

THESE

Présentée à

L'Université des Sciences et Technologies de Lille

Pour obtenir le grade de

Docteur de l'Université des Sciences et Technologies de Lille
Mention : Instrumentation, analyses avancées

Et de

Doctor de l'Universitat de Barcelona
en Química Analítica

Présentée par

Lionel BLANCHET

Méthodes de résolution dédiées à l'étude spectroscopique de processus photoinduits. Adaptation aux spécificités des spectres résolus en temps

Rapporteurs : **Ricard BOQUE**, Professeur, Universitat Rovira i Virgili, Tarragone

Pascal PERNOT, Chargé de Recherche CNRS, Université de Paris 11

Examineurs : **Lutgarde BUYDENS**, Professeur, Radboud Universiteit Nijmegen, Nimègue

Romà TAULER, Professeur, CSIC, Barcelone

Anna DE JUAN, Professeur, Universitat de Barcelona, Barcelone

Cyril RUCKEBUSCH, Maître de Conférences HDR, Université de Lille 1

Jean-Pierre HUVENNE, Professeur, Université de Lille 1

THESE

Présentée à

L'Université des Sciences et Technologies de Lille

Pour obtenir le grade de

Docteur de l'Université des Sciences et Technologies de Lille
Mention : Instrumentation, analyses avancées

Et de

Doctor de l'Universitat de Barcelona
en Química Analítica

Présentée par

Lionel BLANCHET

**Méthodes de résolution dédiées à l'étude spectroscopique de
processus photoinduits. Adaptation aux spécificités des spectres
résolus en temps**

Sommaire

Abréviations	13
Notations mathématiques	14
Introduction	17
Chapitre I : Analyse multivariée de données spectroscopiques en suivi de procédés.....	23
1. Mesures, données et outils statistiques.....	25
1.1 Mesures et données	25
1.2 Nature et effet du bruit affectant la mesure.....	31
2. Analyse exploratoire de matrices de données de suivi de procédé	34
2.1 Analyse en Composantes Principales (ACP)	34
2.2 Rang de matrices expérimentales	37
2.3 EFA	41
2.4 SIMPLISMA	46
3. Méthodes de modélisation souple de procédés (soft-modelling)	49
3.1 Ambiguïtés de rotation	49
3.2 Résolution Multivariée de Courbes par Moindres Carrés Alternés (MCR-ALS)	50
4. Méthode de modélisation rigide de procédés	58
5. Méthode de modélisation hybride souple et rigide de procédés	64
6. Conclusion.....	68
Chapitre II : Développements méthodologiques pour la spectroscopie résolue en temps	71
1. Déficience de rang induite par la structure des données de différence	73
2. CAS 1: données simulées avec un schéma réactionnel de rang plein	74
2.1 Estimation théorique du rang	74
2.2 Données simulées	76
2.3 Estimation du rang par SVD	78
2.4 Analyse factorielle évolutive (EFA)	79
2.5 Résolution multivariée hybride	81
3. CAS 2 : données simulées avec un schéma réactionnel de rang plein affecté d'une contribution non cinétique.....	85
3.1 Estimation théorique du rang	85
3.2 Simulations de données	85
4. CAS 3 : données simulées avec un schéma réactionnel induisant une déficience de rang 91	
5. Conclusion.....	93

Chapitre III : Contribution à l'étude du centre réactionnel de *Rhodobacter sphaeroides* par spectroscopie de différence IRTF *rapid-scan* et méthodes de résolution de courbes

multivariées	95
1. Système biochimique	97
1.1 Photosynthèse et photochimie	97
1.2 Photosynthèse bactérienne	100
2. Matériels et méthodes.....	105
3. Analyse exploratoire	108
3.1 Détection du rang et estimation initiales	109
4. Modélisation souple du procédé. Application de MCR-ALS	113
4.1 Résolutions individuelles	113
4.2 Résolution en augmentation de matrices.....	116
5. Résolution avec la méthode hybride HS-MCR.....	120
5.1 Choix d'un modèle cinétique	120
5.2 Résultats de l'application de la méthode hybride HS-MCR	123
6. Conclusion.....	128

Chapitre IV : Étude de la photochimie de la benzophénone par spectroscopie d'absorption

transitoire.....	131
1. Principes de la photochimie	133
1.1 Principes	133
1.2 Mécanisme d'excitation	134
1.3 Mécanisme de relaxation.....	136
1.4 Propriétés des états excités	138
2. Systèmes photochimiques	140
2.1 Photochimie de la benzophénone	140
2.2 Photochimie de la méthoxybenzophénone	142
3. Matériels et méthodes.....	144
3.1 Spectroscopie d'absorption transitoire	144
3.2 Mesures réalisées.....	146
4. Analyse exploratoire	148
4.1 Rang des données concernant la benzophénone	148
4.2 Rang des données concernant la méthoxybenzophénone	150
5. Résolution MCR-ALS.....	153
5.1 Analyse de la benzophénone	153
5.2 Résolution sur une augmentation de matrices.....	155
5.3 Analyse des résidus	157
5.4 Analyse de la méthoxybenzophénone	159
6. Resolution HS-MCR	163
7. Conclusion et perspectives sur cette application.....	167

Chapitre V : Caractérisation de structures de bruit en spectroscopie résolue en temps et intégration de cette information dans l'algorithme de résolution.....	169
1. Estimation et analyse de l'erreur	171
2. Caractérisation des contributions structurées à l'erreur	178
3. Pondération des données en fonction de l'erreur	181
3.1 Simulation de données affectées par une erreur structurée	182
3.2 Cas réel	192
4. Conclusion et perspectives	193
Conclusion.....	195
Bibliographie	199
Síntesis: Análisis multivariante de espectros resueltos en el tiempo	205
Análisis de datos multivariantes.....	210
Multivariate curve Resolution – Alternating Least Squares (MCR-ALS).....	210
<i>Hard modelling</i> de datos cinéticos	212
Modelado híbrido	212
Conclusión.....	213
Modelado de datos de espectroscopia de diferencia	214
Rango de datos espectroscópicos de diferencia	214
Modelado híbrido de datos de diferencia	216
Conclusiones	218
Publicaciones relacionadas con esta sección.....	219
Estudio del centro de reacción de <i>Rhodobacter sphaeroides</i> utilizando espectroscopia IRTF rapid-scan y quimiometría.....	220
Fotosíntesis bacteriana	220
Soft-modelling.....	220
Hard- and Soft-Modelling	222
Conclusiones	225
Publicaciones relacionadas con esta sección.....	225
Estudio de la fotoquímica de la benzofenona utilizando espectroscopia de absorción transitoria y quimiometría	227
Relajación de la benzofenona.....	227
Conclusiones	230
Publicaciones relacionadas con esta sección.....	230
Caracterización de la estructura del ruido y de su influencia en la resolución de datos	232
Determinación de la estructura del ruido y de la presencia de contribuciones correlacionadas.....	232
Utilización de MCR-WALS (resolución multivariante de curvas por mínimos cuadrados alternados ponderados).....	233

Publicaciones relacionadas con esta sección.....	240
Conclusión.....	241
Referencias	243

Table des figures

Figure 1 : Représentation schématique d'une mesure prenant la forme a) scalaire b) vectorielle c) matricielle.....	25
Figure 2 : Représentation a)d'un spectre et b) d'une matrice de spectres D_{ex} suivant un procédé cinétique.....	26
Figure 3 : Décomposition d'une matrice de données bilinéaires D a) représentation par constituants b) représentation matricielle, à l'erreur E près.....	27
Figure 4 : Représentation schématique d'une mesure générant un cube de données et de la décomposition trilinéaire de celui-ci.	28
Figure 5 : Représentation schématique d'une augmentation de matrices dans a) le sens des concentrations b) le sens des spectres c) les deux sens.	29
Figure 6 : Décomposition bilinéaire d'une augmentation de matrices dans la direction des lignes.	30
Figure 7 : Décomposition bilinéaire d'une augmentation de matrices dans la direction des colonnes.....	30
Figure 8 : Représentation de trois types de bruit a) indépendant homoscédastique b) indépendant hétéroscédastique et c) structure de bruit.....	33
Figure 9 : Décomposition bilinéaire par ACP d'une matrice de données D.....	35
Figure 10: Principe de la décomposition en valeurs singulières de D.	35
Figure 11 : Comparaison de la SVD et de l'algorithme NIPALS	36
Figure 12 : Décomposition par SVD de la matrice D_{ex} en trois paires de <i>scores</i> et <i>loadings</i> ..	36
Figure 13: Décomposition par SVD de la matrice D_{ex} en scores U_{ex} et loadings V_{ex}	37
Figure 14 : Principe de l'Analyse Factorielle Evolutive a) en sens direct b) en sens inverse..	43
Figure 15 : Analyse Factorielle Evolutive de D_{ex} en sens a) direct b) inverse.....	43
Figure 16 : Graphique contenant les informations du EFA <i>forward</i> et EFA <i>backward</i> de D_{ex} , la ligne horizontale en trait noir épais représente le niveau de bruit en dessous duquel on ne considère plus les contributions détectées.....	44
Figure 17 : Profils de concentration C_{ex} de D_{ex} estimés par EFA	45
Figure 18 : Représentation graphique de l'obtention de profils de concentration ou de spectres purs par la méthode SIMPLISMA	46
Figure 19 : Estimations construites par SIMPLISMA a) des spectres purs S_{ex} et b) des profils de concentration C_{ex}	48

Figure 20 : Exemple d'application de la contrainte de non négativité.....	51
Figure 21 : Exemple d'application de la contrainte d'unimodalité.....	51
Figure 22 : Exemple d'application de la contrainte de système fermé.	52
Figure 23 : Schéma de principe de MCR-ALS	55
Figure 24 : Résultats de la décomposition par MCR-ALS de D_{ex} en a) profils de concentrations C_{ex} et b) spectres purs S_{ex}	56
Figure 25 : Illustration de la souplesse d'application des contraintes à une augmentation de matrices dans la direction des colonnes.	57
Figure 26 : Principe de la méthode des moindres carrés non linéaires, la matrice D est utilisée pour estimer les paramètres d'un modèle (ici deux paramètres cinétiques k_1 et k_2) qui sont ensuite utilisés pour calculer C et S.	59
Figure 27 : Schéma de principe de l'algorithme de Newton-Gauss.....	60
Figure 28 : Schéma de principe de l'algorithme de Levenberg-Marquardt	61
Figure 29 : Résultats de la décomposition par la méthode des moindres carrés non linéaires de D_{ex} en a) profils de concentrations C_{ex} et b) spectres purs S_{ex}	62
Figure 30 : Schéma de principe de HS-MCR.....	65
Figure 31 : Représentation de l'effet de la contrainte cinétique sur les profils de concentrations, la zone en grisé correspond aux profils participant au modèle physico chimique.....	66
Figure 32 : Matrices de données simulées a) D_1 b) ΔD_1	77
Figure 33 : Graphique de l'évolution des valeurs propres issu de l'EFA de D_1	80
Figure 34 : Graphique de l'évolution des valeurs propres issu de l'EFA de ΔD_1	80
Figure 35 : Profils de concentrations obtenus par résolution HS-MCR Les lignes pleines, en tirets et en pointillé représentent respectivement la résolution de D_1 , la résolution de ΔD_1 et le profil reconstitué de la première espèce (transparente).....	82
Figure 36 : Spectres purs obtenus par résolution HS-MCR a) pour D_1 , b) pour ΔD_1 représentés en tirets, les lignes pleines correspondent aux spectres de différence calculés à partir des spectres purs issu de la décomposition de D_1	83
Figure 37 : Matrices de données simulées a) ΔD_2 b) ΔD_{2B}	86
Figure 38 : EFA de a) ΔD_2 b) ΔD_{2B}	87
Figure 39 : Décompositions effectuées par HS-MCR des matrices a) ΔD_2 et b) ΔD_{2B}	88
Figure 40 : Profils de concentration de a) ΔD_2 et b) ΔD_{2B} et spectres purs correspondants c) ΔD_2 et d) ΔD_{2B}	89

Figure 41 : Illustration du transfert d'énergie de Forster. Les spectres numérotés correspondent à 1) l'absorption du donneur 2) la fluorescence du donneur 3) l'absorption de l'accepteur 4) la fluorescence de l'accepteur. La zone grisée correspond à la zone de recouvrement entre la fluorescence du donneur et l'absorption de l'accepteur.	98
Figure 42 : Formule générale des bactériochlorophylles.	99
Figure 43 : Structure tridimensionnelle du centre réactionnel de <i>Rhodobacter sphaeroides</i> . A) Représentation des 11 hélices α formant l'enveloppe du centre réactionnel B) les 10 hélices en rouge et en vert forment les deux protéines au positionnement symétrique, perpendiculairement à l'axe cytoplasmique (AS). C) Chaîne redox : P870 est le donneur primaire d'électrons ; d'après [107].	101
Figure 44 : Formule de l'ubiquinone-10 ; dans le centre réactionnel de <i>Rhodobacter sphaeroides</i> Q_A et Q_B sont toutes deux des molécules d'ubiquinone-10.	102
Figure 45 : Formule de l'ubiquinol réduite à partir de Q_B dans le centre réactionnel de <i>Rhodobacter sphaeroides</i> durant la photoexcitation.	103
Figure 46 : Cycle de réduction de la quinone dans les centres réactionnels des bactéries. Q_B est réduite deux fois, puis se lie successivement à deux protons. La molécule d'ubiquinol formé (QH_2) quitte le centre et est remplacée par une quinone.	103
Figure 47 : Transfert d'électrons chez une bactérie pourpre sur une échelle de potentiels redox. P870 est le donneur primaire d'électrons, siège de la séparation de charge ; H_A , une bactériophéophytine, est l'accepteur primaire de l'électron éjecté. Une autre voie redox apporte, via des sulfures S^{2-} , des électrons au P870.	104
Figure 48 : Schéma d'une expérience <i>rapid-scan</i> utilisant une excitation de courte durée... ..	105
Figure 49 : Schéma d'une expérience <i>rapid-scan</i> utilisant une excitation de longue durée.. ..	106
Figure 50 : Spectres IRTF <i>rapid scan</i> obtenus durant le suivi de la photoexcitation de membranes de <i>Rb. Sphaeroides</i> , a) phase d'excitation b) phase de relaxation.	108
Figure 51 : Résultats de l'EFA en sens direct (noir) et inverse (rouge) pour les matrices a) D_1 , b) D_2 , c) D_3 , d) D_4 et e) D_5	110
Figure 52 : EFARD calculée sur la matrice augmentée [D_3 ; D_1].....	111
Figure 53 : EFARD calculée sur la matrice augmentée [D_1 ; D_2].....	112
Figure 54 : Résultats des décompositions individuelles des matrices D_3 , D_4 et D_5 par MCR-ALS en a) C_3 , b) S_3 , c) C_4 , d) S_4 , e) C_5 et f) S_5	114
Figure 55 : Résultats de la décomposition individuelle de la matrice D_2 par MCR-ALS en a) C_2 , et b) S_2	115

Figure 56 : Structure de la matrice augmentée dans la direction des colonnes utilisée lors de la résolution MCR-ALS	116
Figure 57 : Spectres de différence S^T issus de la résolution MCR-ALS de la matrice augmentée $[D_1; D_2; D_3; D_4; D_5]$. De haut en bas spectres attribués aux changements de conformation de la protéine, à la formation d'ubiquinol sous éclaircissement faible et à la formation d'ubiquinol sous éclaircissement fort.	117
Figure 58 : Profils de concentration a) C_1 , b) C_2 , c) C_3 , d) C_4 et e) C_5 en fonction du temps calculés en utilisant MCR-ALS sur la matrice augmentée. Le profil correspondant aux variations de l'environnement protéique est en bleu, à la formation d'ubiquinol sous éclaircissement faible en rouge et à la formation d'ubiquinol sous éclaircissement fort en vert.....	119
Figure 59 : Schéma de l'augmentation de matrices $[D_1; D_2; D_3; D_4; D_5]$ et du type de contraintes appliquées à chaque contribution, la contribution protéique (en bleue) est sous contraintes souples, les deux autres doivent respecter des contraintes cinétiques.	122
Figure 60 : Application de deux modèles interdépendants (excitation et relaxation) pour chacune des matrices D_1 à D_5	123
Figure 61 : Spectres de différence S^T issus de la résolution HS-MCR de la matrice augmentée $[D_1; D_2; D_3; D_4; D_5]$. De haut en bas spectres attribués au changement de conformation de la protéine, à la formation d'ubiquinol sous éclaircissement faible et à la formation d'ubiquinol sous éclaircissement fort.	124
Figure 62 : Profils de concentrations a) C_1 , b) C_2 , c) C_3 , d) C_4 et e) C_5 en fonction du temps calculé en utilisant HS-MCR. Le profil correspondant à la variation de l'environnement protéique est en bleu, à la formation d'ubiquinol sous éclaircissement faible en rouge et à la formation d'ubiquinol sous éclaircissement fort en vert.	125
Figure 63 : Excitation électronique d'une molécule diatomique	135
Figure 64 : Diagramme de Perrin-Jablonski représentant les différentes transitions électroniques possibles.....	137
Figure 65: Formule de la benzophénone	140
Figure 66 : Diagramme d'énergie proposé pour la benzophénone	141
Figure 67 : Formule de la méthoxybenzophénone.	142
Figure 68 : Diagramme énergétique pour les états triplets les plus bas de la 4-MeBP dans des solvants de différentes polarités, d'après [144].....	142
Figure 69 : Schéma de principe de la spectroscopie d'absorption par la méthode pompe-sonde.	145

Figure 70 : Spectres d'absorption transitoire de la benzophénone après photoexcitation à 267 nm (de 0,8 ps, en rouge, à 50 ps, en bleu).....	148
Figure 71 : Résultats de l'EFA en sens direct (trait plein) et inverse (pointillés) pour les matrices a) D_1 , b) D_2 , c) D_3 et d) D_4	150
Figure 72 : Spectres d'absorption transitoire de la méthoxybenzophénone après photoexcitation à 267 nm (de 1,2 à 100 ps) dans a) le cyclohexane b) l'acétonitrile c) un mélange acétonitrile eau 1:3.....	151
Figure 73 : Résultats de l'EFA en sens direct (trait plein) et inverse (pointillés) pour les matrices a) D_{Me}^{CH} , b) D_{Me}^{ACN} et c) $D_{Me}^{ACN/H2O}$	152
Figure 74 : Résultats de l'analyse MCR-ALS de D_1 : a) profils cinétiques b) spectres purs de l'état S_1 (ligne continue) T_2 (tirets) et T_1 (tirets pointillés).....	154
Figure 75 : Profils de concentrations de l'état S_1 (ligne continue bleue) T_2 (pointillés verts) et T_1 (tirets pointillés rouges) pour [D_2 ; D_3 ; D_4] obtenus par MCR-ALS.....	155
Figure 76 : Spectres purs de l'état S_1 (ligne continue bleue) T_2 (pointillés verts) et T_1 (tirets pointillés rouges) pour [D_2 ; D_3 ; D_4] obtenus par MCR-ALS.....	156
Figure 77 : Spectres purs (gauche) et résidus en la direction spectrale (droite) pour des résolutions MCR-ALS un rang allant de 2 à 4.	158
Figure 78 : Résultats cinétiques des résolutions MCR-ALS pour un rang allant de 2 à 4 et résidus correspondants	159
Figure 79 : Résolutions MCR-ALS des spectres d'absorption de la méthoxybenzophénone dans a) le cyclohexane b) l'acétonitrile c) un mélange acétonitrile eau (3:1).....	161
Figure 80 : Résultats de l'analyse HS-MCR de D_1 : a) profils cinétiques b) spectres purs de l'état S_1 T_2 et T_1	164
Figure 81 : a) Profils de concentration et spectres obtenus b) S_2 c) S_3 et d) S_4 ; lors de la résolution HS-MCR de [D_2 ; D_3 ; D_4] en utilisant un modèle global.....	165
Figure 82 : Calcul de la matrice de covariance de l'erreur Σ dans la direction spectrale.	172
Figure 83 : Tranches représentatives du cube Ψ à 586,0 et 387,5 nm. L'échelle de couleur correspond à la valeur de la covariance.....	173
Figure 84 : Tranches représentatives de la matrice Σ à 3,2 et 50 ps. L'échelle de couleur correspond à la valeur de la covariance.....	174
Figure 85 : Matrices de corrélation de l'erreur de mesure sur les longueurs d'onde au temps $t = 3,2$ ps et $t = 50$ ps. L'échelle de couleur correspond à la valeur de la corrélation.	175

Figure 86 : Représentation de toutes les tranches formant le cube de covariance de l'erreur Σ .	176
Figure 87 : <i>Loadings</i> de la matrice Σ obtenus par PARAFAC a) 1 ^{er} mode b) 2 nd mode c) 3 ^{ème} mode.	177
Figure 88 : Cibles sélectionnées par TFA, en bleu le signal moyen et en vert le signal moyen de la source.	179
Figure 89 : Diagonale de a) la matrice de covariance de l'erreur $\bar{\Sigma}$ et b) sa reconstruction basée sur les vecteurs choisis par TFA.	180
Figure 90 : a) Profils de concentration et b) spectres purs utilisés pour construire les matrices D_1 à D_4 .	183
Figure 91 : Matrice de données simulées D_2 .	184
Figure 92 : Résolutions de D_1 obtenues par MCR-ALS a) profils de concentration b) spectres purs et par MCR-WALS c) profils de concentration d) spectres purs.	187
Figure 93 : Résolutions de D_2 obtenues par MCR-ALS a) profils de concentration b) spectres purs ; par MCR-WALS c) profils de concentration d) spectres purs.	188
Figure 94 : Résolutions de D_3 obtenues par MCR-ALS a) profils de concentration b) spectres purs et par MCR-WALS c) profils de concentration d) spectres purs.	189
Figure 95 : Résolutions de D_4 obtenues par MCR-ALS a) profils de concentration b) spectres purs ; par MCR-WALS en utilisant un schéma de pondération complet c) profils de concentration d) spectres purs ; par MCR-WALS en utilisant un schéma de pondération de considérant que l'erreur structurée ayant une forme différente du signal e) profils de concentration f) spectres purs.	191
Figure 96 : a) Profils de concentrations et b) spectres obtenu par MCR-WALS en appliquant un schéma de pondération prenant en compte le bruit indépendant et le bruit structuré dépendant de la source.	192

Liste des tableaux

Tableau 1 : Valeurs singulières et pourcentage de variance capturée par les CPs de la matrice D_{ex} . Les valeurs en gras correspondent aux valeurs singulières associées au signal.	40
Tableau 2 : Premières valeurs singulières de D_1 et ΔD_1 (les valeurs en gras représentent les valeurs singulières associées au signal).....	78
Tableau 3 : Premières valeurs singulières de ΔD_2 et ΔD_{2B}	87
Tableau 4 : Conditions expérimentales des différentes expériences enregistrées sur le même échantillon.	107
Tableau 5 : Premières valeurs singulières issues de la SVD des matrices de données D_1 , D_2 , D_3 , D_4 et D_5	109
Tableau 6 : Constantes de réactions calculées en utilisant HS-MCR	126
Tableau 7 : Règles d'El Sayed pour les composés carbonylés [129], transitions entre les singulets (notés ¹) et les triplet (notés ³).	138
Tableau 8 : Échelles de temps des différents mécanismes photochimiques.	138
Tableau 9 : Premières valeurs singulières des données concernant la benzophénone.	149
Tableau 10 : Premières valeurs singulières des données concernant la méthoxybenzophénone.	151
Tableau 11 : Propriétés spectrocinétiques obtenues lors de l'analyse HS-MCR de spectres d'absorption transitoire de la benzophénone après différentes excitations.....	166
Tableau 12 : Résultats de la TFA sur la matrice de covariance de l'erreur moyenne $\bar{\Sigma}$	179
Tableau 13 : Premières valeurs singulières des matrices de données simulées.	184
Tableau 14 : Lof théoriques et des différents modèles MCR-ALS et MCR-WALS pour les matrices D_1 , D_2 , D_3 et D_4 . Les deux valeurs fournies pour la résolution MCR-WALS de D_4 correspondent aux schémas de pondération (1) et (2) appliqués.....	186

Abréviations

ACP : Analyse en Composantes Principales

BP : Benzophénone

EFA : Evolving Factor analysis

HS-MCR : Hard Soft – Multivariate Curve Resolution

IRTF: InfraRouge à Transformée de Fourier

IC : Internal Conversion

ICA : Independant Component Analysis

ISC : InterSystem Crossing

LMA : Levenberg-Marquardt Algorithm

MCR-ALS : Multivariate Curve Resolution – Alternating Least Squares

MeBP : Methoxybenzophénone

OPA: Orthogonal Projection Analysis

PARAFAC : PARAllel FACtor analysis

PCA : Principal Composant Analysis

SIMPLISMA: SIMPLe to use Interactive Self-modeling Mixture Analysis

SVD : Singular Value Decomposition

TFA : Target-testing Factor Analysis

Notations mathématiques

C : Matrice des profils de concentrations

Γ : Matrice de variance-covariance du signal

D : Matrice de données brutes (2 voies)

D : Matrice de données brutes (3 voies)

\bar{D} : Matrice de données centrée

\hat{D} : Matrice de données reconstruite

$d_{i,j}$: Éléments de **D**

$\hat{d}_{i,j}$: Éléments de **\hat{D}**

Δ : Matrice de variance-covariance du signal diagonalisée

k : Rang de la matrice de données **D**

m : Nombre de lignes de **D**

n : Nombre de colonnes de **D**

P : Matrice des vecteurs propres (loadings)

p : Nombre d'éléments sur la troisième dimension de **D**

R : Matrice de rotation (inversible)

S : Matrice de spectres purs

T : Matrice des coordonnées factorielles (scores)

Introduction

La chimiométrie est la science qui relie les mesures faites sur un système ou un procédé chimique avec son état par l'application de méthodes mathématiques ou statistiques. Cette définition de l'*International Chemometrics Society* est très large mais elle répond au grand nombre de domaines d'application de la chimiométrie. Dans ce travail, la partie de la chimiométrie qui nous intéresse est celle qui fournit des outils mathématiques et statistiques permettant de séparer les informations concernant différentes espèces chimiques pures contenues dans un mélange. Ce type de problème ne constitue qu'une branche de la chimiométrie, elle est généralement désignée par le terme "analyse de mélanges" et les méthodes qui s'y relient le plus sont la "résolution de courbes multivariées" qui appartient à la grande famille de méthodes nommée "analyse de facteurs". Dans le problème d'analyse de mélanges, on doit encore faire une distinction selon que l'on connaisse ou non les espèces chimiques présentes dans l'échantillon analysé. Le second cas de figure est évidemment le plus complexe mais aussi le plus courant en recherche de procédés. Nous verrons comment la chimiométrie permet en association à des techniques instrumentales avancées, telle que la spectroscopie, d'apporter des informations sur des composés chimiques inconnus et parfois impossibles à isoler.

La spectroscopie permet de définir la structure atomique ou moléculaire d'un composé par l'analyse de son spectre. L'analyse de données spectroscopiques implique l'utilisation de méthodes multivariées. En effet chaque mesure consiste au minimum en un vecteur de plusieurs centaines de valeurs. La multiplication de ces mesures génère donc d'importantes quantités de données qui si on veut en tirer le meilleur profit doivent être analysées de manière exhaustive. La solution consistant à suivre uniquement une ou quelques longueurs d'ondes jugées d'intérêt n'est donc pas acceptable de ce point de vue. Il est alors nécessaire d'utiliser des outils d'analyses multivariées qui suivent une loi linéaire, en accord avec la loi de Lambert-Beer, inhérente à la mesure spectroscopique. L'application de ces outils à l'analyse de procédés permettra, en général, une description qualitative et quantitative complète des phénomènes prenant place dans le système d'intérêt. Cette description est possible grâce à l'estimation des profils de concentration cinétiques et des spectres purs de

chaque espèce impliquée dans le procédé. Les procédés qui nous intéressent plus particulièrement ici sont ceux suivis par la spectroscopie résolue en temps. Le but de celle-ci est d'obtenir des informations sur le déroulement d'une réaction cinétique, en identifiant les intermédiaires. On comprend bien que pour suivre un phénomène dans le temps il faut accumuler les mesures à différents stades de ce processus. Le volume de données à traiter devient donc rapidement important. De plus les instrumentations avancées, telles que celles rencontrées en recherche, sont souvent très spécifiques et impliquent certaines difficultés qui jusqu'alors n'ont pas été envisagées. C'est dans ce cadre que s'inscrit cette thèse effectuée en cotutelle en l'Université des sciences et Technologies de Lille (USTL) et l'Université de Barcelone (UB).

L'objectif principal de ce travail est d'apporter de développements chimiométriques aux méthodes d'analyse de données existantes de manière à améliorer l'interprétation des procédés suivis par spectroscopie résolue en temps. Ce travail s'effectue donc en étroite collaboration avec les différents chercheurs utilisant les méthodes spectroscopiques les plus avancées. Les différents points clés peuvent être définis comme suit.

Le premier point se rapporte à la chimiométrie pour la spectroscopie résolue en temps. L'étude de processus dépendant du temps par spectroscopie génère d'importants volumes de données, riches en information nécessitant d'être analysés par des outils chimiométriques. La première étape consiste donc à identifier les algorithmes adaptés à cette tâche et leurs avantages respectifs. Les principes régissant certaines de ces méthodes seront décrits de manière à permettre à la fois leur compréhension et à mettre en lumière les avantages et inconvénients de chacun. Plus spécifiquement, nous nous focaliserons sur la description de plusieurs types de méthodes. Nous évoquerons dans un premier temps la modélisation souple de procédés (*soft-modelling*), qui ne repose que sur le modèle bilinéaire de la mesure spectroscopique et l'obéissance à quelques contraintes générales associées à la mesure instrumentale et la nature chimique des données. Nous détaillerons ensuite les méthodes de modélisation rigide (*hard-modelling*), qui décrivent la variation de la mesure enregistrée par l'ajustement d'un modèle physicochimique. Finalement nous proposerons des méthodes hybrides souple-rigides (*hard- and soft-modelling*), qui combinent les avantages des deux familles d'algorithmes parentes.

Le second point est en rapport avec l'influence de la structure de données de différence. Le suivi de procédés chimiques en temps courts requiert parfois l'utilisation de méthodes de détection différentielle ou de ne considérer que de faibles variations du signal. Les mesures obtenues prennent alors la forme de données de différence. Cette structure à l'avantage de mettre en valeur l'information d'intérêt, c'est à dire la faible variation, rapide, subie par le système. Cependant il est nécessaire de prendre certaines précautions lors de l'analyse et l'interprétation de ces mesures. Une analyse théorique doit être effectuée pour évaluer correctement les particularités de ce type de données. En particulier, nous analyserons comment la perte d'information sur l'état initial du procédé peut affecter l'estimation du nombre de contributions détectables dans un procédé et comment il faut interpréter les profils cinétiques et les spectres purs (de différence) associés à chacune des contributions modélisées.

Un des principaux points clés est à mettre en rapport avec la gestion de données de différence résolue en temps. Les résultats de l'étude théorique précédente indiquent qu'il est nécessaire de développer de nouveaux outils pour travailler sur ce genre de système. Une stratégie tirant le meilleur profit de l'information disponible et de l'organisation intrinsèque des données résolues en temps doit être mise au point. Elle doit permettre d'obtenir une estimation fiable de l'évolution des concentrations des différentes espèces impliquées dans un procédé chimique ainsi que les spectres purs permettant de les identifier. Ce besoin se traduit par l'incorporation de modèles cinétiques dans des méthodes de modélisation hybride qui prennent en compte la première contribution du procédé (non détectable), définie comme espèce transparente, mais qui ne l'associent pas aux spectres purs récupérés. On prêtera une attention particulière aux procédés photochimiques qui se déroulent en étapes successives d'excitation-relaxation et qui impliquent, par conséquent, l'introduction des modèles différents et opposés pour les deux phases du procédé. Un ajustement cinétique séquentiel est proposé afin de respecter l'interconnexion entre ces deux phases. En effet, les résultats obtenus lors de la modélisation de l'étape d'excitation sont ceux qui fournissent les conditions initiales de l'ajustement pour l'étape de relaxation, inconnues d'après les données brutes.

Le point d'intérêt suivant pour ce travail concerne la caractérisation du bruit structuré. Les données expérimentales sont affectées par différents bruits et soumises à différents types d'erreurs de mesure. Il n'est pas possible de connaître exactement l'erreur commise sur une mesure donnée, cependant il peut s'avérer utile d'évaluer de manière plus générale l'erreur commise sur un type d'instrumentation ou pour une expérience donnée. La spectroscopie

résolue dans le temps doit donc être considérée dans ce cadre, en particulier lorsque des erreurs corrélées sont suspectées.

Un autre point de développement méthodologique concerne la gestion des différents types de bruits instrumentaux. Les méthodes chimiométriques doivent s'adapter de manière à prendre en compte l'erreur expérimentale évaluée. En fonction de la nature de l'erreur il est nécessaire d'utiliser des approches différentes. Le niveau de confiance placée sur les différentes mesures doit permettre d'éviter de considérer des informations peu fiables car trop affectées par le bruit. Les résultats obtenus dans l'étude effectuée sur la nature du bruit de données de spectroscopie résolue en temps, et plus particulièrement de données de spectroscopie d'absorption transitoire, ont permis d'adapter un schéma de pondération basé sur la structure du bruit dans le but d'améliorer les méthodes de modélisation souple. Il s'agit de la première étude de l'effet sur l'analyse de procédés de la pondération basée sur des structures de bruit corrélé.

Ces différents points clés ont permis de mieux comprendre les spécificités de la spectroscopie résolue dans le temps et donc d'améliorer l'analyse des données obtenues par cette méthode.

L'organisation de cette thèse est la suivante. Le premier chapitre est le résultat d'un travail bibliographique et de l'apprentissage pratique des algorithmes les plus utilisés en analyse de procédés. C'est l'occasion d'exposer les méthodes chimiométriques utilisées lors de l'analyse de données spectroscopiques. Les principes fondamentaux seront détaillés ici, notamment le principe de bilinéarité, le concept de rang et les méthodes d'analyse de facteurs et de modélisation de procédés les plus utilisées et qui constituent l'état de l'art dans cette branche de la chimiométrie. Ces notions sont nécessaires à la compréhension des apports proposés par le chapitre II qui regroupe les développements théoriques et méthodologiques concernant l'adaptation de ces méthodologies à l'analyse des données de différence. Des outils sont proposés pour évaluer les conséquences de cette situation mais aussi pour la gérer. Ce chapitre représente donc un des piliers méthodologiques du travail effectué et répond à une partie des interrogations rencontrées lors de l'analyse de données issues des différentes instrumentations résolues en temps présentes au Laboratoire de Spectrochimie Infrarouge et Raman (LASIR).

Les deux chapitres suivants (III et IV) regroupent les applications de ces nouveaux outils chimiométriques à l'étude par spectroscopie de deux systèmes chimiques. Le chapitre III concerne une des étapes du processus de photosynthèse chez une bactérie. Nous verrons que cet exemple est particulièrement pertinent pour illustrer les difficultés inhérentes aux données de différence et comment la méthodologie proposée au chapitre II permet de résoudre ces problèmes. Il est important de noter qu'il s'agit d'un exemple de procédé complexe, où les particularités des données de différence s'ajoutent à la présence des contributions interférentes au signal enregistré (qui ne suivent pas un modèle cinétique). Le chapitre IV est une illustration de l'intérêt des méthodes chimiométriques (basées ou non sur un modèle) dans l'étude de la réactivité par spectroscopie ultrarapide. Le cas étudié ici est la photochimie associée à la relaxation de la benzophénone. Ce cas est relié à la détection et modélisation des espèces intermédiaires à l'échelle de la femtoseconde qui ont un recouvrement spectral très intense avec des espèces voisines. Nous avons dû proposer des stratégies de validation des résultats proposés, qui combinent des outils de diagnostic chimiométrique et des expériences photochimiques supplémentaires.

Le chapitre V constitue le deuxième pilier théorique de cette thèse et concerne un autre aspect fondamental en spectroscopie (résolue en temps ou non): la gestion du bruit. Nous verrons comment une analyse rigoureuse de l'erreur de mesure permet de détecter des structures parasites de bruit corrélé, dont la cause physique sera identifiée. Un nouveau schéma de pondération, tenant compte de la corrélation du bruit dans la direction spectrale, sera ensuite proposé pour prendre en compte les informations obtenues et améliorer la fiabilité de l'analyse.

Finalement il est important de rappeler que cette thèse s'inscrit dans un cadre international puisque qu'il s'agit d'une cotutelle entre l'Université des sciences et Technologies de Lille (USTL) et l'Université de Barcelone (UB). C'est pourquoi un chapitre de synthèse est proposé en espagnol ; comme son nom l'indique il reprend l'ensemble des travaux de manière succincte en s'appuyant sur les différentes publications en relation avec cette thèse.

**Chapitre I : Analyse multivariée de données
spectroscopiques en suivi de procédés**

L'analyse de données spectroscopiques évolutives, et en particulier les données résolues en temps, vise à isoler les signaux des différents constituants d'un mélange et à obtenir leurs concentrations respectives et leurs profils spectraux. Deux approches distinctes sont disponibles, l'une se base sur l'utilisation de modèles physico-chimiques des procédés étudiés, l'autre sur une analyse de la variation du signal enregistré. Ce dernier domaine est relié aux méthodes de résolution de courbes multivariées. Ces méthodes ont pris racine dans l'analyse factorielle il y a plusieurs décennies. Les premiers travaux [1-3] dans ce domaine apparurent en 1971 lorsque Lawton et Sylvestre introduisirent pour la première fois en chimie une méthode de résolution de spectres UV/Visible basée sur l'analyse en composantes principales et sous contrainte de non négativité [4]. Cette méthode n'était applicable qu'à des systèmes ne comprenant que deux composantes. Il fallu attendre 1985 pour qu'elle soit étendue à des systèmes à trois composantes [5] et a été généralisé en 1987 à k composantes [6]. Le nombre d'applications et d'algorithmes a ensuite rapidement progressé les différentes évolutions visant à prendre en compte le maximum d'informations physico-chimiques ou statistiques. L'objectif de ce chapitre est de définir un ensemble de concepts nécessaires à la compréhension du travail de recherche exposé dans ce manuscrit. Nous commencerons par évoquer quelques concepts généraux d'analyse multivariée. Nous présenterons ensuite des outils mathématiques permettant de traiter des données multivariées puis plus spécifiquement les outils adaptés à des données spectroscopiques évolutives, c'est-à-dire les deux grandes familles de méthodes de modélisation de procédés, la modélisation rigide (*hard-modelling*), qui repose sur le modèle physicochimique et la modélisation souple (*soft-modelling*), dont l'objectif est la description de l'évolution du signal mesuré.

1. Mesures, données et outils statistiques

1.1 Mesures et données

Les techniques de mesures instrumentales actuelles sont de plus en plus complexes et génèrent des quantités de données de plus en plus importantes. La nature de ces données est dépendante du système et de l'instrument utilisé. Une mesure peut produire un simple scalaire x (1×1), un vecteur \mathbf{d} ($1 \times n$) ou une matrice \mathbf{D} ($m \times n$) selon les situations, comme le montre la Figure 1.

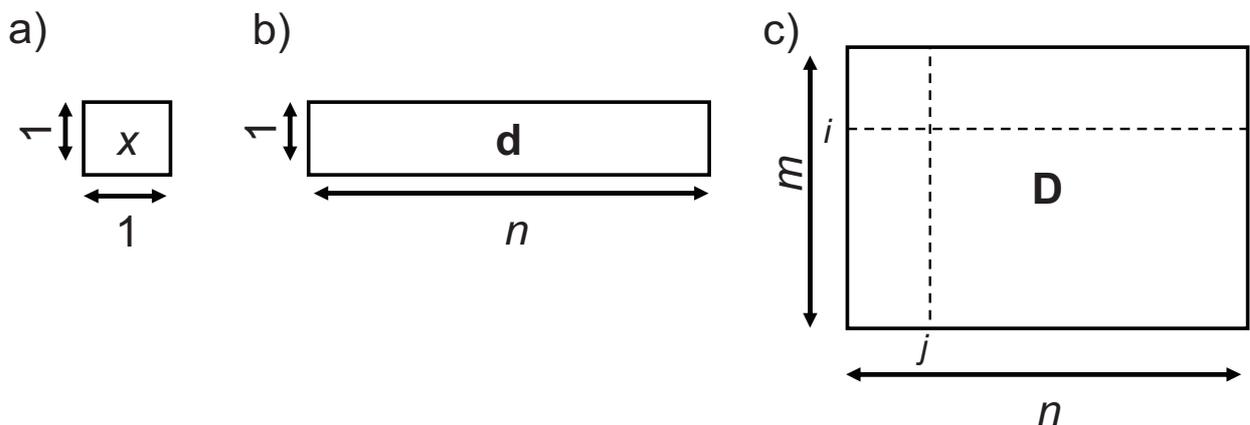


Figure 1 : Représentation schématique d'une mesure prenant la forme a) scalaire b) vectorielle c) matricielle.

Une mesure d'un spectrophotomètre (un spectre) correspondra par exemple à un vecteur \mathbf{d} de n éléments correspondant aux n différentes longueurs d'onde mesurées. Une matrice de données \mathbf{D} est facilement obtenue en regroupant les m différents spectres enregistrés durant une expérience. De nombreux exemples de ce type de matrices seront utilisés par la suite durant l'étude de données spectroscopiques résolues en temps. Les lignes de la matrice de données \mathbf{D} sont alors organisées en fonction du temps.

Au cours de ce chapitre nous utiliserons comme illustration un jeu de données spectroscopiques UV-visible \mathbf{D}_{ex} , publiés par ailleurs [7]. Il concerne le suivi d'une réaction cinétique en deux étapes dans un système fermé. Un spectre seul et l'ensemble des spectres sont représentés Figure 2. Chaque spectre constitue une mesure formée ici de 251 scalaires (pour autant de longueurs d'onde). L'ensemble de la matrice est formé de 271 spectres enregistrés à différents stades du procédé étudié.

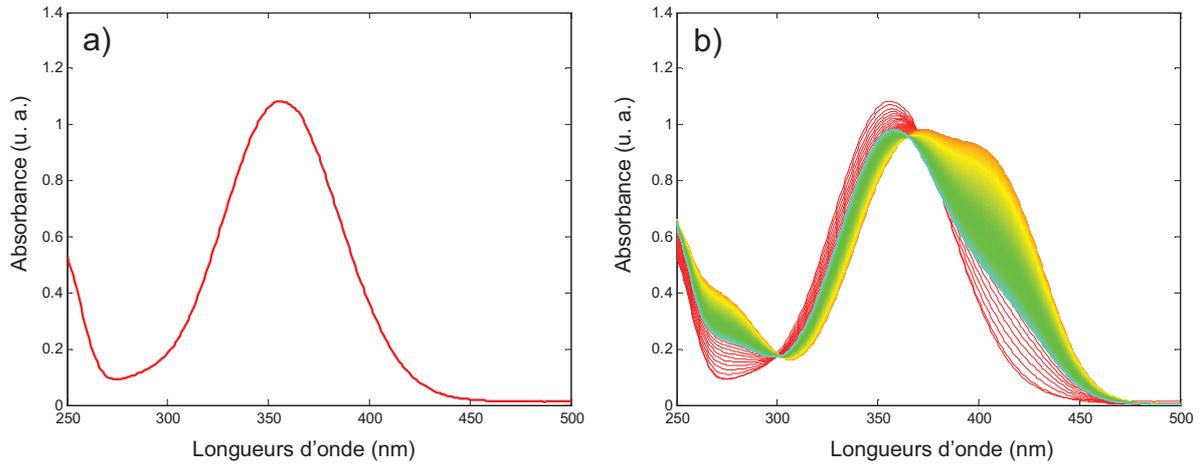


Figure 2 : Représentation a) d'un spectre et b) d'une matrice de spectres \mathbf{D}_{ex} suivant un procédé cinétique.

On admet que ces données sont régies par la loi de Beer-Lambert, (équation (1)) où l'absorbance A suit un modèle linéaire dépendant du coefficient d'extinction molaire ε de la concentration de l'analyte c , et du chemin optique parcouru par le faisceau optique l (dans une gamme de concentrations déterminée).

$$A = \varepsilon c l \quad (1)$$

Les données suivant cette règle sont dites bilinéaires ce qui va nous permettre d'utiliser différents outils mathématiques durant leur analyse. Un système est linéaire si un rapport de dépendance simple est établi entre les n mesures des variables d et c .

$$d_i = a_1 c_i + a_0 + e_i \quad i = 1, \dots, n \quad (2)$$

Ce qui peut s'écrire sous forme vectorielle comme :

$$\mathbf{d} = a_1 \mathbf{c} + a_0 + \mathbf{e} \quad (3)$$

Le terme \mathbf{e} représente l'erreur résiduelle de la décomposition. Le concept de bilinéarité est facilement extrapolable du concept de linéarité. La seule différence est que la relation linéaire doit être accomplie dans les deux directions d'une matrice bidimensionnelle \mathbf{D} ($m \times n$). Pour un système avec k constituants, chaque élément d_{ij} d'une telle matrice doit pouvoir être décomposé en termes additifs reliés à chaque constituant, qui incluent l'information sur la concentration du constituant à la ligne i , (c_i) et sur son absorptivité de la longueur d'onde j (s_{ij}), de la manière suivante :

$$d_{ij} = c_{i1} s_{1j} + c_{i2} s_{2j} + \dots + e_{ij}; \quad i = 1, \dots, m; \quad j = 1, \dots, n \quad (4)$$

Ces différents termes peuvent être réécrit sous forme vectorielle :

$$\mathbf{D} = \mathbf{c}_1 \mathbf{s}_1 + \mathbf{c}_2 \mathbf{s}_2 + \dots + \mathbf{E} ; \quad (5)$$

Ceci peut être écrit sous forme matricielle :

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (6)$$

Cette notation montre déjà l'intérêt de l'approche matricielle des données bilinéaires puisque l'on peut voir que la matrice de données \mathbf{D} est décomposée en deux matrices \mathbf{C} et \mathbf{S} qui représentent respectivement les profils de concentration et les spectres des constituants purs présents dans le mélange étudié. L'absorbance de chaque point d'un spectre d ou d'une matrice \mathbf{D} peut alors être reliée aux différents composants présents dans l'échantillon. Cette décomposition peut être représentée de deux façons comme le montre la Figure 3 pour une décomposition correspondant à deux espèces chimiques A et B.

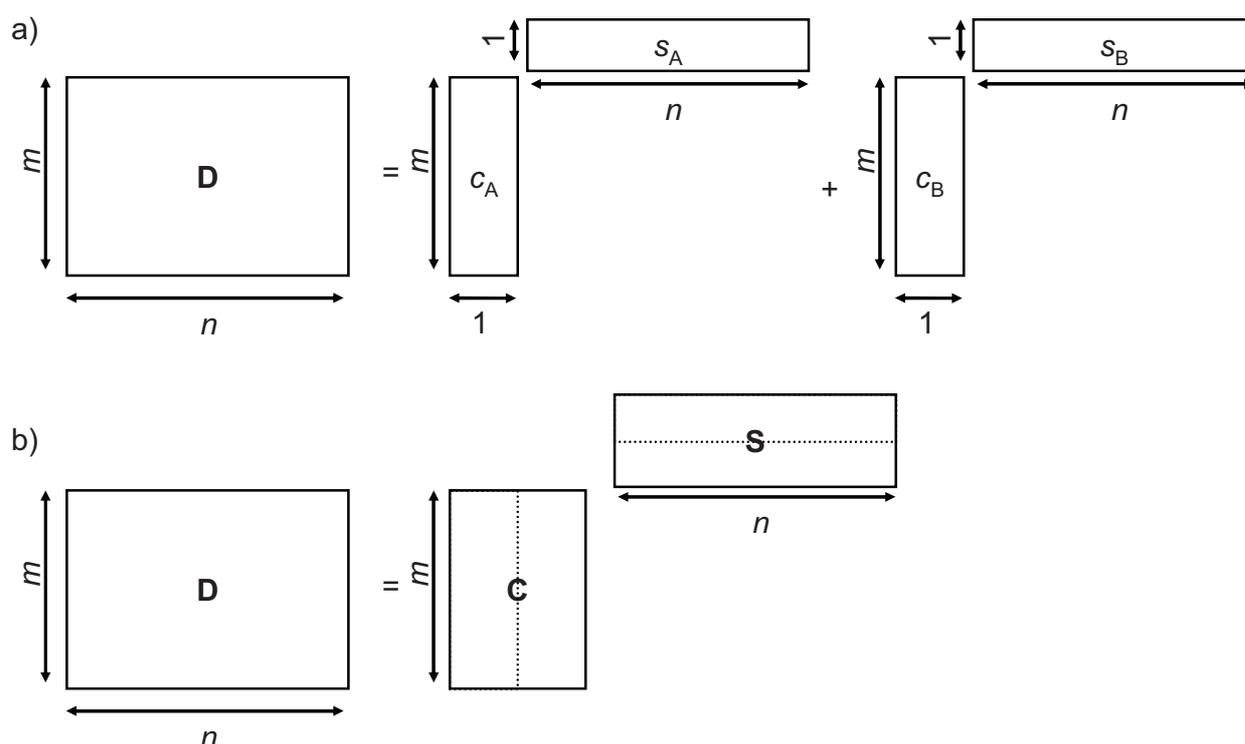


Figure 3 : Décomposition d'une matrice de données bilinéaires \mathbf{D} a) représentation par constituants b) représentation matricielle, à l'erreur \mathbf{E} près

Les données bilinéaires, courantes en sciences expérimentales, et les décompositions correspondantes sont généralement l'objet de méthodes telles que l'analyse en composantes principales et les méthodes de résolution de courbes multivariées que nous verrons respectivement aux sections 2.1, 2.3 et 2.4.

Certaines techniques, telles que la fluorescence d'excitation-émission, génèrent des données plus complexes, rassemblant p matrices et formant des structures cubiques $\underline{\mathbf{D}}$ ($m \times n \times p$) formées par des matrices de taille identique. De façon analogue à la Figure 1, la Figure 4

représente schématiquement un cube de données et la décomposition trilinéaire de celui-ci, lorsqu'elle est possible [8].

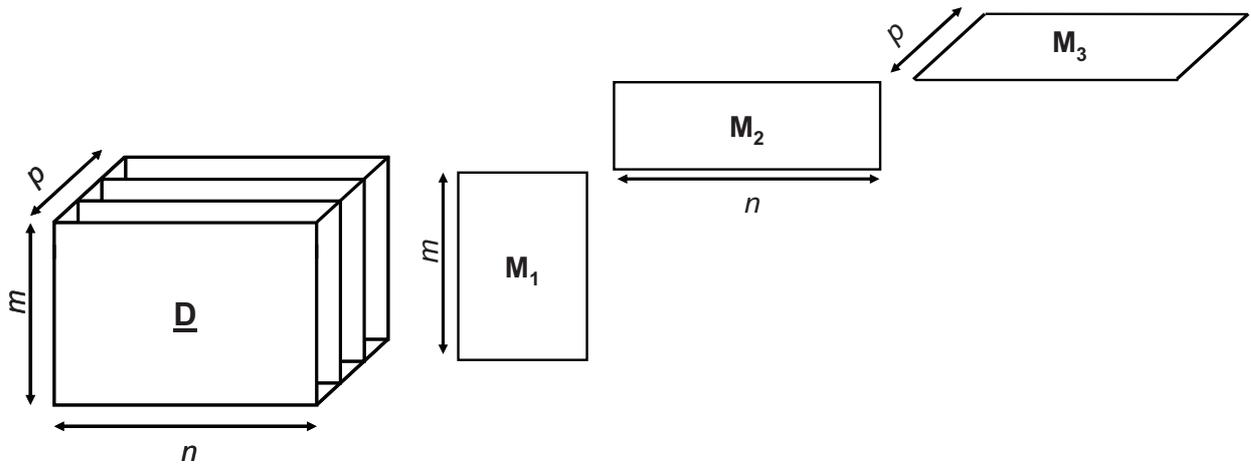


Figure 4 : Représentation schématique d'une mesure générant un cube de données et de la décomposition trilinéaire de celui-ci.

Ce type de données est décomposé en utilisant des méthodes trilinéaires du type de PARAFAC [9, 10] qui fournit trois matrices reliées aux trois directions (modes) du cube, M_1 , M_2 et M_3 . Le modèle trilinéaire demande que les matrices puissent être organisées comme un cube, c'est-à-dire qu'elles doivent avoir la même taille et ses lignes et colonnes doivent représenter le même type de variable. De plus, les profils des trois modes pour un seul constituant doivent être identiques dans chacune des matrices du cube. Cela nous donne un modèle puissant et robuste, mais très restrictif sur la structure naturelle des données.

Une dernière possibilité existe lorsque plusieurs expériences sont effectuées et que celles-ci partagent certains points communs. Les matrices de données obtenues peuvent alors être assemblées de différentes façons, il s'agit de matrices augmentées ou de structures dites '*multiset*'. L'analyse est alors effectuée de façon simultanée pour ces différentes expériences, on parle également d'analyse multi-expériences. Deux matrices de données peuvent être concaténées dans la direction des lignes (*row-wise augmented matrix*) comme le montre la Figure 5a, ce qui correspondrait, par exemple, à un suivi cinétique par deux spectroscopies différentes (pour lequel les profils de concentration sont communs). On peut également envisager l'augmentation dans la direction des colonnes (*column-wise augmented matrix*), comme le montre la Figure 5b. Cette fois il pourrait s'agir du suivi de deux cinétiques impliquant les mêmes constituants mais effectuées dans des conditions expérimentales différentes. Enfin si certaines caractéristiques sont communes dans les deux directions on peut effectuer l'augmentation dans les deux directions à la fois (*row- and column-wise augmented*

data matrices) comme sur la Figure 5c. Ces différents arrangements de matrices sont plus flexibles que la construction d'un cube de données car ils permettent l'analyse simultanée de matrices de données n'ayant pas nécessairement la même taille dans toutes les directions ou ne nécessitant pas que les profils obtenus dans la direction augmentée soient identiques en forme et ou même de différentes natures chimiques. Autrement dit l'augmentation de matrices n'impose pas de respecter la trilinearité mais seulement la bilinéarité des données.

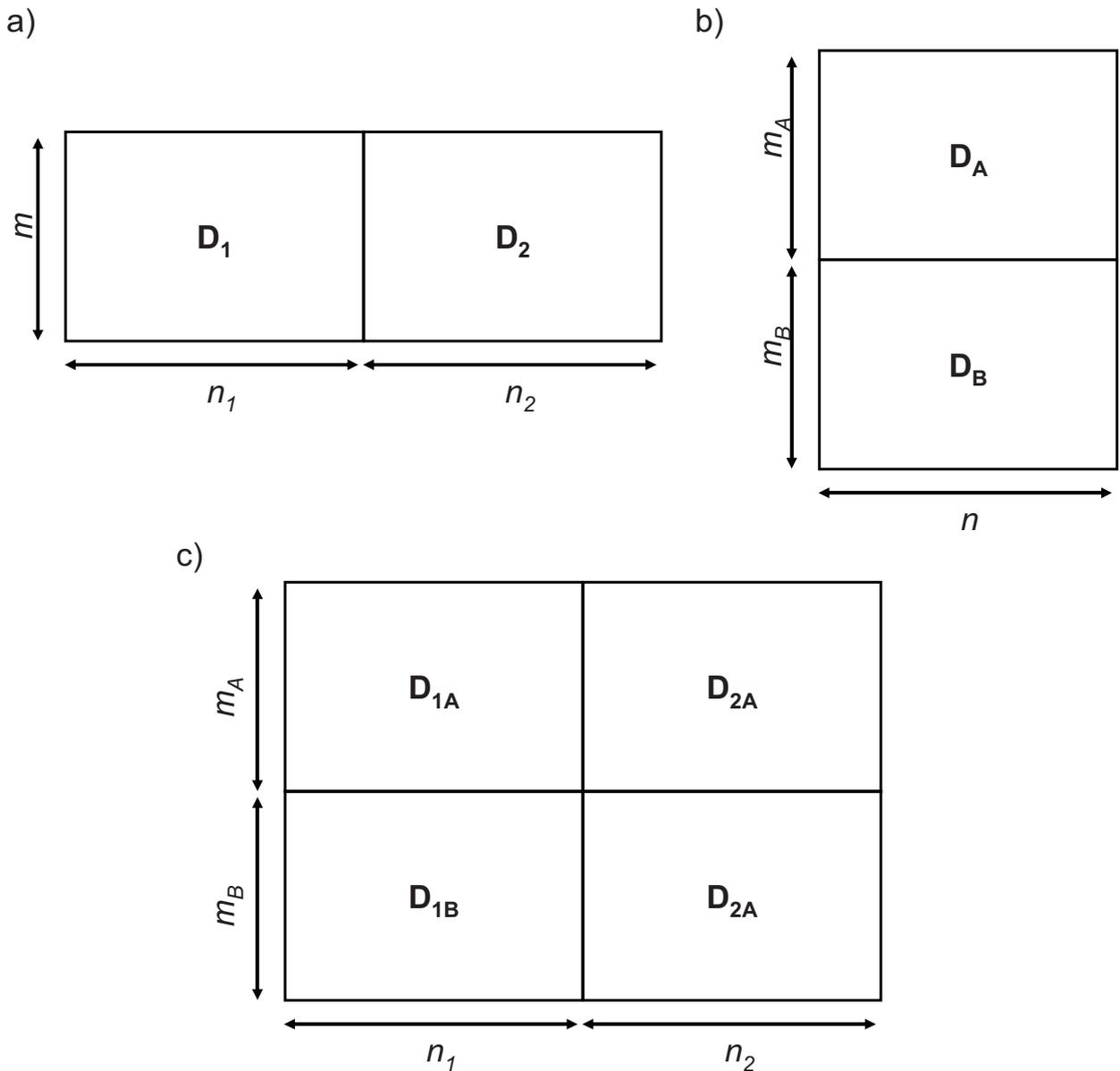


Figure 5 : Représentation schématique d'une augmentation de matrices dans a) le sens des concentrations b) le sens des spectres c) les deux sens.

Ces différents arrangements de données peuvent donc être décomposés en accord avec un modèle bilinéaire comme le montre la Figure 6. L'accumulation des données permet non seulement d'augmenter la quantité d'information utilisée mais elle apporte également un gain

qualitatif dans la résolution. Dans le cas d'une matrice augmentée selon les lignes, on obtiendrait une seule matrice de concentration (on utilise différentes techniques mais on suit toujours le même procédé) et une matrice de spectres augmentée, où chaque sous-matrice contiendrait les spectres purs propres à chaque technique instrumentale.

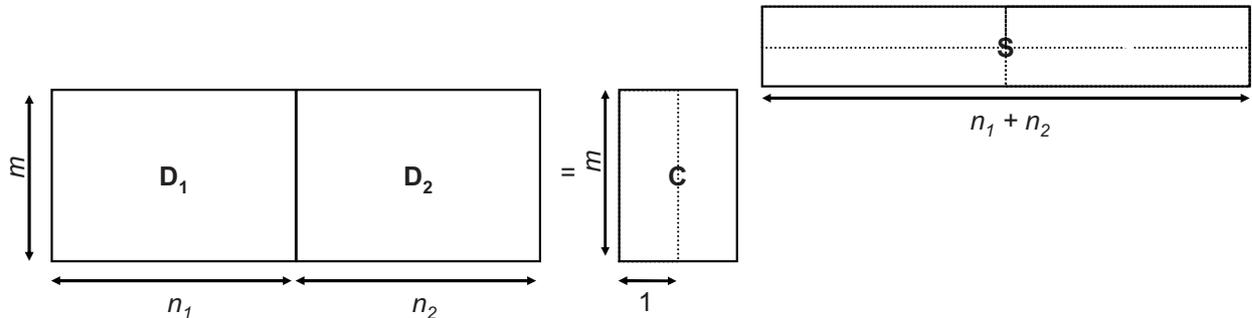


Figure 6 : Décomposition bilinéaire d'une augmentation de matrices dans la direction des lignes.

La même idée peut être développée pour l'augmentation dans le sens des colonnes, où les spectres purs obtenus des différentes matrices sont identiques, comme le montre la Figure 7, mais où différents profils de concentration sont possibles selon les sous-matrices de C.

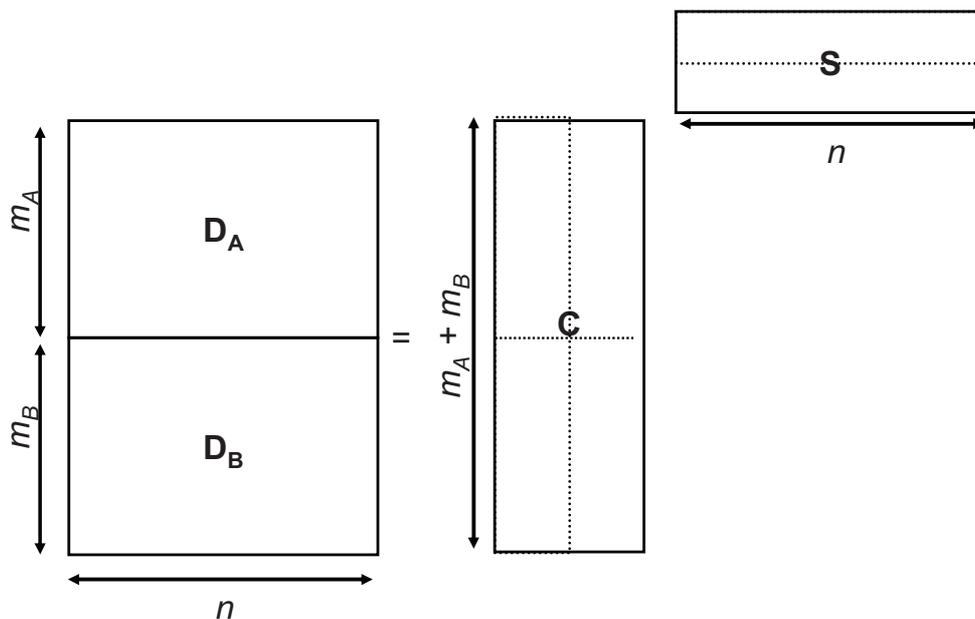


Figure 7 : Décomposition bilinéaire d'une augmentation de matrices dans la direction des colonnes.

La nature des données est donc dépendante de l'expérience réalisée et de l'instrumentation utilisée, en suivi de procédés les arrangements *multiset* sont les plus courants.

Un autre aspect est à prendre en compte lors de l'analyse de ce type de données spectroscopiques, quelque soit la méthode considérée: le bruit total affectant la mesure et la nature des différents bruits présents.

1.2 Nature et effet du bruit affectant la mesure

Toutes les techniques instrumentales sont affectées, à différents niveaux par le bruit. On considère comme un *bruit* toute contribution au signal qui n'est pas reliée à la nature chimique de l'échantillon ou du procédé suivi. Le bruit peut être vu d'un point de vue expérimental ou statistique. Une vision expérimentale du bruit permet de définir différentes sources de variation qui influencent les mesures spectroscopiques, telles que:

- le bruit de photon : la lumière interagit de façon discrète avec le détecteur. Sous une illumination parfaitement constante, le nombre de photons frappant deux photosites est différent, car la "capture" d'un photon est (dans l'état actuel des connaissances) aléatoire. Ce bruit augmente lorsque le nombre de photons augmente, néanmoins le rapport signal sur bruit varie comme la racine carrée du nombre de photons.
- le bruit thermique : les photons frappant le détecteur délogent des électrons du semi-conducteur, lesquels s'accumulent dans le puits de potentiel du photosite. Cependant certains déplacements d'électrons sont produits par l'agitation thermique dans le substrat du détecteur. Les électrons concernés s'accumulent dans les photosites. Le nombre d'électrons thermiques dans un photosite dépend de la température du détecteur, des propriétés du substrat dans la région du photosite et du temps d'intégration (le temps qui s'est écoulé depuis que le photosite a été vidé). Ceci crée un biais thermique. Mais comme la production d'électrons thermique est statistiquement aléatoire, elle génère un bruit thermique aléatoire.
- le bruit de transfert de lecture : les électrons d'un photosite sont déplacés de puits de potentiel en puits de potentiel jusqu'à ce qu'ils arrivent dans un dernier puits où ils sont "lus". Certains électrons sont perdus durant le transfert et la lecture est sujette à des bruits électroniques.
- le bruit de quantification (ou numérisation) : le nombre d'électrons captés est converti en une valeur numérique par un convertisseur analogique numérique. Le pas de ce convertisseur n'est pas d'un électron. Donc, deux nombres différents d'électrons peuvent être convertis en la même valeur.

Une approche alternative consiste à définir le bruit de mesure de manière statistique. Les différents bruits affectant les mesures multivariées sont souvent considérés indépendants et identiquement distribués, tels que des bruits gaussiens. Les structures les plus courantes sont

représentées Figure 8. Les deux premiers cas sont des exemples de bruits indépendants, c'est-à-dire que le bruit affectant un canal n'a pas de relation avec le bruit des canaux voisins. La Figure 8a montre un bruit homoscedastique, c'est-à-dire, de moyenne nulle et d'écart type égal sur les différents canaux. Ce type de bruit est le moins problématique, il peut être éliminé (ou réduit) par différentes procédures de débruitage [11, 12]. Le bruit thermique correspond à ce type de distribution. La Figure 8b est reliée au bruit hétéroscedastique, souvent présent en sciences expérimentales. La moyenne de ce bruit est également centrée sur zéro mais l'écart type est propre à chaque canal de mesure et souvent corrélé à l'intensité du signal mesuré [13] (il faut noter que seule l'amplitude du bruit est corrélée à l'intensité du signal, mais pas son signe, ce qui peut donner une contribution positive ou négative d'une façon complètement aléatoire). Ce cas correspond au bruit de photons par exemple. Ces bruits peuvent être au moins en partie supprimés soit en accumulant les données soit en utilisant par exemple la transformée de Fourier (si les fréquences du signal et du bruit sont suffisamment différentes). On notera aussi que dans ces cas les bruits des différents canaux ne sont pas corrélés les uns aux autres.

La réalité peut être très différente et mettre en jeu d'autres structures de bruits. La Figure 8c est un exemple de bruit structuré. Ici le bruit affectant des canaux voisins n'est plus indépendant mais corrélé. Il s'agit du type de bruit le plus complexe à gérer. Il peut provenir d'une large gamme d'effets comme une dérive instrumentale, un bruit de pompe optique [14] ou encore un défaut de conversion digitale [15]. Les données spectrocinétiques peuvent être affectées par ce type de bruit aussi bien dans la direction spectrale que dans la direction temporelle.

L'influence globale de ces erreurs doit être évaluée. En effet une analyse correcte des données implique une connaissance de la nature de l'erreur l'affectant, qu'elle soit ou non corrélée [16]. Ces différentes structures compliquent l'analyse des données enregistrées. En effet, on a vu que les données spectroscopiques suivent une loi bilinéaire. En l'absence de bruit il serait donc possible d'utiliser directement les outils de l'algèbre linéaire pour décomposer les matrices de données en matrices de profils de concentration et de spectre purs. L'analyse de données expérimentales ne le permet pas puisque la présence de bruit perturbe les résultats obtenus. Ainsi des outils plus adaptés ont été conçus capables de séparer le bruit du modèle obtenu ou, pour des cas plus difficiles, de minimiser son effet sur les résultats en incluant l'information sur sa structure lors de l'analyse de données.

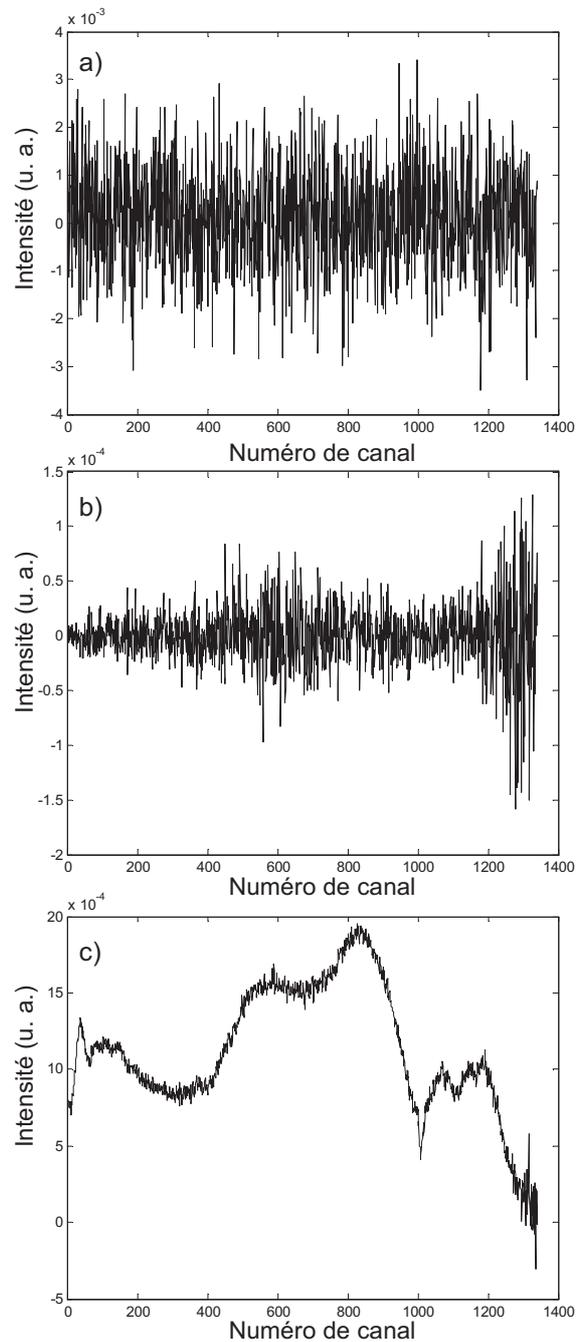


Figure 8 : Représentation de trois types de bruit a) indépendant homoscédastique b) indépendant hétéroscédastique et c) structure de bruit.

Nous allons présenter dans la section suivante les algorithmes de référence pour travailler sur des données bilinéaires. L'idée commune de ces méthodes est que la part significative (du point de vue chimique) du signal enregistré représente une variance plus élevée que celle attribuée au bruit des données.

2. Analyse exploratoire de matrices de données de suivi de procédé

2.1 Analyse en Composantes Principales (ACP)

L'objectif de l'analyse en composantes principales (ACP) [13, 17, 18] est d'exprimer la structure de variation d'une ensemble de variables descriptives d'une matrice de données (les longueurs d'onde des spectres obtenus d'un procédé, par exemple) en utilisant un nombre restreint de variables définies comme des combinaisons linéaires des variables originales. Ces combinaisons linéaires sont appelées composantes principales (CP). La variance totale d'une matrice de données à n variables peut souvent être condensée autour de k composantes principales (avec $k < n$); ou encore k directions de variance maximale. Les composantes obtenues sont classées par importance de la variance expliquée, ce qui veut dire que la première CP expliquera davantage de variance que la seconde qui en expliquera elle même plus que la suivante, et ainsi de suite.

L'ACP impose que les composantes principales soient orthogonales entre elles. La même information ne peut donc pas être partagée entre deux CPs. Cette condition assure l'unicité de la solution. On verra par la suite que cette condition pose des problèmes lors de l'interprétation chimique des résultats car les CPs obtenues sont des facteurs abstraits.

L'ACP peut être définie aussi comme une méthode d'extraction des variations systématiques d'un jeu de données [19]. L'ACP cherche à reconstruire la structure de variance/covariance des données dans un modèle représentant les variations significatives et où le bruit aléatoire (supposé indépendant et identiquement distribué) est considéré comme une erreur.

Les données \mathbf{D} se décomposent donc en un produit (d'Hadamard) de deux matrices, \mathbf{T} ($m \times k$) la matrice des coordonnées factorielles (appelés *scores*) et \mathbf{P} ($k \times n$) la matrice des vecteurs propres (appelés *loadings*), plus une matrice de résidus \mathbf{E} .

$$\mathbf{D} = \mathbf{TP}^T + \mathbf{E} \quad (7)$$

La dimension k du nouvel espace est déterminée par le rang de la matrice \mathbf{D} , c'est-à-dire pour le chimiste par le nombre d'espèces détectables. La Figure 9 correspond à la représentation graphique usuelle de cette décomposition bilinéaire.

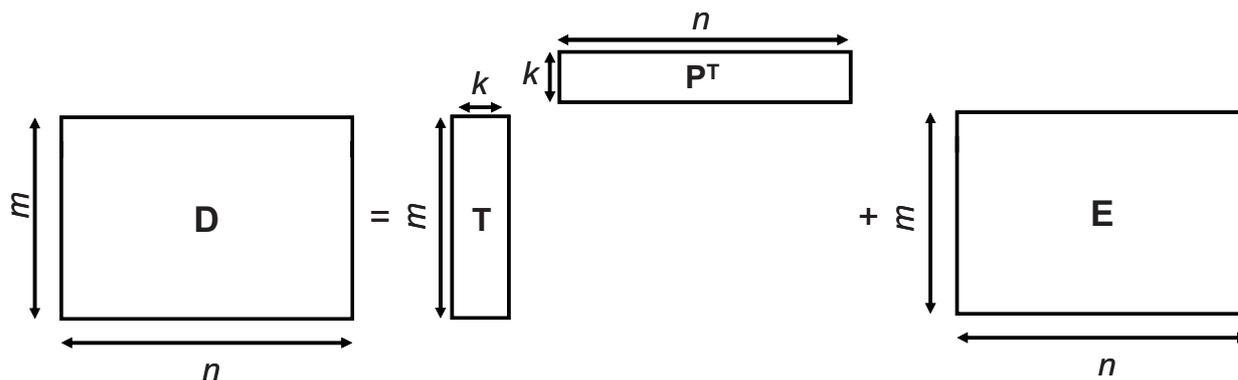


Figure 9 : Décomposition bilinéaire par ACP d'une matrice de données D .

L'algorithme le plus classique pour calculer l'ACP est NIPALS [13]. C'est cet algorithme que nous venons implicitement de présenter en utilisant les matrices de *scores* et de *loadings*. L'algorithme SVD (*Singular Value Decomposition*) [20] est plus souvent utilisé pour l'ACP et décompose la matrice D selon l'équation suivante :

$$D = U \Sigma V^T \quad (8)$$

Ici U et V sont deux matrices orthonormales telles que $U^T U = I$ et $V^T V = I$. La diagonale de la matrice Σ regroupe les valeurs singulières de D . La Figure 10 permet de mieux visualiser cette décomposition.

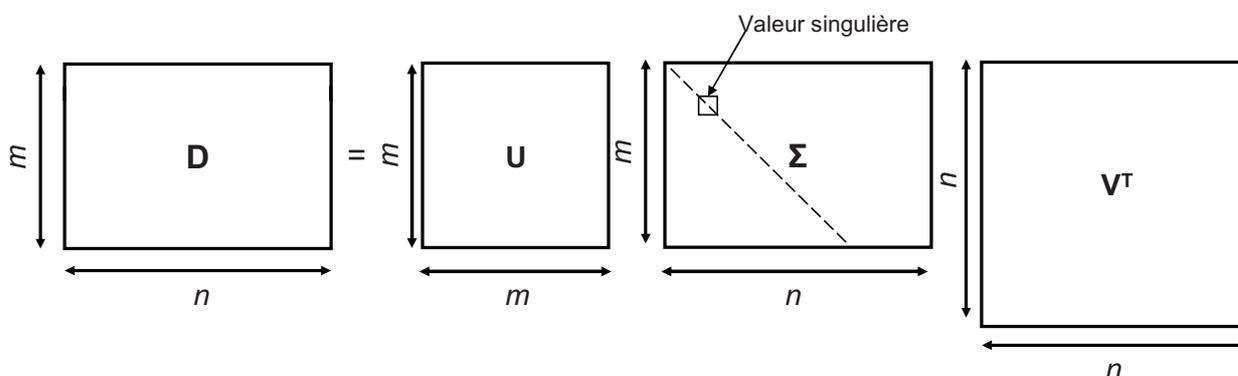


Figure 10: Principe de la décomposition en valeurs singulières de D .

La connexion entre l'algorithme SVD et l'algorithme NIPALS est expliqué par la Figure 11. Cette figure permet de voir que lorsque seulement k facteurs significatifs sont considérés, le produit des matrices U et Σ de la SVD correspond à la matrice de *scores* T de l'ACP. La matrice V est équivalente à la matrice des *loadings* P .

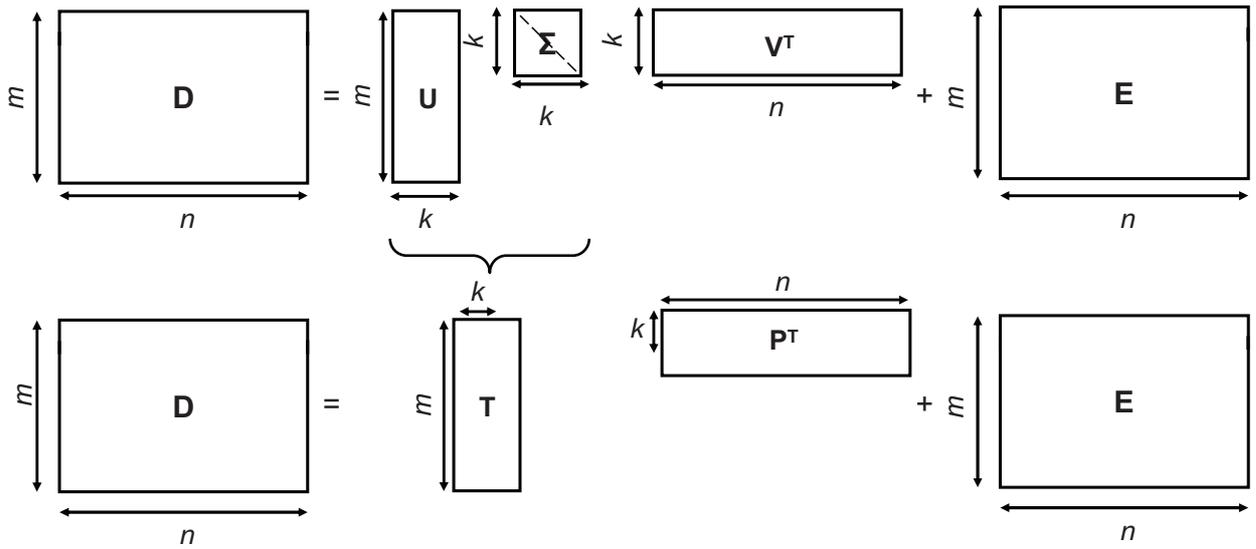


Figure 11 : Comparaison de la SVD et de l'algorithme NIPALS .

Dans le cas de l'exemple D_{ex} , le procédé de trois contributions, la décomposition bilinéaire par SVD prend la forme décrite par la Figure 12, où u_1 , u_2 et u_3 représentent les trois premiers *scores* calculés par la SVD et v_1 , v_2 et v_3 les *loadings* correspondants. La représentation graphique de ces profils 'abstrait' fournit une information qualitative sur le procédé.

