



HAL
open science

Analyse numérique des spectres de fluorescence 3D issus de mélanges non linéaires

Xavier Luciani

► **To cite this version:**

Xavier Luciani. Analyse numérique des spectres de fluorescence 3D issus de mélanges non linéaires. Traitement du signal et de l'image [eess.SP]. Université du Sud Toulon Var, 2007. Français. NNT : . tel-00287255

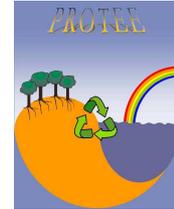
HAL Id: tel-00287255

<https://theses.hal.science/tel-00287255>

Submitted on 11 Jun 2008

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

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



THESE

présentée

devant l'Université du Sud Toulon Var

pour obtenir

le grade de DOCTEUR DE L'UNIVERSITÉ DU SUD TOULON VAR

par

Luciani XAVIER

Équipe d'accueil : Laboratoire - PROTEE

École Doctorale : USTV

Composante universitaire : UFR DE SCIENCES ET TECHNIQUES

Titre de la thèse :

Analyse numérique des spectres de fluorescence 3D issus de mélanges non linéaires

soutenue le 14 décembre 2007 devant la commission d'examen constituée de :

| | | |
|-----------------------|------------------------------------|------------------------|
| M. Pierre COMON | Directeur de recherche CNRS (UNSA) | Rapporteur. |
| M. Frédéric THERAULAZ | HDR (Université de provence) | Rapporteur. |
| M. David BRIE | Pr. (Université Henri Poincaré) | Examineur. |
| M. André BOIS | HDR (Université de provence) | Directeur de thèse. |
| M. Stéphane MOUNIER | HDR (USTV) | Co-directeur de thèse. |
| M. Claude JAUFFRET | Pr. (USTV) | Président. |

A ma grand-mère

Remerciements

Quatre années séparent ces quelques mots de mes débuts comme doctorant au sein du laboratoire PROTEE. Lorsque l'on a finalement connu que des bancs d'école, le temps lui-même semble s'y être assis. Ainsi une durée de quatre ans renvoie précisément pour moi au temps passé au collège, la période des copains et des expériences qui vous marquent. Pour des raisons proches, ces quatre ans de thèse constitueront à n'en pas douter ma nouvelle référence.

D'une manière générale et en marge de tout aspect scientifique, je tiens à saluer en premier lieu les membres des équipes ISO et CAPTE. Le plaisir que j'ai aujourd'hui à rédiger ces lignes doit beaucoup à la qualité des rapports entretenus avec la plupart d'entre eux et dont j'ai pu apprécier l'altruisme, la finesse d'esprit, et l'absence d'autosatisfaction.

Parmi eux, je remercie tout particulièrement : André Bois, mon directeur de thèse, pour la qualité de son encadrement d'un point de vue scientifique et humain, illustrée par la pertinence de ses remarques et la grande liberté de nos échanges ; Roland Redon, sur lequel j'ai pu compter à tout moment pour orienter et accompagner mes recherches. Tous deux m'ont appris la rigueur nécessaire à tout travail scientifique et m'ont transmis leur goût de la physique expérimentale ce qui n'était pas une mince affaire ; Stéphane Mounier pour m'avoir aidé à réaliser la validation expérimentale de nos approches et proposé plusieurs applications géochimiques, au prix des nombreuses séances de mesures et autres manipulations expérimentales que je lui ai imposées, mais également pour son respect d'un certain idéal universitaire. Cette thèse n'aurait pu avoir lieu sans la curiosité intellectuelle de ces trois chercheurs issus de domaines différents, Je leur suis donc particulièrement reconnaissant d'avoir mis leur ouverture d'esprit et leurs compétences diverses au service de ce travail interdisciplinaire. Je remercie enfin mes trois encadrant d'avoir trouvé le juste équilibre entre une certaine prévenance et l'importante autonomie qu'il m'a été laissée. J'ajoute que celle-ci n'aurait eu aucun sens sans leur très grande disponibilité. Le bon déroulement de cette thèse a été facilité par la complicité de plusieurs autres membres du laboratoire que je voudrais saluer ici : Marc Ripert qui m'a fait profiter de ses conseils avisés au laboratoire comme à l'IUT, de sa bonne humeur et finalement

de son amitié ; Michel Grimaldi pour sa disponibilité et son aide malgré nos thèmes de recherche différents. Je les remercie également tous deux d'avoir facilité mon intégration au sein du laboratoire et du plaisir que j'ai eu à refaire le monde avec eux ; Michel Raynaud pour l'intérêt porté à mon travail et ses compétences, spontanément mises à mon service ; Annaëlle Huiyu Zhao dont l'aide dans la dernière ligne droite me fut particulièrement précieuse.

Je remercie Yves Lucas de m'avoir accueilli dans son laboratoire et permis de travailler dans les meilleures conditions et bien sûr mes « camarades » doctorants (de PROTEE ou d'ailleurs) et notamment Habib, Nadège et Nicolas pour leur solidarité et leur amitié. Enfin je salue les membres de l'équipe EBMA que j'ai pu rencontrer et qui m'ont toujours soutenu et conseillé amicalement ainsi que tous ceux qui m'ont fait l'honneur d'assister à la soutenance de cette thèse.

Je voudrais exprimer toute ma gratitude à Messieurs Frédéric Théraulaz et Pierre Comon d'avoir accepté de rapporter cette thèse, dont l'objet sortait pour partie de leurs domaines de recherche respectifs. Je suis également très honoré de leur participation à la commission d'examen ainsi que de celle de Messieurs David Brie et Claude Jauffret que je remercie pour l'intérêt qu'ils portèrent à mon travail.

Mes études supérieures ne me prédestinaient pas à faire une thèse, aussi, je suis particulièrement reconnaissant à messieurs Philippe Courmontagne de L'ISEN et Claude Jauffret de m'avoir initié à la recherche en DEA et appuyé dans cette voie.

Enfin, ces quatre années ne faisant pas exceptions, j'ai pu bénéficier dans cette étape du soutien indéfectible de Maud et de mes parents. J'ai conscience que tous trois ont discrètement partagé avec moi les efforts consentis et les périodes de doute ou de stress. Je les remercie donc pour leur constance et leur délicatesse.

Table des matières

| | |
|--|-----------|
| Notations mathématiques | 7 |
| Introduction | 9 |
| 1 Spectroscopie de fluorescence | 13 |
| 1.1 Rappels historiques | 13 |
| 1.2 Mécanisme d'absorption | 14 |
| 1.2.1 Niveaux d'énergie et transitions électroniques | 14 |
| 1.2.2 Loi de Beer-Lambert et spectres d'absorption | 16 |
| 1.3 Emission de fluorescence | 18 |
| 1.3.1 Mécanismes de désactivation | 18 |
| 1.3.2 Spectres de fluorescence d'une solution | 21 |
| 1.3.2.1 Fonctionnement du spectrofluorimètre. | 21 |
| 1.3.2.2 Mesure du spectre d'émission. | 22 |
| 1.3.2.3 Spectres d'excitation. | 23 |
| 1.3.2.4 Matrices d'Excitation-Emission de Fluorescence. | 24 |
| 1.3.2.5 Correction des spectres | 24 |
| 1.3.2.6 Remarques générales sur les spectres. | 26 |
| 1.3.2.7 Influence du milieu sur la fluorescence. | 26 |
| 1.4 La diffusion | 26 |
| 1.5 Phénomènes non linéaires | 29 |
| 1.5.1 Mise en évidence du problème | 29 |
| 1.5.2 Limites de l'approximation linéaire et effet d'écran primaire | 30 |
| 1.5.3 Phénomène de réabsorption, effet d'écran secondaire | 33 |
| 1.5.4 Remarque finale | 33 |
| 2 Analyse numérique des spectres de fluorescence, état de l'art | 35 |
| 2.1 Introduction à l'analyse linéaire des mélanges fluorescents | 36 |
| 2.2 Décompositions bilinéaires | 38 |
| 2.2.1 Analyse en composantes principales | 39 |
| 2.2.2 Moindre carrés récursifs (MCR) | 41 |
| 2.2.3 OPA-Simplisma | 42 |
| 2.2.4 Analyses factorielles à fenêtre | 44 |
| 2.2.4.1 Construction de la carte de rang par EFA | 44 |
| 2.2.4.2 Résolution spectrale par EFA | 45 |

| | | |
|----------|---|-----------|
| 2.2.4.3 | Résolution spectrale par WFA/OPR | 46 |
| 2.2.4.4 | Conclusion | 48 |
| 2.3 | Décompositions trilineaires | 48 |
| 2.3.1 | Bilinéarisation | 48 |
| 2.3.1.1 | ACP d'un jeu de MEEF | 48 |
| 2.3.1.2 | OPA-A et ALS | 48 |
| 2.3.1.3 | Conclusion | 49 |
| 2.3.2 | Décomposition bilinéaire conjointe | 49 |
| 2.3.3 | PARAFAC | 50 |
| 2.3.3.1 | Modélisation trilineaire | 51 |
| 2.3.3.2 | Algorithme de décomposition | 52 |
| 2.3.3.3 | Utilisation pratique | 55 |
| 2.3.4 | Trilinear EFA | 57 |
| 2.3.4.1 | Principe | 57 |
| 2.3.4.2 | Avantages et limites | 58 |
| 2.3.5 | Lagrange Operator Algorithm | 58 |
| 2.4 | Approches non linéaires et effet d'écran | 58 |
| 3 | Modélisation physique des effets d'écran | 61 |
| 3.1 | De l'intérêt de corriger l'effet d'écran | 61 |
| 3.2 | Modélisation de l'effet d'écran | 62 |
| 3.2.1 | Approximations et notations | 62 |
| 3.2.2 | Modèle théorique | 66 |
| 3.3 | Validation expérimentale | 67 |
| 3.3.1 | Solution pure | 68 |
| 3.3.2 | Mélanges à deux composants | 69 |
| 3.3.3 | Mélanges à trois composants | 69 |
| 4 | Correction des effets d'écran | 73 |
| 4.1 | Introduction | 73 |
| 4.2 | Méthode de correction par faible dilution | 76 |
| 4.2.1 | principe | 76 |
| 4.2.2 | Bases de test | 77 |
| 4.2.3 | Résultats expérimentaux | 79 |
| 4.2.3.1 | Mélanges à 2 composants | 79 |
| 4.2.3.2 | Mélanges à 3 Composants | 91 |
| 4.2.3.3 | Matière organique | 100 |
| 4.3 | Méthode de correction par optimisation | 116 |
| 4.3.1 | Objectif et principe de la méthode | 116 |
| 4.3.2 | Estimation statistique d'une base de E | 117 |
| 4.3.3 | Estimation d'une base de E par optimisation | 119 |
| 4.3.4 | Utilisation pratique et premiers résultats | 120 |
| 4.3.4.1 | Nombre de mélanges nécessaires à l'estimation de la base | 120 |
| 4.3.4.2 | Résultats sur les mélanges synthétiques | 121 |

| | | |
|----------|--|------------|
| 4.3.4.3 | Résultats sur les mélanges réels | 124 |
| 4.4 | Conclusion | 127 |
| 5 | Applications au suivi de la matière organique | 131 |
| 5.1 | Exemple 1 : Sols du Cameroun | 132 |
| 5.1.1 | Prétraitement des données | 132 |
| 5.1.2 | Décomposition par PARAFAC | 132 |
| 5.1.3 | Première analyse des résultats | 133 |
| 5.2 | Exemple 2 : Baie de Sepetiba | 140 |
| | Conclusion | 141 |
| A | Exemple d'application 2 | 143 |
| B | Produit de Khatri-Rao | 155 |
| B.1 | Définitions et propriétés | 156 |
| B.2 | Lien avec PARAFAC | 157 |
| | Bibliographie | 164 |
| | Table des figures | 165 |

Notations mathématiques

Ce mémoire comporte de nombreuses formules et équations mathématiques. Nous récapitulons ici les notations et les règles communes à ces expressions.

- Les variables scalaires ou les fonctions sont représentées par des lettres minuscules : x .
- Les vecteurs sont représentés par des lettres minuscules en gras : \mathbf{x} .
- Les matrices sont représentées par des lettres majuscules en gras : \mathbf{X} .
- Les cubes de données sont représentés par des lettres majuscules en gras et soulignées : $\underline{\mathbf{X}}$.
- L'opérateur de transposition est signifié par un t à gauche : ${}^t\mathbf{X}$.
- $\|\cdot\|$ représente la norme euclidienne.
- La matrice identité de taille N est notée \mathbf{I}_N .

Note : Afin de ne pas multiplier les notations ou d'éviter des confusions, certains vecteurs du chapitre 4 seront représentés par des majuscules.

Quelques règles générales :

- Par défaut, les vecteurs sont des vecteurs colonnes.
- Les dimensions d'une matrice ou d'un cube sont données entre parenthèses : $\mathbf{X}, (P, Q)$ signifie que \mathbf{X} est une matrice de P lignes et Q colonnes.
- De nombreux paramètres sont indicés. Généralement, l'indice courant est représenté par une lettre minuscule et sa valeur maximale par la même lettre en majuscule.

Il est parfois utile de faire référence à une ligne ou une colonne d'une matrice. Ainsi,

- \mathbf{x}_n représente le vecteur correspondant à la colonne n de la matrice \mathbf{X} .
- ${}^t\mathbf{x}_n$ représente le vecteur correspondant à la ligne n de la matrice \mathbf{X} .

Nous utiliserons également différents produits :

- Le produit entre deux scalaires n'est pas symbolisé ou est parfois représenté par le signe \times .
- Le produit matriciel usuel n'est pas symbolisé.
- \bullet représente le produit de Hadamard.
- \otimes représente le produit de Kronecker.
- \odot représente le produit de Khatri-Rao.

Enfin, selon le contexte, une même grandeur pourra être vue comme une fonction, un paramètre indicé ou un vecteur. Les parenthèses sont alors réservées aux fonctions ou aux matrices et les crochets aux vecteurs.

On notera par exemple :

- $h(\lambda_j)$ la valeur de la fonction h pour la longueur d'onde λ_j .
- h_j ou $\mathbf{h}[j]$ la coordonnée j du vecteur \mathbf{h} .

Au fil des chapitres ou des sections, une même lettre pourra être réutilisée pour symboliser différentes variables. Cependant, certaines lettres sont réservées aux mêmes grandeurs physiques tout au long du mémoire. Ainsi,

- λ désigne la longueur d'onde (à l'exception de la section 2.2.1).
 - I désigne une intensité lumineuse mesurée.
 - Φ désigne le rendement quantique.
 - ε désigne le coefficient d'absorption molaire (lié au spectre d'excitation).
 - γ désigne le spectre d'émission normé.
-

Introduction

Qu'il nous paraisse inquiétant ou ludique, mystérieux ou attirant, le phénomène de fluorescence attise autant notre curiosité qu'il capte notre regard. Cette manifestation transitoire résulte des interactions entre la lumière et la matière. Lorsqu'ils sont éclairés par une lumière incidente, certains éléments réémettent, une partie de l'énergie à différentes longueurs d'onde. L'intensité de cette lumière de fluorescence varie en fonction des longueurs d'onde de la lumière incidente et de la lumière émise. L'allure de ces variations forme le spectre de fluorescence de l'élément éclairé et peut se mesurer à l'aide d'un spectrofluorimètre. Notons d'emblée que le cadre de notre recherche sera limité aux mesures de fluorescence réalisées en solution, dans l'UV et le visible avec une source d'excitation continue et corrigée de ses fluctuations temporelles. Au niveau moléculaire, le spectre de fluorescence est propre au groupement fluorescent (ou fluorophore) dont il est issu, bien qu'il dépende également d'éléments extérieurs. Les applications scientifiques et techniques sont donc nombreuses, qu'il s'agisse d'étudier les propriétés de la matière ou d'identifier des groupements fluorescents. Par exemple, en chimie de l'environnement, la mesure de ces spectres permet non seulement de caractériser des prélèvements inconnus mais constitue également un traceur de la matière organique.

Dans de nombreux cas, le signal mesuré est une combinaison des spectres de fluorescence des différents fluorophores composant la solution, alors qu'il ne s'agit pas de l'information d'intérêt. En effet, il est bien plus intéressant de disposer des spectres individuels des composants du mélange ou encore des rapports de concentration. Par conséquent, de nombreux outils d'analyse numérique permettant d'approcher ces résultats avec plus ou moins de précision, sont régulièrement proposés ou améliorés dans la littérature. Ces différents algorithmes ont des bases algébriques ou statistiques et reposent sur des décompositions linéaires ou multilinéaires des signaux du mélange. L'hypothèse de départ est alors la suivante : le spectre du mélange est une combinaison linéaire des spectres de fluorescence des différents composants fluorescents du mélange et les coefficients de cette combinaison sont proportionnels aux concentrations.

Lorsque l'on étudie le dispositif de mesure, il apparaît clairement qu'il ne s'agit là que d'un modèle mathématique pratique résultant d'une série d'approximation au premier ordre. Ce modèle linéaire de fluorescence est classique et très largement employé pour sa simplicité. Par ailleurs, il donne dans certains cas de très bons résultats. On parle alors de mélanges linéaires. Cependant, cette approximation peut s'avérer désastreuse lorsqu'elle n'est pas justifiée. D'une manière un peu abusive, nous nous contenterons pour l'heure d'affirmer que ce modèle est uniquement valable pour de faibles concentrations, c'est-à-dire pour des solutions fortement diluées. Dans le cas

contraire, les effets non linéaires dus à l'interaction entre les composants doivent être pris en compte. En d'autres termes, il convient de corriger les distortions non linéaires observées sur le spectre mesuré avant d'appliquer l'une des méthodes de décomposition. Les travaux rapportés dans ce mémoire s'inscrivent dans cette démarche.

Ces phénomènes sont connus et modélisés. Cependant, leurs effets sont trop souvent négligés et les méthodes de correction sont contraignantes ou peu efficaces. Nous proposons dans ce rapport quelques approches simples permettant de limiter les effets non linéaires sur l'analyse des mélanges fluorescents. Mais auparavant, il convient de poser les bases du problème et d'effectuer un large tour d'horizon des différentes techniques d'analyse des signaux de fluorescence.

