatic hydrolysis of forage cell walls, J. Jung, D.R. Buxton, R.D. Hatfield, J. Ralph (editors), *Forage wall structure and digestibility*, 445-484, 1993.

- Widjaja E., Zheng W., Huang Z., Classification of colonic tissues using near-infrared Raman spectroscopy and support vector machines, *Int. J. Onco.* 32 (2008) 653-662.
- Williams P., P. Geladi, G. Fox, M. Manley, Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis, *Anal. Chim. Acta.* 653 (2009) 121-130.
- Wold H., P. Geladi, K. Esbensen, J. Ohman, Multi-way principal components and PLS analysis, *J. chemometr.* 1 (1987) 41-56.
- Wold S., N. Kettaneh, K. Tjessem, Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection, *J. chemometer.* 10 (1996) 463-482.
- Wold S., M. Sjöström, L. Eriksson, PLS-regression: A basic tool of chemometrics, *J. Chemom. Intell. Lab. Sys.* 58 (2001) 109-130.
- Xiao B., X.F. Sun, R.C. Sun, Chemical, structural and thermal characterization of alkali soluble lignins and hemicelluloses and cellulose from maize stems, rye straw and rice straw. *Polym. Degrad. Stabil.* 74 (2001) 307-319.
- Yan Z., S. Chen, H. Wang, B. Wang, J. Jiang, Biosynthesis of bacterial cellulose/multi-walled carbon nanotubes in agitated culture, *Carbohydr. Polym.* 74 (2008) 659-665.
- Zhang Y., L. Pitkanen, J. Douglade, M. Tenkanen, C. Remond, C. Joly, Wheat bran arabinoxylans, Chemical structure and film properties of three isolated fractions, *Carbohydr. Polym.* 86 (2011-a) 852-859.
- Zhang Y., T. Culhaoglu, B. Pollet, C. Melin, D. Denoue, Y. Barriere, S. Baumberger, V. Mechin, Impact of lignin structure and cell wall reticulation on maize cell wall degradability, *J. Agr. Food Chem.* 59 (2011-b) 10129-10135.
- Zhu Y.M., Mutual information-based registration of temporal and stereo retinal images using constrained optimization, *Comput Meth Prog Bio.* 86 (2007) 210-215.
- Zitova B., J. Flusser, Image registration methods a survey, *J. Image vision comput.* 21 (2003) 977-1000.

Annexe A

Cette annexe montre la propriété de monotonie de l'algorithme analyse inter-batterie de Tucker à trois entrées. La convergence de l'algorithme est donc prouvée.

Considérons $(\mathbf{c}_X^{(h)}, \mathbf{c}_{Y(z)}^{(h)})$, $(h=0, 1, 2, \dots)$ comme une séquence générée par l'algorithme,

Cette annexe montre l'inégalité ci-dessous:

$$\sum_{z=1}^Z \mathbf{cov}^2(\mathbf{c}_X^{(h)}, \mathbf{c}_{Y(z)}^{(h)}) \leq \sum_{z=1}^Z \mathbf{cov}^2(\mathbf{c}_X^{(h+1)}, \mathbf{c}_{Y(z)}^{(h+1)}) \quad \text{pour } h = 0, 1, 2, 3, \dots$$

Considérons le problème de maximisation suivant :

$$\text{Maximiser } \sum_{z=1}^Z \alpha_{(z)}^{(h)} \mathbf{cov}(\mathbf{c}_X, \mathbf{c}_{Y(z)}) \quad (\text{avec } \|\mathbf{u}\|=1 \text{ et } \|\mathbf{v}\|=1) \quad (1)$$

$$\text{où } \alpha_{(z)}^{(h)} = \mathbf{cov}(\mathbf{c}_X^{(h)}, \mathbf{c}_{Y(z)}^{(h)})$$

A partir de l'inégalité (1)

$$\sum_{z=1}^Z \alpha_{(z)}^{(h)} \mathbf{cov}(\mathbf{c}_X^{(h)}, \mathbf{c}_{Y(z)}^{(h)}) = \mathbf{u}' \mathbf{X}' (\sum_{z=1}^Z \alpha_{(z)}^{(h)} \mathbf{Y}_{(z)}) \mathbf{v}$$

Il résulte que la solution du problème (1) sont $\mathbf{u}^{(h+1)}$ et $\mathbf{v}^{(h+1)}$ décrit dans l'étape 6 de l'algorithme (partie D section I.1.2).