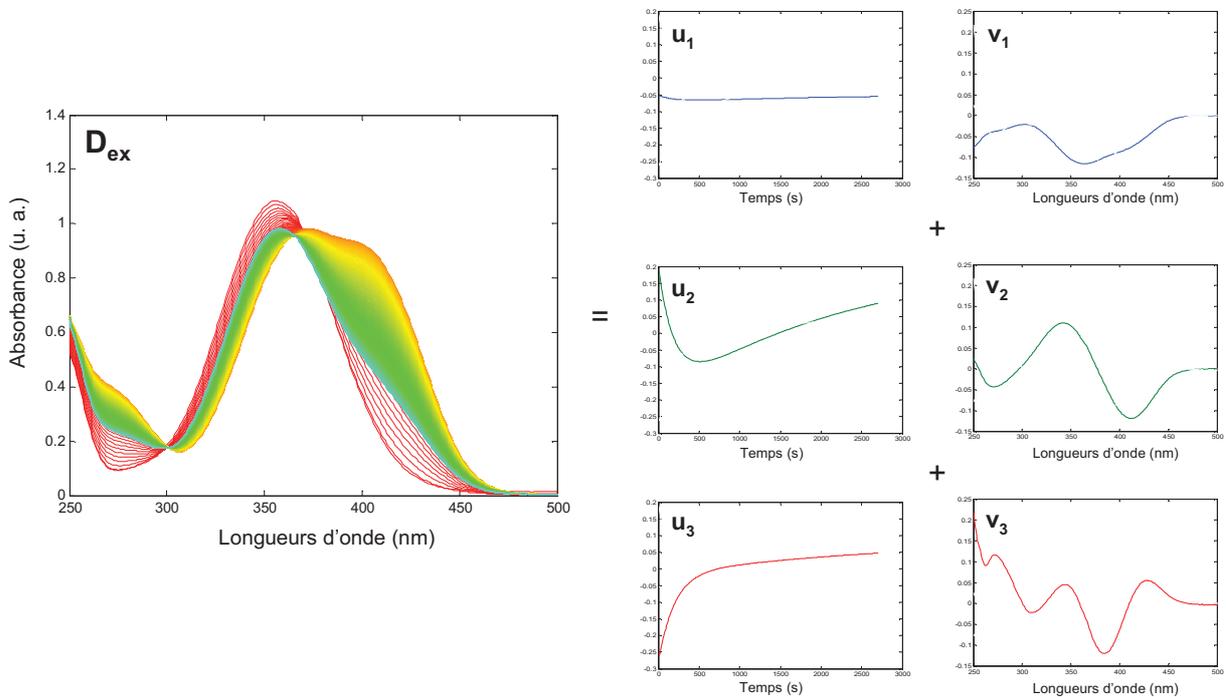


Figure 12 : Décomposition par SVD de la matrice D_{ex} en trois paires de *scores* et *loadings*.

Cette décomposition peut s'écrire de façon matricielle comme le produit des matrices \mathbf{U}_{ex} et \mathbf{V}_{ex} (aux facteurs Σ près) et donc il est possible de les représenter ensemble comme le montre la Figure 13.

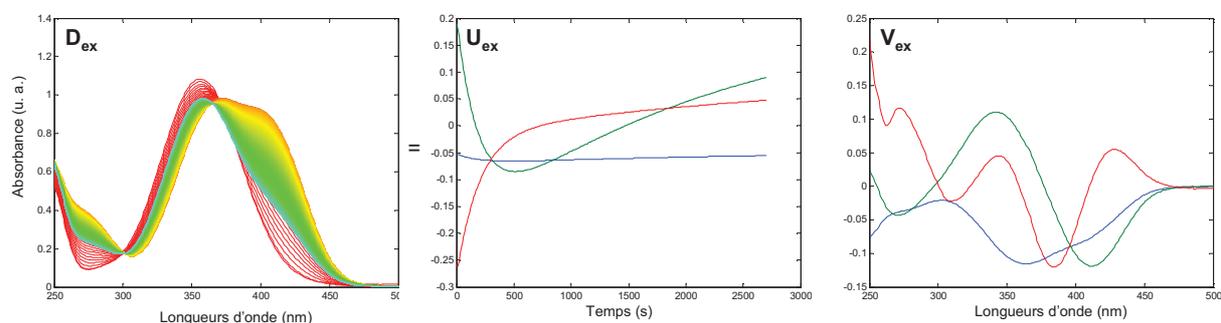


Figure 13: Décomposition par SVD de la matrice \mathbf{D}_{ex} en scores \mathbf{U}_{ex} et loadings \mathbf{V}_{ex} .

Les *loadings* obtenus décrivent correctement les données puisque chacune des zones spectrales d'intérêt est reproduite par l'un ou l'autre. Le *loading* bleu à la forme d'un spectre moyen inversé et est ainsi caractérisé par les bandes à 357 nm et 420 nm et le vert et le rouge par des bandes à 350, 420, 275 nm,..., c'est-à-dire par des régions spectrales où on observe des variations importantes au long du procédé. Les *scores*, ou profils de concentration abstraits, montrent aussi des allures caractéristiques de profils cinétiques, avec décroissances et croissances typiques. L'information qualitative fournie par la SVD sur le procédé est indubitable, mais il faut dire que, à cause de l'orthogonalité des composantes principales, il n'est pas possible d'établir une correspondance directe entre ces profils abstraits et la vraie évolution des profils de concentration et des spectres des espèces chimiques. Pour ceci, il faut utiliser aussi des méthodes basées sur la décomposition bilinéaire optimale des données, mais qui imposent des conditions plus flexibles et adaptées à la nature chimique du système sur les profils obtenus. Ces méthodes seront introduites aux sections 3 et 4 de ce chapitre.

Dans le contexte d'analyse de procédés, l'ACP est surtout utilisée pour répondre une importante question préliminaire: combien de contributions sont présentes dans le procédé étudié? C'est sur ce point que se concentrent les sections suivantes.

2.2 Rang de matrices expérimentales

La première information d'intérêt, lors d'une analyse de données spectroscopiques, est le nombre d'espèces chimiques impliquées dans la réaction étudiée. On parle alors de rang chimique par analogie au rang mathématique. En algèbre linéaire le rang d'une matrice ou

d'un ensemble de vecteurs est la dimension du sous-espace vectoriel engendré par cet ensemble. Autrement dit le rang représente le nombre de spectres "purs" (vecteurs propres) ou de profils de concentration nécessaires pour expliquer l'ensemble des mesures, chaque enregistrement n'étant qu'une combinaison linéaire de ces vecteurs propres.

Dans le cas idéal, les données enregistrées ne sont pas affectées par le bruit. Le rang chimique correspond alors à la définition du rang mathématique. Il suffit donc d'utiliser l'algèbre linéaire et de calculer le rang "mathématique" de la matrice de spectres pour connaître le nombre d'espèces présentes. Evidemment ce cas est éloigné de la réalité. Le bruit est toujours présent dans des données expérimentales. Prenons l'exemple d'une matrice \mathbf{D} constituée de 30 spectres de 100 longueurs d'onde qui appartiennent à des échantillons où seulement une espèce chimique est présente. En l'absence de bruit et selon la loi de Beer-Lambert, \mathbf{D} peut être décrite par une seule contribution constituée du produit de la concentration de l'échantillon par le spectre du produit pur. Le rang chimique de \mathbf{D} est égal à 1, là où le rang mathématique est souvent bien supérieur (au maximum égal à 30) à cause du bruit expérimental. Les données enregistrées peuvent alors être décrites par l'équation (9).

$$\mathbf{D} = \mathbf{c}\mathbf{s}^T + \mathbf{E} \quad (9)$$

Le vecteur \mathbf{c} (30×1) contient les concentrations de l'espèce chimique dans les 30 échantillons et \mathbf{s} le spectre pur de l'espèce absorbante. \mathbf{E} est la matrice associée au bruit. Cette équation peut être résumée sous la forme:

$$\mathbf{D} = \hat{\mathbf{D}} + \mathbf{E} \quad (10)$$

La matrice de données reconstruite $\hat{\mathbf{D}}$ a un rang égal à 1 puisqu'elle ne représente que la variation systématique de \mathbf{D} , ce qui définit le pseudo-rang de \mathbf{D} . Le rang chimique correspond à ce pseudo-rang [21] et se définit finalement comme le nombre de sources faisant varier le signal chimique et distinguables de la contribution du bruit. Il reste que l'estimation correcte du nombre d'espèces est délicate et conditionne l'efficacité de toutes méthodes d'analyse multivariée. L'intérêt de l'ACP réside dans sa capacité à séparer et quantifier les différentes sources orthogonales de variance dans un jeu de données, représentées par les différentes composantes principales. C'est le pourcentage de variance capturée ou des paramètres qui y sont reliés, tels que les valeurs propres ou les valeurs singulières, qui permettront d'attribuer une CP à une variation significative du signal ou au bruit [20].

Estimation du rang chimique

L'estimation du rang d'une matrice de données spectroscopiques peut être envisagée sous plusieurs angles. L'approche la plus simple repose sur une connaissance *a priori* du système étudié. Néanmoins les informations concernant le nombre d'espèces impliquées et/ou d'étapes réactionnelles sont souvent soit manquantes ou incomplètes soit justement l'information recherchée. Un grand nombre de techniques [21-28] a donc été mis au point pour estimer le rang à partir des données expérimentales uniquement. La méthode la plus populaire repose sur l'Analyse en Composantes Principales et sur l'hypothèse que les signaux d'intérêt provenant des différentes espèces chimiques sont des sources de variance plus importantes que les bruits.

L'étude de la matrice Σ provenant de la SVD offre des informations sur le rang chimique de la matrice de données \mathbf{D} . Il faut dire que les valeurs singulières se définissent comme la racine carrée des valeurs propres et celles-ci ont un rapport direct avec le pourcentage de variance capturée par une composante principale [20]. Ceci explique pourquoi l'étude des valeurs singulières est une des méthodes les plus utilisées pour estimer le nombre d'espèces chimiques d'un procédé. On considère que les valeurs singulières associées à une contribution spectroscopique doivent être plus importantes que celles associées au bruit expérimental [21]. Il est alors possible d'estimer le nombre d'espèces chimiques contribuant significativement au signal enregistré.

Dans le cas de notre exemple \mathbf{D}_{ex} , le système étudié peut se résumer par deux réactions:



Le nombre de contributions impliquées est donc égal à trois. Le Tableau 1 présente les premières valeurs singulières (les premières valeurs de la diagonale de la matrice Σ) et le pourcentage de variance expliquée par les CP de la matrice \mathbf{D}_{ex} . On constate que les trois premières valeurs singulières sont nettement supérieures aux suivantes et cela coïncide avec les conclusions qu'on peut tirer du pourcentage de variance capturée. Le rang chimique est donc bien égal à trois. La distinction entre les valeurs singulières associées au signal et celles associées au bruit peut se faire soit en inspectant les valeurs obtenues directement (ou graphiquement) soit en utilisant des critères statistiques de comparaison [21].

Composante principale	Valeurs singulières	% de variance capturée
1	136,7067	98,63
2	16,0341	1,36
3	1,8552	0,02
4	0,0923	0
5	0,0252	0
6	0,0085	0
7	0,0068	0
8	0,006	0
9	0,0045	0
...

Tableau 1 : Valeurs singulières et pourcentage de variance capturée par les CPs de la matrice \mathbf{D}_{ex} . Les valeurs en gras correspondent aux valeurs singulières associées au signal.

Déficiences de rang

Une déficience de rang correspond à une situation où le rang estimé est inférieur au nombre d'espèces chimiques présentes dans le mélange étudié. Cette déficience peut se produire pour différentes raisons [29, 30]. Les mélanges en réaction peuvent être dirigés par un nombre clairement plus faible de réactions que d'espèces présentes. Le centrage des données peut aussi induire une déficience de rang si la ligne de base des données brutes évolue durant la mesure. L'existence de dépendances linéaires entre les spectres purs des espèces ou les profils de concentrations induit également une déficience de rang.

Amrhein *et al.* ont beaucoup travaillé sur le phénomène de déficience de rang dans le contexte de suivi de procédés [21]. Ils ont conclu que pour des systèmes chimiques qui incluent plusieurs réactions simultanément, l'estimation du rang pour des systèmes fermés peut être définie par une équation en fonction du nombre d'espèces n_E et de réactions n_R :

$$\text{rang}(\mathbf{D}) = \min(n_R + 1, n_E) \quad (13)$$

Différentes solutions ont été proposées pour lever les déficiences de rang ; parmi celles-ci nous nous concentrerons sur l'augmentation de matrices [30, 31]. On a vu que l'augmentation de matrices consiste à regrouper plusieurs matrices provenant de systèmes différents et ayant en commun plusieurs ou l'ensemble de leurs espèces chimiques. L'analyse simultanée de différents mélanges des mêmes composés dans des conditions physico-

chimiques différentes permet d'extraire des informations fiables sur la nature des différents procédés impliqués. Cet ensemble d'informations permet d'éliminer des relations linéaires entre profils de concentration ou spectres qui peuvent être présentes dans une matrice particulière mais qui ne s'étendent pas à toute la matrice augmentée. Il est important de noter que la cause de la déficience de rang dans une matrice de données peut être reliée soit à la direction des concentrations soit à celle des spectres. La direction provoquant la déficience de rang doit être augmentée avec de l'information complémentaire pour supprimer cet effet. Donc si la déficience de rang est reliée aux profils de concentration (une évolution parallèle de plusieurs espèces par exemples), on utilisera une matrice augmenté par colonnes pour la supprimer et si elle est associée aux spectres, une matrice augmentée par lignes.

L'augmentation des matrices s'avère aussi utile pour des déficiences de rang ou estimations peu fiables du rang pour des spectres suivis dans le temps et affectés par des niveaux de bruit élevés. Le rang de ces derniers est généralement difficile à estimer à cause de l'influence du bruit, l'analyse simultanée de plusieurs séries de mesures différentes permet d'exalter la variance associée aux espèces chimiques. Il est alors plus aisé de les distinguer de la variance associée au bruit.

Comme on l'a dit précédemment, les solutions proposées par l'ACP sont souvent difficiles à relier à la chimie du problème posé. En effet les *scores* et *loadings* obtenus sont des solutions mathématiques qui ne correspondent pas forcément à la réalité du procédé. On obtient par exemple des *scores* négatifs alors qu'ils devraient être assimilables à des concentrations. Il est donc nécessaire de corriger ces résultats en introduisant soit de "l'information chimique" soit des restrictions interdisant certaines solutions. Différentes méthodes ont été mises au point pour tendre vers l'estimation de facteurs acceptables du point de vue de la chimie du problème. Nous proposons de détailler deux d'entre elles, Evolving Factor Analysis et SIMPLISMA, qui permettent d'estimer la matrice des profils de concentration \mathbf{C} dans le premier cas ou d'estimer \mathbf{C} et la matrice des spectres purs \mathbf{S} dans le second cas.

2.3 EFA

Une première approche pour estimer la matrice des profils de concentration \mathbf{C} , en accord avec le modèle bilinéaire présentée dans le paragraphe 1.1., de données

spectroscopiques résolues en temps est l'Analyse Factorielle Evolutive (EFA, *Evolving Factor Analysis*) [32-35]. Il s'agit d'une méthode basée sur le calcul successif des valeurs singulières (ou valeurs propres) de sous matrices de la matrice de données \mathbf{D} . Cette approche tire bénéfice de l'ordre intrinsèque des spectres dans les matrices qui représentent le suivi d'un procédé. La première étape consiste à évaluer l'évolution des valeurs propres de \mathbf{D} dans le sens direct de l'évolution du procédé (*forward EFA*) en calculant successivement les valeurs propres d'une sous-matrice composée d'abord des deux premières lignes de \mathbf{D} (c'est-à-dire des deux premiers temps), puis des trois premières lignes et ainsi de suite jusqu'à les n lignes de la matrice. Ceci revient à effectuer des ACP successives comme le montre la Figure 14a. Cela permet de suivre l'émergence des valeurs propres au long du procédé. Lorsqu'une nouvelle espèce absorbante commence à apparaître, une valeur propre ressort du groupe des valeurs propres de l'erreur, et augmente de façon proportionnelle à la contribution de cette espèce dans le jeu de données. Les valeurs propres ne peuvent qu'augmenter au fur et à mesure de l'analyse puisque la variance totale augmente à chaque ajout de ligne. En somme, le *forward EFA* indique où se situe l'émergence des nouvelles espèces au long d'un procédé. La même procédure est suivie dans le sens inverse (*backward EFA*) de l'évolution selon le principe exposé par la Figure 14b Ceci permet d'observer non pas les apparitions des différentes contributions mais cette fois leurs disparitions

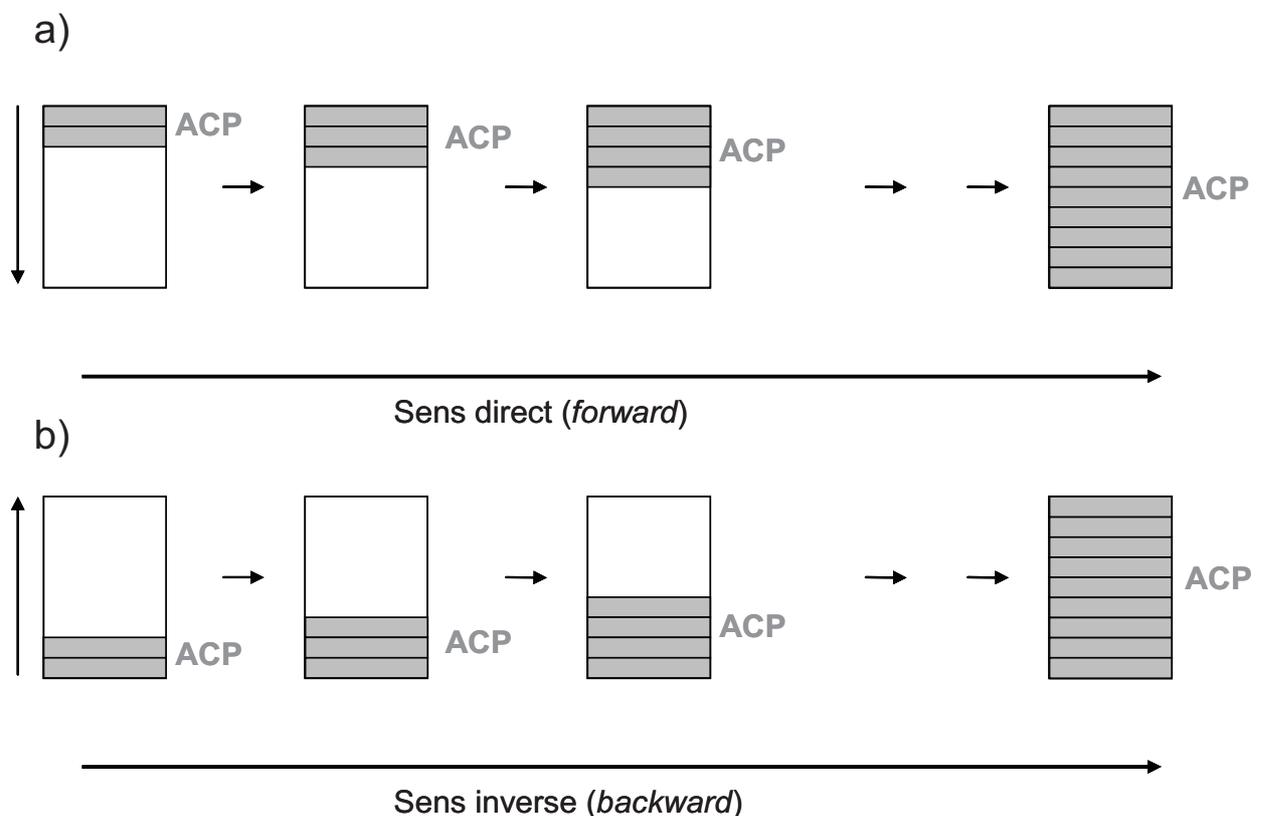


Figure 14 : Principe de l'Analyse Factorielle Evolutive a) en sens direct b) en sens inverse.

Le résultat de l'EFA *forward*, présenté Figure 15a, permet d'abord d'évaluer le rang global de D_{ex} . En effet les trois contributions significatives se détachent des autres valeurs propres représentant le bruit. Plus important, il est possible d'estimer à quel moment apparaissent les différentes espèces. Dans notre exemple, trois contributions sont détectables, la première, en bleu, est présente dès le début, la seconde, en vert, apparaît aux alentours de la ligne 6 de la matrice D , enfin la dernière, en rouge, n'apparaît qu'à la ligne 80. Le résultat de l'EFA *backward* est présentée Figure 15b. Seules trois contributions sont visibles, comme dans l'EFA *forward*. On obtient cette fois des informations sur les disparitions des espèces à la ligne 25 pour l'espèce bleue et à la ligne 248 pour l'espèce verte. Enfin la disparition de la troisième contribution (en rouge) n'est pas visible, ce qui veut dire qu'elle reste à la fin du procédé.

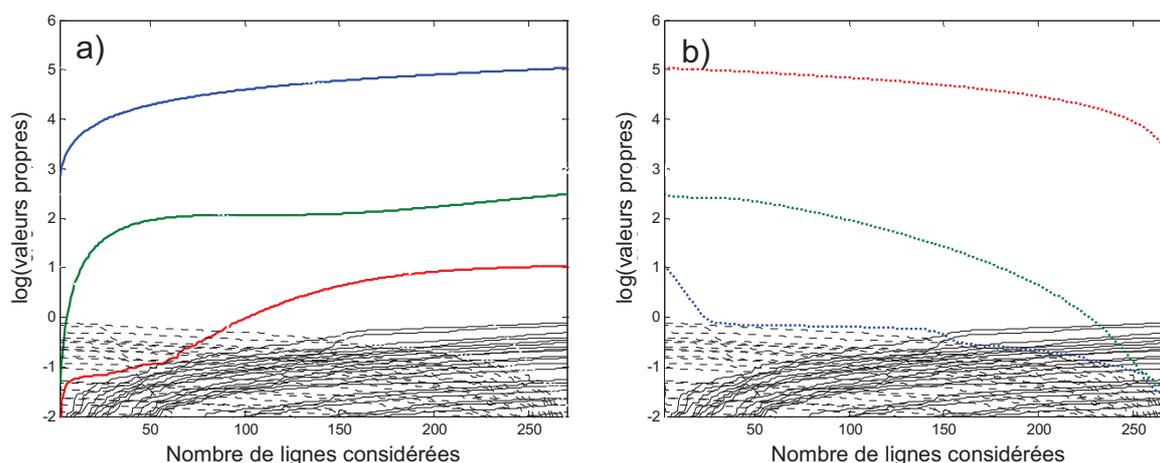


Figure 15 : Analyse Factorielle Evolutive de D_{ex} en sens a) direct b) inverse.

Partant de l'hypothèse que la première espèce à apparaître est en suivi de procédés la première à disparaître, l'analyse simultanée des résultats de l'EFA *forward* et *backward* permet de construire une estimation des profils de concentration des trois contributions détectées. Dans ce but, on utilise un graphique EFA qui contient les résultats d'EFA *forward* et *backward*, c'est-à-dire toute l'information disponible sur le procédé; comme le montre la Figure 16. La ligne noire en trait gras représente le niveau de bruit estimé on considère qu'au dessous de cette limite les contributions observées ne sont plus significatives. On obtient l'estimation des profils de concentration en unissant le premier profil qui apparaît avec le premier qui disparaît, qui selon notre hypothèse représente l'apparition et la disparition de la même espèce. Les profils suivants sont obtenus de manière similaire toujours en supposant

que la $n^{\text{ème}}$ espèces apparaissant sera aussi la $n^{\text{ème}}$ à disparaître. L'exemple de la seconde contribution en vert est particulièrement parlant. S'il n'y a pas croisement de lignes (au début et à la fin du procédé), on utilise uniquement les lignes de plus forte intensité, qui représentent mieux soit la disparition (pointillée bleue) ou soit l'apparition (rouge pleine) des espèces.

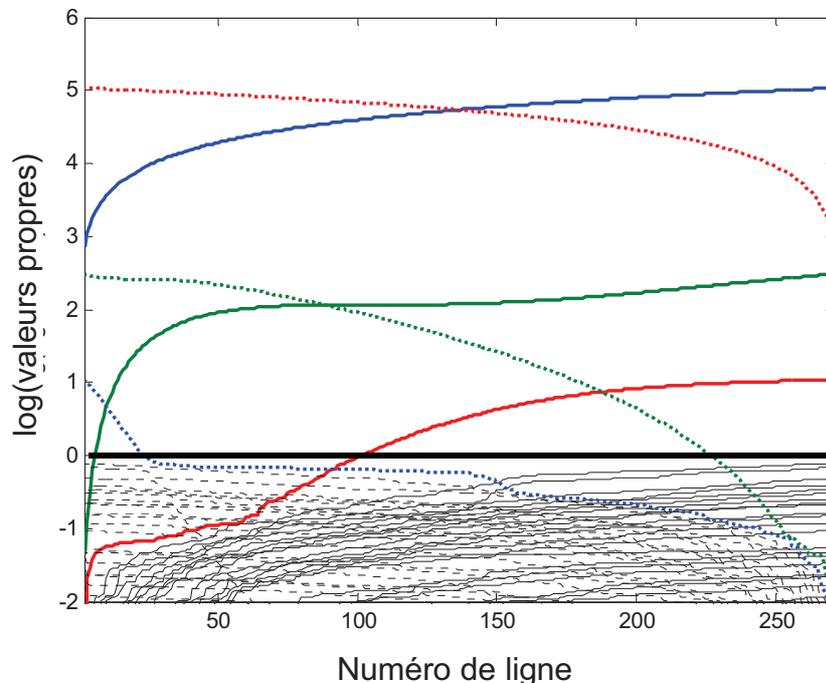


Figure 16 : Graphique contenant les informations de l'EFA *forward* et EFA *backward* de \mathbf{D}_{ex} , la ligne horizontale en trait noir épais représente le niveau de bruit en dessous duquel on ne considère plus les contributions détectées..

Le résultat final de l'estimation de profils de concentration est présenté Figure 17. En outre des informations sur la sélectivité sont maintenant disponibles: la première espèce est visiblement la seule présente au premier temps, la troisième est elle aussi la seule présente dans les dernier temps, on peut même estimer que avant la ligne 30 la dernière espèce est absente. Ces informations permettent d'identifier les régions sélectives du système et les régions où le rang local est inférieur au rang total. Cette information peut s'avérer utile lors de l'interprétation des résultats ou peut être introduite dans la résolution lors de l'analyse par l'algorithme MCR-ALS qui sera détaillé au paragraphe 3.2. Il faut noter que dans les figures précédentes, on utilise normalement comme axe x la variable de contrôle du procédé, c'est-à-dire temps, pH,... Cela s'avère important lorsque les mesures du procédé ne sont pas faites de façon équidistante au long du procédé (on a préservé ici le numéro de ligne comme axe par l'équidistance de l'exemple).

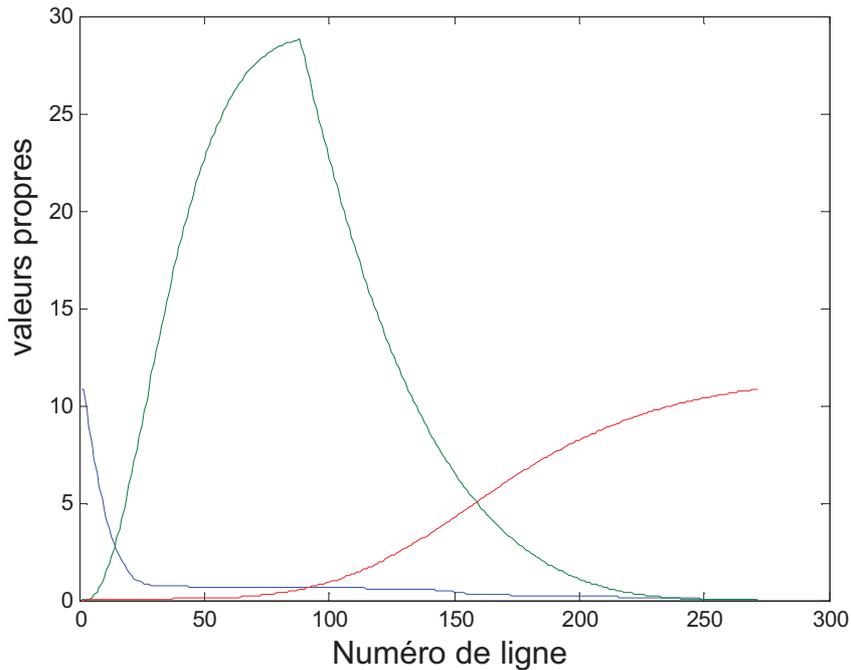


Figure 17 : Profils de concentration C_{ex} de D_{ex} estimés par EFA

L'EFA est une méthode à l'origine de nombreuses variantes, parmi celles-ci citons l'Analyse Factorielle Evolutive à Fenêtre Mobile de Taille Fixée (FSMWEFA, *Fixed Size Moving Window Evolving Factor Analysis*) [36-38] et l'Analyse Factorielle Evolutive Exhaustive (E-EFA, *Exhaustive Evolving Factor Analysis*) [39]. Cette dernière méthode cherche à vérifier que l'évolution des espèces est séquentielle. On notera plus particulièrement l'existence d'une version permettant de travailler simultanément avec plusieurs matrices ; ce qui permet gérer certains problèmes de déficience de rang.

En effet l'analyse classique par EFA de données déficientes de rang ne fournit pas d'informations fiables. Certaines contributions peuvent être noyées dans le bruit expérimental ou mêlées à une autre contribution. Or les déficiences de rang peuvent être levées par augmentation de matrices dans la direction de la déficience [31], par exemple dans la direction des concentrations lorsque deux réactions coévoluent parallèlement. Une méthode basée sur le principe de l'augmentation de matrices et l'EFA a été proposée pour lever la déficience de rang et par conséquent permettre la construction d'estimation des profils de concentration fiables [40]. Il s'agit de l'Analyse Factorielle Evolutive pour matrices Déficientes de Rang (EFARD, *Evolving Factor Analysis for Rank Deficient matrix*) [31, 40]. Nous évoquerons ce point dans le contexte concret des données analysées dans les chapitres suivants.

L'EFA, mise au point pour l'analyse de données chromatographiques [33] et l'étude de systèmes en réaction a été utilisée soit directement [41], soit en estimation initiale d'autres méthodes [42-44]. Les résultats obtenus peuvent en effet être améliorés. C'est ce dernier point qui nous intéressera plus particulièrement dans la section 3.

2.4 SIMPLISMA

La seconde approche possible pour construire une estimation de **S** ou de **C** en accord avec la nature chimique des procédés étudiés est d'utiliser SIMPLISMA (*SIMPL*e to use *Interactive Self-Modelling Algorithm*) [45-47]. Cet algorithme repose également sur le modèle bilinéaire présenté dans le paragraphe 1.1 et sur la sélection de ce que l'on appelle des variables pures. Une variable temporelle pure est un stade du procédé ayant une variance unique au regard de la concentration d'un seul analyte du mélange et ne présentant pas de variance vis-à-vis des autres composés. De façon analogue il est possible de chercher une variable pure le long de la direction spectrale. Selon le cas, on construit soit une estimation des profils de concentration soit une estimation des spectres, comme le montre la Figure 18. En suivi de procédés la direction temporelle est implicitement organisée par l'évolution de la réaction. Les variables pures le sont plus dans la direction des concentrations parce que ces profils sont normalement moins superposés que les spectres des espèces chimiques, SIMPLISMA est alors plus efficace pour estimer la matrice **S**.

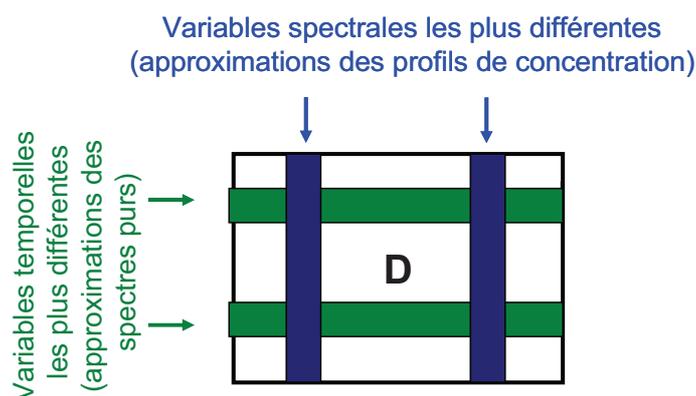


Figure 18 : Représentation graphique de l'obtention de profils de concentration ou de spectres purs par la méthode SIMPLISMA

La pureté p_i sert de mesure de sélectivité pour un analyte particulier. Si SIMPLISMA s'applique dans la direction des variables temporelles, la pureté se calcule comme le ratio de

l'écart type σ_i et de la moyenne μ_i pour chaque spectre d_i d'une matrice de données bilinéaires $\mathbf{D}(m \times n)$.

$$p_i = \frac{\sigma_i}{\mu_i} \quad (14)$$

Afin d'éviter d'attribuer une importante valeur de pureté à des spectres (ou des temps) dont la moyenne serait faible, c'est-à-dire des spectres de bruit, un offset est inclut à la moyenne μ_i . Le nouveau dénominateur μ_i' est calculé selon l'équation (15).

$$\mu_i' = \mu_i + (\text{offset} / 100) * \max(\mu_i) \quad i = 1, \dots, m \quad (15)$$

Le processus suivi par SIMPLISMA peut être décrit en cinq étapes, plaçons nous dans le cas de la construction d'une estimation de \mathbf{S} :

1. Sélection de la variable (temporelle) la plus pure, c'est à dire présentant la pureté p_i maximale (équation (14)), elle indique le premier spectre sélectionné.
2. Normalisation des spectres.
3. Sélection de la seconde variable la plus pure, pour cela :

- a. Calcul du poids w_i de chaque objet :

$$w_i = \det(\mathbf{Y}_i^T \mathbf{Y}_i) \quad i = 1, \dots, m \quad (16)$$

\mathbf{Y}_i est la matrice formé par le spectre sélectionnée à l'étape 1 et le spectre \mathbf{d}_i .

- b. Recalcul de la pureté.

$$p_i' = w_i p_i \quad (17)$$

- c. Sélection de la variable la plus pure.

4. Sélection de la troisième variable la plus pure, pour cela :

- a. Calcul du nouveau poids de chaque chaque objet :

$$w_i' = \det(\mathbf{Y}_i^T \mathbf{Y}_i) w_i \quad i = 1, \dots, m \quad (18)$$

\mathbf{Y}_i est la matrice formé par les variables sélectionnées à l'étape 1 et 3 et le spectre \mathbf{d}_i .

- b. Recalcul de la pureté.

$$p_i'' = w_i' p_i \quad (19)$$

- c. Sélection de la variable la plus pure.

5. Suite des étapes de recalcul du poids et de sélection d'une variable jusqu'à ce que la variable sélectionnée (son spectre associé) ne soient plus discernable des sélections précédentes (ou que le rang connu soit atteint).

Les spectres obtenus en appliquant SIMPLISMA à notre exemple D_{ex} sont présentés Figure 19a. Ils forment la matrice S_{ex} . De manière générale on peut constater que les principaux motifs spectraux sont représentés, notamment à 276 nm, 356 nm, 375 nm et 402 nm. Il est aussi possible de voir que deux des trois spectres proposés correspondent au premier et au dernier spectre de la matrice D_{ex} . Cette configuration est courante en suivi de procédé puisque le premier et le dernier spectre sont très différents l'un de l'autre (et de la moyenne) ils sont considérés comme les plus purs et donc sélectionnés par cet algorithme. SIMPLISMA a aussi été utilisé ici pour évaluer les profils de concentration, la matrice C_{ex} obtenue est présentée Figure 19b. Ces profils sont difficiles à interpréter et sont en fait des combinaisons linéaires des vrais profils de concentration purs. Ceci est dû au fait qu'il n'y ait aucune longueur d'onde (et, par conséquent, aucune trace cinétique associée) qui soit spécifique à une espèce chimique. Comme on a déjà dit, ce manque de sélectivité des longueurs d'onde fait que, en général, SIMPLISMA s'applique sur la direction des variables temporelles (ou variables du procédé).

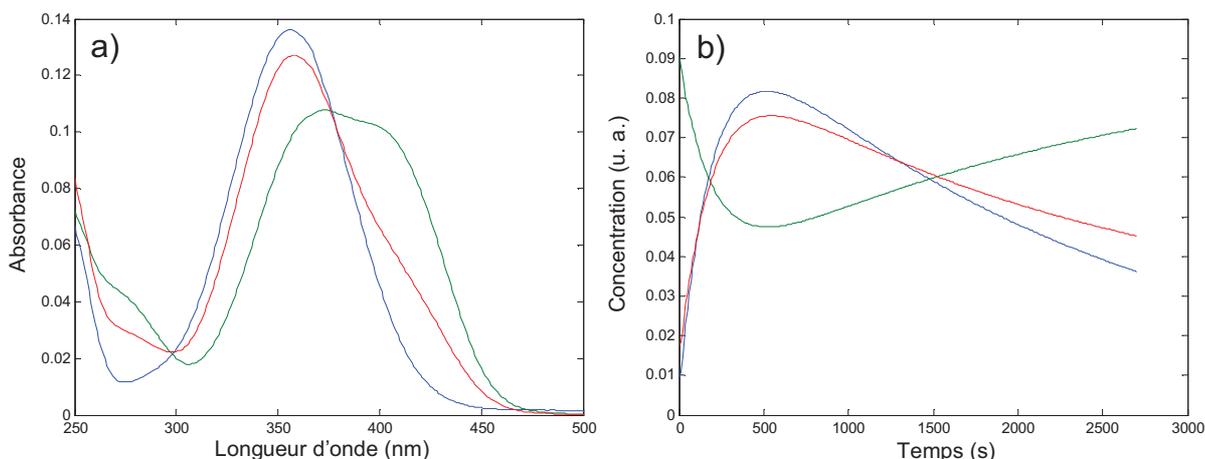


Figure 19 : Estimations construites par SIMPLISMA a) des spectres purs S_{ex} et b) des profils de concentration C_{ex} .

L'ACP mène à une décomposition bilinéaire unique de la matrice de données D mais les *scores* et *loadings* obtenus ne sont pas directement interprétables chimiquement. L'EFA et SIMPLISMA permettent quant à eux de construire des estimations des matrices C et S , en général interprétables du point de vue chimique. Ces solutions ne sont cependant pas toujours satisfaisantes et peuvent être améliorées. Ceci est possible en utilisant des méthodes de modélisation de procédés souples, tels que l'algorithme de résolution multivariée de courbes par moindres carrés alterné, MCR-ALS, détaillé au paragraphe 3

3. Méthodes de modélisation souple de procédés (soft-modelling)

Les résultats obtenus par les méthodes exploratoires présentées précédemment, assimilables aux matrices \mathbf{C} et \mathbf{S} , ne sont pas toujours complètement satisfaisants. Ces solutions pourraient être améliorées en introduisant des contraintes lors de la résolution. C'est l'idée principale des méthodes de *soft modelling* que nous qualifierons par analogie de méthodes de modélisation souple. Les informations utilisées pour contraindre la résolution ne nécessitent pas forcément une connaissance approfondie du système étudiée. Il n'y a donc pas d'utilisation explicite d'un modèle physico-chimique, par opposition au *hard-modelling* que nous verrons au paragraphe 4. La décomposition bilinéaire de la matrice de données \mathbf{D} par les méthodes de résolution souple a néanmoins l'inconvénient de fournir un ensemble de solutions équivalentes, puisque la condition d'orthogonalité de l'ACP est levée. On parle d'ambiguïtés de rotation des solutions obtenues. Nous commencerons donc par présenter cet aspect avant de nous concentrer sur l'algorithme de Résolution Multivariée de Courbes par Moindres Carrés Alternés (MCR-ALS, *Multivariate Curve Resolution Alternating Least Squares*) [48].

3.1 Ambiguïtés de rotation

La décomposition des données multivariées en un produit de deux matrices \mathbf{C} et \mathbf{S}^T s'effectue selon l'équation (6), rappelée ci-dessous, et a pour critère d'optimisation la minimisation des résidus \mathbf{E} .

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (6)$$

Quelque soit la méthode utilisée ; cette décomposition peut être affectée par un facteur d'ambiguïté. En effet l'équation (20) montre que l'introduction d'une matrice non-singulière \mathbf{R} peut changer la décomposition sans faire varier \mathbf{E} .

$$\mathbf{D} = \mathbf{C}(\mathbf{R}\mathbf{R}^{-1})\mathbf{S}^T + \mathbf{E} \quad (20)$$

$$\mathbf{C}' = \mathbf{C}\mathbf{R} \quad (21)$$

$$\mathbf{S}'^T = \mathbf{R}^{-1}\mathbf{S}^T \quad (22)$$

Les matrices \mathbf{C}' et \mathbf{S}' et \mathbf{C} et \mathbf{S} ne diffèrent que par rotation et correspondent strictement aux mêmes résidus \mathbf{E} . Le pourcentage de variance expliquée ne varie pas. On dit que les solutions subissent une rotation, d'où le terme généralement utilisé d'ambiguïtés de rotation [4, 49]. On parle également d'ambiguïtés d'intensité lorsque \mathbf{R} est une matrice diagonale ; dans ce cas,

seulement l'échelle des profils de concentration et des spectres est variée. L'effet des ambiguïtés de rotation sur la forme et l'interprétation des spectres et profils de concentration peut être dramatique et une série de contraintes a été mise au point permettant d'interdire certaines solutions et donc de réduire (ou d'éliminer) les rotations possibles.

3.2 Résolution Multivariée de Courbes par Moindres Carrés Alternés (MCR-ALS)

MCR-ALS vise à améliorer la justesse des résultats d'une décomposition bilinéaire en réduisant les ambiguïtés de rotation (c'est-à-dire les bandes de solutions possibles) par l'application de contraintes. Cet algorithme a en effet été appliqué avec succès dans des domaines aussi variés que la chromatographie [50, 51], l'électroanalyse [52, 53], les spectroscopies [54] infrarouge [55-59], Raman [60, 61], UV-Visible [62-64], de fluorescence [44, 65], RMN [66] ou de masse [67]. MCR-ALS permet de contraindre les solutions à respecter certaines règles basiques de la chimie. On entend par là des vérités communes à de nombreux systèmes (comme par exemple la non négativité des concentrations). Nous commencerons par définir ces contraintes.

Le terme contrainte désigne des propriétés mathématiques ou chimiques que doivent respecter les solutions d'une résolution multivariée. Elles permettent de minimiser l'effet des ambiguïtés de rotation. Les contraintes peuvent s'appliquer aussi bien aux profils de concentration qu'aux spectres. Les contraintes courantes en résolution de courbes multivariées sont présentées ci-dessous.

- **Non négativité** : il s'agit de la contrainte la plus généralement applicable aux données chimiques [68, 69]. Elle restreint les valeurs des profils de concentration ou les valeurs des spectres à être supérieures ou égales à zéro (voir Figure 20). Il est possible de l'appliquer de plusieurs manières. La plus directe, mais aussi la moins satisfaisante, est de remplacer les valeurs négatives par zéro. L'autre solution consiste à appliquer l'algorithme des moindres carrés non négatifs (*Non Negative Least Squares* NNLS) [70] ou les adaptations de cet algorithme nécessitant moins de temps de calcul et développées plus récemment pour traiter des volumes de données plus importants [71, 72]. Cette contrainte s'applique généralement aux profils de concentration et à des

réponses instrumentales qui sont, par nature, positives (par exemple, spectres d'absorption UV,...).

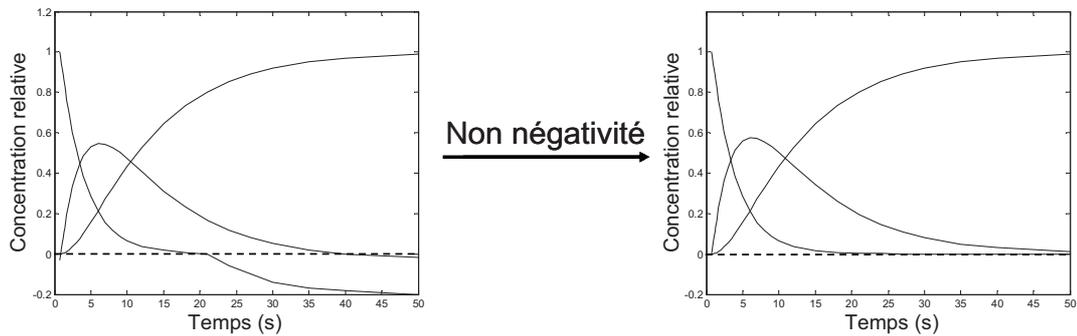


Figure 20 : Exemple d'application de la contrainte de non négativité.

- **Unimodalité** : cette contrainte peut s'appliquer à des profils ne possédant qu'un maximum, c'est souvent le cas pour des profils cinétiques (voir Figure 21). Plusieurs algorithmes existent [73] mais la forme la plus appliquée est aussi la plus intuitive. Le maximum est recherché. Les autres valeurs étant sensées diminuer de façon monotone (à un facteur de bruit instrumental près), les maxima locaux sont remplacés par la valeur moyenne des deux mesures voisines.

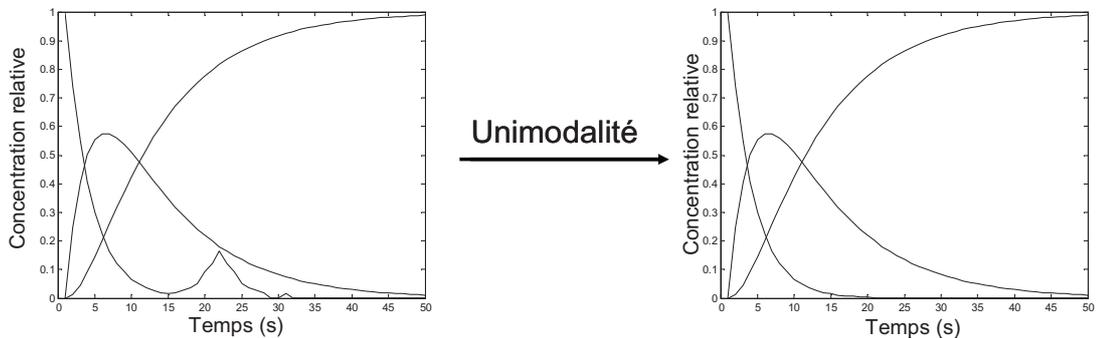


Figure 21 : Exemple d'application de la contrainte d'unimodalité.

- **Système fermé (ou clos)** (balance des masses) : cette contrainte peut être appliquée lorsque la concentration totale de matière est constante (Figure 22). Elle supprime les ambiguïtés d'intensité et normalise les profils de concentration.

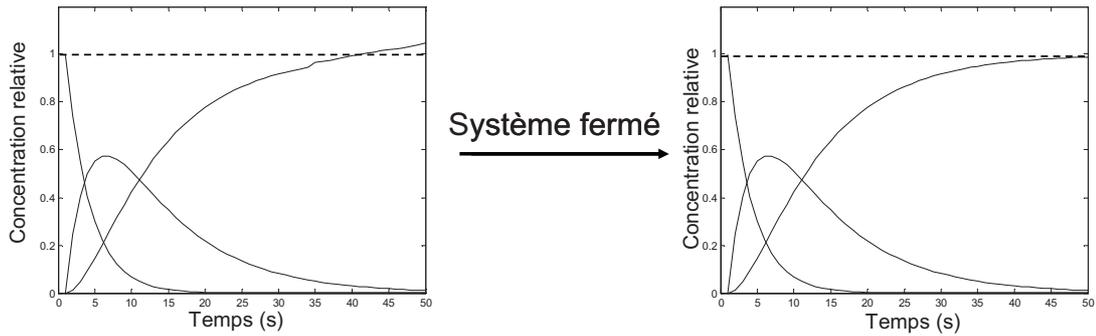


Figure 22 : Exemple d'application de la contrainte de système fermé.

- **Sélectivité et rang local** : lorsqu'une seule espèce contribue au signal dans une zone donnée, spectrale ou temporelle, et que cette information est connue, il est possible d'utiliser cette information sur le rang pour faciliter la résolution. La contrainte de sélectivité permet d'identifier plus facilement l'espèce en question et indirectement les autres espèces impliquées ailleurs dans la réaction. Il est aussi possible d'utiliser la connaissance de l'absence d'une ou plusieurs espèces (rang local) lors de la résolution. Cet information peut être connue *a priori* ou être obtenue par méthodes d'analyse de rang local, telles qu'EFA.
- **Egalité** : cette contrainte permet d'introduire une connaissance éventuelle de certains spectres purs ou de profils de concentration. Elle agit comme une fonction de pénalisation à chaque itération, forçant les solutions calculées à se rapprocher de l'information introduite [70, 74, 75]. Cette dernière contrainte est la traduction d'une connaissance partielle du système étudié. Elle est toutefois considérée comme une contrainte sans *a priori* car elle n'induit pas un modèle complet (au sens de la modélisation rigide).

MCR-ALS est une des méthodes itératives les plus populaires du fait de sa grande flexibilité. L'application de contraintes en est la clé de voûte. Maintenant que celles-ci ont été définies, il est temps de décrire le fonctionnement de cet algorithme.

MCR-ALS décompose la matrice de données \mathbf{D} en un produit de deux matrices \mathbf{C} et \mathbf{S}^T contenant respectivement les profils de concentration et les spectres des espèces purs. Les résidus de cette décomposition sont rassemblés dans la matrice d'erreur \mathbf{E} , comme le montre l'équation (6).

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (6)$$

La matrice \mathbf{D} , de rang k , est composée de m spectres de n longueurs d'onde, la matrice \mathbf{C} est donc de dimension $(m \times k)$ et \mathbf{S}^T de dimension $(k \times n)$. Cette décomposition peut être résumée en cinq étapes :

- 1) Détermination du nombre de composants présents (du rang chimique) dans \mathbf{D} . Cette étape est généralement réalisée en utilisant la décomposition en valeurs singulières (SVD) ou, dans certains cas, peut répondre à une connaissance préalable du système.
- 2) Construction d'une estimation initiale. MCR-ALS étant un algorithme itératif il faut choisir un point de départ avant de lancer le calcul. L'estimation initiale fournie peut aussi bien être une estimation des profils de concentration, \mathbf{C} , construite par EFA par exemple, que des spectres purs, \mathbf{S}^T , calculée par SIMPLISMA ou d'autres méthodes non aléatoires. L'algorithme proposé correspond au cas où l'estimation fournie est \mathbf{C} .
- 3) Calcul par moindres carrés, sous contraintes, d'une estimation de \mathbf{S}^T à partir de \mathbf{D} et de l'estimation de \mathbf{C} selon l'équation (23)

$$\mathbf{S}^T = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{D} \quad (23)$$

- 4) Calcul par moindres carrés sous contraintes d'une nouvelle estimation de \mathbf{C} , plus juste du point de vue chimique que l'estimation initiale, à partir de \mathbf{D} et de l'estimation de \mathbf{S}^T selon l'équation (24).

$$\mathbf{C} = \mathbf{D} \mathbf{S} (\mathbf{S}^T \mathbf{S})^{-1} \quad (24)$$

- 5) Reconstruction de \mathbf{D} en utilisant \mathbf{C} et \mathbf{S}^T de manière à évaluer les résidus \mathbf{E} et donc le manque d'ajustement (*Lack of fit Lof*) et retour à l'étape 3) si la convergence n'est pas atteinte. Un critère de convergence est préétabli définissant un seuil de différence entre deux valeurs de *Lof* pour deux itérations consécutives.

Deux critères permettent d'évaluer la qualité de la solution obtenue, le pourcentage de variance expliquée R^2 (défini par l'équation (25)) et le *Lof* calculé selon l'équation (26).

$$R^2 = 100 \times \frac{\sum_i \sum_j \hat{d}_{i,j}^2}{\sum_i \sum_j d_{i,j}^2} \quad (25)$$

$$Lof = 100 \times \sqrt{\frac{\sum_i \sum_j (d_{i,j} - \hat{d}_{i,j})^2}{\sum_i \sum_j d_{i,j}^2}} \quad (26)$$

Les termes $d_{i,j}$ et $\hat{d}_{i,j}$ désignent les éléments de la matrice \mathbf{D} et de la reconstruction $\hat{\mathbf{D}}$ de celle-ci à partir des solutions \mathbf{C} et \mathbf{S} de MCR-ALS. Le fonctionnement de l'algorithme MCR-ALS est résumé sous forme schématique par la Figure 23.

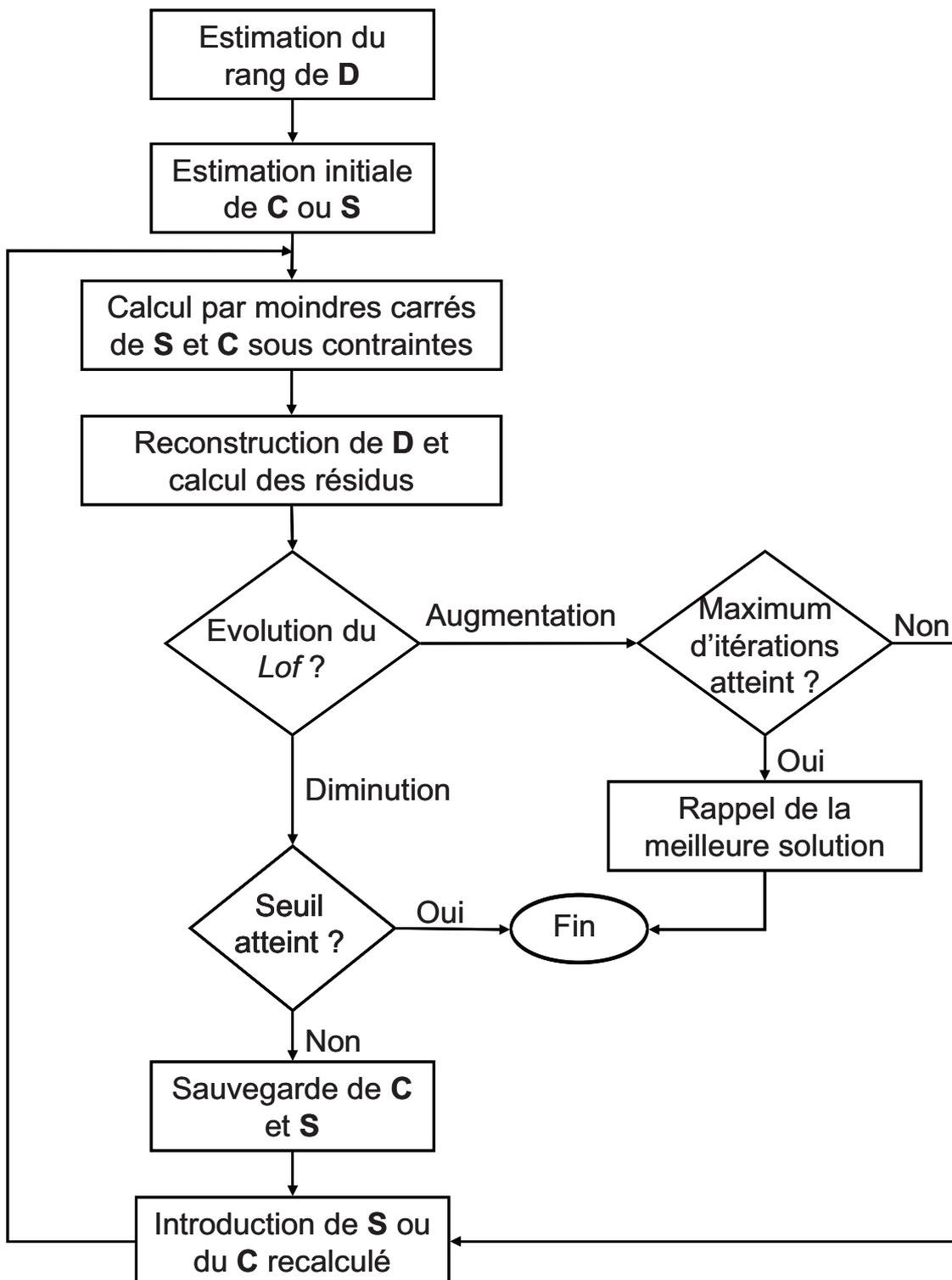


Figure 23 : Schéma de principe de MCR-ALS

La décomposition proposée par MCR-ALS de \mathbf{D}_{ex} est présentée Figure 24. Les contraintes appliquées dans ce cas sont la non négativité de \mathbf{C}_{ex} et \mathbf{S}_{ex} ainsi que l'unimodalité des profils de concentration. Le modèle obtenu explique 99,99% de la variance et le *Lack of fit* est égal à 0,08%. L'estimation initiale utilisée est celle provenant de l'EFA, des résultats identiques peuvent être obtenus dans ce cas en utilisant l'estimation de \mathbf{S}_{ex} par SIMPLISMA.

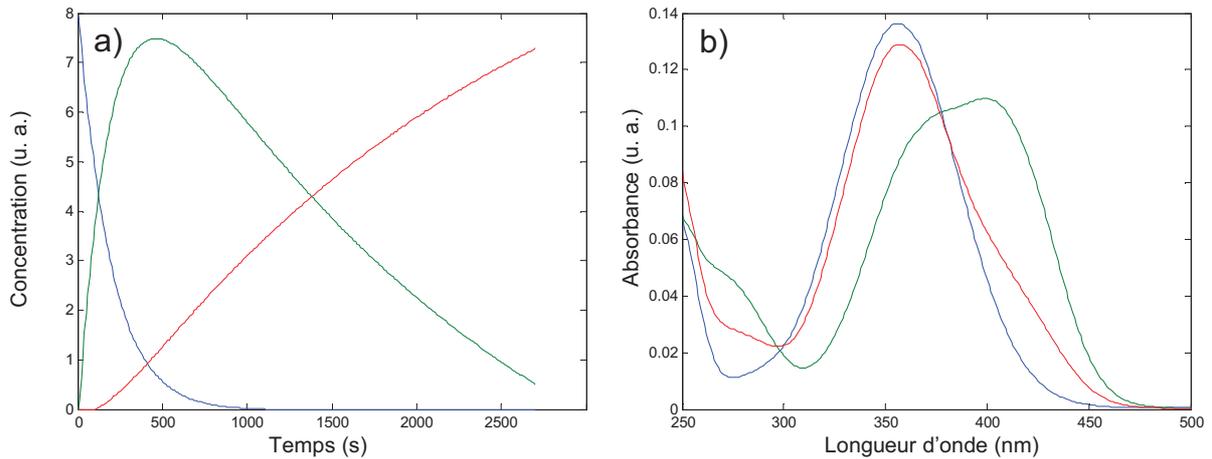


Figure 24 : Résultats de la décomposition par MCR-ALS de \mathbf{D}_{ex} en a) profils de concentrations \mathbf{C}_{ex} et b) spectres purs \mathbf{S}_{ex} .

Les contraintes disponibles dans MCR-ALS ne sont pas le seul moyen de réduire les ambiguïtés de rotation. Il est en effet possible d'appliquer cette méthode à des systèmes multi-expériences (matrices augmentées ou structures *multiset*). Le surcroît d'information présent dans les différentes matrices de données limite le nombre de solutions possibles et donc réduit les ambiguïtés. Dans ce cas des contraintes différentes peuvent être appliquées à chacune des matrices \mathbf{C} et \mathbf{S} , voire à seulement une partie d'entre elles (à quelques sous matrices \mathbf{C}_i et/ou \mathbf{S}_i). On peut imaginer comme exemple une résolution utilisant les contraintes locales telles que représentées par la Figure 25. Ici un spectre est contraint à être égal à un spectre connu, les deux autres doivent respecter la non négativité. Une des deux sous matrices constituant \mathbf{C} (ici $\mathbf{C} = [\mathbf{C}_A ; \mathbf{C}_B]$) correspond à un système fermé et deux des profils de concentration doivent respecter l'unimodalité.

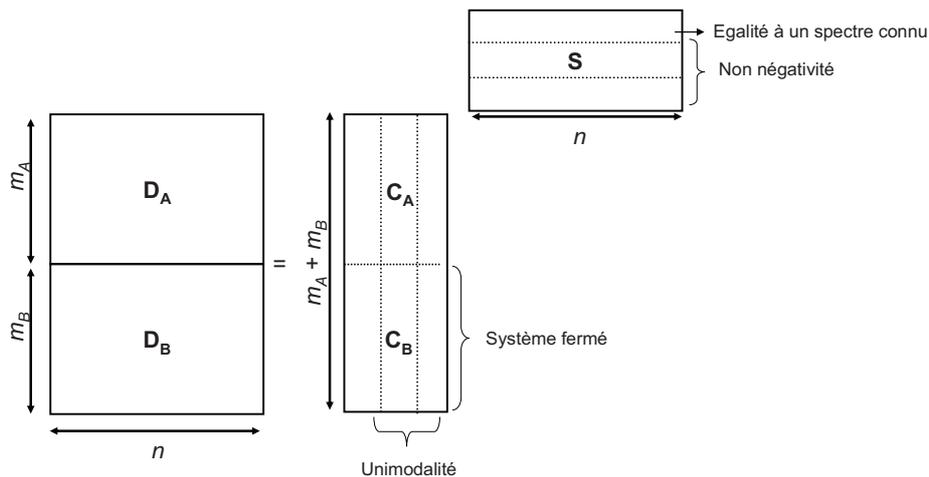


Figure 25 : Illustration de la souplesse d'application des contraintes à une augmentation de matrices dans la direction des colonnes.

Les profils obtenus après l'application de contraintes peuvent être uniques dans des cas favorables où la sélectivité ou des conditions de rang local le permettent [76]. La plupart des résolutions, même contraintes, sont soumises à des degrés d'ambiguïté variables [48]. Il est alors nécessaire de considérer un ensemble de solutions équivalentes. Cet ensemble est connu comme les bandes de faisabilité. Plusieurs procédures ont été proposées pour estimer ces bandes [77-79]. Il a été démontré qu'il est possible d'estimer les limites de ces bandes en utilisant une fonction d'optimisation basée sur le ratio du signal d'une espèce sur le signal total [80] et donc d'observer l'effet des contraintes souples [81] ou de l'augmentation de matrices [82] sur la résolution. En partie dans l'objectif de diminuer les ambiguïtés, MCR-ALS a dernièrement été modifié de manière à rendre possible l'application de contraintes rigides, notamment des modèles cinétiques. Nous reviendrons sur cet aspect dans la section 4 de ce chapitre.

Cette section illustre les capacités des méthodes de résolution souples, mais une autre approche appelée modélisation rigide (*hard modelling*) de procédés est possible. Le concept de base de cette approche est différent de celui de MCR-ALS puisqu'il s'agit de s'appuyer sur la connaissance *a priori* du schéma réactionnel du système.

4. Méthode de modélisation rigide de procédés

Une approche alternative au *soft modelling* de l'analyse de données spectroscopiques résolues en temps consiste à utiliser l'information chimique disponible pour construire un modèle physico-chimique. On parle alors de *hard modelling*, ici littéralement traduit par modélisation rigide. Différentes méthodes basées sur des modèles physico-chimiques ont été proposées pour calculer les constantes d'équilibre [68, 83-90]. Dans notre cas, on suppose que les données respectent un modèle bilinéaire permettant de décomposer la matrice \mathbf{D} selon l'équation (6) en deux matrices \mathbf{C} et \mathbf{S}^T contenant les profils de concentration et les spectres purs des espèces impliquées. La qualité d'une résolution rigide repose sur trois piliers : la qualité du modèle cinétique choisi, celle de l'ajustement des paramètres (c'est-à-dire des constantes de réaction) et celle de l'ajustement des données. La première étape consiste logiquement à choisir le modèle chimique qui s'ajuste le mieux aux données expérimentales. Ce choix repose sur la connaissance *a priori* du système étudié (ou sur le résultat d'autre analyse telle que la modélisation souple). Ensuite les paramètres de ce modèle doivent être optimisés : on recherche les paramètres minimisant les résidus \mathbf{E} . Les profils de concentration reconstruits à partir des paramètres estimés permettent de reconstruire les spectres purs.

La modélisation paramétrique peut aussi être envisagée dans la direction spectrale [84, 91]. Ce type de modèle nécessite cependant davantage de paramètres qu'un modèle cinétique. La première étape consiste à réaliser un ajustement correct des données avec le nombre correct de formes spectrales et d'amplitudes, les spectres (d'absorption ou de fluorescence) pouvant être décrits comme une superposition de bandes de formes standards spécifiques (Gaussienne, Lorentzienne, de Voigt, de Pearson ...). Ces bandes sont elles mêmes descriptibles à travers quelques paramètres. Ces formes sont dénommées Courbes d'Amplitudes de Bandes (BAC, *Band Amplitude Curves*). Les BAC peuvent ensuite être utilisées dans un modèle cinétique [91] ou thermodynamique [92]. Un grand nombre d'algorithmes est donc disponible, nous nous contenterons d'en décrire un basé sur une modélisation cinétique, qui sera utilisé par la suite dans cette thèse.

Le principe de la méthode de modélisation rigide qui utilise l'algorithme des moindres carrés non linéaires est résumé par la Figure 26. L'ajustement des paramètres du modèle

permet de construire la matrice **C**. La matrice **S** peut ensuite être estimée en utilisant **D** et **C** par moindres carrés (cf. équation (23)).

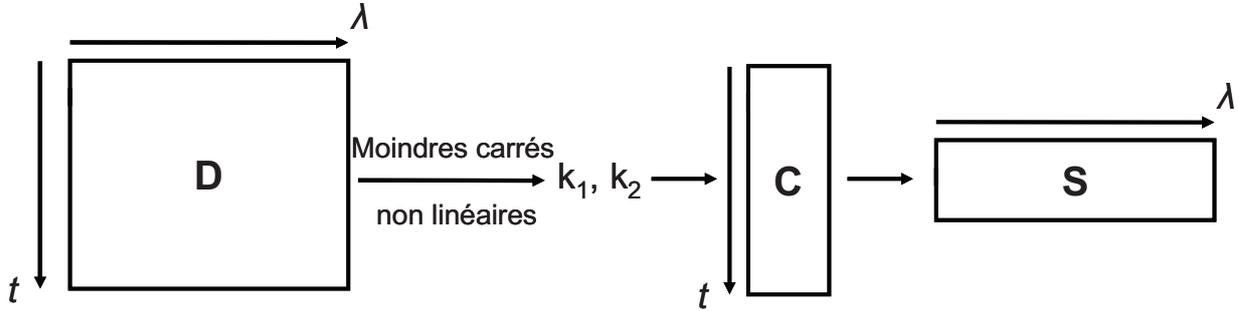


Figure 26 : Principe de la méthode des moindres carrés non linéaires, la matrice **D** est utilisée pour estimer les paramètres d'un modèle (ici deux paramètres cinétiques k_1 et k_2) qui sont ensuite utilisés pour calculer **C** et **S**.

La modélisation rigide de données cinétiques repose sur l'ajustement non linéaire des paramètres du modèle. En effet l'estimation initiale de la matrice **C** est construite à partir de la connaissance des concentrations initiales du système et d'une estimation des paramètres qui régissent le modèle physico-chimique, c'est-à-dire, des constantes cinétiques. L'objectif est finalement d'obtenir un jeu de paramètres optimisés permettant de reconstruire une estimation $\hat{\mathbf{D}}$ aussi proche que possible de la matrice **D**. Ceci revient à minimiser la matrice de résidus **E** et peut être mis en équation sous la forme :

$$\min \|\mathbf{D} - \hat{\mathbf{D}}\| \quad (27)$$

Or la matrice reconstruite $\hat{\mathbf{D}}$ peut s'écrire :

$$\hat{\mathbf{D}} = \mathbf{C}\mathbf{S}^T \quad (28)$$

$$\hat{\mathbf{D}} = \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{D} \quad (29)$$

Donc le critère d'optimisation s'écrit :

$$\min \|\mathbf{D} - \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{D}\| \quad (30)$$

$$\min \|\mathbf{D} (\mathbf{I} - \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T)\| \quad (31)$$

Cette dernière équation met en lumière le fait que l'ajustement peut reposer uniquement sur les données originales et les profils de concentration générés par le modèle. Il n'est pas nécessaire de prendre en compte la dimension spectrale. La matrice de profils de concentration utilisée ne dépend que des paramètres du modèle choisi. Ceci simplifie le

problème, l'optimisation itérative ne s'effectuant plus que sur les paramètres non linéaires [93, 94]. La fonction d'optimisation minimisant les résidus ne dépend plus que d'un faible nombre de paramètres : les constantes de réactions.

La minimisation de résidus s'effectue en utilisant l'algorithme de Levenberg-Marquardt (ALM) [95]. Celui-ci fait le lien entre l'algorithme de Newton-Gauss (ANG) et la méthode de descente de gradient. L'ALM, bien que plus lent, est plus robuste que l'ANG et permet de retrouver la solution à partir d'estimations initiales éloignées du minimum final. L'ALM se base sur l'algorithme de Newton-Gauss. Commençons par détailler celui-ci. Une estimation $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ du jeu de paramètres du modèle doit être fournie pour lancer l'optimisation. A chaque itération ce vecteur β sera remplacé par une nouvelle estimation $(\beta + \delta)$. Les fonctions $f(\beta + \delta)$ définissant le modèle sont approximées dans le but de déterminer δ .

$$f(\beta + \delta) \approx f(\beta) + J\delta \quad (32)$$

où J est la matrice Jacobienne de f pour les paramètres β . Le gradient de la somme des carrés des résidus (ssq) devient nul lorsqu'un minimum de résidus est atteint. Le terme à droite de l'équation (32) peut alors être différencié ce qui mène à l'équation (33).

$$(J^T J)\delta = J^T [D - f(\beta)] \quad (33)$$

Ceci correspond à un ensemble d'équations linéaires, qui peut être résolues pour δ . L'algorithme de Newton-Gauss peut être schématisé par la Figure 27.

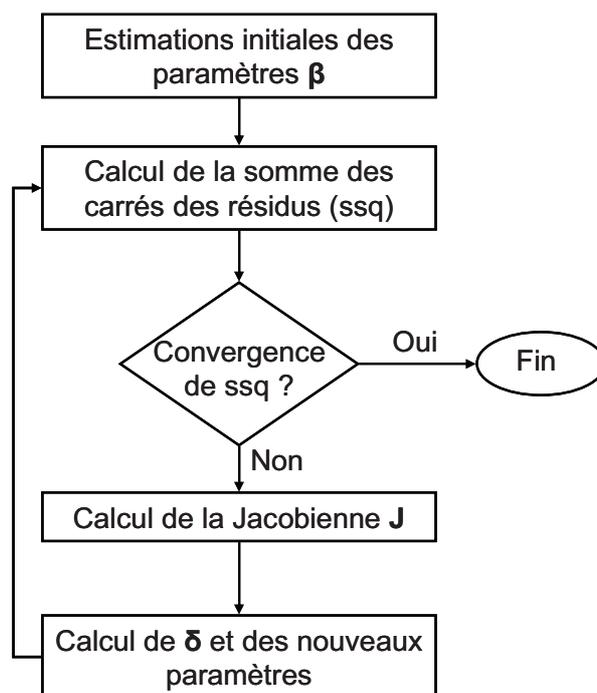


Figure 27 : Schéma de principe de l'algorithme de Newton-Gauss

La modification introduite par Marquardt consiste à ajouter un terme scalaire m_p permettant d'éviter une divergence de l'optimisation.

$$(\mathbf{J}^T \mathbf{J} + m_p \mathbf{I}) \delta = \mathbf{J}^T [\mathbf{D} - f(\boldsymbol{\beta})] \quad (34)$$

La valeur m_p permet de diriger l'ALM soit vers la méthode de Gauss-Newton pour les cas où la somme des carrés des résidus diminue rapidement (m_p faible) soit vers la méthode de descente de gradient (m_p élevé). L'algorithme de Levenberg-Marquardt peut être schématisé par la Figure 28.

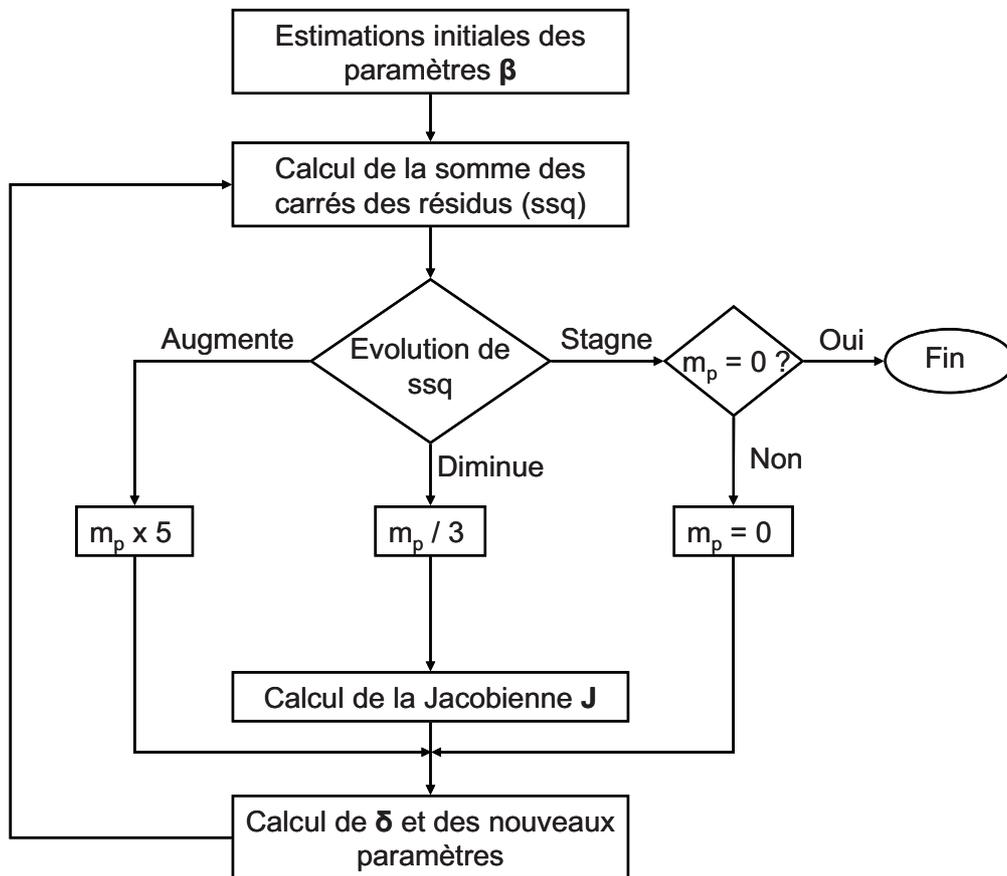


Figure 28 : Schéma de principe de l'algorithme de Levenberg-Marquardt

L'algorithme permet en outre d'estimer l'écart-type des constantes de réactions $std(\beta_i)$, en utilisant les résidus (res) et leur Jacobienne \mathbf{J} comme le montre l'équation (35).

$$std(\beta_i) = \sqrt{(\mathbf{J}^T \mathbf{J})^{-1}} std(res) \quad (35)$$

Cette méthode a été appliquée pour décomposer la matrice \mathbf{D}_{ex} selon le modèle défini par les équations (11) et (12). Les résultats obtenus sont présentés Figure 29. Ils correspondent à une décomposition dont le Lof est égal à 0,4 % et le R^2 à 99,998 %. Les profils de concentration et spectres obtenus sont en accord avec les résultats obtenus en résolution souple. On remarque cependant que l'avancement de la réaction n'est pas le même : l'espèce intermédiaire est encore présent à hauteur de 30% dans le mélange au dernier temps là où en résolution souple elle avait quasiment disparue. Ceci peut être dû aux ambiguïtés de rotation affectant MCR-ALS. Les constantes de réaction obtenues sont $k_1 = 0,29 \text{ min}^{-1}$ et $k_2 = 0,03 \text{ min}^{-1}$.

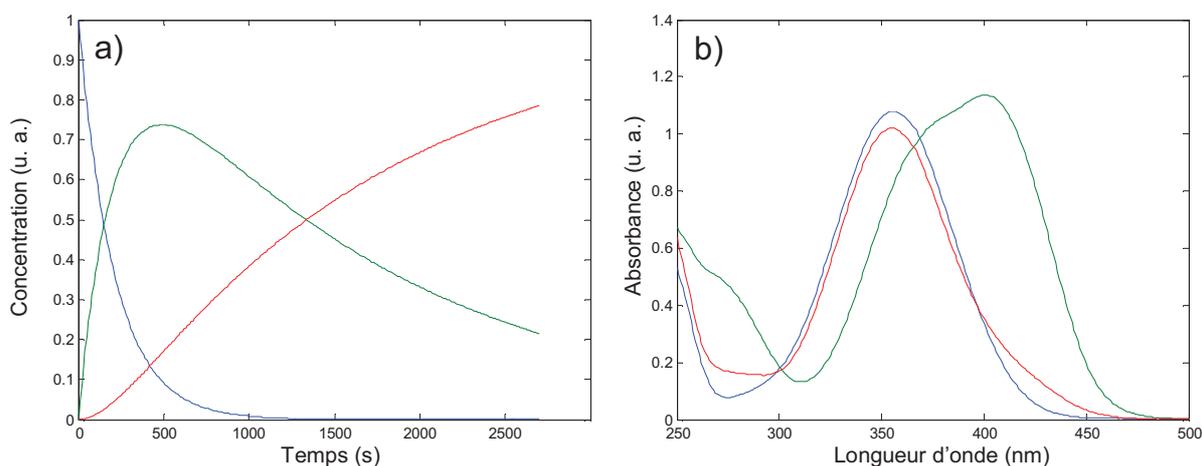


Figure 29 : Résultats de la décomposition par la méthode des moindres carrés non linéaires de \mathbf{D}_{ex} en a) profils de concentrations \mathbf{C}_{ex} et b) spectres purs \mathbf{S}_{ex} .

Cet exemple est bien adapté pour illustrer l'intérêt de la modélisation rigide. Cependant diverses applications (notamment celles détaillées dans les chapitres 3 et 4) ne permettent pas d'appliquer un modèle cinétique aussi simplement. Plusieurs problèmes peuvent en effet se poser: le premier est évidemment que la connaissance dont on dispose *a priori* n'est en général pas suffisante pour postuler un modèle. La seconde difficulté possible, et plus insurmontable, est que le système chimique dans son ensemble ne suit pas forcément un modèle cinétique. En effet les méthodes rigides cherchent à exprimer toute la variance des données spectroscopiques originales en utilisant le modèle cinétique postulé (voir équation (31)). S'il existe des sources de variance qui ne sont pas reliées à la réaction, comme c'est le

cas lors d'un décalage de ligne de base ou en présence d'interférences absorbantes, l'ajustement cinétique mène à des résultats erronés.