Le premier chapitre de cette thèse n'a comme autre objectif que de définir la notion de spectre de fluorescence. Nous rappelons dans un premier temps le mécanisme d'absorption de la lumière par la matière, à l'origine, entre autres, du phénomène de fluorescence. Nous aboutirons alors à la définition du spectre d'absorption et à la loi de Beer-Lambert. Cette relation, fondamentale pour la suite de notre étude, permet de décrire les variations de l'intensité de la lumière transmise à travers un milieu absorbant. Nous décrirons ensuite le mécanisme de fluorescence, quantifié par la notion de spectre d'émission. Par ailleurs, notre travail répond à des contraintes physiques très précises dues au système de mesure. Une part importante de ce chapitre est donc consacrée au fonctionnement du spectrofluorimètre (en absorption et en émission), faisant apparaître clairement la distinction entre le signal mesuré et la grandeur physique correspondante. Nous distinguons également les trois types de spectres de fluorescence mesurables : les spectres d'excitation, les spectres d'émission et les spectres d'excitation-émission. Nous présentons enfin les principaux phénomènes venant altérer ces mesures. Nous décrivons d'abord rapidement les mécanismes de diffusion de la lumière, en insistant sur leurs effets sur les spectres. Nous évoquons ensuite la notion d'effets d'écrans, effets non linéaires intrinsèques au système de mesure et qui ne peuvent plus être négligés lorsque l'absorbance de la solution est trop importante. Nous donnons de nombreux exemples justifiant la prise en compte de ces effets dans l'analyse des signaux de fluorescence.

Le second chapitre est consacré à la description des méthodes d'analyse des signaux de fluorescence issus de mélanges. La totalité de ces approches reposent sur le modèle linéaire. En outre, les méthodes de corrections de l'effet d'écran, qu'il s'agisse des approches classiques ou de celles que nous proposons, ne sont qu'un préalable à l'application de ces algorithmes (nous parlerons à ce propos de « linéarisation » des spectres). Il est donc indispensable d'en connaître les avantages et les limites. Cet état de l'art est divisé en trois parties. Nous traitons d'abord le cas de données bilinéaires, dites du premier ordre, à travers quelques unes des nombreuses méthodes consacrées. En marge de ces méthodes nous rappelons les principes de l'analyse en composantes principales (ACP), très largement employée dans le domaine. Dans une seconde partie présentant les algorithmes trilineaires, nous insisterons plus encore sur la description de l'algorithme de décomposition PARAFAC. Celui-ci est particulièrement adapté à notre problème et fait aujourd'hui référence. Enfin nous présentons dans la dernière section le traitement classique de l'effet d'écran.

La correction des effets d'écran fera plus largement l'objet des deux chapitres suivants. Nous montrons en préambule du troisième chapitre l'inefficacité de PARAFAC lorsqu'il s'agit de traiter des spectres présentant un fort effet d'écran. Nous établissons ensuite un modèle approximatif des effets d'écran adapté au traitement des spectres 3D des mélanges. La formulation que nous proposons est proche du modèle classique des effets d'écran mais nous en apportons une mise en équations simple tout en soulignant ses limites théoriques. Enfin, la mise à l'épreuve de ce modèle est effectuée sur des spectres issus de différents mélanges parfaitement connus. Nous montrons par l'exemple que ce modèle décrit correctement les fortes distortions non linéaires présentes dans des solutions très concentrées.

Nous proposons dans le quatrième chapitre de ce mémoire trois méthodes de correction de l'effet d'écran. Une méthode de linéarisation, dite « correction par faible dilution » nécessitant une mesure supplémentaire, et deux méthodes purement numériques. Ces deux dernières ne sont pour l'heure que des approches mais les fondements théoriques sont établis et les premiers résultats sont prometteurs. En revanche, notre première démarche a été testée sur trois jeux de données expérimentales représentatifs, composés de mélanges connus de deux ou trois fluorophores ou d'échantillons de matière organique. Nous lui consacrons donc la plus large partie de ce chapitre.

Enfin, le cinquième et dernier chapitre de ce mémoire propose deux exemples d'applications du traitement numérique des spectres de fluorescence au suivi environnemental de la matière organique. En particulier, la première application illustre la chaîne de traitement complète dans le cas d'une étude réelle, de la mesure des spectres à l'analyse géochimique des résultats. Nous insisterons à cette occasion sur le traitement de l'effet d'écran par notre algorithme et sur la décomposition des spectres linéarisés par PARAFAC.

Chapitre 1

Spectroscopie de fluorescence

1.1 Rappels historiques

Le phénomène de luminescence caractérise l'émission de lumière résultant de l'excitation électronique de la matière, par opposition au phénomène d'incandescence qui découle d'une augmentation de la température. Différents types de luminescence peuvent être observés suivant le mode d'excitation de la matière. Ainsi, lorsque l'excitation se fait par absorption de photons, on parle de photoluminescence. Ce phénomène englobe la fluorescence, la fluorescence retardée et la phosphorescence que nous distinguerons plus loin. Notons enfin, qu'il existe un autre type très important d'interaction lumière-matière : la diffusion, mécanisme sur lequel nous reviendrons.

Ces phénomènes facilement observables dans la nature sont connus et utilisés depuis de nombreux siècles. Le terme phosphorescent (« qui transporte de la lumière ») désignait au Moyen-Age les objets brillant dans l'obscurité après exposition à la lumière. En ce qui concerne la fluorescence, bien que le phénomène soit connu depuis le 16^{ème} siècle, le terme ne fut introduit qu'en 1852 par le physicien George Gabriel Stokes, à l'occasion d'un célèbre article [76]. Stokes y démontrait expérimentalement que le phénomène consistait en une émission de lumière résultant de l'absorption de lumière. Il postula également que la lumière est émise à une longueur d'onde supérieure à celle de la lumière absorbée. Cette conjecture, caractéristique des phénomènes de photoluminescence, est toujours connue sous le nom de loi de Stokes, bien qu'elle fût formulée pour la première fois par Edmond Becquerel dix ans auparavant [10].

Le phénomène de fluorescence met donc en jeu deux mécanismes successifs : l'absorption d'un photon par la molécule fluorescente (fluorophore) suivie de l'émission d'un photon de plus basse énergie par ce même fluorophore. Depuis l'expérience de Stokes, ces deux mécanismes furent abondamment étudiés et sont maintenant théorisés et quantifiés. En particulier, par les spectres d'absorption et les spectres de fluorescence. Les travaux présentés dans ce mémoire sont basés sur ces deux notions dont nous pouvons donner les définitions intuitives suivantes :

Le spectre d'absorption (resp. d'émission) d'un composant peut être entendu comme une mesure de la capacité du composant à absorber (resp. émettre) des photons à différentes longueurs d'onde. Ce chapitre sera donc consacré à préciser ces définitions. Nous nous limiterons ici à la spectroscopie des composants en solution, et ce, dans le

domaine UV-visible.

1.2 Mécanisme d'absorption

La lumière est constituée d'ondes électromagnétiques caractérisées par leur longueur d'onde λ ou leur fréquence ν ($\nu = c/\lambda$). En mécanique quantique, on associe au rayonnement d'une onde monochromatique (λ, ν), un corpuscule (photon) de masse nulle et d'énergie inversement proportionnelle à la longueur d'onde : $E = h\nu$ où h est la constante de Planck. Nous allons voir comment ces quanta d'énergie peuvent être absorbés par la molécule.

1.2.1 Niveaux d'énergie et transitions électroniques

Nous savons depuis le début du vingtième siècle et l'avènement de la théorie quantique qu'il est impossible de connaître exactement le mouvement des électrons d'un atome au sens classique (vecteurs position et quantité de mouvement). Néanmoins, il est possible, pour chaque électron de l'atome, de connaître la région de l'espace où celui-ci a une très forte probabilité de présence. Cette région constitue l'orbitale atomique de l'électron. Il existe différents types d'orbitales, caractérisés par l'énergie électrostatique des électrons présents sur l'orbitale. A chaque orbitale correspond un niveau d'énergie électronique. Les niveaux d'énergie électronique accessibles sont quantifiés et notés S_n . L'ensemble des orbitales atomiques possibles est donc dénombrable. La répartition des électrons sur les différentes orbitales détermine la configuration électronique de l'atome. Cette répartition est régie par des lois que nous ne détaillerons pas ici. Nous retiendrons simplement que sans apport d'énergie extérieure, l'atome se trouve dans l'état fondamental S_0 . Cet état est stable et correspond à la configuration électronique d'énergie minimale. L'atome quitte cet état s'il reçoit une énergie suffisante. Dans ce cas, un électron d'une orbitale périphérique peut passer à une orbitale inoccupée de plus haute énergie. Lors de cette transition électronique, l'atome se trouve dans un état excité instable.

Les différents atomes d'une molécule mettent en commun des électrons pour former des liaisons chimiques. Celles-ci correspondent à une combinaison d'orbitales atomiques, définissant ainsi une orbitale moléculaire. On peut extrapoler au niveau moléculaire certaines des notions précédentes et conserver les mêmes notations. Un apport d'énergie suffisant permet des transitions électroniques entre orbitales moléculaires, la molécule passant alors de l'état stable à un état excité ou d'un état excité à un état excité de niveau d'énergie supérieure.

Par ailleurs, des niveaux intermédiaires d'énergie vibrationnelle et rotationnelle, liés aux mouvements relatifs des noyaux, s'intercalent entre les différents niveaux d'énergie électroniques. On notera S_{n-k} le niveau vibrationnel k de S_n . Il est important de noter les ordres de grandeur très différents des transitions entre les niveaux. Les transitions entre deux niveaux successifs sont de l'ordre de 1 à quelques eV pour une transition électronique, de 10^{-2} à 0.1 eV pour une transition vibrationnelle et de 10^{-3} eV pour une transition rotationnelle. En revanche l'écart énergétique entre deux ni-

veaux intermédiaires de même nature est constant.

Le rapport entre le nombre de molécules N_i et N_j se trouvant respectivement aux niveaux d'énergie E_i et E_j de dégénérescence g_i et g_j suit la loi statistique de Boltzmann : $N_i/N_j = \frac{g_i}{g_j} e^{-\frac{E_i - E_j}{kT}}$ où k est la constante de Boltzmann ($k = 1.3810^{-23} J.K^{-1}$) et T la température en Kelvin. En conséquence, d'après les ordres de grandeur précédents, la probabilité de trouver un nombre significatif de molécules dans un niveau excité est quasiment nulle à température ambiante.

L'absorption d'un photon de longueur d'onde λ par la molécule autorise une transition électronique si son énergie $\frac{hc}{\lambda}$ est suffisante. La molécule passe alors de l'état S_{0-0} à un état excité S_{n-k} . La différence d'énergie entre les deux états est égale à l'énergie du photon absorbé. Comme à chaque valeur de l'énergie correspond une longueur d'onde, ce mécanisme, illustré par le diagramme de Jablonski (figure 1.1), nous conduit à la notion d'étalement du spectre d'absorption. En effet, en supposant que nous imposions les deux contraintes suivantes à la molécule : a) seuls les niveaux d'énergie électronique lui sont accessibles. b) Elle n'en possède que deux : l'état fondamental et un état excité. Dans ce cas, elle ne peut absorber que des photons transportant une quantité d'énergie bien précise, égale à l'écart entre ces deux niveaux. Suivant ces deux restrictions, son spectre d'absorption ne serait formé que d'une seule raie correspondant à cette énergie. Si l'on supprime la seconde contrainte, du fait de l'écart énergétique important entre les niveaux électroniques, on obtient un spectre de raies discontinues. Cependant, les niveaux d'énergie intermédiaires sont très proches les uns des autres, cela assure la quasi continuité du spectre.

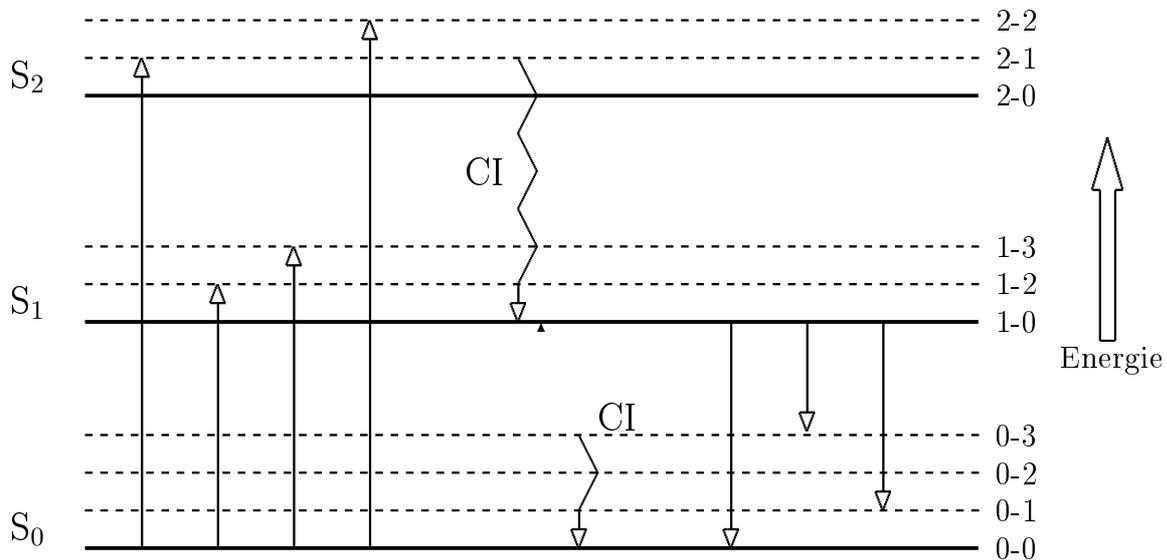


FIG. 1.1 – Diagramme de Jablonski

Exemples de transition électronique vers des niveaux d'énergie supérieurs après absorption de photons. L'étalement du spectre d'absorption est constitutif de l'existence des niveaux d'énergie intermédiaires entre les différents niveaux électroniques.

1.2.2 Loi de Beer-Lambert et spectres d'absorption

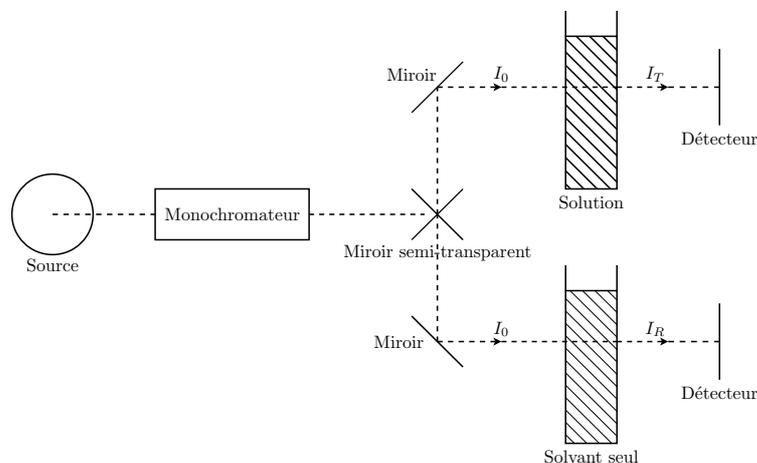


FIG. 1.2 – Principe de mesure du spectre d'absorption

Sur ce schéma de principe, le signal d'intérêt I_T et le signal de correction, I_R sont mesurés simultanément. En pratique, on effectue généralement les deux mesures successivement.

L'amplitude du spectre d'absorption (ou absorbance), à une longueur d'onde donnée, $A(\lambda)$ est mesurée expérimentalement à l'aide d'un spectrophotomètre. Le dispositif est représenté sur la figure 1.2. L'ensemble source/monochromateur permet de balayer en excitation une partie de la gamme UV-visible. L'intensité lumineuse transmise, I_T , est mesurée par le système de détection. Une étude simplifiée de la propagation du faisceau dans la cuve permet de relier l'intensité transmise à l'absorbance. Chaque molécule possède une section efficace d'absorption σ (cm^{-2}) qui dépend de la longueur d'onde.

Soit une couche élémentaire de solution d'épaisseur dl et contenant dN molécules on a :

$$dN = N_a c S dl \quad (1.1)$$

où S est la section efficace du faisceau excitateur, c la concentration de la solution et N_a le nombre d'Avogadro. La section efficace d'absorption de la couche élémentaire est donc égale à σdN . La probabilité de capture d'un photon est donc simplement égale à la fraction de la lumière absorbée par la couche :

$$-\frac{dI}{I} = \frac{\sigma dN}{S} = N_a \sigma dl \quad (1.2)$$

On définit alors le coefficient d'absorption : $\alpha = N_a \sigma c$. En intégrant cette relation sur une couche de solution d'épaisseur l , on obtient la loi de Beer-Lambert reliant l'intensité lumineuse sortant de la couche, $I_T(\lambda)$, à l'intensité du faisceau excitateur $I_0(\lambda)$ et au coefficient d'absorption :

$$I_T(\lambda) = I_0(\lambda) e^{-l\alpha(\lambda)} \quad (1.3)$$

| Composant | ε ($L.mol^{-1}.cm^{-1}$) | Composant | ε ($L.mol^{-1}.cm^{-1}$) |
|--------------------|--|-----------------|--|
| Benzene | 200 | Acridine | 12000 |
| Phenol | 2000 | Biphenyl | 16000 |
| Carbazole | 4200 | Bianthryl | 24000 |
| 1-Naphtol | 5400 | Acridine orange | 30000 |
| Indole | 5500 | Perylene | 34000 |
| Fluorene | 9000 | Eosin Y | 90000 |
| Anthracene | 10000 | Rhodamine B | 105000 |
| Sulfate de quinine | 10000 | | |

TAB. 1.1 – Exemples de coefficients d'absorption molaire de différents fluorophores. Les valeurs sont données à la longueur d'onde qui correspond à l'absorption maximale de la bande de plus bas niveau d'énergie. Par ailleurs, elles sont approximatives car elles dépendent du solvant employé.