Par conséquent :

$$\sum_{z=1}^Z \mathbf{cov}^2(\mathbf{c}_X^{(h)}, \mathbf{c}_{Y(z)}^{(h)}) = \sum_{z=1}^Z \alpha_{(z)}^{(h)} \mathbf{cov}(\mathbf{c}_X^{(h)}, \mathbf{c}_{Y(z)}^{(h)}) \leq \sum_{z=1}^Z \alpha_{(z)}^{(h)} \mathbf{cov}(\mathbf{c}_X^{(h+1)}, \mathbf{c}_{Y(z)}^{(h+1)}) \quad (2)$$

L'inégalité (2) vient du fait que $\mathbf{u}^{(h+1)}$ et $\mathbf{v}^{(h+1)}$ sont les solutions de (1).

En appliquant l'inégalité de Cauchy-Shwartz :

$$\sum_{z=1}^Z \alpha_{(z)}^{(h)} \mathbf{cov}(\mathbf{c}_X^{(h+1)}, \mathbf{c}_{Y(z)}^{(h+1)}) \leq \sqrt{\sum_{z=1}^Z \alpha_{(z)}^{(h)2}} \sqrt{\sum_{z=1}^Z \mathbf{cov}(\mathbf{C}_X^{(h+1)}, \mathbf{C}_{Y(z)}^{(h+1)})^2} \quad (3)$$

En combinant les inégalités (2) et (3), on obtient :

$$\sum_{z=1}^Z \mathbf{cov}^2(\mathbf{c}_X^{(h)}, \mathbf{c}_{Y(z)}^{(h)}) \leq \sum_{z=1}^Z \mathbf{cov}^2(\mathbf{c}_X^{(h+1)}, \mathbf{c}_{Y(z)}^{(h+1)})$$

Contributions scientifiques

Revue internationale avec comité de lecture

1. Allouche F., M. Hanafi, F. Jamme, P. Robert, F. Guillon, M.F. Devaux, Coupling hyperspectral image data having different spatial resolutions by extending multivariate inter-battery Tucker analysis, *Chemometer. Intell. Lab. Syst.* 113 (2010) 43-51.
2. Allouche F., M. Hanafi, F. Jamme, P. Robert, F. Guillon, M.-F. Devaux, Coupling hyperspectral image data having different spatial resolutions using Multiple co-inertia analysis, *Chemometer. Intell. Lab. Syst.* <http://dx.doi.org/10.1016/j.chemolab.2012.04.004>.

Conférence avec comité de lecture

1. Allouche F., M.F. Devaux, F. Jamme, P. Robert, F. Guillon, Maize cell walls heterogeneity according to cell types using hyperspectral imaging: Infrared Microspectroscopy, 5th SOLEIL Users' meeting, Paris, 2010, (communication par poster).
2. Allouche F., M. Hanafi, F. Jamme, P. Robert, F. Guillon, M.F. Devaux, Relating Hyperspectral images having different spatial resolutions, *Chimiométrie*, Paris, 2010, (communication orale).
3. Devaux M.F., F. Allouche, F. Jamme, P. Robert, F. Guillon, Spatial and spectral normalisation of hyperspectral images. *Chimiométrie*, Paris, 2010, (communication par poster).
4. Allouche F., M.F. Devaux, F. Jamme, P. Robert, F. Guillon, Polymer distribution mapping in maize stems using hyperspectral imaging: Mid-Infrared, Fluorescence (synchrotron source) and Raman microspectroscopy. 6th SOLEIL Users' meeting, Paris, 2011, (communication orale).
5. Allouche F., M. Hanafi, F. Jamme, P. Robert, F. Guillon, M.F. Devaux, Inter-battery Tucker Analysis in the presence of three way data table and application in Hyperspectral images, 43^{ème} Journées de Statistique, Tunis, 2011, (communication orale).