De plus il est important de noter que l'information spectrale n'est jamais utilisée de façon active dans les approches de modélisation rigide. L'optimisation recherche l'ajustement optimum de toute la variation de \mathbf{D} en n'utilisant que les profils de concentrations \mathbf{C} découlant des paramètres β du modèle cinétique. Les informations disponibles dans le domaine spectral (de sélectivité par exemple) ne sont pas directement utilisables. Ce défaut est particulièrement critique lors de l'analyse d'une matrice augmentée dans la direction des colonnes, puisque l'on perd le lien constitué par les spectres \mathbf{S} communs entre les matrices de données expérimentales. Cette perte oblige la modélisation rigide, quand elle est appliquée à de structures *multiset*, à reposer sur un modèle cinétique global qui décrive le comportement de toutes les expériences analysées ou, à la limite, sur des modèles individuels dont on connaisse les relations.

Une approche hybride entre la modélisation souple et la modélisation rigide a été proposée pour répondre à ces problèmes. Cette méthode est décrite dans la partie suivante de ce chapitre.

5. Méthode de modélisation hybride souple et rigide de procédés

La modélisation souple présente l'avantage de ne nécessiter aucun modèle concernant les données mais les résultats obtenus peuvent être affectés par les ambiguïtés de rotation et ne fournissent pas directement de paramètres physicochimiques. La modélisation rigide est quant à elle complètement tributaire du modèle choisi et ne permet donc pas d'analyser un système dont notre connaissance serait partielle ou qui présente des contributions au signal qui ne peuvent pas être expliquées par le modèle. Les deux approches sont conceptuellement opposées mais leurs avantages respectifs sont complémentaires. L'algorithme hybride de résolution de courbes multivariées allie les qualités de ces deux familles (HS-MCR, *Hard-Soft-Multivariate Curve Resolution*) [96]. Celui-ci est basé sur la structure de MCR-ALS. Un nouveau type de contraintes est implémenté : il est possible d'utiliser un modèle cinétique. Mais contrairement à l'approche rigide, on applique le modèle cinétique avec l'objectif d'ajuster la matrice des profils de concentration **C** et pas la matrice de données **D**. Cette différence est importante car la matrice **C** contient l'information de toutes les espèces chimiques contribuant au signal de façon complètement séparée, ce qui n'est pas le cas dans la matrice **D**, où le signal enregistré est le mélange des contributions. En effet, la séparation de l'information du système en différents profils de concentration permet de ne travailler que sur certaines colonnes de **C** ; c'est à dire sur les contributions qui obéissent individuellement au modèle cinétique proposé. Les autres espèces (intérférents) ne sont soumises qu'à des contraintes souples. Le modèle cinétique n'étant donc pas forcément appliqué à l'ensemble des phénomènes se produisant dans l'échantillon, il ne s'agit pas non plus d'une contrainte rigide au sens strict. Cette méthode a été appliquée avec succès au suivi de système chimiques [42] ou biologiques en présence d'espèces interférentes [97, 98].

D'un point de vue technique cette méthode utilise l'estimation des profils de concentration calculée lors de la quatrième étape de MCR-ALS pour effectuer un ajustement non linéaire de l'ensemble ou d'une partie des profils de **C**. Une boucle supplémentaire vient alors se greffer à MCR-ALS dans laquelle s'applique l'algorithme de Levenberg-Marquardt. Les profils calculés par le modèle sont ensuite réinjectés pour le calcul postérieur de **S** sous contraintes souples. Les solutions sont donc améliorées de façon itérative à la fois par des contraintes souples et par un modèle cinétique dont les paramètres sont de mieux en mieux ajustés. Le fonctionnement de cet algorithme peut être résumé par la Figure 30.

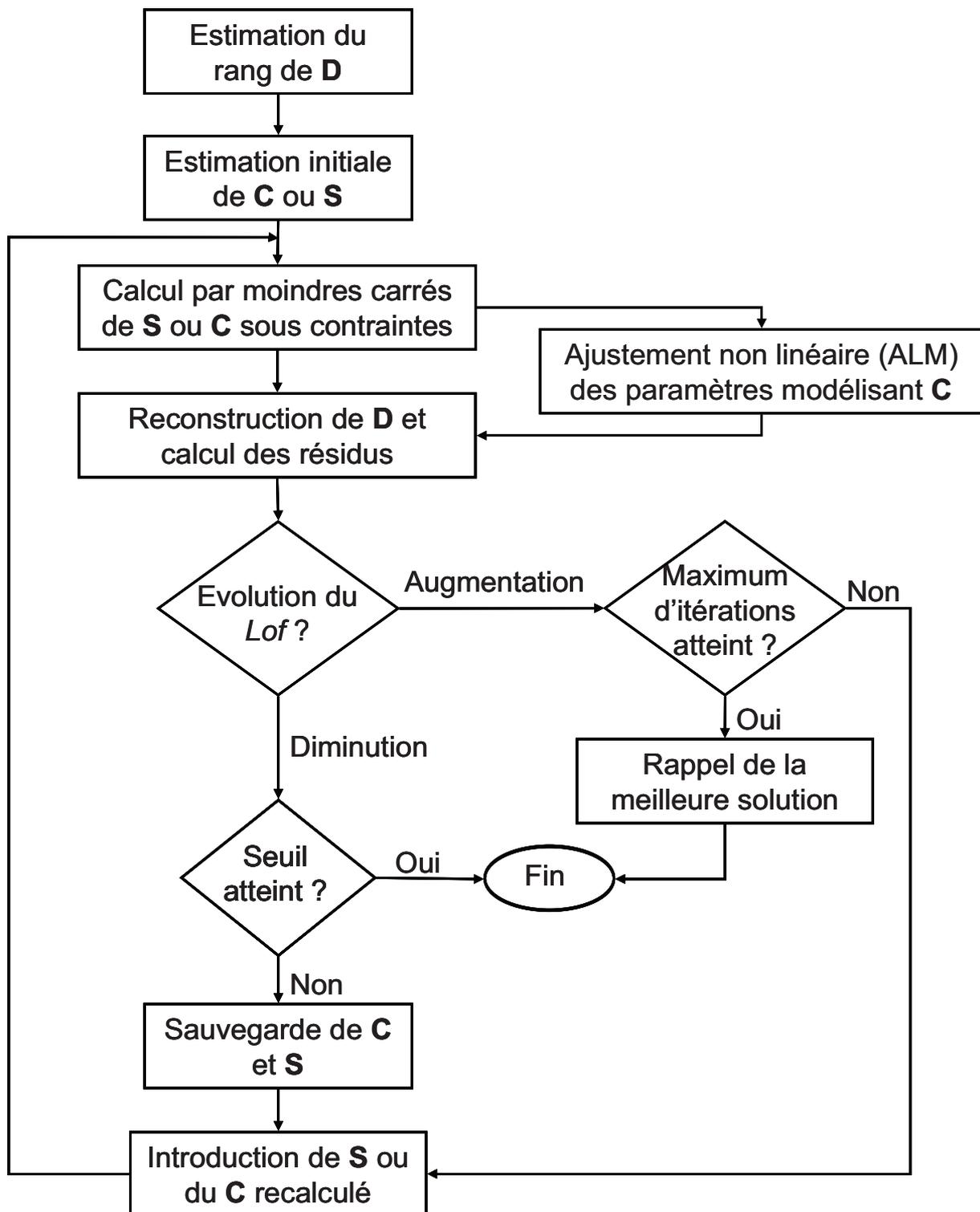


Figure 30 : Schéma de principe de HS-MCR

Deux critères d'optimisation de la résolution sont utilisés. Le premier ne concerne que l'optimisation des paramètres du modèle cinétique. Au lieu de minimiser la différence entre la matrice de données \mathbf{D} et la matrice de données reconstruite, on cherche maintenant à minimiser la différence entre les profils de concentration \mathbf{C}_{ALS} obtenus par la partie *soft modelling* de l'algorithme et les profils de concentration \mathbf{C}_{kin} reconstruits en utilisant les lois cinétiques du modèle choisi. Ceci peut être résumé par l'équation suivante :

$$\min \|\mathbf{C}_{\text{ALS}} - \mathbf{C}_{\text{kin}}\| \quad (36)$$

Seules les colonnes prises en compte par le modèle cinétique sont utilisées durant cette minimisation, comme le montre la Figure 31.

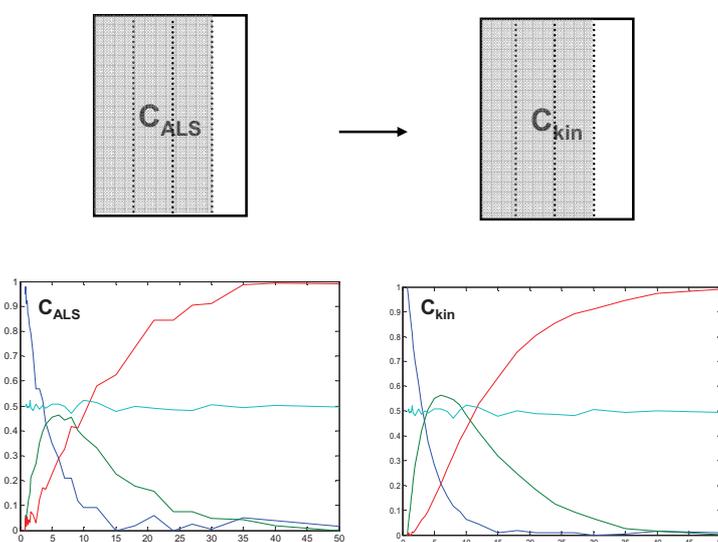


Figure 31 : Représentation de l'effet de la contrainte cinétique sur les profils de concentrations, la zone en gris correspond aux profils participant au modèle physico chimique.

Le second critère concerne la boucle générale de l'algorithme, c'est en fait le même que celui de MCR-ALS : on évalue le *Lof*. Cette évaluation s'effectue en utilisant une matrice de données reconstruite à partir de \mathbf{C} et de \mathbf{S} . La matrice des spectres est donc directement impliquée dans l'optimisation, ce qui n'est pas le cas en modélisation rigide.

L'utilisation de cet algorithme pour l'analyse de matrices augmentées présente un avantage supplémentaire. En effet grâce à la boucle de MCR-ALS l'information contenue dans chacune des expériences peut être mise en commun non seulement grâce à un modèle physicochimique global (comme en modélisation rigide par le modèle cinétique) mais aussi par l'utilisation active de la seule matrice \mathbf{S} commune à toutes les expériences analysées (voir Figure 7). En réalité, l'existence d'un seul de ces liens est nécessaire pour réaliser

efficacement l'analyse simultanée des expériences. Il est alors possible d'analyser plusieurs expériences ne partageant pas la même cinétique mais impliquant les mêmes espèces chimiques, c'est-à-dire d'appliquer des modèles locaux ou même d'analyser simultanément des données qui obéissent un modèle physicochimique avec des données qui ne suivent pas aucun modèle ou dont le modèle est inconnu.

Les avantages de cette méthode hybride peuvent être évalués face à ses deux méthodes parentes. Tout d'abord en comparaison à MCR-ALS :

- les solutions fournies sont bien moins affectées par les ambiguïtés de rotation ;
- la présence d'un modèle permet de prendre en compte d'éventuelles contributions transparentes dans le domaine spectral, ce qui est impossible dans MCR-ALS. En effet, en modélisation souple, chaque contribution doit être associée à un signal détectable ;
- les paramètres physico-chimiques d'intérêt tels que les constantes de réactions du procédé étudié sont obtenus.

Les avantages de HS-MCR par rapport aux méthodes de modélisation rigides sont les suivants :

- il est possible de ne soumettre aux contraintes cinétiques qu'une partie des espèces présentes et donc de laisser certaines espèces hors du modèle cinétique ; on peut donc modéliser des systèmes en réaction en présence d'interférents qui contribuent au signal enregistré.
- les spectres purs sont contraints de façon explicite (par exemple à la non négativité), ce qui permet de supprimer l'ambiguïté sur les constantes de réaction pour un procédé suivant plusieurs étapes successives. Il est en effet possible que les constantes obtenues soit inversées par rapport à l'ordre des réactions on parle d'ambiguïté « *slow fast* ») [99].
- La gestion d'une augmentation de matrices permet d'appliquer un modèle commun à toutes les expériences mais aussi d'appliquer des modèles individuels pour chaque matrice [100] ou de laisser des expériences sans aucun modèle.

6. Conclusion

Les méthodes de modélisation souple et rigide représentent les deux principaux outils pour l'interprétation des procédés suivis par spectroscopie résolue en temps. Chacun d'entre eux a été appliqué avec succès à ce type de problématique mais chacun présente également ses propres inconvénients. La résolution de courbes multivariées basée sur la méthode hybride HS-MCR semble très prometteuse car elle permet de tirer profit du meilleur des deux méthodes (souple et rigide). Celle-ci sera utilisée dans la suite de cette thèse pour analyser diverses données expérimentales. Nous reviendrons sur ces applications dans les chapitres 4 et 5. Cependant certains développements méthodologiques sont nécessaires. En effet, les données spectroscopiques peuvent avoir des structures particulières, il peut s'agir par exemple de spectres de différence. Ces données impliquent certaines difficultés spécifiques qui demandent une réponse adaptée. Nous détaillerons donc dans le chapitre suivant une approche originale permettant de modéliser correctement les données de spectroscopie de différence.

Chapitre II : Développements méthodologiques pour la spectroscopie résolue en temps

La spectroscopie de différence met en avant les variations les plus faibles du signal reflétant l'évolution de l'échantillon ou du procédé étudié, mais induit inévitablement une perte d'information. Il est nécessaire de caractériser les conséquences de la structure différentielle des données sur le nombre d'espèces détectables (rang) et sur l'approche à utiliser pour l'analyse et l'interprétation des résultats de ce genre de mesures. En spectroscopie de différence, l'état initial du système ne peut plus être caractérisé directement. D'un point de vue chimiométrique, cette perte d'information se transforme en un problème de déficience de rang. Cette dernière peut éventuellement s'ajouter à d'autres déficiences de rang propres au système chimique réactionnel. C'est notamment le cas lorsque plusieurs espèces évoluent de façon parallèle [101]. Une analyse basée sur une augmentation de matrices peut souvent lever la déficience de rang liée au schéma réactionnel [31, 40, 55] mais, pour celle inhérente à la spectroscopie de différence, une autre approche doit être utilisée.

Nous envisagerons donc dans un premier temps différents cas possibles correspondant à diverses situations communes en suivi de procédés. Chacune d'entre elles sera détaillée d'un point de vue théorique puis des solutions, basées sur l'utilisation et le développement de l'algorithme hybride de modélisation rigide et flexible HS-MCR, seront proposées pour s'affranchir des problèmes inhérents aux données de différence.

1. Déficience de rang induite par la structure des données de différence

Tout d'abord il est nécessaire de définir de façon rigoureuse l'influence sur le rang de la structure des données de différence. Une description théorique de l'effet de la spectroscopie de différence sur les données sera fournie dans ce chapitre. Ce travail a été publié en 2007 [102]. Cette problématique est à mettre en relation avec le travail de Amrhein *et al.* [29] concernant les phénomènes de déficience de rang en suivi de procédés causés par les schémas réactionnels. Dans ce dernier cas, les auteurs du travail [6] ont en effet montré que le rang d'une matrice de données **D** peut être estimé pour des systèmes clos en spectroscopie directe selon l'équation (13).

$$\text{rang}(\mathbf{D}) = \min(\text{nombre de réactions} + 1, \text{nombre d'espèces absorbantes}) \quad (13)$$

Il est alors possible de distinguer les systèmes déficients de rang à cause des réactions qui s'y déroulent et les données déficientes de rang à cause de la nature de la mesure enregistrée. En général, un jeu de données est déficient de rang si le rang estimé par l'équation est inférieur au nombre réel d'espèces. On peut trouver des données qui sont déficientes de rang à cause de la mesure, à cause des réactions ou qui sont affectées simultanément par les deux types de problèmes.

Des exemples variés ont été considérés ; ils sont représentatifs de l'ensemble des situations possibles en suivi temporel de procédés chimiques en spectroscopie de différence. Ce travail formel est un préalable nécessaire à la bonne interprétation des résolutions obtenues et surtout au développement d'une approche chimiométrique hybride mêlant résolution souple et ajustement cinétique. Trois situations seront considérées, elles serviront de base à l'extension de notre démonstration à des systèmes réels plus complexes (voir chapitre 3). Tout au long de cette étude nous considérons des systèmes chimiques clos, ceux-ci étant les plus courants. Les deux premiers concernent des schémas réactionnels qui induisent un rang plein (Cas 1) et en présence d'un interférent (Cas 2). Ils seront complètement détaillés et illustrés. Le Cas 3 se rapporte à l'étude de données de différence résultant d'un schéma réactionnel induisant une déficience de rang.

2. CAS 1: données simulées avec un schéma réactionnel de rang plein

2.1 Estimation théorique du rang

Prenons le cas simple d'une réaction cinétique de premier ordre en deux étapes successives $A \xrightarrow{k_1} B \xrightarrow{k_2} C$. Le système défini par ce modèle est de rang plein. En effet l'équation (13) donne dans ce cas un rang égal à 3, ce qui correspond au nombre d'espèces présentes. La concentration totale c_T peut être écrite, lorsque le système est clos, comme la somme des concentrations de tous les composés du procédé.

$$c_T = c_{A,i} + c_{B,i} + c_{C,i} \quad (37)$$

où $c_{A,i}$, $c_{B,i}$ et $c_{C,i}$ représentent les concentrations des espèces A, B et C pour chaque point i de l'avancement de la réaction.

Les données bilinéaires peuvent se décomposer en contributions de concentrations et spectrales comme on l'a vu au chapitre précédent. L'équation (38) décrit alors le premier spectre \mathbf{d}_0 enregistré durant cette réaction selon la loi de Beer-Lambert.

$$\mathbf{d}_0 = c_{A,0} \cdot \mathbf{s}_A + c_{B,0} \cdot \mathbf{s}_B + c_{C,0} \cdot \mathbf{s}_C \quad (38)$$

où \mathbf{s}_A , \mathbf{s}_B et \mathbf{s}_C représentent les spectres purs de A, B et C, et $c_{A,0}$, $c_{B,0}$ et $c_{C,0}$ les concentrations des différentes espèces au temps $t = 0$.

Les données de différence correspondent comme leur nom l'indique à la différence entre deux spectres : l'un de référence et l'autre mesuré à un instant i . Les équations suivantes décrivent le spectre de différence enregistré au temps i .

$$\Delta \mathbf{d}_i = \mathbf{d}_i - \mathbf{d}_0 \quad (39)$$

$$\Delta \mathbf{d}_i = (c_{A,i} \cdot \mathbf{s}_A + c_{B,i} \cdot \mathbf{s}_B + c_{C,i} \cdot \mathbf{s}_C) - (c_{A,0} \cdot \mathbf{s}_A + c_{B,0} \cdot \mathbf{s}_B + c_{C,0} \cdot \mathbf{s}_C) \quad (40)$$

$$\Delta \mathbf{d}_i = (c_T - c_{B,i} - c_{C,i}) \cdot \mathbf{s}_A + c_{B,i} \cdot \mathbf{s}_B + c_{C,i} \cdot \mathbf{s}_C - (c_T - c_{B,0} - c_{C,0}) \cdot \mathbf{s}_A - c_{B,0} \cdot \mathbf{s}_B - c_{C,0} \cdot \mathbf{s}_C \quad (41)$$

$$\Delta \mathbf{d}_i = (c_{B,i} - c_{B,0}) \cdot (\mathbf{s}_B - \mathbf{s}_A) + (c_{C,i} - c_{C,0}) \cdot (\mathbf{s}_C - \mathbf{s}_A) \quad (42)$$

En notation matricielle, l'équation (42) devient :

$$\Delta \mathbf{D} = [\mathbf{c}_B - c_{B,0} \quad \mathbf{c}_C - c_{C,0}] \begin{bmatrix} \mathbf{s}_B - \mathbf{s}_A \\ \mathbf{s}_C - \mathbf{s}_A \end{bmatrix} \quad (43)$$

Dans cette équation, la matrice des spectres de différence, $\Delta \mathbf{D}$, est décrite par deux contributions, produits des profils de concentration, avec la même forme que \mathbf{c}_B et \mathbf{c}_C , et de deux spectres de différence, $(\mathbf{s}_B - \mathbf{s}_A)$ et $(\mathbf{s}_C - \mathbf{s}_A)$. Ceux-ci relient les composés B et C au composé A, présent dès le début de réaction. La première conclusion est que le rang de la matrice de spectres de différence, égal à deux, ne respecte plus l'équation 1. Le rang détecté est donc inférieur au nombre réel d'espèces chimiques. On rappelle que le rang de la matrice de données directes suivant le même mécanisme réactionnel est en accord avec le nombre d'espèces.

Le raisonnement exposé ci dessus peut être étendu à des procédés constitués de réactions successives, comptant un plus grand nombre d'espèces (N), comme le montrent les équations (44) et (45). Il est ainsi démontré que le suivi de procédés de rang plein par spectroscopie de différence génère des matrices de données déficientes de rang.

$$\Delta \mathbf{d}_i = (c_{B,i} - c_{B,0})(\mathbf{s}_B - \mathbf{s}_A) + (c_{C,i} - c_{C,0})(\mathbf{s}_C - \mathbf{s}_A) + \dots + (c_{N,i} - c_{N,0})(\mathbf{s}_N - \mathbf{s}_A) \quad (44)$$

$$\Delta \mathbf{D} = [\mathbf{c}_B - c_{B,0} \quad \mathbf{c}_C - c_{C,0} \dots \mathbf{c}_N - c_{N,0}] \begin{bmatrix} \mathbf{s}_B - \mathbf{s}_A \\ \mathbf{s}_C - \mathbf{s}_A \\ \vdots \\ \mathbf{s}_N - \mathbf{s}_A \end{bmatrix} \quad (45)$$

Nous avons montré pour le cas d'une réaction chimique de premier ordre et, par extension, pour n'importe quel schéma réactionnel induisant un rang plein en spectroscopie directe, que les données de différence en spectroscopie sont nécessairement déficientes de rang. En données de différence, l'équation (13) permettant d'estimer le rang d'une matrice de données doit donc être réécrite comme suit :

$$\text{rang}(\mathbf{D}) = \min(\text{nombre de réactions}, \text{nombre d'espèces absorbantes} - 1) \quad (46)$$

Ce résultat, qui paraît intuitif, devait être validé de façon rigoureuse avant de poursuivre l'analyse des données de différence. Il semble en effet naturel qu'en soustrayant

un spectre à l'ensemble de la matrice une partie de l'information disparaisse. Cependant son effet sur le rang aurait pu être différent et/ou moins clair. D'autre part nous avons pu redémontrer que la forme des profils de concentration des différentes espèces n'est pas affectée par cette opération. Cet aspect est d'un grand intérêt lorsqu'il est possible d'appliquer un modèle cinétique.

L'approche théorique exposée précédemment fournit des indications quant à l'interprétation à donner aux résolutions chimiométriques obtenues sur les données de différence. La simulation proposée dans le paragraphe suivant permet d'illustrer ces conclusions et de présenter les méthodes mises en place pour compenser cette déficience de rang.

2.2 Données simulées

La matrice de données directes \mathbf{D}_1 est construite à partir de spectres réels UV-visible et de profils de concentration construits en utilisant un modèle cinétique d'ordre 1, $A \xrightarrow{k_1} B \xrightarrow{k_2} C$. Les constantes de réaction utilisées sont $k_1 = 0,19 \text{ s}^{-1}$ et $k_2 = 0,11 \text{ s}^{-1}$. Chaque matrice est constituée de 30 spectres composés de 301 absorbances. Un bruit de nature homoscedastique représentant 1,5% du signal maximum a été ajouté.

La matrice $\Delta\mathbf{D}_1$ est calculée en soustrayant le premier spectre de \mathbf{D}_1 (c'est-à-dire le spectre de l'espèce A) à toutes les lignes de \mathbf{D}_1 . Ces deux matrices sont présentées dans la Figure 32. Il est intéressant de comparer les deux matrices. Les bandes positives caractéristiques des espèces formées (B et C), qui apparaissent, sont reconnaissables aussi bien dans \mathbf{D}_1 que dans $\Delta\mathbf{D}_1$, à 265 ou 360 nm par exemple. La partie négative des spectres de différence peut quant à elle être reliée aux bandes caractéristiques de l'espèce A, qui disparaît, notamment entre 290 et 320 nm.

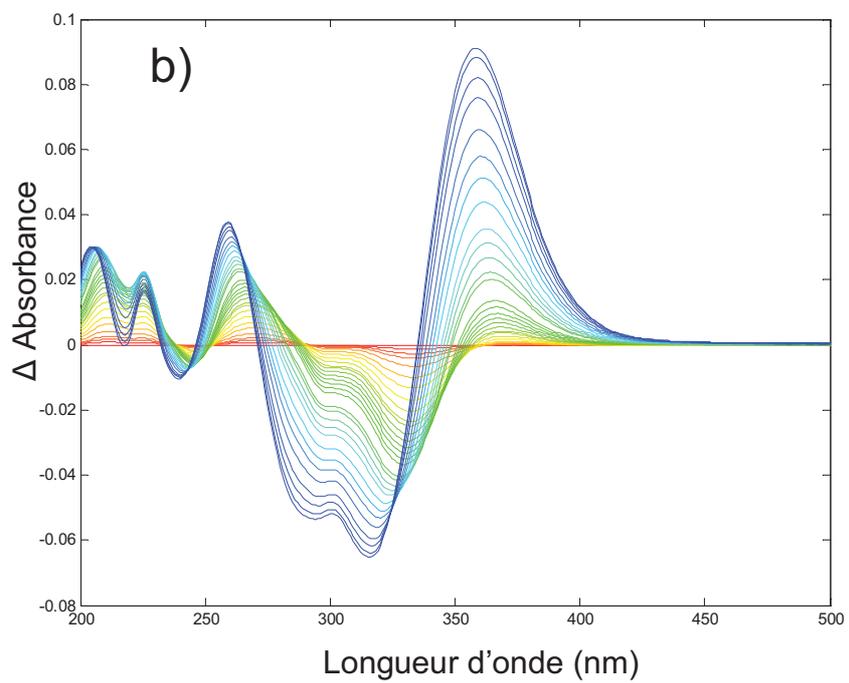
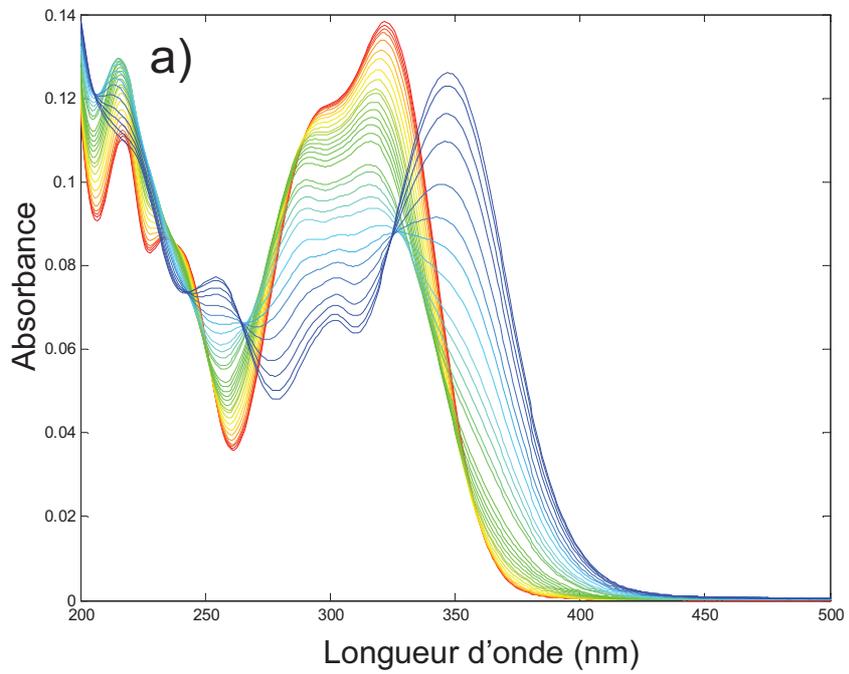


Figure 32 : Matrices de données simulées a) D_1 b) ΔD_1 .

2.3 Estimation du rang par SVD

La décomposition en valeurs singulières (SVD) est généralement utilisée pour estimer le rang d'une matrice de données [21]. Les premières valeurs singulières de \mathbf{D}_1 et $\Delta\mathbf{D}_1$ sont présentées dans le Tableau 2.

\mathbf{D}_1	$\Delta\mathbf{D}_1$
62,939	16,308
11,385	0,6186
0.3981	0,0226
0,0216	0,0216
0,0213	0,0211
0,0206	0,0202
0,0203	0,0199

Tableau 2 : Premières valeurs singulières de \mathbf{D}_1 et $\Delta\mathbf{D}_1$ (les valeurs en gras représentent les valeurs singulières associées au signal).

On note dans ce tableau que pour \mathbf{D}_1 trois valeurs singulières (en gras) se distinguent significativement des suivantes, représentant le bruit. Le rang de la matrice \mathbf{D}_1 est estimé à trois et correspond au rang théorique de cette simulation. Ceci est cohérent avec les profils cinétiques issus du modèle $A \xrightarrow{k_1} B \xrightarrow{k_2} C$.

L'analyse des valeurs singulières de $\Delta\mathbf{D}_1$ ne permet de distinguer que deux valeurs singulières significativement différentes du bruit. Le rang estimé n'est donc que de deux, même si trois espèces sont mises en jeu dans ce procédé. La matrice $\Delta\mathbf{D}_1$ est donc déficiente de rang, ce qui est en accord avec l'approche mathématique présentée précédemment.

2.4 Analyse factorielle évolutive (EFA)

L'analyse factorielle évolutive (EFA) [33] permet d'obtenir des informations additionnelles sur le rang des données de spectroscopie directe ou de différence; en effet, l'EFA nous montre l'émergence et la disparition des espèces détectables dans notre jeu de données. Les résultats pour \mathbf{D}_1 sont présentés Figure 33. Trois profils émergent clairement du niveau de bruit dans le sens direct et dans le sens inverse. Ils permettent de reconstruire une estimation correcte (non présentée ici) des profils de concentration utilisés pour construire cette simulation [13, 21]. On note clairement qu'il y a une espèce présente en proportion importante dès le début du procédé (ligne continue bleue).

En comparaison, l'EFA de $\Delta\mathbf{D}_1$ ne présente que deux profils significatifs, comme il est possible d'observer Figure 34. Ceci est cohérent avec l'analyse SVD. Il est cependant intéressant de constater que les profils obtenus traduisent le fait que le spectre au premier temps ($t = 0$) est nul. Les deux autres contributions émergent aux mêmes points sur l'axe x qui correspond à la variable liée à l'évolution du procédé (pour nous le temps) dans les deux figures. Il est aussi possible d'observer que la ligne qui indiquerait la disparition de la première espèce (ligne pointillée bleue de la Figure 2) est absente. Il est intéressant de noter ici que les valeurs propres sont plus faibles que dans le cas de l'analyse de \mathbf{D}_1 . Ceci correspond à la réduction du rapport signal sur bruit. Le bruit reste en effet équivalent entre les deux matrices \mathbf{D}_1 et $\Delta\mathbf{D}_1$, mais le signal de $\Delta\mathbf{D}_1$ est inévitablement plus faible.

L'information perdue en spectroscopie de différence concerne donc l'espèce initiale. Celle-ci est en quelque sorte transférée sur les autres profils qui sont reliés à des combinaisons de spectres, $(\mathbf{s}_B - \mathbf{s}_A)$ et $(\mathbf{s}_C - \mathbf{s}_A)$. Malgré ce fait, il est possible d'utiliser un modèle cinétique pour rendre compte du comportement de l'espèce A dans le procédé en utilisant l'information contenue dans l'évolution des deux contributions spectrales détectables.

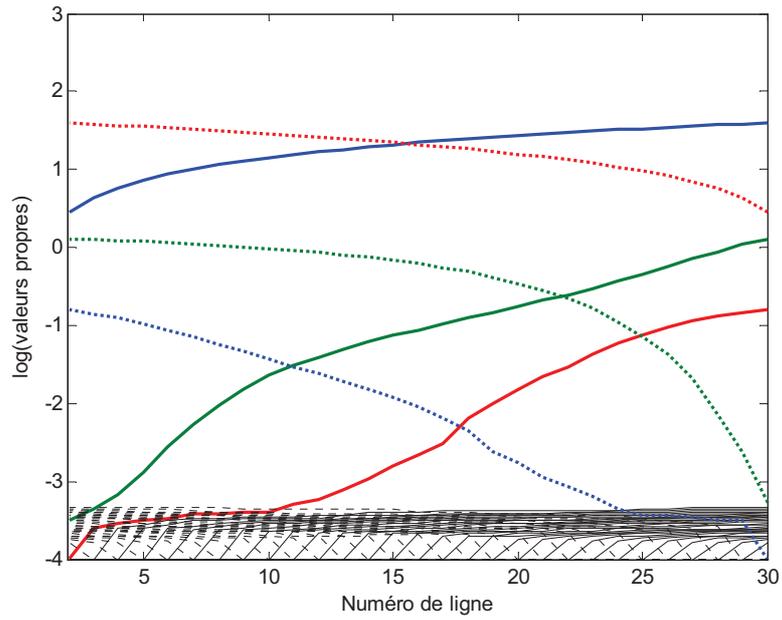


Figure 33 : Graphique de l'évolution des valeurs propres issu de l'EFA de D_1 .

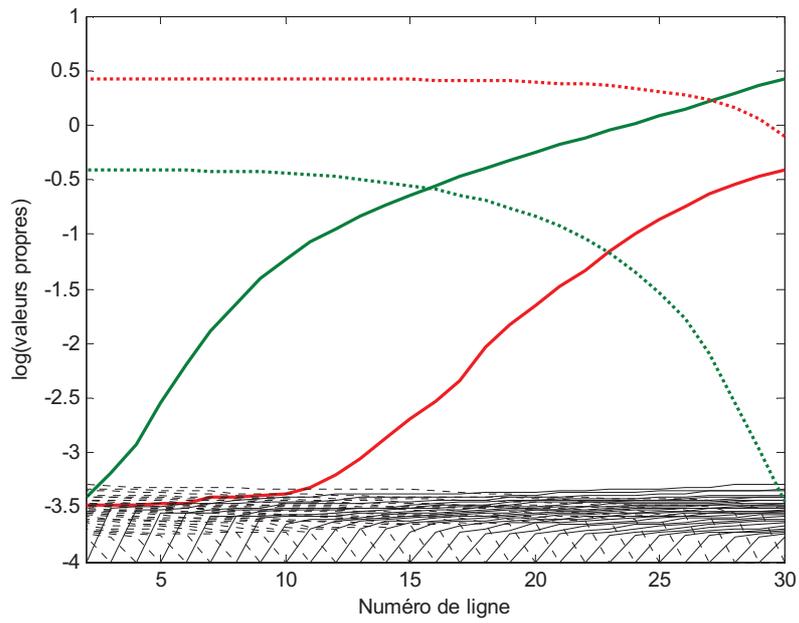


Figure 34 : Graphique de l'évolution des valeurs propres issu de l'EFA de ΔD_1 .

2.5 Résolution multivariée hybride

L'information concernant la (ou les) première(s) espèce(s) n'est pas directement accessible. Le profil de concentration \mathbf{c}_A de la première espèce ne peut pas être exprimé explicitement à partir des données brutes; néanmoins, la présence de cette espèce chimique influe nécessairement sur le procédé. Elle doit donc être prise en compte lors de l'interprétation des résolutions des spectres de différence et lors de l'établissement d'un modèle cinétique décrivant le procédé. En effet l'analyse correcte de ces données simulées nécessite de prendre en compte l'espèce A. À ce stade, il est important de noter que les profils de concentration $(\mathbf{c}_B - \mathbf{c}_{B,0})$ et $(\mathbf{c}_C - \mathbf{c}_{C,0})$ de l'équation (43) ont la même forme que \mathbf{c}_B et \mathbf{c}_C . Ils obéissent donc à la même cinétique que le système de départ, $A \xrightarrow{k_1} B \xrightarrow{k_2} C$, c'est-à-dire qu'ils évoluent selon les mêmes constantes cinétiques. Si l'on se réfère au fonctionnement des algorithmes des moindres carrés non-linéaires et de HS-MCR, présentés dans le chapitre précédant, on peut noter que l'ajustement cinétique ne s'effectue qu'en utilisant la matrice de profils de concentration \mathbf{C} . Il est donc techniquement possible d'ajouter un profil de concentration \mathbf{c}_A à la matrice \mathbf{C} uniquement dans le but d'effectuer un ajustement cinétique en accord avec le modèle cinétique complet sans qu'une forme spectrale y soit associée (on parle d'une espèce transparente). Une fois l'ajustement du modèle effectué on obtient les profils de concentration des différentes espèces visibles ainsi que celui correspondant à cette contribution supplémentaire. Les spectres purs sont ensuite obtenus par moindres carrés, directement ou sous contraintes souples en n'utilisant que les profils de concentration ayant une contribution sur le signal mesuré (dans ce cas, les profils de B et C). Les résultats finaux fournis par l'algorithme HS-MCR sont deux profils de concentration, $(\mathbf{c}_B - \mathbf{c}_{B,0})$ et $(\mathbf{c}_C - \mathbf{c}_{C,0})$, associés aux contributions spectrales $(\mathbf{s}_B - \mathbf{s}_A)$ et $(\mathbf{s}_C - \mathbf{s}_A)$ et un profil de concentration, \mathbf{c}_A , qui n'a pas de spectre associé.

L'introduction d'un modèle rigide avec une ou plusieurs espèces initiales transparentes dans l'algorithme HS-MCR permet contourner le problème de déficience de rang des données de différence. En effet, cela permet de reconstituer tous les profils de concentration du modèle cinétique du procédé et, par extension, les constantes cinétiques correctes. Pour illustrer ce résultat, la Figure 35 propose de comparer les profils de concentration issus des résolutions de \mathbf{D}_1 et $\Delta\mathbf{D}_1$. Les deux résolutions sont basées sur le modèle cinétique complet utilisé pour construire la simulation. La non négativité des concentrations est appliquée dans les deux cas.

Les *lack of fit* et R^2 obtenus sont respectivement pour \mathbf{D}_1 de 1,43% et 99,98% et pour $\Delta\mathbf{D}_1$ de 5,3% et 99,71%. Les profils de concentration de \mathbf{D}_1 (lignes pleines) correspondent aux profils simulés utilisés, avec un coefficient de corrélation supérieure à 0,99. Plus important, les profils de concentration de $\Delta\mathbf{D}_1$ (lignes en tirets) sont aussi cohérents avec \mathbf{D}_1 et la simulation. Ceci démontre que, dans le cas de schémas réactionnels induisant un rang plein, l'information cinétique est reconstruite par le modèle malgré la structure des données de différence. Le profil de concentration de l'espèce A est obtenu dans l'étape de l'ajustement cinétique (ligne en pointillés), mais c'est l'information concernant les espèces qui contribuent au signal qui dirige l'ajustement vers les valeurs correctes des constantes cinétiques du modèle. Tous les profils de concentration sont ainsi connus et les spectres purs peuvent être associés, indirectement, à chacune des espèces du procédé.

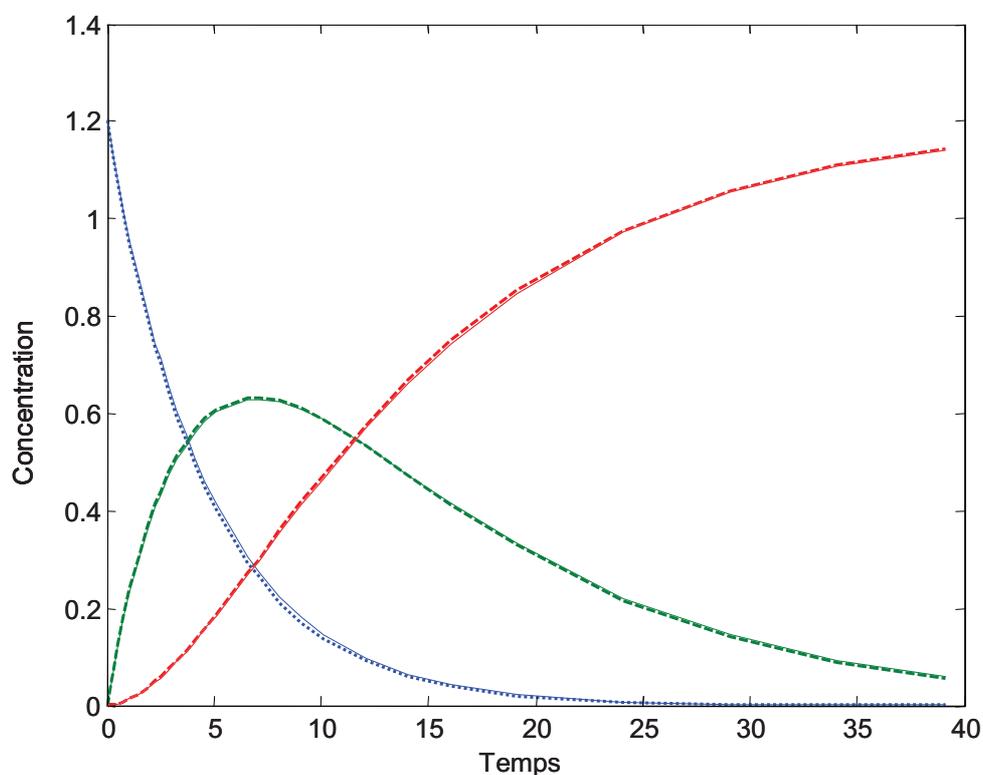


Figure 35 : Profils de concentration obtenus par résolution HS-MCR Les lignes pleines, en tirets et en pointillés représentent respectivement la résolution de \mathbf{D}_1 , la résolution de $\Delta\mathbf{D}_1$ et le profil reconstitué de la première espèce (transparente).

Le nombre et la forme des spectres résolus sont beaucoup plus influencés par la nature des données (spectroscopie directe ou de différence). Les spectres utilisés pour construire la simulation sont retrouvés correctement lors de la résolution de \mathbf{D}_1 , comme le montre la Figure 36a. Les spectres issus de la résolution de $\Delta\mathbf{D}_1$ doivent, d'après les calculs précédents, être

$(s_B - s_A)$ et $(s_C - s_A)$. La Figure 36b permet d'attester la validité de ces calculs et de l'approche chimiométrique ; en effet les profils spectraux (lignes en tirets) issus de la résolution de ΔD_1 correspondent parfaitement aux spectres de différence calculés à partir des spectres simulés des espèces A, B et C (lignes pleines).

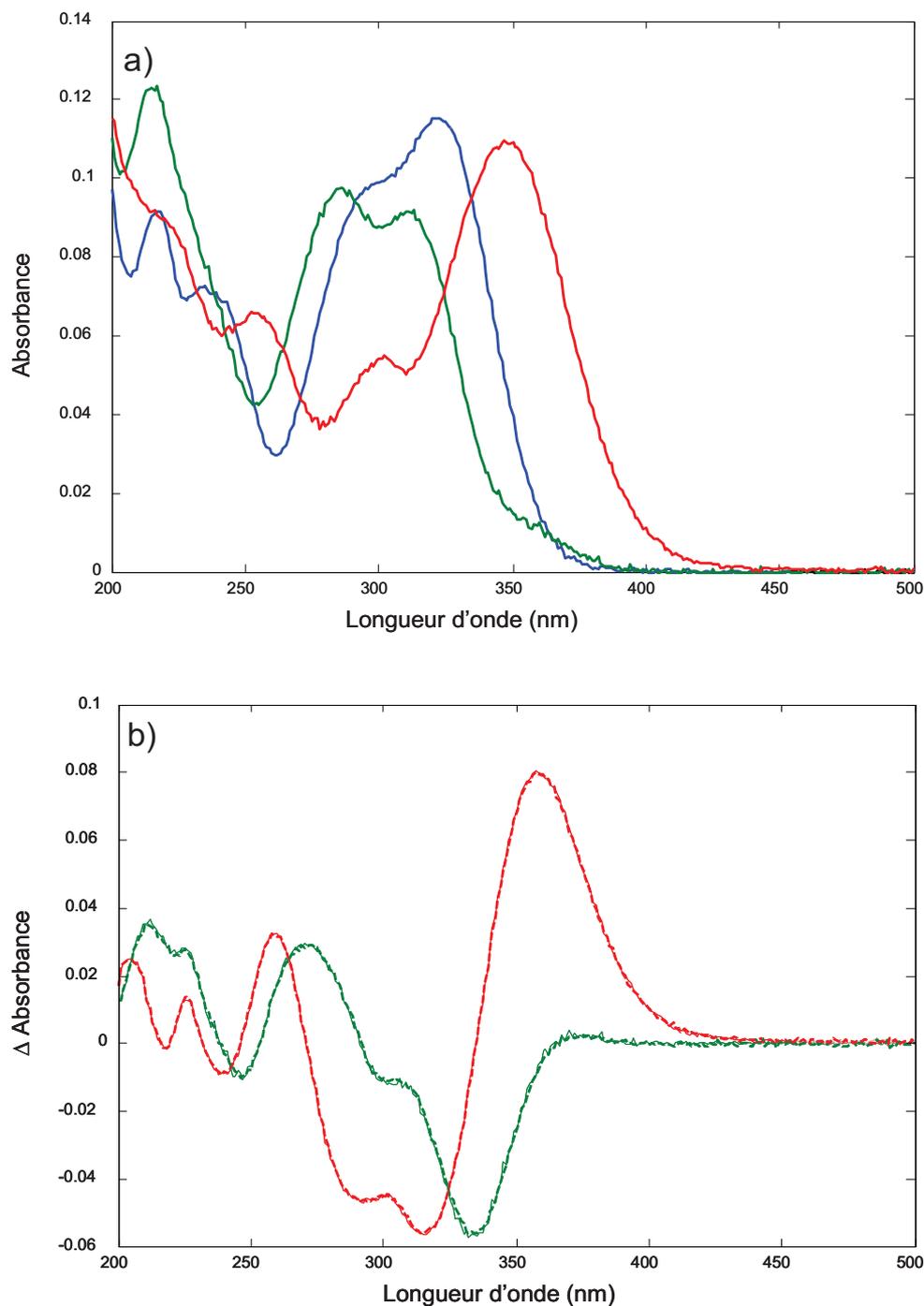


Figure 36 : Spectres purs obtenus par résolution HS-MCR a) pour D_1 , b) pour ΔD_1 représentés en tirets, les lignes pleines correspondent aux spectres de différence calculés à partir des spectres purs issus de la décomposition de D_1 .

La déficience de rang implicite aux données de différence peut donc être contournée (et non pas supprimée) par l'utilisation d'une contrainte cinétique qui incorpore des espèces initiales transparentes. Dans l'exemple présenté, nous avons choisi d'utiliser l'algorithme HS-MCR mais il faut noter que d'autres méthodes de modélisation cinétique auraient pu fonctionner. La seule condition est que l'algorithme doit prévoir la possibilité d'utiliser une ou plusieurs espèces transparentes dans le modèle. La situation n'est pas la même dans le cas de l'exemple suivant qui traite de la présence d'espèces interférentes ayant une signature spectrale. HS-MCR présente alors l'avantage unique de pouvoir gérer ce type d'espèce et de pouvoir encore utiliser une contrainte basée sur un modèle cinétique.

3. CAS 2 : données simulées avec un schéma réactionnel de rang plein affecté d'une contribution non cinétique

3.1 Estimation théorique du rang

Une contribution non cinétique au signal peut être considérée comme un interférent, qui ne doit pas être pris en compte dans un modèle cinétique puisqu'il évolue d'une façon complètement indépendante. Le CAS 2 est donc, du point de vue théorique, similaire au CAS 1. La seule différence réside dans la présence de cet interférent. Cet espèce reste exclue du raisonnement théorique sur le rang des procédés et est détectée comme une espèce additionnelle. Le développement proposé précédemment reste alors valide pour les autres espèces. L'intérêt de cette section est de démontrer que HS-MCR est capable de gérer des situations plus complexes, qui ne rentrent pas dans le cadre de la modélisation rigide classique.

3.2 Simulations de données

Les jeux de données \mathbf{D}_2 et $\Delta\mathbf{D}_2$ ont été simulés à partir de spectres proche infrarouge réels et de profils de concentration construits artificiellement en utilisant un modèle cinétique identique au précédent aux constantes de réactions près qui sont maintenant $k_1 = 0,21 \text{ ms}^{-1}$ et $k_2 = 0,09 \text{ ms}^{-1}$. Chaque matrice est constituée de 30 spectres de 100 longueurs d'onde chacun. Un bruit de nature homoscedastique a été ajouté, représentant en intensité 1% du signal maximum. La matrice $\Delta\mathbf{D}_2$ est calculée en soustrayant le premier spectre de \mathbf{D}_2 à toutes les lignes de la matrice \mathbf{D}_2 . La même démonstration que précédemment pourrait être faite sur cette simulation mais ici nous nous intéresserons à un cas plus complexe. Une contribution différentielle supplémentaire I est ajoutée à $\Delta\mathbf{D}_2$, elle joue le rôle d'un interférent ne contribuant pas à la cinétique du procédé. La matrice obtenue est nommée $\Delta\mathbf{D}_{2B}$. Les deux matrices de spectres de différence sont exposées dans la Figure 37. Cette nouvelle contribution ne suit pas de modèle cinétique. Il n'est donc pas possible d'utiliser les méthodes traditionnelles de modélisation rigide. Le bruit introduit étant connu il nous est possible de dire qu'une décomposition parfaite de ces deux matrices de données mène à un *Lof* de 9,29 % pour $\Delta\mathbf{D}_2$ et 11,43 % pour $\Delta\mathbf{D}_{2B}$.

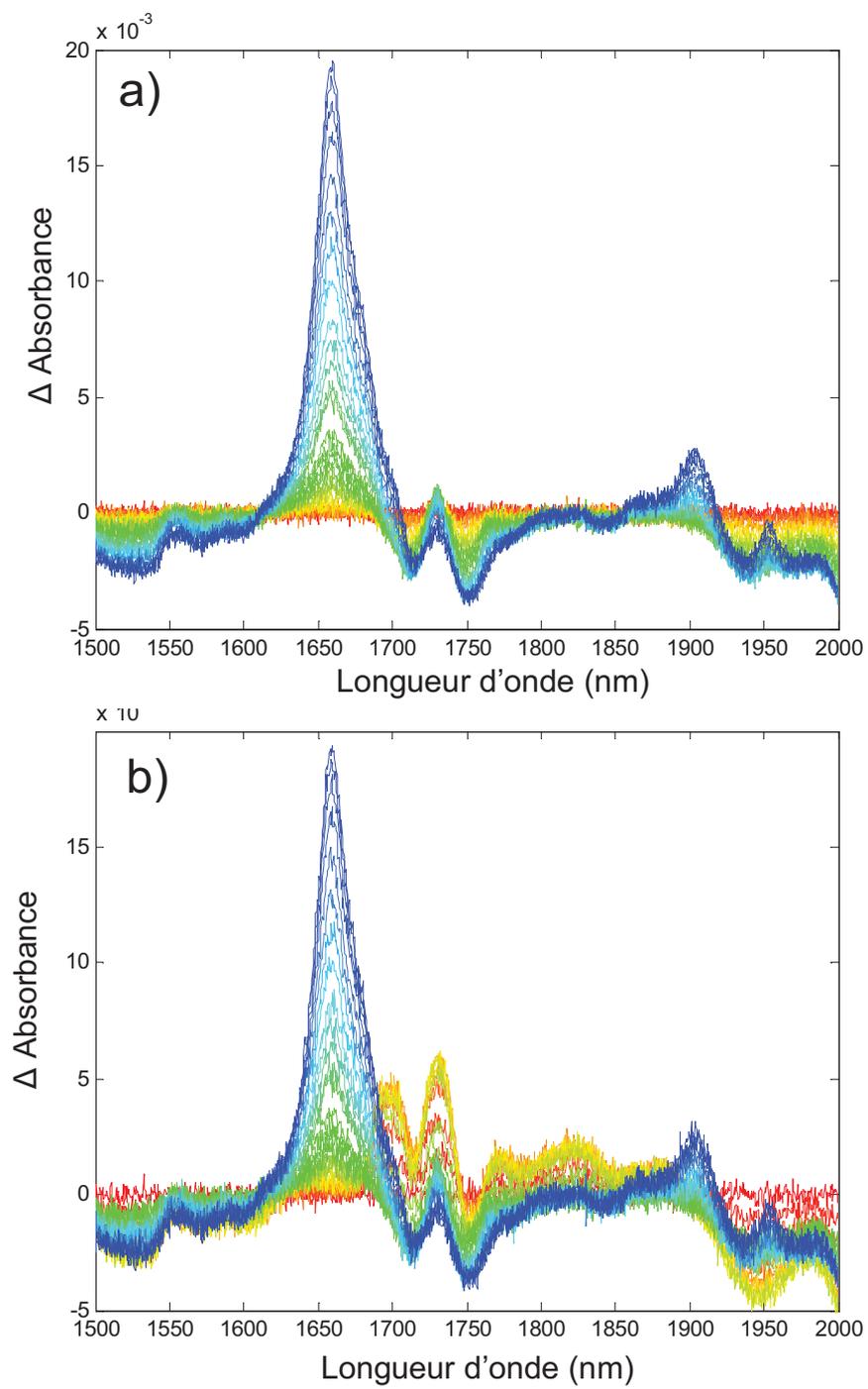


Figure 37 : Matrices de données simulées a) $\Delta \mathbf{D}_2$ b) $\Delta \mathbf{D}_{2B}$.

Les premières étapes de l'analyse restent les mêmes, il est nécessaire d'estimer le rang de la matrice et de construire une estimation initiale des profils de concentrations (ou des spectres purs). Les valeurs singulières de $\Delta\mathbf{D}_2$ et $\Delta\mathbf{D}_{2B}$ sont présentées dans le Tableau 3. La matrice $\Delta\mathbf{D}_2$ correspond au cas de figure précédent, on détecte donc un rang égal à deux, pour $\Delta\mathbf{D}_{2B}$ le rang est de trois. Cela traduit bien la détection d'une espèce supplémentaire, en accord avec l'observation possible à 1740 nm sur la Figure 37. La contribution de l'interfèrent significativement plus importante que le bruit est donc détectée lors de l'analyse SVD.

$\Delta\mathbf{D}_2$	$\Delta\mathbf{D}_{2B}$
1,6308	2,5013
0,6186	1,3136
0,0226	0,5662
0,0211	0,0326
0,0202	0,0273
0,0199	0,0263
0,0198	0,0259

Tableau 3 : Premières valeurs singulières de $\Delta\mathbf{D}_2$ et $\Delta\mathbf{D}_{2B}$.

L'EFA de ces deux matrices de données, présentée Figure 38, est cohérente avec la SVD. En effet pour $\Delta\mathbf{D}_2$ deux profils sont observés alors que pour $\Delta\mathbf{D}_{2B}$ trois profils émergent du niveau de bruit. Il est important de noter que l'ajout d'un interfèrent dans cette simulation peut perturber la construction d'une estimation initiale de \mathbf{C} . En effet, l'EFA postule lors de la reconstruction des profils de concentration que la première contribution à apparaître sera aussi la première à disparaître, ce n'est pas nécessairement le cas pour une contribution liée à un interfèrent.

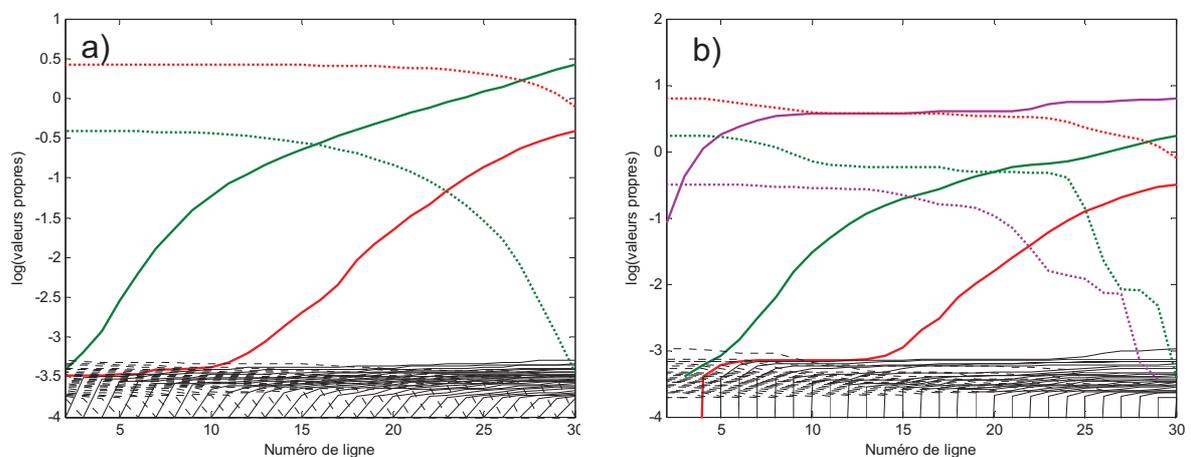


Figure 38 : EFA de a) $\Delta\mathbf{D}_2$ b) $\Delta\mathbf{D}_{2B}$

L'application d'un modèle cinétique global n'est pas possible dans le cas de $\Delta\mathbf{D}_{2B}$, en tout cas pas sur l'ensemble des profils de concentration. C'est ici que HS-MCR se distingue des méthodes rigides [103]. En effet les méthodes classiques de modélisation rigide travaillent à partir de la matrice de données \mathbf{D} , là où une approche hybride ajuste le modèle cinétique sur une estimation de \mathbf{C} . Ceci permet de sélectionner les espèces (ou colonnes de la matrice \mathbf{C}) entrant dans le modèle, et d'en laisser certaines en dehors [104]. De plus il est également possible de contraindre la matrice \mathbf{S} et donc d'utiliser explicitement la dimension spectrale. Les contraintes appliquées lors de la résolution HS-MCR sont résumées par la Figure 39.

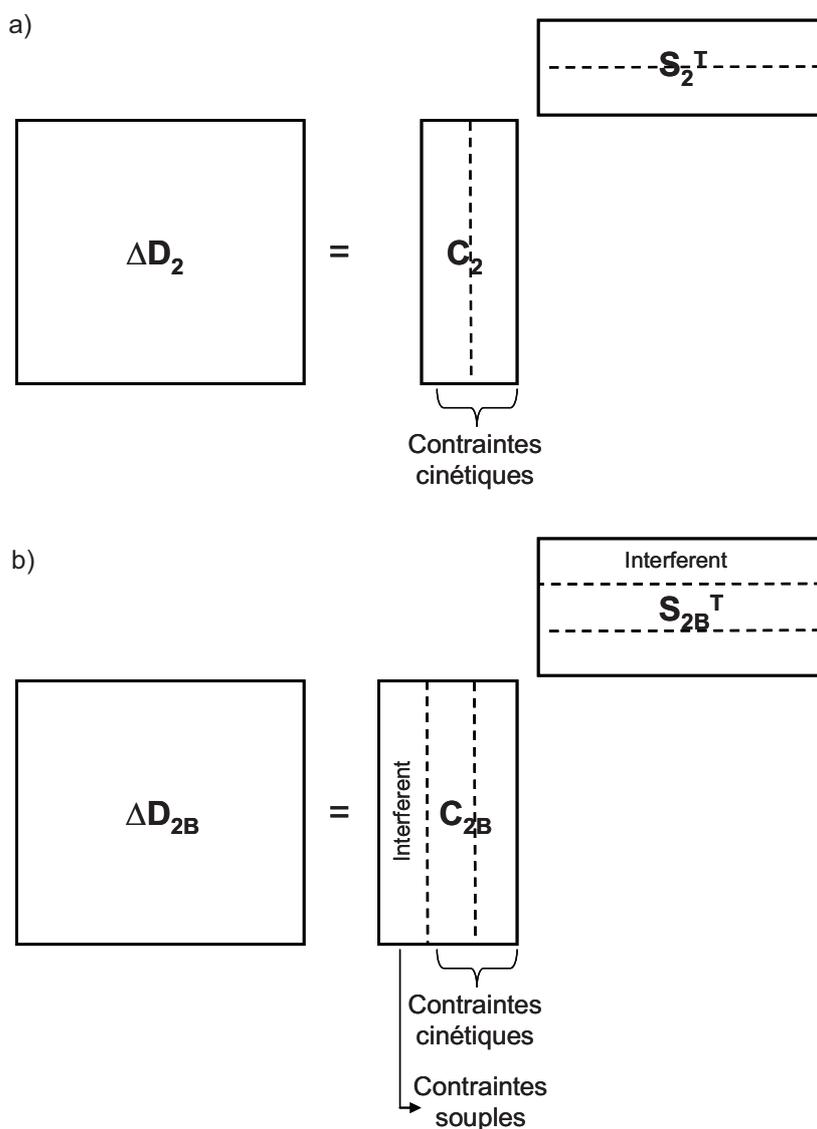


Figure 39 : Décompositions effectuées par HS-MCR des matrices a) $\Delta\mathbf{D}_2$ et b) $\Delta\mathbf{D}_{2B}$.

La résolution des jeux de données $\Delta\mathbf{D}_2$ et $\Delta\mathbf{D}_{2B}$ a donc été calculée en utilisant HS-MCR. Le même modèle cinétique $A \xrightarrow{k_1} B \xrightarrow{k_2} C$ a été utilisé pour les deux matrices mais dans le cas de $\Delta\mathbf{D}_{2B}$ une espèce supplémentaire est prise en compte lors des itérations sous contraintes souples, c'est-à-dire en dehors de la modélisation cinétique. La contrainte de non négativité des profils de concentrations a été appliquée. La résolution de $\Delta\mathbf{D}_2$ a un $Lof = 9,01\%$ et un $R^2 = 99,18\%$. Celle de $\Delta\mathbf{D}_{2B}$ a un $Lof = 11,12\%$ et un $R^2 = 98,76\%$, ces valeurs sont en accord avec les valeurs théoriques de Lof calculées lors de la construction des simulations. La Figure 40 présente les profils de concentrations et les spectres obtenus. On retrouve dans les deux situations les mêmes profils de concentrations pour les espèces A, B et C et les mêmes constantes de réaction $k_1 = 0,20 \text{ ms}^{-1}$ et $k_2 = 0,08 \text{ ms}^{-1}$. L'espèce supplémentaire est détectée et le profil de concentration résolu correspond à celui utilisé pour construire la simulation.

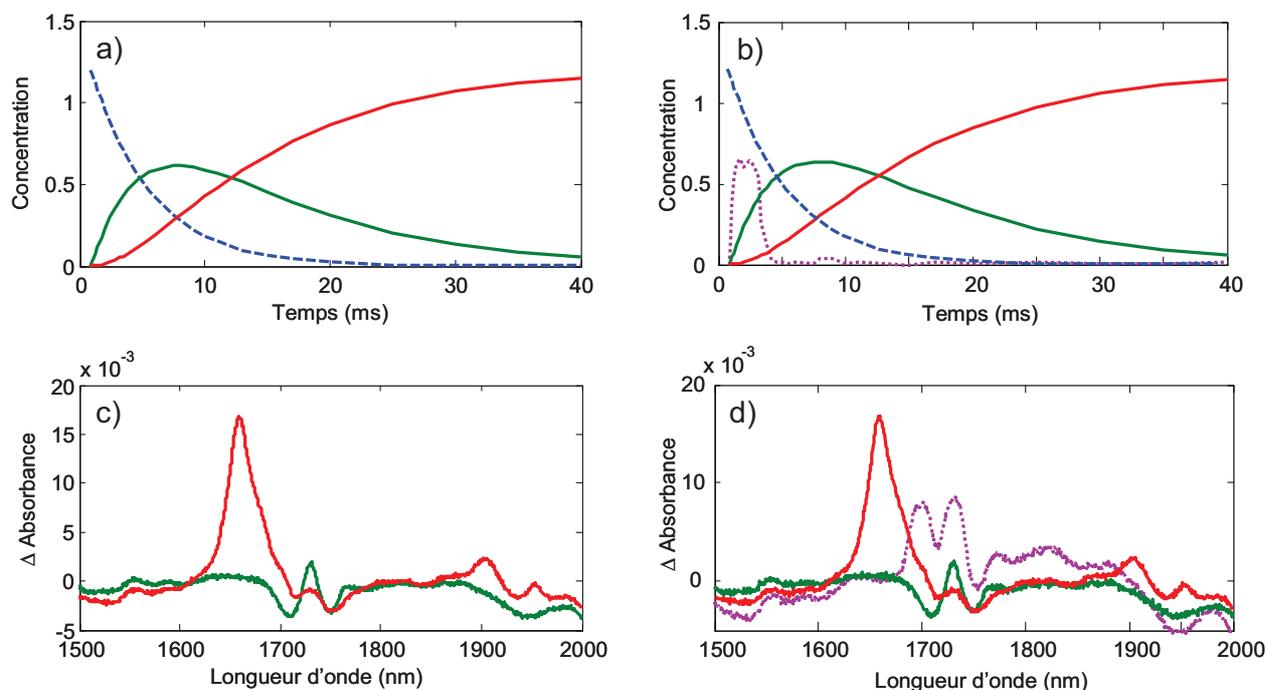


Figure 40 : Profils de concentration de a) $\Delta\mathbf{D}_2$ et b) $\Delta\mathbf{D}_{2B}$ et spectres purs correspondants c) $\Delta\mathbf{D}_2$ et d) $\Delta\mathbf{D}_{2B}$.

La qualité de l'information spectrale peut être estimée de la même manière, les spectres $(\mathbf{s}_B - \mathbf{s}_A)$ et $(\mathbf{s}_C - \mathbf{s}_A)$ sont identiques dans les deux cas, et le spectre de différence de l'interfèrent I est conforme à celui utilisé pour la simulation. Ce résultat a été obtenu en n'utilisant que les estimations initiales de l'EFA et le modèle cinétique partiel, le spectre de I n'a jamais été introduit dans la résolution. La présence de cet interfèrent ne perturbe donc pas la résolution si sa présence est connue et donc prise en compte lors de l'application des

contraintes dans HS-MCR. L'application du modèle cinétique permet encore dans ces conditions de compenser les spécificités structurelles des données de différence.

Deux obstacles à la résolution de données de différence peuvent donc être contournés en utilisant HS-MCR : la déficience de rang inhérente à ce type de données et la présence d'espèces interférentes. La question se pose alors de l'effet de la spectroscopie de différence dans le cas d'un procédé qui suit un schéma réactionnel déjà déficient de rang, tel qu'un système avec des réactions parallèles.

4. CAS 3 : données simulées avec un schéma réactionnel induisant une déficience de rang

Un exemple de schéma réactionnel induisant une déficience de rang peut être rencontré dans un procédé où deux réactions chimiques $A_1 \rightarrow A_2$ et $B_1 \rightarrow B_2$ se déroulent simultanément. L'équation (13) proposée par Amrhein conduit à une valeur du rang égal à 3 (d'après le terme "deux réactions +1") alors que quatre espèces sont présentes.

$$\text{rang}(\mathbf{D}) = \min(\text{nombre de réactions} + 1, \text{nombre d'espèces absorbantes}) \quad (13)$$

Ce type de procédé ne peut donc générer que des données déficientes de rang, même pour des données que ne seraient pas en différence. Si on reste dans l'hypothèse que les deux systèmes sont clos, les équations (47) et (48) peuvent être écrites.

$$c_{T_A} = c_{A_1} + c_{A_2} \quad (47)$$

$$c_{T_B} = c_{B_1} + c_{B_2} \quad (48)$$

De plus, lorsque les deux systèmes clos coexistent, une relation de proportionnalité entre les concentrations totales de chaque système c_{T_A} et c_{T_B} est vérifiée

$$c_{T_A} = k \cdot c_{T_B} \quad (49)$$

La loi de Beer-Lambert permet comme précédemment de décrire n'importe quel spectre \mathbf{d}_i à partir des concentrations, c_{A_1} , c_{A_2} , c_{B_1} et c_{B_2} , et des spectres purs \mathbf{s}_{A_1} , \mathbf{s}_{A_2} , \mathbf{s}_{B_1} et \mathbf{s}_{B_2} correspondants. Le premier spectre \mathbf{d}_0 s'écrit alors :

$$\mathbf{d}_0 = c_{A_1,0} \cdot \mathbf{s}_{A_1} + c_{A_2,0} \cdot \mathbf{s}_{A_2} + c_{B_1,0} \cdot \mathbf{s}_{B_1} + c_{B_2,0} \cdot \mathbf{s}_{B_2} \quad (50)$$

$$\mathbf{d}_0 = (k \cdot c_{T_B} - c_{A_2,0}) \cdot \mathbf{s}_{A_1} + c_{A_2,0} \cdot \mathbf{s}_{A_2} + (c_{T_B} - c_{B_2,0}) \cdot \mathbf{s}_{B_1} + c_{B_2,0} \cdot \mathbf{s}_{B_2} \quad (51)$$

Les spectres de différence correspondants, $\Delta \mathbf{d}_i$, peuvent alors être écrits de la forme suivante :

$$\Delta \mathbf{d}_i = ((k \cdot c_{T_B} - c_{A_2,i}) \cdot \mathbf{s}_{A_1} + c_{A_2,i} \cdot \mathbf{s}_{A_2} + (c_{T_B} - c_{B_2,i}) \cdot \mathbf{s}_{B_1} + c_{B_2,i} \cdot \mathbf{s}_{B_2}) - ((k \cdot c_{T_B} - c_{A_2,0}) \cdot \mathbf{s}_{A_1} + c_{A_2,0} \cdot \mathbf{s}_{A_2} + (c_{T_B} - c_{B_2,0}) \cdot \mathbf{s}_{B_1} + c_{B_2,0} \cdot \mathbf{s}_{B_2}) \quad (52)$$

$$\Delta \mathbf{d}_i = (c_{A_2,i} - c_{A_2,0}) \cdot (\mathbf{s}_{A_2} - \mathbf{s}_{A_1}) + (c_{B_2,i} - c_{B_2,0}) \cdot (\mathbf{s}_{B_2} - \mathbf{s}_{B_1}) \quad (53)$$

Ou encore, sous forme matricielle:

$$\Delta \mathbf{D} = \begin{bmatrix} c_{A_2} - c_{A_2,0} & c_{B_2} - c_{B_2,0} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{A_2} - \mathbf{s}_{A_1} \\ \mathbf{s}_{B_2} - \mathbf{s}_{B_1} \end{bmatrix} \quad (54)$$

On observe dans ce cas de procédé impliquant une déficience de rang que le rang des données de différence est diminué d'une unité par rapport au résultat attendu selon l'équation (13). La matrice des spectres de différence est là encore décrite par deux contributions qui dépendent de c_{A_2} et c_{B_2} , et par les spectres de différence correspondants, $(\mathbf{s}_{A_2} - \mathbf{s}_{A_1})$ et $(\mathbf{s}_{B_2} - \mathbf{s}_{B_1})$. Il est important de garder à l'esprit qu'il n'est pas possible de retrouver explicitement les profils de concentration des espèces présentes en début de réaction, A_1 et B_1 , mais que celles-ci doivent être considérées durant le choix d'un modèle cinétique et l'interprétation des résultats de la décomposition. Les profils de concentration de l'équation (53) indiquent non seulement que les mécanismes cinétiques des deux réactions simultanées sont conservés mais aussi que les systèmes A et B peuvent être décrits indépendamment aussi bien dans la direction spectrale que la direction des concentrations. En d'autres termes le couple formé par c_{A_2} et $(\mathbf{s}_{A_2} - \mathbf{s}_{A_1})$ n'inclut pas de contribution du système B, et réciproquement. Comme nous l'avons vu précédemment, ce raisonnement peut être généralisé à des systèmes plus complexes, comprenant plus d'espèces ou de plus nombreuses étapes réactionnelles. Par souci de brièveté, ce cas n'est pas illustré par une simulation puisque les données réelles du chapitre III expriment des situations plus complexes ou la validité de la formulation théorique de ce chapitre est confirmée.

5. Conclusion

Un développement mathématique a été proposé pour rationaliser la détermination du rang des données de différence et l'interprétation des résultats obtenus lors de résolutions multivariées. Les résultats obtenus sont testés par différentes simulations. Chaque cas, incluant ou non une espèce interférente, a montré la validité des conclusions théoriques.

Les données de différence sont systématiquement affectées par une déficience de rang. Quelque soit le procédé étudié (CAS1, CAS2 ou CAS3) ces données ont un rang inférieur d'une unité aux données non différentielles. Une stratégie basée sur l'utilisation de l'algorithme hybride HS-MCR est proposée en réponse à ce problème. L'utilisation d'un modèle cinétique permet de compenser la perte de rang en utilisant le(s) profil(s) de concentration de la (des) première(s) espèce(s) chimique(s) impliquée(s) dans le procédé à travers la contrainte cinétique. Cette espèce peut en effet être prise en compte dans l'ensemble des équations définissant le modèle cinétique, même si elle n'est pas associée à un signal dans le domaine spectral. Les profils de concentration obtenus sont alors équivalents à ceux obtenus avec une technique non différentielle.

Il est cependant important de garder à l'esprit que les spectres purs issus de ce type de résolution doivent être analysés comme des spectres de différence. Les outils présentés jusqu'ici seront particulièrement utiles pour l'étude détaillée dans le chapitre suivant qui concerne l'analyse du centre réactionnel d'une bactérie photosynthétique par spectroscopie IRTF *rapid-scan* de différence.

Chapitre III : Contribution à l'étude du centre réactionnel de *Rhodobacter sphaeroides* par spectroscopie de différence IRTF *rapid-scan* et méthodes de résolution de courbes multivariées

L'objectif de ce chapitre est d'appliquer les développements méthodologiques du chapitre précédent à l'analyse du système photochimique que constitue le centre réactionnel d'une bactérie photosynthétique. Le centre réactionnel est le lieu où l'énergie lumineuse captée par la bactérie est convertie en énergie chimique. Nous commencerons par détailler le système chimique en décrivant les phénomènes d'absorption et de transport de l'énergie lumineuse vers le centre réactionnel. Les processus prenant place à ce stade sont encore imparfaitement compris, il peut être intéressant de proposer de nouvelles approches d'analyse de données. Elles pourront aider à la compréhension des mécanismes permettant à l'énergie lumineuse d'induire un transfert d'électrons et la formation de certains produits dans le centre réactionnel.

Nous détaillerons ensuite le schéma expérimental basé sur la spectroscopie de différence InfraRouge à Transformée de Fourier (IRTF) *rapid-scan*. Cet outil, en place au Service de BioEnergétique (SBE) du Commissariat à l'Energie Atomique (CEA), atteint une résolution temporelle suffisamment basse (quelques centaines de millisecondes) pour suivre une partie des processus prenant place dans le centre réactionnel et donne accès à un grand nombre de nouvelles informations à travers l'étude des vibrations moléculaires.

Nous verrons ensuite comment les méthodes de résolution de courbes multivariées peuvent être appliquées et comment elles améliorent l'interprétation des processus photochimiques qui ont lieu. Nous détaillerons également les développements méthodologiques spécifiques à cette application. Ce chapitre suivra le plan suivant. Une approche exploratoire des données sera proposée. Elle permettra de définir certaines spécificités des données et les outils applicables à ce cas. Une résolution et les interprétations correspondantes des données spectroscopiques enregistrées seront proposées. La validité de notre approche sera confirmée par des critères statistiques et par l'interprétation chimique des résultats obtenus.

1. Système biochimique

La lumière solaire est la source d'énergie indispensable au maintien de tous les processus de la vie. La conversion de cette énergie lumineuse en énergie chimique est rendue possible par les différents mécanismes de la photosynthèse prenant place dans les plantes, les algues et certaines bactéries. La photosynthèse se décompose en deux grandes étapes se déroulant dans deux sites différents de la même cellule. La première a lieu dans une membrane spécifique de la cellule. Il s'agit d'un phénomène photochimique qui utilise l'énergie lumineuse pour produire un gradient de protons entre les deux cotés de la membrane. La seconde étape consiste à utiliser ce gradient lors de la synthèse d'ATP (Adénosine triphosphate) à partir d'ADP (Adénosine diphosphate) et de phosphate inorganique. L'ATP, stockée par la cellule, intervient lors de la biosynthèse des sucres.

1.1 Photosynthèse et photochimie

On l'aura compris les toutes premières phases de la photosynthèse sont des phénomènes photochimiques auxquels nous nous intéresserons plus particulièrement. Une fois que l'énergie lumineuse a été absorbée par différents pigments, elle est ensuite transmise vers le centre réactionnel. Un processus spécifique aux chromophores (molécules absorbant dans le visible) intervient ici : le transfert d'énergie entre molécules fluorescentes (aussi appelé FRET pour *Förster resonance energy transfer*) [105]. Ce mécanisme est à la base de la photosynthèse [106]. Ce type de transfert d'énergie peut être vu comme un simple transfert radiatif, mais il ne s'agit pas d'une molécule qui fluoresce puis d'une autre qui absorbe cette émission. Il s'agit plutôt d'un mécanisme de couplage dipôle-dipôle des champs électrostatiques entre une molécule excitée et une molécule accepteuse d'énergie. Ce transfert ne peut s'effectuer qu'entre deux molécules proches l'une de l'autre, typiquement distantes de moins de 10 nm. Les deux molécules n'ont pas besoin d'être identiques pour que ce phénomène se produise, il peut donc s'effectuer entre des pigments différents. La condition la plus importante, pour que ce transfert d'énergie ait lieu, est que le spectre d'absorption de l'accepteur recouvre au moins une partie du spectre de fluorescence du donneur comme le montre la Figure 41.

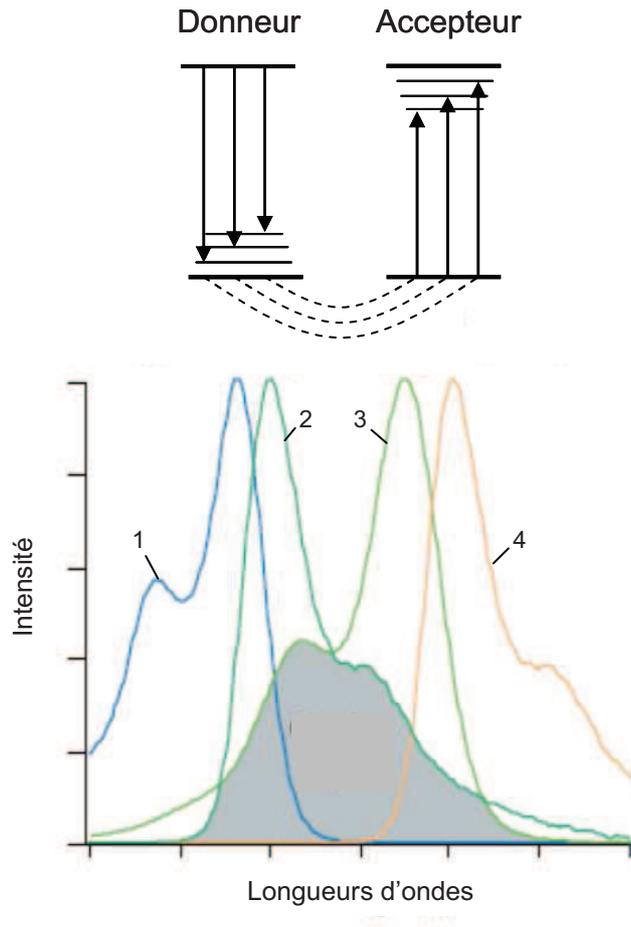


Figure 41 : Illustration du transfert d'énergie de Förster. Les spectres numérotés correspondent à 1) l'absorption du donneur 2) la fluorescence du donneur 3) l'absorption de l'accepteur 4) la fluorescence de l'accepteur. La zone grisée correspond à la zone de recouvrement entre la fluorescence du donneur et l'absorption de l'accepteur.