En écrivant l'équation précédente sous la forme suivante :

$$I_T(\lambda) = I_0(\lambda)10^{-l\frac{\alpha(\lambda)}{\ln(10)}} \quad (1.4)$$

celà nous permet d'introduire la définition de l'absorbance A :

$$A(\lambda) = \log\left(\frac{I_0(\lambda)}{I_T(\lambda)}\right) = l\frac{\alpha(\lambda)}{\ln(10)} \quad (1.5)$$

Le traitement du signal de fluorescence étant basé sur la loi de Beer-Lambert, il sera souvent utile par la suite de faire apparaître la concentration dans la relation précédente. On définit pour cela le coefficient d'absorption molaire $\varepsilon = \frac{\alpha(\lambda)}{c}$ (généralement exprimé en $L.mol^{-1}.cm^{-1}$) donc :

$$I_T(\lambda) = I_0(\lambda)e^{-lc\varepsilon(\lambda)} = I_0(\lambda)e^{-2.3A(\lambda)} \quad (1.6)$$

Le tableau 1.1 donne l'ordre de grandeur du coefficient d'absorption molaire pour plusieurs fluorophores classiques.

Cette définition du spectre d'absorption entraîne plusieurs remarques. Elle fait intervenir des paramètres expérimentaux (c, l). De plus, le coefficient d'absorption dépend du solvant. En conséquence, parler du spectre d'absorption d'un fluorophore est abusif. Il s'agit du spectre d'absorption d'une solution mesurée dans des conditions données. Par ailleurs, dans le cas d'un mélange de différents composants (ε_n, c_n), les absorbances s'additionnent et la loi de Beer-Lambert peut s'écrire :

$$I_T(\lambda) = I_0(\lambda)e^{-\sum_n lc_n\varepsilon_n(\lambda)} \quad (1.7)$$

Nous exposerons et proposerons dans les chapitres 2 et 4 différentes méthodes permettant de déterminer les spectres d'absorption de chaque composant d'un mélange. Compte tenu de la première remarque et, afin d'estimer la qualité des résultats de chaque méthode, il est nécessaire de travailler dans des conditions pour lesquelles les spectres d'absorption de référence ont déjà été mesurés. Nous verrons qu'il en sera de

même pour les spectres de fluorescence.

La méthode de mesure de l'absorbance décrite précédemment est généralement complétée par une procédure de correction afin de s'affranchir de l'absorbance de la cuve et du solvant. En effet, l'absorbance mesurée est celle de l'ensemble cuve-solvant-fluorophore. Il est donc recommandé d'effectuer une seconde mesure en remplaçant la solution par le solvant seul comme sur la figure 1.2. L'intensité I_R mesurée permet de déterminer l'absorbance corrigée du fluorophore A_c :

$$A_c(\lambda) = \log \left(\frac{I_R(\lambda)}{I_T(\lambda)} \right) \quad (1.8)$$

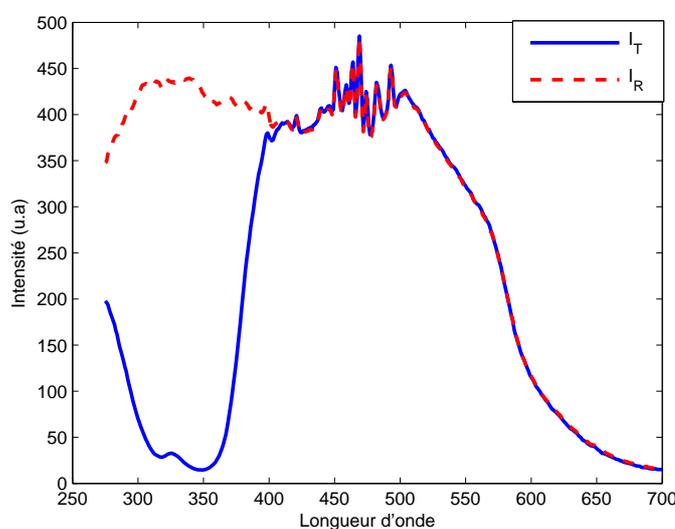


FIG. 1.3 – Mesures de la lumière transmise pour une solution de sulfate de quinine très concentrée (109 ppm) et pour le solvant seul

Nous utiliserons souvent par la suite, le développement au premier ordre de la loi de Beer-Lambert. La valeur de l'absorbance pourra être utilisée afin d'estimer la pertinence de cette approximation. Ce point important sera développé en fin de chapitre à la section 1.5.

Les figures 1.3 et 1.4 présentent les intensités $I_T(\lambda)$ et $I_R(\lambda)$ mesurées sur une solution de sulfate de quinine ainsi que le spectre absorption qui en découle. On notera les valeurs très élevées de l'absorbance.

1.3 Emission de fluorescence

1.3.1 Mécanismes de désactivation

Après l'absorption d'un photon, la molécule se trouve dans un état énergétique excité instable. Le retour spontané à l'état fondamental peut se faire selon différents

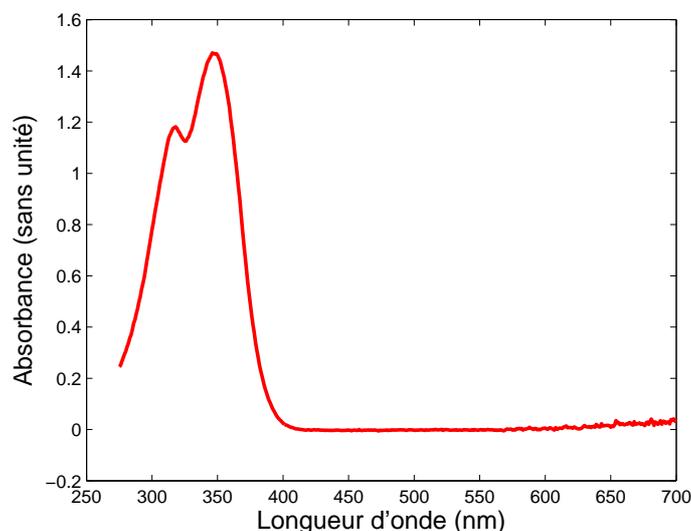


FIG. 1.4 – Spectre d'absorption du sulfate de quinine

processus de désactivation résumés sur le diagramme de Jablonski de la figure 1.5. Suivant leur nature, on parle de chemin chimique ou de chemin physique. La fluorescence n'est qu'un des chemins de désactivation possibles. Il convient donc de décrire, même succinctement, les autres mécanismes.

Conversion interne. Le processus de conversion interne désigne la transition non radiative vers un niveau d'énergie électronique inférieur. Ce retour au plus bas niveau d'énergie vibrationnelle se fait par relaxation vibrationnelle. En solution, cet excès d'énergie peut également être dissipé par collision avec les molécules de solvant voisines. Ce processus est le chemin le plus probable de retour à l'état excité S_1 . La transition de S_1 à S_0 par conversion interne est également possible mais dans ce cas, ce n'est plus forcément le chemin prépondérant.

Fluorescence. Le phénomène de fluorescence traduit la relaxation de l'état excité vers l'état fondamental S_0 par émission de photons. A l'exception de quelques éléments chimiques, cette relaxation n'a lieu qu'à partir de S_1 . De plus l'état excité S_{1-k} perd rapidement l'énergie vibrationnelle excédentaire. L'émission de photon se fait donc presque exclusivement à partir de l'état excité S_{1-0} vers les différents niveaux d'énergie vibrationnelle de l'état fondamental (règle de Kasha). Comme dans le cas de l'absorption, l'écart entre S_{1-0} et S_{0-0} ainsi que le grand nombre de niveaux intermédiaires se traduisent par un étalement spectral et une quasi-continuité du spectre d'émission de fluorescence. Cela explique l'origine des observations de Stokes et la loi éponyme. En effet, la plus grande énergie d'émission (et donc la plus petite longueur d'onde du spectre d'émission) ne peut être qu'inférieure ou égale à l'énergie entre S_{1-0} et S_{0-0} ce qui correspond à la plus petite énergie d'excitation (et donc à la plus grande longueur d'onde du spectre d'absorption). Cependant en pratique, on observe dans la

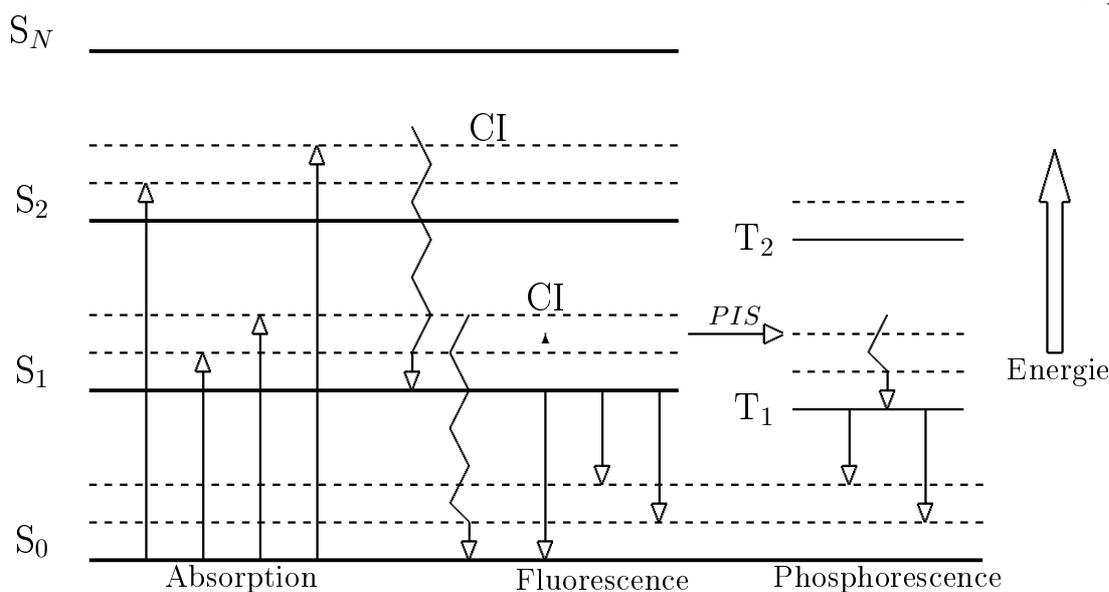


FIG. 1.5 – Diagramme de Jablonski complet

Les principaux mécanismes de désactivation (retour à l'état fondamental) après absorption sont illustrés : conversion interne, fluorescence et phosphorescence. L'étalement du spectre de fluorescence s'explique par les différents niveaux d'énergie vibrationnelle de l'état fondamental S_0

plupart des cas un léger chevauchement des deux spectres. Cela est dû au fait qu'à température ambiante, conformément aux lois statistiques de Boltzmann, une petite fraction de molécules se trouvent avant l'absorption d'un photon à un état d'énergie vibrationnelle supérieur à S_{0-0} , élargissant ainsi le spectre d'absorption vers les longueurs d'onde supérieures.

Les écarts entre les niveaux d'énergie vibrationnelle sont similaires pour chaque niveau d'énergie électronique. Par conséquent, le spectre d'émission de fluorescence a une allure proche de celle du spectre d'absorption, mais symétrique par rapport à l'énergie (longueur d'onde) de transition entre S_{0-0} et S_{1-0} . On parle de spectre miroir.

Nous terminerons par une remarque importante pour la suite. D'après la règle de Kasha, l'allure intrinsèque du spectre d'émission de fluorescence ne dépend pas de la longueur d'onde d'excitation.

Passage inter-système, phosphorescence. Lors du passage très rapide de l'état fondamental à l'état S_1 après excitation d'un électron, celui-ci n'a pas le temps de changer de spin. Cependant, une fois l'état singulet S_1 atteint, la molécule peut passer à un état triplet T_1 (passage inter-système). Le retour à l'état fondamental peut se faire par transition non radiative ou par émission de photon. Ce dernier cas porte le nom de phosphorescence. Il se distingue de la fluorescence par sa durée de vie généralement beaucoup plus longue et son domaine de longueur d'onde plus élevé car le plus bas niveau d'énergie vibrationnelle de T_1 est plus faible que celui de S_1 . On retiendra essentiellement qu'en solution, ce chemin est minoritaire par rapport aux mécanismes de désexcitation non radiatifs et que par conséquent, il est rarement observé.

Fluorescence retardée. Il est possible à partir de l'état T_1 de retourner à l'état S_1 . Pour cela la molécule doit gagner de l'énergie soit par collision, soit par absorption d'un autre photon, soit par transfert d'énergie entre deux états T_1 . La fluorescence observée apparaît alors plus tardivement. En revanche, les longueurs d'onde d'émission sont identiques à celles de la fluorescence spontanée.

Transfert d'énergie. L'énergie absorbée par la molécule, peut aussi être transmise à une molécule qui n'absorbe pas de photon. On a alors un phénomène de transfert d'énergie, qui peut se produire à distance par interaction dipôle-dipôle ou quadripôle-quadripôle ou par collision avec un accepteur.

On peut associer à chacun de ces chemins un rendement quantique ϕ_i défini par le rapport entre le nombre de molécules désactivées par le chemin i et le nombre total de photons absorbés. Chaque photon absorbé produit un état excité et chaque état excité retourne naturellement à l'état fondamental. Ainsi, ϕ_i représente la probabilité de désactivation par le chemin i et donc, $\sum_i \phi_i = 1$. Dans toute la suite, on notera ϕ le rendement quantique de fluorescence.

1.3.2 Spectres de fluorescence d'une solution

Nous avons précédemment donné un sens physique à l'étalement et l'amplitude du spectre d'émission de fluorescence. Nous allons maintenant en décrire le principe de mesure. Le spectrofluorimètre est l'outil de mesure classique des spectres de fluorescence d'une solution. On distingue trois types de spectre de fluorescence : les spectres d'excitation, d'émission et d'excitation-émission. Les spectres d'un fluorophore lui sont spécifiques. Par conséquent, l'analyse spectroscopique de fluorescence est basée sur le traitement de ces signaux.

Les principes de mesure présentés dans cette section sont valables pour tout type de solution. Cependant, nous insisterons sur l'étude des solutions dont l'absorbance est suffisamment faible pour justifier l'approximation linéaire de la loi de Beer-Lambert, diluées dans un solvant non absorbant. Bien que l'objet de notre travail soit justement de nous affranchir de telles contraintes pour estimer les spectres réels, ce cas nous intéresse particulièrement pour deux raisons. La mesure de spectres fidèles aux spectres réels nécessite ce cadre restrictif. Or, pour être validé expérimentalement nos résultats devront être comparés à ces spectres de référence. De plus, la plupart des méthodes actuelles d'analyse des mélanges reposent sur cette approximation.

1.3.2.1 Fonctionnement du spectrofluorimètre.

La figure 1.6 montre le schéma de principe d'un spectrofluorimètre standard. La source de lumière excitatrice est formée d'une lampe et d'un monochromateur. La lampe au xénon permet une excitation relativement stable de 250 nm jusqu'à l'infrarouge. En dessous de 250 nm, l'intensité de la lampe s'écroule ce qui a tendance à dégrader le rapport signal sur bruit. Le monochromateur d'excitation permet de

sélectionner une longueur d'onde d'excitation avec une certaine précision. Le faisceau excitateur est alors dirigé vers la cuve contenant la solution. Cependant, une partie est déviée vers un système de comptage de photons qui servira à la correction des spectres d'excitation. Le signal d'émission de fluorescence est collecté et amplifié par le photomultiplicateur, placé derrière un monochromateur, à 90° du faisceau incident afin que celui-ci ne perturbe pas la mesure. Les monochromateurs sont motorisés et balayent les gammes de longueurs d'onde définies par l'utilisateur. Les différents réglages des monochromateurs permettent de mesurer les trois types de spectres de fluorescence.

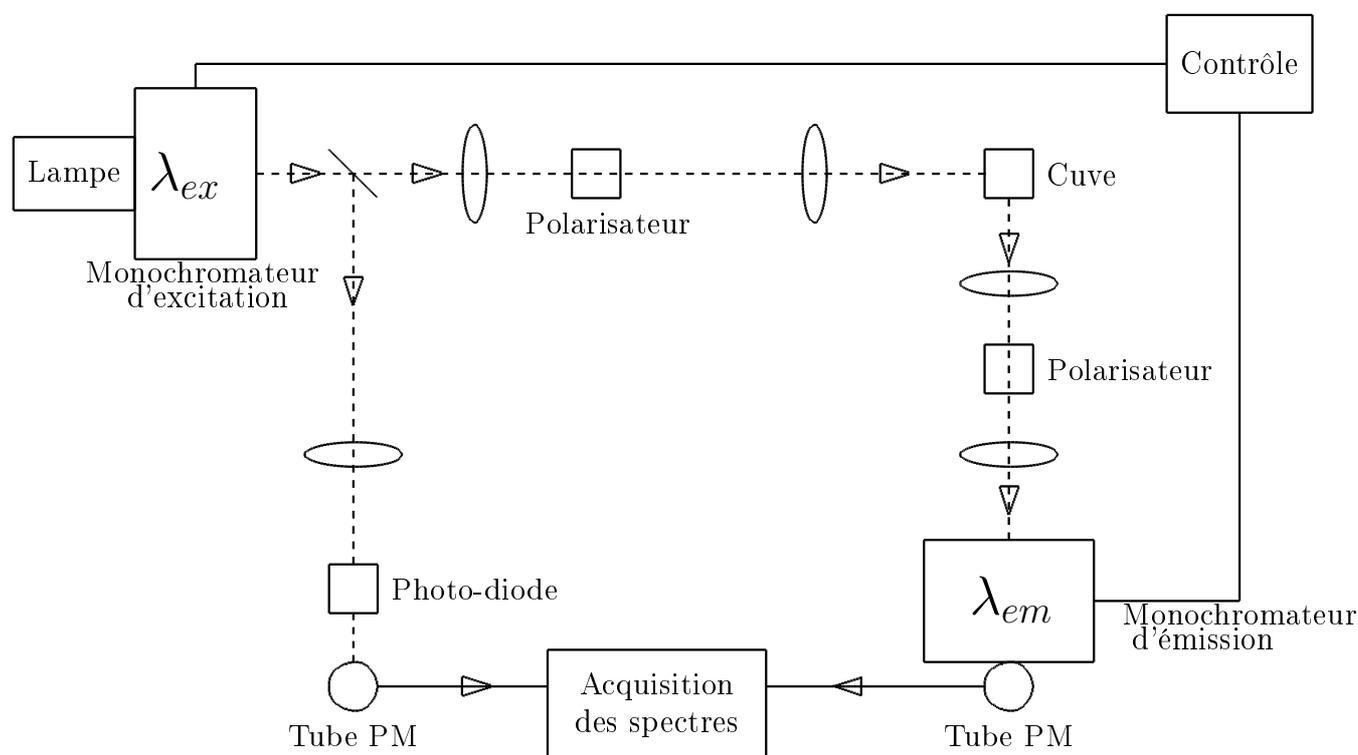


FIG. 1.6 – Schéma de principe du spectrofluorimètre

Les traits pointillés représentent le chemin optique, les traits pleins, les liaisons entre les composants de l'appareil.