Conférence nationale sans comité de lecture

1. Allouche F., M. Hanafi, F. Jamme, P. Robert, F. Guillon , M.F. Devaux, Analyse de co-inertie multiple en présence de tableaux à trois entrées et applications en analyse d'images hyperspectrales, Journées Calcul parallèle et statistique spatiale, Paris, 2011, (communication orale).
2. Allouche F., M. Hanafi, F. Jamme, P. Robert, F. Guillon , M.F. Devaux, Analyse multiboc d'images hyperspectrales multimodales multirésolutions: cartographie des tiges de maïs en rayonnement synchrotron, Journées des jeunes chercheurs, Nantes, 2011, (communication orale).

Thèse de Doctorat

Fatma ALLOUCHE

Analyse multibloc d'images hyperspectrales multimodales et multirésolutions : cartographie des parois de tiges de maïs par microspectroscopies en rayonnement synchrotron

Résumé

Le couplage de différentes techniques de microspectroscopie est un outils puissant pour caractériser *in situ* des matériaux biologiques complexes. Dans ce projet de recherche, nous proposons de coupler trois techniques d'imagerie hyperspectrale complémentaires en utilisant les microspectroscopies infrarouge, de fluorescence et Raman. Ce travail est réalisé pour caractériser la composition chimique des parois cellulaires de tiges de maïs dans un contexte de production de bio-carburant. Les méthodes multibloc ont été retenues pour analyser les images hyperspectrales couplées. En pratique, chaque microspectroscopie fonctionne avec sa propre résolution. Chaque spectre infrarouge correspond à un ensemble de spectres de fluorescence et de Raman. La structure de données préservant la résolution spatiale est obtenue suite au dépliement de l'image infrarouge. La structure de données résultante consiste en un tableau à deux entrées couplé à deux tableaux à trois entrées. Une telle structure de données ne peut pas être analysée directement avec une méthode multibloc classique. Une idée naturelle est de réduire la résolution des images pour obtenir des tableaux de données à deux entrées. L'inter-batterie de Tucker et l'analyse de co-inertie multiple sont deux méthodes d'analyse multivariée utilisées comme outils pour relier respectivement deux et plusieurs tableaux de données à deux entrées. Ces deux méthodes peuvent être étendues si la structure de données contient des tableaux à trois entrées. Les approches, développées dans cette thèse pour préserver la résolution spatiale, sont basées sur l'analyse inter-batterie de Tucker et l'analyse de co-inertie multiple en considérant les tableaux à trois entrées comme étant un ensemble de tableaux à deux entrées.

Mots clés

imagerie hyperspectrale, microspectroscopie infrarouge, microspectroscopie de fluorescence, microspectroscopie Raman, analyse inter-batterie de Tucker, analyse de co-inertie multiple, analyse d'images multivariées, imagerie synchrotron.

Abstract

Coupling different techniques in microspectroscopy is an efficient approach to characterise complex biological material. In this research project, we propose to couple infrared, fluorescence and Raman hyperspectral images. This work is carried out to characterise maize stem cell wall composition in the context of biofuel production. Investigations have been attempted in coupling spectral data by means of multiblock methods. Multiblock techniques have been assumed to be equally applicable and effective in the analysis of complex biological samples. In practice, each microspectroscopy operates with its own resolution. Each infrared spectrum corresponds to a set of fluorescence and Raman spectra. A data structure that preserved the spatial resolution, is obtained by unfolding infrared hyperspectral image. The resulting data structure consists in a two-way data table with a spatial way and a spectral way, paired to a three-way data table with two spatial ways and a spectral way. Such data structure can not be analysed with classical multibloc methods. A natural idea is to reduce the resolution of the images to obtain data tables with two ways. Inter-battery Tucker analysis and multiple co-inertia analysis are used as a tool to relate respectively two datatables and several data tables having two ways. Models of decomposition of these two methods can be extended if the data structure contains three-way data tables. The approaches developed in this thesis to preserve spatial resolution are based on the analysis of inter-battery Tucker analysis and multiple co-inertia analysis by considering three-way datatables as a set of two-way datatables.

Key Words

hyperspectral imaging, infrared microspectroscopy, fluorescence microspectroscopy, Raman microspectroscopy, inter-battery Tucker analysis, multiple co-inertia analysis, multivariate image analysis, synchrotron imaging