L'énergie captée par les différents pigments des organismes photosynthétiques peut donc être transmise par ce biais aux molécules environnantes. Ceci permet d'initier un transfert d'électrons (et de protons) qui mènera à la formation de réserve d'énergie chimique.

Pigments photosynthétiques

Le rôle principal des pigments est de capturer l'énergie lumineuse pour initier le processus de photosynthèse. On reviendra par la suite sur l'exception des pigments impliqués dans la formation de paires spéciales. Les pigments les plus courants et les plus connus sont les chlorophylles. On distingue les chlorophylles rencontrées dans les organismes oxygéniques des bactériochlorophylles propres aux bactéries phototrophes. La formule générale de ces dernières est fournie Figure 42. Elles sont constituées d'un noyau

tétrapyrrollique (porphyrine) quasiment plan et d'une chaîne carbonée aussi appelée queue phytol.

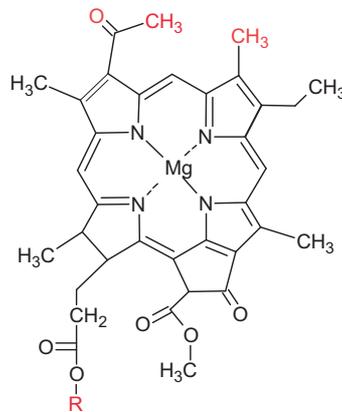


Figure 42 : Formule générale des bactériochlorophylles.

D'autres pigments sont mis en jeu dans les complexes antennaires des systèmes photosynthétiques, les caroténoïdes et les phycobilines [107]. Leurs spectres d'absorption sont différents de ceux des chlorophylles ce qui offre deux avantages. Le premier est d'absorber des longueurs d'ondes non « utilisées » par les chlorophylles. Le second est de permettre un flux d'électrons vers la paire spéciale du centre réactionnel [108].

Conversion de la lumière

La paire de molécules spéciale est donc formée par deux chlorophylles, qui constituent un centre photochimique P. Une fois à l'état excité P* cette structure a la propriété d'expulser un électron (par séparation de charges). Ces étapes sont résumées par les équations (55) et (56) :



L'électron éjecté est transféré à un accepteur d'électrons M (une pheophytine) soit au total :



La paire radicalaire [P⁺, M⁻] doit avoir une durée de vie très courte pour éviter une recombinaison de charges, la charge de M⁻ est donc rapidement transmise à un autre accepteur d'électrons. La forme P⁺ est ensuite réduite par l'apport d'un électron fourni par le milieu externe (H₂O ou H₂S).

D'un point de vue énergétique, le donneur primaire P non excité a un potentiel redox d'environ + 400 mV. L'absorption d'un photon de 840 nm par exemple abaisse ce potentiel à environ - 1000 mV, ce qui fait de P* un réducteur très puissant. L'accepteur d'électron M en relation avec une chaîne redox va donc transmettre cette charge jusqu'à un accepteur terminal (une adénine nicotinamide dinucléotide (NAD⁺) ou une quinone comme dans le cas qui nous intéresse) [109].

1.2 Photosynthèse bactérienne

Au cours de l'évolution des organismes vivants, la photosynthèse a été le processus principal de l'élaboration de la matière organique. Phénomène complexe, elle implique des mécanismes physiques et biochimiques par lesquels les plantes et certaines bactéries utilisent l'énergie solaire pour assimiler le gaz carbonique de l'atmosphère, l'incorporer et synthétiser ensuite les molécules organiques. On se limitera ici à l'étude des bactéries photosynthétiques. Parmi celles-ci le groupe des bactéries pourpres est le plus étudié [107] car il s'agit des organismes phototrophes les plus simples. Ils possèdent des centres réactionnels d'un seul type (I ou II, c'est à dire incluant soit des accepteurs de type fer-soufre soit des quinones) dont les caractères sont primitifs par rapport aux centres réactionnels des plantes.

Ces bactéries sont des organismes unicellulaires procaryotes dont l'habitat est typiquement aquatique et anaérobie. Chaque espèce ne renferme qu'un seul type de bactériochlorophylle, a ou b. Leur coloration rouge pourpre provient des caroténoïdes rencontrés dans ces organismes. Deux groupes de bactéries pourpres peuvent être distingués les sulfureuses ou non-sulfureuses [107]. Nous nous intéresserons plus particulièrement à une espèce appartenant au second groupe : *Rhodobacter sphaeroides*.

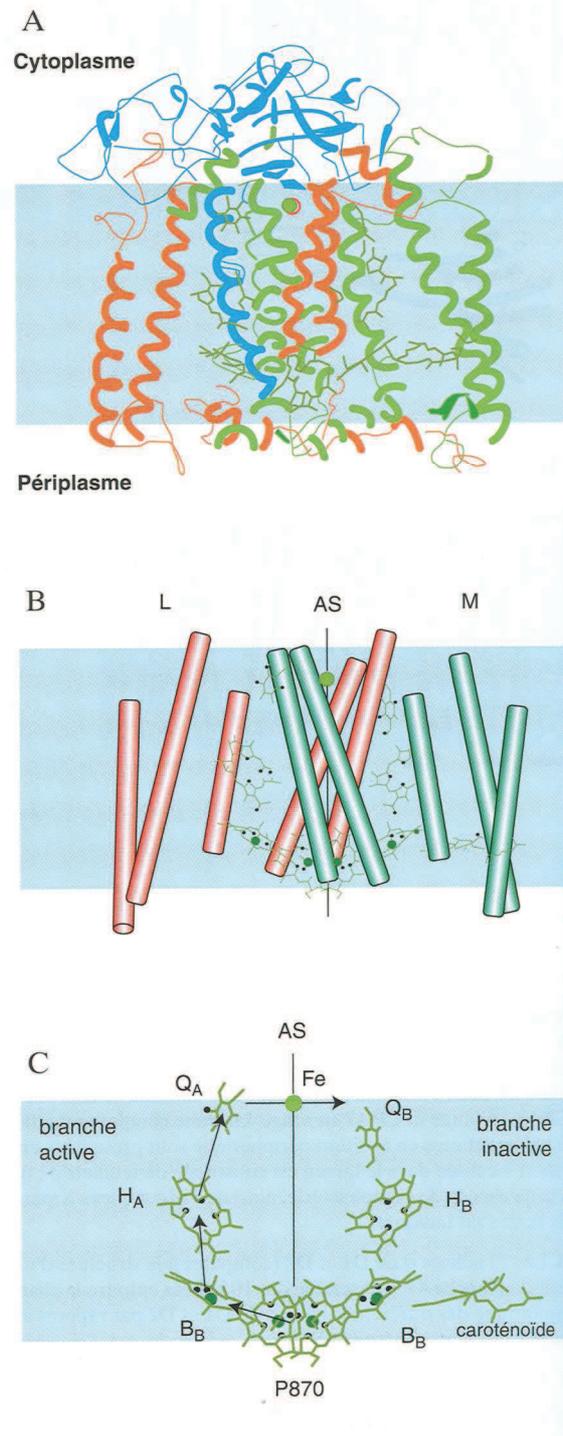


Figure 43 : Structure tridimensionnelle du centre réactionnel de *Rhodospira rubra*. A) Représentation des 11 hélices α formant l'enveloppe du centre réactionnel B) les 10 hélices en rouge et en vert forment les deux protéines au positionnement symétrique, perpendiculairement à l'axe cytoplasmique (AS). C) Chaîne redox : P870 est le donneur primaire d'électrons ; d'après [107].

Le centre réactionnel de cette espèce est de type II (donc contenant des quinones). Sa structure tridimensionnelle a été établie de longue date après cristallisation par la technique de diffraction des rayons X [110-112]. Ce système est donc considéré comme un des centres réactionnels "modèles". Il est composé de trois protéines d'environ 25kDa. Deux de ces protéines sont symétriques par rapport à un axe perpendiculaire à la membrane cytoplasmique. Quatre bactériochlorophylles a, deux bactériophéophytines a (H), deux ubiquinones (Q) et un atome de fer non hémique sont liées à ces protéines. Deux des bactériochlorophylles a forment la paire spéciale P qui est le siège d'une séparation de charge et constitue le donneur primaire d'électron (P870). Les deux autres molécules de ce type dites bactériochlorophylles accessoires sont appelées B_A et B_B. Les lettres A et B en indice se réfèrent à deux positions symétriques dans le centre réactionnel, comme le montre la Figure 43.

La formation du donneur primaire d'électron P870*, sous éclairage, est suivie de l'expulsion d'un électron : $P^* \rightarrow P^+ + e^-$. La disposition symétrique de deux des protéines forment deux chaînes de composés (A et B) comme le montre la Figure 43. Une seule de ces deux chaînes est active (A). La bactériophéophytine a (H_A) fonctionne comme accepteur primaire unique d'électrons ; puis l'électron migre vers la quinone Q_A, puis de Q_A à Q_B. Le transfert d'électrons entre H_A et Q_A est très rapide (≈ 200 ps). Il est beaucoup plus lent de Q_A à Q_B (≈ 200 μ s) [113].

La présence de deux quinones Q_A et Q_B comme accepteurs d'électrons terminaux du centre réactionnel est caractéristique des centres de type II et du photosystème II des plantes supérieures [107]. Chez *Rhodobacter sphaeroides*, Q_A et Q_B sont des ubiquinones-10, cette molécule est représentée Figure 44.

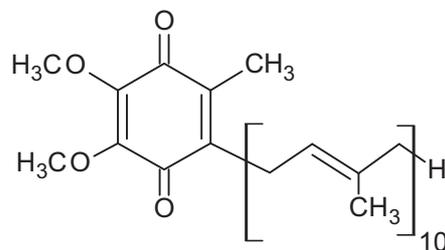


Figure 44 : Formule de l'ubiquinone-10 ; dans le centre réactionnel de *Rhodobacter sphaeroides* Q_A et Q_B sont toutes deux des molécules d'ubiquinone-10.

Le transfert d'électrons entre Q_A et Q_B se fait parallèlement à la membrane cytoplasmique, à la différence du transfert entre P870 et Q_A. L'ubiquinone Q_A, située dans un environnement

très hydrophobe, est isolée du milieu cytoplasmique. Elle fonctionne comme un transporteur d'électrons, alors que les quinones Q_B sont également des transporteurs d'hydrogène. Les deux électrons acceptés successivement par Q_B (et Q_BH^- [114]) sont neutralisés par deux protons [115-117]. De ce fait la molécule Q_B , réduite et protonée en ubiquinol Q_BH_2 , fonctionne comme un accumulateur de deux charges négatives [118], cédées deux par deux à la réserve de quinones [119] et de là au complexe bc_1 . La structure de l'ubiquinol est représentée Figure 45.

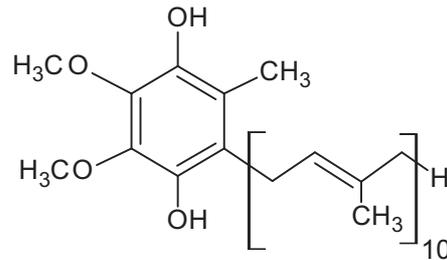


Figure 45 : Formule de l'ubiquinol réduite à partir de Q_B dans le centre réactionnel de *Rhodobacter sphaeroides* durant la photoexcitation.

Le fonctionnement de Q_B est parfois qualifié de portail électronique (*electron gate*). L'ubiquinol Q_BH_2 est ensuite réoxydée par une enzyme membranaire (NAD^+ déshydrogénase). Ce mécanisme cyclique agit donc comme une pompe à protons [117]. Ce cycle est décrit Figure 46.

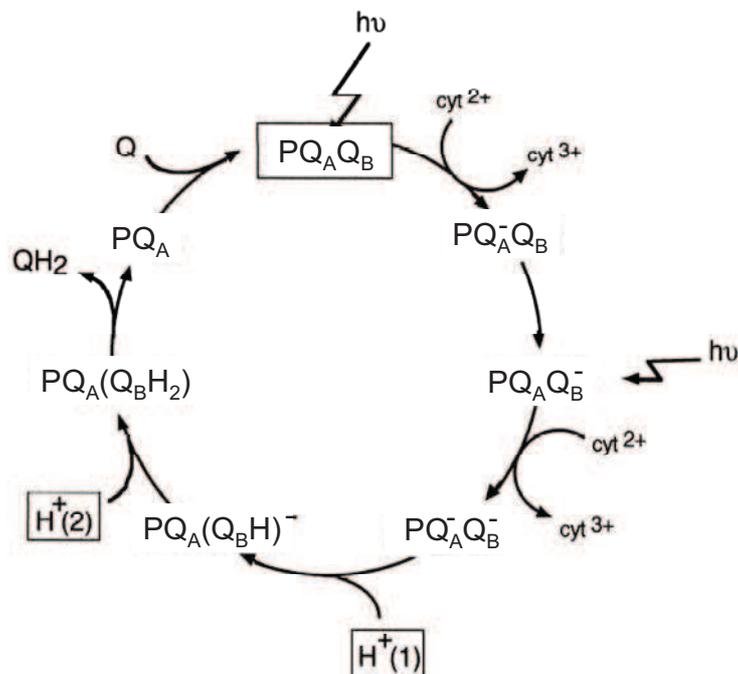


Figure 46 : Cycle de réduction de la quinone dans les centres réactionnels des bactéries. Q_B est réduite deux fois, puis se lie successivement à deux protons. La molécule d'ubiquinol formé (QH_2) quitte le centre et est remplacée par une quinone.

Le gradient de protons formé est à la base de la formation d'ATP (Adénosine triphosphate), elle même utilisée pour alimenter les réactions métaboliques de la cellule. Les principales étapes de ce processus sont résumées par la Figure 47.

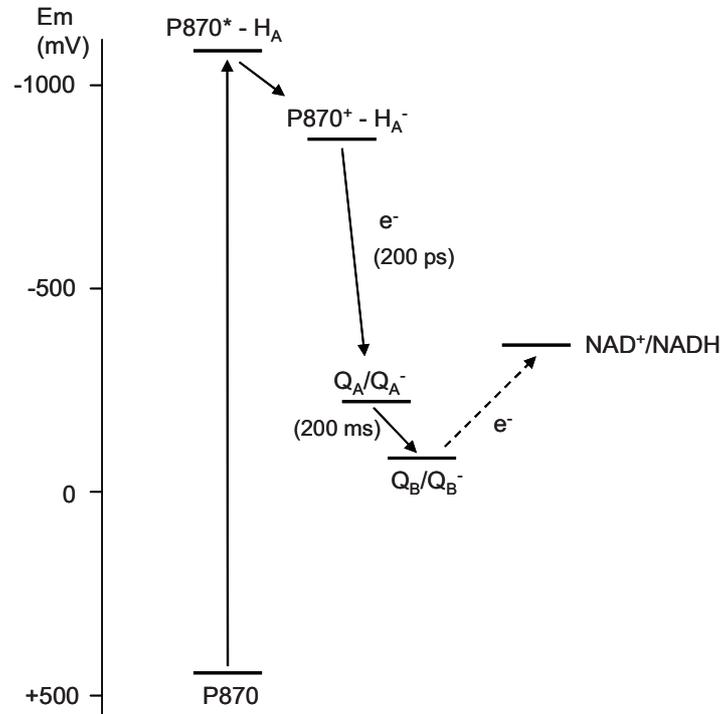


Figure 47 : Transfert d'électrons chez une bactérie pourpre sur une échelle de potentiels redox. P870 est le donneur primaire d'électrons, siège de la séparation de charge ; H_A , une bactériophéophytine, est l'accepteur primaire de l'électron éjecté. Une autre voie redox apporte, via des sulfures S^{2-} , des électrons au P870.

Certains aspects de la formation d'ubiquinol et du processus d'expulsion de cette molécule hors du centre réactionnel restent mal connus, notamment le transfert d'électrons entre Q_A et Q_B et le mode de formation de l'ubiquinol Q_BH_2 . Il peut être intéressant de proposer un suivi par spectroscopie IRTF *rapid-scan* et une analyse multivariée des données résultantes pour expliquer ces phénomènes.

2. Matériels et méthodes

Les expériences présentées ici ont été réalisées au CEA de Saclay par le Dr. Alberto Mezzetti [118, 119]. Les membranes photosynthétiques contenant les centres réactionnels de *Rhodobacter sphaeroides* ont été préparées comme décrit par Clayton *et al.* [120] et purifiées par gradient de sucrose. La suspension obtenue a été centrifugée. Le centrifugat a été pressé et scellé entre deux fenêtres de fluorure de calcium CaF₂. Les spectres ont été enregistrés avant et après illumination de l'échantillon par une lampe tungstène-halogène 250 W (OrielTM), en utilisant le mode *rapid-scan*. Ce mode a pour but d'accélérer l'enregistrement des interférogrammes en spectroscopie IRTF pour atteindre des échelles temporelles plus basses. Le miroir mobile se déplace à la vitesse maximale permise par le moteur ($\approx 10 \text{ cm.s}^{-1}$). Seul un "demi-interférogramme" est enregistré en mode *rapid-scan*. Les deux bras de l'interférogramme étant symétriques, il est possible de n'en enregistrer qu'un. La constante de temps des détecteurs devient alors le facteur critique. Les mesures IRTF ont été faites sur un spectromètre Bruker IFS 88 équipé avec un détecteur MCT-A dont la constante de temps est de l'ordre de quelques nanosecondes et avec une lame séparatrice en KBr. Plusieurs interférogrammes de référence sont enregistrés à des délais croissants, chacun représentant l'état du système à une position temporelle. Le cycle de mesures est répété pour améliorer le rapport signal sur bruit. La transformée de Fourier est calculée après la mesure. La Figure 48 schématise cette procédure.

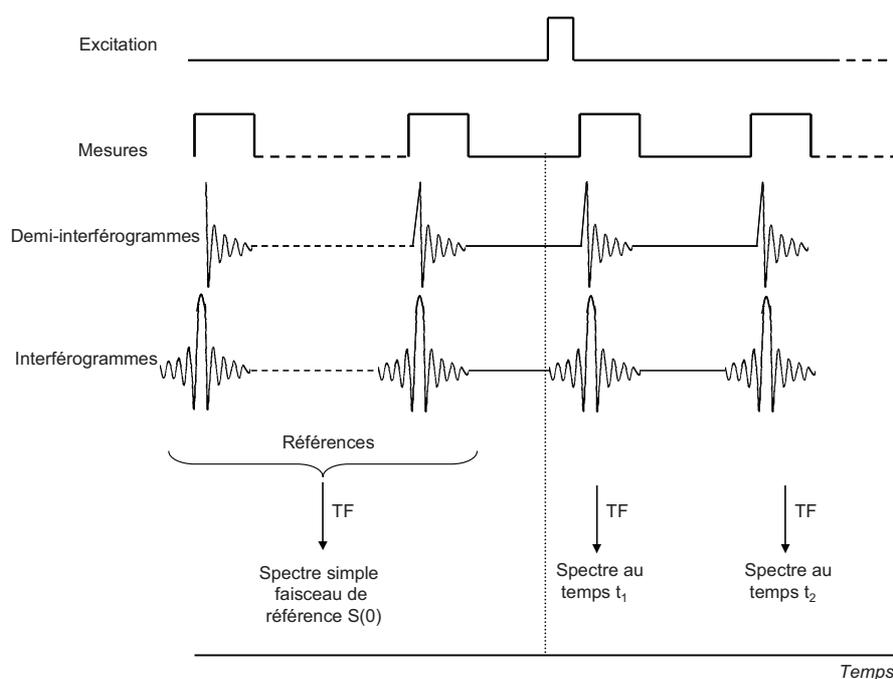


Figure 48 : Schéma d'une expérience *rapid-scan* utilisant une excitation de courte durée.

Ce schéma correspond à l'observation successive d'excitations de courtes durées. Une autre option permet d'observer les effets d'une excitation plus longue [119]. De nouveau plusieurs interférogrammes sont enregistrés avant le début de l'excitation. Dans notre cas les scans de référence ont été enregistrés dans l'obscurité. Ils sont constitués de 200 interférogrammes, puis la transformée de Fourier permet de construire le spectre simple faisceau $S(0)$ de l'état initial. Ensuite la lampe est allumée et les mesures résolues en temps commencent, ce qui décrit la phase de photoexcitation du système. Les mesures se poursuivent une fois la lampe excitatrice éteinte. Ceci permet d'observer la relaxation du système. Cette procédure est résumée par la Figure 49.

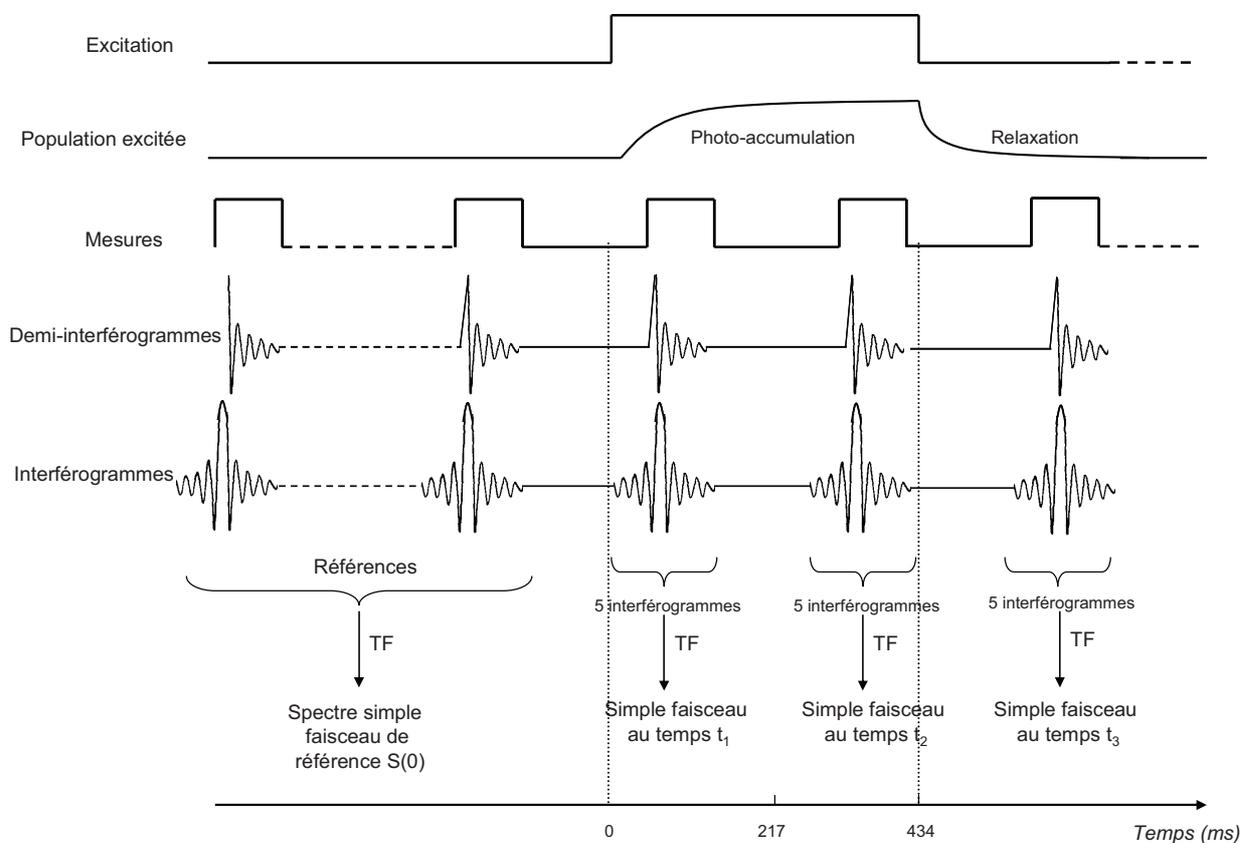


Figure 49 : Schéma d'une expérience *rapid-scan* utilisant une excitation de longue durée.

Un jeu de données résulte de nombreuses répétitions d'une même expérience, d'où la nécessité de travailler avec des phénomènes réversibles. Dans chaque expérience, les cinq premiers interférogrammes sont enregistrés dans l'obscurité pour estimer le niveau de bruit. Ensuite l'échantillon est éclairé, dans notre cas durant 4,3 s. Les interférogrammes acquis durant l'éclairement sont moyennés par groupe de cinq avant de calculer leur transformée de Fourier. Les spectres simple faisceau $S(t)$ représentent donc une valeur moyenne sur une fenêtre temporelle de 217 ms. La mesure continue ensuite dans l'obscurité de la même

manière durant 55 s pour suivre la relaxation du système. Le système reste ensuite dans l'obscurité pendant plusieurs minutes de façon à assurer une relaxation complète avant la répétition du cycle de mesure. Les spectres de différence sont calculés à partir des spectres simple faisceau selon l'équation (58).

$$\Delta A = -\log(S(t)/S(0)) \quad (58)$$

Cinq matrices de données ont ainsi pu être obtenues. Elles se différencient par le nombre d'accumulations utilisées ainsi que par l'intensité de l'excitation lumineuse. Le Tableau 4 résume les différentes conditions expérimentales utilisées.

Expérience	Photoexcitation (en % du filtre de transmission)	Nombre d'accumulations par spectre
D₁	8	150
D₂	5	750
D₃	2,2	1350
D₄	2,2	1050
D₅	2,2	1050

Tableau 4 : Conditions expérimentales des différentes expériences enregistrées sur le même échantillon.

Le nombre plus bas d'accumulations aux expériences **D₁** et **D₂** s'explique par la plus grande intensité de lumière incidente, qui pourrait détériorer l'échantillon si l'exposition était plus longue.

3. Analyse exploratoire

A titre d'exemple, les spectres de différence d'absorption IRTF *rapid-scan* (formant la matrice D_3) des membranes de *Rhodobacter (Rb.) sphaeroides* sont présentés Figure 50. La partie de gauche représente la phase d'excitation (0 à 4,3 s) de l'échantillon et celle de droite la phase de relaxation (de 4,3 à 59 s). Les changements observés décrivent la réponse de l'échantillon à l'éclairement puisque le spectre de référence $S(0)$ est enregistré dans l'obscurité. Les couleurs utilisées correspondent aux différents temps, allant du bleu vers le rouge en excitation et du rouge vers le bleu en relaxation.

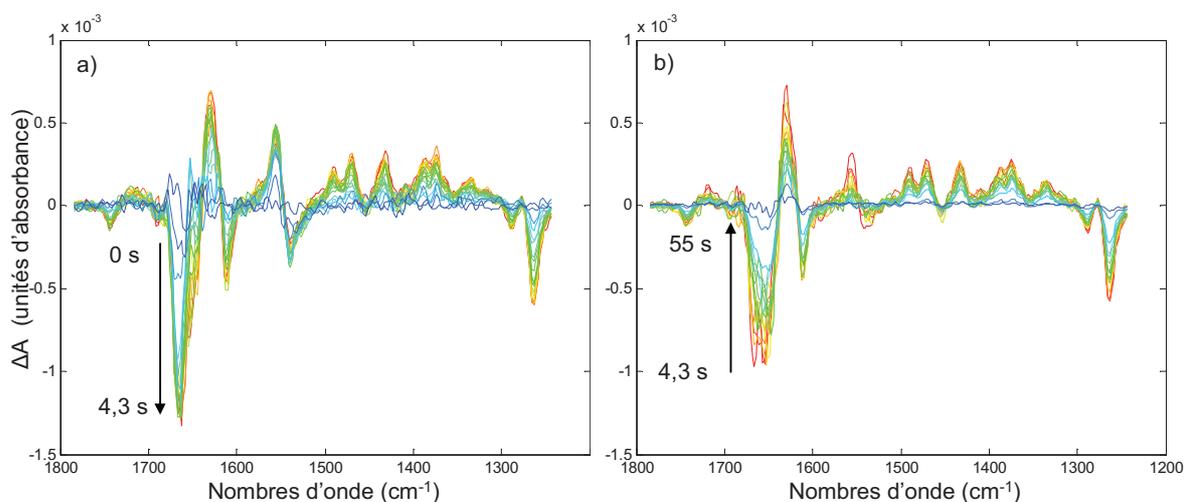


Figure 50 : Spectres IRTF *rapid scan* obtenus durant le suivi de la photoexcitation de membranes de *Rb. Sphaeroides*, a) phase d'excitation b) phase de relaxation.

La principale observation possible Figure 50a durant la phase d'excitation est l'émergence de bandes positives et négatives, en particulier dans les régions Amide I ($1700-1600\text{ cm}^{-1}$) et Amide II ($1580-1520\text{ cm}^{-1}$). La bande négative à 1264 cm^{-1} peut être associée aux modes de vibration de C–O–C des groupes méthoxy de l'ubiquinone [121]. Cette bande négative est généralement utilisée comme indicateur du nombre de molécules d'ubiquinone consommées par la réaction [118, 119]. On observe la formation correspondante d'ubiquinol grâce aux bandes situées entre 1520 et 1350 cm^{-1} . La relaxation du système peut être observée Figure 50b. Les bandes observées précédemment décroissent progressivement vers un retour à l'état initial du système.

3.1 Détection du rang et estimation initiales

Les premières valeurs obtenues lors de la décomposition en valeurs singulières des matrices \mathbf{D}_1 à \mathbf{D}_5 sont présentées dans le Tableau 5. Les premières valeurs singulières sont associées au signal et l'objectif est de les séparer des suivantes associées au bruit. Les valeurs en gras représentent les valeurs singulières considérées comme significatives et on constate que l'estimation du rang pour les cinq matrices considérées fournit une fourchette allant de 2 à 3 contributions significatives. On peut également noter que les valeurs singulières associées au bruit sont pour \mathbf{D}_1 plus grandes que dans le cas des autres matrices. Cela traduit le fait que la matrice \mathbf{D}_1 est affectée par un niveau de bruit plus important, puisque le nombre d'accumulations utilisées pour former \mathbf{D}_1 est beaucoup plus faible que pour les autres jeux de données (voir Tableau 4).

\mathbf{D}_1	\mathbf{D}_2	\mathbf{D}_3	\mathbf{D}_4	\mathbf{D}_5
0,0388	0,0278	0,0196	0,0175	0,0145
0,0077	0,0071	0,0050	0,0044	0,0039
0,0047	0,0025	0,0014	0,0018	0,0017
0,0039	0,0017	0,0012	0,0015	0,0015
0,0036	0,0015	0,0011	0,0013	0,0014
0,0031	0,0013	0,0010	0,0012	0,0013
0,0030	0,0012	0,0009	0,0010	0,0011
...

Tableau 5 : Premières valeurs singulières issues de la SVD des matrices de données \mathbf{D}_1 , \mathbf{D}_2 , \mathbf{D}_3 , \mathbf{D}_4 et \mathbf{D}_5 .

Les résultats de l'EFA [33] des cinq matrices sont rassemblés Figure 51. Les lignes noires représentent le résultat de l'EFA *forward* et les lignes rouges l'EFA *backward*. On peut noter dans tous les cas l'absence de contributions en début de procédé, typique des spectres de différence. Le cas le plus simple est rencontré dans les matrices \mathbf{D}_3 , \mathbf{D}_4 et \mathbf{D}_5 où deux contributions sont clairement détectées au dessus du niveau de bruit (estimé ici à $-5,5$ sur l'échelle logarithmique des valeurs singulières). Ces deux contributions apparaissent au niveau des deuxième et septième lignes environ pour disparaître au niveau des vingt quatrième et trente cinquième lignes. Les résultats de l'EFA de \mathbf{D}_2 sont différents de ceux des matrices \mathbf{D}_3 , \mathbf{D}_4 et \mathbf{D}_5 . Deux contributions majeures apparaissent clairement et de façon synchrone à celles observées pour \mathbf{D}_3 . Cependant une troisième contribution est visible, même si elle n'est pas aussi clairement séparée du niveau de bruit. Nous montrerons la validité de l'existence de cette contribution par la suite en utilisant une approche chimométrique légèrement différente [59].

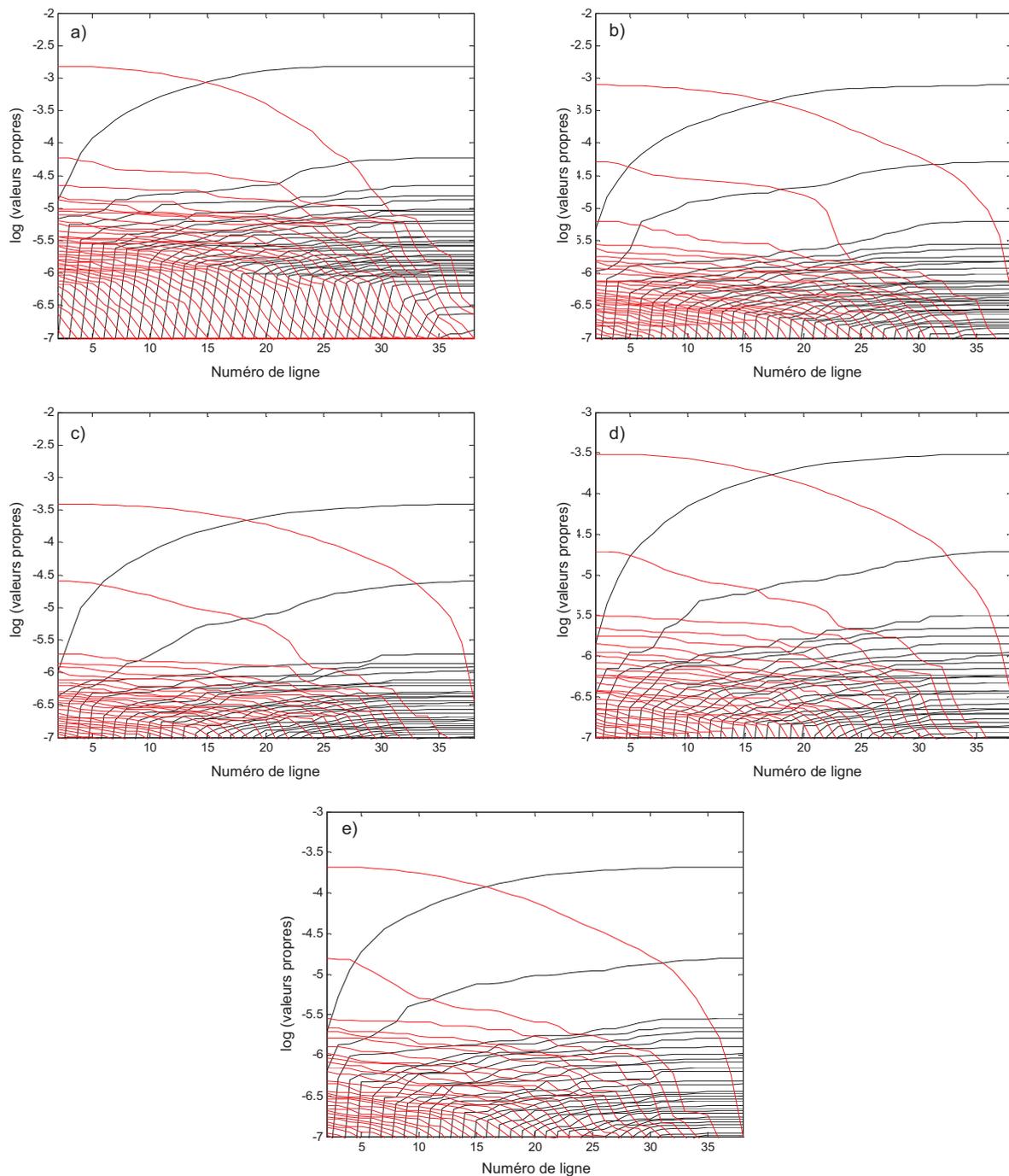


Figure 51 : Résultats de l'EFA en sens direct (noir) et inverse (rouge) pour les matrices a) D_1 , b) D_2 , c) D_3 , d) D_4 et e) D_5 .

Enfin le cas de D_1 est le moins clair puisqu'il a été enregistré avec peu d'accumulations. Le rapport signal sur bruit est donc dégradé. De plus, l'importante intensité lumineuse utilisée pour exciter l'échantillon est suspectée de provoquer un mécanisme réactionnel plus complexe impliquant un nombre plus important de contributions. L'EFA présentée Figure 51a montre deux contributions ressortant du bruit, mais ce même bruit atteint un niveau plus élevé que dans les autres expériences (près de -4,5 sur l'échelle des

logarithmes de valeurs propres) qui couvre probablement les contributions supplémentaires. On peut alors proposer d'utiliser l'algorithme d'Analyse Factorielle Evolutive pour matrices Déficientes de Rang (EFARD, *Evolving Factor Analysis for Rank Deficient matrices*) [31, 40]. Cette méthode permet en effet de détecter plus facilement des contributions communes entre deux matrices et des contributions qui peuvent être différentes. Il est important de noter que la méthode EFARD travaille comme l'EFA classique, mais sur des paires de matrices augmentées. Si une des matrices est déficiente de rang, elle est toujours analysée en second et attachée à la matrice de rang complet, ce qui revient à faire l'EFA *forward* en direction $\mathbf{D}_i \rightarrow \mathbf{D}_j$ et l'EFA *backward* en direction $\mathbf{D}_j \leftarrow \mathbf{D}_i$. On travaille de la même façon quand les deux matrices sont de rang plein. La Figure 52 correspond aux résultats de l'EFARD sur la matrice augmentée $[\mathbf{D}_3; \mathbf{D}_1]$. On peut voir aussi bien dans le sens *forward* que dans le sens *backward*, lorsque la limite de la matrice \mathbf{D}_3 est dépassée, l'apparition d'une troisième contribution qui jusqu'alors n'était pas clairement observée dans \mathbf{D}_1 . Celle-ci est mise en valeur par l'augmentation de données qui aide à distinguer la nouvelle contribution puisque les deux autres contributions communes aux deux matrices sont mieux définies dans la matrice \mathbf{D}_3 (moins bruitée). Pour confirmer que cette troisième contribution est la même que celle détectée dans \mathbf{D}_2 , une approche identique basée sur l'EFARD est utilisée sur une augmentation de données $[\mathbf{D}_2; \mathbf{D}_1]$.

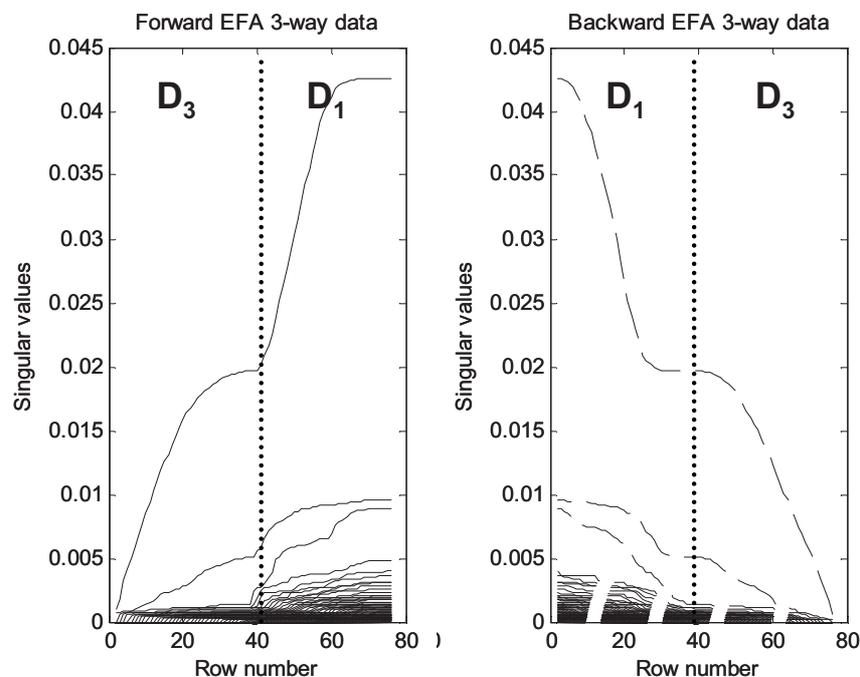


Figure 52 : EFARD calculée sur la matrice augmentée $[\mathbf{D}_3; \mathbf{D}_1]$

La Figure 53 présente les résultats *forward* et *backward* de l'EFARD sur la matrice augmentée composée de $[D_2; D_1]$. Dans ce cas, on ne détecte que trois contributions différentes du bruit tout au long de l'analyse EFA. D_2 et D_1 étant de rang trois, cela indique que les trois espèces détectées sont communes aux deux expériences.

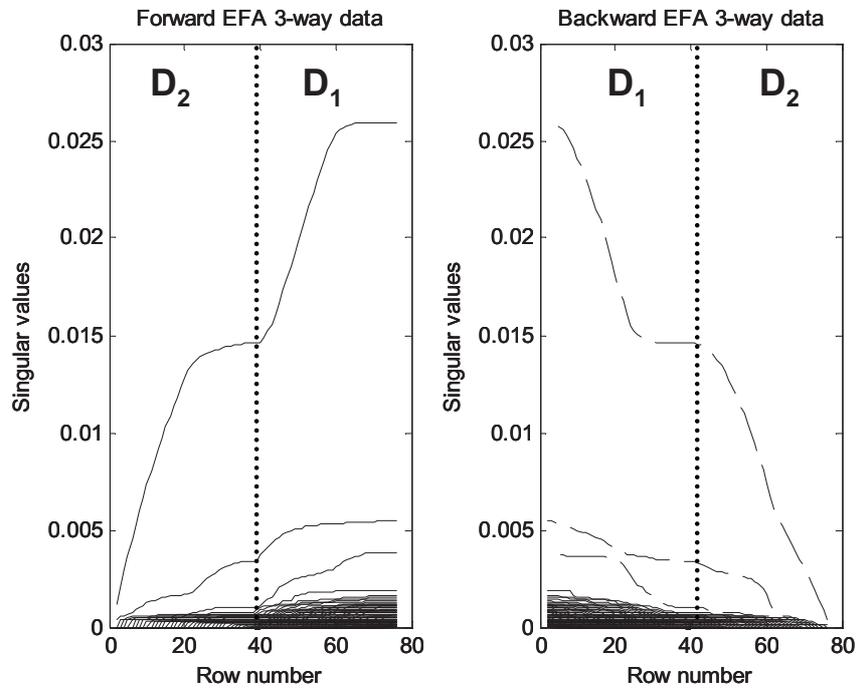


Figure 53 : EFARD calculée sur la matrice augmentée $[D_1; D_2]$.

Des analyses SVD et EFA, on peut conclure que les conditions d'éclairage semblent influencer sur le nombre de contributions détectables. Les matrices D_1 et D_2 soumises à une illumination plus importante ont un rang estimé à trois là où D_3 , D_4 et D_5 n'ont qu'un rang égal à deux. Cette conclusion se base sur l'EFARD des augmentations de matrices $[D_3; D_1]$ et $[D_2; D_1]$. Les approches EFA et EFARD seront systématiquement utilisées par la suite sur l'ensemble des données.

4. Modélisation souple du procédé. Application de MCR-ALS

4.1 Résolutions individuelles

L'analyse des données a été effectuée matrice par matrice dans un premier temps. Les résultats de la décomposition MCR-ALS obtenus pour les matrices \mathbf{D}_3 , \mathbf{D}_4 et \mathbf{D}_5 sont présentés Figure 54. Le cas des matrices \mathbf{D}_1 et \mathbf{D}_2 sera discuté par la suite. Les contraintes utilisées lors de ces résolutions sont la non négativité et l'unimodalité des profils de concentration. Les *lack of fit* et la variance expliquée (R^2) sont directement reportés sur les graphiques. Il est intéressant de noter que les résultats de ces trois matrices sont comparables aussi bien pour les profils de concentration que pour les spectres. Ceci est en accord avec le fait que ces trois séries de mesures aient été enregistrées dans les mêmes conditions expérimentales. Les trois matrices de profils de concentration sont composées de deux contributions. La première, représentée en bleu, évolue très rapidement après le début de l'excitation jusqu'à atteindre une valeur maximale (aux alentours de $2,5 \cdot 10^{-3}$ unités arbitraires sur la Figure 54). Une fois la lampe éteinte (après 4,3 s) cette contribution disparaît très rapidement. L'interprétation complète de l'ensemble des résultats sera détaillée dans la section suivante (Résolution en augmentation de matrices). Le profil de concentration de la seconde contribution, en vert, évolue plus lentement. La valeur maximale en concentration de cette contribution n'est atteinte qu'à la fin de la période d'excitation lumineuse et le retour à l'équilibre dure près d'une minute. Le spectre correspondant est caractérisé de nouveau par la présence de bandes dans la zone Amide I mais aussi par différents pics entre 1500 et 1250 cm^{-1} . On remarquera plus particulièrement le pic négatif à 1264 cm^{-1} qui, on l'a vu précédemment, est associé à la consommation d'ubiquinone.

La similitude des résultats, aussi bien pour les profils de concentration que pour les spectres, sur ces différentes matrices permet de supposer que les espèces impliquées sont les mêmes dans les trois cas. Une stratégie d'augmentation de matrice dans la direction des concentrations est alors possible et encouragée pour réduire les ambiguïtés de rotation et d'augmenter la justesse de la résolution. Le choix d'une augmentation mettant en commun les spectres est également privilégié car cette stratégie permettra d'introduire les matrices \mathbf{D}_1 et \mathbf{D}_2 dans la résolution. En effet ces deux matrices étant soumises à des excitations plus intenses, la cinétique des procédés observés sera probablement affectée.

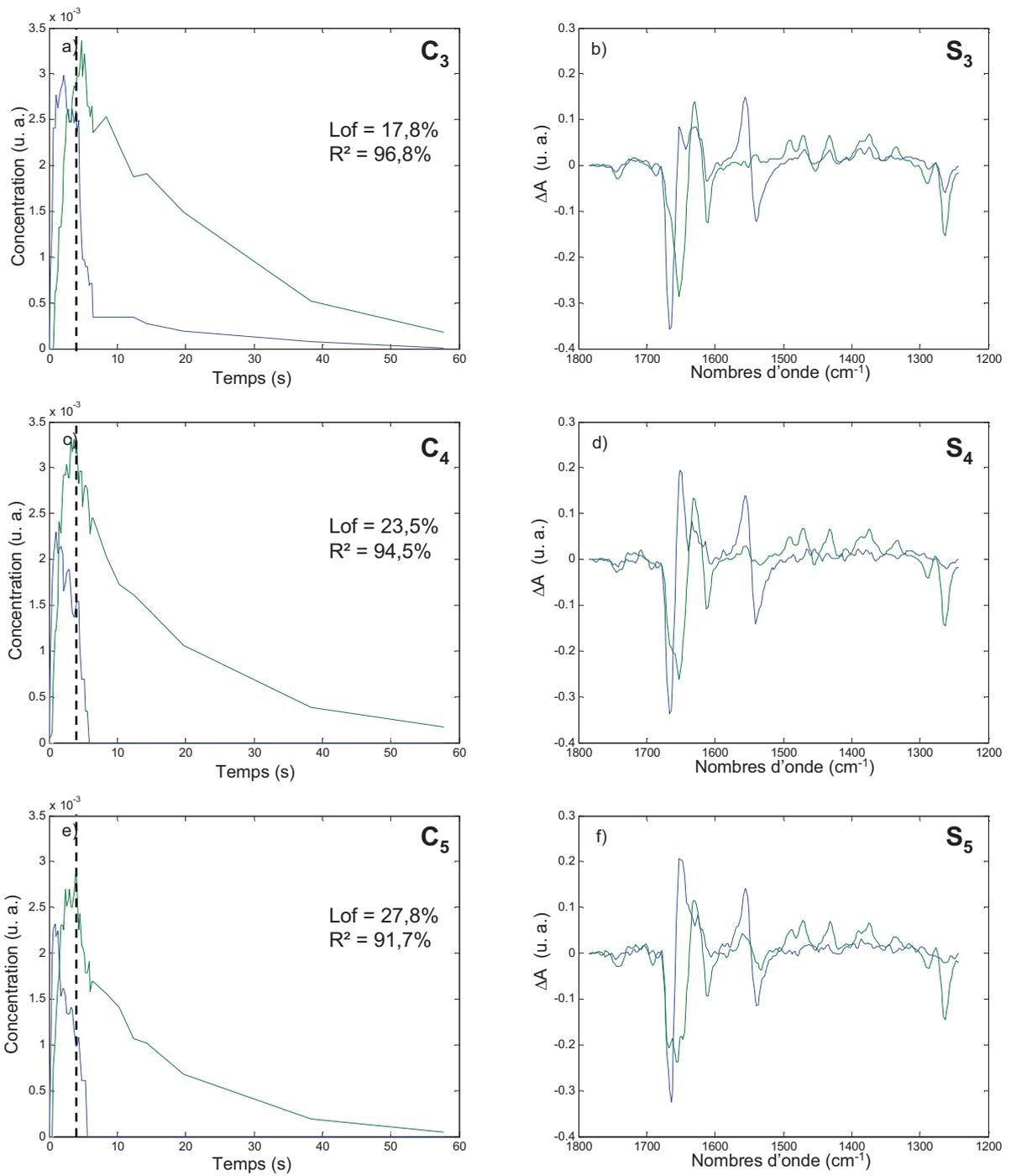


Figure 54 : Résultats des décompositions individuelles des matrices D_3 , D_4 et D_5 par MCR-ALS en a) C_3 , b) S_3 , c) C_4 , d) S_4 , e) C_5 et f) S_5 .

Le résultat de la décomposition de la matrice \mathbf{D}_2 par MCR-ALS est présenté Figure 55. Les mêmes contraintes que précédemment ont été utilisées. Les deux premières espèces détectées sont comparables à celles observées dans les résolutions précédentes, même si dans les deux cas des pics sont discernables entre 1500 et 1250 cm^{-1} ce qui n'était pas vrai pour la première espèce dans \mathbf{D}_3 , \mathbf{D}_4 et \mathbf{D}_5 . La troisième espèce a un profil de concentration qui évolue très rapidement durant la phase d'excitation et les premiers temps de la relaxation. Le spectre correspondant est assez similaire à celui de la deuxième espèce en dehors de la bande positive visible à 1562 cm^{-1} .

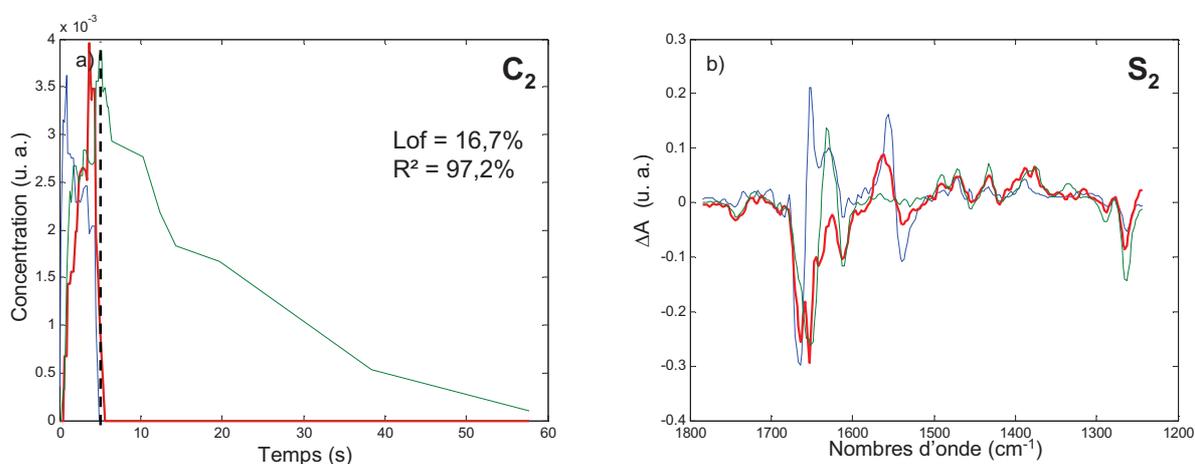


Figure 55 : Résultats de la décomposition individuelle de la matrice \mathbf{D}_2 par MCR-ALS en a) C_2 , et b) S_2 .

Les résultats correspondant à la matrice \mathbf{D}_1 seule ne sont pas présentés ici car le rapport signal sur bruit de ces données ne permet pas d'obtenir une résolution acceptable. Cependant le résultat de l'EFARD permet d'affirmer que \mathbf{D}_1 et \mathbf{D}_2 partagent les mêmes espèces, donc si les contributions détectables dans \mathbf{D}_2 permettent l'augmentation il paraît raisonnable d'y inclure \mathbf{D}_1 . C'est l'objet de la section suivante.

4.2 Résolution en augmentation de matrices

Les résolutions individuelles présentées précédemment permettent de tirer de premières conclusions même si l'ajustement des différents modèles n'est pas parfait. On peut notamment constater que l'augmentation de matrices en colonne est envisageable. L'un des intérêts d'analyser plusieurs matrices simultanément est que la résolution obtenue sera semi quantitative, c'est-à-dire que les niveaux de concentration observés dans les différentes expériences seront directement comparables pour chaque espèce en termes relatifs. Le fait que les espèces chimiques soient communes entre expériences (et, par conséquence, S^T) mais que les cinétiques soient différentes entre expériences (C_i variables) explique que les données soient analysées ensemble sous la forme d'une matrice augmentée dans le sens des colonnes [$D_1; D_2; D_3; D_4; D_5$] comme le montre la Figure 56.

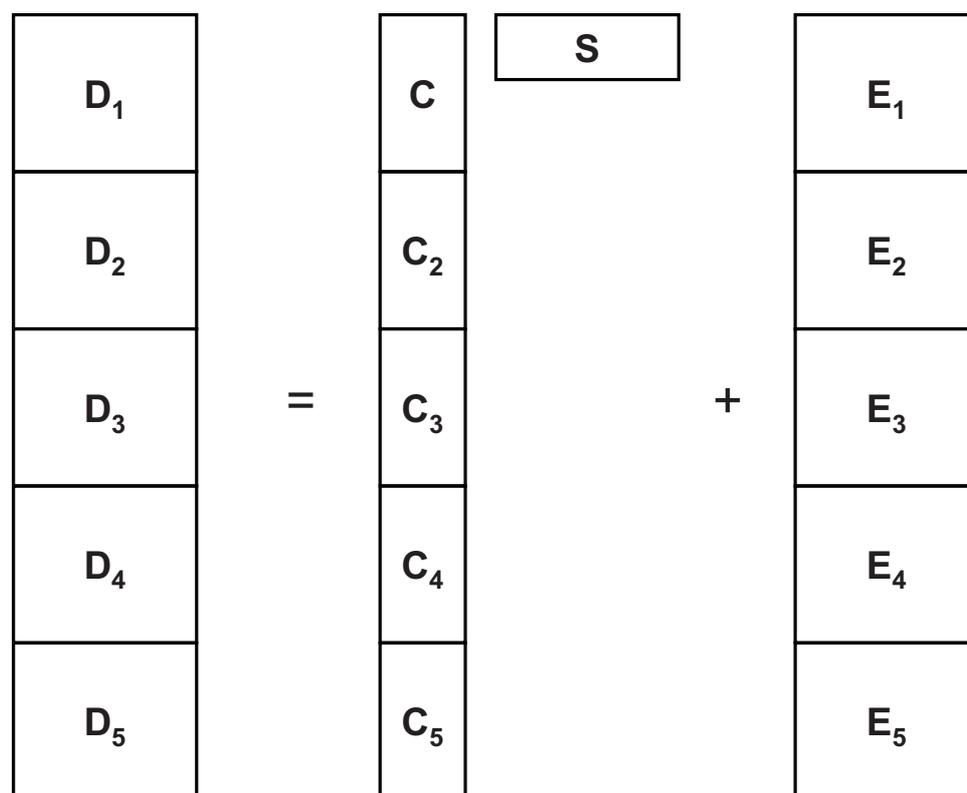


Figure 56 : Structure de la matrice augmentée dans la direction des colonnes utilisée lors de la résolution MCR-ALS

Les résultats obtenus en utilisant MCR-ALS sont présentés Figure 57 pour les spectres purs S^T et Figure 58 pour les profils de concentration C_1 , C_2 , C_3 , C_4 et C_5 en fonction du temps. Les contraintes utilisées lors de cette résolution sont la non négativité et l'unimodalité des profils de concentration. Dans ce cas le *Lack of fit* est égal à 27,4%, le pourcentage de variance expliquée atteint 93% et les résidus correspondent à un bruit indépendant, plus intense dans la bande Amide I. Ceci est cohérent avec les observations expérimentales.

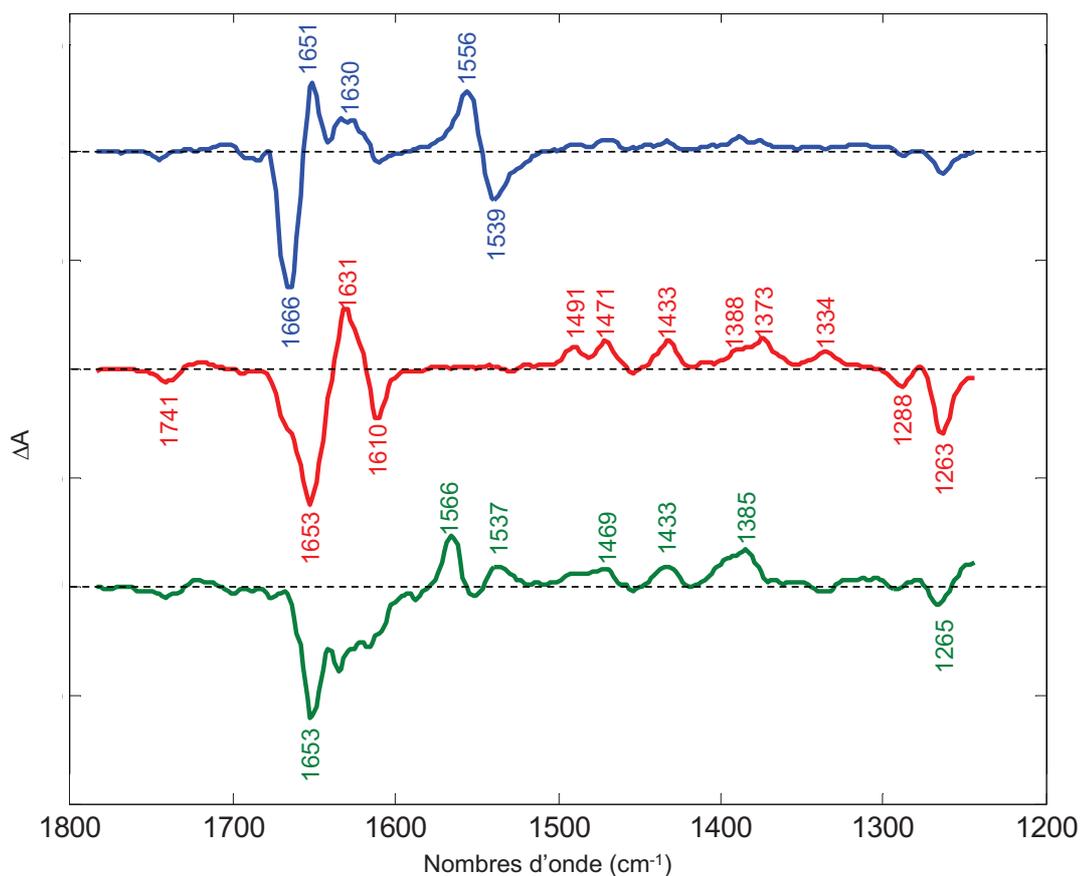


Figure 57 : Spectres de différence S^T issus de la résolution MCR-ALS de la matrice augmentée [D_1 ; D_2 ; D_3 ; D_4 ; D_5]. De haut en bas spectres attribués aux changements de conformation de la protéine, à la formation d'ubiquinol sous éclaircissement faible et à la formation d'ubiquinol sous éclaircissement fort.

La région la plus caractéristique pour le premier spectre (en bleu) est la bande Amide II où une bande différentielle est observable, composée d'un pic négatif à 1539 cm^{-1} et d'un positif à 1556 cm^{-1} . Des contributions importantes sont aussi présentes dans la bande Amide I (vers 1665 cm^{-1}). Ces deux observations peuvent être reliées à un important réarrangement protéique. Les profils de concentration indiquent en outre que celui-ci se produit très rapidement en réponse à l'excitation lumineuse (en environ 800 ms) puis stagne jusqu'à l'extinction de la lampe après 4,3 s. L'interprétation affirmant que le premier spectre n'est

relié qu'aux changements de conformation des protéines formant l'environnement du centre réactionnel est appuyé par le fait qu'aucune des bandes caractéristiques de la formation d'ubiquinol ne soient clairement décelables.

Les bandes de formation de l'ubiquinol sont visibles dans le second spectre (en rouge). Les bandes positives à 1373 cm^{-1} , 1433 cm^{-1} , 1471 cm^{-1} et 1491 cm^{-1} sont caractéristiques de la formation d'ubiquinol QH_2 . Ceci est confirmé par la présence de bandes négatives à 1288 cm^{-1} et 1263 cm^{-1} correspondant à l'oxydation (consommation) de molécules d'ubiquinone. Pour ce spectre une contribution est encore détectable dans la bande Amide I (entre 1610 et 1680 cm^{-1}) mais aucune autre contribution n'est décelable dans la bande Amide II. Les chaînes protéiques environnantes peuvent donc encore contribuer à ce spectre de différence [122] sans affecter la bande Amide II [123]. Le profil de concentration correspondant à ce spectre est présenté Figure 58. Il montre la formation d'ubiquinol, qui augmente lentement durant la photoexcitation, atteignant son maximum à la fin de la période d'éclairement (à 4,3 s). Dans l'obscurité la concentration en ubiquinol décroît lentement vers l'état original qu'elle atteint aux alentours de 50 s. Ceci reflète le processus de réoxydation de l'ubiquinol par des donneurs d'électrons extérieurs. Il est intéressant de noter que l'intensité de cette contribution décroît régulièrement de \mathbf{D}_3 à \mathbf{D}_5 . Ces trois matrices ont été enregistrées en utilisant la même intensité lumineuse. Nous avons proposé d'attribuer cette décroissance a donc été attribuée à un effet de fatigue de l'échantillon dû aux illuminations répétées.

Les bandes de formation de QH_2 peuvent être retrouvées dans le troisième spectre mais sont alors accompagnées d'un pic à 1566 cm^{-1} . L'hypothèse la plus probable est que ce spectre correspond à une voie alternative de formation d'ubiquinol impliquant un réarrangement protéique légèrement différent. De récentes études ont montré qu'une partie importante de la réserve de quinone de la cellule se distribue dans des réserves secondaires situées dans le super-complexe protéique entourant le centre réactionnel [124, 125]. Une seconde réserve de quinones serait donc impliquée, ce qui nécessiterait un réarrangement protéique différent pour permettre la circulation des molécules d'ubiquinone et d'ubiquinol. De plus, ceci est en accord avec le fait que cette contribution ne soit observée que sous un éclairage plus intense. Les profils de concentration des deux contributions reliées à la formation d'ubiquinol (rouge et verte) présentent une allure cinétique et il peut être intéressant d'appliquer un modèle cinétique lors de la résolution.

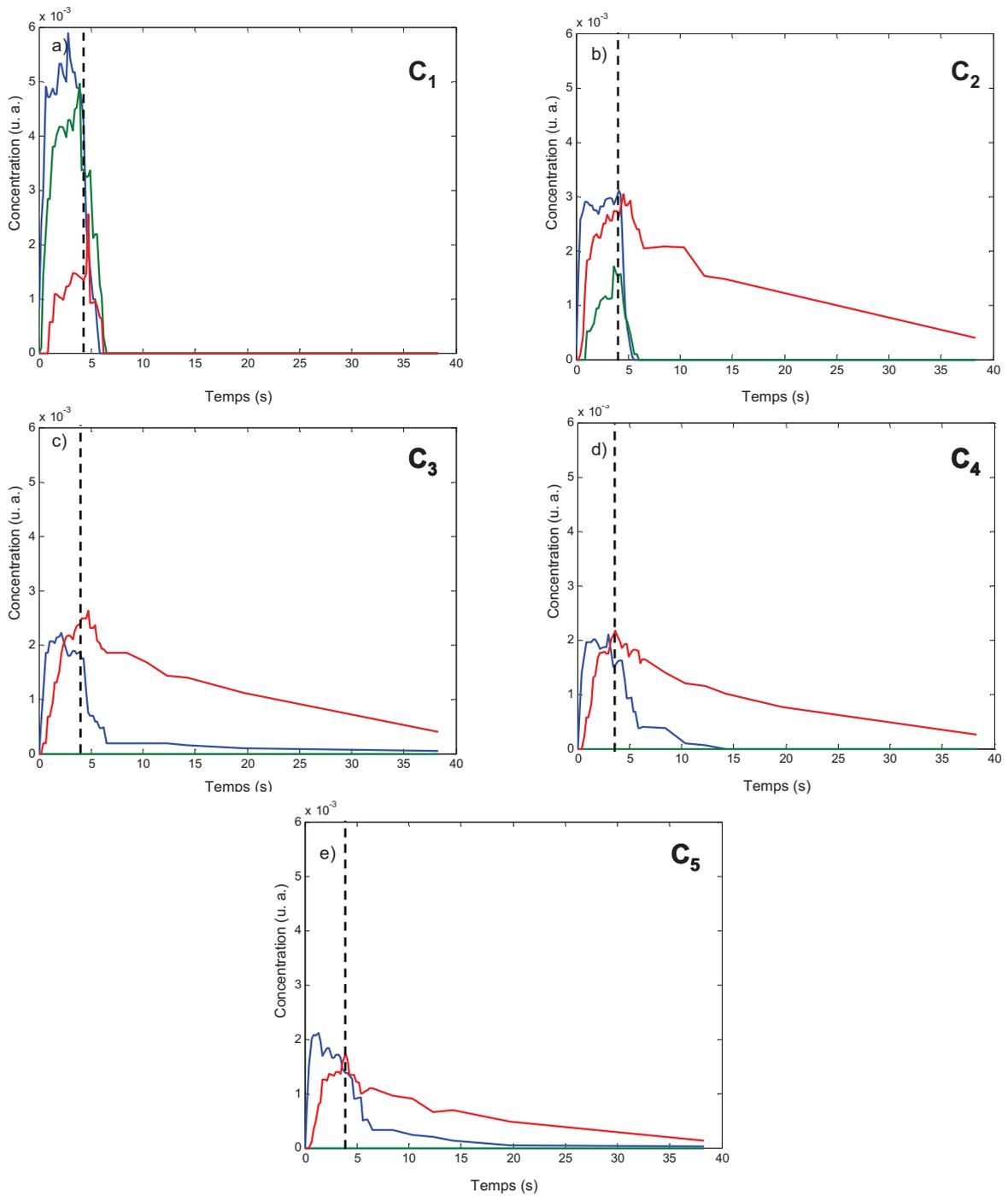


Figure 58 : Profils de concentration a) C₁, b) C₂, c) C₃, d) C₄ et e) C₅ en fonction du temps calculés en utilisant MCR-ALS sur la matrice augmentée. Le profil correspondant aux variations de l'environnement protéique est en bleu, à la formation d'ubiquinol sous éclairage faible en rouge et à la formation d'ubiquinol sous éclairage fort en vert.

5. Résolution avec la méthode hybride HS-MCR

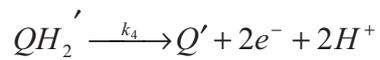
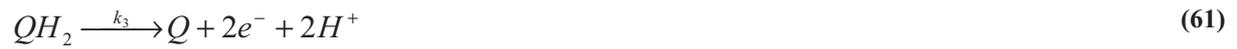
5.1 Choix d'un modèle cinétique

Les profils de concentration détectés en utilisant l'EFA et MCR-ALS peuvent en partie être soumis à un modèle cinétique et donc l'utilisation d'une méthode hybride de modélisation de procédés pourrait être le choix optimal pour ce genre de problème. En effet, la contribution attribuée au réarrangement des protéines entourant le centre réactionnel pose un problème. En principe la réponse de ces protéines à la photoexcitation est très rapide. La validité de modèles chimiques simples pour un tel phénomène, impliquant des molécules complexes, n'est pas parfaitement assurée. L'algorithme hybride de résolution de courbes multivariées (HS-MCR, *Hard-Soft-Multivariate Curve Resolution*) [96] offre donc des possibilités intéressantes. La contribution des protéines peut être analysée sous contraintes souples simultanément à l'application d'un modèle cinétique pour décrire les deux voies de formation de l'ubiquinol. On considère cette contribution protéique comme s'il s'agissait d'un interférent, hors du modèle, comme on l'a vu au CAS2 du chapitre II [102]. Une difficulté supplémentaire doit être évoquée, il est nécessaire de distinguer deux périodes dans chacune des expériences : la photoexcitation et la relaxation. En effet les conditions expérimentales sont différentes non seulement entre les matrices mais aussi entre le début et la fin de chaque expérience. En effet durant 4,3 secondes l'échantillon est photo-excité puis il est laissé en relaxation dans l'obscurité. Les modèles cinétiques pour ces deux phases doivent donc être différents et être connectés l'un à l'autre.

Durant la phase d'excitation, seuls les profils de concentration correspondant aux deux voies d'oxydation de l'ubiquinone en ubiquinol peuvent être soumis à un modèle cinétique. L'ubiquinone Q, étant présente initialement dans le centre réactionnel, n'a pas de contribution mesurable en spectroscopie de différence telle que l'IRTF *rapid-scan*. On peut seulement observer sa disparition en suivant le pic négatif à 1264 cm^{-1} . Le point de départ du modèle cinétique doit donc être une espèce transparente (sans spectre). Cette espèce sera impliquée dans les deux réactions menant à la formation d'ubiquinol QH₂. On distingue les deux procédés grâce à la présence (ou non) de la bande dans la zone Amide II à 1566 cm^{-1} . Le modèle cinétique de la phase d'excitation, régi par deux constantes cinétiques k_1 et k_2 , peut donc être résumé comme suit :



La phase de relaxation répond aux mêmes contraintes : seules les réactions concernant l'ubiquinol et l'ubiquinone peuvent être modélisées cinétiquement. Le modèle appliqué ici est rigoureusement l'inverse du précédent avec des constantes de réactions k_3 et k_4 plus petites car la relaxation est plus lente.



(62)

Les contraintes appliquées aux données expérimentales lors de cette résolution sont en partie souples (pour la contribution associée aux protéines) et rigides (à travers l'application du modèle ci-dessus), le schéma proposé Figure 59 permet de mieux comprendre comment sont gérées les différentes contributions, les profils de concentration et les spectres correspondants.

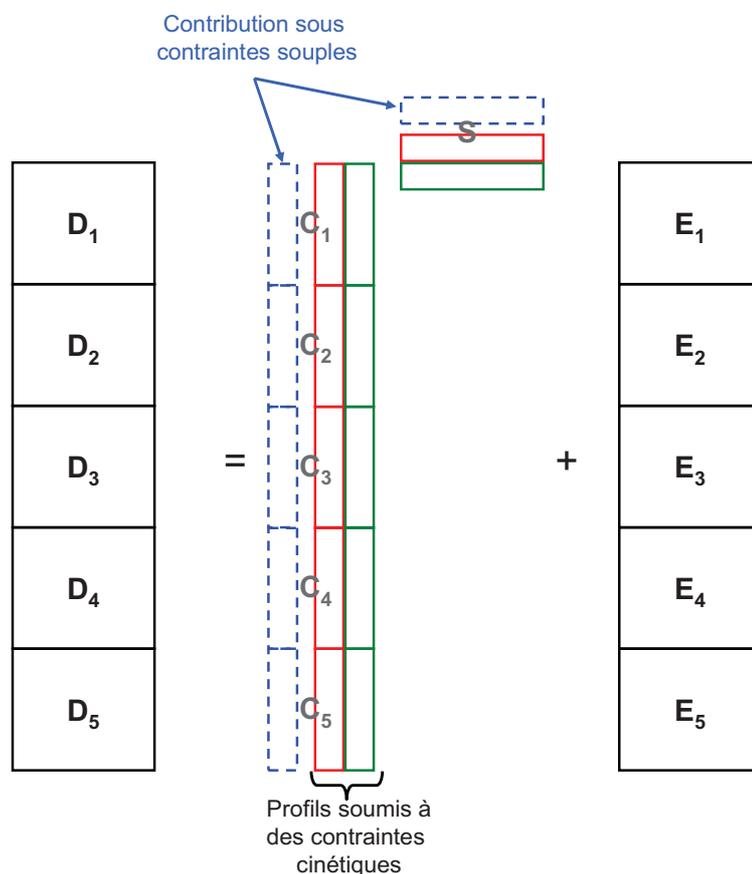


Figure 59 : Schéma de l'augmentation de matrices [D_1 ; D_2 ; D_3 ; D_4 ; D_5] et du type de contraintes appliquées à chaque contribution, la contribution protéique (en bleue) est sous contraintes souples, les deux autres doivent respecter des contraintes cinétiques.

Cependant une difficulté supplémentaire doit être prise en compte. En effet l'application d'une contrainte cinétique dans HS-MCR nécessite non seulement de choisir un modèle réactionnel mais aussi de fournir les concentrations initiales de chacune des espèces impliquées. Ce n'est pas un problème pour la phase de photoexcitation (début du procédé) mais cela devient critique pour la phase de relaxation. En effet, les concentrations initiales du modèle de relaxation doivent correspondre aux concentrations finales de la phase d'excitation, or les concentrations initiales du modèle de relaxation sont dépendantes du résultat de la modélisation de la phase de photoexcitation. Il faut donc envisager soit un modèle prévoyant à la fois l'excitation et la relaxation soit un ajustement séquentiel, où la phase d'excitation est

modélisée en premier, puis la phase de relaxation. Cette dernière option permet de récupérer à chaque itération les dernières concentrations de la période d'exposition à la lumière pour les injecter en tant que concentration au temps zéro de la période de relaxation. La souplesse des approches basées sur l'algorithme MCR-ALS permet de gérer cette situation. L'algorithme HS-MCR a été modifié de façon à prendre en compte cette spécificité. Le schéma de la Figure 60 montre comment chaque matrice de données doit être considérée comme deux sous matrices suivant deux modèles différents mais interdépendants. Cette division a été appliquée pour toutes les matrices simultanément, on peut donc considérer que l'augmentation de matrices, dont les résultats sont présentés dans la section suivante, est constituée de dix sous matrices : cinq pour la phase d'excitation et cinq pour la phase de relaxation. Cette approche originale n'est possible qu'en utilisant une approche hybride. Les méthodes de modélisation rigide nécessitent en effet de fixer tous les paramètres avant le début de l'ajustement

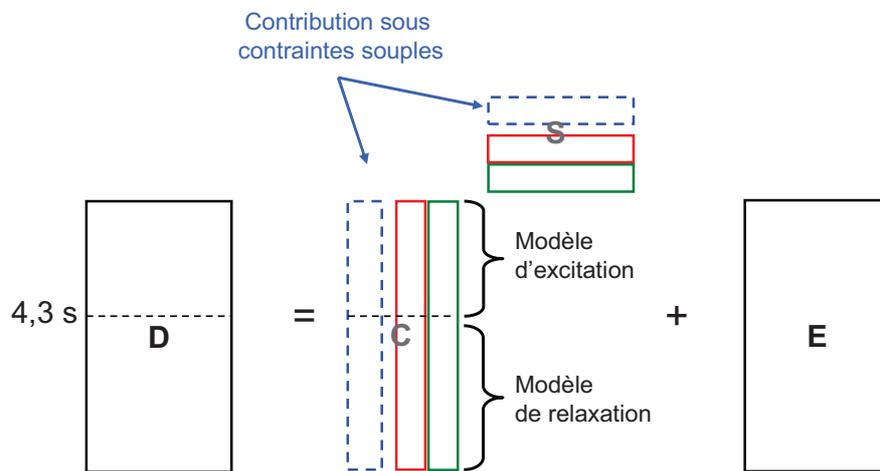


Figure 60 : Application de deux modèles interdépendants (excitation et relaxation) pour chacune des matrices D_1 à D_5 .

5.2 Résultats de l'application de la méthode hybride HS-MCR

La résolution HS-MCR présentée ci-dessous a été calculée, comme lors de la résolution MCR-ALS, sur la matrice augmentée dans le sens des colonnes [D_1 ; D_2 ; D_3 ; D_4 ; D_5] [126]. Les spectres purs S^T obtenus sont présentés Figure 61 et les profils de concentrations C_1 , C_2 , C_3 , C_4 et C_5 en fonction du temps Figure 62. Dans ce cas le pourcentage de variance expliqué atteint 92,4 % et les résidus correspondent à un bruit indépendant, plus important dans la bande Amide I. Les solutions obtenues par HS-MCR sont en accord avec celles obtenues initialement par MCR-ALS ce qui est un bon indicateur de la validité de la résolution et du modèle cinétique choisi. Le pourcentage de variance expliquée

reste comparable à celui obtenu par MCR-ALS et les résultats sont interprétables chimiquement, ce qui sans valider formellement le modèle choisi ne permet pas de l'écarter. L'interprétation des spectres et de profils de concentration obtenus en utilisant HS-MCR est identique à celle proposée pour les résultats de MCR-ALS. Il est cependant possible de discuter de deux points supplémentaires : les valeurs des constantes cinétiques obtenues et la robustesse de l'analyse.