1.3.2.2 Mesure du spectre d'émission.

Le spectre d'émission de fluorescence d'un fluorophore $F(\lambda)$ reflète la distribution de probabilité des différentes transitions entre S_{1-0} et les différents niveaux d'énergie vibrationnelle de S_0 . A chaque différence d'énergie correspond la longueur d'onde des photons émis. $F(\lambda)$ représente donc également le rapport entre le nombre de photons émis à la longueur d'onde λ et le nombre total de photons absorbés.

$$\int_0^{\infty} F(\lambda) d\lambda = \phi \quad (1.9)$$

Dans le cas le plus simple, le spectrofluorimètre mesure l'intensité de fluorescence à différentes longueurs d'onde d'émission pour une longueur d'onde d'excitation donnée λ_e .

Pour une solution faiblement concentrée ne contenant qu'un seul fluorophore, le spectre mesuré I_F est proportionnel au spectre réel et à l'intensité absorbée I_A :

$$I_F(\lambda_f) = kI_A F(\lambda_f) \quad (1.10)$$

$I_{A\lambda_e}$ représente l'intensité de la lumière absorbée à la longueur d'onde λ_e . L'équation 1.10 représente une approximation linéaire de la loi de Beer-Lambert dont les limites seront discutées à la fin de ce chapitre et au chapitre 3. Par ailleurs, le coefficient de proportionnalité k dépend de trop de paramètres liés au dispositif de mesure pour pouvoir être déterminé simplement. Cela nous serait d'ailleurs inutile, l'important étant que la forme du spectre d'émission soit conservée par la mesure.

Le rendement quantique de fluorescence dépend du fluorophore. Sa valeur, souvent inconnue, jouera un rôle dans le traitement numérique des mélanges. C'est pourquoi, on introduit le spectre d'émission normé $\gamma(\lambda_f) = \frac{F(\lambda_f)}{\phi}$. On a alors :

$$I_F(\lambda_f) = kI_A \phi \gamma(\lambda_f) \quad (1.11)$$

Dans le cas d'un mélange faiblement concentré de N fluorophores, on mesure une combinaison linéaire des spectres individuels :

$$I_F(\lambda_f) = k \sum_{n=1}^N I_{A_n} \phi_n \gamma_n(\lambda_f) \quad (1.12)$$

1.3.2.3 Spectres d'excitation.

Le réglage inverse du spectrofluorimètre permet de mesurer l'intensité de fluorescence I_{EF} à différentes longueurs d'onde d'excitation pour une longueur d'onde d'émission donnée. Pour une solution peu concentrée contenant un seul fluorophore, cette mesure constitue le spectre d'excitation du composant étudié.

Reécrivons l'équation 1.10 avec la longueur d'onde d'émission fixée et en supposant que toute l'épaisseur de la solution l contribue uniformément au signal de fluorescence :

$$I_{EF}(\lambda_e) = kI_A(\lambda_e) F_{\lambda_f} \quad (1.13)$$

où k est une constante. Par définition,

$$I_A(\lambda_e) = I_0(\lambda_e) - I_T(\lambda_e) \quad (1.14)$$

et,

$$I_T(\lambda_e) = I_0(\lambda_e) e^{-lc\varepsilon(\lambda_e)} \quad (1.15)$$

il vient :

$$I_{EF}(\lambda_e) = kI_0(\lambda_e)F_{\lambda_f}(1 - e^{-lc\varepsilon(\lambda_e)}) \quad (1.16)$$

Par ailleurs, nous verrons plus loin que l'on peut facilement s'affranchir des fluctuations de I_0 de la source en fonction de la longueur d'onde. On a alors $I_0(\lambda_e) = I_0$. Pour une solution faiblement concentrée, le développement au premier ordre de l'équation 1.16 donne :

$$I_{EF}(\lambda_e) \simeq kI_0lcF_{\lambda_f}\varepsilon(\lambda_e) = KA(\lambda_e) \quad (1.17)$$

Dans le cas d'un mélange peu concentré de N fluorophores, cela implique que l'absorbance du mélange est une simple combinaison linéaire des absorbances des N constituants fluorescents :

$$I_{EF}(\lambda_e) = kI_0l \sum_{n=1}^N c_n F_{\lambda_f, n} \varepsilon_n(\lambda_e) \quad (1.18)$$

1.3.2.4 Matrices d'Excitation-Emission de Fluorescence.

En utilisant successivement les deux monochromateurs du spectrofluorimètre, il est possible de mesurer les spectres d'émission pour différentes longueurs d'onde d'excitation. On obtient ainsi la matrice d'émission-excitation de fluorescence (MEEF). L'intensité mesurée à (λ_f, λ_e) peut s'écrire :

$$I(\lambda_f, \lambda_e) = kI_0\phi\gamma(\lambda_f)(1 - e^{-lc\varepsilon(\lambda_e)}) \quad (1.19)$$

soit pour les faibles concentrations,

$$I(\lambda_f, \lambda_e) = kI_0\phi_F\gamma(\lambda_f)lc\varepsilon(\lambda_e) \quad (1.20)$$

Dans le cas d'un mélange de N fluorophores répondant à l'approximation des faibles concentrations, on obtient une relation généralisant celles des deux paragraphes précédents (1.12 et 1.18) :

$$I_{3D}(\lambda_e, \lambda_f) = klI_0 \sum_{n=1}^N \Phi_n \gamma_n(\lambda_f) \varepsilon_n(\lambda_e) c_n \quad (1.21)$$

Un exemple MEEF de matière organique est donné sur la figure 1.7.

1.3.2.5 Correction des spectres

Les composants du spectrofluorimètre présentent des fluctuations liés aux longueurs d'onde utilisées. Ainsi, l'écart entre un comportement idéal, qui serait parfaitement indépendant des longueurs d'onde choisies, et leurs performances réelles nécessitent souvent une correction des mesures.

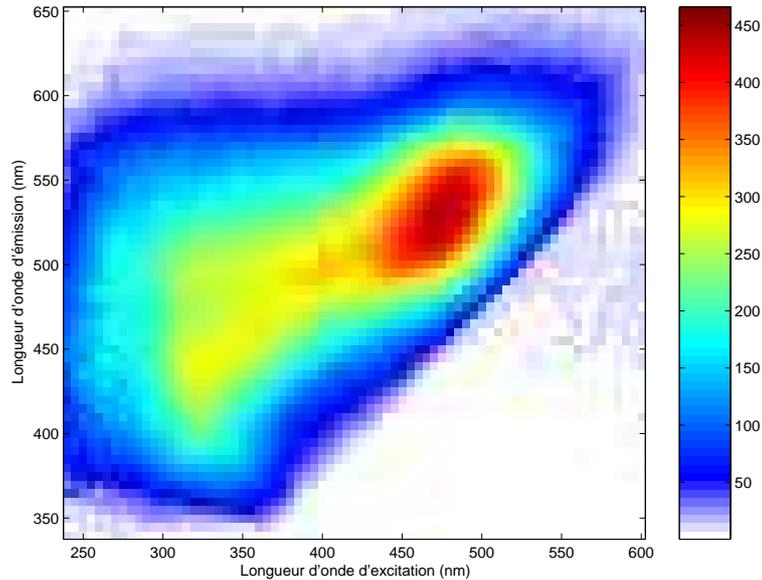


FIG. 1.7 – Exemple de MEEF de matière organique

Nous adoptons dans ce mémoire l'orientation usuelle (longueurs d'onde d'excitation en abscisse).

Correction en excitation La cellule d'excitation est principalement composée de la lampe et d'un monochromateur. L'intensité du faisceau délivrée par cet ensemble varie en fonction de la longueur d'onde d'excitation. Ces fluctuations perturbent donc la mesure d'un spectre d'excitation. Le spectre corrigé, I_{EFc} , sera donc égal au rapport entre le spectre mesuré et une fonction de correction $C_{ex}(\lambda_e)$.

$$I_{EFc}(\lambda_e) = \frac{I_{EF}(\lambda_e)}{C_{ex}(\lambda_e)} \quad (1.22)$$

C_{ex} peut être déterminée grâce au dispositif de comptage de photon. Le dispositif n'est rien d'autre qu'une cuve triangulaire contenant solution de référence (généralement, de la rhodamine) très concentrée et dont l'intensité du spectre d'émission ne dépend pas de la longueur d'onde d'excitation. La solution absorbe la totalité du faisceau exciteur et réémet vers le photomultiplicateur une image des fluctuations de la cellule d'excitation. Il faut noter que le rendement du photomultiplicateur dépend lui aussi de la longueur d'onde et introduit donc une nouvelle distortion. Cependant, lors du calcul de $I_{EFc}(\lambda_e)$ par 1.22, celle-ci est compensée par la distortion due au photomultiplicateur d'émission qui est supposée identique.

Correction en émission De la même manière, la mesure du spectre d'émission doit être corrigée des fluctuations de la réponse des éléments sensibles aux variations de la longueur d'onde d'émission. Il s'agit cette fois de l'ensemble formé par le monochromateur et le photomultiplicateur.

On considère généralement que les fluctuations sont reproductibles, les facteurs correctifs utilisés sont donc toujours les mêmes. La correction est automatique et dans

toute la suite, tous les spectres auxquels nous ferons référence auront été corrigés des fluctuations d'appareil.

1.3.2.6 Remarques générales sur les spectres.

Les relations 1.11 et 1.19 montrent que les spectres d'émission et d'excitation mesurés sont des images fidèles des spectres d'émission et d'absorption réels des solutions faiblement concentrées.

Par ailleurs, la spectroscopie de fluorescence est une technique beaucoup plus sensible que la mesure de l'absorbance. En effet, selon la relation 1.8, la mesure d'une faible absorbance nécessite la comparaison de deux signaux de forte amplitude ce qui entraîne une grande imprécision. En tenant compte de ces deux remarques, le spectre d'absorption de référence du fluorophore étudié sera obtenu par la mesure du spectre d'excitation d'une solution faiblement concentrée, à la longueur d'onde d'émission correspondant au maximum du spectre d'émission.

Dans le cas de l'approximation linéaire, seule l'amplitude du spectre d'émission dépend de la longueur d'onde d'excitation et réciproquement. On utilise la convention suivante pour représenter les MEEF : chaque colonne représente une mesure du spectre d'émission à une longueur d'onde d'excitation donnée et chaque ligne une mesure du spectre d'excitation à une longueur d'onde d'émission donnée.

Lors de la mesure avec le spectrofluorimètre, l'échantillonnage de la matrice en longueur d'onde peut être défini indépendamment en excitation et en émission.

Enfin, on parle généralement de spectre 3D pour désigner une MEEF et de spectres 2D pour les spectres d'émission, d'excitation ou d'absorption.

1.3.2.7 Influence du milieu sur la fluorescence.

De nombreux facteurs peuvent affecter les spectres de fluorescence d'un composant en solution. La partie centrale de notre travail étant consacrée à l'étude des effets de concentrations, nous ne nous étendrons pas ici sur ces facteurs. Les autres principaux paramètres influents sont la température, le pH et le quenching de fluorescence.

La température modifie la répartition statistique de Boltzmann. Une augmentation de température diminue généralement le rendement de la photoluminescence. A l'inverse, White a montré que celui-ci augmente avec le pH de la solution. La forme du spectre de fluorescence dépend également du pH. Le quenching est un effet d'inhibition de la fluorescence par certaines molécules qui affecte le rendement quantique ainsi que la durée de vie de fluorescence.

1.4 La diffusion

L'absorption de photons n'est pas le seul mécanisme d'interaction de la lumière avec la matière. Le phénomène de diffusion caractérise les nombreuses collisions entre le flux de photons et les molécules du milieu. Ce type d'interaction est prépondérant

en solution. Il convient d'en évaluer les effets sur les mesures spectrales. La nature du choc, élastique ou inélastique, conduit alors à distinguer deux types de diffusion.

Diffusion Rayleigh. La plupart des chocs sont élastiques et le photon diffusé ne change donc pas d'énergie après la collision. Les radiations observées à la même longueur d'onde que le faisceau excitateur portent le nom de diffusion Rayleigh. En pratique, on observe un élargissement de la raie spectrale de quelques nanomètres autour de la longueur d'onde d'excitation, élargissement dû aux niveaux d'énergie vibrationnelle.

Diffusion Raman. On observe également sur un spectre de diffusion deux raies de moindre amplitude, situées de part et d'autre de celle de la diffusion de Rayleigh. Elles correspondent aux collisions inélastiques des photons avec le milieu. Il y a diffusion des photons avec perte (ou gain) d'une certaine quantité d'énergie gagnée (ou cédée) par la molécule. La diffusion Raman est caractérisée par ce changement de longueur d'onde. Pour comprendre la position de ces raies, supposons que les photons excitateurs aient une fréquence ν_0 . Après collision inélastique, les photons diffusés ont une énergie $h\nu_d$ et l'on a :

$$h\nu_d - h\nu_0 = E_i - E_j \quad (1.23)$$

E_i : énergie de la molécule avant la collision.

E_j : énergie de la molécule après la collision.

On distingue alors deux cas de figure possibles :

- Certains photons cèdent de l'énergie à la molécule ($E_i < E_j$) et dans ce cas $\nu_d < \nu_0$. On observe alors une raie dite "Stockes" à une longueur d'onde supérieure à celle du faisceau excitateur.
- D'autres photons reçoivent de l'énergie de la molécule ($E_i > E_j$) et dans ce cas $\nu_d > \nu_0$. D'où, une raie dite "Anti-Stockes" à une longueur d'onde inférieure à celle du faisceau excitateur.

Aux températures ordinaires, les molécules sont situées dans l'état électronique fondamental et ont une énergie vibrationnelle minimale. Dans ces conditions, c'est le plus souvent le photon qui cède de l'énergie à la molécule lors d'un choc inélastique. On observera donc rarement la raie Anti-Stockes.

Correction de la diffusion. La figure 1.8 représente la même MEEF que la figure 1.7 mais avant correction de la diffusion. Cela illustre clairement l'impact de ce phénomène qui masque le signal de fluorescence et la nécessité de le corriger. Notons que l'on observe également sur ces spectres les harmoniques du Rayleigh et du Raman correspondant aux longueurs d'onde d'excitation deux fois plus élevées. L'intensité du (spectre) Rayleigh est généralement beaucoup plus élevée que l'intensité maximale du spectre de fluorescence. En revanche, le décalage entre excitation et émission, imposé par la loi de Stockes, minimise les risques de chevauchement avec le spectre d'émission des fluorophores. Ainsi, le Rayleigh peut masquer le signal d'intérêt mais ne perturbe principalement que le bord du spectre.

Le (spectre) Raman pose en quelque sorte le problème inverse. Son amplitude, plus

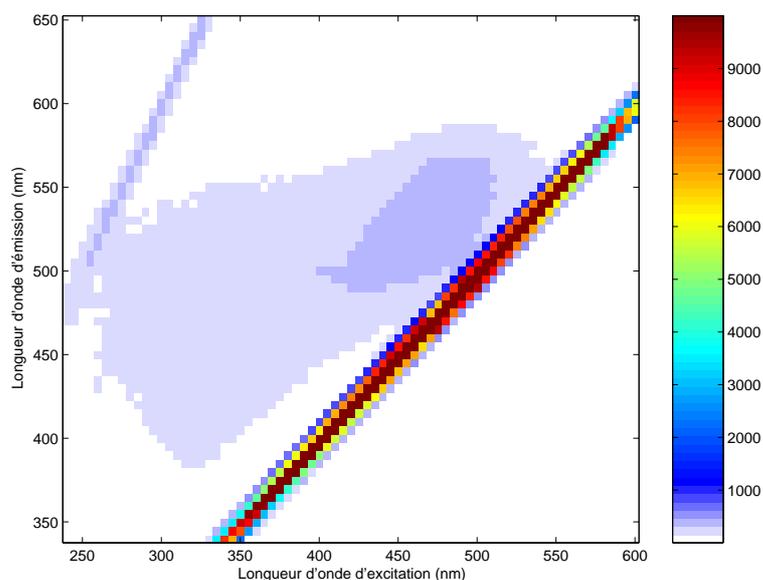


FIG. 1.8 – Impacte de la diffusion sur une MEEF

Pour la concentration employée ici, le spectre de diffusion et ses harmoniques masquent totalement le signal de fluorescence.

modeste, ne masque pas le signal, mais sa position peut entraîner une importante distortion.

Les méthodes d'analyse spectrale des mélanges sont sensibles à ces distorsions. En effet, la majorité d'entre elles sont basées sur l'approximation linéaire de la loi de Beer-Lambert, qui ne peut rendre compte du phénomène non linéaire de diffusion. Par conséquent, la correction de la diffusion doit se faire au préalable. Les spectres de diffusion, répondent à des phénomènes connus et reproductibles. Ils sont donc facilement localisables et peuvent être modélisés. De nombreuses méthodes permettent de filtrer ces pics parasites.

Une approche classique repose sur l'observation suivante, valable pour les solutions usuelles utilisées en spectroscopie : le nombre de molécules du (des) fluorophore(s) est toujours négligeable devant celui du solvant. La contribution de celles-ci au spectre de diffusion observé est donc également négligeable, tout comme l'est celle du solvant au spectre de fluorescence. Il suffit alors de mesurer dans les mêmes conditions le spectre du solvant seul et de le soustraire à celui de la solution.

Par opposition, il existe plusieurs méthodes de filtrage numérique ne nécessitant aucune mesure supplémentaire. L'étude comparative de toutes ces méthodes n'entre pas dans le cadre de cette thèse. Nous avons utilisés pour notre part la méthode de Zepp [82] qui donnait de bons résultats sur nos MEEF.

Il peut arriver que la perturbation due au Raman soit sans conséquence pour la suite de l'étude ou tout simplement qu'elle soit imperceptible. Dans ce cas la méthode la plus simple pour s'affranchir du Rayleigh consiste à « couper » sur les spectres de fluorescence la zone où $\lambda_{ex} > \lambda_{em}$.

Le traitement des spectres 2D fut beaucoup plus simple. Les seules spectres employés

dans ce travail sont ceux de fluorophores purs. Dans ce cas, un choix adéquat de la longueur d'onde d'excitation (pour les spectres d'émission) ou d'émission (pour les spectres d'excitation) autorise un décalage suffisant entre les pics de diffusion et les spectres de fluorescence. Un exemple est donné figure 1.9.

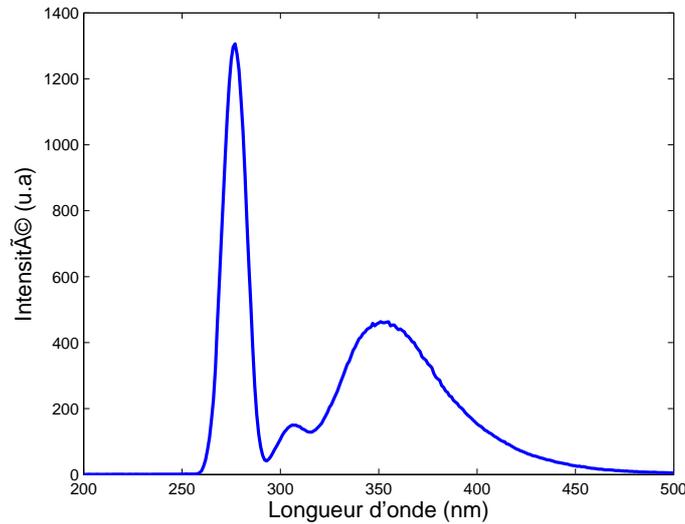


FIG. 1.9 – Exemple de diffusion 2D

distortion du spectre d'émission du triptophane par la diffusion Rayleigh (pic à 275 nm) et la diffusion Raman (pic à 305 nm).

1.5 Phénomènes non linéaires

1.5.1 Mise en évidence du problème

L'équation 1.24 que nous rappelons ici, relie la concentration, les spectres d'excitation et le spectres d'émission des différents fluorophores d'un mélange.

$$I_{3D}(\lambda_e, \lambda_f, \mathbf{c}) = klI_0 \sum_{n=1}^N \Phi_n \gamma_n(\lambda_f) \varepsilon_n(\lambda_e) c_n \quad (1.24)$$

où \mathbf{c} est le vecteur des concentrations du mélange : $\mathbf{c} = \{c_n\}_{n=1\dots N}$. Ce modèle simplifié appréhende la cuve comme un volume élémentaire de solution au sein duquel chaque fluorophore absorbe et émet une certaine quantité de lumière sans interaction possible : une sorte de boîte noire au sein de laquelle s'opère un mélange linéaire des signaux de fluorescence parvenant intégralement au détecteur. Cette relation multilinéaire est à la base des méthodes d'analyse spectrale. Cependant, il s'agit là d'une vision simpliste des phénomènes mesurés.

Définition : On qualifiera ainsi de « non linéaire » tout phénomène influençant la mesure du spectre de fluorescence et qui ne peut pas être modélisé par l'ajout d'un ou

plusieurs composants supplémentaires dans l'approximation linéaire de la loi de Beer-Lambert.

Nous mettons en avant dans cette section les effets non linéaires causés par des concentrations élevées, regroupés sous le nom d'effet d'écran ou de filtre interne. Ces déviations peuvent être dues à l'imprécision du développement au premier ordre dans le cas de solutions à forte absorbance, ou à la réabsorption de la lumière de fluorescence de la solution par la solution elle-même. La prise en compte de ces effets nécessite un modèle physique plus général que celui de l'équation 1.24 et nous renvoyons à ce propos au chapitre 3. En revanche, les autres phénomènes non linéaires cités au paragraphe 1.3.2.7 ne seront pas traités dans ce travail, nous nous sommes donc efforcés d'en limiter les effets. Les résultats expérimentaux comparatifs exposés dans ce mémoire ont été réalisés à pH et températures constants. De même, les effets de la diffusion sont systématiquement corrigés.

Le rôle de la concentration peut être mis en évidence simplement. La courbe de la figure 1.10 représente l'évolution du maximum du spectre d'émission mesurée pour des solutions graduellement concentrées de fluorescéine. L'augmentation de l'intensité de fluorescence avec la concentration semble continuellement freinée et la courbe s'éloigne très rapidement d'un modèle linéaire. On constate également que celle-ci décroît fortement au delà d'une concentration limite. Ce résultat classique [49], semble paradoxal de prime abord puisqu'une augmentation importante de la concentration du fluorophore conduit à une diminution de l'intensité de la fluorescence observée!

Mathématiquement, ces données sont bien modélisables par une fonction de type $f(x) = Gxe^{-gx}$, nous verrons au chapitre 3 qu'un tel modèle se justifie physiquement.

La figure suivante (1.11) témoigne de la non linéarité du spectre d'émission d'un mélange concentré. Il s'agit ici d'un mélange test de sulfate de quinine à 54.38 ppm et de fluorescéine à 41.58 ppm pour lequel nous avons comparé le spectre d'émission avec la combinaison linéaire des spectres d'émissions du sulfate de quinine et de la fluorescéine la plus proche. De même la figure 1.12 met en évidence la déformation du spectre 3D du sulfate de quinine sous l'effet de la concentration.

1.5.2 Limites de l'approximation linéaire et effet d'écran primaire

L'approximation linéaire se justifie pour des absorbances suffisamment faibles. Afin de préciser cela, on peut calculer, par exemple, l'écart entre l'intensité absorbée par une solution, I_A et son approximation linéaire en fonction de l'absorbance de la solution. Rappelons que :

$$I_A = K(1 - e^{-2.3A}) \quad (1.25)$$

La figure 1.13 représente ainsi la déviation relative entre le modèle linéaire $y = 2.3x$ et la fonction $f(x) = (1 - \exp(-2.3x))$ où x représente l'absorbance.

Ainsi, une absorbance de 0.1 entraîne une erreur supérieure à 10%.

Pour une solution ne contenant qu'un seul fluorophore, l'absorbance est le produit de trois paramètres :

- La longueur du chemin optique, fixée par le matériel utilisé et généralement

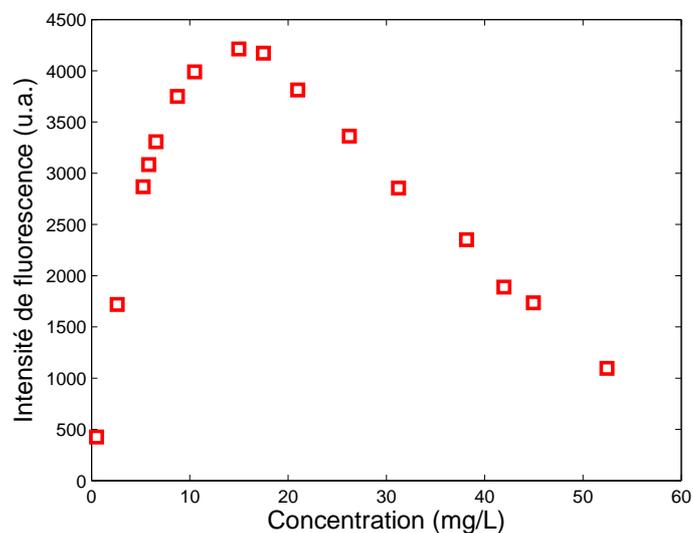


FIG. 1.10 – Evolution du maximum du spectre d'émission mesuré pour différentes solutions de fluorescéine graduellement concentrées

Au delà de 10 mg/L, l'évolution s'écarte franchement du modèle linéaire et à partir de 15 mg/L l'intensité de fluorescence décroît.

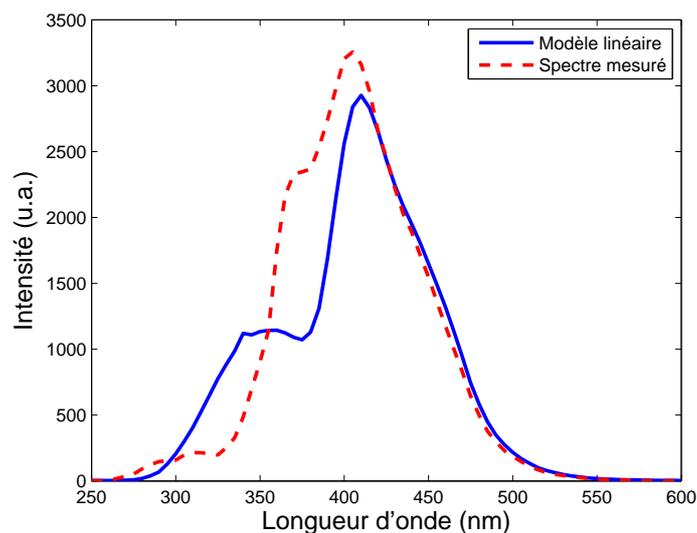


FIG. 1.11 – Modélisation linéaire du spectre d'émission d'un mélange concentré
A la longueur d'excitation utilisée, l'intensité de fluorescence des deux composants est comparable.

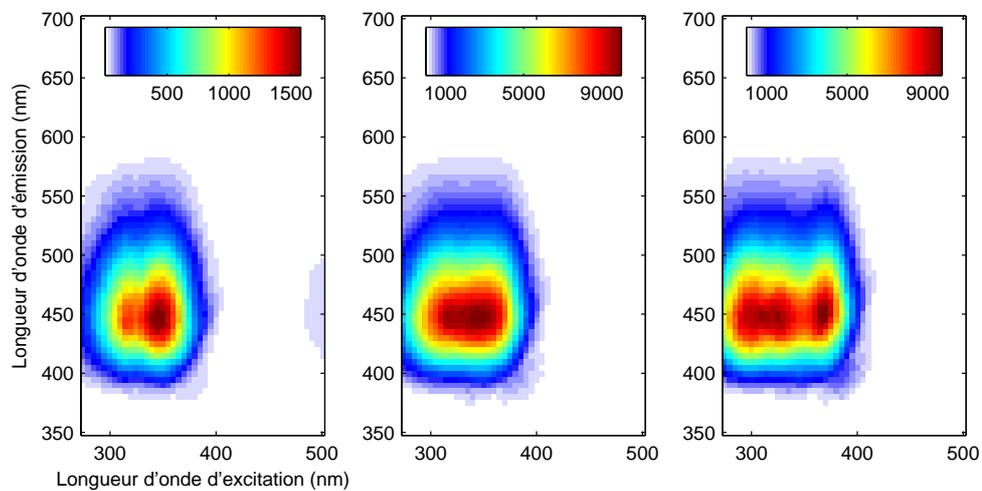


FIG. 1.12 – Influence de la concentration sur le spectre 3D du sulfate de Quinine
De la solution la moins concentrée à la plus concentrée (de gauche à droite : 3.54 ppm, 51.54 ppm
et 108.75 ppm)

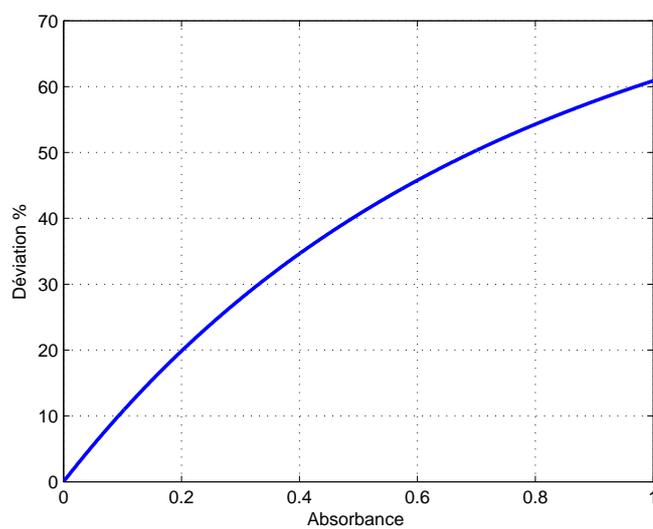


FIG. 1.13 – Déviation relative entre l'intensité absorbée et son approximation linéaire
en fonction de l'absorbance

égale à la largeur de la cuve.

- Le coefficient d'absorption molaire, caractéristique du fluorophore.
- La concentration du fluorophore, liée à la solution.

Pour une concentration et un seuil de linéarité acceptables donnés, la concentration limite n'est donc pas la même pour tous les fluorophores.

Des considérations géométriques imposées par le dispositif de mesure sont alors à prendre en compte. En effet, le système de détection ne reçoit qu'une partie de la lumière émise par la cuve, provenant principalement de sa partie centrale. Plus l'absorbance est élevée, plus l'intensité du faisceau excitateur parvenant au centre de la cuve est faible.

Cet effet non linéaire causé par l'absorption progressive du faisceau excitateur, porte le nom d'effet d'écran primaire (EEP), il explique en partie les résultats expérimentaux précédents.

1.5.3 Phénomène de réabsorption, effet d'écran secondaire

L'autre explication est liée au phénomène de réabsorption ou effet d'écran secondaire (EES). Ce phénomène traduit l'absorption de la lumière de fluorescence par la solution. Il ne peut y avoir de réabsorption que si certaines molécules de la solution absorbent à des longueurs d'onde où d'autres émettent. Le décalage de Stokes entre les spectres d'absorption et d'émission d'un fluorophore limite le phénomène d'autoabsorption d'un fluorophore. En revanche, les bandes d'absorption et d'émission des composants étant assez larges (souvent supérieures à 100 nm), il est courant d'observer des chevauchements entre les spectres des différents composants d'un mélange. C'est par exemple le cas de notre mélange concentré de fluorescéine et de sulfate de quinine (figure 1.14). Les molécules de fluorescéine peuvent absorber la fluorescence du sulfate de quinine. Il s'ensuit alors une modification non linéaire du spectre illustrée par la figure 1.11. Par ailleurs s'il s'agit du cas de figure le plus général, un tel phénomène n'est pas limité aux mélanges de fluorophores. En effet, le (les) composant(s) absorbant tout ou partie de la lumière de fluorescence n'est (ne sont) pas nécessairement fluorescent(s). On parle dans ce cas de chromatophore(s), et les effets produits sont comparables. En pratique, lors de l'acquisition d'un spectre de fluorescence, l'effet de la réabsorption sur la mesure est négligeable à condition que l'absorbance de la solution soit suffisamment faible. Dans le cas contraire, il doit être pris en compte dans le modèle et corrigé, ce que nous ferons aux chapitres 3 et 4.

1.5.4 Remarque finale

Dans le cas des faibles concentrations le modèle multilinéaire, dont la résolution fera l'objet du chapitre suivant, synthétise les trois relations bilinéaires existant entre les différents paramètres :

$$I_{3D}(\lambda_e, \lambda_f, \mathbf{c}) = klI_0 \sum_{n=1}^N \Phi_n \gamma_n(\lambda_f) \varepsilon_n(\lambda_e) c_n \quad (1.26)$$

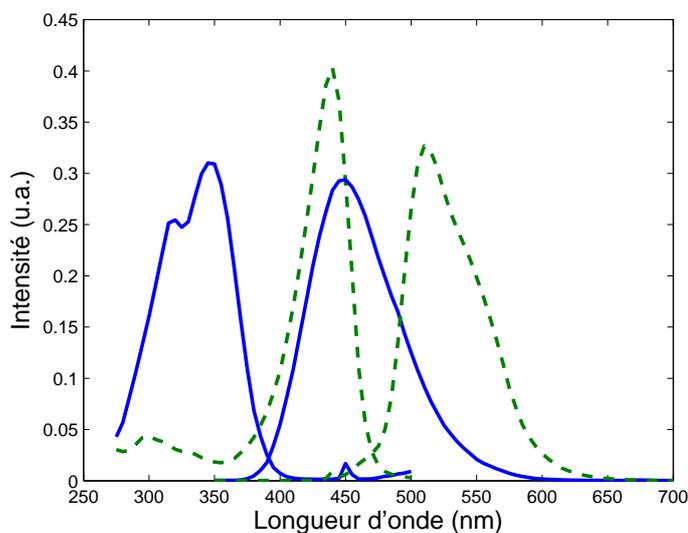


FIG. 1.14 – Spectres du sulfate de quinine et de la fluorescéine : mise en évidence du chevauchement spectral

En traits pleins, les spectres du sulfate de quinine.

- les spectres d'émission liés linéairement aux concentrations :

$$I(\lambda_f, \mathbf{c}) = \sum_{n=1}^N u_n \gamma_n(\lambda_f) c_n \quad (1.27)$$

où les u_n sont constants par rapport à λ_f et \mathbf{c} .

- les spectres d'excitation liés linéairement aux concentrations :

$$I(\lambda_e, \mathbf{c}) = \sum_{n=1}^N v_n \varepsilon_n(\lambda_e) c_n \quad (1.28)$$

où les v_n sont constants par rapport à λ_e et \mathbf{c} .

- les spectres d'émission liés linéairement aux spectres d'excitation :

$$I(\lambda_e, \lambda_f) = \sum_{n=1}^N w_n \gamma_n(\lambda_f) \varepsilon_n(\lambda_e) \quad (1.29)$$

où les w_n sont constants par rapport à λ_e et λ_f .