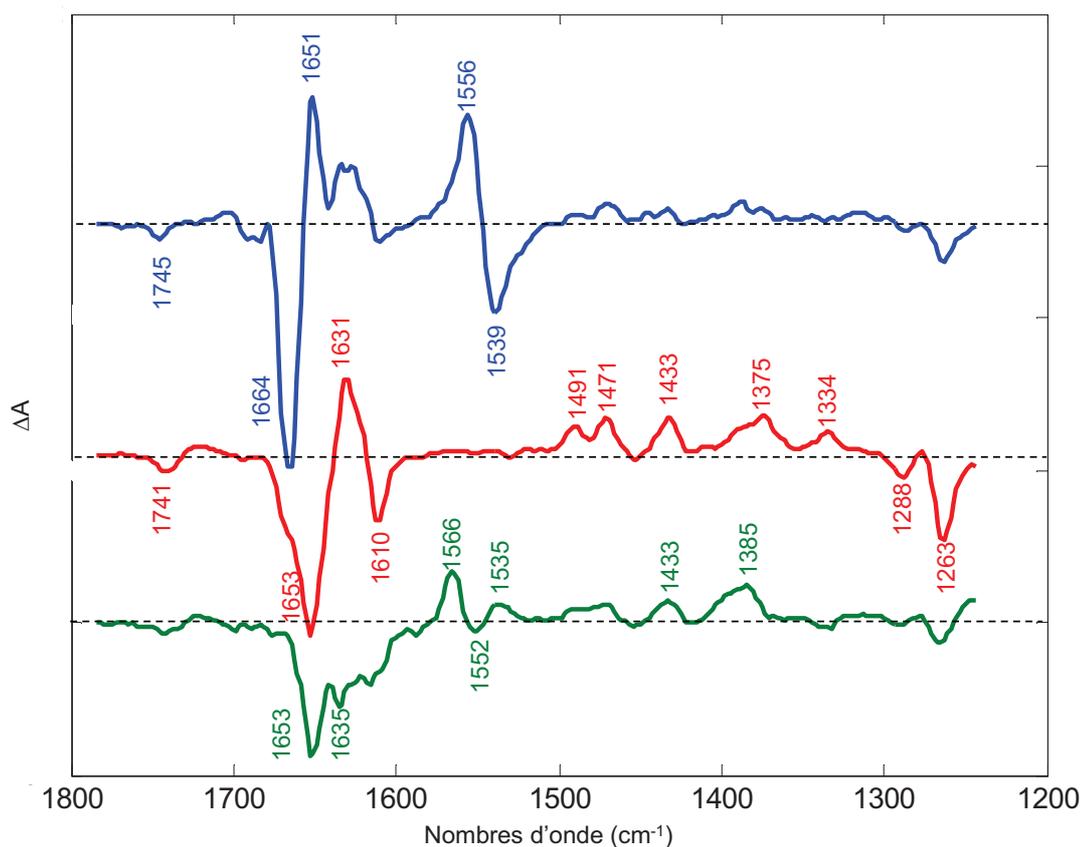


Figure 61 : Spectres de différence S^T issus de la résolution HS-MCR de la matrice augmentée [D_1 ; D_2 ; D_3 ; D_4 ; D_5]. De haut en bas spectres attribués au changement de conformation de la protéine, à la formation d'ubiquinol sous éclairage faible et à la formation d'ubiquinol sous éclairage fort.

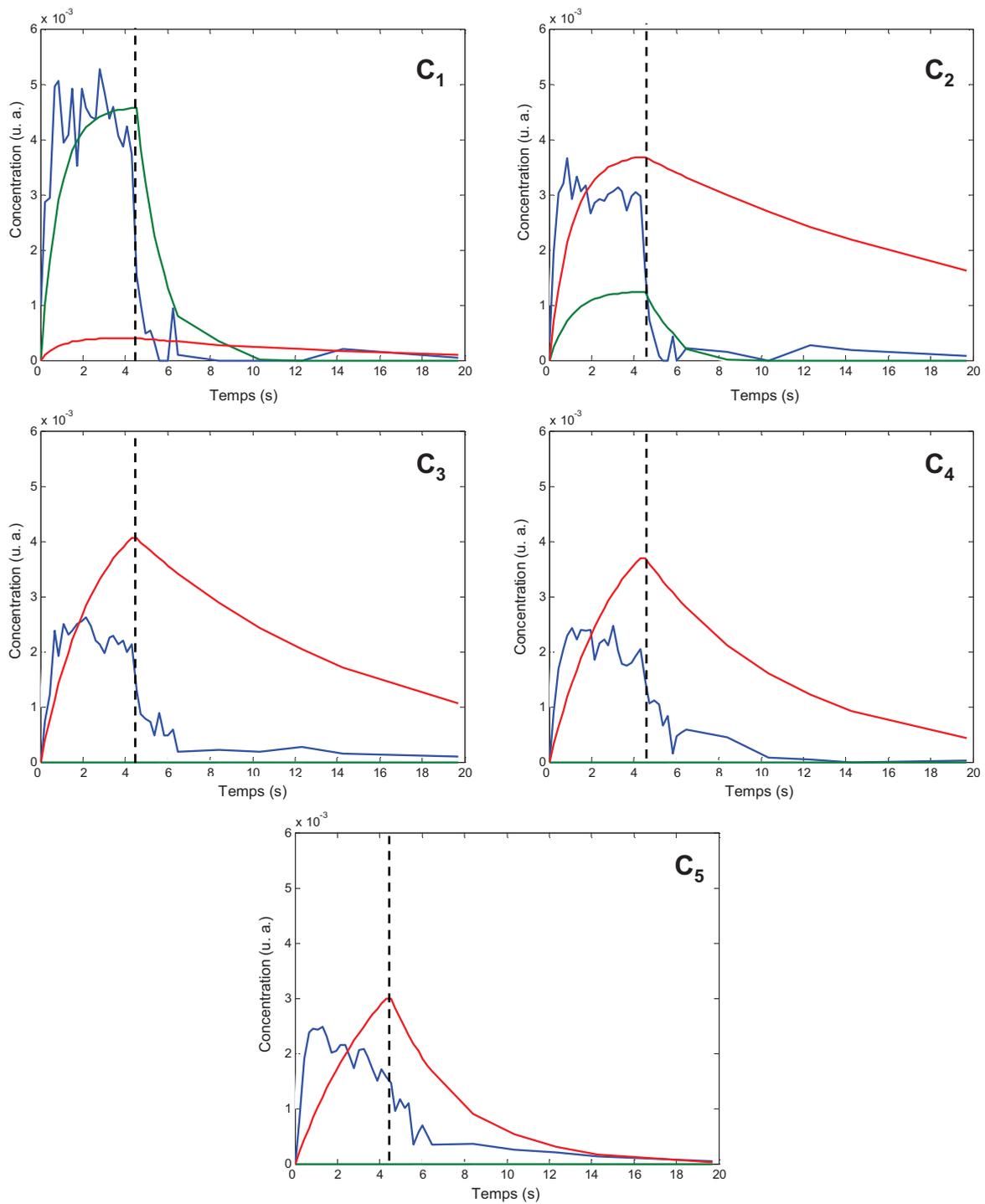


Figure 62 : Profils de concentrations a) C₁, b) C₂, c) C₃, d) C₄ et e) C₅ en fonction du temps calculé en utilisant HS-MCR. Le profil correspondant à la variation de l'environnement protéique est en bleu, à la formation d'ubiquinol sous éclairage faible en rouge et à la formation d'ubiquinol sous éclairage fort en vert.

L'un des intérêts de l'application d'une contrainte de modélisation rigide est que celle-ci donne accès aux constantes cinétiques des deux processus de formation de l'ubiquinone. Ces constantes sont regroupées dans le

Tableau 6. L'influence de l'intensité lumineuse est visible lors des phases d'excitation : plus la luminosité est intense (comme dans **D₁** et **D₂**) plus les processus de formation d'ubiquinol (considérés ensemble) sont favorisés (la somme $k_1 + k_2$ augmente). Cependant on constate que la seconde voie devient prédominante lorsque l'intensité de l'excitation lumineuse augmente, c'est-à-dire dans le cas de **D₁** ($k_2 > k_1$). En outre les constantes de réaction k_i des matrices **D₃**, **D₄** et **D₅** (mêmes conditions d'excitation) reflètent de nouveau l'effet de fatigue de l'échantillon. La phase de relaxation est quant à elle dépendante de l'état du système atteint à la fin de l'excitation. Elle est beaucoup plus rapide lorsque le second processus de formation de l'ubiquinol est mis en jeu. On peut aussi constater que la fatigue de l'échantillon tend à accélérer sa relaxation. Il est aussi important de noter comment l'intensité des profils de concentration de la Figure 62 confirme que, sous éclairage intense, la deuxième voie de formation de l'ubiquinol (profil vert) est clairement préférée à celle qui est présente à des intensités basses d'éclairage (profil rouge).

Expérience	Excitation		Relaxation	
	k_1	k_2	k_3	k_4
D₁	$0,09 \pm 0.02 \text{ s}^{-1}$	$1,06 \pm 0.08 \text{ s}^{-1}$	$0,09 \pm 0.01 \text{ s}^{-1}$	$0,82 \pm 0.01 \text{ s}^{-1}$
D₂	$0,73 \pm 0.05 \text{ s}^{-1}$	$0,25 \pm 0.03 \text{ s}^{-1}$	$0,05 \pm 0.01 \text{ s}^{-1}$	$0,69 \pm 0.01 \text{ s}^{-1}$
D₃	$0,39 \pm 0.01 \text{ s}^{-1}$	-	$0,09 \pm 0.01 \text{ s}^{-1}$	-
D₄	$0,31 \pm 0.01 \text{ s}^{-1}$	-	$0,14 \pm 0.01 \text{ s}^{-1}$	-
D₅	$0,22 \pm 0.01 \text{ s}^{-1}$	-	$0,29 \pm 0.04 \text{ s}^{-1}$	-

Tableau 6 : Constantes de réactions calculées en utilisant HS-MCR

Pour finir, un autre avantage de l'inclusion de contraintes rigides est l'augmentation de la robustesse du modèle proposé. La résolution est en effet moins sensible à la présence de bruit et aux ambiguïtés de rotation, ces deux facteurs n'étant pas compatibles avec les modèles cinétiques. La combinaison de l'augmentation de matrices et de la modélisation hybride permet d'assurer la cohérence des résultats obtenus sur les différentes expériences. Il est à noter qu'il est possible d'imposer un modèle commun à toutes les matrices [100], ceci améliore encore la robustesse du modèle proposé lorsque le système étudié est parfaitement

répétable. Ce n'est pas le cas d'un échantillon biologique car sa nature et l'effet de fatigue induisent des variations incontrôlables et/ou imprévisibles. L'application de modèles locaux a donc été préférée.

6. Conclusion

L'analyse de spectres de différence IRTF résolus en temps concernant l'étude du centre réactionnel de *Rhodobacter sphaeroides* comporte des difficultés spécifiques. Ces dernières, liées à la nature intrinsèque des données de différence et à la présence d'interférents, imposent d'adapter les outils de modélisation existants. En effet les techniques de modélisation paramétriques, potentiellement utiles pour l'analyse de données cinétiques, ne sont pas valides en présence d'interférents. Le concept d'interférents est ici élargi à tout type de contribution ne suivant pas rigoureusement un modèle physico-chimique. Ceci peut prendre la forme d'un composé absorbant inerte ou de procédés évoluant à des échelles de temps différentes de celle du procédé d'intérêt. L'approche hybride HS-MCR offre dans ce cas différents avantages qui permettent de surmonter ces difficultés. L'application de contraintes souples permet de gérer les contributions interférentes, comme le changement de conformation des protéines entourant le centre réactionnel de *Rhodobacter sphaeroides*. Les contraintes rigides permettent de proposer une modélisation cinétique d'une partie des procédés observés et cela permet de réduire l'influence des ambiguïtés de rotation sur la résolution. Des informations supplémentaires sont alors fournies sur la nature du modèle et sur les constantes de réaction correspondantes.

Certains développements algorithmiques ont été nécessaires à l'analyse de ces données. Le premier est la gestion de données de différence, il a été exposé dans le Chapitre II. Le second point plus spécifique à ce sujet est la possibilité de modéliser de façon cohérente des phases successives d'excitation et de relaxation en utilisant des contraintes cinétiques.

Ces adaptations répondent aux besoins rencontrés lors du suivi d'une partie du processus de photosynthèse de *Rhodobacter sphaeroides* par IRTF *rapid-scan*. Une interprétation *soft-modelling* a d'abord été proposée [59] puis améliorée [102] par l'application de cette méthode [126]. Deux processus compétitifs de formation de l'ubiquinol ont ainsi pu être détectés. L'influence de l'intensité d'éclairement a pu être démontrée : lorsque l'intensité est suffisamment importante il semble que l'échange ubiquinone-ubiquinol s'effectue avec une seconde réserve de quinones située dans le super complexe protéique entourant le centre réactionnel.

**Chapitre IV : Étude de la photochimie de la benzophénone
par spectroscopie d'absorption transitoire**

L'objectif de ce chapitre est d'appliquer les développements méthodologiques détaillés dans le chapitre II à l'analyse de la relaxation de la benzophénone après une photoexcitation. Ce phénomène est l'un des plus importants processus de la photochimie organique car il a un rôle fondamental dans la photoréactivité des cétones aromatiques. Pourtant le mécanisme suivi entre les différents états électroniques dérivés de cette excitation n'est pas clair et certains aspects font débat, tels que le mécanisme de formation des états triplets et le croisement intersystème observé pour la benzophénone aux temps courts, qui est d'après la théorie connue impossible. Nous proposons d'allier la chimométrie à la spectroscopie d'absorption transitoire pour lever le voile sur cette question.

Les principes généraux de la photochimie, nécessaires à la compréhension du système constitué par la benzophénone, seront résumés dans une première partie. Une description plus spécifique des systèmes chimiques utilisés pour cette étude sera ensuite proposée. Les méthodes instrumentales seront présentées avant de passer à l'analyse chimométrique des données obtenues.

L'analyse en elle-même sera détaillée comme suit. Une première approche exploratoire des données sera proposée et permettra de définir les spécificités des données et les outils applicables à ce cas. Ensuite nous proposerons les résolutions souples correspondant aux deux systèmes utilisés. Finalement, d'après les résultats obtenus avec la résolution souple, des modèles cinétiques seront proposés et l'application de ceux-ci sera discutée. La validité des résultats obtenus sera cette fois vérifiée par différentes approches chimométriques et en utilisant un second système chimique comparable à la benzophénone.

1. Principes de la photochimie

La photochimie est l'étude des réactions chimiques initiées par la lumière. Les effets de la lumière sur des substances pures ont été observés pour la première fois par Scheele en 1775 qui avait noté que les sels d'argent noircissaient lors d'une exposition au soleil. Ce n'est qu'en 1820 que Grotthaus (puis Draper en 1843) reconnut ce phénomène en tant que réaction chimique catalysée par un rayonnement lumineux. La photochimie est donc, comme la spectroscopie, à l'interface de la chimie et de la physique, voire de la biologie comme on l'a vu précédemment avec la photosynthèse.

Les principes de la photochimie seront résumés dans les paragraphes suivants, nous détaillerons plus particulièrement les mécanismes d'excitation puis de relaxation. Certaines propriétés des états excités seront ensuite évoquées.

1.1 Principes

Les réactions photochimiques mettent en jeu des états électroniques excités, ce qui constitue une différence importante par rapport aux états fondamentaux vibrationnellement excités. Cette différence a deux conséquences importantes du point de vue de la compréhension de la réaction chimique. D'une part, une excitation monochromatique peut conduire à la préparation d'un état électronique excité d'énergie contrôlée dont on peut suivre la réactivité et, d'autre part, l'utilisation de lasers pulsés délivrant des impulsions aussi brèves que quelques femtosecondes permet l'étude de la dynamique de réactions chimiques ultrarapides.

Dans une réaction photochimique, on distingue trois étapes. La première est l'étape d'absorption de la lumière, c'est-à-dire l'excitation. La seconde étape est constituée par les processus photophysiques primaires. Ce sont les processus qui mettent en jeu les états électroniques excités : fluorescence, phosphorescence, transferts d'énergie... La troisième étape est constituée par les processus secondaires : il s'agit de l'ensemble des réactions qui ont lieu à partir des intermédiaires réactionnels formés.

La première loi de la photochimie, connue sous le nom de loi de Grotthaus-Draper, énonce que la lumière doit être absorbée par une substance chimique pour qu'une réaction photochimique ait lieu. La seconde loi, la loi de Stark-Einstein, veut que pour chaque photon

(quantum) absorbé par un système chimique, une seule molécule soit excitée. Cette loi est connue comme la loi de photoéquivalence et a été proposée par Albert Einstein durant le développement de la théorie quantique de la lumière. La lumière fournit l'énergie d'activation nécessaire à la réaction, mais elle permet aussi de changer la symétrie de la configuration électronique d'une molécule. Ceci donne accès à des chemins réactionnels spécifiques, comme les décrivent les règles de sélection de Woodward-Hoffmann [127].

L'efficacité avec laquelle un processus photochimique donné se produit est définie par son rendement quantique. Beaucoup de processus photochimiques sont complexes et peuvent être en concurrence avec des pertes d'énergie non productives. Le rendement quantique est donc généralement défini pour un événement particulier. Par conséquent nous pouvons définir le rendement quantique comme le nombre de moles de réactif consommées (ou de produit formées) par mole de photons (ou einstein).

1.2 Mécanisme d'excitation

L'absorption d'un rayonnement lumineux par une molécule introduit une énergie suffisante pour casser ou réorganiser la plupart des liaisons covalentes. La lumière requise pour une réaction photochimique peut provenir de différentes sources. Au début du vingtième siècle, la lumière solaire était directement utilisée. De nos jours, en fonction des composés étudiés et de l'information désirée, on utilise les lampes à incandescence (visible ou IR), des lampes à vapeur de mercure (émettant selon la pression à 185 - 255 nm, 255 -1000 nm et 220 -1400 nm), des sources flash ou des lasers. Certaines études sur des systèmes sensibles peuvent nécessiter l'utilisation de lumières monochromatiques.

Dans cette section nous nous intéresserons à la nature et au comportement des états électroniques excités, formés par l'absorption d'un photon par un groupe fonctionnel photochromique (ayant l'aptitude de se colorer sous un rayonnement lumineux). Une telle excitation provoque un changement d'occupation des orbitales moléculaires, un accroissement d'énergie et des changements dans les liaisons et la distribution de charge.

Commençons par considérer les cas simples de molécules diatomiques comme Cl_2 ou Br_2 . Le diagramme présenté Figure 63 illustre l'excitation électronique initiale. Le niveau fondamental et le niveau excité sont représentés comme des niveaux énergétiques peuplés par

des états vibrationnels et des états rotationnels (non représentés). La réorganisation électronique se produisant lorsque le niveau fondamental est excité par l'absorption d'un photon est beaucoup plus rapide que n'importe quel mouvement du noyau moléculaire qui pourrait suivre ce processus. En d'autres mots, les décalages électroniques se produisent comme si le noyau plus lourd restait immobile [128]. Cette conséquence de l'approximation de Born-Oppenheimer mène à la formulation du principe de Franck-Condon : la transition électronique est beaucoup plus rapide qu'un mouvement de vibration de telle sorte que la distance internucléaire r demeure inchangée au cours de la transition [129, 130].

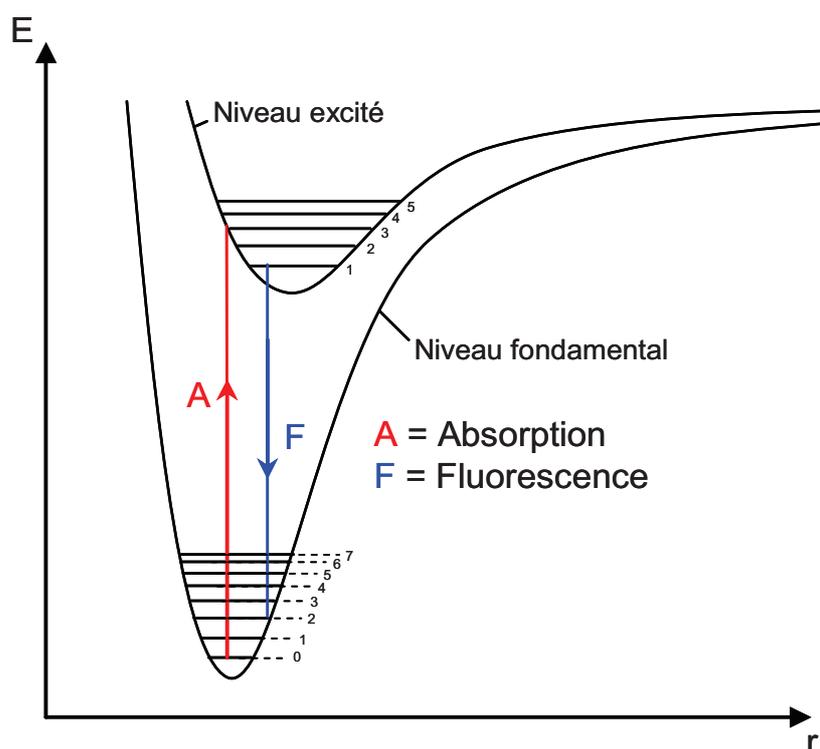


Figure 63 : Excitation électronique d'une molécule diatomique

Considérons le passage d'un photon à travers le nuage électronique d'une molécule. En supposant que le nuage ait un diamètre de 0,3 nm, la perturbation électromagnétique le traverse en 10^{-18} s. On admet que la transition, le saut électronique s'opère dans un temps très court de l'ordre de 10^{-15} - 10^{-17} secondes. Considérons maintenant une vibration moléculaire. Ce mouvement intramoléculaire du réseau nucléaire est beaucoup plus lent. En réalité, les vibrations les plus rapides s'effectuent en quelque 10^{-13} s. Il s'ensuit que la géométrie de la molécule demeure inchangée pendant un saut électronique. Cette remarque est également valable en ce qui regarde les mouvements de rotation.

Du point de vue de la mécanique quantique, cela veut dire qu'un saut électronique ne se produit que si les fonctions d'onde associées aux deux états se chevauchent. On interprète cette condition en termes classiques en disant que la distance internucléaire r ne doit pas changer lors du saut électronique.

1.3 Mécanisme de relaxation

A température ambiante quasiment toutes les molécules se trouvent au niveau fondamental. Le principe de Franck-Condon implique que l'excitation se traduise par une transition verticale, et donc que les niveaux vibrationnels de l'état excité se peuplent. Plusieurs événements peuvent ensuite avoir lieu [127] :

- l'énergie vibrationnelle peut être perdue sous forme de chaleur, relaxant l'état excité vers le niveau fondamental de vibration.
- l'état excité peut retourner au niveau fondamental en émettant un photon. Si ceci se produit depuis un niveau vibrationnel nul, la fréquence (ou l'énergie) de la lumière émise sera plus basse que celle initialement absorbée (voir Figure 63). Cette émission est appelée fluorescence et se produit rapidement après l'excitation. Le terme phosphorescence se réfère à une désexcitation plus lente impliquant d'autres états excités (états triplets).
- si un état vibrationnel élevé est atteint la molécule peut se casser en deux radicaux.

L'écrasante majorité des composés organiques est constituée de plus de deux atomes, le diagramme d'énergie des molécules diatomiques doit donc être ajusté au nombre croissant de liaisons. Une façon de le faire consiste à conserver la dimension énergétique et négliger la dimension spatiale (r). Le diagramme de Perrin-Jablonski ci-dessous, Figure 64, en est un exemple. Ce diagramme représente les états électroniques d'une molécule type, la sous structure vibrationnelle de chacun de ces états et les transitions mises en jeu en photochimie. Une importante caractéristique exprimée par ce diagramme est qu'il est probable que plus d'un état électronique excité existe pour une molécule donnée, ici six sont dessinés. Chaque état électronique aura un groupe d'états vibrationnels (et rotationnels). Les transitions entre les états électroniques se produisent souvent aux niveaux vibrationnels élevés qui se relaxent ensuite par perte de chaleur (énergie de translation) à des niveaux vibrationnels plus bas. Il est important de se souvenir que ce diagramme est surtout valable pour les molécules en phase

condensée et ne s'applique pas aux molécules isolées (phase gazeuse et jets moléculaires) car celles-ci ne peuvent pas perdre leur énergie par collision avec leurs voisins.

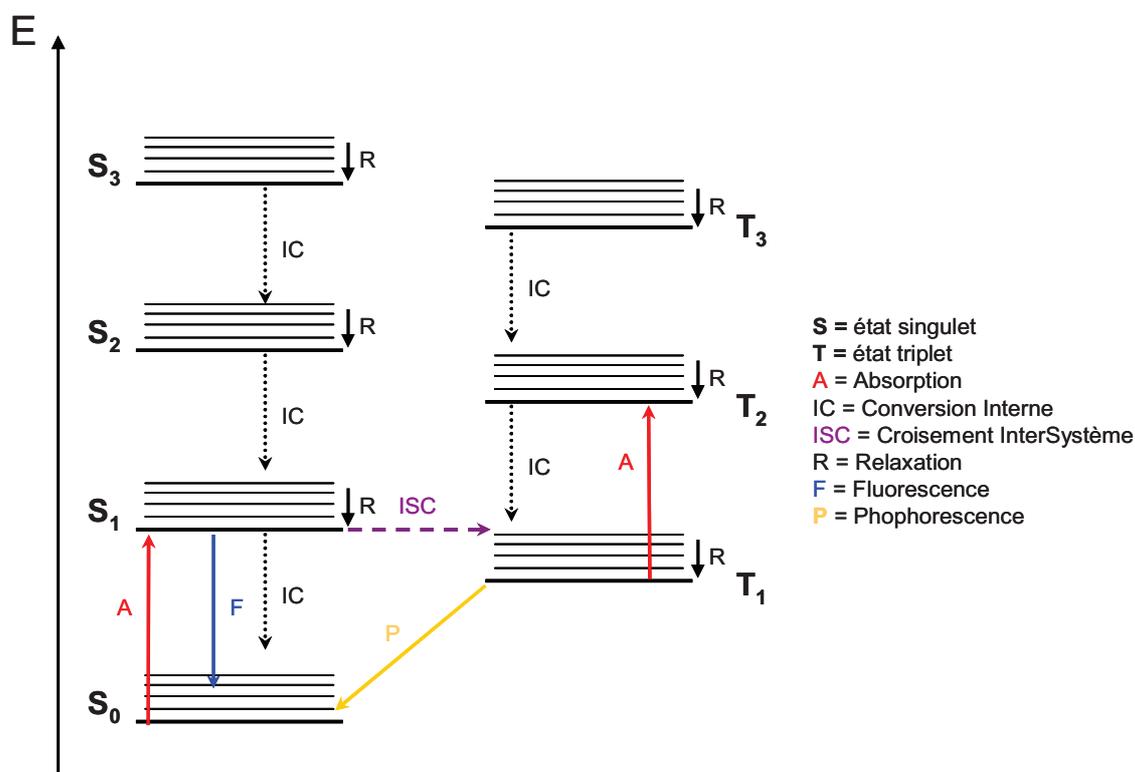


Figure 64 : Diagramme de Perrin-Jablonski représentant les différentes transitions électroniques possibles.

Les états excités peuvent être qualifiés de singulet ou de triplet en fonction du moment angulaire de leurs spins électroniques. Les électrons de la plupart des composés organiques (non métalliques) sont pairs (les spins sont opposés) dans les orbitales liantes et non-liantes. L'état fondamental de ces composés est diamagnétique et a donc un spin global nul. Ces états n'ont qu'un seul niveau énergétique dans un champ magnétique donné et sont donc appelés singulet. Les états électroniques dans lesquels deux électrons avec un spin identique occupent deux orbitales différentes (selon le principe d'exclusion de Pauli) ont un spin de 1 ($2 * \frac{1}{2}$) et sont paramagnétiques. Dans un champ magnétique, ces états ont trois niveaux d'énergies (+1, 0 et -1), ils sont donc appelés triplets. La distinction entre les états singulet et triplet est importante car l'absorption d'un photon mène toujours à un état de même multiplicité, c'est-à-dire d'un état singulet à un état singulet ou d'un état triplet à un état triplet. La plupart des états fondamentaux étant des singulets (à l'exception notable du dioxygène), les espèces excitées initialement formées sont en général des singulets.

La conversion interne (IC) des états excités vers des états de plus basse énergie de même multiplicité se produit rapidement et est accompagnée d'une perte d'énergie (relaxation). Alternativement, un état excité peut retourner à l'état fondamental en émettant un photon (désexcitation radiative). C'est le cas de la fluorescence si la désexcitation se produit entre états singulets. La conversion d'un état singulet en un état triplet de plus basse énergie (ou vice-versa) est appelée croisement intersystème (ISC) et est plus lente que la conversion interne. L'application de la règle d'or de Fermi aux ISC permet de déduire les règles d'El Sayed, présentées dans le Tableau 7. Ces règles ne sont pas absolues mais les transitions interdites ne s'effectuent qu'à des vitesses beaucoup plus faibles (de 100 à 1000 fois) que celles qui sont autorisées.

$S_1 \rightarrow T_1$	$^1(n, \pi^*) \leftrightarrow ^3(n, \pi^*)$	interdit
	$^1(\pi, \pi^*) \leftrightarrow ^3(n, \pi^*)$	interdit
	$^{1,3}(n, \pi^*) \leftrightarrow ^{3,1}(\pi, \pi^*)$	permis

Tableau 7 : Règles d'El Sayed pour les composés carbonyles [131], transitions entre les singulets (notés ¹) et les triplets (notés ³).

Enfin, la désexcitation radiative provenant des états triplets est appelée phosphorescence et est généralement lente. Les échelles de temps de ces différents phénomènes sont fournies par le Tableau 8.

	Transition	Temps (s)
Absorption (excitation)	$S_0 \rightarrow S_n$	10^{-15}
Conversion Interne (IC)	$S_n \rightarrow S_1$	10^{-14} à 10^{-11}
Relaxation vibrationnelle (R)	$S_n^* \rightarrow S_n$	10^{-12} à 10^{-10}
Croisement InterSystème (ISC)	$S_1 \rightarrow T_1$	10^{-11} à 10^{-6}
Fluorescence (F)	$S_1 \rightarrow S_0$	10^{-9} à 10^{-6}
Phosphorescence (P)	$T_1 \rightarrow S_0$	10^{-3} à 100

Tableau 8 : Échelles de temps des différents mécanismes photochimiques.

1.4 Propriétés des états excités

La photochimie d'une molécule dépend des propriétés des états excités formés suite à l'absorption d'un rayonnement lumineux. Ces états ont un excédent d'énergie qui leur permet de réagir plus facilement, et parfois différemment, que le niveau fondamental. En phase condensée, les états énergétiques les plus élevés ont des durées de vie très courtes, on ne considère donc que les états excités de basse énergie (S_1 , S_2 , T_1 et T_2). Ces états électroniques

peuvent être considérés comme des espèces physico-chimiques à part entière, avec une géométrie souvent différente de l'état fondamental, un moment dipolaire et une polarisabilité caractéristiques. Les propriétés physico-chimiques peuvent aussi varier :

- la géométrie de la molécule varie de façon à minimiser l'hypersurface de potentiel du nuage électronique excité ;
- le moment dipolaire d'une molécule polaire peut changer en fonction de l'excitation. C'est particulièrement le cas des molécules dont la partie conjuguée (ou aromatique) est substituée par un groupement donneur ou attracteur d'électrons.
- les propriétés acido-basiques peuvent se modifier lorsque la molécule passe à l'état excité (davantage pour l'état S_1 que pour T_1) ;
- un état excité est toujours plus oxydant (vis-à-vis d'un réducteur donné) ou plus réducteur (vis-à-vis d'un oxydant donné) que l'état fondamental.

Ces variations des propriétés physico-chimiques affectent nécessairement les interactions avec d'autres molécules. Ceci est à l'origine de nombreux processus photochimiques secondaires.

2. Systèmes photochimiques

Les spectroscopies résolues en temps jouent un rôle crucial dans la compréhension des processus physicochimiques ultrarapides. La dynamique des espèces transitoires (états excités, intermédiaires réactionnels) est caractérisée par une résolution temporelle très élevée grâce aux spectroscopies électroniques. Ces méthodes vont être ici appliquées à l'analyse de la photochimie d'une molécule organique bien étudiée [132-138] : la benzophénone. L'état des connaissances sur les mécanismes de relaxation de cette molécule sera exposé dans un premier temps. Une molécule parente, la méthoxybenzophénone, sera évoquée avant de passer à la description des méthodes instrumentales utilisées pour étudier ces deux systèmes.

2.1 Photochimie de la benzophénone

La benzophénone (BP), représentée Figure 65, est une molécule « modèle » en photochimie du fait du nombre d'études la concernant et de ses nombreuses utilisations, notamment comme protection de la lumière UV pour des composés organiques (parfums, colorants ou plastiques) [139].

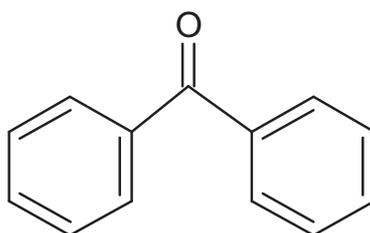


Figure 65: Formule de la benzophénone

Une des particularités rendant cette molécule si intéressante est que l'écart énergétique entre les états singulet et triplet de la benzophénone est relativement faible et le couplage spin-orbite y est important. Ceci signifie que le rendement quantique de formation de l'état triplet par photoexcitation de l'état fondamental S_0 est proche de cent pour cent. La relaxation du système est donc dominée par l'état triplet. Le processus d'excitation est relativement bien connu. L'absorption d'énergie lumineuse par la benzophénone à l'état fondamental S_0 permet de peupler les états singulets supérieurs (S_1 , S_2 etc.). Les transitions sont dépendantes de la longueur d'onde de la pompe : pour une excitation comprise en 230 et 310 nm la benzophénone suit une transition électronique $\pi \rightarrow \pi^*$, c'est-à-dire $S_0 \rightarrow S_2$. Une énergie plus faible, entre 310 et 400 nm provoque une transition $n \rightarrow \pi^*$ qui correspond à $S_0 \rightarrow S_1$ [138]. L'état S_1 peut aussi être peuplé par conversion interne (IC) depuis S_2 .

Néanmoins, les étapes menant à la formation du triplet de plus basse énergie T_1 sont encore mal comprises. Il est généralement admis que l'état triplet T_1 se peuple par croisement intersystème (ISC) depuis S_1 [140]. Plusieurs études ont été menées sur cette étape. Diverses conclusions ont été proposées [1, 3, 4]. Premièrement la décroissance de l'état S_1 est associée à la formation de l'état triplet T_1 . Cette dynamique peut être décrite par un temps caractéristique d'environ 10 ps quelque soit la longueur d'onde d'excitation. Cependant aucun de ces travaux ne permet de valider l'existence de l'ISC entre S_1 et T_1 . En effet ces deux états sont de nature (n,π^*) , donc d'après les règles d'El Sayed [131] une transition directe est interdite. Il est alors légitime de suspecter l'existence d'un état intermédiaire (π,π^*) permettant de la transition électronique depuis l'état S_1 vers T_1 . Deux hypothèses sont possibles, la première suppose que le peuplement de T_1 s'effectue par croisement intersystème depuis l'état $S_2(\pi,\pi^*)$ [132]. La formation de triplet serait alors dépendante de la longueur d'onde d'excitation, si celle-ci est trop grande, l'état S_2 n'est pas peuplé et l'ISC vers T_1 ne serait pas possible. Ceci n'est pas confirmé par l'expérience [137, 141]. La seconde hypothèse est donc plus probable. Elle suppose que l'état S_1 effectue un croisement intersystème vers un état $T_2(\pi,\pi^*)$ [132]. L'existence d'un état triplet correspondant à cette hypothèse permet de construire un modèle réactionnel en accord avec les règles d'El Sayed : un croisement intersystème $S_1(n,\pi^*) \rightarrow T_2(\pi,\pi^*)$ serait suivi par une conversion interne $T_2(\pi,\pi^*) \rightarrow T_1(n,\pi^*)$ comme le montre la Figure 66. Yabumoto *et al.* ont récemment observé la bande d'absorption T_2 - T_1 en spectroscopie IRTF résolue dans le temps [141].

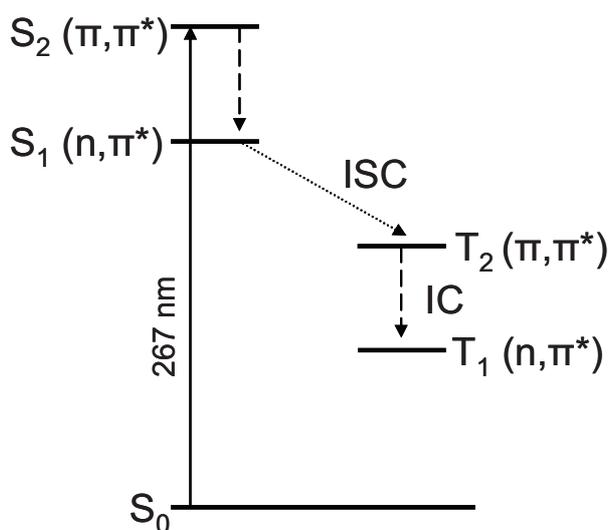


Figure 66 : Diagramme d'énergie proposé pour la benzophénone

2.2 Photochimie de la méthoxybenzophénone

Les travaux qui suivent ont pour but de mettre en évidence, par spectroscopie UV-visible et chimométrie, l'existence d'un état intermédiaire entre S_1 et T_1 et de déterminer son spectre, sa nature et sa durée de vie. La validité des méthodes chimométriques et de l'hypothèse photochimique avancée seront appuyées par une confirmation chimique issue de l'analyse d'une molécule parente de la benzophénone : la méthoxybenzophénone (MeBP), présentée Figure 67.

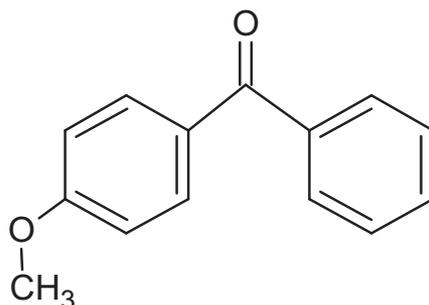


Figure 67 : Formule de la méthoxybenzophénone.

Il est donc nécessaire de décrire les particularités de cette molécule. La nature du triplet de plus basse énergie de la benzophénone n'est pas sensible au solvant et reste donc toujours (n,π^*) (sauf en présence de groupes donneurs d'électrons tels que NH_2 ou $N(CH_3)_2$) [142, 143]. Ce n'est pas le cas de la méthoxybenzophénone [144] pour laquelle la nature du triplet de plus basse énergie est dépendante du solvant. La Figure 68 montre que lorsque la polarité du solvant augmente l'écart entre $T_2(\pi,\pi^*)$ et $T_1(n,\pi^*)$ se réduit jusqu'à s'annuler pour un mélange acétonitrile (ACN) eau 3:1 puis la valeur des énergies des triplets (π,π^*) et (n,π^*) s'inverse et le triplet (π,π^*) devient T_1 lorsque la polarité du solvant est encore augmentée.

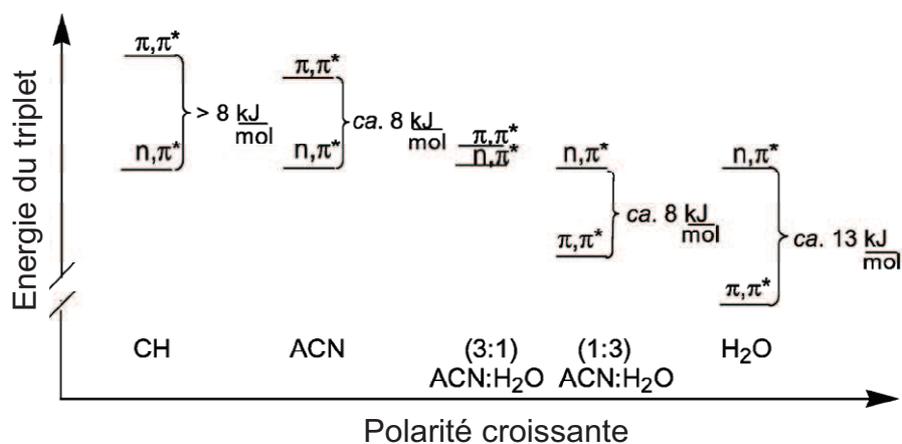


Figure 68 : Diagramme énergétique pour les états triplets les plus bas de la 4-MeBP dans des solvants de différentes polarités, d'après [144].

Cette propriété permet de choisir des voies de peuplement de l'état T_1 différentes. En effet, pour des solvants apolaires, la méthoxybenzophénone est comparable à la benzophénone, et un mécanisme $S_1(n,\pi^*) \rightarrow T_2(\pi,\pi^*) \rightarrow T_1(n,\pi^*)$, avec un intermédiaire T_2 devrait être impliqué. Pour des solvants polaires, du fait de l'inversion des états énergétiques des triplets, l'ISC directe entre S_1 et T_1 est possible selon les règles d'El Sayed puisqu'il s'agit d'un croisement $S_1(n,\pi^*) \rightarrow T_1(\pi,\pi^*)$. L'analyse du peuplement de l'état T_1 de la benzophénone nécessite l'enregistrement de données d'absorption transitoire à la fois sur la BP mais aussi sur la MeBP. Le principe de ces mesures est détaillé dans la section suivante.

3. Matériels et méthodes

3.1 Spectroscopie d'absorption transitoire

Le principe général d'une méthode de spectroscopie d'absorption transitoire pompe-sonde est de faire passer les molécules d'un système d'un état fondamental à un état excité par le biais d'une impulsion d'excitation (impulsion pompe). L'effet de l'excitation est suivi par une impulsion d'observation (impulsion sonde). On peut introduire un décalage temporel entre les deux impulsions ce qui permet de suivre l'évolution du signal en fonction du temps.

Les faisceaux de pompe et de sonde peuvent avoir des longueurs d'onde identiques, ou non, en fonction des énergies nécessaires pour les transitions électroniques. L'effet de la pompe sur le système peut alors être étudié de deux façons différentes :

- par l'observation des modifications du système après l'excitation par la pompe et par la comparaison avec son état avant l'excitation, c'est le cas des techniques d'absorption transitoire. La pompe est alors formée par un continuum de lumière blanche ;
- par l'étude de nouveaux effets produits par la sonde elle-même comme la fluorescence induite.

Dans une expérience de spectroscopie d'absorption transitoire résolue en temps, on étudie la différence d'absorption du système après l'excitation à l'aide de la sonde. L'impulsion de pompe définit le temps zéro. Un troisième faisceau, provenant de la sonde, est souvent utilisé comme référence passant à travers le système "non perturbé", comme le montre la Figure 69. Le délai entre les impulsions de pompe et de sonde est contrôlé par la longueur ΔL de la ligne à retard.

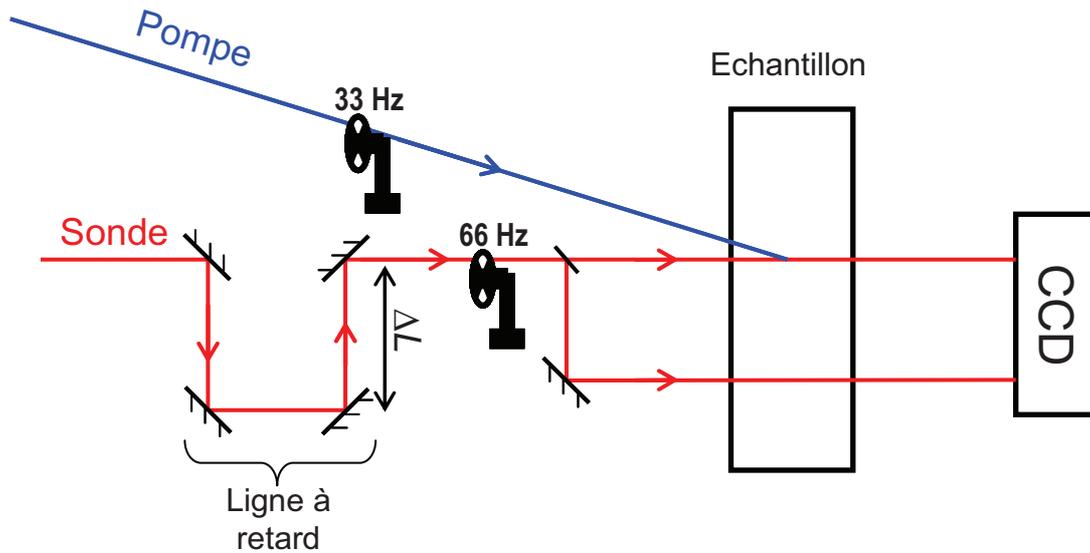


Figure 69 : Schéma de principe de la spectroscopie d'absorption par la méthode pompe-sonde.

L'absorbance ΔA de l'échantillon est alors recalculée pour chaque longueur d'onde λ et chaque temps t selon l'équation (63).

$$\Delta A(\lambda, t) = \log \left(\frac{I_0^{\text{sonde}}(\lambda) - I^{\text{noir}}(\lambda)}{I_0^{\text{ref}}(\lambda) - I^{\text{noir}}(\lambda)} \frac{I_{\text{pompe}}^{\text{ref}}(\lambda) - I^F(\lambda)}{I_{\text{pompe}}^{\text{sonde}}(\lambda, t) - I^F(\lambda)} \right) \quad (63)$$

Où I_0^{sonde} et I_0^{ref} représentent respectivement les intensités de la sonde et de la référence en l'absence de pompe, $I_{\text{pompe}}^{\text{sonde}}$ et $I_{\text{pompe}}^{\text{ref}}$ sont les intensités de la sonde et de la référence en présence d'excitation, I^{noir} est le bruit de fond recueilli en l'absence de la sonde, de la référence et de la pompe. Finalement I^F correspond à l'émission spontanée de l'échantillon induite par la pompe. La variation d'absorbance calculée s'affranchit donc au mieux des différentes sources de lumières parasites et de possibles variations des profils spectraux de la sonde et de la référence sur leurs chemins optiques respectifs.

Les expériences d'absorption transitoire pompe-sonde ont été réalisées au laboratoire par Stéphane Aloïse et Julien Réhault, en utilisant un laser Ti:saphir amplifié (BM Industries) délivrant 0,8 mJ et des impulsions de 80 fs réglable entre 766 et 800 nm, avec une fréquence de répétition de 1kHz. Les impulsions pompe à 383 nm sont obtenues par doublage de fréquence de la fondamentale à 766 nm alors que les impulsions à 267 nm sont obtenues en triplant une fréquence fondamentale de 800 nm. Des énergies supérieures à 20 μJ sont fournies dans les deux cas. Le faisceau de sonde est généré en focalisant 1 μJ de la fondamentale sur un cristal de CaF_2 d'un millimètre d'épaisseur. Le continuum de lumière

blanche obtenu couvre l'UV-visible et le proche infrarouge. Le faisceau de sonde est divisé en un faisceau de mesure et un de référence avant de traverser l'échantillon. Ces deux faisceaux sont enregistrés sur deux caméras CCD. L'absorption transitoire est obtenue en comparant le signal à la référence pour différentes échelles de temps. Le délai entre la pompe et la sonde peut être fixé de 500 fs à 1.5 ns (en utilisant une ligne à retard optique micrométrique). La résolution temporelle de ce système atteint 300 fs. Les faisceaux de pompe et de sonde sont focalisés sur les fenêtres en CaF₂ d'une cellule à circulation de 2 mm. Les différents produits utilisés proviennent de Sigma-Aldrich : benzophénone, 4-méthoxybenzophénone, acétonitrile (ACN), méthanol (MeOH), dichlorométhane (CH₂Cl₂) et cyclohexane.

3.2 Mesures réalisées

Les mesures effectuées sur la benzophénone ont été faites dans l'acétonitrile après excitation à 267 nm. Les données expérimentales concernant la benzophénone sont nommées comme suit. **D**₁ correspond au meilleur rapport signal sur bruit obtenu, dix mille séquences pompe-sonde ont été moyennées dans ce cas. Les dimensions de **D**₁ sont (30 × 1340) où les 30 lignes de **D**₁ sont les spectres mesurés à 1340 longueurs d'ondes dans la gamme 697,1 – 318 nm pendant un intervalle temporel allant de 0,8 à 50 ps. **D**₂, **D**₃ et **D**₄ ont été obtenus en répétant la même expérience avec des concentrations en benzophénone différentes (entre 0,1 et 5 mM), les conditions optiques sont légèrement différentes et le nombre d'accumulations moyennées est plus faible. Ces matrices correspondent à différents niveaux de rapport signal sur bruit allant d'un niveau acceptable **D**₂ à un mauvais rapport **D**₄. **D**₂, **D**₃ et **D**₄ ont pour dimensions (25 × 1340). Les spectres ont été mesurés entre 1 et 50 ps dans la gamme spectrale 285,2 – 664,9 nm. Ces matrices seront utilisées de manière individuelle ou simultanément en utilisant une augmentation de matrices en colonnes. Ces mesures ne sont pas corrigées de la dispersion de vitesse de groupe (GVD), cependant elles sont (à une exception près) toute enregistrée après 1 ps se qui limite l'effet de ce problème sur les données.

Les mesures effectuées sur la méthoxybenzophénone ont été faites dans différents solvants après excitation à 267 nm. Les données correspondantes sont nommées comme suit : **D**_{Me}^{CH} pour les mesures effectuées dans le cyclohexane ; **D**_{Me}^{ACN} pour les mesures effectuées dans l'acétonitrile ; **D**_{Me}^{ACN/H2O} pour les mesures effectuées dans un mélange acétonitrile eau 1:3. La

concentration typique de la méthoxybenzophénone dans les différentes solutions est de 0,5 mM.

4. Analyse exploratoire

Les spectres d'absorption transitoire D_1 de la benzophénone après une photoexcitation à 267 nm sont présentés Figure 70. Il est possible d'observer dans un premier temps que deux zones spectrales évoluent de manières distinctes. La première se situe entre 310 et 380 nm et décroît au cours du temps. La seconde zone, entre 400 et 600 nm, évolue de façon plus complexe. Une bande croît en fonction du temps vers 520 nm alors que cette croissance est biens moins évidente à la bande centrée à 575 nm. Or d'après la littérature l'absorption de l'état S_1 de la BP est maximum à $\lambda_{\max} = 575$ nm. L'état triplet doit avoir lui un maximum d'absorption à une longueur d'onde λ_{\max} inférieure [138, 145], ce qui correspond à la bande centrée à 520 nm. Ces différents changements sont observés entre 0,8 ps et 50 ps.

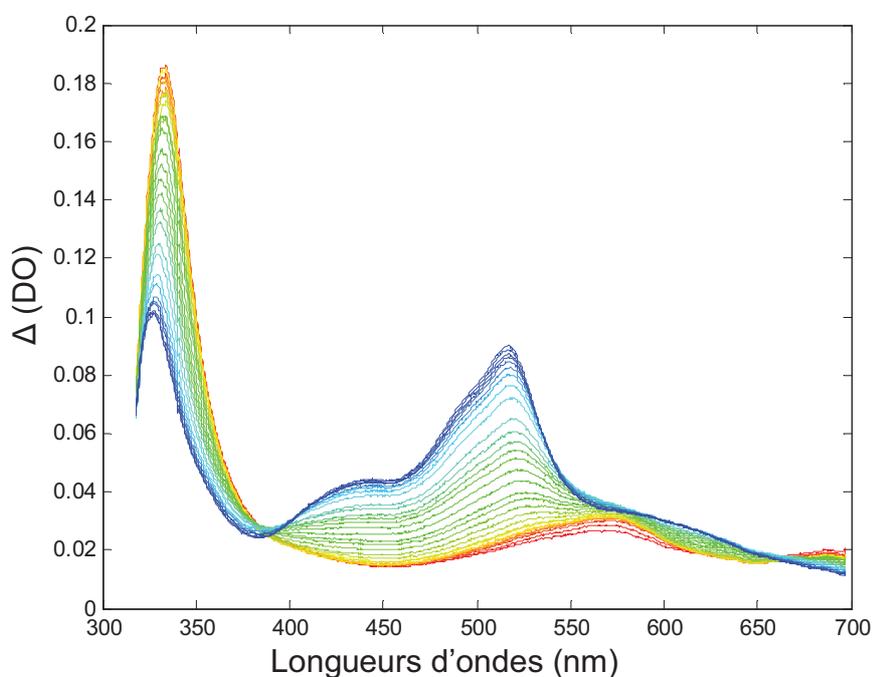


Figure 70 : Spectres d'absorption transitoire de la benzophénone après photoexcitation à 267 nm (de 0,8 ps, en rouge, à 50 ps, en bleu).

4.1 Rang des données concernant la benzophénone

L'estimation du rang par décomposition en valeurs singulières de ce type de données permet de détecter trois facteurs significatifs lorsque l'expérience est menée dans des conditions optimales. Il est en effet intéressant de constater que le rang détectable est en partie dépendant de la concentration de la solution utilisée (en réalité, cela dépend du rapport signal

sur bruit qui décroît en même que quand la concentration le fait). Le Tableau 9 regroupe les premières valeurs singulières des différents jeux de données concernant la benzophénone.

D₁	D₂	D₃	D₄	[D₂; D₃; D₄]
9,3378	4,1906	3,4809	1,3436	5,608
2,4792	1,0896	0,9487	0,3792	1,4962
0,1722	0,1068	0,1073	0,0458	0,1698
0,0701	0,0353	0,0416	0,0204	0,0946
0,0563	0,0226	0,0137	0,0159	0,0616
0,0253	0,0166	0,0132	0,0112	0,0452
...

Tableau 9 : Premières valeurs singulières des données concernant la benzophénone.

Ces décompositions en valeurs singulières permettent d'affirmer que, pour les matrices **D₁**, **D₂** et **D₃**, il est nécessaire de considérer au moins trois facteurs significatifs. Ceci est particulièrement clair dans le cas de **D₁** qui peut être considérée comme la matrice la plus fiable à cause du rapport signal sur bruit plus favorable. Les matrices **D₂** et **D₃** ont aussi un rang égal à trois, mais seulement deux facteurs se détectent d'une façon claire dans **D₄**. Nous attribuons ceci à la faible concentration en benzophénone et au mauvais rapport signal sur bruit de ce jeu de données. Les analyses suivantes permettront de démontrer l'existence d'un troisième facteur commun à toutes les matrices, mais il est déjà possible de calculer la décomposition en valeurs singulières de la matrice augmentée [**D₂**; **D₃**; **D₄**]. La matrice **D₁** est exclue de cette augmentation car elle a été enregistrée à des longueurs d'onde légèrement différentes. Trois facteurs significatifs sont détectés. Cette approche sera mise à profit durant la résolution de données par MCR-ALS.

L'EFA permet également de construire des estimations initiales des profils de concentration [146] Ici les résultats présentés Figure 71 montrent que le rang des quatre matrices est supérieur à 2, dans certains cas tel que **D₃** on pourrait même suspecter la présence d'une quatrième contribution. Les résolutions ultérieures montreront que le rang n'est en réalité que de 3.

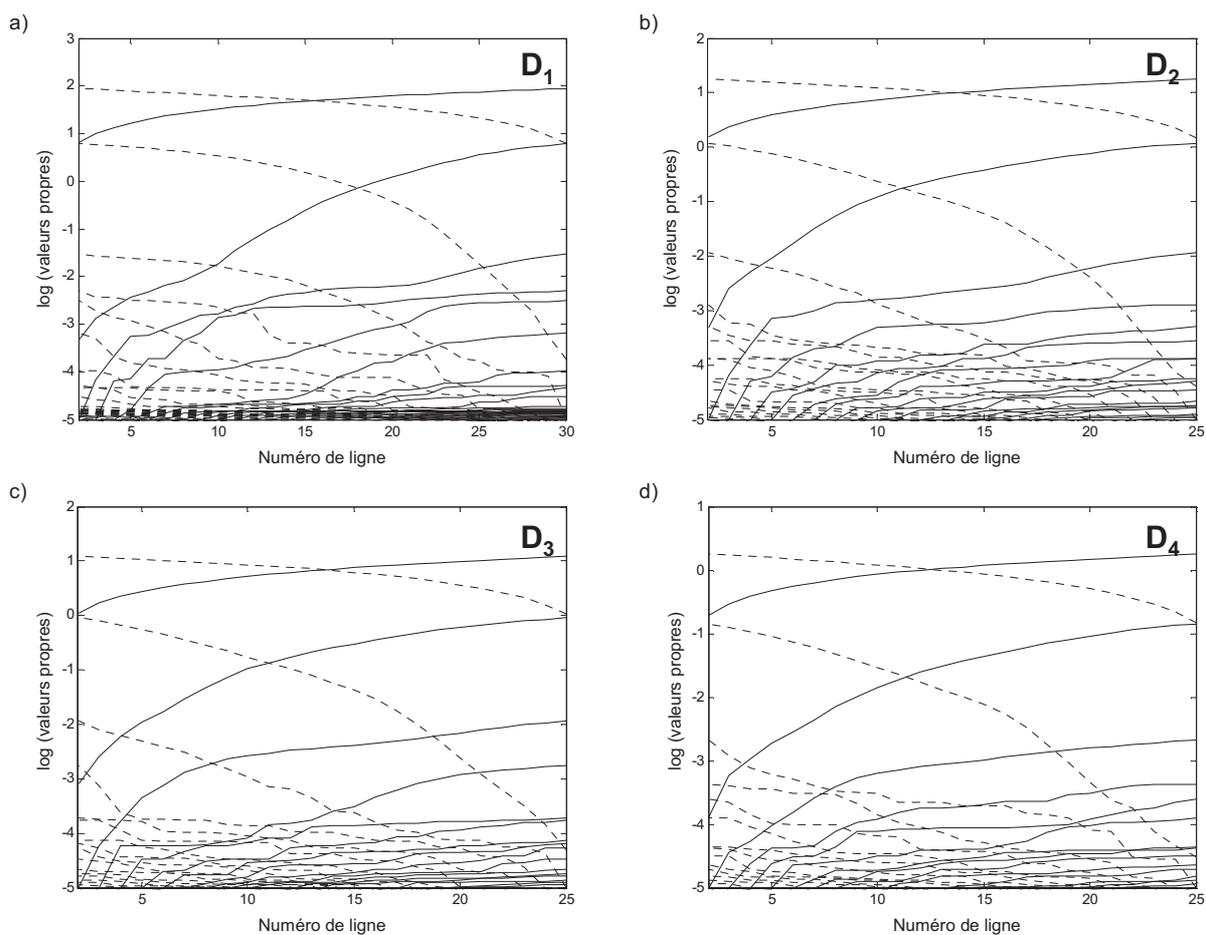


Figure 71 : Résultats de l'EFA en sens direct (trait plein) et inverse (pointillés) pour les matrices a) D_1 , b) D_2 , c) D_3 et d) D_4 .

4.2 Rang des données concernant la méthoxybenzophénone

Un premier examen visuel de l'allure des données présentées Figure 72 indique clairement une influence du solvant sur les formes spectrales enregistrées. Il est possible d'observer, dans le cyclohexane, la croissance de la bande de l'état triplet $T_1(n,\pi^*)$ au environ de 525 nm. Celle-ci coïncide avec la décroissance de la bande de l'état singulet $S_1(n,\pi^*)$ autour de 340 nm. Dans l'acétonitrile un épaulement supplémentaire est visible à 440 nm alors que dans le mélange ACN-eau seule la croissance du pic de $T_1(\pi,\pi^*)$ à 440 nm est observée.

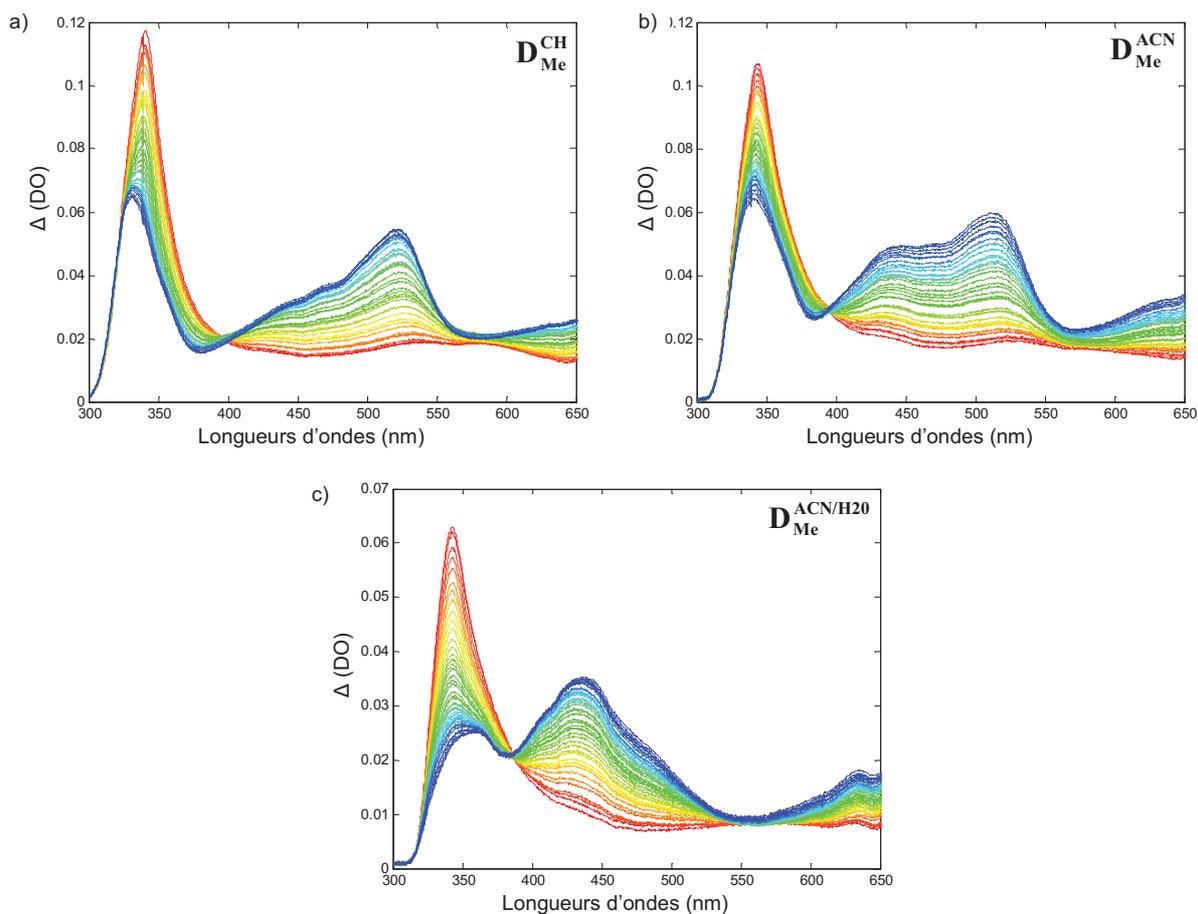


Figure 72 : Spectres d'absorption transitoire de la méthoxybenzophénone après photoexcitation à 267 nm (de 1,2 à 100 ps) dans a) le cyclohexane b) l'acétonitrile c) un mélange acétonitrile eau 1:3.

Les données concernant la méthoxybenzophénone ont été enregistrées dans des conditions de concentration et de rapport signal sur bruit optimales. L'estimation du rang due au bruit observé précédemment sur la matrice \mathbf{D}_4 n'est donc plus envisageable. Le Tableau 10 regroupe les premières valeurs singulières des différents jeux de données concernant la méthoxybenzophénone.

\mathbf{D}_{Me}^{CH}	\mathbf{D}_{Me}^{ACN}	$\mathbf{D}_{Me}^{ACN/H2O}$
6,9627	7,6652	4,0386
1,4442	1,4193	0,7178
0,1085	0,0772	0,0542
0,0214	0,0311	0,0149
0,0177	0,0175	0,0086
0,0119	0,0165	0,0068
...

Tableau 10 : Premières valeurs singulières des données concernant la méthoxybenzophénone.

Les analyses par SVD indiquent un rang de trois pour les matrices D_{Me}^{CH} et D_{Me}^{ACN} correspondant aux solvants apolaires. Les mécanismes de photoexcitation de la MeBP sont donc comparables à ceux de la BP dans ces deux cas. La matrice $D_{Me}^{ACN/H2O}$ est différente, la détection d'un troisième facteur est moins certaine. Les résultats de l'EFA proposés Figure 73 vont dans le même sens : un éventuel troisième facteur détecté dans $D_{Me}^{ACN/H2O}$ serait plus proche du bruit que dans les deux autres matrices. Ceci ne pouvant plus être attribué à un effet de la concentration il est probable que la photochimie de la MeBP dans un mélange d'acétonitrile et d'eau (donc plus polaire) soit différente des autres cas.

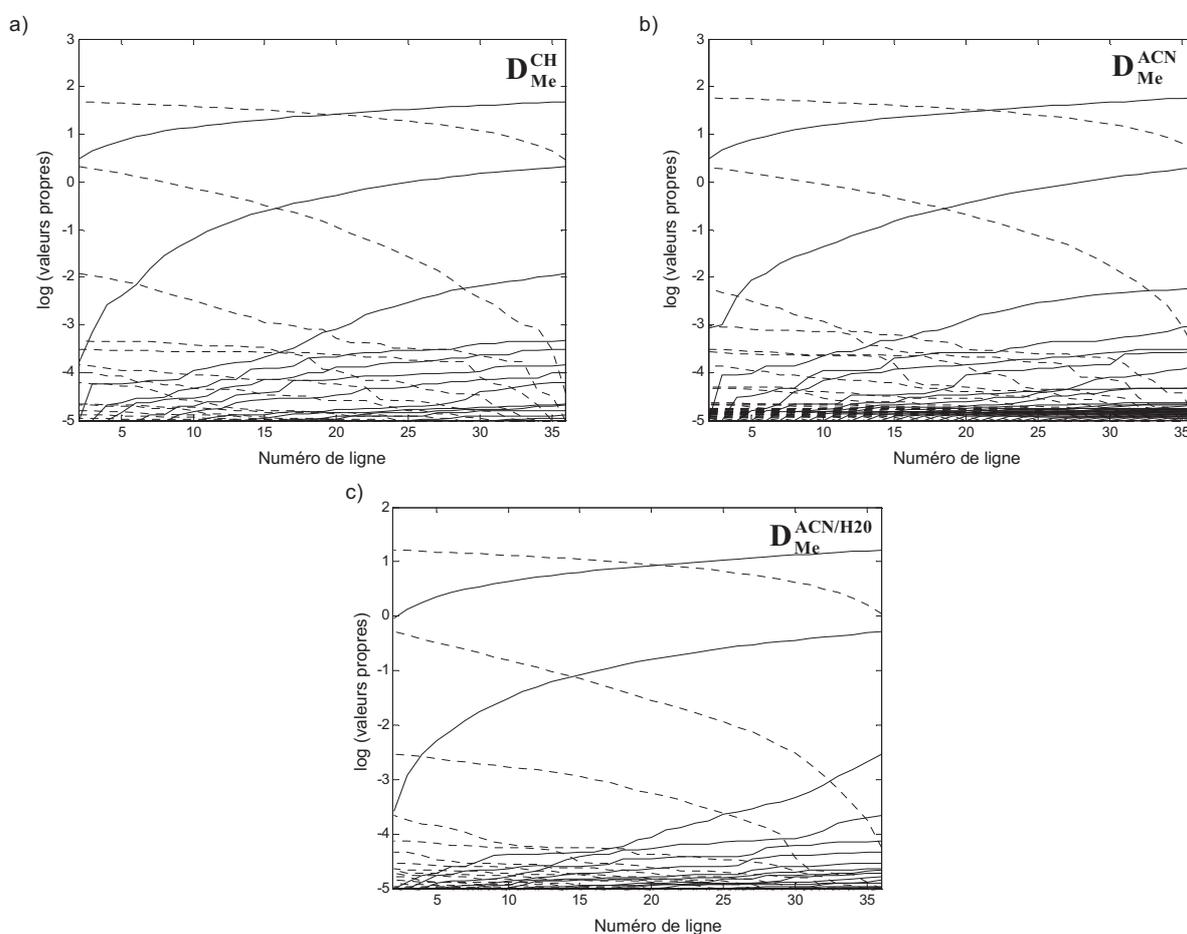


Figure 73 : Résultats de l'EFA en sens direct (trait plein) et inverse (pointillés) pour les matrices a) D_{Me}^{CH} , b) D_{Me}^{ACN} et c) $D_{Me}^{ACN/H2O}$.

5. Résolution MCR-ALS

Les résolutions MCR-ALS proposées ci-dessous se divisent en deux groupes : le premier concernant la benzophénone et le second la méthoxybenzophénone. Nous commencerons par montrer que trois états sont distinguables dans chacune des matrices \mathbf{D}_1 à \mathbf{D}_4 . Nous nous concentrerons ensuite sur l'analyse de \mathbf{D}_1 . La seconde série d'analyses sur la MeBP permettra ensuite de confirmer l'interprétation avancée.

5.1 Analyse de la benzophénone

Les résolutions MCR-ALS des matrices individuelles se basent sur une estimation initiale de l'EFA. Les résultats obtenus pour les matrices \mathbf{D}_1 à \mathbf{D}_3 sont comparables, ceux de \mathbf{D}_1 sont présentés Figure 74. Trois espèces sont détectées. Les profils de concentration correspondent visuellement à ceux d'un mécanisme cinétique. La seule contrainte imposée ici est la non négativité des concentrations et des spectres (aucun *bleaching* n'ayant été observé sur les données brutes). Les valeurs de *Lof* et de R^2 correspondantes montrent que cette résolution ajuste très bien les données puisque celles ci sont respectivement égales à 0,99 % et 99,99 %. La première contribution, représentée en ligne continue, est attribuée à l'état S_1 de la BP. En effet cet état se dépeuple en 10 ps environ et le spectre présente un maximum local d'absorbance à 577 nm. Ces deux observations sont en accord avec les caractéristiques de S_1 fournies par la littérature [138, 145]. La troisième contribution est identifiée grâce à la bande ayant un maximum d'absorption à 517 nm et correspond au triplet T_1 . Le temps caractéristique d'environ 10 ps [141] et le rendement quantique de l'ISC proche de 100 % confirment cette affirmation.

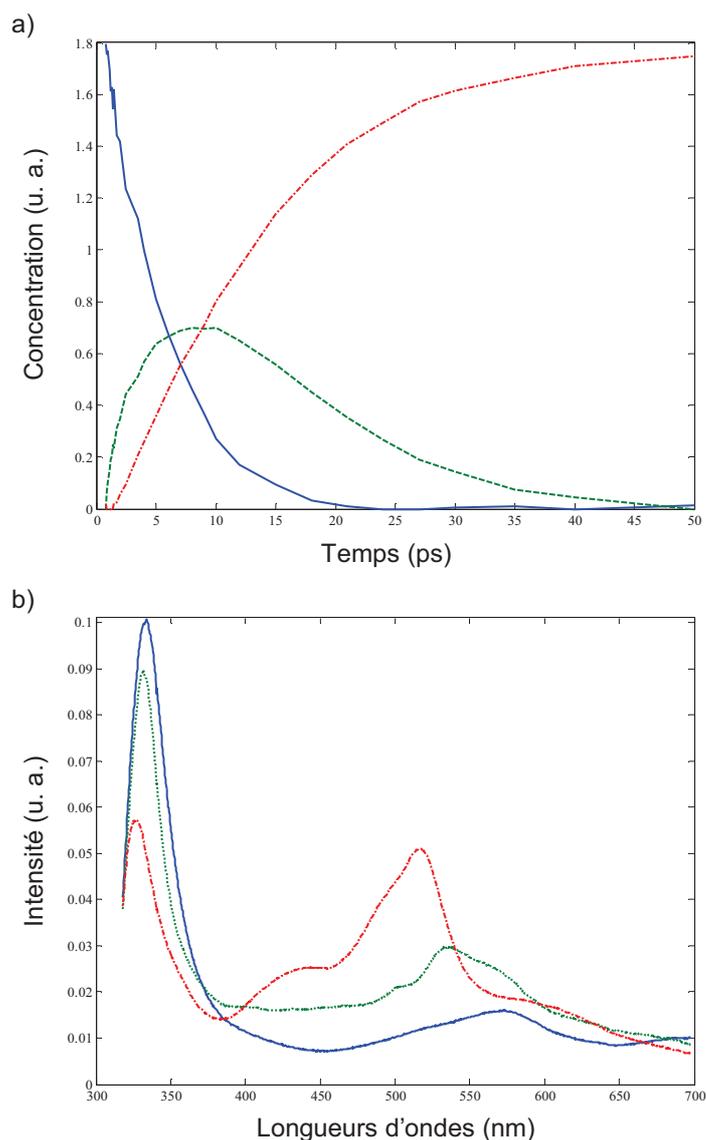


Figure 74 : Résultats de l'analyse MCR-ALS de D_1 : a) profils cinétiques b) spectres purs de l'état S_1 (ligne continue) T_2 (tirets) et T_1 (tirets pointillés).

Le principal intérêt de cette résolution repose sur l'identification de la seconde contribution. Cette contribution atteint un maximum en concentration entre 8 et 10 ps, avant que sa population ne décroisse en faveur de T_1 . Les profils de concentration semblent indiquer qu'il s'agit d'un état intermédiaire permettant de passer d'un état $S_1(n,\pi^*)$ à un état $T_1(n,\pi^*)$. Nous avons vu précédemment que l'hypothèse la plus probable correspondant à ce cas est que cet intermédiaire soit un état triplet $T_2(\pi,\pi^*)$. Le spectre obtenu correspond à un état triplet (bande à 535 nm) mais il peut sembler linéairement dépendant des deux autres contributions. Il est donc nécessaire de valider ce résultat. Les différentes approches envisagées sont détaillées ci-dessous.

5.2 Résolution sur une augmentation de matrices

Deux approches chimiométriques sont utilisées ici pour étayer la résolution présentée précédemment et donc l'hypothèse affirmant l'existence d'un état intermédiaire $T_2(\pi, \pi^*)$ lors de la formation du triplet. La première consiste à analyser conjointement l'ensemble des données disponibles sous la forme d'une augmentation de matrices. La seconde se base sur l'analyse des résidus des résolutions MCR-ALS.

Les trois matrices \mathbf{D}_2 , \mathbf{D}_3 et \mathbf{D}_4 ont été enregistrées de façon à rendre possible l'augmentation de matrices. Le principal intérêt ici est de permettre la détection de l'intermédiaire dans toutes les expériences, y compris \mathbf{D}_4 . En effet nous avons vu précédemment qu'il n'est pas facile de détecter la présence d'une troisième contribution lors de l'analyse individuelle de \mathbf{D}_4 . Un autre avantage découle du caractère semi quantitatif de la résolution qui permettra de comparer les niveaux relatifs de chaque espèce dans les différentes matrices. La Figure 75 présente les profils de concentration issus de la résolution MCR-ALS de la matrice augmentée $[\mathbf{D}_2; \mathbf{D}_3; \mathbf{D}_4]$. Les spectres sont fournis par la Figure 76. Les valeurs de Lof et de R^2 correspondantes à cette résolution sont respectivement égales à 2,91 % et 99,94 %. Ces valeurs étaient attendues puisque le niveau de bruit est plus important dans ces données et l'augmentation permet d'écarter davantage de bruit expérimental dans la matrice des résidus, ce qui a pour effet d'augmenter légèrement le Lof .

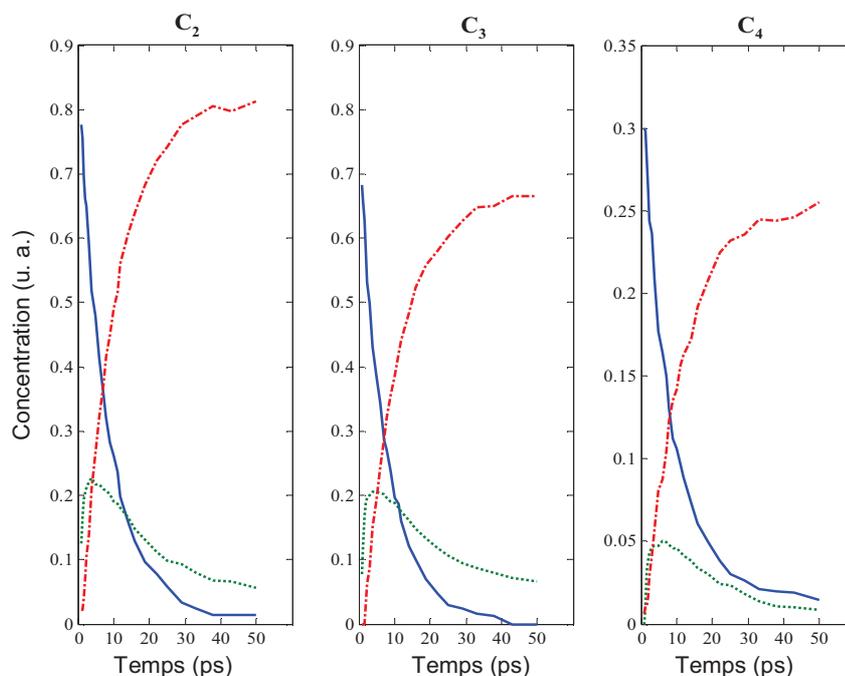


Figure 75 : Profils de concentrations de l'état S_1 (ligne continue bleue) T_2 (pointillés verts) et T_1 (tirets pointillés rouges) pour $[\mathbf{D}_2; \mathbf{D}_3; \mathbf{D}_4]$ obtenus par MCR-ALS.

Les profils cinétiques des trois contributions se retrouvent dans les différents résultats y compris dans ceux de \mathbf{D}_4 . Ceci confirme bien notre hypothèse concernant le rang de cette matrice : le nombre de contributions détectable clairement n'était que de deux à cause de la faible concentration (et donc du faible rapport signal sur bruit) de l'échantillon. De plus, l'échelle de concentrations (relatives) entre les trois expériences permet de voir la basse concentration de l'espèce intermédiaire dans \mathbf{D}_4 , ce qui explique sa détection difficile quand on ne dispose que de cette expérience. L'augmentation de matrices permet ici de retrouver l'intermédiaire à des niveaux plus faibles. Le rang chimique des différentes matrices est donc toujours égal à 3.

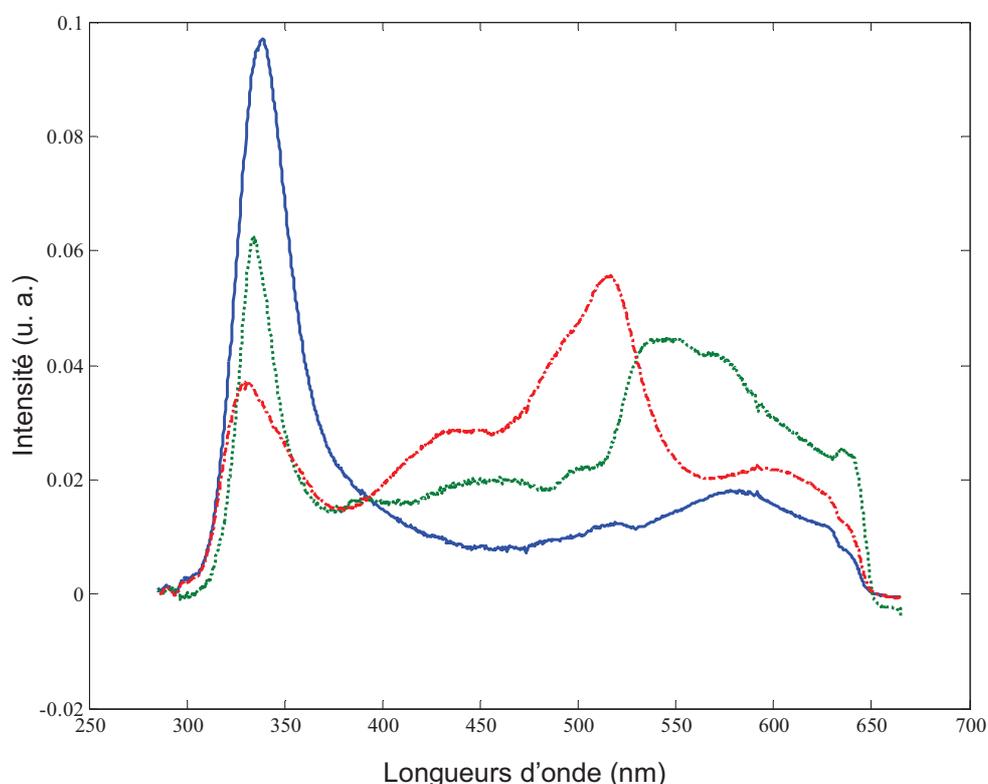


Figure 76 : Spectres purs de l'état S_1 (ligne continue bleue) T_2 (pointillés verts) et T_1 (tirets pointillés rouges) pour $[\mathbf{D}_2; \mathbf{D}_3; \mathbf{D}_4]$ obtenus par MCR-ALS.