Revenons maintenant sur les résultats expérimentaux présentés au début de cette section. Ceux-ci attestent, que la simple augmentation des concentrations suffit à rendre non linéaire chacune de ces trois relations.

Chapitre 2

Analyse numérique des spectres de fluorescence, état de l'art

Le chapitre précédent nous a permis de caractériser les différents types de spectres de fluorescence. Nous avons également rappelé les principes de mesure et les quelques relations mathématiques qui en découlent. Néanmoins, les signaux mesurés ne sont pas toujours exploitables en l'état. Le traitement numérique des signaux de fluorescence a trois principaux rôles :

- la correction de la mesure.
- l'analyse des mélanges.
- la classification d'échantillons.

Correction des mesures. Les spectres mesurés, peuvent présenter des distortions par rapport aux spectres réels. En supposant que le spectrofluorimètre est correctement calibré (voir pour cela le paragraphe 1.3.2.5), deux sortes de mesures peuvent nécessiter des corrections.

Quelle que soit la nature du signal, toute mesure expérimentale est perturbée par un bruit de mesure aléatoire et peut être caractérisée par son rapport signal sur bruit. Bien évidemment, c'est également le cas des signaux de fluorescence. Lorsque les fluctuations parasites ne peuvent pas être négligées, on peut appliquer des techniques de filtrage classiques en traitement du signal. Le rapport signal sur bruit des spectres mesurés avec le spectrofluorimètre est généralement satisfaisant. C'est notamment le cas des solutions concentrées qui nous intéressent. Toutefois, afin d'effectuer la correction de l'effet d'écran, nous serons amenés au chapitre 4 à utiliser un algorithme de filtrage adapté à nos images. Celui-ci est basé sur une approche multirésolution du signal. Il s'agit d'une application directe de l'algorithme à trous [39], écrite au laboratoire pour l'occasion.

Les autres causes de distortions, effet d'écran et diffusion, sont spécifiques à la mesure de fluorescence ou aux techniques de spectroscopie. Leur importance a déjà été soulignée au chapitre précédent (paragraphe 1.4 et 1.5). Par ailleurs, le traitement de l'effet d'écran fera l'objet des chapitres 3 et 4.

Analyse des mélanges. La connaissance des spectres de fluorescence permet généralement d'identifier un fluorophore. Dans le cas d'un mélange de plusieurs fluorophores nous n'avons accès qu'à la mesure des spectres du mélange. Un traitement postérieur est nécessaire afin de déterminer les spectres de chaque composant de ce mélange et éventuellement leurs concentrations relatives.

Les problèmes inverses sont généralement difficiles à résoudre et des simplifications sont souvent nécessaires. Dans cette optique, le type de mélange le plus simple mathématiquement est formé par une combinaison linéaire des contributions de chaque composant. C'est pourquoi le modèle linéaire des mélanges de fluorophores (1.24) est largement employé. Nous présentons dans ce chapitre les principales méthodes et algorithmes d'analyse linéaire.

Classification d'échantillons. La classification d'échantillons en fonction de leurs spectres de fluorescence est au centre de nombreux travaux en chimie analytique ([29] et [37] ne sont que deux exemples récents parmi une très vaste bibliographie). Dans la plupart des études actuelles, ce problème englobe le problème précédent. En effet, les techniques précédentes fournissent la base de fluorophores élémentaires composant le jeu d'échantillon ainsi qu'une estimation de l'influence de chaque fluorophore sur le signal de fluorescence des différents échantillons. La classification s'opère à partir de ces résultats, comme l'illustre les exemples d'application proposés au chapitre 5. Cependant, il n'est pas toujours nécessaire de décomposer les signaux de fluorescence sur la base « physique » des fluorophores élémentaires. Il existe également des méthodes de classification purement statistiques généralement basées sur l'analyse en composantes principales du jeu de données. Il nous a paru important de faire ici la distinction entre les deux approches. Notons que l'avènement des méthodes trilineaires a marginalisé ce type d'étude.

2.1 Introduction à l'analyse linéaire des mélanges fluorescents

Remarque préalable : Les méthodes suivantes s'appliquent sur des échantillons dilués ou sur des MEEF corrigées des interactions entre fluorophores. C'est pourquoi nous confondrons ici spectres d'excitation et spectres d'absorption (cf chapitre 1).

Jusqu'à la section 2.4 de ce chapitre, nous considérons que l'approximation de l'équation 1.24 est satisfaisante. Le spectre (excitation, émission, 3D) d'un mélange sera donc vu comme une combinaison linéaire des spectres de chaque composant du mélange. Notre objectif sera d'estimer les spectres individuels et leur poids respectif. En pratique, on dispose souvent de plusieurs échantillons correspondant aux mêmes composants mais dont la distribution des poids varie. L'absence d'un composant dans un mélange est modélisé par un poids nul. Chaque échantillon étant associé à un mélange, on confondra donc souvent les deux termes.

Les modèles linéaires, de part leur simplicité, se retrouvent dans la plupart des do-

maines scientifiques, en particulier en chimométrie. D'ailleurs, aucune des approches que nous présentons ici n'est spécifique aux signaux de fluorescence. Notons que l'analyse numérique des mélanges linéaires ayant fait l'objet d'une importante production scientifique, il existe de très nombreuses méthodes de résolution, dont il est impossible de dresser ici une liste exhaustive. En outre, ces méthodes sont régulièrement améliorées et comptent parfois plusieurs variantes. Par ailleurs, leur grande variété autorise de nombreux critères de classement (méthodes itératives ou non itératives, bilinéaires ou multilinéaire, basées ou non sur l'optimisation d'un critère quadratique, méthodes de détection ou de prédiction...). En réalité, toutes ne répondent pas exactement au même type de problème, et en pratique, le choix s'effectue le plus souvent en fonction des données disponibles que par les qualités de tel ou tel algorithme. Dans le cas des signaux de fluorescence, le modèle linéaire général fait intervenir jusqu'à trois variables pour chaque fluorophore : leur spectre d'émission, leur spectre d'absorption, leur concentration. Dès lors, nous pouvons distinguer trois cas de figure.

Données d'ordre 0. On dispose uniquement du spectre 2D, \mathbf{x} (excitation ou émission) d'un seul échantillon du mélange linéaire. Soit \mathbf{A}_p la matrice contenant les spectres des composants du mélange en colonnes et \mathbf{m}_p le vecteur de leur poids respectif dans le mélange, tous deux inconnus, le modèle est alors le suivant :

$$\mathbf{x} = \mathbf{A}_p \mathbf{m}_p \quad (2.1)$$

On parle de données d'ordre 0 et de décomposition linéaire. C'est bien sûr, le cas le moins favorable et sans autres informations sur la nature du mélange, il paraît difficile de tirer des conclusions élaborées sur ses constituants (dans certains cas il peut même être impossible d'affirmer qu'il s'agit d'un mélange).

Données du premier ordre. On dispose cette fois d'un spectre 2D de plusieurs échantillons. Chaque ligne de la matrice de données \mathbf{X} est formée par le spectre de l'échantillon correspondant ${}^t\mathbf{x}$. Le modèle précédent devient alors :

$$\mathbf{X} = \mathbf{M}_p {}^t\mathbf{A}_p \quad (2.2)$$

On parle dans ce cas de données du premier ordre et de décomposition bilinéaire. Ce modèle est également valable si l'on dispose du spectre 3D d'un unique échantillon. Dans ce cas, \mathbf{X} est une matrice d'excitation-émission. Les colonnes de \mathbf{M}_p et ${}^t\mathbf{A}_p$ contiennent respectivement les spectres d'émission et d'excitation des composants du mélange.

Données du second ordre. Enfin, si l'on dispose du spectre 3D de plusieurs échantillons, on parle de données du second ordre et de décomposition trilinéaire. Les éléments du cube de données \mathbf{X} suivent l'équation 1.24 dont la reformulation matricielle sera donnée à la section 2.3.3.

Nous traiterons uniquement des données du premier ou du second ordre. La seconde partie de ce chapitre décrit les principales techniques d'analyse bilinéaire. Nous

détaillons également dans cette partie les principaux outils mathématiques à la base de ces méthodes. Nous verrons que la plupart des notions abordées seront réutilisées dans le cas trilinéaire. Celui-ci fait l'objet d'une troisième partie dont une large part est consacrée à l'algorithme PARAFAC qui constitue aujourd'hui l'outil d'analyse des mélanges fluorescents le plus abouti. Dans tous les cas nous ne disposons d'aucune information *a priori* sur les constituants fluorescents du (des) mélange(s).

2.2 Décompositions bilinéaires

La matrice de donnée \mathbf{X} (P, Q) pourra contenir indifféremment :

- P spectres 2D (excitation ou émission) de taille Q correspondant à P échantillons différents.
- P spectres d'émission de taille Q d'un même mélange, mesurés à P longueurs d'onde d'excitation.
- P spectres d'excitation de taille Q d'un même mélange, mesurés à P longueurs d'onde d'émission.

Dans toute la suite, on suppose que les mélanges sont des combinaisons de N fluorophores et que $P > N$. D'un point de vue algébrique, l'équation 2.2 s'interprète comme la décomposition des vecteurs lignes de la matrice \mathbf{X} sur la base des vecteurs colonnes de \mathbf{A}_p . \mathbf{M}_p représente alors les coordonnées de ces P vecteurs dans cette nouvelle base. Si \mathbf{X} est de rang R , \mathbf{A}_p sera donc une matrice (Q, R) et \mathbf{M}_p une matrice (P, R) . Idéalement, R est égal au nombre de composants constituant le (les) mélange(s). En pratique, à cause du bruit, les relations entre les lignes de X ne sont pas parfaitement linéaires, le rang est donc généralement supérieur à N ce qui entraîne une surévaluation de la dimension de la base de décomposition si l'on veut respecter l'équation 2.2. Nous cherchons plutôt à déterminer une base d'un sous espace de dimension R_s modélisant les spectres des composants du mélange et rejetant le bruit. C'est à dire, deux matrices \mathbf{A} et \mathbf{M} de dimensions respectives (Q, R_s) et (P, R_s) et telles que :

$$\mathbf{X} = \mathbf{M}^t \mathbf{A} + \mathbf{E} \quad (2.3)$$

où \mathbf{E} (P, Q) représente l'écart résiduel. La dimension R_s du sous espace est appelé pseudo rang de X , dans notre cas le pseudo rang (ou rang chimique) est supposé égal à N . Il est important de noter que \mathbf{A} est la restriction de \mathbf{A}_p au sous espace des spectres. Surévaluer la dimension de ce sous espace améliorera toujours l'adéquation du modèle avec les données mais n'aura pas de sens physique. Une difficulté récurrente de l'analyse des mélanges est donc d'évaluer le nombre de composants approprié.

De plus, la décomposition de X suivant l'équation 2.3 n'est pas unique puisque d'un point de vue strictement mathématique, n'importe quelle base du même espace conviendra. En effet, si $(\mathbf{M}, {}^t\mathbf{A})$ est un couple solution alors, quelle que soit la matrice \mathbf{T} (R_s, R_s) inversible, il est évident que $(\mathbf{MT}, \mathbf{T}^{-1t}\mathbf{A})$ est également solution. Par conséquent, il convient d'imposer certaines contraintes supplémentaires afin de déterminer la solution physique $(\mathbf{M}, {}^t\mathbf{A})$. L'analyse en composantes principales (ACP) ne permet pas d'aboutir directement à ce résultat, néanmoins cette décomposition est à la base de la plupart des approches. C'est pourquoi, la partie suivante en rappelle les principes.

2.2.1 Analyse en composantes principales

Nous avons adopté dans notre présentation un point de vue algébrique, bien naturel lorsqu'il s'agit de calcul matriciel. Il a été mentionné que la présence de bruit aléatoire devait être prise en compte. L'équation 2.2 s'interprète également en termes statistiques. Chaque ligne de \mathbf{X} représente les coordonnées d'une observation dans la base canonique des Q variables (des longueurs d'onde dans notre cas). L'espace engendré par les variables est donc \mathbb{R}^Q . On note E le sous espace de \mathbb{R}^Q engendré par les composants du (des) mélange(s). Nous devons déterminer à partir des observations bruitées contenues dans \mathbf{X} le sous espace le plus proche de E .

Principe de l'ACP. L'ACP, ou décomposition de Karhunen-Loeve, fournit l'unique base orthogonale de \mathbb{R}^Q , (u_1, \dots, u_Q) respectant la contrainte suivante : Quel que soit l'entier q , $1 \leq q < Q$, la famille de q vecteurs orthogonaux qui minimise l'écart quadratique moyen entre les observations et leur projection dans le sous espace qu'elle engendre est (u_1, \dots, u_q) . Ainsi, u_1 sera la direction de l'espace décrivant le mieux les observations, (u_1, u_2) le plan décrivant le mieux les observations... jusqu'à l'hyperplan (u_1, \dots, u_{Q-1}) . u_Q est alors le complément orthogonal de (u_1, \dots, u_{Q-1}) dans \mathbb{R}^Q . On montre alors que minimiser l'écart quadratique moyen entre observation et projection revient à maximiser la variance des projections et que (u_1, \dots, u_Q) est la base de \mathbb{R}^Q formée par les vecteurs propres de la matrice ${}^t\mathbf{X}\mathbf{X}$. Par ailleurs, quel que soit l'entier q , $1 \leq q < Q$ la variance totale (ou cumulée) des projections sur (u_1, \dots, u_q) , σ_q^2 , est égale à un facteur près à la somme des valeurs propres $\lambda_1, \dots, \lambda_q$ associées à u_1, \dots, u_q . Ce qui conduit au résultat fondamental de l'ACP :

$$\frac{\sigma_q^2}{\sigma_Q^2} = \frac{\sigma_q^2}{\sigma_X^2} = \frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^Q \lambda_i} \quad (2.4)$$

où σ_X^2 est la variance des observations. Les valeurs propres permettent donc de mesurer la proportion de la variance des observations décrite dans un sous espace de dimension donnée. C'est à dire, la qualité de l'approximation des observations par leur projection dans ce sous espace.

Application aux spectres de fluorescence. Dans le cas idéal d'observations non bruitées, \mathbf{X} est de rang N et par conséquent, seules les N premières valeurs propres sont non nulles. Pour des observations peu bruitées comme les signaux de fluorescence, la contribution du bruit à la variance totale est négligeable et les valeurs propres d'ordre supérieur à N seront négligeables par rapport aux N premières. L'ACP nous permet donc de trouver le pseudo rang de la matrice de données (*i.e.* le nombre de composants du (des) mélange(s)) ainsi qu'une base orthogonale du sous espace de dimension N décrivant le mieux les observations. On peut légitimement supposer que les spectres réels des composants peuvent se décomposer fidèlement dans cette base.

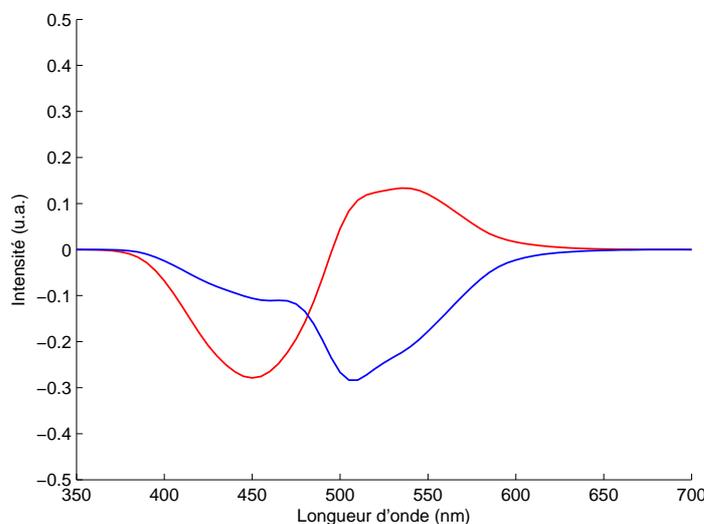


FIG. 2.1 – Analyse en composantes principales d'une matrice d'excitation-émission

Les vecteurs propres de cette base utilisés pour la décomposition sont appelés composantes principales de \mathbf{X} . En analyse factorielle, on utilise souvent la terminologie anglo-saxonne : on appelle *loadings*, les vecteurs de la base de décomposition et *scores*, les coordonnées des observations dans cette base. L'exemple de la figure 2.1 illustre la décomposition de la matrice d'excitation émission d'un mélange de fluorescéine et de sulfate de quinine sur la base des composantes principales. Les longueurs d'onde d'émission sont les variables et les spectres d'émission constituent autant d'observations que de longueurs d'onde d'excitation. Les deux premières composantes principales expriment plus de 99% de la variance totale. En ce sens, l'espace qu'elles engendrent est très proche de celui engendré par les spectres d'émissions du sulfate de quinine et de la fluorescéine.

La contrainte d'orthogonalité permet notamment de quantifier rigoureusement les contributions de chaque composante principale. Cependant, il est clair qu'utilisée seule, l'ACP ne fournira pas une bonne approximation des spectres recherchés, qui sont rarement orthogonaux. Lorsqu'il s'agit d'analyser une MEEF, l'ACP ne constitue donc qu'une première étape. Celle-ci est alors suivie de la recherche d'une transformation linéaire entre la base principale et celle formée par les spectres (d'émission ou d'excitation) réels des différents constituants du mélange. Pour cela, on rajoute généralement des contraintes (non négativité ou unimodalité des spectres) permettant de restreindre l'espace des solutions.

Un second problème, inhérent à la plupart des méthodes d'analyse factorielle concerne la détection des signaux de faible amplitude. Si l'un des composants a une réponse faible par rapport aux autres, et donc une faible contribution à la variance totale, il risque d'être assimilé à du bruit. En revanche, L'ACP est souvent utilisée [29][61] pour classer des échantillons en fonction de leur spectre de fluorescence. La projection sur

les composantes principales élimine les informations superflues, et peut mettre en valeur les différentes classes d'échantillons (cf chapitre 5). Par ailleurs, pour certaines études il n'est pas nécessaire de connaître les spectres réels [29]. Enfin, c'est également un outil simple pour la détection des observations aberrantes (outliers).

Décomposition en valeurs singulières. Il s'agit d'une décomposition liée à l'ACP que l'on retrouvera par la suite. Toute matrice \mathbf{X} (P, Q) de rang R peut se décomposer de la manière suivante :

$$\mathbf{X} = \mathbf{U}\mathbf{S}^t\mathbf{V} \quad (2.5)$$

où \mathbf{U} (P, R) et \mathbf{V} (Q, R) sont deux matrices unitaires et \mathbf{S} (R, R) la matrice diagonale formée par les valeurs singulières. Cette décomposition porte le nom de décomposition en valeurs singulières (DVS). Les valeurs singulières sont égales à la racine carrée des valeurs propres de ${}^t\mathbf{X}\mathbf{X}$. Notons que la décomposition en valeurs singulières de ${}^t\mathbf{X}\mathbf{X}$ est identique à sa décomposition en valeurs propres. Cette décomposition est souvent utilisée pour trouver une base orthogonale des éléments de X . Par la suite, on se limitera aux N premières valeurs singulières où N est le rang chimique de \mathbf{X} . On a alors :

$$\mathbf{X} = \mathbf{U}\mathbf{S}^t\mathbf{V} + \mathbf{E} \quad (2.6)$$

2.2.2 Moindre carrés récursifs (MCR)

L'algorithme des moindres carrés récursif [43] estime les matrices \mathbf{A} et \mathbf{M} de l'équation 2.2 de manière itérative. Comme son nom l'indique, cette méthode est basée sur l'estimateur des moindres carrés dont nous rappelons brièvement les principes dans le cas linéaire. Supposons une relation linéaire du type :

$$\mathbf{R} = \mathbf{Q}\mathbf{O} + \mathbf{E} \quad (2.7)$$

où l'on cherche à estimer la matrice des paramètres \mathbf{O} , connaissant \mathbf{R} et \mathbf{Q} avec un bruit aléatoire \mathbf{E} . Un estimateur non biaisée (si \mathbf{E} est centré) et de variance minimale sera alors :

$$\hat{\mathbf{O}} = \mathit{argmin}_{\mathbf{O}}(\|\mathbf{R} - \mathbf{Q}\mathbf{O}\|^2) \quad (2.8)$$

On montre alors que,

$$\hat{\mathbf{O}} = ({}^t\mathbf{Q}\mathbf{Q})^{-1}{}^t\mathbf{Q}\mathbf{R} \quad (2.9)$$

Dans notre cas, nous ne connaissons ni \mathbf{O} ni \mathbf{Q} . L'idée est donc de remplacer successivement \mathbf{O} puis \mathbf{Q} par leurs estimés $\hat{\mathbf{O}}$ et $\hat{\mathbf{Q}}$ et de recommencer jusqu'à la convergence du modèle. L'algorithme de résolution par les moindres carrés récursifs est le suivant :

1. Initialisation de $\hat{\mathbf{Q}}$
2. Estimation de \mathbf{O} , par 2.9
3. Calcul de l'écart quadratique entre \mathbf{R} et $\hat{\mathbf{Q}}\hat{\mathbf{O}}$

4. Actualisation de $\widehat{\mathbf{Q}} : \widehat{\mathbf{Q}} = \mathbf{R}^t \widehat{\mathbf{O}} (\widehat{\mathbf{O}}^t \widehat{\mathbf{O}})^{-1}$
5. On réitère les étapes 2 à 4 jusqu'à convergence du modèle ou jusqu'à ce que le nombre d'itérations maximal ait été atteint.

Le critère de convergence ou critère d'arrêt est généralement l'écart absolu entre les résidus de deux itérations successives inférieur à un ε fixé (par ex., $\varepsilon = 10^{-6}$). En pratique, quelques centaines d'itérations suffisent pour obtenir la convergence. Cette approche, quoique brutale, a un intérêt majeur : il est très facile d'implémenter certaines contraintes, comme l'unimodalité des spectres ou la non négativité des éléments de \mathbf{O} et \mathbf{Q} . Le respect de ces contraintes restreint significativement l'espace des solutions. En outre, la version pondérée de l'estimateur des moindres carrés permet de contrôler l'influence des données peu fiables sur l'estimation. En contrepartie, rappelons qu'il existe une infinité de solutions minimisant l'écart quadratique et même sous contraintes, il est difficile de savoir si le résultat correspond bien à la solution physique. Remarquons également, qu'il est nécessaire de fournir en entrée les dimensions des matrices à estimer. Il faudra donc estimer au préalable le nombre de composants par une autre technique (par exemple, l'ACP) si celui-ci est inconnu. Nous développons ce point dans la description de PARAFAC (2.3.3).

2.2.3 OPA-Simplisma

A l'origine, l'Orthogonal Projection Approach (OPA) fut développée pour l'étude des impuretés dans les spectres de chromatographie [22], puis rapidement étendue à l'estimation des spectres des composants d'un mélange à partir de son chromatogramme [21] ou de sa MEEF [28]. Supposons donc la matrice d'excitation-émission, \mathbf{X} , (P, Q) d'un mélange de N composants. Toutes les longueurs d'onde d'excitation, n'excitent pas de la même manière chaque composant. Par conséquent, il existe pour chaque composant n , une longueur d'onde d'excitation, parmi celles balayées par les lignes de \mathbf{X} , pour laquelle le spectre d'émission du mélange est le plus proche du spectre d'émission du composant n . La première étape de l'OPA va être de déterminer ces N longueurs d'onde d'excitation et les spectres d'émission correspondants. Pour cela, on introduit la notion de dissemblance entre les spectres. Les auteurs proposent un critère géométrique pour évaluer la dissemblance d_i entre un spectre \mathbf{x}_i et un spectre de référence \mathbf{x}_r . Soit \mathbf{Y}_i la matrice ($Q, 2$) formée par les deux spectres en question. Alors,

$$d_i = \det({}^t \mathbf{Y}_i \mathbf{Y}_i) = \|\mathbf{x}_r\| \|\mathbf{x}_i\| \sin(\alpha_i) \quad (2.10)$$

où α_i est l'angle entre \mathbf{x}_i et \mathbf{x}_r . Géométriquement, d_i représente l'aire du parallélogramme formé par les deux vecteurs, plus ceux-ci sont proches, plus cette aire est faible. On mesure de la même façon la dissemblance entre \mathbf{x}_i et un jeu de vecteurs de référence. Par exemple, si l'on veut comparer \mathbf{x}_i à n spectres de référence, \mathbf{Y}_i sera formée des $n+1$ spectres. Lorsqu'il y a deux spectres de référence d_i mesure le volume défini par les trois spectres.

Afin de déterminer les spectres d'émission recherchés, on calcule successivement la dissemblance de chaque spectre d'émission \mathbf{x}_i de \mathbf{X} avec un jeu de spectres de référence. On trace ensuite la courbe des dissemblances en fonction des longueurs d'onde

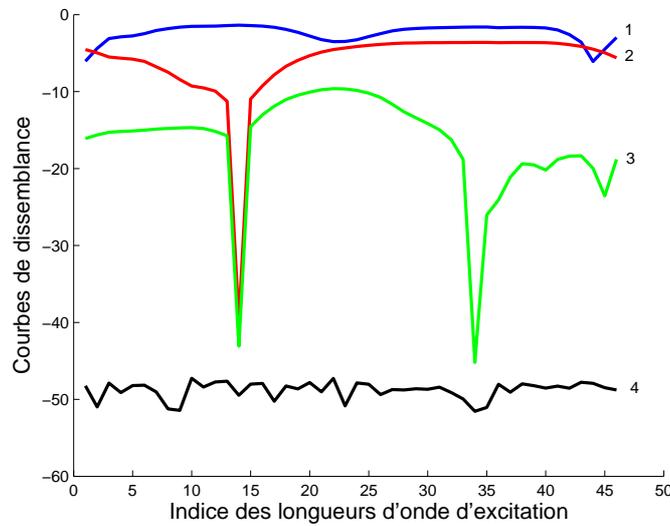


FIG. 2.2 – Courbes de dissemblance d'un mélange synthétique de trois fluorophores

d'excitation. A la première itération, le jeu de référence est uniquement constitué par le spectre d'émission moyen \mathbf{x}_{r0} , on a donc P matrices \mathbf{Y}_i de taille $(Q, 2)$. A chaque nouvelle itération, le spectre le plus dissemblable \mathbf{x}_{r1} est sélectionné et inclus dans les \mathbf{Y}_i . Le critère d'arrêt est fourni par la comparaison des courbes de dissemblance successives. En toute logique, lorsque le même nombre de spectres que de composants du mélange ont été sélectionnés, les valeurs des dissemblances devraient fortement chuter et leur courbe présenter un profil aléatoire. La figure 2.2 montre le logarithme des quatre premières courbes de dissemblance calculées à partir d'une FEEM d'un mélange synthétique de trois composants. On constate effectivement une forte différence de niveau entre les trois premières courbes et la quatrième qui est par ailleurs, la seule à présenter un profil aléatoire. Ceci permet de conclure sur le bon nombre de composants du mélange. Les spectres sélectionnés constituent une première estimation des spectres purs. Pour un mélange composé de peu de composants, cette approximation peut être assez bonne. D'une manière générale, les spectres des composants excités aux plus basses et aux plus hautes longueurs d'onde seront correctement estimés. Le traitement des spectres intermédiaires est plus délicat, suivant le degré de chevauchement entre les spectres d'excitation et les amplitudes relatives du signal de fluorescence des différents composants, l'estimation sera (ou non) satisfaisante.

L'OPA est donc complétée par une phase de correction itérative des spectres estimés. En réalité, il s'agit simplement d'appliquer la méthode des MCR en utilisant les spectres sélectionnés précédemment pour l'initialisation de l'algorithme. Les essais effectués sur nos mélanges n'ont pas montré d'amélioration substantielle des spectres sélectionnés par l'OPA. Signalons l'existence d'un algorithme similaire, SIMPLISMA [80], que nous n'avons pas testé. Cette méthode est également basée sur le principe de sélection des spectres dans la matrice de données, mais utilise un critère un peu différent.

2.2.4 Analyses factorielles à fenêtre

Nous avons vu comment déterminer le nombre global de composants à l'aide de l'analyse en composantes principales. Cependant, il est assez rare que tous les composants contribuent au signal de chaque observation. On peut alors construire une carte du rang (rank map) de la matrice de données \mathbf{X} correspondant au calcul du rang de plusieurs sous matrices (fenêtres) formées de quelques lignes consécutives de \mathbf{X} . Le choix des sous matrices utilisées conditionne directement l'allure de la carte et son interprétation. Deux méthodes sont régulièrement citées : la Fixed Size Moving Window-Evolving Factor Analysis (FSMW-EFA) [44] et l'Evolving Factor Analysis (EFA) [72]. En revanche, le calcul local du rang est toujours basé sur l'analyse en composantes principales. Par ailleurs, il existe également une large famille d'algorithmes d'analyse tirant profit de cette étude locale du rang. Plusieurs études comparatives de ces méthodes et combinaisons de méthodes peuvent être trouvées dans la littérature [46][21].

Dans le cas de l'analyse des MEEF, la carte du rang doit permettre de déterminer l'étalement spectral de chaque composant. A partir de ces données, deux méthodes de résolution sont couramment employées. La première porte le même nom que la méthode de cartographie local du rang qui lui fut originellement associée : l'EFA [72]. La seconde a donné lieu à deux algorithmes équivalents : Window Factor Analysis (WFA) [51] et Orthogonal Projection Resolution (OPR).

2.2.4.1 Construction de la carte de rang par EFA

Les spectres d'excitation des différents composants d'un mélange se chevauchent souvent mais leur étalement spectral respectif est généralement décalé de plusieurs nm . La figure 1.14, par exemple, illustre ce phénomène. L'émission de fluorescence est liée à l'excitation. En balayant les longueurs d'onde d'excitation successives, on modifie progressivement l'influence des différents composants sur le spectre d'émission du mélange. Il est difficile de connaître le nombre de composants contribuant au spectre d'émission du mélange à une longueur d'onde d'excitation donnée à partir de ce seul spectre. En revanche l'analyse en composantes principales permet de connaître le nombre de composants effectifs sur une certaine plage de longueurs d'onde d'excitation. Dans le cas de l'EFA, on utilise deux séries de fenêtres d'analyse antagonistes contenant des lignes successives. Pour la première série, toutes les fenêtres débutent par la première ligne de \mathbf{X} . La fenêtre initiale ne contient que la première ligne. Ensuite, chaque fenêtre est formée en ajoutant une ligne à la fenêtre précédente, jusqu'à la dernière fenêtre contenant la matrice entière. A chaque fenêtre correspond l'indice d'une ligne de \mathbf{X} égale à la taille de la fenêtre. Pour chaque indice on calcule les valeurs singulières de la fenêtre d'analyse. La première partie de la carte de rang est constituée des courbes d'évolution du logarithme des N plus grandes valeurs singulières en fonction de la longueur d'onde d'excitation. La figure 2.3 montre l'allure caractéristique de ces courbes.

Les premières lignes ne contenant que du bruit présentent de faibles valeurs singulières. Lorsque l'on atteint le début du spectre d'excitation du premier composant, celui-ci apparaît sur le spectre de fluorescence et par conséquent la première valeur

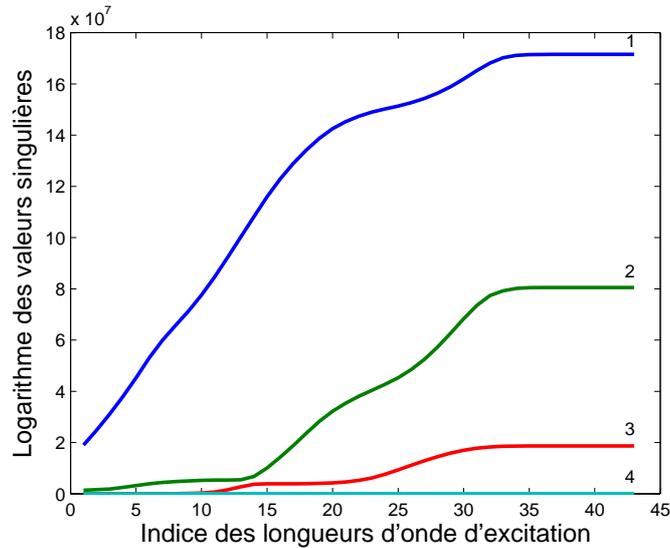


FIG. 2.3 – Courbes EFA, analyse allée

Les courbes sont obtenues à partir d'un mélange synthétique de trois composants

singulière augmente, puis, à chaque fois que le spectre d'excitation d'un nouveau composant est atteint, une nouvelle valeur singulière croît. Cette première analyse permet seulement de localiser le début de chaque spectre d'excitation recherché. La seconde série de fenêtres est construite exactement de la même manière mais en commençant par la dernière ligne de la matrice et en remontant. On obtient alors la famille de courbes miroirs illustrée par la figure 2.4. Cela procure une indication sur la fin de chaque spectre d'excitation. En supposant qu'aucun composant n'a un étalement spectral en excitation imbriqué dans celui d'un autre composant, le recoupement de ces deux jeux de courbes croisées à la manière de la figure 2.5 permet de déterminer l'étalement spectral de chaque composant.

2.2.4.2 Résolution spectrale par EFA

Le résultat précédent définit également le domaine spectral pour lequel un composant donné n'est pas excité. A l'intérieur de ce domaine son spectre d'excitation est nul. Pour chaque colonne n de la matrice \mathbf{M} des spectres d'excitation, les éléments de ce domaine forment la fenêtre d'excitation nulle notée : \mathbf{m}_n^0 . Par ailleurs, la décomposition en valeurs singulières de \mathbf{X} donne :

$$\mathbf{X} = \mathbf{U}\mathbf{S}^t\mathbf{V} + \mathbf{E} = \mathbf{M}^t\mathbf{A} + \mathbf{E} \tag{2.11}$$

D'où l'on déduit :

$$\mathbf{M} = \mathbf{U}\mathbf{S}({}^t\mathbf{A}\mathbf{V})^{-1} = \mathbf{U}\mathbf{R} \tag{2.12}$$

avec $R = \mathbf{S}({}^t\mathbf{A}\mathbf{V})^{-1}$ Soit, pour le composant n :

$$\mathbf{m}_n = \mathbf{U}\mathbf{r}_n \tag{2.13}$$

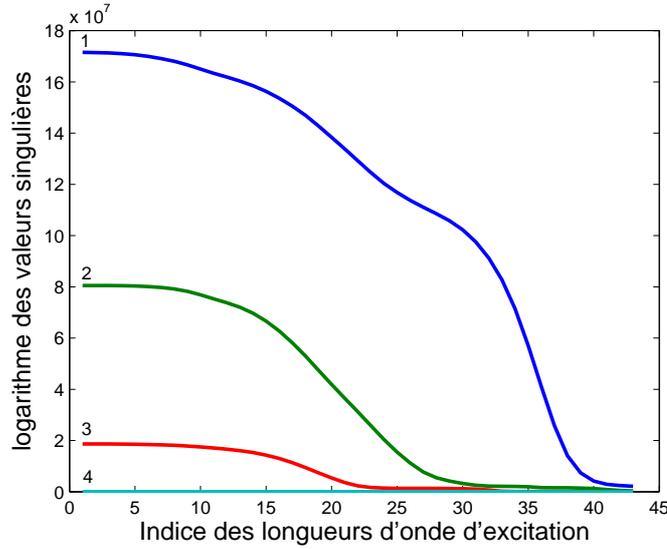


FIG. 2.4 – Courbes EFA, analyse retour

On note alors \mathbf{U}_n^0 la matrice formée par les lignes de U correspondant à la fenêtre d'excitation nulle de n et on a alors :

$$\mathbf{m}_n^0 = \mathbf{U}_n^0 \mathbf{r}_n = \mathbf{0} \quad (2.14)$$

Par construction, le rang de \mathbf{U}_n^0 est $N - 1$. Il faut donc fixer arbitrairement un des éléments de \mathbf{r}_n afin de résoudre l'équation. On impose en général $\mathbf{r}_n(1) = 0$. Le calcul successif des \mathbf{r}_n pour chaque composant du mélange permet de déterminer les matrices \mathbf{M} et \mathbf{A} grâce aux équations 2.13 et à l'estimateur des moindres carrés.