La qualité des spectres récupérés est affectée par la diminution du rapport signal sur bruit (par rapport à \mathbf{D}_1). On retrouve cependant les caractéristiques permettant d'identifier les trois contributions comme étant les états S_1 , T_2 et T_1 de la benzophénone. Le spectre de l'état intermédiaire T_2 est le plus difficile à détecter. Il est important de noter que la résolution a été faite en utilisant les contraintes les moins restrictives, c'est-à-dire, seulement la non-négativité dans les deux directions. Cela implique que les solutions obtenues sont plus affectées par

l'ambiguïté de rotation que si d'autres contraintes avaient été appliquées. Mais à ce stade, l'objectif n'est que de déterminer quel est le rang le plus approprié et, par conséquent, quel est le modèle possible pour effectuer par la suite une modélisation cinétique de plus grande qualité.

5.3 Analyse des résidus

La seconde approche proposée pour confirmer le rang et, dans ce contexte, la présence d'une troisième espèce intermédiaire, se base sur la comparaison des résidus des différentes résolutions MCR-ALS utilisant un nombre différent d'espèces chimiques (rang) pour construire le modèle. Il faut rappeler que les résidus se trouvent dans la matrice d'erreur \mathbf{E} du modèle MCR ($\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E}$). Étant donné que \mathbf{E} a la même taille et la même structure que la matrice de données originale \mathbf{D} , un graphique des lignes de la matrice \mathbf{E} indiquera la structure des résidus dans la direction spectrale et un graphique des colonnes de \mathbf{E} la structure de résidus dans la direction temporelle. La présence ou l'absence de structure dans la distribution des résidus et la taille des résidus permettent de déterminer si l'estimation du nombre d'espèces chimiques est correcte. En effet, lorsqu'il est possible d'observer des résidus avec des formes rappelant des variations spectrales où cinétiques, cela signifie probablement qu'une partie informative du signal enregistrée a été écartée hors le modèle et donc, le nombre d'espèces chimiques (rang) sous-estimé. Ce fait nous indique qu'il est nécessaire d'inclure une espèce chimique de plus dans notre modèle. Par contre, si la forme et la taille des résidus ne varient pas significativement après l'incorporation d'une espèce additionnelle, cela signifie que cette espèce n'est pas vraiment nécessaire pour expliquer la variation des données.

Reprenons le cas de \mathbf{D}_1 , où l'information du signal est plus claire. Nous avons estimé par SVD un rang égal à 3. Il ne s'agit que d'une estimation, il est donc nécessaire de tester la résolution pour des estimations de rang voisines, de deux et de quatre, et de comparer les résultats. Les résultats de ces trois résolutions sont représentées Figure 77 (pour les spectres purs et les résidus du modèle dans la direction spectrale), et Figure 78 (pour les profils de concentration cinétiques et les résidus du modèle en la direction temporelle). Si on ne considère que les profils récupérés, les solutions à deux ou trois espèces paraissent valides et on ne pourrait pas rejeter non plus celle à quatre espèces. Cependant l'analyse des résidus offre quelques indices intéressants. On peut en effet remarquer une structure claire dans les résidus d'une résolution basée sur deux contributions. Dans la direction spectrale, cette

structure est plus importante entre 450 à 600 nm. On peut donc envisager la possibilité qu'une contribution chimique supplémentaire soit présente. La résolution à trois contributions confirme cette supposition notamment grâce à l'espèce qui apparaît, avec des bandes principalement placées sur la zone structurée des résidus observés auparavant. Les résidus du modèle à trois espèces sont moins importants et moins structurés: il ne reste plus qu'un massif aux basses longueurs d'onde. Le fait d'ajouter une quatrième contribution n'a plus d'influence aussi significative sur les résidus, même si les résidus sont de nouveau un peu plus réduits.

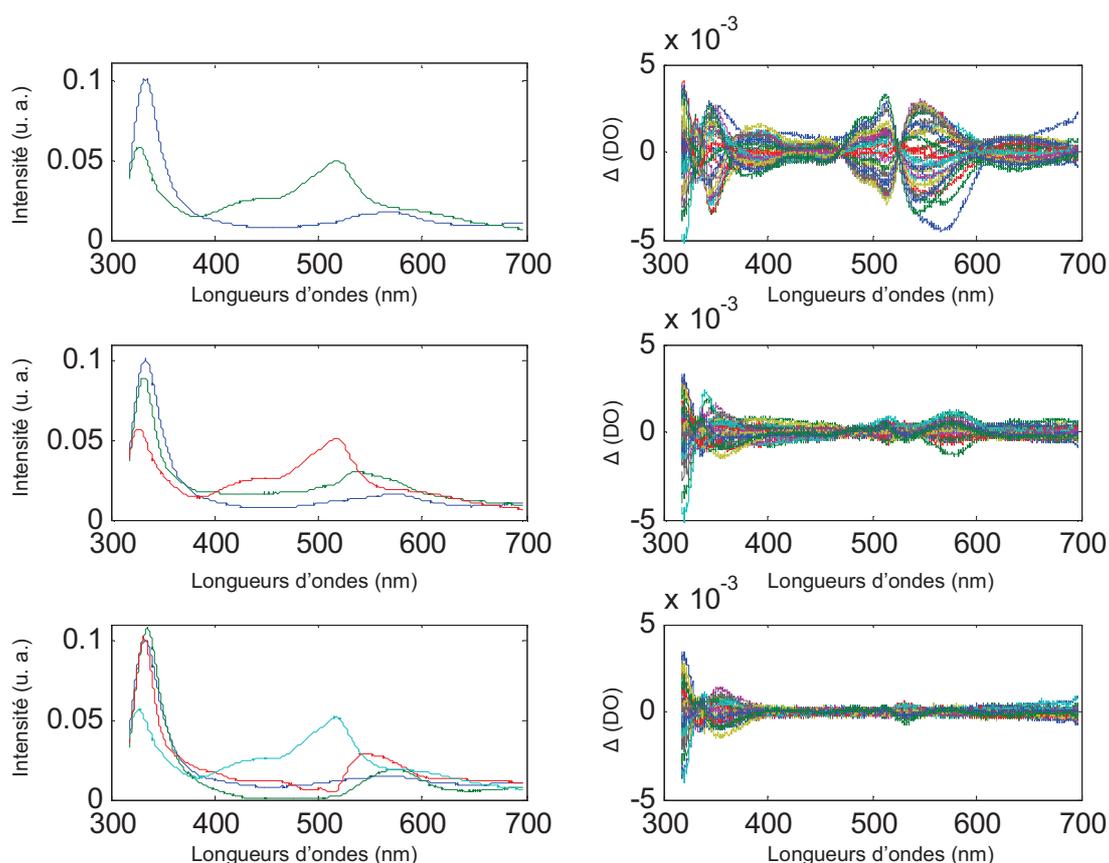


Figure 77 : Spectres purs (gauche) et résidus en la direction spectrale (droite) pour des résolutions MCR-ALS un rang allant de 2 à 4.

La même méthodologie peut être appliquée dans la direction temporelle. La structure des résidus permet d'avancer certaines hypothèses. On peut en effet constater que les résidus de la résolution à deux contributions sont structurés principalement dans deux zones: entre 4 et 17 ps et entre 28 et 50 ps. Il manque donc au moins une contribution. Les résultats obtenus avec trois contributions le confirment et l'inclusion d'une quatrième n'améliore pas les résultats significativement. Le rang de \mathbf{D}_1 est donc bien égal à trois.

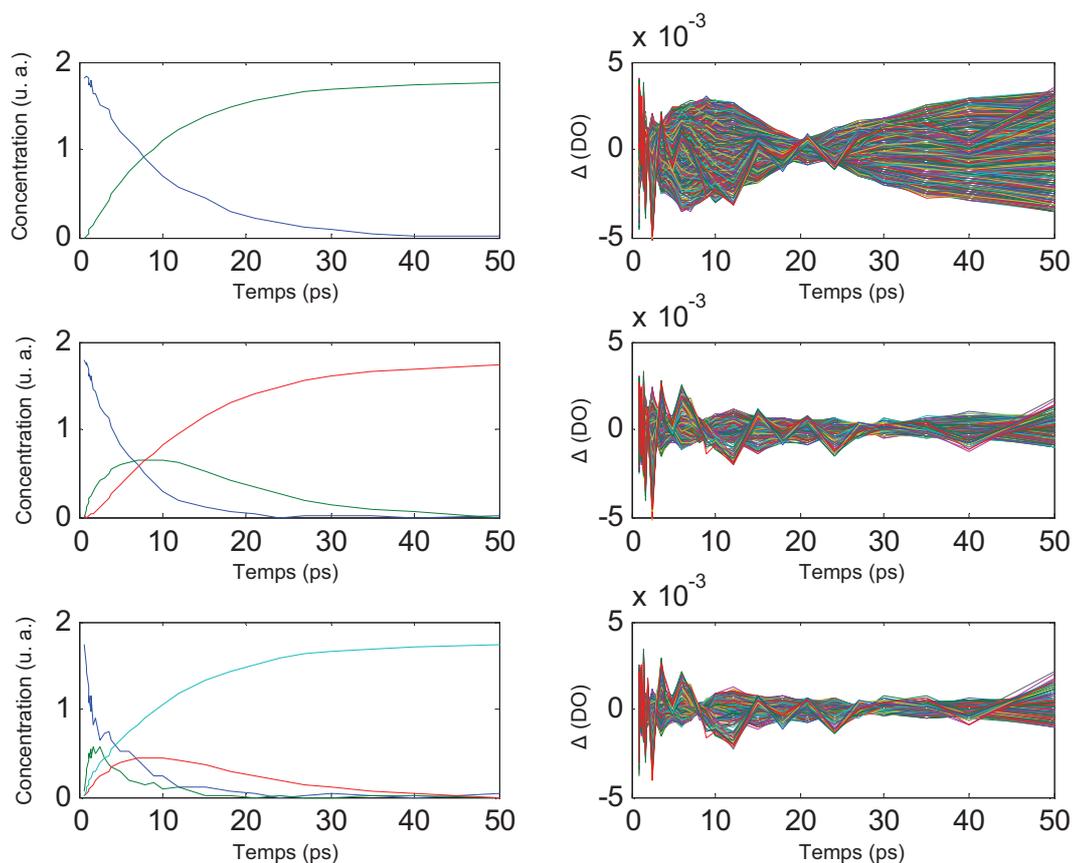


Figure 78 : Résultats cinétiques des résolutions MCR-ALS pour un rang allant de 2 à 4 et résidus correspondants

Les différents arguments présentés dans cette section constituent un faisceau de preuves attestant l'existence de trois contributions significatives. Le rang est confirmé de façon certaine, ce qui exclut la possibilité d'une transition directe $S_1 \rightarrow T_1$. Les règles d'El Sayed ne sont donc pas prises en défaut dans cet exemple. L'hypothèse formulée précédemment et résumée par la Figure 66 est en accord avec ces résultats. Les trois contributions détectées sont donc associées aux états S_1 , T_2 et T_1 de la BP.

5.4 Analyse de la méthoxybenzophénone

Dans les sections 5.2 et 5.3, nous avons validé les résultats obtenus pour la modélisation de la photochimie de la benzophénone d'un point de vue chimiométrique. Il est aussi important d'apporter, si possible, un autre type d'information qui permette une validation basée sur des faits chimiques, c'est pourquoi le système de la méthoxybenzophénone a été étudié [147]. Il a en effet été démontré que les niveaux

énergétiques des différents états triplets de la méthoxybenzophénone sont sensibles au solvant utilisé [144]. Cette propriété permet de tester différents schémas en accord avec les règles d'El Sayed. Le niveau énergétique de l'état triplet (π,π^*) de la MeBP diminue lorsque la polarité augmente. Le niveau T_2 se rapproche donc de l'état triplet $T_1(n,\pi^*)$ et finit par le rejoindre puis passe sous ce niveau devenant alors le nouveau niveau T_1 de la MeBP. C'est notamment le cas dans un mélange d'acétonitrile et d'eau (1:3). Les spectres d'absorption transitoire de la MeBP dans les trois solvants différents présentés par la Figure 72 ont été analysés par MCR-ALS. Les résolutions pour les expériences réalisées dans chacun des solvants sont proposées Figure 79.

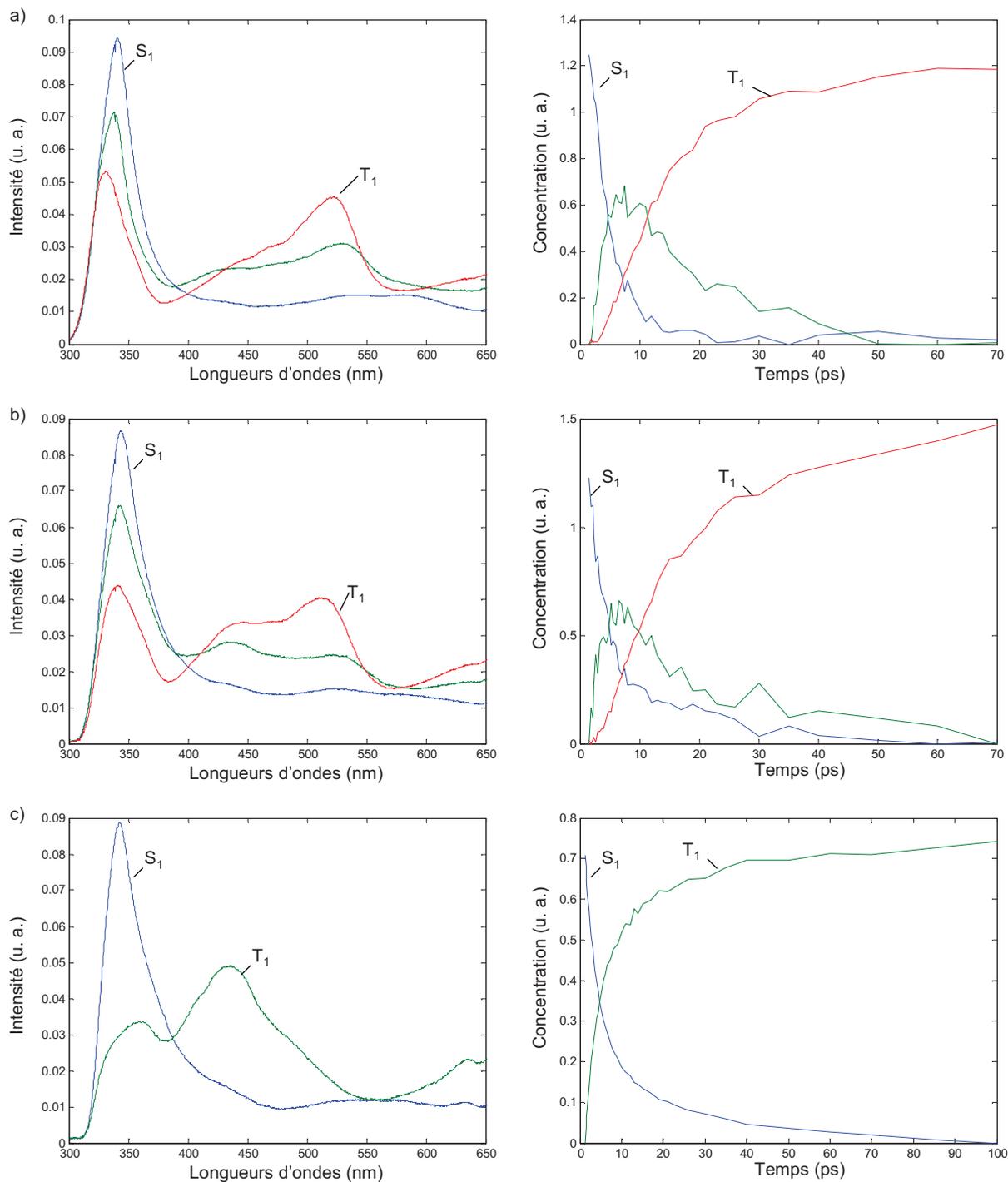


Figure 79 : Résolutions MCR-ALS des spectres d'absorption de la méthoxybenzophénone dans a) le cyclohexane b) l'acétonitrile c) un mélange acétonitrile eau (3:1).

Les deux premières résolutions concernent la MeBP dans le cyclohexane et l'acétonitrile. Dans ces deux cas le niveau d'énergie de l'état triplet (π,π^*) est supérieur à celui de l'état triplet (n,π^*). La transition de S_1 à T_1 passe donc par un intermédiaire T_2 comme pour la benzophénone. Les profils cinétiques obtenus valident ce schéma. Les spectres obtenus pour la MeBP dans le cyclohexane peuvent être associés aux états S_1 , T_2 et T_1 grâce à leurs

maxima d'absorbance respectifs à 589 nm, 527 nm et 522 nm. Par contre, la différence est notable si on considère la résolution du même système dans un mélange d'acétonitrile et d'eau (1:3). La présence d'un intermédiaire n'est pas détectée ni modélisée, ce qui confirme le mécanisme $S_1 \rightarrow T_1$ puisque l'inversion énergétique des états triplets permet maintenant un croisement intersystème direct qui ne transgresse pas les règles d'El Sayed.

Il existe donc une corrélation claire entre la présence d'un intermédiaire et le niveau d'énergie du triplet T_2 de la MeBP. De plus la benzophénone dans l'acétonitrile et la méthoxybenzophénone dans le cyclohexane ont la même distribution énergétique des états $S_1(n,\pi^*)$, $T_2(\pi,\pi^*)$ et $T_1(n,\pi^*)$. Dans ces conditions les spectres de l'espèce intermédiaire de ces deux composants sont comparables. L'hypothèse, proposant que l'intermédiaire observée durant la relaxation de la BP soit un état triplet $T_2(\pi,\pi^*)$, est étayée par cette étude de la photochimie de la MeBP.

L'analyse du processus de photorelaxation de la BP peut donc se poursuivre en se basant sur cette hypothèse. Nous verrons dans la partie suivante comment prendre en compte ce schéma lors de la résolution.

6. Resolution HS-MCR

Les méthodes de résolution de courbes proposées jusqu'à présent ne supposent pas d'implémenter des connaissances chimiques du système. L'analyse des différentes données disponibles, concernant aussi bien la benzophénone que la méthoxybenzophénone, permet d'utiliser le diagramme énergétique proposé Figure 66 comme point de départ pour une analyse hybride incluant un modèle cinétique. Selon les résultats obtenus aux sections précédentes, le schéma réactionnel considéré est :



Il faut noter que l'application de HS-MCR dans cette étude est préférée à une méthode de modélisation rigide classique car elle fournit la possibilité d'appliquer la contrainte de non-négativité sur les spectres, ce qui peut potentiellement éviter l'ambiguïté liée aux modèles cinétiques de réactions consécutives de premier ordre, nommée *slow-fast* ou *flip-flop*, qui permet d'échanger la séquence des constantes cinétiques sans induire une perte d'ajustement des données [99]. Les profils de concentration et les spectres correspondant à l'application de ce modèle à **D**₁ sont présentés Figure 80.

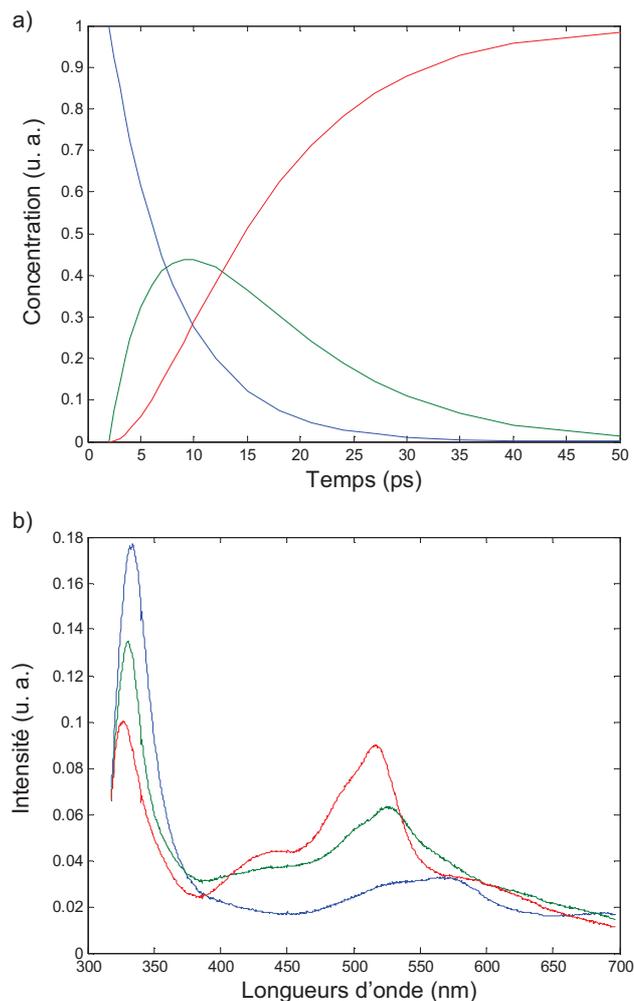


Figure 80 : Résultats de l'analyse HS-MCR de \mathbf{D}_1 : a) profils cinétiques b) spectres purs de l'état S_1 , T_2 et T_1 .

Les résultats obtenus sont cohérents avec ceux de l'analyse MCR-ALS, ce qui montre l'adéquation du modèle choisi. Les trois contributions peuvent être reliées aux états S_1 , T_2 et T_1 de la même façon que précédemment grâce aux pics à 577, 528 et 517 nm. Le *Lof* obtenu est égal à 0,59% et les constantes de réactions correspondants à ce modèle sont $k_1 = 0,154 \pm 0,002 \text{ ps}^{-1}$ et $k_2 = 0,093 \pm 0,001 \text{ ps}^{-1}$.

Pour confirmer cette thèse, nous avons utilisé aussi HS-MCR avec le même modèle global pour analyser la matrice augmentée formée par les trois expériences [\mathbf{D}_2 ; \mathbf{D}_3 ; \mathbf{D}_4] [103]. Les profils récupérés (Figure 81) sont bien meilleurs que ceux des Figures 13 et 14 à cause de l'application d'une contrainte plus rigide et confirment que, même dans des situations moins favorables du rapport signal sur bruit, cette troisième espèce peut être correctement modélisé. Le *Lof* obtenu est égal à 4,2% et les constantes de réaction sont $k_1 = 0,147 \pm 0,006 \text{ ps}^{-1}$ et $k_2 = 0,103 \pm 0,005 \text{ ps}^{-1}$.

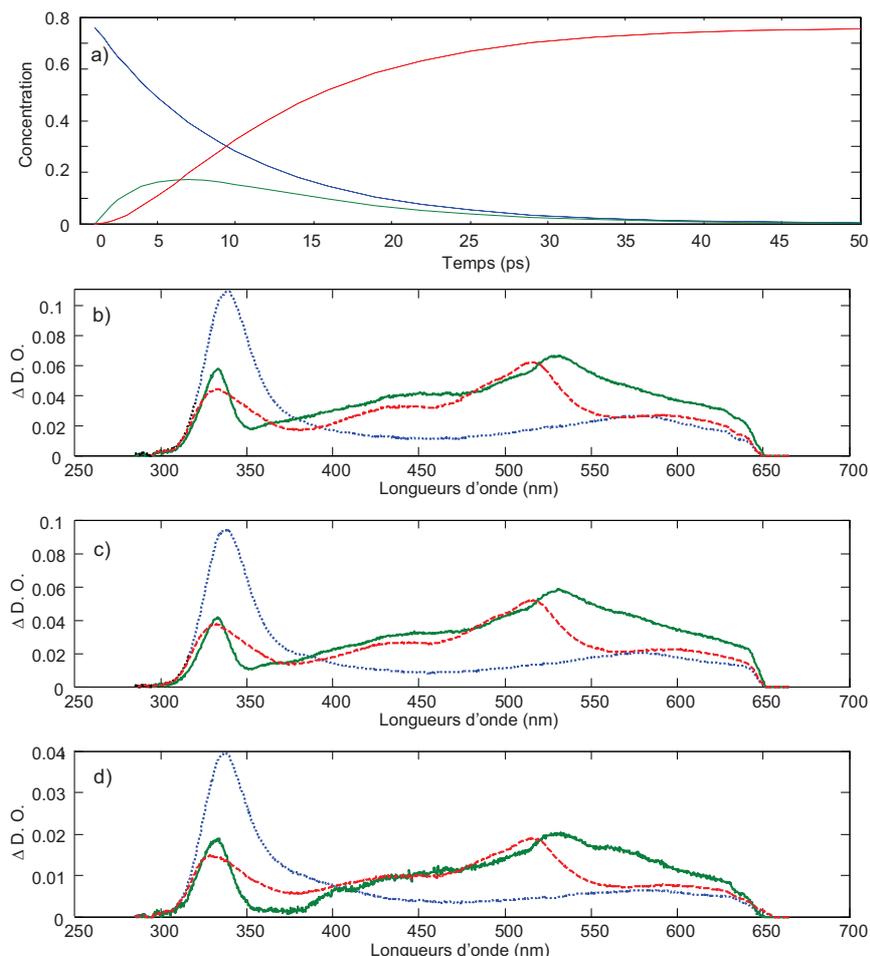


Figure 81 : a) Profils de concentration et spectres obtenus b) S_2 c) S_3 et d) S_4 ; lors de la résolution HS-MCR de $[D_2; D_3; D_4]$ en utilisant un modèle global.

Dans le but d'apporter plus d'information sur la photochimie de la benzophénone, nous avons utilisé cette approche pour calculer les constantes de réaction du même modèle pour des données correspondant au suivi de la relaxation de la benzophénone dans différents solvants, après différentes photoexcitations. Ces valeurs sont regroupées dans le Tableau 11 en fonction de la longueur d'onde d'excitation utilisée (λ_{exc}) et du solvant. Un résultat significatif est obtenu dans la découverte d'un effet de la longueur d'onde d'excitation : les écarts maxima entre les deux bandes ($\Delta\lambda_{max}$) sont deux fois plus grands pour $T_2 \rightarrow T_1$ quand on utilise une longueur d'onde plus courte. Cette différence est due au fait que seul l'état vibrationnel le plus bas de S_1 est peuplé suite à l'excitation à 383 nm alors qu'à 267 nm plusieurs états vibrationnels sont peuplés. Ces états sont alors en plus grande interaction avec les états triplets. Par contre, on n'observe pas d'effet clair sur la photochimie de la benzophénone par rapport aux différents solvants utilisés.

Solvant	λ_{exc} (nm)	$\Delta\lambda_{max}$ (nm)		$S_1 \rightarrow T_2$ k_1 (fs ⁻¹)	$T_2 \rightarrow T_1$ k_2 (fs ⁻¹)	LOF (%)
		$S_1 \rightarrow T_2$	$T_2 \rightarrow T_1$			
ACN	267	41	13	154 ± 2	93 ± 1	0,59
ACN	383	40	5	133 ± 2	99 ± 11	0,92
MeOH	383	42	3	167 ± 4	117 ± 3	1,21
CH ₂ Cl ₂	383	41	4	146 ± 3	103 ± 3	0,7

Tableau 11 : Propriétés spectrocinétiques obtenues lors de l'analyse HS-MCR de spectres d'absorption transitoire de la benzophénone après différentes excitations.

Malgré le bon accord entre les résultats chimométriques et les comparaisons avec le système connu de la méthoxybenzophénone, il ne faut pas écarter la possibilité que le véritable chemin réactionnel soit plus complexe que le modèle utilisé ici. Des réactions parallèles sont en effet probables et des états vibrationnels différents peuvent être inclus sous une même contribution apparente. Il faut noter que les résultats obtenus avec les méthodes chimométriques utilisées reflètent la variation de la mesure enregistrée. Par conséquent, si des états réactionnels ont des spectres extrêmement similaires ou s'ils évoluent simultanément en fonction du temps, ils peuvent être modélisés comme une seule contribution. Cela dit, il est bon de noter que les constantes de réactions observées sont cohérentes avec celles attendues c'est-à-dire avec les échelles de temps connues des croisements intersystèmes. Il est important de remarquer que le résultat proposé ici est bien meilleur en terme d'ajustement qu'une résolution n'impliquant pas d'intermédiaire ($S_1(n,\pi^*) \rightarrow T_1(n,\pi^*)$) non présentée ici) et que le traitement chimométrique multivariée est beaucoup plus robuste que des ajustements classiques qui ne travaillent qu'avec l'interprétation des traces cinétiques enregistrés à une seule longueur d'onde.

7. Conclusion et perspectives sur cette application

L'analyse de données d'absorption transitoire rencontre différents obstacles spécifiques à ce type de méthode. Certains sont à relier à la complexité de l'instrumentation utilisée, d'autres à la nature même du signal. Les méthodes d'analyses chimiométriques peuvent apporter des réponses innovantes à ces problèmes complexes. Ce chapitre a permis de démontrer la faisabilité d'une résolution multivariée de courbes sur des données de ce type, où le recouvrement spectral est très important. L'existence d'un état intermédiaire lors de la transition $S_1(n,\pi^*) \rightarrow T_1(n,\pi^*)$ de la benzophénone a pu ainsi être démontrée là où des moyens physico-chimiques classiques ne peuvent l'isoler. Un faisceau de preuves tend à démontrer que cet intermédiaire est un triplet $T_2(\pi,\pi^*)$.

D'un point de vue purement chimiométrique, les principales difficultés de ce travail concernent l'évaluation correcte du rang et le degré élevé de recouvrement spectral (et de colinéarité) des données. L'utilisation des différentes méthodes de résolution (EFA et MCR-ALS) nous a permis de déterminer le rang. Les résolutions obtenues à ce stade peuvent encore être critiquées car les spectres purs des différents états sont très proches les uns des autres. Cependant l'application de contraintes cinétiques permet d'apporter un surcroît d'information. En effet, en plus d'améliorer la robustesse de la résolution en limitant les ambiguïtés rotationnelles cette approche nous permet d'obtenir les constantes de réaction des différents phénomènes observés. On a pu voir que les résultats obtenus sont conformes aux temps de vie généralement observés en photochimie. De plus cette stratégie nous a permis d'identifier le modèle physico-chimique le plus probable.

Chapitre V : Caractérisation de structures de bruit en spectroscopie résolue en temps et intégration de cette information dans l'algorithme de résolution

Les méthodes analytiques et donc la spectroscopie sont affectées par différents bruits. Le terme bruit représente un ensemble de phénomènes physicochimiques parasites et menant à des erreurs lors de la mesure. Ces bruits sont généralement réduits par une optimisation instrumentale ou par une accumulation de données. Certaines techniques instrumentales modernes ayant atteint leurs limites technologiques n'ont d'autres alternatives que d'accumuler sur de longues périodes des répétitions de mesure pour réduire l'influence du bruit. L'importance de l'erreur de mesure dans ce type de données est parfois encore mal connue. De plus elle est suspectée de générer différents problèmes lors de la résolution multivariée de courbes, notamment de surajustement. Ce dernier terme correspond aux situations dans lesquelles la résolution prend en compte une partie du bruit expérimental dans les résultats. Nous appliquerons d'abord une méthodologie permettant de caractériser les différents bruits affectant la mesure. Nous proposerons ensuite une nouvelle adaptation d'un outil chimiométrique permettant d'intégrer ce type d'informations dans la résolution.

Les données utilisées ici se rapportent à la photorelaxation de la benzophénone, comme au chapitre précédent. Cependant il s'agit ici de réplicats de la même expérience obtenus dans des conditions beaucoup moins favorables d'accumulations. L'objectif est d'abaisser le rapport signal sur bruit puisqu'à présent on s'intéresse à la connaissance des bruits et de leurs effets sur l'analyse de données.

Les travaux qui suivent visent à caractériser de manière statistique puis physicochimique les erreurs affectant un ensemble de données spectroscopiques. Ces erreurs peuvent soit être associées à des bruits soit à des défauts de mesures. Les connaissances acquises pourront ensuite être utilisées pour améliorer les résolutions de courbes multivariées en y incluant une estimation de l'erreur de mesure dans un schéma de pondération.

1. Estimation et analyse de l'erreur

L'estimation de la structure complète de l'erreur est possible en utilisant la méthodologie proposée par l'équipe de P. Wentzell [148]. Celle-ci se base sur l'analyse de matrices de covariance de l'erreur. Il est nécessaire d'enregistrer une série de r matrices \mathbf{D}_k ($k = 1, \dots, r$) de dimension m (temps) \times n (longueurs d'ondes), chacune correspondant à un réplicat de la même expérience. La mesure de l'erreur de chaque réplicat, \mathbf{E}_k , est estimée comme suit :

$$\mathbf{E}_k = \mathbf{D}_k - \bar{\mathbf{D}} \quad (65)$$

où $\bar{\mathbf{D}}$ est la matrice qui contient la moyenne de toutes les mesures effectués. L'ensemble de ces matrices d'erreur forme le cube $\underline{\mathbf{E}}$ de dimension $m \times n \times r$. L'erreur de mesure d'un spectre ($1 \times n$), enregistré à un temps i , peut être étudiée à partir de la matrice de covariance de l'erreur ($n \times n$), Σ_i , calculée selon l'équation (66).

$$\Sigma_i = \frac{1}{(r-1)} \sum_{k=1}^r (\mathbf{d}_k - \bar{\mathbf{d}})^T (\mathbf{d}_k - \bar{\mathbf{d}}) \quad (66)$$

où \mathbf{d}_{ik} représente le spectre au temps i du réplicat k et $\bar{\mathbf{d}}$ est le spectre moyen de toutes les réplicats enregistrées au temps i dans les différentes expériences. La matrice Σ_i contient, le long de sa diagonale, la variance de l'erreur de mesure à chaque longueur d'onde. Les éléments hors diagonale donnent la covariance des erreurs entre les différents canaux de mesure. L'expérience étant enregistrée pour m temps on obtient au final un cube $\underline{\Sigma}$ de dimension $m \times n \times n$, qui représente la covariance de l'erreur dans le premier mode des données, c'est-à-dire la dimension spectrale (voir Figure 82). De manière analogue, pour la direction temporelle, l'erreur de mesure d'une trace cinétique ($m \times 1$), colonne enregistré à la longueur d'onde j , peut être étudiée avec la matrice de covariance de l'erreur, Ψ_j ($m \times m$), calculée selon:

$$\Psi_j = \frac{1}{(r-1)} \sum_{k=1}^r (\mathbf{d}_{jk} - \bar{\mathbf{d}})^T (\mathbf{d}_{jk} - \bar{\mathbf{d}}) \quad (67)$$

où maintenant $\bar{\mathbf{d}}$ est la trace cinétique moyenne de toutes les réplicats enregistrées à la longueur d'onde j dans les différentes expériences. Ensuite, un second cube de covariance $\underline{\Psi}$ ($n \times m \times m$) peut être obtenu avec les matrices Ψ_j provenant de chacune des n colonnes de la matrice \mathbf{D} . Ce cube sera relié au second mode de la matrice : la direction temporelle. L'étude

des cubes $\underline{\Sigma}$ et $\underline{\Psi}$ va fournir des informations sur la présence ou l'absence de bruit structuré dans le mode spectrale et temporelle, respectivement.

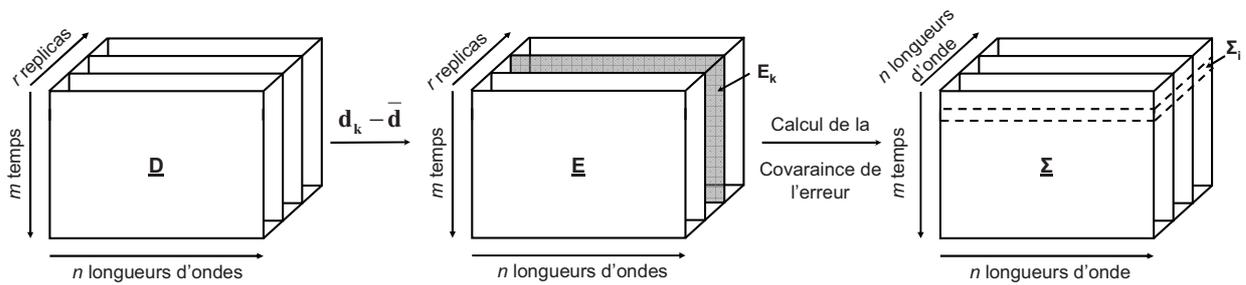


Figure 82 : Calcul de la matrice de covariance de l'erreur $\underline{\Sigma}$ dans la direction spectrale.

En présence de bruit structuré, les sources de l'erreur de mesure peuvent être analysées en suivant la méthodologie déjà mentionnée [148]. Les cubes $\underline{\Sigma}$ et $\underline{\Psi}$ peuvent être analysés en utilisant PARAFAC [9] qui propose une décomposition suivant un modèle trilineaire, comme on l'a déjà vu au chapitre 1. Les *loadings* obtenus donnent un premier aperçu qualitatif des contributions structurées de l'erreur.

Cette approche a été appliquée à l'analyse de l'erreur affectant les mesures effectuées sur l'expérience de la benzophénone. Cette expérience décrite au chapitre précédent a été reproduite trente fois. Chaque répétition correspond à 50 accumulations par spectre là où précédemment on pouvait atteindre 10000 cycles pompe-sonde. Ces 50 accumulations sont moyennés pour ne produire qu'un spectre. Les mesures s'effectuent à 30 temps différents et à 1340 longueurs d'onde. Les données forment donc finalement un cube \underline{D} de dimension $30 \times 30 \times 1340$ (pour 30 spectres, 30 réplicats et 1340 longueurs d'onde).

Les cubes de variance-covariance $\underline{\Sigma}$ et $\underline{\Psi}$ ont ensuite été calculés. Commençons par analyser l'erreur de mesure dans la direction temporelle, donc le cube $\underline{\Psi}$. Deux des sous matrices Ψ_i de $\underline{\Psi}$ sont proposées à la Figure 83. Elles correspondent à la matrice de variance-covariance de l'erreur pour les longueurs d'onde $\lambda = 586,0$ nm et $\lambda = 387,5$ nm. Deux observations peuvent être faites à propos de ce graphique. La première est que seul les valeurs de la diagonale se détachent nettement (en termes d'intensité). Ces valeurs prouvent la présence de bruit de mesure. Les variations de l'intensité des valeurs le long de la diagonale suggèrent que le bruit est de nature hétéroscédastique. Les différentes tranches de $\underline{\Psi}$ ne

montrent pas de structure en dehors de la diagonale, ce qui suggère que l'erreur est indépendante dans la direction temporelle. L'analyse complète de l'erreur est alors grandement facilitée car si une structure est présente elle ne peut plus l'être que dans la direction spectrale et donc observable sur $\underline{\Sigma}$.

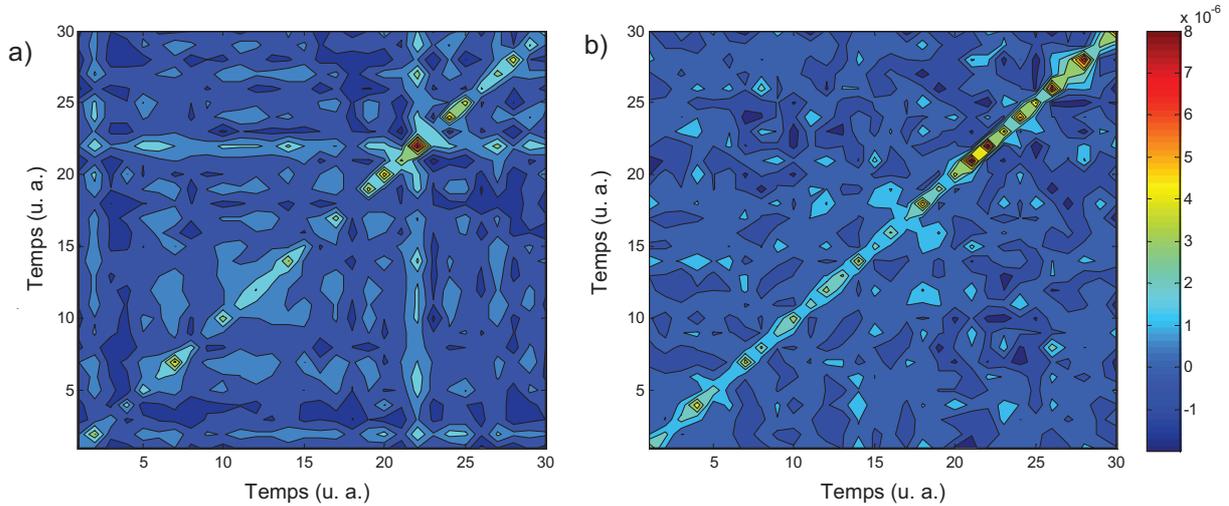


Figure 83 : Tranches représentatives du cube $\underline{\Psi}$ à 586,0 et 387,5 nm. L'échelle de couleur correspond à la valeur de la covariance.

Des conclusions similaires peuvent être tirées de l'étude des matrices de corrélation correspondantes, présentées Figure 84. En effet dans les deux cas les diagonales se détachent nettement d'un fond ou il n'est pas possible d'observer de structures de corrélation claires. La précision de ces estimations des corrélations est de $\pm 0,18$, c'est-à-dire l'inverse de la racine carrée du nombre de répliquats (ici 30).

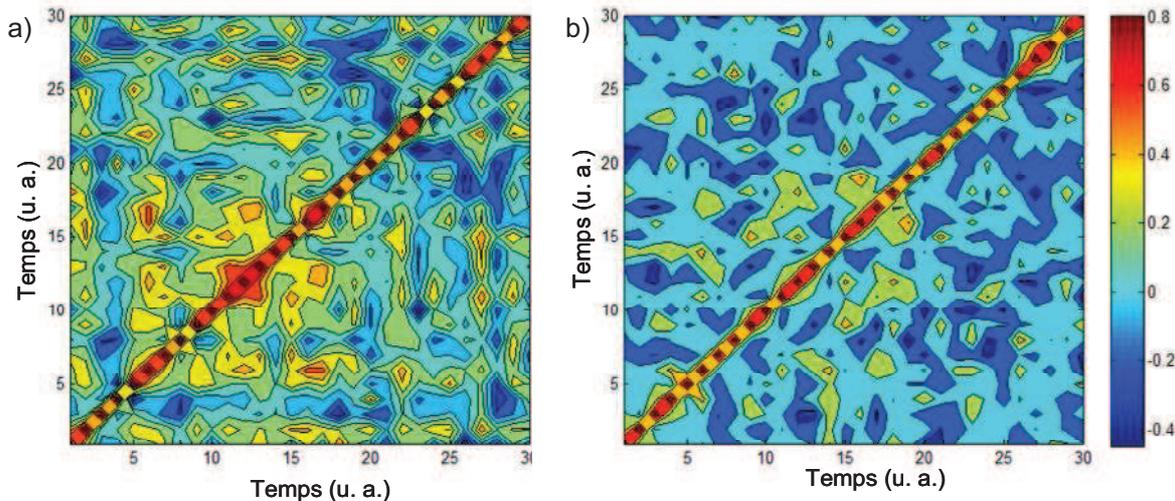


Figure 84 : Matrices de corrélation de l'erreur de mesure sur les temps à 586,0 et 387,5 nm. L'échelle de couleur correspond à la valeur de la corrélation.

La Figure 85a et la Figure 85b présentent deux tranches de la matrice de covariance de l'erreur $\underline{\Sigma}$ correspondant aux temps 3,2 et 50 ps. Deux points doivent être notés ici. Premièrement les structures visibles en dehors de la diagonale révèlent que l'erreur de mesure n'est pas indépendante. Ensuite il est possible d'observer que cette structure change entre les deux temps présentés. Deux zones spectrales montrent une corrélation de l'erreur, une est reliée aux longueurs d'onde inférieures à 400 nm et l'autre à celles supérieures à 400 nm.

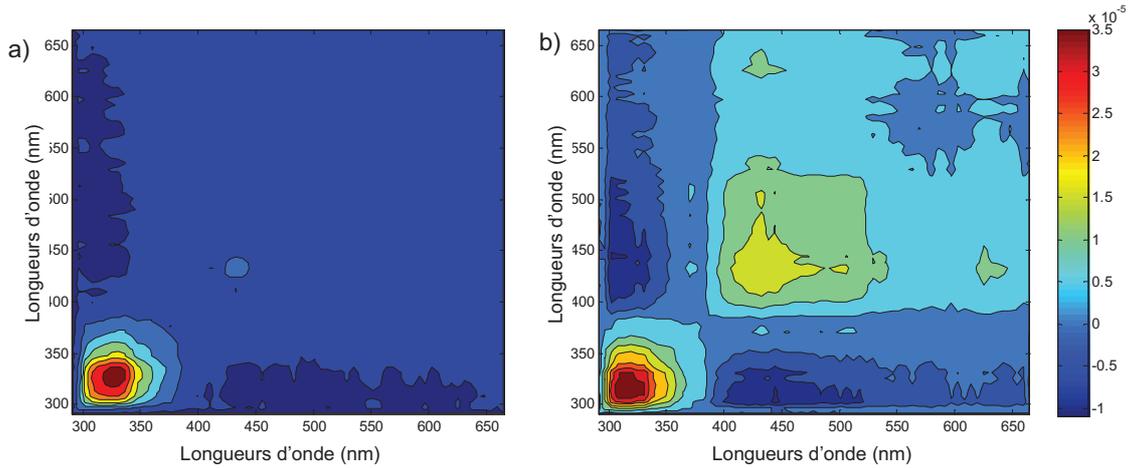


Figure 85 : Tranches représentatives de la matrice $\underline{\Sigma}$ à 3,2 et 50 ps. L'échelle de couleur correspond à la valeur de la covariance.

Les matrices de corrélation correspondantes sont présentées Figure 86. Les deux zones de bruit structuré décrites précédemment sont de nouveau visibles. Elles se séparent clairement l'une de l'autre Ceci permet de suspecter l'influence de deux facteurs distincts sur l'erreur de mesure. Même si l'interprétation qualitative de l'erreur est très claire en utilisant la matrice de corrélation, cette représentation travaille avec une échelle normalisée entre -1 et 1. Une partie de l'information sur l'intensité des erreurs est perdue, c'est pourquoi la représentation des matrices de covariances sera conservée par la suite dans ce chapitre.

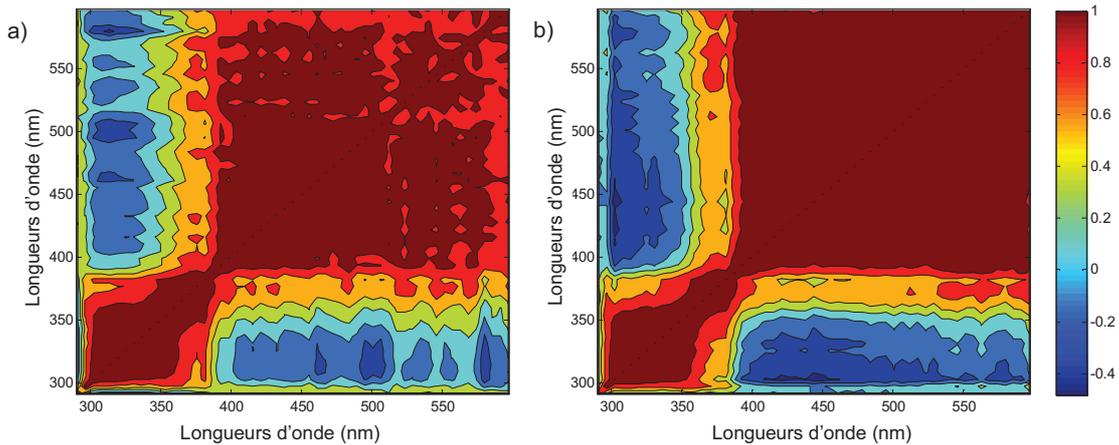


Figure 86 : Matrices de corrélation de l'erreur de mesure sur les longueurs d'onde au temps $t = 3,2$ ps et $t = 50$ ps. L'échelle de couleur correspond à la valeur de la corrélation.

L'erreur de mesure dans la direction spectrale répond donc à une structure, qui d'après les premières observations, semble être constituée de différents facteurs puisqu'elle influe de manière distincte dans deux zones. En effet si l'on compare toutes les tranches de la matrice $\underline{\Sigma}$, comme le permet la Figure 87, il est possible de voir que les deux zones peuvent atteindre des niveaux de covariance importants, apparemment indépendamment l'une de l'autre. La qualité de cette représentation a du être adaptée au grand nombre de sous matrices Σ_i .

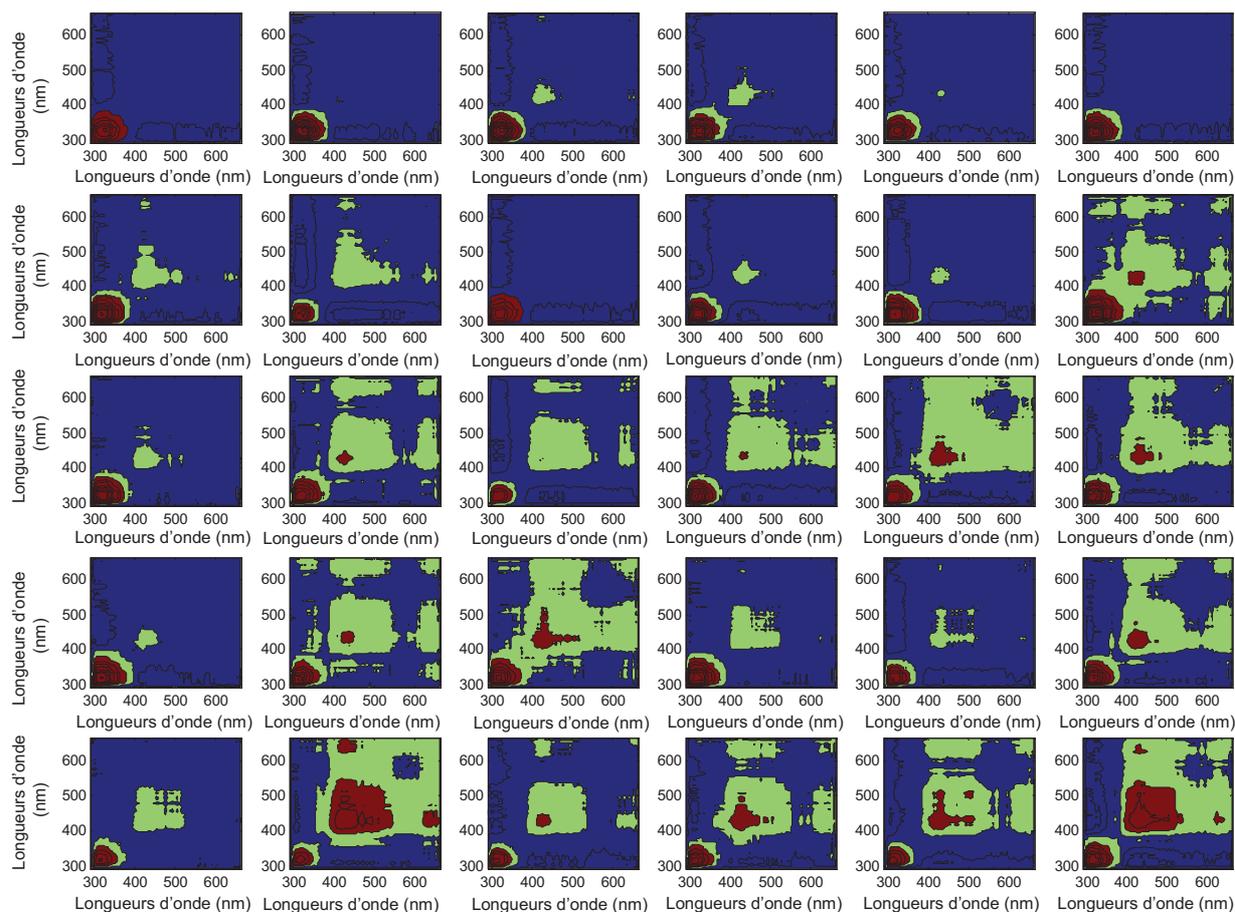


Figure 87 : Représentation de toutes les tranches formant le cube de covariance de l'erreur $\underline{\Sigma}$.

L'analyse et la modélisation des structures de l'erreur dans la dimension spectrale ainsi que leurs évolutions nécessitent donc une analyse chimiométrique plus poussée. La décomposition du cube $\underline{\Sigma}$ par PARAFAC fournit plusieurs informations. Plusieurs modèles PARAFAC ont été testés en variant le nombre de contributions utilisées. Deux contributions permettent de reconstruire la matrice $\underline{\Sigma}$ avec un pourcentage de variance capturée $R^2 = 99,99\%$. Les *loadings* des trois modes obtenus sont regroupés Figure 88. Les deux premiers modes sont symétriques tout comme chacune des sous matrices Σ_i . Les deux profils des *loadings* du mode spectral (lignes bleue et verte des Figure 88a et Figure 88b) ont des formes différentes et correspondent chacun à l'une des deux zones spectrales discutées précédemment. Les *loadings* du troisième mode (Figure 88c) semblent aléatoires au regard des profils obtenus. Cette décomposition montre clairement que deux facteurs de bruit structuré influencent l'erreur de mesure dans la direction spectrale, cependant PARAFAC n'offre pas d'interprétation physicochimique des *loadings* obtenus, pour cela d'autres outils chimiométriques peuvent être utilisés.

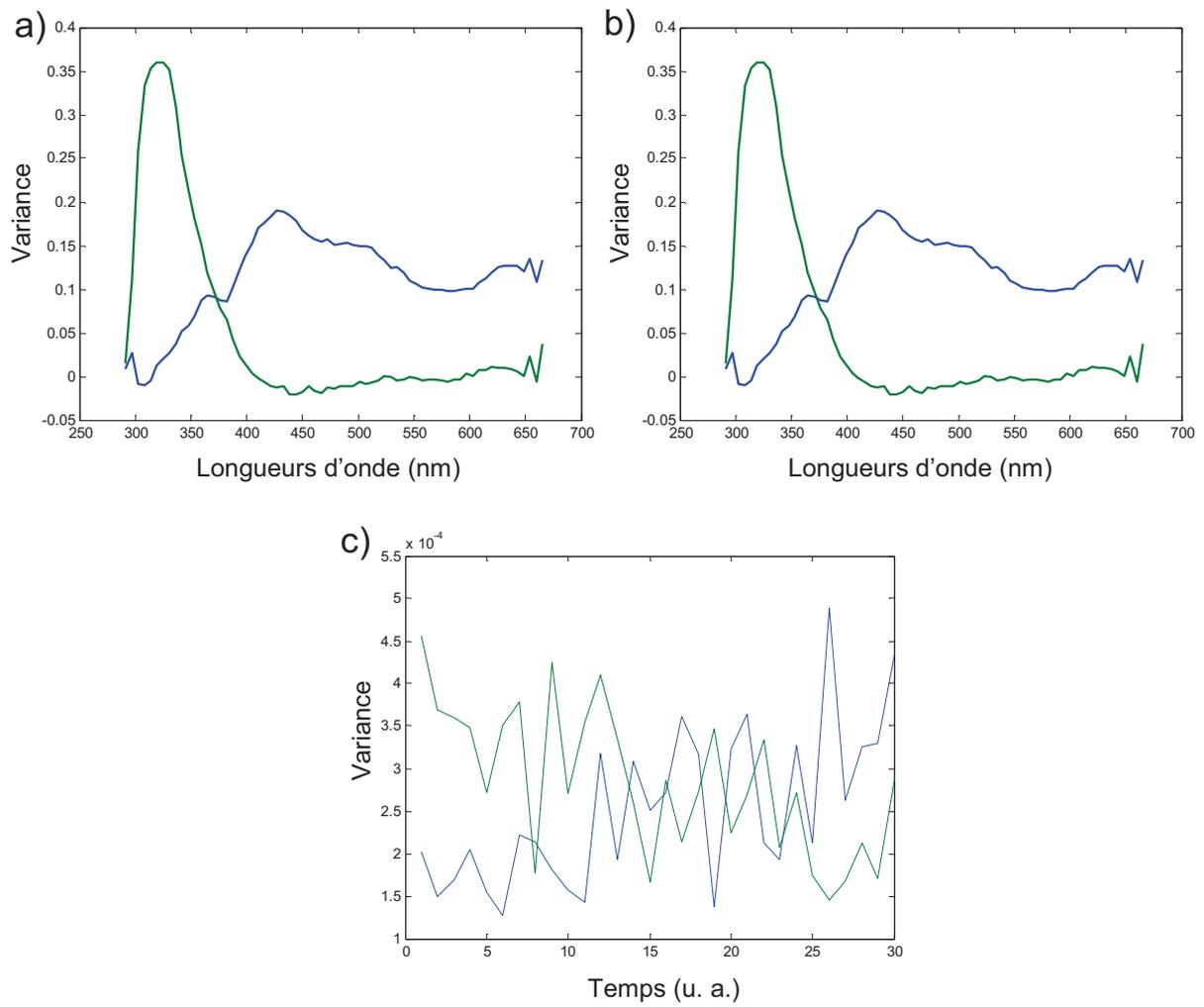


Figure 88 : Loadings de la matrice Σ obtenus par PARAFAC a) 1^{er} mode b) 2nd mode c) 3^{ème} mode.

2. Caractérisation des contributions structurées à l'erreur

Les contributions de l'erreur structurée observées en utilisant PARAFAC sont rarement interprétables en termes physico-chimiques. L'étape suivante consiste donc à tester différents facteurs suspectés. Pour ce faire, sur la matrice de covariance moyenne de l'erreur $\bar{\Sigma}$ (moyenne des m temps) est calculée. Si une erreur structurée avait été détectée dans la direction temporelle, on procéderait de manière analogue avec la matrice $\bar{\Psi}$ (moyenne des n longueurs d'onde). Ces matrices moyennes permettent de travailler avec les algorithmes de tests de cibles.

Le test de cibles est un groupe de méthodes permettant de relier des solutions en facteurs mathématiques abstraits à des facteurs physico-chimiques réels. L'idée générale est de tester par exemple si un facteur réel (vecteur ou profil cible) est proche ou non du sous-espace défini par les données. Les modèles développés à partir de cette méthode sont donc explicites. L'idée générale de l'algorithme d'Analyse Factorielle par Test de Cibles (TFA, *Target testing Factor Analysis*) [21] est de vérifier si une cible appartient au sous-espace défini par l'ACP. Dans le contexte de l'analyse de l'erreur, les cibles sont sources potentielles d'erreur qui peuvent être représentées par un profil et desquelles on veut tester l'appartenance à l'espace défini par la matrice de covariance de l'erreur $\bar{\Sigma}$.

On se place ici dans un cas simple où seules deux cibles, \mathbf{a} et \mathbf{b} , sont retenues. Une fois les cibles correctes identifiées, les contributions de chacune d'entre-elles à la covariance de l'erreur peuvent être estimées comme suit :

$$\begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} = [\mathbf{a} \ \mathbf{b}]^+ \Sigma_i ([\mathbf{a} \ \mathbf{b}]^T)^+ \quad (68)$$

Les coefficients σ_a^2 et σ_b^2 représentent la contribution des deux facteurs \mathbf{a} et \mathbf{b} à la covariance de l'erreur ; le terme σ_{ab} correspond à l'interaction des deux facteurs \mathbf{a} et \mathbf{b} lorsque la structure de bruit ne suit pas un modèle bilinéaire. Une matrice de covariance « modèle » Σ_{calc} peut alors être recalculée :

$$\Sigma_{\text{calc}} = [\mathbf{a} \ \mathbf{b}] \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} [\mathbf{a} \ \mathbf{b}]^T \quad (69)$$

L'intérêt de cette modélisation est de permettre de recalculer une estimation de l'erreur pour d'autres expériences sans avoir à les répéter. Il suffit de réadapter les facteurs utilisés au

nouveau cas. Les estimations obtenues peuvent ensuite être utilisées dans des schémas de pondération des données dans les méthodes de résolution de courbes multivariées.

L'analyse TFA a donc été appliquée à notre cas. Différentes cibles ont été envisagées. Parmi celles-ci on trouve : le signal moyen, le signal de source moyen, l'écart-type du signal et du signal de source, un offset, des déviations de ligne de base. Le Tableau 12 résume les principales cibles utilisées et la qualité de la cible pour décrire les données (sous la forme de l'angle entre la cible et l'espace de données). Un angle faible signifie que le vecteur testé est proche du plan décrit par les données. On constate que les deux angles minimaux, c'est-à-dire les deux meilleurs accords, sont fournis pour le signal moyen (à ne pas confondre avec l'hétéroscédasticité) et la moyenne du signal de source. Une comparaison graphique de ces cibles, présentées Figure 89, avec les *loadings* de PARAFAC, Figure 88 , confirme ces conclusions.

Cible	Angle (°)
Signal Moyen	3,9
Source (moyenne)	13,1
Offset	13,5
1/lambda	14
Lambda	16,3
Source (écart-type)	16,8

Tableau 12 : Résultats de la TFA sur la matrice de covariance de l'erreur moyenne $\bar{\Sigma}$.

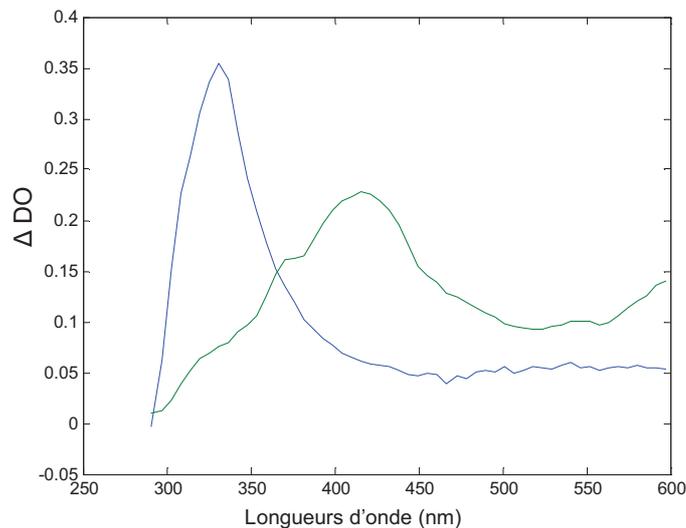


Figure 89 : Cibles sélectionnées par TFA, en bleu le signal moyen et en vert le signal moyen de la source.

Ces deux vecteurs ont ensuite été utilisés pour calculer les coefficients de l'équation (68), on obtient les valeurs suivantes :

$$\begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} = \begin{bmatrix} 0,0004 & 0,0002 \\ 0,0002 & 0,0003 \end{bmatrix} \quad (70)$$

On peut constater que les deux facteurs σ_a^2 et σ_b^2 ont des influences comparables mais surtout que l'interaction σ_{ab} de ces deux facteurs est presque aussi importante. La matrice de covariance de l'erreur a ensuite été reconstruite en accord avec l'équation (69). La Figure 90 permet de comparer la diagonale de la matrice reconstruite avec la diagonale de la matrice de covariance de l'erreur $\bar{\Sigma}$ originale.

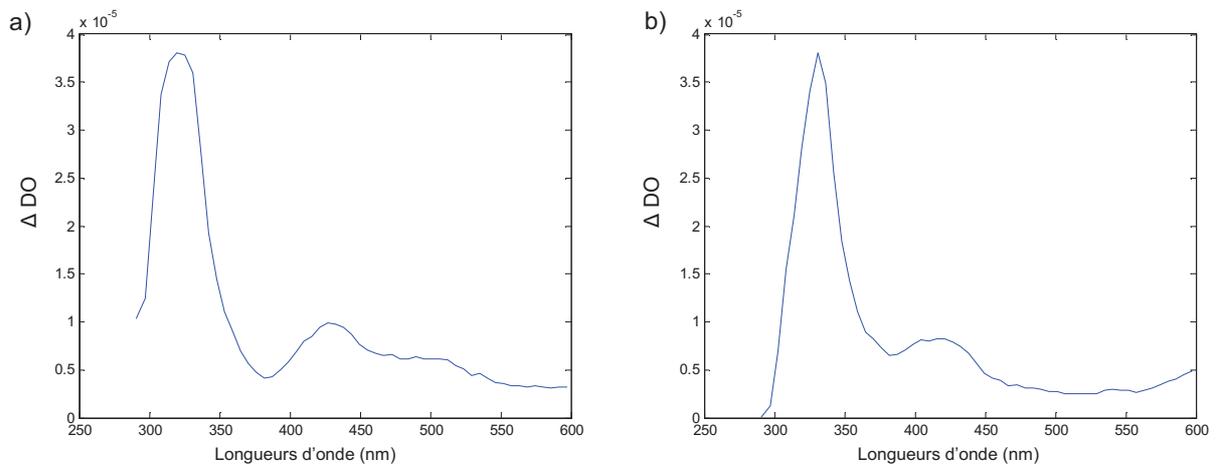


Figure 90 : Diagonale de a) la matrice de covariance de l'erreur $\bar{\Sigma}$ et b) sa reconstruction basée sur les vecteurs choisis par TFA

L'analyse TFA effectuée ici a deux intérêts, le premier est d'identifier les facteurs physico-chimiques responsables des bruits structurés observés. Le second est de permettre une modélisation de ces bruits qui pourra être réutilisée et/ou adaptée lors de la pondération des données. Cette pondération a pour but de donner moins d'importance au signal lorsque celui-ci est plus influencé par le bruit et donc d'améliorer la qualité de la résolution. C'est ce que nous verrons dans la section suivante.

3. Pondération des données en fonction de l'erreur

Les informations obtenues sur la nature du ou des bruits affectant la mesure peuvent être exploités de deux façons. La première est d'optimiser de manière concrète l'instrumentation utilisée de manière à réduire l'influence des facteurs parasites identifiés. La seconde consiste à prendre en compte ces informations lors du traitement chimométrique des données enregistrées. La modification de MCR-ALS proposé par Wentzell en 2006 [149] se révèle alors être un outil précieux. L'algorithme est rebaptisé MCR-WALS pour *Multivariate Curve Resolution using Weighted Alternating Least Squares*. Les données originales \mathbf{D} sont pondérées à chaque itération ligne par ligne et colonne par colonne en fonction des matrices de covariance de l'erreur $\mathbf{\Sigma}$ et $\mathbf{\Psi}$ selon les équations (71) et (72) respectivement :

$$\hat{\mathbf{x}}_i = \mathbf{x}_i \mathbf{\Sigma}_i^{-1} \mathbf{S}^T (\mathbf{S} \mathbf{\Sigma}_i^{-1} \mathbf{S}^T)^{-1} \mathbf{S} \quad (71)$$

$$\hat{\mathbf{x}}_j = \mathbf{C} (\mathbf{C}^T \mathbf{\Psi}_j^{-1} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{\Psi}_j^{-1} \mathbf{x}_j \quad (72)$$

Le reste de l'algorithme est ensuite parfaitement similaire à MCR-ALS, sauf pour le calcul des résidus qui doit lui aussi être pondéré. Jusqu'à présent, l'algorithme MCR-WALS a été appliqué avec grand succès à des données montrant une hétéroscédasticité importante, telles que les données provenant de puces à ADN [149]. Même si la formulation théorique envisageait la possibilité de pondérer des données ayant un bruit structuré et en proposait la stratégie [150], l'effet d'introduire ce bruit sur la résolution des données n'a pas été effectivement testé. C'est cela que nous proposons de faire sur des données simulées et réelles.

Jusqu'à présent, pour prendre en compte les bruits indépendants hétéroscédastiques observés dans les puces à ADN ou d'autres données similaires, la pondération des lignes et des colonnes de données se faisait en utilisant des matrices diagonales de covariance de l'erreur. En effet, toute l'information intéressante se concentre sur la diagonale à cause de la nature indépendante de l'erreur. Ce schéma de pondération permettait de travailler avec une seule matrice de covariance $\mathbf{\Sigma}$ et une seule matrice $\mathbf{\Psi}$ si on considère que le schéma hétéroscédastique se répétait dans toutes les lignes et toutes les colonnes de \mathbf{D} , respectivement, ou avec autant de matrices $\mathbf{\Sigma}_i$ et $\mathbf{\Psi}_j$ que de lignes et colonnes si le bruit indépendant variait entre lignes et colonnes. On pouvait aussi avoir des situations intermédiaires, telles qu'utiliser une seule matrice $\mathbf{\Sigma}$ et toutes les matrices $\mathbf{\Psi}_j$, ou vice versa. Le choix dépendait de la structure du bruit dans les deux directions.

Les données présentées dans ce chapitre sont différentes. Les erreurs structurées observées en spectroscopie d'absorption transitoire nécessitent une pondération plus complète et plus complexe. L'analyse effectuée dans la section précédente a permis de montrer que la direction temporelle n'est pas affectée par une erreur structurée. La pondération des colonnes de la matrice ne nécessitent alors que d'utiliser des matrices diagonales Ψ_j , variables selon la colonne dans l'équation (72).

Au contraire l'erreur dans la direction spectrale est affectée par deux contributions structurées. Il est alors nécessaire de considérer l'ensemble de la matrice de covariance de l'erreur Σ lors de la pondération. Ceci permet de prendre en compte le poids des erreurs indépendantes (sur la diagonale) et la corrélation de l'erreur entre les différents canaux (éléments hors diagonale). Comme la forme de la structure se répète dans toutes les lignes de la matrice de données, on utilisera la même matrice de covariance Σ pour pondérer toutes les lignes de \mathbf{D} .

L'intérêt de cette approche va être mis à l'épreuve dans le paragraphe suivant en utilisant des données simulées. L'objectif est de connaître et de contrôler parfaitement les bruits impliqués et d'adapter les schémas de pondération aux différents cas envisagés.

3.1 Simulation de données affectées par une erreur structurée

La principale difficulté pour valider cette méthode d'analyse du bruit est que par définition on ne connaît pas parfaitement le bruit, sauf si les données utilisées sont simulées. Un ensemble de trois spectres réels et de trois profils de concentration d'un modèle cinétique simple a donc été utilisé pour former une matrice de données. Les spectres purs sont issus des résolutions présentées au chapitre précédent et les profils de concentration ont été construits en utilisant un modèle du premier ordre $A \rightarrow B \rightarrow C$. Ils sont présentés Figure 91.

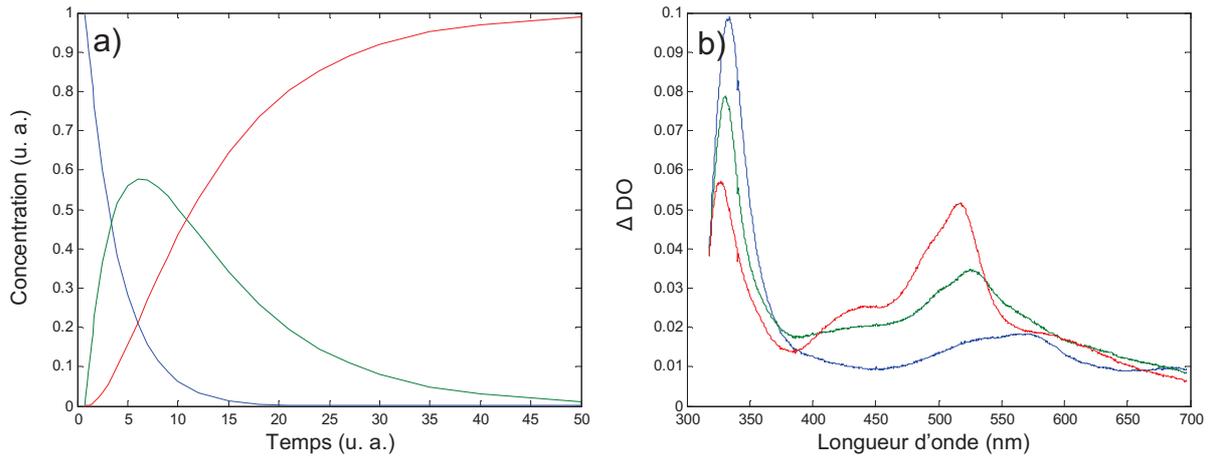


Figure 91 : a) Profils de concentration et b) spectres purs utilisés pour construire les matrices \mathbf{D}_1 à \mathbf{D}_4 .

On ajoute ensuite à celle-ci différentes combinaisons de bruits. La matrice \mathbf{D}_1 correspond au cas où seuls des bruits homoscédastiques et hétéroscédastiques sont ajoutés. \mathbf{D}_2 est affectée par les mêmes bruits indépendants et en plus une erreur structurée : celle qui correspond à l'influence de la source. L'erreur de la source a été générée en prenant la cible utilisée à la section 2 et pour chaque ligne a été multipliée par un facteur d'échelle et signe aléatoire. Cette erreur n'est corrélée que dans la direction spectrale. \mathbf{D}_3 est affectée par des bruits indépendants et une erreur structurée ayant la même forme que le signal moyen. Cette forme a été "rendue aléatoire" comme l'erreur de la source. Enfin \mathbf{D}_4 correspond au cas où les bruits indépendants et les deux erreurs structurées sont présents. Le rapport d'intensité (en valeur absolue) entre ces deux erreurs a été choisi de manière à privilégier l'une des erreurs (1/3 pour l'erreur de la même forme que le signal et 2/3 pour la seconde). Les erreurs sont ensuite multipliées aléatoirement par un facteur ± 1 et ensuite par un facteur d'échelle aléatoire.

Chaque matrice de données obtenue est constituée de 30 spectres de 1340 longueurs d'onde. La matrice \mathbf{D}_2 est présentée Figure 92, le bruit est ici suffisamment important pour être visible directement sur les spectres, cependant la présence d'une erreur structurée ne peut pas être attestée sans l'analyse présentée précédemment. Les bruits présents étant connus on peut calculer un *Lof* théorique. Il est par exemple égal à 4,86 % pour \mathbf{D}_2 .

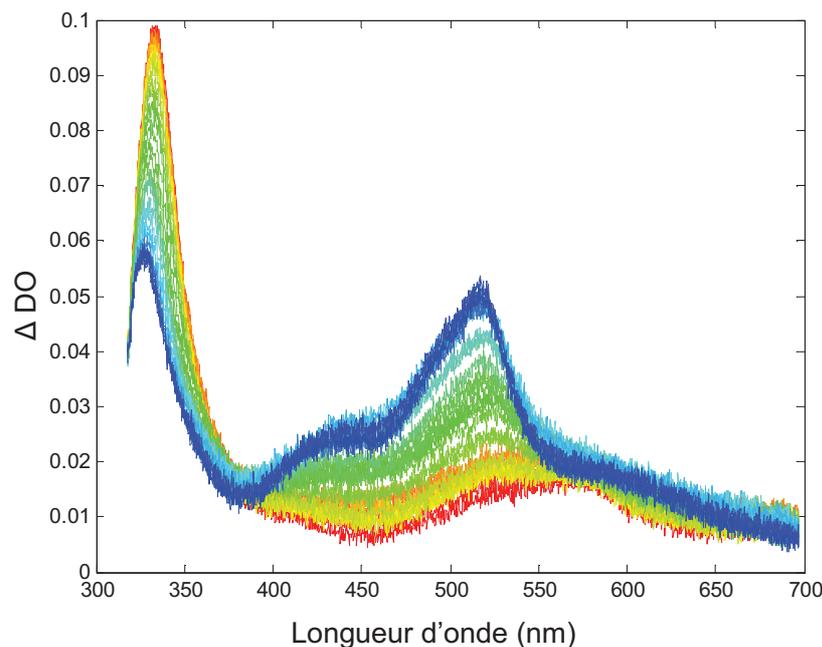


Figure 92 : Matrice de données simulées D_2 .

Les premières étapes de la résolution sont toujours les mêmes : on estime d'abord le rang par SVD. Les premières valeurs singulières pour les différentes matrices sont regroupées dans le Tableau 13. Trois contributions se détachent du bruit dans les différents cas.

D_1	D_2	D_3	D_4
5,2665	5,2679	5,2488	5,2244
1,2533	1,2905	1,2527	1,2428
0,1102	0,1425	0,1008	0,1276
0,0609	0,0776	0,0203	0,0720
0,0590	0,0413	0,0199	0,0100
0,0576	0,0405	0,0198	0,0100
0,0569	0,0403	0,0196	0,0098

Tableau 13 : Premières valeurs singulières des matrices de données simulées.

Comme prévu, le rang est clairement égal à trois dans les matrices D_1 et D_3 , la première ne montrant que du bruit indépendant et le second ayant un bruit de même forme que le signal. Dans ce deuxième cas, le bruit est complètement incorporé aux valeurs singulières significatives puisqu'il n'est pas défini par une forme différente de celle du signal chimique. Dans D_2 et D_4 , on propose encore un rang de trois, même si la quatrième valeur singulière est un peu plus grande que les suivantes. Ceci peut être expliqué par le fait que la forme structurée de la contribution de la source est différente de celle du signal mesuré.

Ensuite l'EFA permet de construire une estimation initiale des profils de concentration. La présence d'une erreur structurée peut parfois rendre cette étape plus difficile : il ne faut pas interpréter la structure de l'erreur comme une contribution significative. Ceci est particulièrement important pour des niveaux élevés de bruit.

Ces différentes simulations ont pour but de tester la pertinence de la pondération des données par une estimation de l'erreur les affectant. Les différentes résolutions obtenues seront comparées et détaillées par la suite. Pour chacune d'elles, nous avons appliqué comme contraintes la non négativité des spectres et des concentrations. Les schémas de pondération utilisés par MCR-WALS sont les suivants :

D₁ : matrices de covariance diagonales Σ_i et Ψ_j différentes pour chaque ligne et chaque colonne, respectivement.

D₂ : matrices de covariance diagonales Ψ_j différentes pour chaque colonne et une seule matrice de covariance complète Σ , qui est générée à partir de la forme de la source.

D₃ : matrices de covariance diagonales Ψ_j différentes pour chaque colonne et une seule matrice de covariance complète Σ , générée à partir de la forme du signal moyen.

D₄ : deux schémas de pondération différents sont proposés. Pour chacun d'entre eux, nous avons utilisés des matrices de covariance diagonales Ψ_j différentes pour chaque colonne et dans un cas (1) une seule matrice de covariance complète Σ , qui est générée à partir de la combinaison des formes de la source et du signal moyen, c'est-à-dire, comme le bruit vrai et dans le second cas (2) une seule matrice de covariance complète Σ , qui est générée uniquement à partir de la forme de la source.

Nous nous concentrerons désormais sur l'effet de la pondération sur l'ajustement des données et sur la récupération des profils de concentration et des spectres. Cependant il est plus commode de rassembler ici premièrement les différentes valeurs de *Lof* obtenues dans chacun de ces cas. Le Tableau 14 résume donc les *Lof* obtenus pour les résolutions des matrices **D₁**, **D₂**, **D₃** et **D₄** en utilisant MCR-ALS et MCR-WALS. Les *Lof* théoriques sont également fournis. On peut constater que dans la plupart des cas le surajustement observé lors de l'utilisation de MCR-ALS est (au moins partiellement) compensé en utilisant MCR-WALS et un schéma de pondération adapté à la nature de l'erreur

	D₁	D₂	D₃	D₄
Lof théorique (%)	4,9	6,1	6,3	5
Lof MCR-ALS (%)	4,7	3,7	1,7	1,3
Lof MCR-WALS (%)	5	5,9	1,7	5,24 / 4,49

Tableau 14 : Lof théoriques et des différents modèles MCR-ALS et MCR-WALS pour les matrices **D₁**, **D₂**, **D₃** et **D₄**. Les deux valeurs fournies pour la résolution MCR-WALS de **D₄** correspondent aux schémas de pondération (1) et (2) appliqués.

La résolution obtenue à partir de **D₁** en utilisant MCR-ALS et MCR-WALS est présentée Figure 93. Les solutions sont contraintes à la non négativité des spectres et des concentrations. Le bruit ajouté dans cette simulation génère théoriquement un *Lof* de 4,9%. Les profils de concentration et les spectres obtenus sont comparables entre eux et avec les originaux utilisés pour construire la simulation. Dans les deux cas la résolution est correcte comme l'indique les *Lof* obtenus, respectivement 4,7 % et 5,0%, où seul MCR-ALS surajuste très légèrement. L'effet est visible sur les spectres purs résolus plus bruités que ceux issus de MCR-WALS. La présence de bruits indépendants entre dans le cadre des hypothèses sur lesquelles est construit MCR-ALS. Ce type de bruit est identifié en tant que tel et est rejeté plus facilement dans la matrice des résidus **E**, comme le fait MCR-WALS. Il est important de noter que l'hypothèse de base de MCR-ALS est que le bruit est indépendant et homoscédastique. Cela explique que, lorsque l'hétéroscédasticité est plus sévère que dans cet exemple, l'on puisse constater un effet bénéfique plus clair de MCR-WALS par rapport au MCR-ALS classique [149].

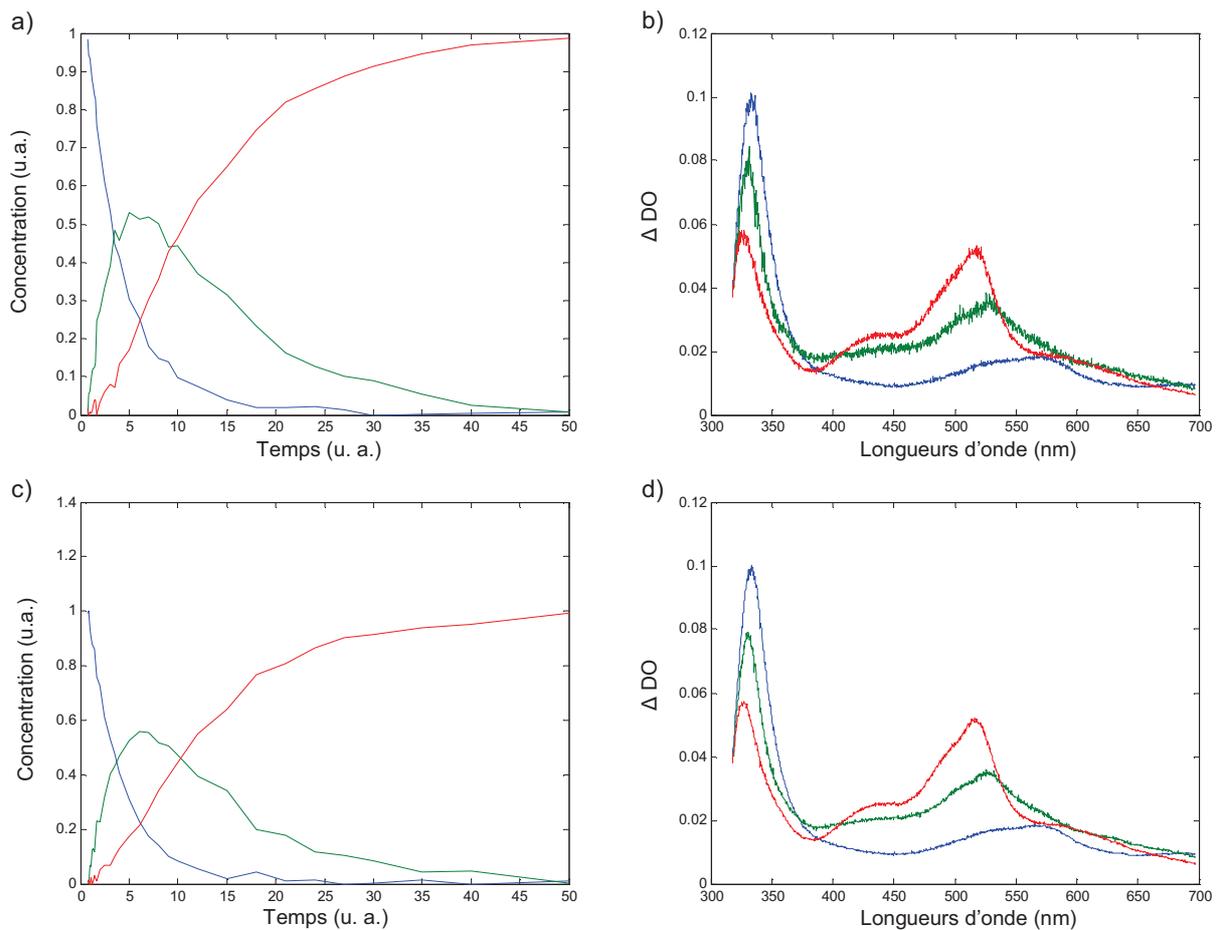


Figure 93 : Résolutions de D_1 obtenues par MCR-ALS a) profils de concentration b) spectres purs et par MCR-WALS c) profils de concentration d) spectres purs.

Le cas de D_2 est différent. En effet, l'erreur structurée ajoutée ne correspond pas à un bruit tel que ceux prévus lors de la conception de MCR-ALS. L'erreur totale doit mener à un *Lof* de 6,1%. La résolution obtenue à partir de D_2 et en utilisant MCR-ALS est présentée Figure 94. On constate que les profils de concentration et les spectres sont déformés à cause de la présence du bruit structuré. L'algorithme a interprété le bruit structuré comme un signal significatif et ne l'a donc pas rejeté dans la matrice de résidus E . On se trouve donc aussi dans une situation de surajustement clair. Ceci est visible sur le *Lof* qui atteint 3,7 % soit plus de 2,5 % de mieux que ce qui est théoriquement possible. La résolution obtenue en utilisant MCR-WALS est présentée Figure 94. Les profils de concentration et les spectres purs ne sont plus perturbés par la présence de bruit structuré ; la pondération des données a donc permis d'éviter le phénomène de surajustement et de déformation des profils. D'ailleurs le *Lof* obtenu est en accord avec la valeur théorique puisqu'il atteint 5,9%.

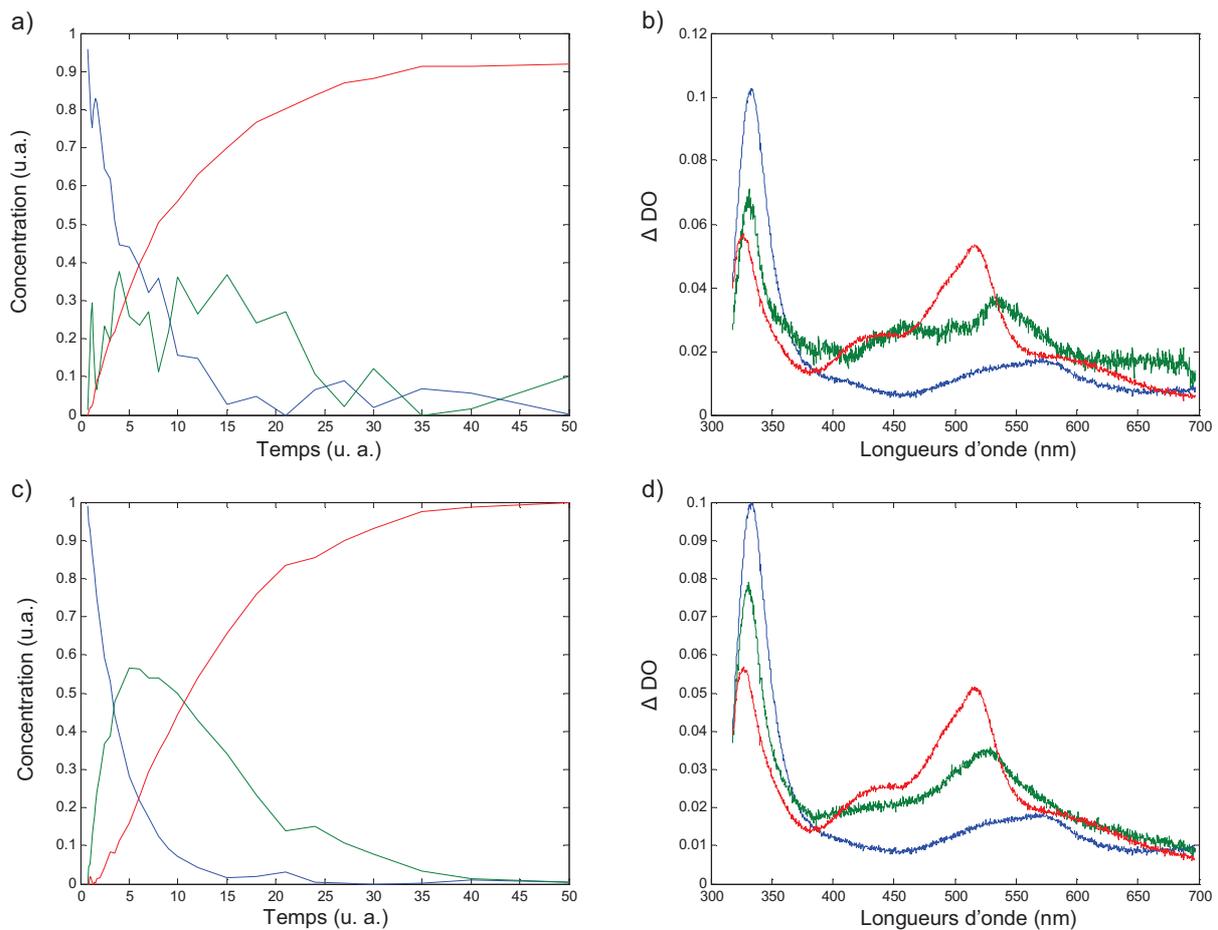


Figure 94 : Résolutions de D_2 obtenues par MCR-ALS a) profils de concentration b) spectres purs ; par MCR-WALS c) profils de concentration d) spectres purs.

Le cas de D_3 est fondamentalement le même que celui de D_2 à un détail près concernant la forme de l'erreur structurée. En effet l'allure de celle-ci est cette fois comparable à celle du signal. La pondération va donc donner moins de poids aux zones où le signal est important, l'algorithme risque alors de valoriser les zones où le signal est plus faible et donc moins utile. Cependant ce problème n'est critique que dans des jeux de données où les différences d'intensité sont considérables, ce qui n'est pas le cas sur cet exemple et, en général, dans des données spectroscopiques. Les résolutions respectives de MCR-ALS et MCR-WALS de D_3 sont présentées Figure 95. Dans les deux cas les profils de concentration obtenus sont affectés par le bruit. L'ajout de cette erreur structurée est donc clairement un handicap pour la résolution : cette structure est identifiée comme faisant partie du signal et mène à une modélisation incorrecte. L'application de la pondération dans ce cas n'aide pas à améliorer la résolution, mais elle n'empire pas non plus la situation. Les *Lof* théoriques, de MCR-ALS et de MCR-WALS sont ici respectivement de 6,3%, 1,7% et 1,7%. Les mauvais

résultats de la pondération confirment les considérations théoriques prévues [150], qui décrivaient cette erreur comme facilement identifiable mais impossible à gérer dans la pratique puisque la forme du signal et de l'erreur sont identiques et donc indistinguables l'un de l'autre. Il est possible de faire l'analogie avec la même situation en cas de déficience de rang causés par la présence de profils identiques.

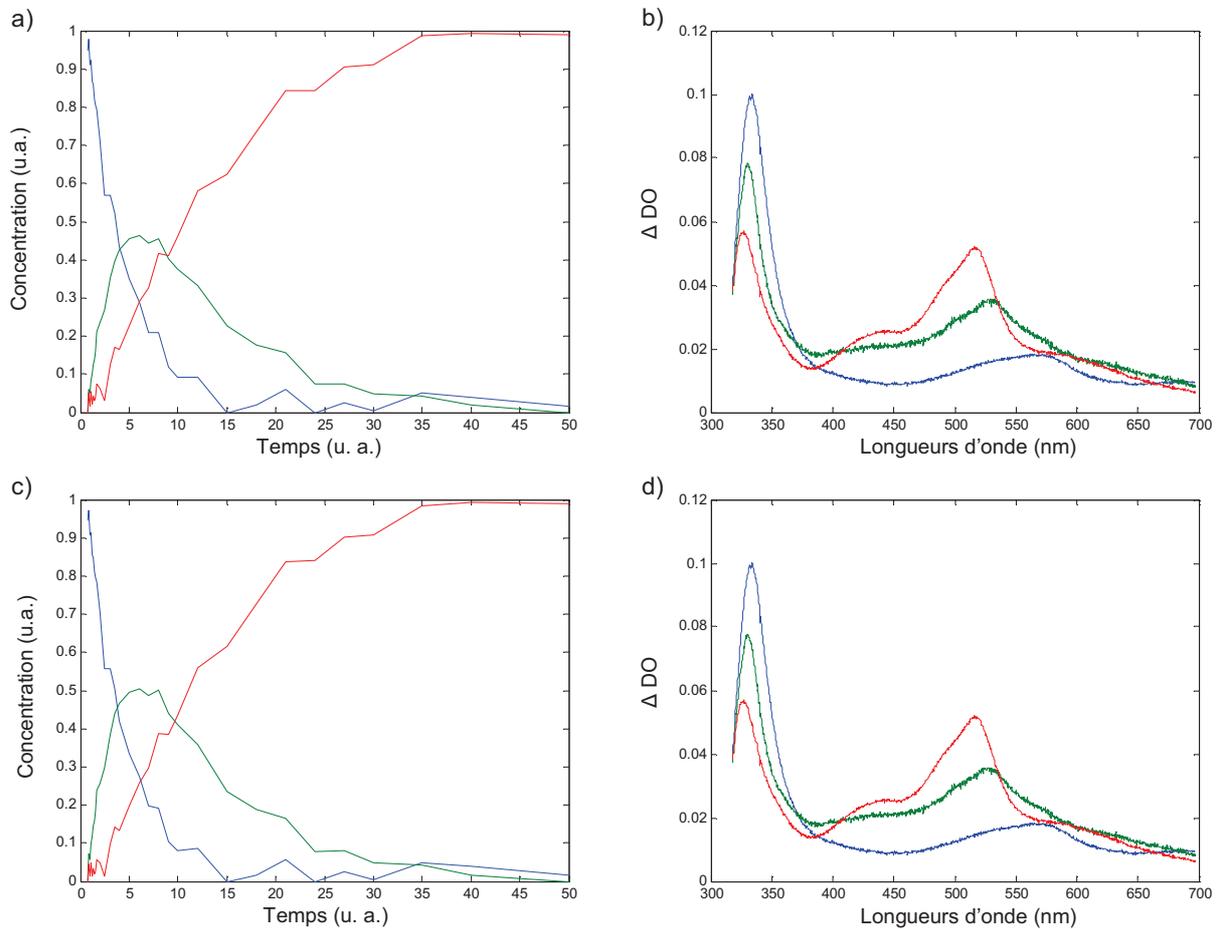


Figure 95 : Résolutions de D_3 obtenues par MCR-ALS a) profils de concentration b) spectres purs et par MCR-WALS c) profils de concentration d) spectres purs.

Finalement un dernier cas est envisagé. D_4 correspond à la situation observée sur les données réelles : une partie de l'erreur provient de bruits indépendants et le reste de deux erreurs structurées dont une ayant la forme du signal. On vient de voir que la pondération ne permet pas de compenser la dernière forme d'erreur cependant il est intéressant de vérifier que le bénéfice observé dans le cas de D_2 est toujours observable si les deux erreurs sont présentes. La décomposition par MCR-ALS est calculée comme précédemment sous contraintes de non négativité des spectres et des concentrations mais MCR-WALS a été appliqué de deux manières distinctes. La première consiste à pondérer en utilisant une estimation correcte de l'erreur c'est-à-dire comprenant les deux erreurs structurées. La

seconde ne pondère qu'en fonction du bruit indépendant et de l'erreur structurée différente du signal (celle présente dans \mathbf{D}_2). Ceci nous permettra de savoir s'il est utile d'utiliser l'information complète disponible sur l'erreur ou s'il est suffisant d'identifier la deuxième source d'erreurs corrélées. La Figure 96 présente ces trois résolutions. De nouveau MCR-ALS surajuste les données menant à des résultats absurdes et à un *Lof* de 1,3 %. Le premier schéma de pondération améliore la qualité de la résolution avec un *Lof* de 5,24%. Les spectres comme les profils de concentration sont comparables aux originaux. Il est intéressant de noter que le second schéma de pondération mène à des résultats très similaires (aux ambiguïtés d'intensité près). Le *Lof* obtenu est alors égal à 4,49%. La pondération incluant l'erreur structurée ayant la forme du signal n'est donc pas plus utile que celle qui utilise seulement la contribution de la source. Il est donc possible de conclure que dans ce cas le second schéma de pondération est suffisant pour améliorer la qualité de la résolution. Cette conclusion est plus importante qu'il n'y paraît puisque, lorsque le bruit structuré est formé par plus d'une contribution, le point critique est d'établir correctement quel est le poids de chacune de ces contributions par rapport à l'autre. Ceci est possible pour une expérience concrète répliquée mais ne s'avère pas facilement extrapolable pour de nouvelles expériences. Étant donné qu'une des contributions a souvent la forme du signal, on peut penser à des schémas de pondération qui l'ignorent et, par conséquence, qui évitent la détermination du poids relatif des contributions dans des bruits multistrués.

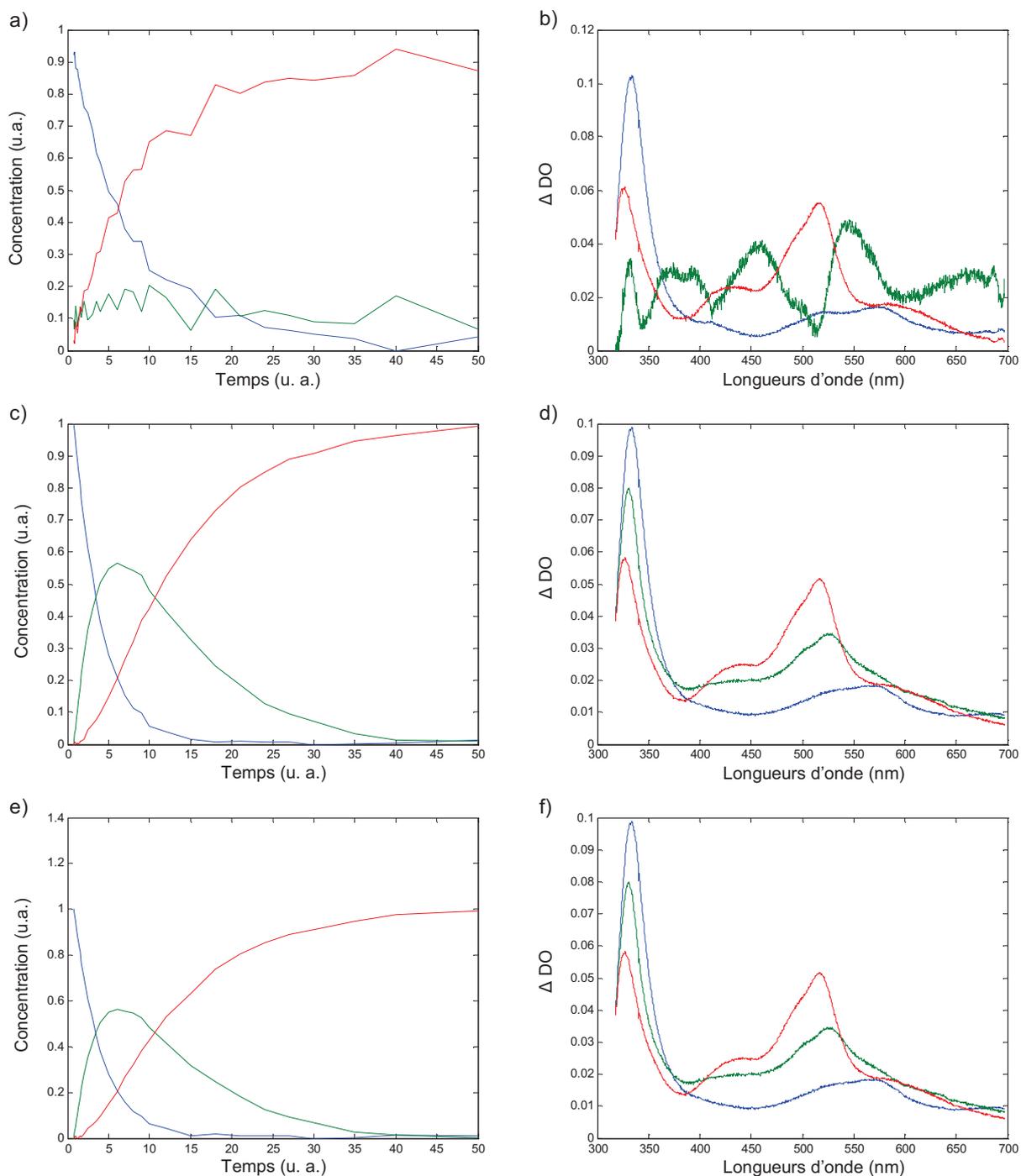


Figure 96 : Résolutions de D_4 obtenues par MCR-ALS a) profils de concentration b) spectres purs ; par MCR-WALS en utilisant un schéma de pondération complet c) profils de concentration d) spectres purs ; par MCR-WALS en utilisant un schéma de pondération de considérant que l'erreur structurée ayant une forme différente du signal e) profils de concentration f) spectres purs.

3.2 Cas réel

Les résultats précédents ont été obtenus sur des simulations. L'analyse du bruit a par contre été effectuée sur des données réelles, même s'il faut noter que ces dernières ont été volontairement enregistrées de manière à observer un important niveau de bruit. Il peut alors être intéressant d'appliquer cette méthodologie aux données réelles présentées dans le chapitre IV. Nous avons donc choisi de calculer la décomposition de la matrice \mathbf{D}_1 , c'est-à-dire de la matrice ayant le meilleur rapport signal sur bruit. L'algorithme MCR-WALS a été appliqué en utilisant un schéma de pondération prenant en compte les bruits indépendants et le bruit structuré dépendant de la source lumineuse. Les contraintes appliquées sont la non négativité des spectres et des concentrations. La Figure 97 présente le résultat de cette décomposition. Le Lof et le R^2 obtenu sont respectivement de 1,41% et 99,98%. Cette résolution est accord avec celles précédemment obtenues en utilisant MCR-ALS et HS-MCR ($Lof = 0,59\%$). Le Lof et le R^2 sont comparables, ce qui indique que l'influence du bruit était plus faible sur ces données de meilleure qualité et surtout ceci valide les résultats présentés au chapitre précédent.

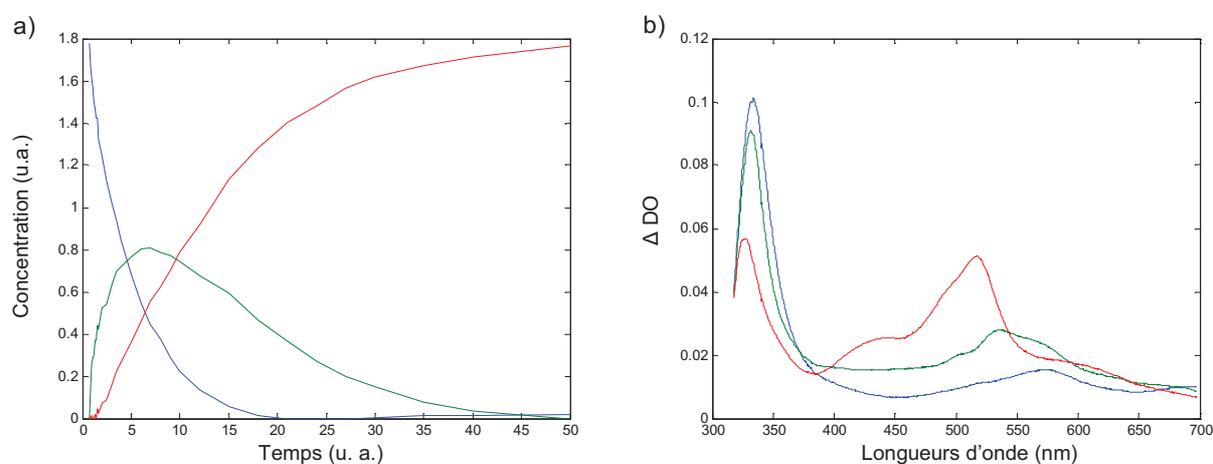


Figure 97 : a) Profils de concentrations et b) spectres obtenu par MCR-WALS en appliquant un schéma de pondération prenant en compte le bruit indépendant et le bruit structuré dépendant de la source.

4. Conclusion et perspectives

Nous avons appliqué une méthode permettant à partir de répliqués d'estimer et de caractériser les facteurs physico-chimiques générant du bruit sur des données de spectroscopie transitoire. Le calcul des matrices de covariance de l'erreur est particulièrement utile pour détecter la présence éventuelle d'erreurs corrélées. Ces informations sont exploitées dans une variante de MCR-ALS, MCR-WALS, qui utilise la connaissance de la structure du bruit pour pondérer les données. Celle-ci est alors capable d'extraire l'information chimique d'intérêt même pour des données fortement bruitées ou présentant des structures de bruit particulières.

Les simulations présentées montrent que la pondération des données en fonction d'une connaissance *a priori* des erreurs permet d'empêcher le surajustement lors de la résolution et la déformation des profils résolus. La robustesse de l'analyse en est améliorée. Cependant il est important de se souvenir que cette approche nécessite une connaissance fiable des erreurs présentes pour ne pas appliquer des schémas de pondération erronés. Il est donc nécessaire de passer par l'étape d'analyse de l'erreur ce qui implique d'enregistrer un nombre important de répliqués. On a pu constater que la pondération utilisant la matrice de covariance complète donne des résultats très prometteurs pour gérer des bruits structurés qui diffèrent de la forme du signal. Lorsque l'erreur structurée a une forme comparable au signal, on sait la détecter sans pour autant pouvoir en tirer profit lors de la résolution.

Il faut cependant avouer que la plupart des données spectroscopiques en suivi de procédés ont une qualité suffisante pour ne pas noter significativement les effets de la pondération, ce qui explique les résultats positifs obtenus avec le MCR-ALS classique. L'étude du bruit devrait toujours être faite pour avoir une connaissance plus profonde de la mesure et pour pouvoir tester correctement l'influence d'un schéma de pondération sur la qualité des résultats obtenus.

Conclusion

L'objectif général de cette thèse était de proposer des méthodes d'analyse multivariée adaptée aux données issues des spectroscopies résolues en temps. Deux axes de recherche se sont dégagés: la gestion de données de différence et l'étude de l'influence du bruit.

Le premier est en relation avec la structure différentielle de certaines données issues des spectroscopies résolues en temps. Nous avons pu montrer que les données de spectroscopie de différence sont systématiquement affectées d'une déficience de rang. Cette dernière peut éventuellement s'ajouter à une autre déficience de rang due au schéma réactionnel du procédé étudié. La formalisation de ce problème nous a permis de vérifier que la cinétique suivie par les contributions au signal reste la même que celle suivie par les espèces chimiques impliquées dans le mélange étudié. Nous avons donc proposé d'utiliser cette propriété pour contourner les difficultés posées par la déficience de rang. L'utilisation d'un modèle cinétique qui inclue les espèces initiales comme transparentes permet en effet de reconstruire tous les profils de concentration et d'obtenir les spectres de différence des espèces intermédiaires et finales du procédé. Un algorithme de modélisation hybride (HS-MCR) est proposé en ce sens. Il permet à la fois de gérer le problème de la déficience de rang à travers l'implémentation d'un modèle cinétique mais aussi de gérer d'éventuelles interférences absorbantes.

Cette stratégie a été appliquée avec succès à l'étude de données concernant le centre réactionnel de la bactérie *Rhodobacter sphaeroides*. Nous avons pu montrer que les méthodes de résolution de courbes multivariées souples (MCR-ALS) permettent d'obtenir une décomposition interprétable des données enregistrées en spectroscopie de différence IRTF *rapid-scan*. Nous avons ensuite adapté l'algorithme hybride aux nécessités spécifiques de la modélisation d'expériences suivant tour à tour l'excitation et la relaxation d'un échantillon. Les modèles obtenus permettent de constater que le procédé photosynthétique de formation d'ubiquinol peut emprunter deux voies distinctes en fonction du niveau de l'excitation lumineuse. Dans les deux cas ce processus s'accompagne d'un réarrangement de l'environnement protéique du centre réactionnel, qui se produit à une échelle de temps beaucoup plus rapide que celle suivie par la spectroscopie IRTF *rapid-scan*, et qui est décrite comme une contribution additionnelle au signal modélisé avec des contraintes souples.

Les méthodes de résolution de courbes multivariées ont ensuite été appliquées à l'analyse de données de spectroscopie d'absorption transitoire. L'objectif était de lever l'incertitude concernant l'existence d'un intermédiaire photophysique (prévue par la théorie) lors de la formation de l'état triplet T_1 de la benzophénone après une photoexcitation. Les résultats obtenus en utilisant aussi bien les méthodes de *soft modelling* (MCR-ALS) que l'algorithme hybride (HS-MCR) semblent indiquer que cet intermédiaire existe. Les spectres purs obtenus sont associables à un état triplet T_2 . Cette hypothèse est appuyée par l'étude menée en parallèle sur la méthoxybenzophénone. La nature des états excités de cette molécule étant dépendante du solvant utilisé, nous avons pu valider les outils chimiométriques utilisés et aussi attester de l'existence de l'intermédiaire.

Le second axe de travail concerne l'influence du bruit aussi bien sur la mesure que sur les méthodes de résolution utilisées. Les méthodes de résolution de courbes multivariées reposent en général sur l'hypothèse que le bruit affectant une mesure expérimentale est de type homoscédastique. Une étude approfondie de l'erreur présente dans des données de spectroscopie d'absorption transitoire montre que la réalité est tout autre. Nous avons ainsi pu montrer l'existence de deux facteurs contribuant à l'émergence de bruits structurés: un corrélé à la forme du signal et l'autre provenant de la source lumineuse utilisée par la mesure. Une adaptation de l'algorithme MCR-ALS a été développée pour prendre en compte ces bruits structurés lors de l'analyse multivariée. L'idée générale repose sur une pondération des données expérimentales en fonction d'une estimation de l'erreur de mesure. Plusieurs simulations de cas réalistes montrent l'intérêt de cette approche. Nous avons également pu valider les résultats des chapitres précédents obtenus sur des données moins affectés par le bruit.

Les différents apports méthodologiques développés au long de ce travail de thèse permettent d'obtenir des résolutions multivariées plus robustes et donc des informations plus fiables concernant les profils de concentration et les spectres purs d'espèces chimiques impliquées dans des procédés suivis par spectroscopie résolue en temps. La détection et l'identification d'espèces intermédiaires sont particulièrement intéressantes puisque grâce à elles nous pouvons accéder à une meilleure compréhension des mécanismes réactionnels.

Bibliographie

- [1] P.C. Jurs, B.R. Kowalski, T.L. Isenhour, *Anal. Chem.*, 41 (1969) 21.
- [2] P.C. Jurs, B.R. Kowalski, T.L. Isenhour, C.N. Reilley, *Anal. Chem.*, 41 (1969) 690.
- [3] B.R. Kowalski, P.C. Jurs, T.L. Isenhour, C.N. Reilley, *Anal. Chem.*, 41 (1969) 695.
- [4] W.H. Lawton, E.A. Sylvestre, *Technometrics*, 13 (1971) 617.
- [5] O.S. Borgen, B.R. Kowalski, *Anal. Chim. Acta*, 174 (1985) 1.
- [6] O.S. Borgen, N. Davidsen, M.Y. Zhu, O. Oyen, *Mikrochim. Acta*, 2 (1987) 63.
- [7] S. Bijlsma, D.J. Louwerse, A.K. Smilde, *J. Chemometrics*, 13 (1999) 311.
- [8] E. Comas, J. Ferré, F.X. Rius, *Anal. Chim. Acta*, 515 (2004) 23.
- [9] R. Bro, *Chemom. Intell. Lab. Syst.*, 38 (1997) 149.
- [10] R.A. Harshman, *UCLA Working Papers Phonet.*, 16 (1970) 1.
- [11] B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, D.L. Massart, *Handbook of Chemometrics and Qualimetrics : Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier, Amsterdam, 1998.
- [12] B.K. Alsberg, A.M. Woodward, M.K. Winson, J. Rowland, D.B. Kell, *Analyst*, 122 (1997) 645.
- [13] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics : Handbook of Chemometrics and Qualimetrics: Part A*, Elsevier, Amsterdam, 1997.
- [14] L.D. Rothman, S.R. Crouch, J.D.J. Ingle, *Anal. Chem.*, 47 (1975) 1226.
- [15] J. Galban, S. De Marcos, I. Sanz, C. Ubide, J. Zuriarrain, *Anal. Chem.*, 79 (2007) 4763.
- [16] P. Pernot, in: M. Mostafavi, T. Gustavsson (Eds.), *Réactions ultrarapides en solutions*, CNRS Editions, Paris, 2006.
- [17] D.T. Larose, *Data mining methods and models*, Wiley, Hoboken, 2006.
- [18] S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.*, 2 (1987) 37.
- [19] H. Martens, T. Naes, *Multivariate calibration*, John Wiley & sons, Chichester, 1989.
- [20] G.H. Golub, C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1996.
- [21] E.R. Malinowski, *Factor Analysis in Chemistry*, John Wiley & Sons Inc, New York, 2002.
- [22] P.R. Peres-Neto, D.A. Jackson, K.M. Somers, *Comput. Stat. Data Anal.*, 49 (2005) 974.
- [23] H. Shen, Y. Liang, O.M. Kvalheim, R. Manne, *Chemom. Intell. Lab. Syst.*, 51 (2000) 49.
- [24] C.J. Xu, Y. Liang, Y. Li, Y.P. Du, *Analyst*, 128 (2003) 75.
- [25] E.R. Malinowski, *J. Chemom.*, 18 (2004) 387.
- [26] K. Faber, B.R. Kowalski, *Anal. Chim. Acta*, 337 (1997) 57.
- [27] E. Levina, A.S. Wagaman, A.F. Callender, G.S. Mandair, M.D. Morris, *J. Chemom.*, 21 (2007) 24.
- [28] R.B. Cattell, *Multivar. Behav. Res.*, 1 (1966) 245.
- [29] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, *Chemom. Intell. Lab. Syst.*, 33 (1996) 17.

- [30] J. Saurina, S. Hernández-Cassou, R. Tauler, A. Izquierdo-Ridorsa, *J. Chemom.*, 12 (1998) 183.
- [31] C. Ruckebusch, A. de Juan, L. Duponchel, J.P. Huvenne, *Chemom. Intell. Lab. Syst.*, 80 (2006) 209.
- [32] H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbuehler, *Anal. Chim. Acta*, 193 (1987) 287.
- [33] M. Maeder, *Anal. Chem.*, 59 (1987) 527.
- [34] H.R. Keller, D.L. Massart, *Chemom. Intell. Lab. Syst.*, 12 (1992) 209.
- [35] H.R. Keller, D.L. Massart, Y.Z. Liang, O.M. Kvalheim, *Anal. Chim. Acta*, 263 (1992) 29.
- [36] H.R. Keller, D.L. Massart, *Chemom. Intell. Lab. Syst.*, 11 (1991) 175.
- [37] Z.D. Zeng, Y.Z. Liang, Y.L. Wang, X.R. Li, L.M. Liang, Q.S. Xu, C.X. Zhao, B.Y. Li, F.T. Chau, *J. Chromatogr., A*, 1107 (2006) 273.
- [38] Z.D. Zeng, C.J. Xu, Y.Z. Liang, B.Y. Li, *Chemom. Intell. Lab. Syst.*, 69 (2003) 89.
- [39] A.C. Whitson, M. Maeder, *J. Chemom.*, 15 (2001) 475.
- [40] A. de Juan, S. Navea, J. Diewok, R. Tauler, *Chemom. Intell. Lab. Syst.*, 70 (2004) 11.
- [41] J. Esteves da Silva, A.A.S.C. Machado, *Chemom. Intell. Lab. Syst.*, 27 (1995) 115.
- [42] J. Diewok, A. de Juan, M. Maeder, R. Tauler, B. Lendl, *Anal. Chem.*, 75 (2003) 641.
- [43] J. Diewok, A. De Juan, R. Tauler, B. Lendl, *Appl. Spectrosc.*, 56 (1992) 40.
- [44] S. Navea, A. de Juan, R. Tauler, *Anal. Chem.*, 74 (2002) 6031.
- [45] W. Windig, J. Guilment, *Anal. Chem.*, 63 (1991) 1425.
- [46] W. Windig, D.A. Stephenson, *Anal. Chem.*, 64 (1992) 2735.
- [47] F. Cuesta-Sanchez, B. Van den Bogaert, S.C. Rutan, D.L. Massart, *Chemom. Intell. Lab. Syst.*, 34 (1996) 139.
- [48] R. Tauler, *Chemom. Intell. Lab. Syst.*, 30 (1995) 133.
- [49] R. Tauler, A. Smilde, B. Kowalski, *J. Chemom.*, 9 (1995) 31.
- [50] E. Peré-Trepat, A. Hildebrandt, D. Barceló, S. Lacorte, R. Tauler, *Chemom. Intell. Lab. Syst.*, 74 (2004) 293.
- [51] S. Mas, G. Fonrodona, R. Tauler, J. Barbosa, *Talanta*, 71 (2007) 1455.
- [52] A. Alberich, J.M. Diaz-Cruz, C. Ariño, M. Esteban, *Analyst*, 133 (2007) 112.
- [53] M. Esteban, C. Ariño, J.M. Diaz-Cruz, M.S. Diaz-Cruz, R. Tauler, *TrAC*, 19 (2000) 49.
- [54] M. Garrido, F.X. Rius, M.S. Larrechi, *Anal. Bioanal. Chem.*, 390 (2008) 2059.
- [55] M. Garrido, I. Lázaro, M.S. Larrechi, F.X. Rius, *Anal. Chim. Acta*, 515 (2004) 65.
- [56] J. Diewok, A. De Juan, R. Tauler, B. Lendl, *Appl. Spectrosc.*, 56 (2002) 40.
- [57] M. Blanco, A.C. Peinado, J. Mas, *Anal. Chim. Acta*, 544 (2005) 199.
- [58] C. Ruckebusch, L. Duponchel, J.P. Huvenne, J. Saurina, *Anal. Chim. Acta*, 515 (2004) 183.
- [59] L. Blanchet, A. Mezzetti, C. Ruckebusch, J.P. Huvenne, A. de Juan, *Anal. Bioanal. Chem.*, 387 (2007) 1863.
- [60] B.K. Dable, B.J. Marquardt, K.S. Booksh, *Anal. Chim. Acta*, 544 (2005) 71.
- [61] B. Muik, B. Lendl, A. Molina-Diaz, M. Valcarcel, M.J. Ayora-Cañada, *Anal. Chim. Acta*, 593 (2007) 54.
- [62] G. Muñoz, A. de Juan, *Anal. Chim. Acta*, 595 (2007) 198.
- [63] S. Navea, A. de Juan, R. Tauler, *Anal. Chim. Acta*, 446 (2001) 187.
- [64] A. Argemi, J. Saurina, *J. Pharm. Biomed. Anal.*, 44 (2007) 859.
- [65] M. Bosco, M.P. Callao, M.S. Larrechi, *Talanta*, 72 (2007) 800.
- [66] J. Jaumot, V. Marchán, R. Gargallo, A. Grandas, R. Tauler, *Anal. Chem.*, 76 (2004) 7094.
- [67] E. Peré-Trepat, S. Lacorte, R. Tauler, *J. Chromatogr., A*, 1096 (2006) 111.

- [68] M. Maeder, A.D. Zuberbühler, *Anal. Chem.*, 62 (1990) 2220.
- [69] W. Windig, *Chemom. Intell. Lab. Syst.*, 16 (1992) 1.
- [70] C.L. Lawson, R.J. Hanson, *Solving least squares problems*, Prentice-Hall, Englewood Cliffs, 1974.
- [71] R. Bro, S. De Jong, *J. Chemom.*, 11 (1997) 393.
- [72] M.H. Van Benthem, M.R. Keenan, *J. Chemom.*, 18 (2004) 441.
- [73] R. Bro, N.D. Sidiropoulos, *J. Chemom.*, 12 (1998) 223.
- [74] P. Gemperline, E. Cash, *Anal. Chem.*, 75 (2003) 4236.
- [75] M.H. Van Benthem, M.R. Keenan, D.M. Haaland, *J. Chemom.*, 16 (2002) 613.
- [76] R. Manne, *Chemom. Intell. Lab. Syst.*, 27 (1995) 89.
- [77] R.C. Henry, B.M. Kim, *Chemom. Intell. Lab. Syst.*, 8 (1990) 205.
- [78] B.M. Kim, R.C. Henry, *Chemom. Intell. Lab. Syst.*, 49 (1999) 67.
- [79] P.D. Wentzell, J.H. Wang, L.F. Loucks, K.M. Miller, *Can. J. Chem.*, 76 (1998) 1144.
- [80] P.J. Gemperline, *Anal. Chem.*, 71 (1999) 5398.
- [81] R. Tauler, *J. Chemom.*, 15 (2001) 627.
- [82] M. Garrido, M.S. Larrechi, F.X. Rius, R. Tauler, *Chemom. Intell. Lab. Syst.*, 76 (2005) 111.
- [83] S. Norman, M. Maeder, *Crit. Rev. Anal. Chem.*, 36 (2006) 199.
- [84] I.H.M. Van Stokkum, D. Larsen, S., R. Van Grondelle, *Biochim. Biophys. Acta*, 1657 (2004) 82.
- [85] H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbühler, *Talanta*, 32 (1985) 95.
- [86] H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbühler, *Talanta*, 32 (1985) 251.
- [87] K.J. Molloy, M. Maeder, M.M. Schumacher, *Chemom. Intell. Lab. Syst.*, 46 (1999) 221.
- [88] E. Furusjö, L.-G. Danielsson, *Chemom. Intell. Lab. Syst.*, 50 (2000) 63.
- [89] K.M. Mullen, M. Vengris, I.H.M. Van Stokkum, *J. Glob. Optim.*, 38 (2007) 201.
- [90] P. Bugnon, J.-C. Chottard, J.-L. Jestin, B. Jung, G. Laurency, M. Maeder, A.E. Merbach, A.D. Zuberbühler, *Anal. Chim. Acta*, 298 (1994) 193.
- [91] I.H.M. Van Stokkum, T. Scherer, A.M. Brouwer, J.W. Verhoeven, *J. Phys. Chem.*, 98 (1994) 852.
- [92] I.H.M. Van Stokkum, H. Lindsell, J.M. Hadden, P.I. Haris, C. D., M. Bloemendal, *Biochemistry*, 34 (1995) 10508.
- [93] M. Maeder, Y.M. Neuhold, *Practical data analysis in chemistry*, Elsevier, Amsterdam, 2007.
- [94] M. Maeder, Y.M. Neuhold, G. Puxty, P.J. Gemperline, *Chemom. Intell. Lab. Syst.*, 82 (2006) 75.
- [95] D.W. Marquardt, *SIAM J. Appl. Math.*, 11 (1963) 431.
- [96] A. de Juan, M. Maeder, M. Martinez, R. Tauler, *Chemom. Intell. Lab. Syst.*, 54 (2000) 123.
- [97] J.M. Amigo, A. de Juan, J. Coello, S. MasPOCH, *Anal. Chim. Acta*, 567 (2006) 236.
- [98] J.M. Amigo, A. de Juan, J. Coello, S. MasPOCH, *Anal. Chim. Acta*, 567 (2006) 245.
- [99] J. Jaumot, P.J. Gemperline, A. Stang, *J. Chemom.*, 19 (2005) 97.
- [100] S. Mas, A. De Juan, S. Lacorte, R. Tauler, *Anal. Chim. Acta*, 618 (2008) 18.
- [101] C.J. Xu, S. GourvéneC, Y.Z. Liang, D.L. Massart, *Anal. Chim. Acta*, 575 (2006) 1.
- [102] L. Blanchet, C. Ruckebusch, J.P. Huvenne, A. de Juan, *Chemom. Intell. Lab. Syst.*, 89 (2007) 26.
- [103] L. Blanchet, C. Ruckebusch, J.P. Huvenne, A. De Juan, *J. Chemom.* (In press).
- [104] A. de Juan, M. Maeder, M. Martinez, R. Tauler, *Anal. Chim. Acta*, 442 (2001) 337.
- [105] T. Förster, *Ann. Pys.*, 437 (1948) 55.

- [106] B. Ke, *Photosynthesis: photobiochemistry and photobiophysics*, Kluwer academic publishers, Dordrecht, 2001.
- [107] J. Farineau, J.-F. Morot-Gaudry, *La photosynthèse. Processus physiques, moléculaires et physiologiques*, INRA, Paris, 2006.
- [108] R. Emerson, W. Arnold, *J. Gen. Physiol.*, 15 (1932) 391.
- [109] A. Brack, P. Mathis, *La chimie du vivant - De la protéine à la photosynthèse*, Le pommier, Paris, 2000.
- [110] J. Deisenhofer, E. U., K. Miki, R. Huber, H. Michel, *Nature*, 318 (1985) 618.
- [111] J. Deisenhofer, H. Michel, *Biosci. Rep.*, 24 (2004) 323.
- [112] C.R.D. Lancaster, H. Michel, *Photosynth. Res.*, 48 (1996) 65.
- [113] D. Kleinfeld, M.Y. Okamura, G. Feher, *Biochim. Biophys. Acta*, 766 (1984) 126.
- [114] Z. Zhu, M.r. Gunner, *Biochemistry*, 44 (2005) 82.
- [115] A. Mezzetti, E. Nabedryk, J. Breton, M. Okamura, M. Paddock, G. Giacometti, W. Leibl, *Biochim. Biophys. Acta - Bioenergetics*, 1553 (2002) 320.
- [116] M. Paddock, G. Feher, M. Okamura, *FEBS Lett.*, 555 (2003) 45.
- [117] M. Okamura, M. Paddock, M.S. Graige, *Biochim. Biophys. Acta*, 1458 (2000) 148.
- [118] A. Mezzetti, W. Leibl, *Eur. Biophys. J.*, 34 (2005) 921.
- [119] A. Mezzetti, W. Leibl, J. Breton, E. Nabedryk, *FEBS Lett.*, 537 (2003) 161.
- [120] R.K. Clayton, R.T. Wang, *Methods Enzymol.*, 23 (1971) 696.
- [121] J. Breton, E. Nabedryk, *Biochim. Biophys. Acta - Bioenergetics*, 1275 (1996) 84.
- [122] A. Barth, *Prog. Biophys. Mol. Biol.*, 42 (2000) 141.
- [123] M. Iwaki, A. Puustinen, M. Wilkström, P.R. Rich, *Biochemistry*, 45 (2006) 10873.
- [124] F. Comayras, C. Jungas, J. Lavergne, *J. Biol. Chem.*, 280 (2005) 11203.
- [125] F. Comayras, C. Jungas, J. Lavergne, *J. Biol. Chem.*, 280 (2005) 11214.
- [126] L. Blanchet, C. Ruckebusch, A. Mezzetti, J.P. Huvenne, A. De Juan, (Submitted to *J. Phys. Chem. B*).
- [127] P. Suppan, *Chemistry and Light*, Royal Society of Chemistry Cambridge, 1994.
- [128] N.J. Turro, *Modern molecular photochemistry*, University Science Books, Sausalito, 1991.
- [129] J. Franck, *Trans. Far. Soc.*, 21 (1926) 536.
- [130] E. Condon, *Phys. Rev.*, 28 (1926) 1182.
- [131] M.A. El Sayed, *Acc. Chem. Res.*, 1 (1968) 8.
- [132] R.M. Hochstrasser, H. Lutz, G.W. Scott, *Chem. Phys. Lett.*, 24 (1974) 162.
- [133] R.W.J. Anderson, R.M. Hochstrasser, H. Lutz, G.W. Scott, *J. Chem. Phys.*, 61 (1974) 2500.
- [134] D.E. Damschen, C.D. Merritt, D.L. Perry, G.W. Scott, L.D. Talley, *J. Phys. Chem.*, 82 (1978) 2268.
- [135] B.I. Greene, R.M. Hochstrasser, R.B. Weisman, *J. Chem. Phys.*, 70 (1979) 1247.
- [136] P. McGarry, F., C.E. Doubleday, C.-H. Wu, H.A. Staab, N.J. Turro, *J. Photochem. Photobiol. A*, 77 (1994) 109.
- [137] K. Prater, W.L. Freund, R.M. Bowman, *Chem. Phys. Lett.*, 295 (1998) 82.
- [138] B.K. Shah, M.A.J. Rodgers, D.C. Neckers, *J. Phys. Chem. A*, 108 (2004) 6087.
- [139] D.R. Kimbrough, *J. Chem. Educ.*, 74 (1997) 51.
- [140] M.S. Churio, M.A. Grela, *J. Chem. Educ.*, 74 (1997) 436.
- [141] S. Yabumoto, S. Sato, H.-o. Hamaguchi, *Chem. Phys. Lett.*, 416 (2005) 100.
- [142] A.C. Bhasikuttan, A.K. Singh, D.K. Palit, A.V. Sapre, J.P. Mittal, *J. Phys. Chem. A*, 102 (1998) 3470.
- [143] A.K. Singh, A.C. Bhasikuttan, D.K. Palit, J.P. Mittal, *J. Phys. Chem. A*, 104 (2000) 7002.

- [144] F. Boscá, g. Cosa, M.A. Miranda, J.C. Scaiano, *Photochem. Photobiol. Sci.*, 1 (2002) 704.
- [145] M. Ramseier, P. Senn, J. Wirz, *J. Phys. Chem. A*, 107 (2003) 3305.
- [146] C. Ruckebusch, S. Aloïse, L. Blanchet, J.P. Huvenne, G. Buntinx, *Chemom. Intell. Lab. Syst.*, 91 (2008) 17.
- [147] S. Aloïse, C. Ruckebusch, L. Blanchet, J. Rehault, G. Buntinx, J.P. Huvenne, *J. Phys. Chem. A*, 112 (2008) 224.
- [148] M.N. Leger, L. Vega-Montoto, P.D. Wentzell, *Chemom. Intell. Lab. Syst.*, 77 (2005) 181.
- [149] P.D. Wentzell, T.K. Karakach, S. Roy, M.J. Martinez, C.P. Allen, M. Werner-Washburne, *BMC Bioinformatics*, 7 (2006) 343.
- [150] P.D. Wentzell, *Comprehensive Chemometrics*, Elsevier, Amsterdam, In press.

Síntesis: Análisis multivariante de espectros resueltos en el tiempo

La quimiometría es la ciencia que relaciona las medidas realizadas sobre un sistema químico con sus estados a través de la aplicación de la matemática o los métodos estadísticos. Esta definición de la Sociedad Internacional de Quimiometría es muy amplia, pero corresponde a la gran cantidad de aplicaciones que tiene esta disciplina. En esta tesis, la parte de la quimiometría que nos interesa es la que proporciona herramientas matemáticas que permiten separar la información relativa a las distintas especies químicas contenidas en una mezcla. Este tipo de problema es sólo una rama de la quimiometría, que suele llamarse "análisis de mezclas". Los métodos correspondientes se designan con el término "resolución de curvas multivariantes", y pertenecen al grupo de métodos de análisis de factores. En el problema de análisis de mezclas, es necesario distinguir dos casos, el primero, en el que se conoce la identidad de los productos implicados en la mezcla, y el segundo, en que esta información no está disponible. El último caso es, obviamente, más complejo pero también más frecuente en investigación de procesos. Presentaremos cómo la quimiometría combinada con técnicas instrumentales espectroscópicas avanzadas da acceso a una información exhaustiva sobre las especies químicas desconocidas que se encuentran en un sistema en evolución..

La espectroscopia permite definir la estructura atómica o molecular de un compuesto químico mediante el análisis de su espectro. El análisis de los datos espectroscópicos implica a menudo la utilización de métodos multivariantes. Será necesario utilizar herramientas quimiométricas basadas en una ley lineal, ya que ésta es la naturaleza de la medida espectroscópica, tal como la describe la ley de Lambert-Beer. La aplicación de estas herramientas en el análisis de procesos da una descripción cualitativa y cuantitativa de los fenómenos que ocurren en el sistema estudiado. Esta descripción es posible gracias a la recuperación de los perfiles de concentración y espectros puros de cada una de las especies involucradas en el proceso. Nos interesan en particular los procesos seguidos con espectroscopia resuelta en el tiempo. El objetivo es obtener información sobre la secuencia de acontecimientos correspondiente a una cinética de reacción y sobre los posibles intermediarios. Es fácil comprender que para seguir un fenómeno a lo largo del tiempo, es necesario acumular las medidas en las diferentes etapas del proceso. La cantidad de datos a ser tratada crece rápidamente. Además, algunos tipos de instrumentación avanzada, tienen particularidades muy específicas, que llevan a dificultades que no se han contemplado con anterioridad. Esta tesis, realizada en colaboración entre la *Université des Sciences et*

Technologies de Lille (USTL) y la *Universitat de Barcelona* (UB), entra en esta línea de investigación.

El objetivo principal de este trabajo es participar en el desarrollo de los actuales métodos de análisis de datos con el fin de mejorar la interpretación de los procesos seguidos por espectroscopia resuelta en el tiempo. Este trabajo se realiza, por lo tanto, en estrecha colaboración con los científicos usuarios de este tipo de técnicas espectroscópicas. Los puntos clave de la investigación llevada a cabo en esta tesis doctoral se definen a continuación.

El primer punto está relacionado con la descripción de los métodos más apropiados para el análisis de medidas espectroscópicas resueltas en el tiempo y en valorar sus ventajas e inconvenientes. Más concretamente, nos centraremos en la descripción de tres tipos de métodos. En primer lugar, se presentarán métodos de modelado suave (*soft-modeling*), que se basan solamente en un modelo bilineal de la medida espectroscópica y en las restricciones generales asociadas a la instrumentación utilizada o al sistema químico. En segundo lugar, se describirán los métodos de modelado duro (*hard-modeling*) que describen la variación de las medidas únicamente un modelo fisicoquímico. Por último, se propondrá un método híbrido que combina las ventajas de los dos tipos de algoritmos anteriores.

El segundo punto está relacionado con la influencia de la estructura de los datos de diferencia sobre la resolución. El seguimiento de procesos químicos en escalas de tiempo cortas puede requerir el uso de sistemas de detección diferencial, que proporcionan espectros que permiten detectar variaciones débiles de la señal. La estructura de datos de diferencia facilita información interesante, pero se debe analizar e interpretar adecuadamente. Se ha realizado un análisis teórico para evaluar las particularidades de este tipo de datos e indicar cuál es la forma correcta de interpretar la información obtenida. Más concretamente, se ha estudiado la pérdida de información sobre el estado inicial del sistema y sus consecuencias sobre el número de contribuciones detectables y sobre su significado físico.

A partir del estudio teórico de la naturaleza e interpretación de los datos espectroscópicos de diferencia, se presentan ejemplos reales de la gestión de este tipo de datos para la obtención de estimaciones fiables de perfiles de concentración y de espectros puros asociados a procesos. En este sentido, la incorporación de la restricción cinética en los métodos de modelado híbrido considerando la especie(s) inicial(es) como no absorbente

permitirá superar la limitación asociada a la pérdida de información explícita sobre el estado inicial del proceso en estudio. Se prestará una atención especial a los procesos fotoquímicos que tienen lugar en fases sucesivas de excitación y relajación. Eso significa que los dos modelos cinéticos involucrados deben estar conectados el uno al otro. Se propone un ajuste cinético secuencial de ambas etapas del proceso para solucionar este problema. Así, los resultados del modelo de la fase de excitación, es decir, las concentraciones al final de esta fase, proporcionan las condiciones iniciales de la fase de relajación (desconocidas en el inicio del análisis).

La siguiente cuestión tratada en esta tesis se refiere a la caracterización de la estructura de ruido de las medidas espectroscópicas resueltas en el tiempo. Los datos experimentales están afectados por diversas fuentes de error. Si bien no es posible conocer exactamente el error asociado a una medida individual, puede ser interesante evaluar, de manera general, la estructura del error asociado a una instrumentación específica. La espectroscopia resuelta en tiempo debe ser estudiada en este marco, más concretamente cuando se sospecha la existencia de errores correlacionados. Es, pues, necesario un nuevo desarrollo metodológico para tener en cuenta las posibles fuentes de error y su estructura. Dependiendo de la naturaleza del ruido, será necesaria una adaptación de las herramientas quimiométricas utilizadas. Los resultados obtenidos en el estudio del ruido en espectroscopia de absorción transitoria nos permitirán construir un sistema de ponderación basado en la estructura de ruido, que dé más importancia a las medidas más fiables y que incluya la información de la correlación del ruido entre canales espectrales o en la dirección temporal. De esta forma, los resultados del modelado suave serán mejorados tanto por lo que respecta a la calidad de la recuperación de los perfiles de concentración y espectros asociados como a la minimización del riesgo de sobreajuste. En esta tesis se presenta el primer estudio sobre el impacto de la ponderación incluyendo ruido estructurado en análisis de procesos.

Los puntos anteriores ayudan a una mejor comprensión de las especificidades de la espectroscopia resuelta en el tiempo y permiten mejorar el análisis de los datos. El objetivo de este capítulo es resumir los resultados de esta tesis en castellano. Las publicaciones escritas durante esta tesis serán la columna vertebral de esta sección. Por lo tanto, cada apartado (excepto el primero) incluirá una lista de las publicaciones relacionadas, ya publicadas o enviadas a revistas.

Los métodos quimiométricos que se han utilizado en esta tesis se presentarán brevemente en una primera parte. Es la ocasión para discutir sobre los algoritmos más utilizados en análisis de datos espectroscópicos y sobre sus ventajas e inconvenientes. Estas nociones son necesarias para la comprensión de la segunda sección, que está relacionada con el análisis del problema asociado a las medidas en diferencia de los espectros resueltos en el tiempo. Esta característica tiene una influencia en el rango de los datos. Proponemos un estudio de este problema y algunas soluciones. Esta sección es, por lo tanto, uno de los pilares metodológicos de este trabajo y responde a una parte de las preguntas surgidas en el análisis de datos de espectros resultantes de la instrumentación del *Laboratoire de Spectrochimie Infrarouge et Raman* (LASIR).

Las secciones 3 y 4 corresponden a la aplicación de las estrategias metodológicas propuestas en la sección 2 a datos reales. La sección 3 se refiere al estudio del centro de reacción de la bacteria *Rhodobacter sphaeroides*. Este ejemplo es particularmente pertinente para ilustrar las dificultades inherentes a la espectroscopia resuelta en el tiempo y cómo la metodología propuesta en la segunda sección permite resolverlos. Es importante señalar que una especie interferente (que no es objeto de una cinética de reacción) está involucrada en este proceso. La sección 4 trata del estudio de la fotorelajación de la benzofenona. Es una ilustración del interés de los métodos de modelado (sea o no basado en un modelo cinético) para el estudio de la reactividad por espectroscopia de absorción ultrarrápida. El ejemplo propuesto es la relajación de la benzofenona después de una fotoexcitación. En este caso, la detección de especies intermedias de vida muy corta que pueden tener una intensa superposición espectral con el resto de especies es un punto de especial dificultad. Proponemos la validación de la existencia y caracterización de estas especies transitorias a partir de estrategias basadas en herramientas quimiométricas de diagnóstico y en experimentos complementarios de fotoquímica.

Por último, la segunda aportación metodológica, de especial relevancia, se propone en la sección 5. Se trata de la gestión de la estructura de ruido que afecta la espectroscopia resuelta en el tiempo. Se mostrará cómo un análisis riguroso de la estructura del error de medida puede ser usado para construir un sistema de ponderación, que se aplicará para mejorar un análisis *soft modelling*, ya que toma en cuenta la información sobre el ruido presente en los datos.

Análisis de datos multivariantes

La quimiometría ha sido desarrollada para mejorar el análisis de datos químicos complejos. Las medidas espectroscópicas obtenidas en el seguimiento de un proceso no pueden ser interpretadas visualmente y necesitan la aplicación de técnicas quimiométricas. Dentro de las herramientas quimiométricas utilizadas para el análisis de procesos, se puede distinguir entre las que utilizan explícitamente la información contenida en un modelo fisicoquímico (cinético y/o de equilibrio), conocidas como métodos de modelado duro (*hard-modeling*) y las que no asumen ningún tipo de mecanismo o ley de comportamiento, denominadas métodos de modelado suave (*soft-modeling*). Ambos tipos de métodos son capaces de proponer una descomposición del conjunto de datos **D** en un modelo bilineal de dos matrices **C** y **S**, que contienen, respectivamente, los perfiles de concentración y los espectros puros de las especies involucradas en el proceso, como se muestra en la ecuación (73).

$$\mathbf{D} = \mathbf{CS}^T + \mathbf{E} \quad (73)$$

Para finalizar esta sección también se hablará de metodologías híbridas más recientes, que combinan las ventajas de los métodos clásicos de *hard-* y *soft-modeling*.

Multivariate curve Resolution – Alternating Least Squares (MCR-ALS)

Los resultados matemáticos obtenidos mediante análisis de componentes principales (PCA) no reflejan la evolución real de las contribuciones químicas en el proceso estudiado. En el contexto de los métodos de modelado suave, los métodos de resolución multivariante son una mejor alternativa ya que introducen cierta información en el análisis del proceso, que permite obtener soluciones interpretables desde el punto de vista químico. La información que se incluye puede ser, de hecho, de tipo químico. Por ejemplo, es evidente que una concentración no puede ser negativa. Esta restricción debe incluirse en el proceso de resolución para prohibir los perfiles de concentración sin sentido físico [8]. Otros ejemplos que se pueden encontrar son la restricción de la unimodalidad [9] o la condición de sistema cerrado (*closure*). Por lo general estas condiciones son designadas como "restricciones suaves" (*soft-modelling*). Esta es la estrategia utilizada por el algoritmo de Resolución Multivariante de Curvas por Mínimos Cuadrados Alternados, MCR-ALS (*Multivariate Curve Resolution – Alternating Least Squares*), que obtiene los perfiles de concentración y los

espectros puros de los componentes de un proceso a partir de una optimización iterativa bajo restricciones [10]. Sus principales etapas se resumen a continuación:

1. Determinación del número de componentes en el conjunto de datos bilineales \mathbf{D} ;
2. Generación de las estimaciones iniciales de los perfiles de concentración, \mathbf{C} , o de los espectros, \mathbf{S}^T ;
3. Tomando \mathbf{D} y las estimaciones iniciales de \mathbf{C} , cálculo por mínimos cuadrados de los espectros, \mathbf{S}^T , bajo las restricciones elegidas;
4. Tomando \mathbf{D} y la matriz calculada \mathbf{S}^T , cálculo por mínimos cuadrados de los perfiles de concentración, \mathbf{C} , bajo las restricciones elegidas;
5. Cálculo de la reconstrucción de los datos, evaluación de la falta de ajuste (LOF) y regreso al paso 3 hasta que se logre la convergencia (es decir, que la diferencia entre el LOF de dos iteraciones consecutivas sea menor que un valor umbral o que se alcance un número de iteraciones predeterminado).

Los valores de falta de ajuste (LOF) y de varianza explicada (R^2) se utilizan para determinar la calidad de la resolución y se calculan como en las ecuaciones (74) y (75):

$$LOF = \sqrt{\frac{\sum_i \sum_j (d_{i,j} - \hat{d}_{i,j})^2}{\sum_i \sum_j d_{i,j}^2}} \quad (74)$$

$$R^2 = \frac{\sum_i \sum_j \hat{d}_{i,j}^2}{\sum_i \sum_j d_{i,j}^2} \quad (75)$$

$d_{i,j}$ y $\hat{d}_{i,j}$ representan los elementos de \mathbf{D} y de la matriz reconstruida a partir del modelo MCR-ALS, respectivamente.

Los perfiles obtenidos, en general, no son únicos, incluso después de una resolución bajo restricciones, se ven afectados por las ambigüedades de rotación [10]. En realidad, se obtienen un conjunto de soluciones posibles que ajustan los datos con la misma calidad. Existen diversos métodos para estimar el ancho de las bandas de soluciones posibles [11-15]. La estrategia del aumento de matrices permite reducir la influencia de estas ambigüedades [16]. Otra posibilidad es la aplicación de las restricciones de modelo cinético [17]. Esta estrategia permite fusionar los beneficios de los métodos de *soft-modelling* y de *hard-modelling*.

Hard modelling de datos cinéticos

La información disponible sobre un sistema puede incluir algunas características específicas tales como el seguimiento de un modelo químico. Diversos métodos [18, 19] han sido propuestos para la descripción de la variación observada en un conjunto de datos a partir del ajuste de un modelo fisicoquímico, con el consiguiente cálculo de constantes de velocidad o de equilibrio [20-25]. Por lo general estos métodos son denominados de modelado duro (*hard-modelling*). Estos métodos también suponen que los datos siguen un modelo bilineal y pueden descomponerse en dos matrices **C** y **S**. Nos centraremos en el método de los mínimos cuadrados no lineales, que es el que se utilizará en esta tesis. La idea principal de este método es ajustar la variación de los datos experimentales mediante la optimización de los parámetros que controlan la cinética del proceso (es decir, las constantes cinéticas). Una vez estos parámetros son conocidos, es sencillo reconstruir la matriz de perfiles de concentración **C** a partir del modelo propuesto y, a continuación, obtener la matriz **S** utilizando el modelo bilineal de la ecuación (1). La optimización de las constantes cinéticas se hace minimizando los residuales de la ecuación (76).

$$\min \left\| \mathbf{D} \left(\mathbf{I} - \mathbf{C}(\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \right) \right\| \quad (76)$$

Como se ve en esta ecuación, la matriz de residuos sólo depende de la matriz **C** y, en consecuencia, de los parámetros no lineales (constantes de velocidad o de equilibrio) que la definen [26, 27]. La minimización de los residuos se hace usando el algoritmo Levenberg-Marquardt [28]. El algoritmo adapta los parámetros del modelo cinético postulado con el fin de maximizar la varianza explicada. Sin embargo, si todas las fuentes de variación de los datos originales no están directamente relacionadas con el modelo propuesto (por ejemplo, un cambio de línea de base o la presencia de un interferente absorbente) los resultados serán erróneos. Además, es interesante observar que la información espectral no se utiliza nunca explícitamente, de hecho, la optimización sólo funciona con **D** y la cinética de los perfiles **C** resultantes del modelo. Esto explica que cuando se trata de una matriz aumentada multiexperimento, los diversos experimentos deban seguir el mismo modelo global.

Modelado híbrido

Los métodos de modelado suave y duro presentados anteriormente se basan en dos ideas opuestas, una es no hacer hipótesis sobre el modelo de comportamiento de los datos y la otra se basa completamente en el conocimiento y uso activo de tal modelo. Sin embargo, los dos

conceptos no son incompatibles y han dado lugar al método híbrido HS-MCR (Hard- Soft-Multivariate Curve Resolution) [17]. Su estructura básica es la del algoritmo MCR-ALS pero introduce la información de un modelo cinético en el cálculo de la matriz \mathbf{C} bajo restricciones. Así, los perfiles de concentración de la matriz \mathbf{C} (todos o algunos) son ajustados según un modelo cinético mediante el algoritmo Levenberg-Marquardt. La minimización de los residuales se calcula sólo con los perfiles de concentración como se muestra en la ecuación (77).

$$\min \|\mathbf{C}_{\text{ALS}} - \mathbf{C}_{\text{kin}}\| \quad (77)$$

Aquí la matriz \mathbf{C}_{ALS} contiene los perfiles de concentración de los compuestos involucrados en el proceso cinético calculados por *soft-modeling* y la matriz \mathbf{C}_{kin} contiene los perfiles restringidos, que se han obtenido a partir del modelo cinético ajustado. La matriz de perfiles de concentración obtenida (con los perfiles restringidos de acuerdo al modelo y los que se han modelado con restricciones suaves) se utiliza después para calcular \mathbf{S} . Las soluciones obtenidas son iterativamente reforzadas por las restricciones paramétricas y no paramétricas.

Diversas ventajas se obtienen utilizando esta estrategia híbrida, entre ellas podemos citar el hecho de que las soluciones obtenidas están mucho menos afectadas por las ambigüedades de rotación. Por otra parte, la aplicación de un modelo cinético permite considerar algunas especies como "transparentes", es decir, sin una señal espectroscópica. En comparación con los métodos clásicos de modelado duro podemos decir que la información espectral se utiliza explícitamente y es posible dejar fuera del modelo algunas especies o interferentes que no sigan el modelo elegido. Además, cuando se trata de un conjunto de datos multiexperimento, el algoritmo híbrido permite aplicar un modelo global común, modelos específicos para una o más experiencias y también permite el análisis de experimentos que siguen un modelo predefinido y de otros que son modelados con el uso único de restricciones suaves.

Conclusión

Los diferentes métodos presentados en esta sección son utilizados y desarrollados en el ámbito del análisis de datos espectroscópicos procedentes de procesos. En próximas secciones veremos cómo las especificidades del algoritmo híbrido se convierten en una ventaja fundamental en el análisis de los espectros de diferencia.

Modelado de datos de espectroscopia de diferencia

Los métodos de resolución multivariante de curvas han sido desarrollados en diversos campos. Cada uno implica algunas adaptaciones de los algoritmos para tener en cuenta algunas características estructurales o instrumentales de la medida objeto de análisis. Las espectroscopias utilizadas en el LASIR (*Laboratoire de Spectrochimie Infrarouge y Raman*) son en general muy específicas y sólo suelen ser utilizadas en aplicaciones de investigación. Por lo tanto, el objetivo de este trabajo es desarrollar herramientas que se adapten a este tipo particular de instrumentación. El primer proyecto desarrollado está relacionado con el uso de los datos de espectros de diferencia. Se demostrará que la interpretación de estos datos debe hacerse con precaución y no es tan sencilla como para datos de espectroscopia directa. La estructura inherente de las medidas de diferencia induce una pérdida de información que tiene consecuencias sobre los datos adquiridos. Veremos en primer lugar que el principal problema proviene de una deficiencia de rango asociada a la naturaleza de esta medida y se propondrán medidas para minimizar el efecto de este hecho en el análisis e interpretación de los datos.

Rango de datos espectroscópicos de diferencia

La determinación del rango es un aspecto relevante del análisis de los datos. El término "rango" se asocia generalmente con el número de especies químicas que intervienen en el proceso en estudio, pero no siempre es así. Un trabajo teórico completo sobre este tema para la estimación del rango de datos de espectroscopia directa en procesos fue llevado a cabo por Amrhein et al. [29]. En este estudio se concluyó que el rango de un conjunto de datos asociado a un proceso es igual a:

$$\text{Rango}(\mathbf{D}) = \min(\text{número de reacciones} + 1, \text{número de especies absorbentes}) \quad (78)$$

Sin embargo, el caso de los datos de diferencia no ha sido considerado explícitamente. La intuición sugiere que la substracción del primer espectro (o de un espectro de referencia) puede llevar a una pérdida de la información y así afectar el rango.

Se han estudiado diversos casos a fin de determinar la influencia de la estructura diferencial sobre la determinación del rango [30]. Varias situaciones son posibles, pero se pueden reducir fundamentalmente a dos. La primera de ellas se relaciona con un sistema químico (proceso) que no induce deficiencia de rango en espectroscopia directa, la segunda a

un sistema químico ya afectado por una (o muchas) deficiencia(s) de rango debidas a la naturaleza del proceso. Más detalles pueden ser encontrados en el artículo relacionado (Publicación I). Brevemente, nuestras conclusiones fueron que la ecuación de Armhein [29] debe ser adaptada para dar una estimación correcta del rango de una matriz de datos de diferencia $\Delta\mathbf{D}$:

$$\text{Rango}(\Delta\mathbf{D}) = \min(\text{numero de reacciones, numero de especies absorbentes} - 1) \quad (79)$$

El resultado es coherente con nuestra intuición inicial, pero tuvo que ser demostrado matemáticamente. Hemos aclarado que los datos de diferencia $\Delta\mathbf{D}$ pueden ser descritos por la ecuación (80), en el caso de un proceso simple que no induce deficiencia de rango del tipo $A \rightarrow B \rightarrow C \rightarrow \dots \rightarrow N$.

$$\Delta\mathbf{D} = \begin{bmatrix} \mathbf{c}_B - \mathbf{c}_{B,0} & \mathbf{c}_C - \mathbf{c}_{C,0} & \dots & \mathbf{c}_N - \mathbf{c}_{N,0} \end{bmatrix} \begin{bmatrix} \mathbf{s}_B - \mathbf{s}_A \\ \mathbf{s}_C - \mathbf{s}_A \\ \vdots \\ \mathbf{s}_N - \mathbf{s}_A \end{bmatrix} \quad (80)$$

Es importante notar que las contribuciones modeladas tienen perfiles de concentración iguales en forma a los obtenidos por espectroscopia directa (excepto los referentes a especies iniciales, que no se modelan) y que los espectros puros no deben interpretarse como espectros de especies individuales sino como espectros de diferencia de cada componente respecto a la especie inicial. Como se ha visto, la estructura de los datos de diferencia siempre induce a una deficiencia de rango, es decir, el rango es una unidad menor de lo que sería en espectroscopia directa. A continuación, se propone una estrategia que permite compensar la pérdida de información inicial a través de la aplicación de una restricción de modelo duro.

La elaboración de un modelo apropiado para la descripción de un sistema químico debe tener en cuenta cada una de las especies involucradas. El uso de datos de diferencia lleva de alguna manera a perder la información relativa a las especies presentes en el comienzo del experimento. En el marco de espectroscopias resueltas en el tiempo, esta dificultad puede ser compensada mediante el uso de modelos cinéticos. Para compensar la ausencia de señal instrumental asociada a las especies iniciales, éstas serán incluidas en el modelo como especie(s) transparente(s). Es necesario especificar que esta estrategia no rompe la deficiencia de rango, pero permite modelar la cinética del proceso (las contribuciones medibles) con un

modelo coherente con la realidad de los datos experimentales. De este modo, aunque el espectro puro de las especies iniciales no podrá ser conocido, los parámetros del modelo (constantes cinéticas) se recuperarán correctamente

Modelado híbrido de datos de diferencia

Las herramientas de modelado duro se adaptan bien a la problemática de incluir especies transparentes en el modelo cinético, pero hemos preferido aplicar el algoritmo híbrido HS-MCR por las ventajas adicionales que proporciona (ver publicación II) [31]. Brevemente, la introducción de un modelo cinético permite considerar la presencia de una especie inicial transparente en el comienzo de la reacción. Sin embargo, la eventual presencia de cualquier interferente (que no sigue una ley cinética clara) puede causar resultados incorrectos cuando se usan herramientas clásicas de modelado duro.

En la Figura 98 se muestra, para un mismo proceso, una comparación entre los perfiles de concentraciones y los espectros resultantes de: a) los análisis de datos de espectroscopia directa y b) de los datos de diferencia que se derivan a partir de las medidas directas por substracción del espectro inicial. Ambos conjuntos de datos siguen el mismo proceso y, por tanto, siguen la misma cinética. HS-MCR permite reconstruir correctamente los perfiles de concentración en ambos casos gracias al modelo cinético (cuando la espectroscopia es directa, todas las especies son absorbentes y cuando se tratan los datos de diferencia, la especie inicial se define como transparente). Los perfiles de concentración obtenidos son similares, demostrando que la aplicación del modelo cinético conduce a la determinación correcta de los perfiles de concentración, incluso en ausencia de una parte de la información espectroscópica. Sin embargo, en el caso de los datos de diferencia, los espectros puros son espectros de diferencia de cada especie respecto a la inicial y deben interpretarse como tales.

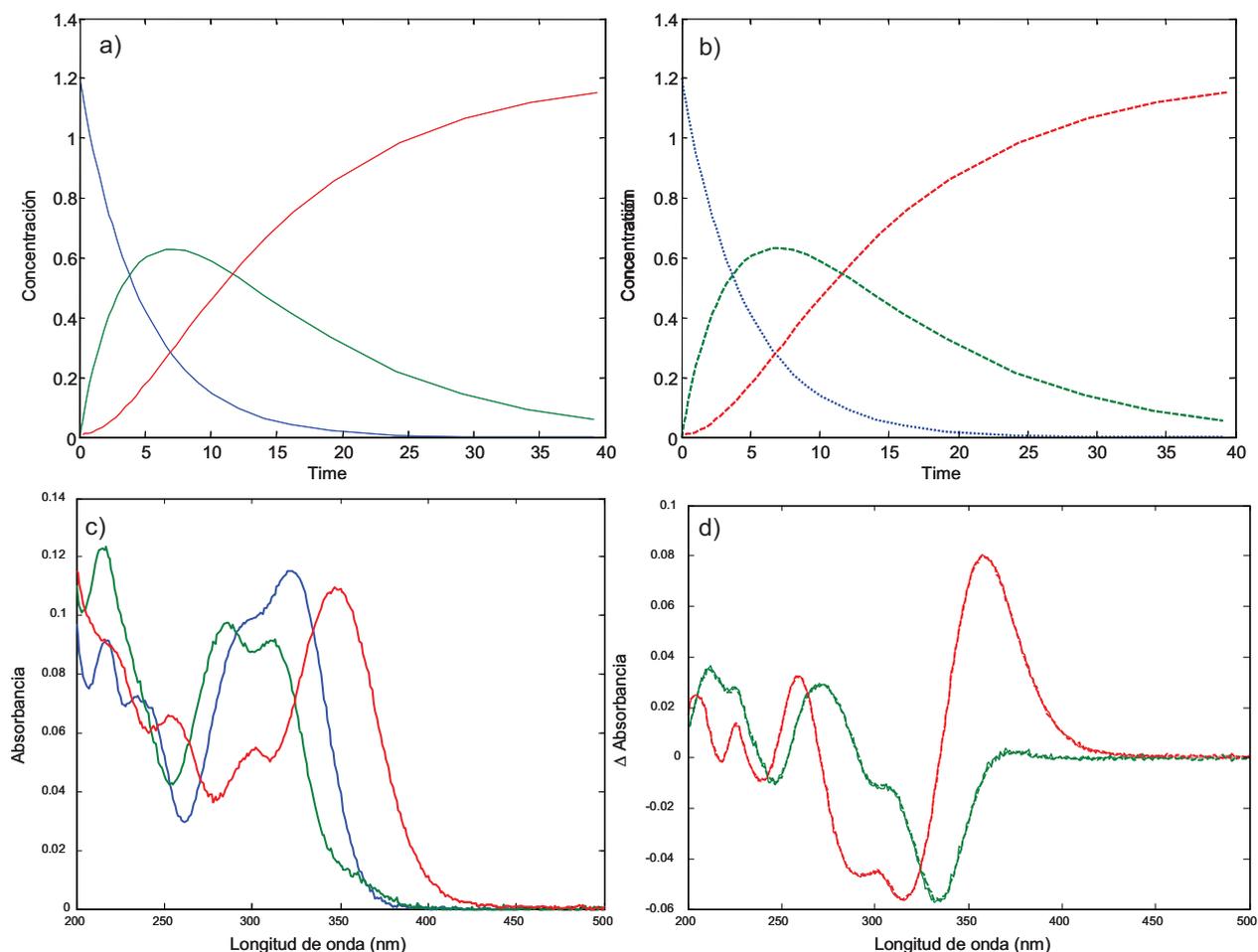


Figura 98: a) y b) Perfiles de concentraciones y c) y d) espectros de puro obtenidos a partir de la resolución de los datos directos y de diferencia siguiendo el mismo proceso, respectivamente.

Se ha simulado un segundo ejemplo para probar la validez de nuestra aproximación en presencia de una especie interferente. El algoritmo híbrido HS-MCR es ahora de uso obligatorio, ya que fusiona las ventajas de los métodos de modelado duro (la introducción del modelo cinético para las especies del proceso) y de modelado suave (la capacidad para hacer frente a compuestos interferentes, que solamente se modelan con el uso de restricciones suaves). En este ejemplo, el primer conjunto de datos de diferencia corresponde al estudio de la cinética de un proceso y el segundo se refiere al mismo proceso pero con la absorción de una nueva contribución que no evoluciona según el modelo cinético. Las resoluciones obtenidas en ambos conjuntos de datos se muestran en la Figura 99. Los perfiles de concentración relacionados con el proceso son en ambos casos recuperados correctamente y en el segundo caso, la contribución interferente es modelada con restricciones suaves. La presencia de este interferente no perturba la resolución de las demás contribuciones.

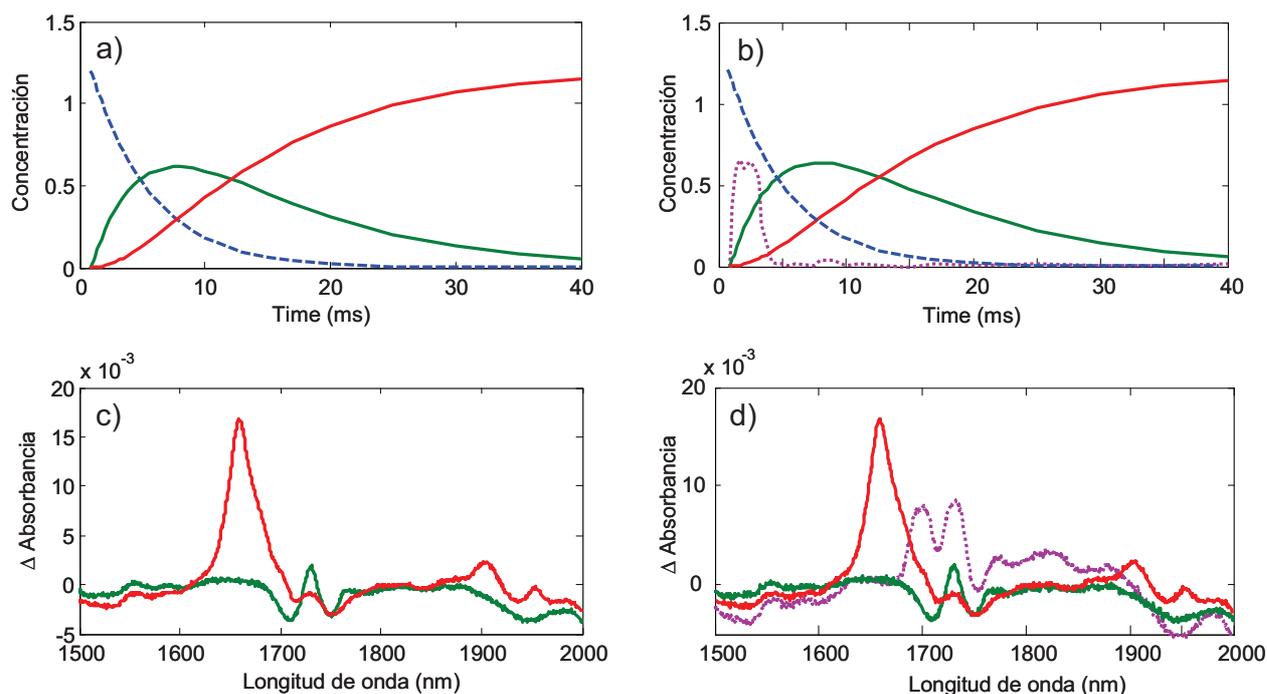


Figura 99 : a) y b) Perfiles de concentración obtenidos, respectivamente, a partir de datos de diferencia en ausencia o presencia de un interferente y c) y d) los espectros puros correspondientes. La línea azul punteada hace referencia a la especie inicial transparente.

El potencial del modelado híbrido se demostrará en las siguientes secciones durante el análisis de datos resueltos en el tiempo de IRTF *rapid-scan* y de espectroscopia de absorción transitoria.

Conclusiones

Esta sección y las publicaciones relacionadas permiten extraer las conclusiones siguientes:

- los datos espectroscópicos de diferencia se ven afectados por una deficiencia de rango, independientemente del proceso químico estudiado;
- la introducción de un modelo cinético permite gestionar esta deficiencia mediante la reintroducción de los perfiles de concentración de las especies iniciales como especies transparentes;
- el uso de un algoritmo híbrido permite modelar correctamente un proceso perturbado por una contribución interferente.

Estos resultados serán útiles en el análisis de dos sistemas químicos reales, como se describe en los apartados 3 y 4 de este capítulo.

Publicaciones relacionadas con esta sección

El trabajo presentado a lo largo de esta sección ha sido objeto de dos artículos y cuatro comunicaciones orales y cuatro pósteres. La lista se proporciona a continuación.

Publicación I: Hybrid Hard- and Soft-Modelling Applied to Difference Spectra

L. Blanchet, C. Ruckebusch, J.P. Huvenne, A. de Juan

Chemometrics and Intelligent Laboratory Systems (Elsevier, Amsterdam)

2007 Volume 89 (1) pages 26-35

Publicación II: Focus on the potential of hybrid hard- and soft- MCR-ALS in Time Resolved Spectroscopy

Lionel Blanchet, Cyril Ruckebusch, Jean-Pierre Huvenne, Anna de Juan

Journal of Chemometrics (Wiley, New York)

DOI: 10.1002/cem.1144

Oral I: Análisis quimiométrica de espectrometría resuelta en el tiempo

L. Blanchet, C. Ruckebusch, J.P. Huvenne, A. de Juan

Séminaire de l'Universitat de Barcelona, Barcelona, Spain, 2006

Oral II: Multivariate curve resolution of time resolved spectra

L. Blanchet, C. Ruckebusch, J.P. Huvenne, A. de Juan

10th Scandinavian Symposium on Chemometrics (SSC10), Lappeenranta, Finland, 2007

Oral III: Adaptation des méthodes de résolution multivariée aux données spectroscopiques résolues en temps

L. Blanchet, C. Ruckebusch, J.P. Huvenne, A. de Juan

Séminaire de 2^{ème} année de thèse au LASIR, Villeneuve d'Ascq, France, 2007

Oral IV: Modélisation hybride (*hard soft*) de données spectroscopiques résolues en temps

L. Blanchet, C. Ruckebusch, J.P. Huvenne, A. de Juan

Chimimétrie 2007, Lyon, France, 2007

Poster I: Detection limit in Multivariate Curve Resolution applied to spectroscopic data.

L. Blanchet, C. Ruckebusch, L. Duponchel, J.P. Huvenne

32nd FACSS and 51th ICASS, Québec, Canada, 2005

Poster II: Quelle limite de détection en résolution de courbes multivariées sur des données spectroscopiques ?

L. Blanchet, C. Ruckebusch, J.P. Huvenne, A. de Juan

Chimimétrie 2005, Villeneuve d'Ascq, France, 2005

Poster III: Gestion d'une déficience de rang de séries de spectres de différence résolus en temps.

L. Blanchet, C. Ruckebusch, J.P. Huvenne, A. de Juan

Chimimétrie 2006, Paris, France, 2006

31th November - 1st December 2006 (Presented by Cyril Ruckebusch)

Poster IV: Análisis híbrido hard-soft modelling para datos de diferencia resuelta en el tiempo.

L. Blanchet, A. Mezzetti, C. Ruckebusch, J.P. Huvenne, A. de Juan

1er Encuentro de Jóvenes Investigadores en Quimiometría, Tarragona, Spain, 2006

Estudio del centro de reacción de Rhodobacter sphaeroides utilizando espectroscopia IRTF rapid-scan y quimiometría

Las conclusiones extraídas sobre el rango de datos de diferencia se explotan en el estudio de un proceso fotoquímico natural. El centro de reacción de la bacteria *Rhodobacter sphaeroides* es el lugar donde la luz absorbida se convierte en energía química. Algunos aspectos de este proceso no son perfectamente conocidos. Un estudio basado en espectroscopia IRTF *rapid-scan* y quimiometría se utiliza para aportar nuevas conclusiones.

Fotosíntesis bacteriana

La fotosíntesis ha sido, a lo largo de la evolución, el principal proceso de formación de la materia orgánica. Este fenómeno complejo es habitualmente conocido en las plantas verdes pero ocurre también en algunas bacterias. El grupo más estudiado de las bacterias fotosintéticas es el de las bacterias púrpura porque son los organismos fototróficos más simples [32]. Cada especie utiliza un sólo tipo de bacterioclorofila y un centro de reacción (RC). El RC de *Rhodobacter sphaeroides*, estudiado aquí, ha sido caracterizado en 3D por difracción de rayos X [33-35]. Se ha propuesto un mecanismo para explicar cómo la energía absorbida por los pigmentos permite oxidar las moléculas de ubiquinona (Q) presentes en el centro de reacción a ubiquinol (QH₂) [36]. La energía se concentra en un par especial de moléculas de bacterioclorofila que inducen una separación de cargas. El electrón formado se transfiere a una molécula de ubiquinona a través de un bacteriofeofitina H_A. La repetición de este fenómeno y la introducción de dos protones permiten convertir la ubiquinona en ubiquinol QH₂, que luego es sustituido por una nueva molécula de ubiquinona.

Soft-modelling

Como se indica en la publicación III, nuestro primer enfoque consistió en analizar los datos sin asumir ni utilizar ningún modelo mecánico. El algoritmo de modelado suave MCR-ALS es aplicado aquí a una matriz de datos aumentada, formada por cinco experimentos realizados a distintas intensidades de iluminación y mostrando una etapa de excitación de la bacteria (bajo la luz) seguida de una etapa de relajación (en la oscuridad). La descomposición obtenida es presentada en la Figura 100. La interpretación de las bandas de los espectros puros indica que la formación de ubiquinol (líneas verde y roja) puede ser

separada del cambio conformacional de la proteína (línea azul) que se produce en el entorno inmediato del centro de reacción.

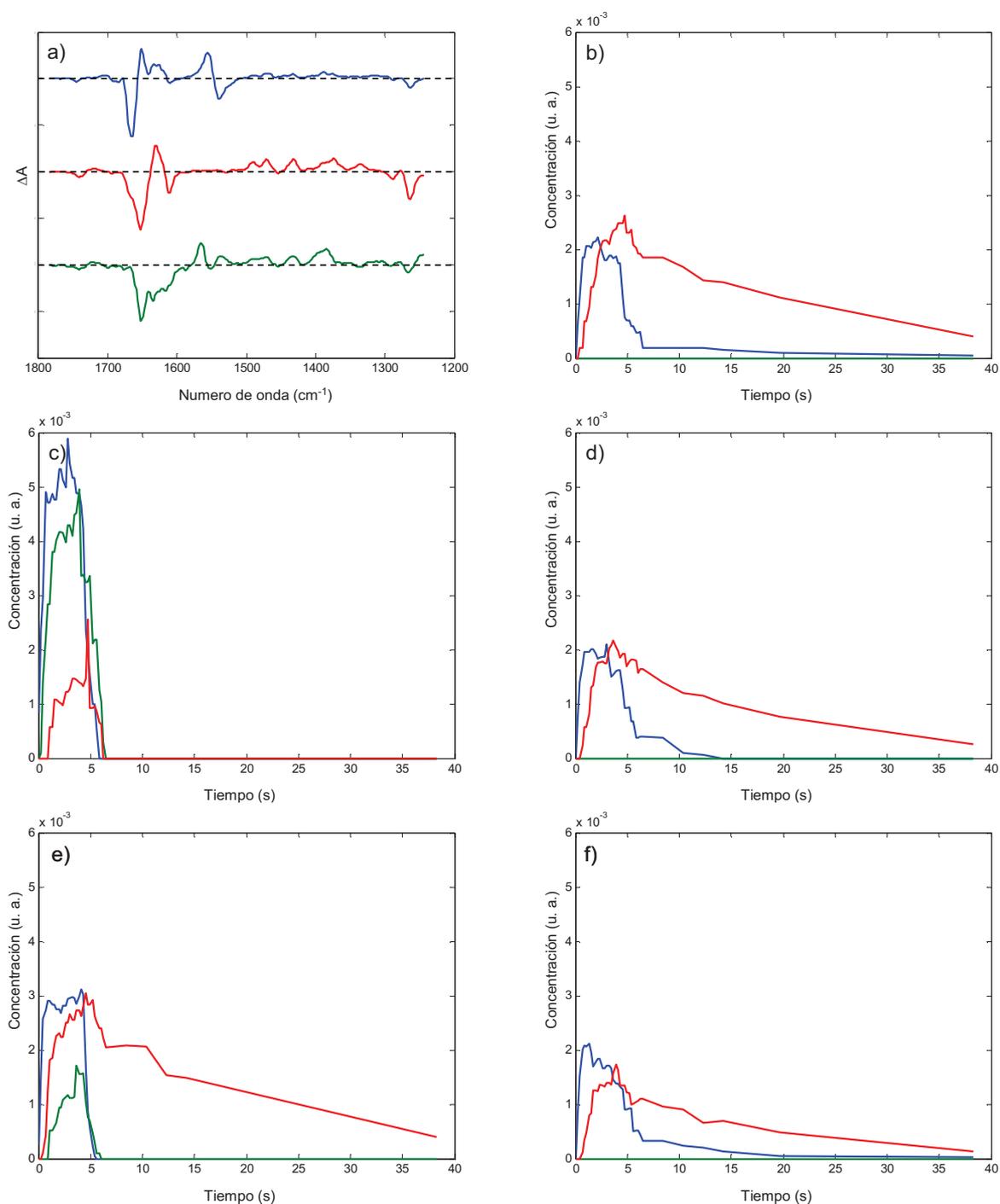


Figura 100 : Resultados obtenidos con un análisis MCR-ALS de la matriz aumentada en la dirección de las columnas. a) espectros puro (de arriba a abajo: del cambio conformacional de la proteína, del proceso de formación de ubiquinol a intensidad luminosa baja y alta). b)-e) perfiles de concentración obtenidos para los cinco experimentos distintos (b,d,f, experimentos a baja intensidad luminosa; e,c: experimentos a alta intensidad luminosa).

Dos efectos pueden ser observados en los perfiles de concentración. El primero está relacionado con la intensidad de la luz utilizada, ya que a intensidades más altas, tiene lugar un segundo proceso, también relacionado con la formación de ubiquinol (línea verde). El análisis nos proporciona una descripción coherente de lo que está ocurriendo en la muestra durante y después de excitación luminosa (formación y desaparición de ubiquinol, respectivamente y cambio conformacional reversible del entorno proteico). Se pueden proponer modelos cinéticos utilizando esta información y probar el uso de restricciones de *hard-modelling*. El segundo efecto observado es la fatiga de la muestra a lo largo del experimento, probablemente debido al estrés provocado por la luz de excitación, que hace que los procesos se den en menor extensión cuando la muestra lleva más tiempo siendo irradiada (experimento $b < d < f$).

Hard- and Soft-Modelling

Nuestras conclusiones preliminares han demostrado que tres contribuciones tienen que ser modeladas. El perfil de concentración correspondiente al cambio de conformación del entorno proteico muestra una fuerte evolución a una escala de tiempo muy corta. Debido a este hecho y a que las proteínas son sistemas complejos, no parece que esta contribución pueda seguir una ley cinética simple en la escala de tiempo estudiada. Por lo tanto, ha sido necesario dejarla fuera de cualquier modelo duro. Los dos procesos relacionados con la formación de ubiquinol son más compatibles con la aplicación de restricciones cinéticas. La publicación IV presenta cómo HS-MCR gestiona simultáneamente estos dos aspectos del modelado (soft para la contribución proteica y hard para las otras dos contribuciones) y la forma secuencial de obtener los modelos de las fases de excitación y relajación de los procesos. Las ecuaciones (81) y (82) describen el modelo de la fase de excitación y las ecuaciones (83) y (84) describen el modelo de la fase de relajación.



Se construyó una matriz aumentada utilizando cinco experimentos diferentes y fueron aplicadas restricciones *soft* y *hard* mediante HS-MCR, como se muestra en la Figura 101.

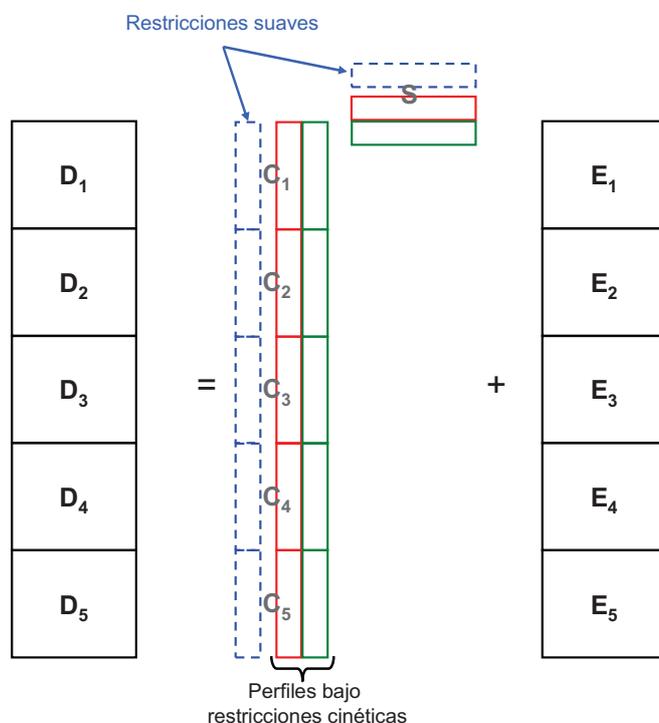


Figura 101 : Esquema de la matriz aumentada en columnas y de las restricciones aplicadas durante el modelado HS-MCR.

Los resultados se presentan en la Figura 102 y las bandas señaladas permiten atribuir cada espectro al cambio de conformación de la proteína o a los procesos de formación del ubiquinol. Las constantes de velocidad de los dos procesos de formación del ubiquinol se han calculado a las diferentes condiciones experimentales y se agrupan en la Tabla 1. Esta nueva información mejora el conocimiento del sistema, lo que lleva a la caracterización de dos procesos, que tienen lugar bajo iluminación intensa o suave. La intensidad de iluminación hace que la preponderancia de ambos procesos se vea modificada.

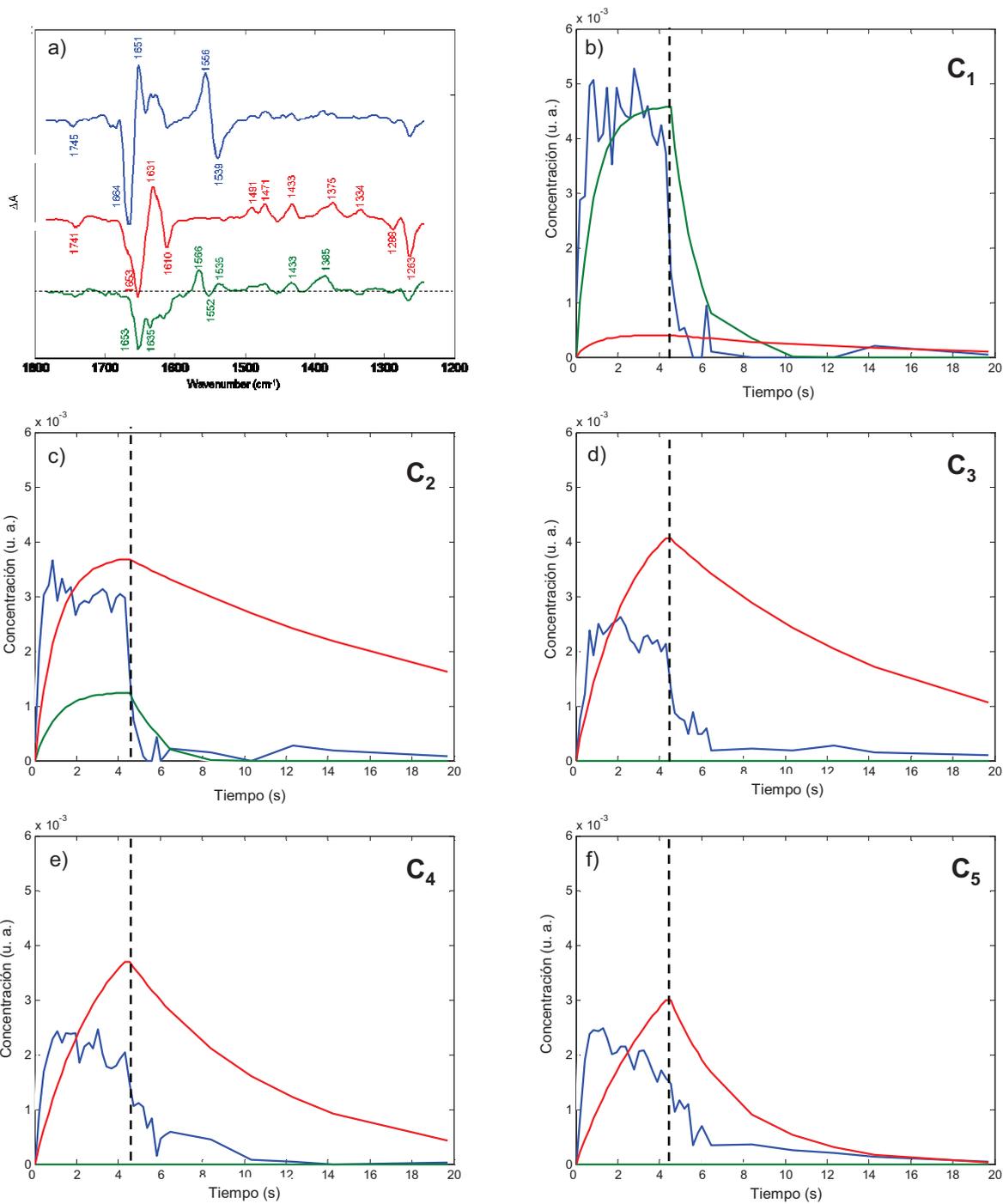


Figura 102 : Resultados obtenidos con un análisis HS-MCR de la matriz aumentada en la dirección de las columnas a) espectros puro (de arriba a abajo: de la proteína, del proceso de formación de ubiquinol en intensidad luminosa baja y alta) y b) hasta e) perfiles de concentración obtenidos para los cinco experimentos distintos.

Matrices	Excitación		Relajación	
	k ₁	k ₂	k ₃	k ₄
D_{hi}	0,09 ± 0.02 s ⁻¹	1,06 ± 0.08 s ⁻¹	0,09 ± 0.01 s ⁻¹	0,82 ± 0.01 s ⁻¹
D_{me}	0,73 ± 0.05 s ⁻¹	0,25 ± 0.03 s ⁻¹	0,05 ± 0.01 s ⁻¹	0,69 ± 0.01 s ⁻¹
D_{low1}	0,39 ± 0.01 s ⁻¹	-	0,09 ± 0.01 s ⁻¹	-
D_{low2}	0,31 ± 0.01 s ⁻¹	-	0,14 ± 0.01 s ⁻¹	-
D_{low3}	0,22 ± 0.01 s ⁻¹	-	0,29 ± 0.04 s ⁻¹	-

Tabla 1: Constantes de velocidad calculadas sobre las 5 matrices analizadas con HS-MCR.

Conclusiones

Esta sección y publicaciones relacionadas permiten extraer las siguientes conclusiones:

- la estrategia basada en HS-MCR es eficaz para la resolución de datos de espectroscopia de diferencia FTIR *rapid-scan*;
- la flexibilidad de HS-MCR permite conectar la información de entrada de los modelos de las fases de relajación con la información de salida de los modelos de las fases de excitación;
- dos procesos de formación de ubiquinol han sido detectado y parecen ser dependientes del nivel de excitación luminosa;
- la fatiga de la muestra biológica repercute en el desarrollo del proceso fotosintético.

Publicaciones relacionadas con esta sección

Los trabajos presentados a lo largo de esta sección han sido objeto de dos artículos y una comunicación oral y tres pósteres. La lista se proporciona a continuación.

Publicación III: Multivariate Curve Resolution of Rapid-Scan FTIR Difference Spectra of Quinone Photoreduction in Bacterial Photosynthetic Membranes

L. Blanchet, A. Mezzetti, C. Ruckebusch, J.P. Huvenne, A. de Juan

Analytical and Bioanalytical Chemistry (Springer, Heidelberg)

2007 Volume 387 (5) pages 1863-1873

Publicación IV: Monitoring and interpretation of photoinduced biochemical processes by rapid-scan FTIR spectroscopy and hybrid hard- and soft-modeling

L. Blanchet, C. Ruckebusch, A. Mezzetti, J.P. Huvenne, A. de Juan

Journal of Physical Chemistry B (American Chemical Society, Washington)

Submitted

Oral V: Static and time-resolved FTIR difference spectroscopy: a valuable tool in the investigation of the mechanism of photo-induced reactions

A. Mezzetti, L. Blanchet, C. Ruckebusch

Conference of the Italian Group of Photochemistry, Salice Terme, Italie, 2006

Poster V: Resolution of Rapid Scan FTIR spectra of the *Rb. sphaeroides* photochemical system.

L. Blanchet, C. Ruckebusch, A. Mezzetti, A. de Juan

The Belgian Chemometrics Society (BCS) Symposium, Brussels, Belgium, 2006

Poster VI: Ubiquinone photo-reduction in photosynthetic membranes by time-resolved FTIR difference spectroscopy: biophysical insights and chemometric data analysis

A. Mezzetti, W. Leibl, L. Blanchet, C. Ruckebusch, A. De Juan, J.P. Huvenne

4th International Conference on Advanced Vibrational Spectroscopy, Corfou, Greece, 2007

Poster VII: Application of time-resolved FTIR difference spectroscopy to photosynthetic systems: biophysical insights and chemometric data analysis

A. Mezzetti, L. Blanchet, C. Ruckebusch, A. De Juan, J.-P. Huvenne, G. Giacometti, R. Spezia, W. Leibl

12th European Conference on the Spectroscopy of Biological Molecules, Bobigny, France, 2007

Estudio de la fotoquímica de la benzofenona utilizando espectroscopia de absorción transitoria y quimiometría

La fotoquímica se define como el estudio de las reacciones químicas inducidas por la absorción de la luz. La utilización de una fuente de excitación monocromática permite llevar a la molécula a un único estado excitado monoenergético, del cual se puede estudiar la reactividad o el mecanismo de reacción seguido. Los procesos fotoquímicos se definen según el momento en que tienen lugar. Así, podemos distinguir fotoprocesos de excitación primaria (por ejemplo, la fluorescencia, fosforescencia) y secundaria (reacciones que llevan a cabo las especies transitorias formadas). Todos estos procesos tienen lugar en escalas de tiempo muy cortas, que van de los 10^{-6} a 10^{-14} segundos. Para lograr la comprensión de eventos químicos tan rápidos, se han desarrollado herramientas conceptuales y experimentales. Los métodos impulsionales son las técnicas más eficientes para iniciar y estudiar estos procesos. La espectroscopia de absorción transitoria se basa en el principio de medidas bomba-sonda (*pump-probe*). La bomba excita la muestra y, después de un tiempo determinado, un segundo impulso, la sonda, permite acceder a información sobre el estado de excitación de la muestra (singulete, triplete,...). Esta técnica se ha aplicado al estudio del cruce intersistema de la benzofenona entre el estado singulete y triplete $S_1 \rightarrow T_1$.

Relajación de la benzofenona

La relajación de los estados excitados de la benzofenona es un proceso fotoquímico importante. Una de las particularidades de esta molécula es que el salto energético entre los estados singuletes y tripletes es relativamente pequeño y se asocia a un importante acoplamiento espín-órbita. El rendimiento cuántico de la formación del triplete desde el estado fundamental S_0 es casi igual a 1. Sin embargo, la transición del singulete excitado al estado triplete estable T_1 es más controvertida: de acuerdo con las normas d'El Sayed [37], la relajación debería hacerse a través de un estado T_2 , que nunca ha sido observado experimentalmente. En el análisis de los experimentos efectuados, la determinación del número de especies es aquí un punto relevante que debe ser validado. Se ha determinado mediante un análisis exhaustivo que tres componentes son detectables en los datos. El trabajo completo está disponible en la publicación V. Brevemente, la validez de la estimación del rango y de los resultados del modelado suave fueron investigados mediante el método EFA y el estudio de los residuales asociados al modelo final. Estos diagnósticos son especialmente

interesantes en un caso como el de la espectroscopia UV resuelta en el tiempo, donde los espectros de especies puras pueden tener un gran solapamiento entre ellos. El trabajo presentado en la Publicación VI tiene como objetivo proponer nuevas indicaciones sobre el comportamiento fotoquímico de la benzofenona utilizando la espectroscopia de absorción transitoria y la quimiometría. MCR-ALS y HS-MCR han sido aplicados tanto al análisis de la benzofenona como al de la metoxibenzofenona. El primer paso es validar las herramientas quimiométricas que se utilizan mediante el estudio de la metoxibenzofenona, que es una molécula modelo de la que se sabe que la naturaleza de sus estados excitados depende del disolvente. Esto significa que, en algunos disolventes, una especie transitoria es necesaria durante la fotorelajación (como en el ciclohexano) y en otros no (mezcla acetonitrilo agua), como se muestra en la Figura 103. Como se observa, el uso de MCR-ALS solamente modela especies transitorias cuando realmente se hallan presentes. Por lo tanto, en caso de detectarse este tipo de especie en la benzofenona respondería a la presencia química de esta forma química y no a un ‘artefacto’ del algoritmo utilizado.

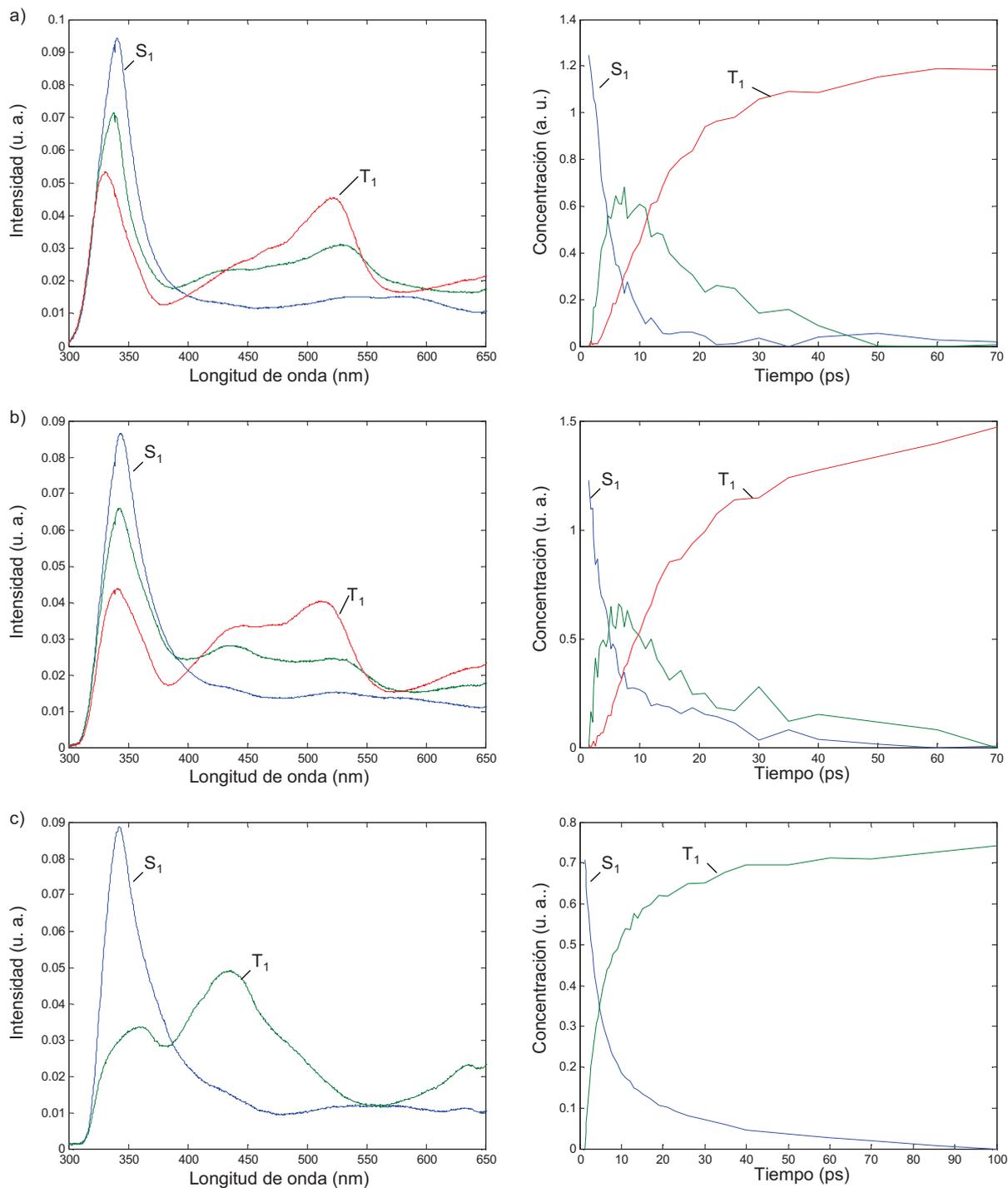


Figura 103: Resoluciones MCR-ALS de los espectros de absorción transitoria de la metoxibenzofenona en a) el ciclohexano b) el acetonitrilo c) una mezcla de agua acetonitrilo (3:1).

En el estudio de la fotorelajación de la benzofenona, se detectó una especie intermedia. Los perfiles de concentración y espectros puros correspondientes se proponen en la Figura 104. La descripción de los datos muestra una falta de ajuste de 0,59% y las constantes cinéticas ajustadas son $k_1 = 0,154 \pm 0.002 \text{ ps}^{-1}$ y $k_2 = 0,093 \pm 0.001 \text{ ps}^{-1}$. Este resultado parece

confirmar la hipótesis del mecanismo $S_1(n,\pi^*) \rightarrow T_2(\pi,\pi^*) \rightarrow T_1(n,\pi^*)$, que transcurriría a través de un estado intermediario triplete y que correspondería a transiciones entre estados excitados que sí respetan las reglas d'El Sayed

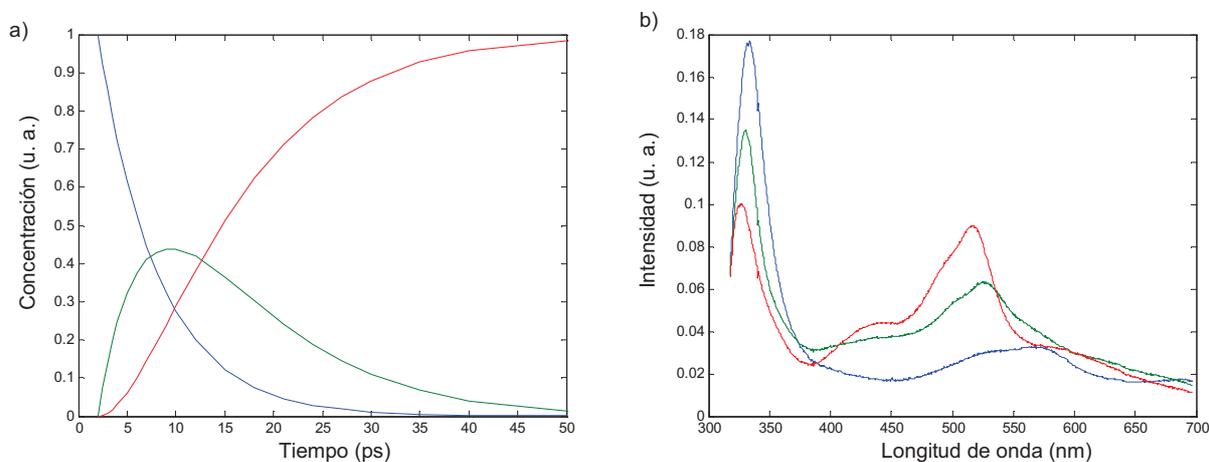


Figura 104: a) perfiles de concentración y b) espectros puros obtenidos por HS-MCR durante el análisis de los espectros de absorción transitoria de la benzofenona.

Conclusiones

Esta sección y las publicaciones relacionadas permiten deducir las conclusiones siguientes:

- los resultados de modelado suave obtenidos por MCR-ALS permiten obtener una primera visión del problema de la fotorelajación de benzofenona;
- el análisis de rango y los resultados de la resolución de curvas multivariantes indican la existencia de un estado intermedio durante la transición entre los estados S_1 y T_1 .
- la validez del método ha sido demostrada mediante el análisis del sistema conocido de la metoxibenzofenona;
- la introducción de una restricción cinética ofrece una estimación de las constantes de velocidad del proceso, que se han encontrado coherente con los resultados de *soft modelling* y con los valores disponibles en la literatura.

Publicaciones relacionadas con esta sección

Los trabajos presentados a lo largo de esta sección han sido objeto de dos artículos y dos comunicaciones orales. La lista se proporciona a continuación.

Publicación V: Reliable Multivariate Curve Resolution of Femtosecond Transient Absorption Spectra

C. Ruckebusch, S. Aloïse, L. Blanchet, J.P. Huvenne, G. Buntinx

Chemometrics and Intelligent Laboratory Systems (Elsevier, Amsterdam)

2008 Volume 91 (1) 17-21

Publicación VI: The benzophenone $S_1(n,p^*) \rightarrow T_1(n,p^*)$ states intersystem crossing reinvestigated by ultrafast absorption spectroscopy and multivariate curve resolution

S. Aloïse, C. Ruckebusch, L. Blanchet, J. Réhault, G. Buntinx, J.P. Huvenne

Journal of Physical Chemistry A (American Chemical Society, Washington)

2008 Volume 112 (2) 224-231

Oral VI : Résolution multivariée de spectres d'absorption transitoire femtoseconde.

C. Ruckebusch, S. Aloïse, L. Blanchet, J.P. Huvenne, G. Buntinx

Chimométrie 2006, Paris, France, 2006

Oral VII: Chemometrics of ultrafast time-resolved spectra

C. Ruckebusch, S. Aloïse, L. Blanchet, J.P. Huvenne

CAC 2008, Montpellier, France, 2008

Caracterización de la estructura del ruido y de su influencia en la resolución de datos

El segundo desarrollo metodológico de esta tesis se propone tener en cuenta las especificidades del ruido atribuible a la instrumentación. De hecho uno de los problemas de los métodos de resolución de curvas y de otros métodos de descomposición bilineal, como PCA, es el sobreajuste de los datos cuando éstos presentan un tipo de ruido que no es de naturaleza puramente aleatoria. Los algoritmos están diseñados para maximizar la varianza explicada lo que puede llevar a incorporar parte de la varianza relacionada con el ruido no aleatorio en la solución final. La descomposición resultante es entonces afectada de manera negativa.

Determinación de la estructura del ruido y de la presencia de contribuciones correlacionadas

La influencia del ruido sobre los resultados de los métodos de análisis de factores es a menudo considerada como no significativa. De hecho, esta asunción es correcta en aquellos casos en que el ruido es tan bajo que puede considerarse irrelevante. No obstante, algunas medidas procedentes de instrumentaciones avanzadas se realizan en condiciones tales que es muy difícil o muy laborioso reducir el nivel de ruido de la medida hasta tal punto que este elemento no deba ser tenido en consideración. Además, los aparatos modernos se basan a menudo en el desarrollo de detectores múlticanales, por ejemplo los espectrómetros utilizando detectores CCD que registran las intensidades en todas las longitudes de onda al mismo tiempo. Esto puede inducir algunos patrones de correlaciones entre los errores en las diferentes longitudes de onda. Este caso es crítico cuando se aplican métodos multivariantes de resolución de curvas porque el algoritmo interpreta tales patrones como una estructura importante de la señal y los incluye en la resolución.

Se ha investigado la estructura del ruido generado por un espectrómetro de absorción transitoria durante el seguimiento de la fotorelajación de la benzofenona. Este es el núcleo de la publicación VII. La estructura del ruido se estudió separadamente en la dirección de las filas (tiempo del proceso) y de las columnas (longitudes de onda) de la matriz de datos por observación de las matrices de covarianza y de correlación del error. No se detectaron correlaciones en la dirección temporal, pero la dirección espectral mostró características más

interesantes, con dos zonas estructuradas, tal y como puede apreciarse en la matriz de correlación (ver publicación VII). Un análisis de la información de las matrices de covarianza en la dirección espectral por PARAFAC y por TFA (Target Factor Analysis) llevó a la identificación de dos factores físicos responsables de los errores de medida: uno con forma idéntica a la de la fuente utilizada y otra a la de la señal promedio.

Utilización de MCR-WALS (resolución multivariante de curvas por mínimos cuadrados alternados ponderados)

Recientemente un procedimiento ponderado ha sido propuesto sobre la base del algoritmo MCR-ALS a fin de tener en cuenta eventuales variaciones de la distribución de ruido en los datos. Originalmente implementado para ruido heterocedástico (independiente); puede ser también aplicado en presencia de ruido estructurado o correlacionado, como el existente en la espectroscopia de absorción transitoria. La idea es calcular durante una descomposición MCR-ALS una forma ponderada de los datos que permite considerar de forma menos importante las zonas con mayor ruido y que permite incorporar la correlación del ruido en una o en las dos direcciones de la matriz en caso de que exista. Para utilizar la ponderación de manera efectiva, es necesaria una estimación correcta de la estructura del ruido. Un análisis exhaustivo de los patrones de error de la medida debe hacerse previamente a partir del estudio de los replicados de un experimento obtenidos con una instrumentación específica.

Se han construido simulaciones para comprobar el efecto del uso del algoritmo MCR-WALS sobre datos que incorporan patrones de ruido conocidos, algunos de ellos contruidos a imitación de los resultados encontrados en el estudio del ruido de los datos de espectroscopia de absorción transitoria. Se han distinguido cuatro casos para poder observar posteriormente el efecto de la ponderación en situaciones reales. La primera situación corresponde a un conjunto de datos \mathbf{D}_1 que sólo se ve afectado por ruido heterocedástico. El segundo \mathbf{D}_2 es más complejo a causa de la adición de un ruido correlacionado con una forma igual a la de la fuente de luz identificada en la caracterización de ruido en espectroscopia de absorción transitoria. La tercera matriz \mathbf{D}_3 también se ve afectada por un ruido correlacionado, pero éste tiene la misma forma que la señal promedio y por lo tanto es más difícil discriminarlo de la señal real. Por último, la cuarta situación \mathbf{D}_4 es comparable a los datos reales, con la presencia de los dos errores correlacionados identificados anteriormente. Las

resoluciones de estos conjuntos de datos se hacen utilizando tanto MCR-ALS como MCR-WALS para apreciar el efecto de la ponderación. Para la utilización de MCR-WALS, los sistemas de ponderación utilizados se han adaptado a la naturaleza del ruido presente en cada conjunto de datos. Dos posibilidades de ponderación han sido probadas para D_4 , una teniendo en cuenta los dos ruidos estructurados y otra, que no tiene en consideración el ruido que tiene la misma forma que la señal. En primer lugar, los resultados referidos a la calidad del ajuste, representados por el parámetro *Lof*, son reagrupados en la Tabla 1. Cada simulación se discutirá posteriormente.

	D_1	D_2	D_3	D_4
Lof teórico (%)	4.9	6.1	6.3	5.0
Lof MCR-ALS (%)	4.7	3.7	1.7	1.3
Lof MCR-WALS (%)	5.0	5.9	1.7	5.24 / 4.49

Tabla 15 : Lof teórico y obtenido por MCR-ALS y por MCR-WALS sobre las cuatro simulaciones propuestas.

La primera simulación se refiere sólo a la adición de ruido independiente heterocedástico. En este caso, el esquema de ponderación utiliza matrices de covarianza del error diagonales en el sentido temporal y espectral, para reflejar la falta de correlación del ruido en ambas direcciones. Los resultados obtenidos por los dos algoritmos están presentados en la Figura 105. Ambas resoluciones son correctas y muy similares. De hecho, a bajo nivel de ruido, el ruido independiente es gestionado correctamente por MCR-ALS. Aquí este algoritmo sobreajusta ligeramente los datos, lo cual es visible en la Figura 105b, en que los espectros puros obtenidos se ven más afectados por el ruido incorporado en la solución que los obtenidos por MCR-WALS, presentados en la Figura 105d. Merece la pena señalar que una base de la hipótesis de MCR-ALS es que el ruido es independiente y homocedástico. Esto explica por qué cuando el nivel de ruido heterocedástico es mayor, MCR-WALS proporciona claramente mejores resultados [38].

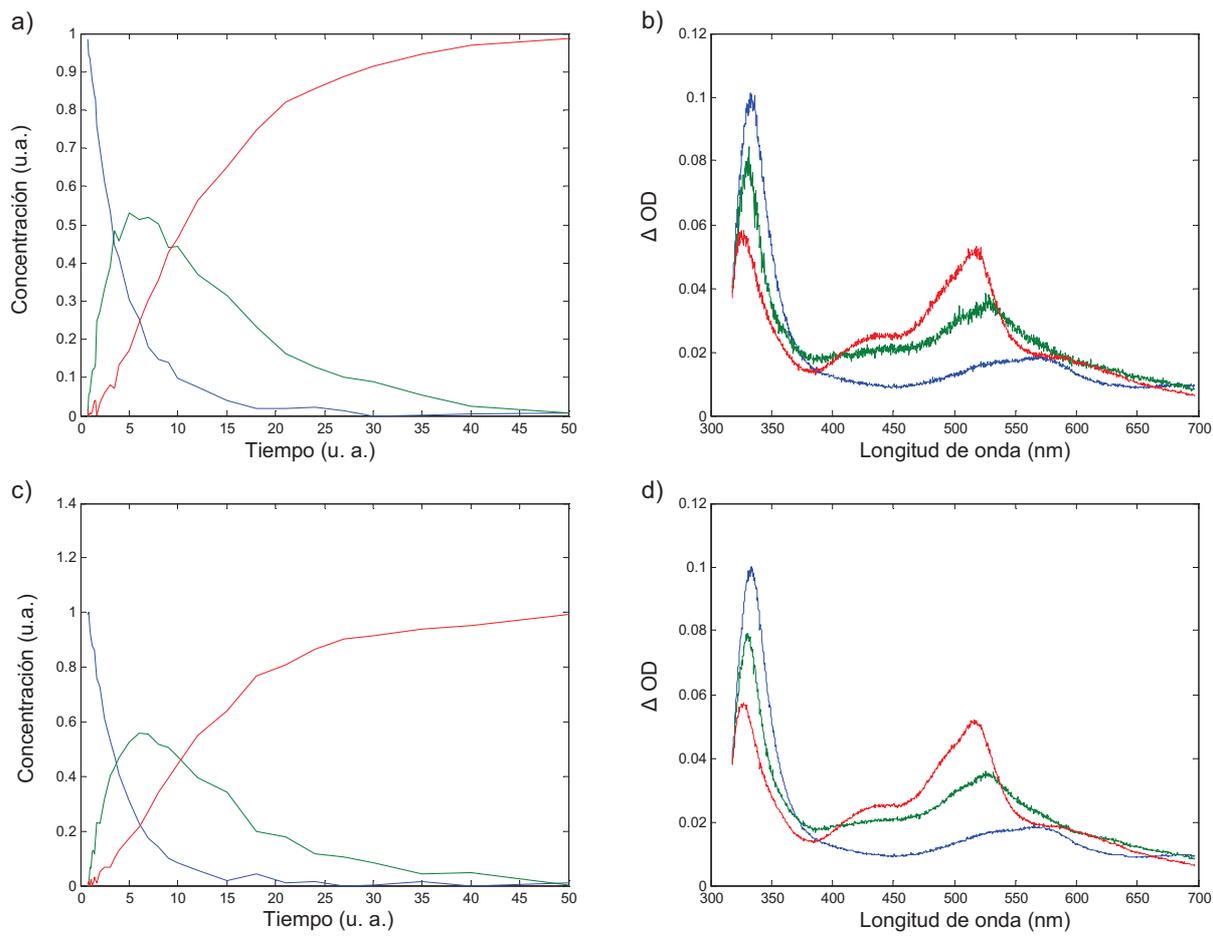


Figura 105: Resolución de D_1 obtenida por MCR-ALS a) perfiles de concentración y b) espectros puros y por MCR-WALS c) perfiles de concentración y d) espectros puros.

La matriz de datos \mathbf{D}_2 es un caso diferente porque se ha introducido un error estructurado. En este caso, el esquema de ponderación en la dirección espectral utiliza la matriz de covarianza del error completa para incorporar la información relativa a la correlación del error, mientras que se sigue utilizando una matriz de covarianza diagonal en la dirección temporal. Este tipo de error no puede ser gestionado adecuadamente por MCR-ALS. El efecto es claramente visible en los perfiles de concentración obtenidos, en la Figura 106a, y los espectros puros, en la Figura 106b. El *Lof* obtenido de 3,7% permite explicar estos resultados; MCR-ALS obtiene un *Lof* mejor que el teóricamente posible, un 6,1%. Esto indica un problema de sobreajuste e incorporación de ruido en la solución, que en este caso se refleja además en una distorsión de los perfiles obtenidos. Al contrario MCR-WALS propone una resolución correcta asociada a un *Lof* de 5,9% de acuerdo con el error añadido a los datos.

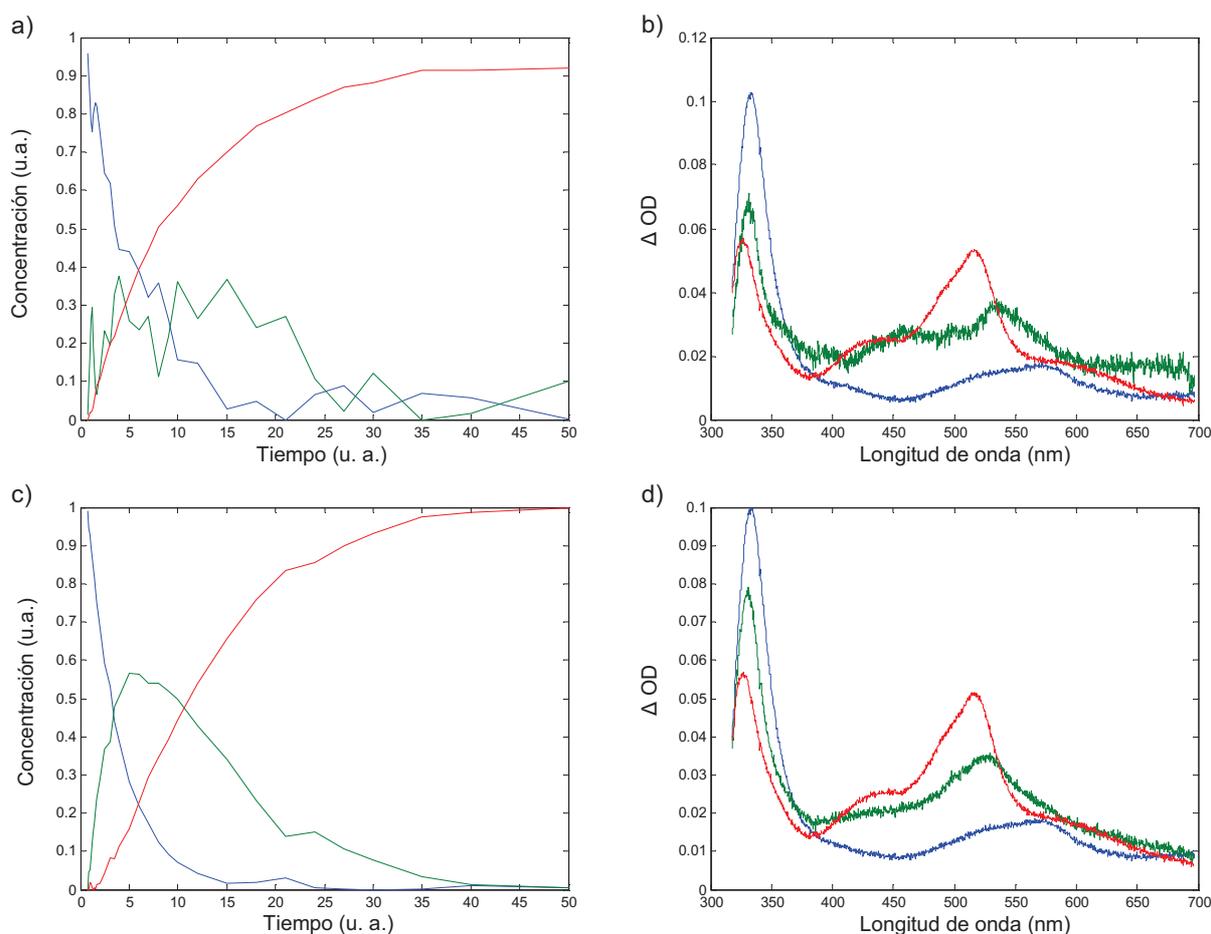


Figura 106: Resolución de \mathbf{D}_2 obtenida por MCR-ALS a) perfiles de concentración y b) espectros puros y por MCR-WALS c) perfiles de concentración y d) espectros puros.

El ruido presente en \mathbf{D}_3 tiene una estructura proporcional a la señal. Esta forma es también problemática para MCR-ALS, pero también para MCR-WALS. Las dos resoluciones obtenidas se presentan en la Figura 107. Ambas se ven afectadas negativamente por ruido, que es particularmente visible en los perfiles de concentración. El Lof teórico, de MCR-ALS y de MCR-WALS son aquí, respectivamente, igual al 6,3%, 1,7% y el 1,7%. Estas observaciones son coherentes con las consideraciones teóricas sobre esta tipología de error, que es fácil de identificar, pero difícil de tratar en la práctica [39]. Por analogía, la eliminación de este error de la estructura de los datos tendría la misma dificultad que tratar de modelar separadamente dos perfiles idénticos en una situación de deficiencia de rango.

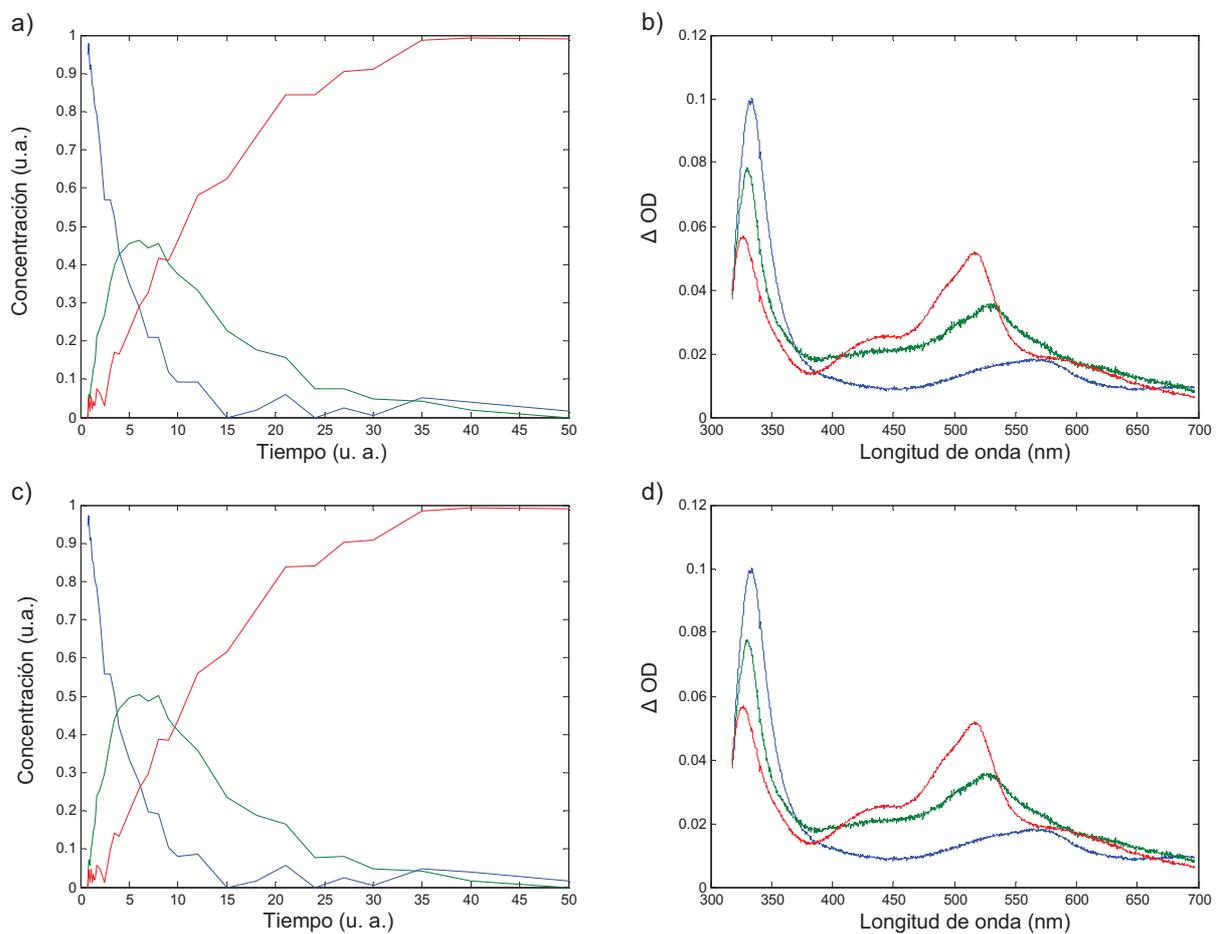


Figura 107: Resolución de \mathbf{D}_3 obtenida por MCR-ALS a) perfiles de concentración y b) espectros puros y por MCR-WALS c) perfiles de concentración y d) espectros puros.

Para finalizar, el último caso, \mathbf{D}_4 , está diseñado para ser una simulación realista de los verdaderos datos analizados anteriormente. Las dos estructuras de ruido utilizadas en \mathbf{D}_2 y \mathbf{D}_3 son añadidas a los datos originales simultáneamente. Aquí, sabiendo nuestros resultados anteriores, no se espera que MCR-ALS proporcione una resolución correcta. Los perfiles de concentración y espectros puros de la Figura 108a y la Figura 108b y muestran una vez más el impacto del problema de sobreajuste, aquí el *Lof* es igual a 1,3% y debería ser igual al 5,0%, y de la distorsión de los perfiles obtenidos. La primera posibilidad con MCR-WALS es poner a prueba un sistema de ponderación teniendo en cuenta los dos ruidos. La Figura 108c y la Figura 108d muestran que este procedimiento mejora la resolución. El *Lof* es aquí igual a 5,2%. La descomposición propuesta por las Figura 108e y Figura 108f es muy similar a la anterior, y ha sido obtenida mediante MCR-WALS y un segundo sistema de ponderación. Sólo una estructura de error es considerada, aquella diferente de la forma de la señal. El *Lof* del 4,5% es comparable a la anterior y al *Lof* teórico. Indica que la segunda estrategia es suficientemente eficaz para proponer una descomposición correcta de la matriz \mathbf{D}_4 . Esta conclusión es importante porque cuando el ruido estructurado está compuesto por más de un factor, es necesario evaluar correctamente el peso de cada uno de ellos en la estructura global. Es conocido que el ruido estructurado proporcional a la señal se encuentra a menudo en datos reales y también se ha comprobado que es fácil de detectar, pero muy difícil de corregir con estrategias de ponderación.. Por lo tanto, si no es necesario tenerlo en cuenta durante la resolución, esto facilitará mucho la construcción de esquemas de ponderación para el caso de ruidos multiestructurados.

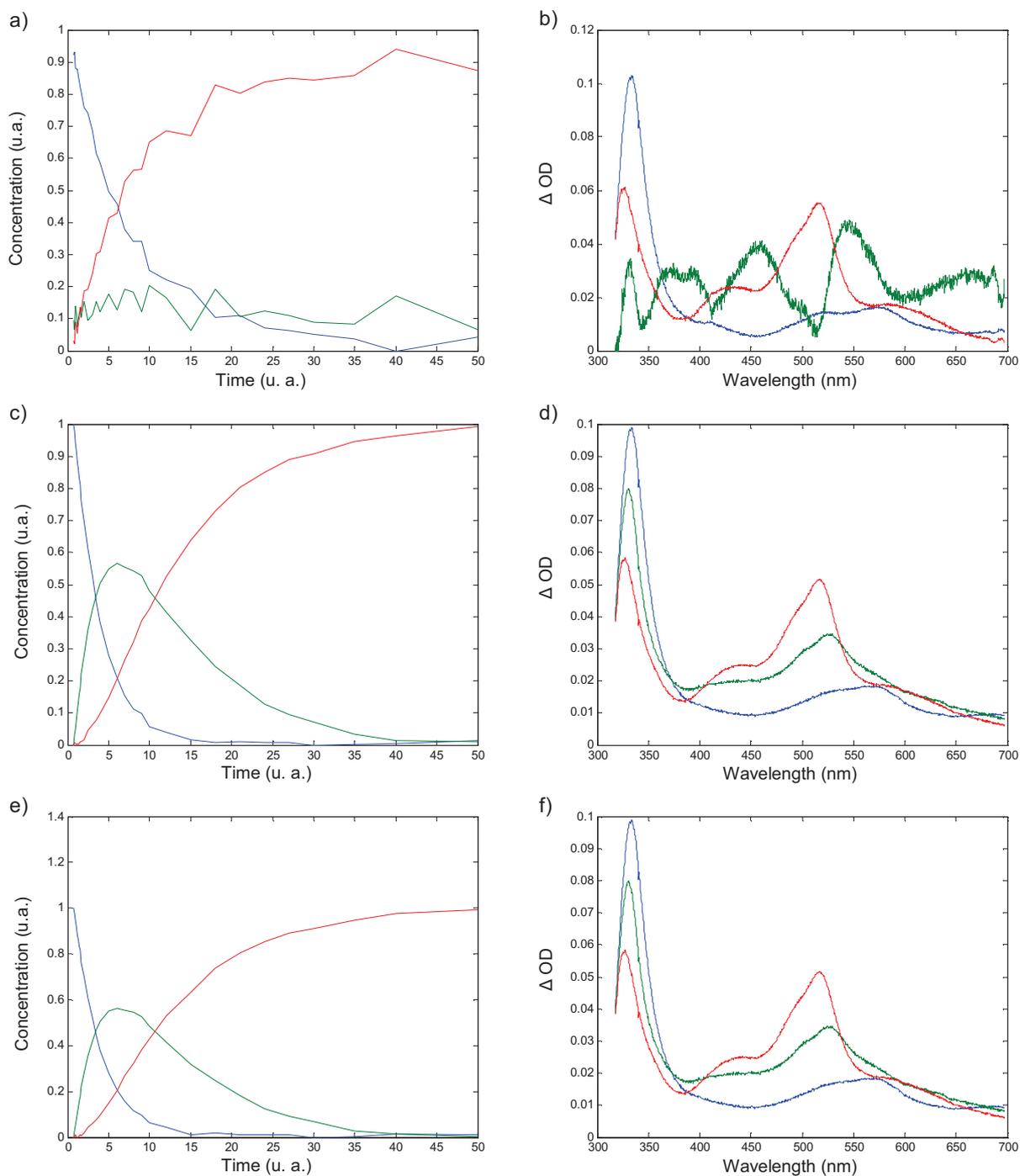


Figura 108: Resolución de D_4 obtenida con MCR-ALS a) perfiles de concentración y b) espectros puros; con MCR-WALS y el primer esquema de ponderación c) perfiles de concentración y d) espectros puros y con MCR-WALS y el segundo esquema de ponderación e) perfiles de concentración y f) espectros puros.

Publicaciones relacionadas con esta sección

Los trabajos presentados a lo largo de esta sección han sido objeto de un artículo y dos comunicaciones orales. La lista se proporciona a continuación.

Publicación VII: Chemometrics description of measurement error structure: study of an ultrafast absorption spectroscopy experiment.

L. Blanchet, J. Réhault, C. Ruckebusch, J.P. Huvenne, R. Tauler, A. de Juan

Analytica Chimica Acta (Elsevier, Amsterdam)

(In press)

Oral VIII: Assessing measurement error structure and weighting in multivariate curve resolution for time-resolved process data

L. Blanchet, C. Ruckebusch, J.P. Huvenne, R. Tauler, A. de Juan

CAC 2008, Montpellier, France, 2008

Oral IX: Influencia de la estructura de error en resolución multivariante

L. Blanchet, C. Ruckebusch, J.P. Huvenne, R. Tauler, A. de Juan

III Workshop en quimiometria, Burgos, Spain, 2008

Conclusión

El objetivo general de esta tesis es proponer métodos multivariantes de análisis adaptados a los datos de espectroscopias resueltas en el tiempo. Se han tenido en cuenta dos aspectos relevantes, uno se refiere a la gestión de datos de diferencia y el otro al estudio de la estructura del ruido y de su influencia en el análisis de datos.

En primer lugar, se ha demostrado que los datos resultantes de espectroscopia de diferencia están sistemáticamente afectados por una deficiencia de rango. Esta deficiencia debida a la naturaleza de la medida instrumental puede añadirse a otras deficiencias de rango causadas por el esquema de reacción del proceso estudiado. La formulación teórica de este problema ha permitido comprobar que los perfiles cinéticos de las contribuciones mesurables en la señal de diferencia son idénticos en forma a los de las especies químicas en la mezcla (excepto por la ausencia del perfil de la especie inicial del proceso). Por lo tanto, se ha propuesto utilizar esta propiedad para compensar las dificultades causadas por la deficiencia de rango. El uso de un modelo cinético, que incluye la primera especie como una contribución transparente, permite reconstruir todos los perfiles de concentración y obtener espectros de diferencia de las especies intermedias y finales del proceso. Se propone la utilización de un algoritmo híbrido (HS-MCR) que gestiona simultáneamente el problema de la deficiencia de rango (a través del modelo cinético) y la eventual presencia de interferentes absorbentes.

Esta estrategia ha sido aplicada con éxito en el análisis de los datos relacionados con el estudio del centro de reacción fotosintético de *Rhodobacter sphaeroides*. Los métodos de resolución multivariante de curvas (MCR-ALS) son eficientes para proponer una descomposición e interpretación de datos espectroscópicos IRTF *rapid-scan*. Además, el algoritmo híbrido ha sido adaptado a las necesidades específicas de los modelos asociados a experimentos que transcurren en fases sucesivas de excitación y relajación. Los resultados obtenidos muestran que la reacción fotosintética de formación de ubiquinol puede tener lugar mediante dos procesos distintos, dependiendo del nivel de excitación luminosa. En ambos casos, esta formación de ubiquinol va acompañada de una reorganización proteica, que tiene lugar en una escala de tiempo mucho menor que la seguida por la espectroscopia IRTF. Por lo tanto, debe ser descrita como una absorción de un interferente y modelada con restricciones suaves.

Los métodos de resolución de curvas multivariantes han sido aplicados al análisis de datos de espectros de absorción transitoria. El objetivo ha sido aclarar la incertidumbre sobre la existencia de un intermediario (previsto por la teoría) durante la formación del estado triplete T_1 de la benzofenona después de una fotoexcitación. Los resultados obtenidos a partir de modelos suaves (MCR-ALS) o de la aplicación de modelos híbridos (HS-MCR) parece indicar que este intermediario existe. Su espectro puro es asociable a un estado triplete T_2 . Esta hipótesis está apoyada por un estudio paralelo realizado en la metoxibenzofenona. La naturaleza de los estados excitados de esta molécula es dependiente del disolvente utilizado, por lo tanto hemos validado las herramientas quimiométricas utilizadas y la existencia de una especie transitoria.

El segundo eje metodológico de este trabajo se refiere al estudio de la influencia del ruido sobre la medida y sobre los métodos de resolución utilizados. La resolución de curvas multivariantes se basa en general en la hipótesis que el ruido que afecta las medidas experimentales es homocedástico. Un estudio sobre el error encontrado en espectroscopia de absorción transitoria muestra una realidad diferente. Dos factores contribuyen a la aparición de ruidos estructurados en la dirección espectral: uno relacionado con la forma de la señal y el otro con la fuente de luz utilizada durante el experimento. Una adaptación del algoritmo MCR-ALS ha sido desarrollada para tener en cuenta estos ruidos estructurados. Más concretamente, un sistema de ponderación que tiene en consideración la naturaleza correlacionada del error ha sido implementado. Varias simulaciones muestran el interés de esta estrategia. Además, se han podido validar los resultados obtenidos en las secciones anteriores con datos menos afectados por el ruido.

Las aportaciones metodológicas desarrolladas a lo largo de esta tesis permiten obtener resoluciones multivariantes más robustas e informaciones más fiables sobre los perfiles de concentración y espectros puros de las especies químicas que intervienen en un proceso seguido por espectroscopia resuelta en el tiempo. La detección e identificación de especies intermedias es particularmente interesante puesto que, gracias a ellas, se puede acceder a una mejor comprensión de las reacciones.

Referencias

- [1] E.R. Malinowski, *Factor Analysis in Chemistry*, John Wiley & Sons Inc, New York, 2002.
- [2] W. Windig, J. Guilment, *Anal. Chem.*, 63 (1991) 1425.
- [3] W. Windig, D.A. Stephenson, *Anal. Chem.*, 64 (1992) 2735.
- [4] F. Cuesta-Sanchez, B. Van den Bogaert, S.C. Rutan, D.L. Massart, *Chemom. Intell. Lab. Syst.*, 34 (1996) 139.
- [5] H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbuehler, *Anal. Chim. Acta*, 193 (1987) 287.
- [6] M. Maeder, *Anal. Chem.*, 59 (1987) 527.
- [7] H.R. Keller, D.L. Massart, *Chemom. Intell. Lab. Syst.*, 12 (1992) 209.
- [8] C.L. Lawson, R.J. Hanson, *Solving least squares problems*, Prentice-Hall, Englewood Cliffs, 1974.
- [9] R. Bro, N.D. Sidiropoulos, *J. Chemom.*, 12 (1998) 223.
- [10] R. Tauler, *Chemom. Intell. Lab. Syst.*, 30 (1995) 133.
- [11] R.C. Henry, B.M. Kim, *Chemom. Intell. Lab. Syst.*, 8 (1990) 205.
- [12] B.M. Kim, R.C. Henry, *Chemom. Intell. Lab. Syst.*, 49 (1999) 67.
- [13] P.D. Wentzell, J.H. Wang, L.F. Loucks, K.M. Miller, *Can. J. Chem.*, 76 (1998) 1144.
- [14] P.J. Gemperline, *Anal. Chem.*, 71 (1999) 5398.
- [15] R. Tauler, *J. Chemom.*, 15 (2001) 627.
- [16] M. Garrido, M.S. Larrechi, F.X. Rius, R. Tauler, *Chemom. Intell. Lab. Syst.*, 76 (2005) 111.
- [17] A. de Juan, M. Maeder, M. Martinez, R. Tauler, *Chemom. Intell. Lab. Syst.*, 54 (2000) 123.
- [18] S. Norman, M. Maeder, *Crit. Rev. Anal. Chem.*, 36 (2006) 199.
- [19] I.H.M. Van Stokkum, D. Larsen, S., R. Van Grondelle, *Biochim. Biophys. Acta*, 1657 (2004) 82.
- [20] M. Maeder, A.D. Zuberbühler, *Anal. Chem.*, 62 (1990) 2220.
- [21] K.J. Molloy, M. Maeder, M.M. Schumacher, *Chemom. Intell. Lab. Syst.*, 46 (1999) 221.
- [22] E. Furusjö, L.-G. Danielsson, *Chemom. Intell. Lab. Syst.*, 50 (2000) 63.
- [23] K.M. Mullen, M. Vengris, I.H.M. Van Stokkum, *J. Glob. Optim.*, 38 (2007) 201.

- [24] H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbühler, *Talanta*, 32 (1985) 95.
- [25] H. Gampp, M. Maeder, C.J. Meyer, A.D. Zuberbühler, *Talanta*, 32 (1985) 251.
- [26] M. Maeder, Y.M. Neuhold, *Practical data analysis in chemistry*, Elsevier, Amsterdam, 2007.
- [27] M. Maeder, Y.M. Neuhold, G. Puxty, P.J. Gemperline, *Chemom. Intell. Lab. Syst.*, 82 (2006) 75.
- [28] D.W. Marquardt, *SIAM J. Appl. Math.*, 11 (1963) 431.
- [29] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, *Chemom. Intell. Lab. Syst.*, 33 (1996) 17.
- [30] L. Blanchet, C. Ruckebusch, J.P. Huvenne, A. de Juan, *Chemom. Intell. Lab. Syst.*, 89 (2007) 26.
- [31] L. Blanchet, C. Ruckebusch, J.P. Huvenne, A. De Juan, *J. Chemom.* DOI: 10.1002/cem.1144.
- [32] J. Farineau, J.-F. Morot-Gaudry, *La photosynthèse. Processus physiques, moléculaires et physiologiques*, INRA, Paris, 2006.
- [33] J. Deisenhofer, E. U., K. Miki, R. Huber, H. Michel, *Nature*, 318 (1985) 618.
- [34] J. Deisenhofer, H. Michel, *Biosci. Rep.*, 24 (2004) 323.
- [35] C.R.D. Lancaster, H. Michel, *Photosynth. Res.*, 48 (1996) 65.
- [36] M. Okamura, M. Paddock, M.S. Graige, *Biochim. Biophys. Acta*, 1458 (2000) 148.
- [37] M.A. El Sayed, *Acc. Chem. Res.*, 1 (1968) 8.
- [38] P.D. Wentzell, T.K. Karakach, S. Roy, M.J. Martinez, C.P. Allen, M. Werner-Washburne, *BMC Bioinformatics*, 7 (2006) 343.
- [39] P.D. Wentzell, 'Maximum Likelihood-based soft-modeling methods' in *Comprehensive Chemometrics*, Elsevier, Amsterdam, In press.

Titre : Méthodes de résolution dédiées à l'étude spectroscopique de processus photoinduits. Adaptation aux spécificités des spectres résolus en temps.

Résumé :

L'analyse pertinente de l'information contenue dans les mesures instrumentales modernes nécessite le développement de méthodes algorithmiques performantes. La spectroscopie n'échappant pas à cette règle, nombre d'outils chimiométriques d'analyse multivariée ont été mis au point en conséquence. L'objectif de cette thèse est de participer à ces développements pour caractériser les spectres et les profils de concentration des composés intervenant lors de réactions photoinduites suivies en spectroscopie résolue en temps.

Un premier travail concerne l'analyse du rang de données de différence. Une déficience de rang systématique de ces données a été mise en évidence. Une approche basée sur un algorithme de modélisation hybride est proposée en réponse à ce problème. En effet l'introduction d'un modèle cinétique compense cette déficience de rang des données.

Le deuxième travail se rapporte à la caractérisation de l'influence du bruit (d'une part sur la mesure et d'autre part sur la résolution multivariée). Une analyse de la structure de corrélation du bruit permet d'expliquer les phénomènes de surajustement des méthodes de modélisation souple. Un second algorithme résout ce problème en pondérant les données en fonction d'une estimation de l'erreur de mesure.

Ces points sont illustrés par deux applications ; premièrement une contribution à l'analyse du centre réactionnel de la bactérie *Rhodobacter sphaeroides* par spectroscopie IRTF et ensuite l'étude de la photorelaxation de la benzophénone par spectroscopie d'absorption transitoire.

Mots-clés : chimiométrie, modélisation, résolution de courbes multivariées, *soft-modelling*, *hard-modelling*, déficience de rang, bruit structuré, spectroscopie de différence,

Title: Resolution methods dedicated to spectroscopic study of photoinduced process. Adaptation to the specificities of time resolved spectra.

Abstract:

Pertinent analysis of the information embedded in modern instrumental measurements requires the development of efficient mathematical and statistical tools. In spectroscopy, various tools have been proposed in such purpose. The aim of this thesis is to participate to these developments in order to characterize the spectra and concentration profiles of the compounds involved in photoinduced reactions monitored by time-resolved spectroscopy.

A first part of this work is related to the rank analysis of difference data. A systematic rank deficiency has been demonstrated in this case. An approach based on a hybrid algorithm is proposed as a solution to this problem. Indeed the introduction of a kinetic model compensates this rank deficiency.

A second part of the work deals with the characterisation of the noise influence (firstly on measurement and then on multivariate resolution). An analysis of the correlations structure of the noise permit to explain the overfitting phenomenon observed with soft modelling methods. A second algorithm solves this problem through a weighting scheme involving an estimation of the measurement error.

These points are illustrated by two applications: first, a contribution to the analysis of the reaction centre of *Rhodobacter sphaeroides* by FTIR spectroscopy; then the study of the benzophenone photorelaxation by transient absorption spectroscopy.

Keywords: chemometrics, modelling, multivariate curve resolution, time-resolved soft-modelling, hard-modelling, rank deficiency, structured noise, difference spectroscopy,