2.2.4.3 Résolution spectrale par WFA/OPR

L'algorithme WFA est également basé sur les fenêtres d'excitation nulle définies précédemment. Celles-ci sont obtenues après analyse d'une carte de rang.

Soit \mathbf{X}_n^0 la sous matrice de X formée par les lignes correspondant à la zone d'excitation nulle du composant n . La décomposition en valeurs singulières de \mathbf{X}_n^0 donne :

$$\mathbf{X}_n^0 = \mathbf{U}_n^0 \mathbf{S}_n^{0t} \mathbf{V}_n^0 + \mathbf{E} \quad (2.15)$$

Les $N - 1$ colonnes de \mathbf{V}_n^0 forment une base orthormale de l'espace engendré par les $N - 1$ spectres d'émission correspondants. On calcule alors la matrice X_n représentant le complémentaire orthogonal de la projection orthogonale des observations dans cette base :

$$\mathbf{X}_n = \mathbf{X}(\mathbf{I}_Q - \mathbf{V}_n^{0t} \mathbf{V}_n^0) = \alpha \mathbf{m}_n^t \mathbf{a}_n^0 \quad (2.16)$$

où, α est une constante, \mathbf{m}_n est le spectre d'excitation du composant n et \mathbf{a}_n^0 un vecteur orthogonal à son spectre d'émission. En théorie, les colonnes de \mathbf{X}_n sont toutes proportionnelles à \mathbf{m}_n . En pratique, l'estimation du spectre d'excitation du composant n peut se faire de différentes manières. En calculant par exemple la moyenne des

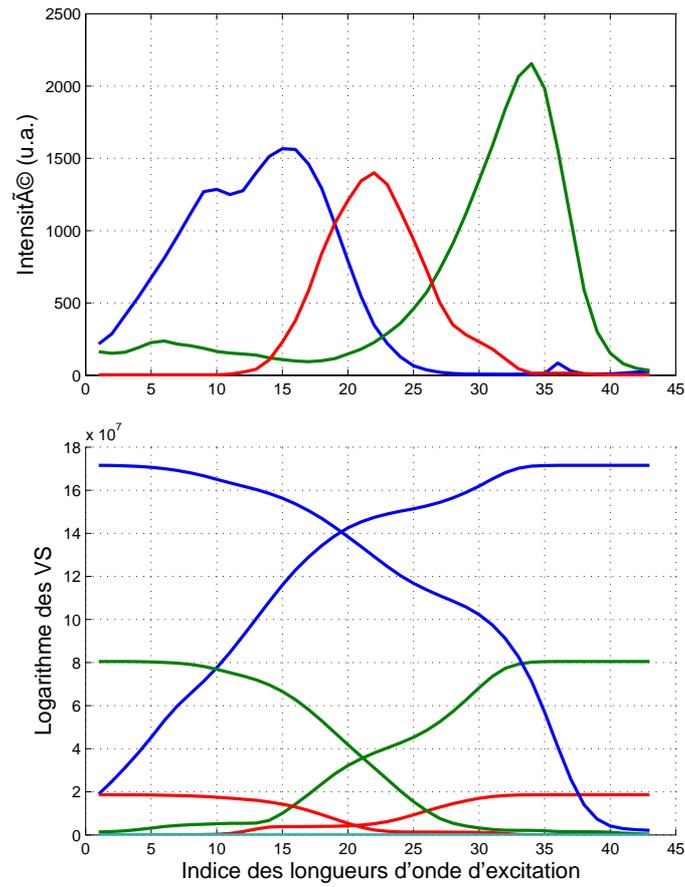


FIG. 2.5 – Comparaison entre les spectres d'excitation des fluorophores composants le mélange et les courbes EFA croisées.

L'analyse du croisement des courbes EFA permet de déterminer l'étalement spectral de chaque composant.

colonnes ou encore, en sélectionnant la colonne de plus grande norme (et *a priori* la moins bruitée). On estime de la même façon les autres colonnes de la matrice \mathbf{M} , puis, la matrice \mathbf{A} à l'aide de l'estimateur des moindres carrés.

2.2.4.4 Conclusion

Ces méthodes reposent sur la répartition progressive des spectres d'excitation des différents composants dans le domaine spectral. Cette propriété est également valable pour les spectres d'émission. Il est donc possible d'inverser le rôle des spectres d'excitation et d'émission.

Dans tous les cas, la première matrice peut servir d'initialisation à la procédure des MCR. Sans la procédure des MCR, les résultats obtenus sont meilleurs que pour l'OPA qui est limitée par les chevauchement de spectres. Néanmoins, plusieurs difficultés apparaissent lorsque certains spectres sont imbriqués ou lorsqu'il existe de fortes différences d'amplitude entre les signaux de fluorescence des différents composants.

2.3 Décompositions trinéaires

Les méthodes suivantes permettent l'analyse groupée de plusieurs MEEF. A l'exception notable de la sous-section 2.3.2, le jeu de données contiendra dans toute la suite, I MEEF (échantillons) \mathbf{X}_i de taille (J, K) . Le nombre total de fluorophores intervenant dans le jeu de données (mais pas nécessairement dans chaque MEEF) est toujours noté N . Pour les applications pratiques qui nous intéressent (cf chapitre 5), les MEEF contiennent généralement une majorité de composants communs.

2.3.1 Bilinéarisation

En arrangeant de manière adéquate le jeu de données, des algorithmes bilinéaires comme l'ALS, l'ACP et l'OPA peuvent s'appliquer directement.

2.3.1.1 ACP d'un jeu de MEEF

La manière la plus simple d'appliquer l'ACP à des données trinéaires comme un jeu de MEEF, consiste à traiter les échantillons comme des observations et les couples de longueurs d'onde d'excitation et d'émission comme des variables. La matrice à traiter sera donc de taille (I, JK) . Le traitement et l'interprétation des résultats est analogue au cas bilinéaire. Pour plus de clarté dans la représentation des résultats, on pourra représenter les composantes principales (vecteurs de taille JK) comme des MEEF.

2.3.1.2 OPA-A et ALS

Ces deux méthodes utilisent la même représentation des données. Les matrices \mathbf{X}_i sont concaténées dans la direction des lignes pour former une matrice \mathbf{X}_c de taille

(IJ, K) .

Les extensions trilinéaires de l'OPA, appelée OPA-A [28] et de l'ALS décomposent la matrice \mathbf{X}_g de la manière suivante :

$$\mathbf{X}_g = \mathbf{M}_g^t \mathbf{A} \quad (2.17)$$

La matrice \mathbf{A} contient en colonnes les spectres d'émission de chaque composant des mélanges. La matrice \mathbf{M}_g contient les I matrices des spectres d'excitation, \mathbf{M}_i , concaténées à l'image de \mathbf{X}_g . Les variations des rapports de concentration entre les différents échantillons sont intégrés à la matrice \mathbf{M}_g . Les colonnes de même indice des matrices \mathbf{M}_i sont donc identiques à un facteur près. On applique ensuite l'une ou l'autre des deux méthodes à la manière du cas bilinéaire pour déterminer les deux matrices. Avec :

$$\mathbf{X}_i = \mathbf{M}_i^t \mathbf{A} \quad (2.18)$$

2.3.1.3 Conclusion

Ces algorithmes sont rapides, et surtout simples à implémenter. Cependant, ils ne tiennent compte à aucun moment de la nature trilinéaire des données. Or, cette contrainte restreint considérablement l'espace des solutions. Ce défaut pénalise fortement ces approches puisqu'il rend l'interprétation des résultats hasardeuse. Seule l'ACP se justifie d'un point de vue statistique, à condition de ne pas s'intéresser à la solution physique du problème de décomposition.

2.3.2 Décomposition bilinéaire conjointe

Nous regroupons sous ces termes les différentes variantes d'une même approche, mettant à profit la trilinearité des données. A l'origine, la méthode du Rank Annihilation Factor Analysis (RAFA) [47] et sa généralisation, Generalized Rank Annihilation Factor Analysis (GRAFA) [68] (appelée depuis Generalized Rank Annihilation Method (GRAM)) traite le problème de la calibration des mélanges. Il s'agit de déterminer les concentrations des différents composants d'un mélange à partir d'un autre échantillon formé des mêmes constituants mais en concentrations connues, sous la contrainte de trilinearité des données. La méthode est simple et rapide puisque basée sur de simples décompositions en vecteurs propres. Elle présente également un avantage décisif sur les approches bilinéaires, puisque la calibration reste possible même lorsque le mélange inconnu comporte des composants non présents dans l'échantillon de calibration (*second order advantage*). Dans [26] les auteurs proposent un résumé des améliorations successives de l'algorithme. Nous présentons ici l'extension de cette approche pour la détermination des spectres d'excitation, d'émission et les concentrations relatives des composants de deux mélanges inconnus. L'algorithme décrit ici fut redécouvert plusieurs fois sous des noms différents : Trilinear Decomposition (TLD) [69], Procrust Rotation (PR) [7][71]... On dispose donc des MEEF de deux échantillons notées \mathbf{X}_1 et \mathbf{X}_2 . On note \mathbf{M} et \mathbf{A} et les matrices des spectres d'excitation et d'émission des N fluorophores présents dans l'un ou l'autre des mélanges. On a donc :

$$\mathbf{X}_1 = \mathbf{M}^t \mathbf{A} \quad (2.19)$$

$$\mathbf{X}_2 = \mathbf{M}\mathbf{D}^t\mathbf{A} \quad (2.20)$$

où D est une matrice diagonale (N, N) contenant les rapports de concentration des fluorophores entre les deux mélanges. Les matrices \mathbf{X}_1 et \mathbf{X}_2 peuvent également se décomposer par ACP :

$$\mathbf{X}_1 = \mathbf{T}_1^t\mathbf{U} \quad (2.21)$$

$$\mathbf{X}_2 = \mathbf{T}_2^t\mathbf{U} \quad (2.22)$$

La matrice \mathbf{U} contient les composantes principales communes aux deux mélanges (déterminée par l'ACP trinéaire) et T_1 et T_2 les coordonnées des observations de \mathbf{X}_1 et \mathbf{X}_2 dans la base principale. La méthode consiste à déterminer la matrice \mathbf{R} inversible telle que :

$$\mathbf{M} = \mathbf{T}_1\mathbf{R}^{-1} \quad (2.23)$$

et

$${}^t\mathbf{A}\mathbf{A} = \mathbf{R}\mathbf{U}^{-1} \quad (2.24)$$

Les équations précédentes conduisent alors à :

$$\mathbf{T}_2 = \mathbf{T}_1\mathbf{R}^{-1}\mathbf{D}\mathbf{R} \quad (2.25)$$

On pose finalement :

$$\mathbf{Q} = \mathbf{R}^{-1}\mathbf{D}\mathbf{R} \quad (2.26)$$

La matrice \mathbf{R} est donc déterminée facilement en calculant les vecteurs propres de \mathbf{Q} . On déduit alors \mathbf{M} et \mathbf{A} de 2.23 et 2.24. Enfin, les concentrations relatives sont données par les valeurs propres de \mathbf{Q} . Cette méthode simple et rapide fournit une solution unique à la décomposition simultanée des deux MEEF correspondant à la solution physique. En revanche, elle reste assez restrictive, puisque le jeu de données ne doit comporter que de deux échantillons (ou encore, en utilisant la symétrie du problème, deux longueurs d'onde d'excitation ou deux longueurs d'onde d'émission). En outre, plusieurs problèmes de résolution numérique peuvent se poser si les valeurs de \mathbf{D} sont inappropriées.

2.3.3 PARAFAC

PARAFAC, (Parallel Factor Analysis) est une méthode d'analyse de données consistant à construire un modèle multilinéaire, le plus proche possible d'un jeu de données expérimentales. Initialement développée par Harshman en psychométrie [30][34][33], cette approche est aujourd'hui employée dans de nombreux domaines de recherche. Son application à la décomposition des MEEF fut popularisée récemment par Bro [13][14]. Bien que la terminologie employée soit proche de celle de l'analyse factorielle bilinéaire classique, il existe de nombreuses spécificités à l'approche multilinéaire. Depuis Harshman, on trouvera d'ailleurs une abondante littérature abordant PARAFAC d'un point de vue purement algébrique.

2.3.3.1 Modélisation trilineaire

Nous disposons donc d'un jeu de données, noté $\underline{\mathbf{X}}$, contenant les mesures d'une grandeur physique $x_{i,j,k}$ fonction de trois paramètres, i, j, k . Un modèle trilineaire de $\underline{\mathbf{X}}$ consiste en une combinaison linéaire de trois variables a, b, c dépendant respectivement de i, j, k et d'un paramètre commun n , telle que :

$$\forall i, j, k \quad x_{i,j,k} = \sum_{n=1}^N a_{in} b_{jn} c_{kn} \quad (2.27)$$

L'erreur due à la mesure et au modèle encourage cependant à introduire un terme résiduel. En pratique, si l'on note $\hat{x}_{i,j,k}$ la mesure et $e_{i,j,k}$ l'écart résiduel on aura :

$$\forall i, j, k \quad \hat{x}_{i,j,k} = \sum_{n=1}^N a_{in} b_{jn} c_{kn} + e_{i,j,k} \quad (2.28)$$

PARAFAC n'est alors rien d'autre qu'une routine numérique permettant d'estimer les variables a_{in} , b_{jn} et c_{kn} à partir des seules mesures $\hat{x}_{i,j,k}$, en minimisant l'écart résiduel avec le modèle trilineaire. Dans le cas de l'analyse de MEEF, $\underline{\mathbf{X}}$ contient les MEEF de I échantillons correspondant à des mélanges de N fluorophores en diverses proportions.

- $\hat{x}_{i,j,k}$ correspond alors à l'intensité de fluorescence de l'échantillon i mesurée au couple de longueurs d'onde d'excitation-émission (j, k) .
- a_{in} représente le facteur de fluorescence (produit de la concentration et du rendement quantique) du fluorophore n dans l'échantillon i .
- b_{jn} représente la valeur du spectre d'émission du fluorophore n à la longueur d'onde j .
- c_{kn} représente la valeur du spectre d'absorption du fluorophore n à la longueur d'onde k .

On associe alors trois vecteurs à chaque fluorophore :

- \mathbf{a}_n représente l'évolution de la concentration du fluorophore n au travers des échantillons.
- \mathbf{b}_n représente le spectre d'émission du fluorophore n .
- \mathbf{c}_n représente le spectre d'excitation du fluorophore n .

PARAFAC nous permet d'estimer l'allure générale des spectres d'absorption et d'émission de chaque fluorophore ainsi que les rapports de concentration d'un même fluorophore pour chaque échantillon.

Notons que le modèle précédent correspond à l'approximation linéaire de la mesure du spectre de fluorescence d'un mélange de N fluorophores (1.24) à une constante multiplicative près. On peut considérer sans changer le résultat de l'algorithme que dans 1.24 la constante est intégrée à l'un des modes. Cette assertion est justifiée dans la section suivante.

Terminologie. Les trois espaces contenant ces trois types de vecteurs sont appelés « modes » : mode des concentrations, mode des émissions, mode des excitations.

Contrairement à la décomposition d'un jeu d'échantillons par ACP, leur rôle dans la décomposition est interchangeable. Mathématiquement, la distinction entre vecteurs *scores* et *loadings* n'a plus de signification. C'est pourquoi dans le descriptif de l'algorithme, nous employons indistinctement le terme de loadings pour désigner les vecteurs des trois modes. Toutefois, pour l'analyse des MEEF on considère généralement que les vecteurs des modes spectraux forment la base de décomposition (*loadings*) tandis que les vecteurs du premier mode regroupent les coordonnées (*scores*) des échantillons dans la base spectrale.

2.3.3.2 Algorithme de décomposition

Note : Les relations établies dans cette sous-section sont valables pour tout jeu de données trilinéaire. Leur application à l'analyse des MEEF est détaillée en italique.

Ecriture matricielle. Le jeu de données $\underline{\mathbf{X}}$ peut se représenter de différentes façons. On associe généralement à chaque mode une direction de l'espace pour former un cube de données indicé par i , j et k . Les loadings sont regroupés dans trois matrices distinctes :

- \mathbf{A} , matrice (I, N) dont les colonnes sont constituées par les vecteurs \mathbf{a}_n
- \mathbf{B} , matrice (J, N) dont les colonnes sont constituées par les vecteurs \mathbf{b}_n
- \mathbf{C} , matrice (K, N) dont les colonnes sont constituées par les vecteurs \mathbf{c}_n

et on note \mathbf{X}_i , la matrice (J, K) correspondant à la section $\underline{\mathbf{X}}_{i..}$ du cube.

Appliquée à l'analyse des MEEF, I , J et K représentent respectivement le nombre d'échantillons du jeu de données, le nombre de longueurs d'onde d'émission et le nombre de longueurs d'onde d'excitation. De même, \mathbf{X}_i est le spectre 3D de l'échantillon i . D'après 1.26, on a bien :

$$\mathbf{X}_i(j, k) = \sum_n^N \mathbf{A}(i, n) \mathbf{B}(j, n)^t \mathbf{C}(n, k) \quad (2.29)$$

$$\mathbf{X}_i = \sum_n^N \mathbf{A}(i, n) \mathbf{b}_n^t \mathbf{c}_n = \sum_n^N \mathbf{A}(i, n) \mathbf{Y}_n \quad (2.30)$$

\mathbf{Y}_n représente alors le spectre 3D du fluorophore n . La MEEF de chaque échantillon s'écrit comme une combinaison linéaire des spectres 3D de chaque fluorophore.

On montre facilement que :

$$\mathbf{X}_i = \mathbf{B} \mathbf{D}_i^t \mathbf{C} \quad (2.31)$$

où \mathbf{D}_i est la matrice diagonale (n, n) contenant \mathbf{a}_i sur sa diagonale. En posant $\mathbf{S}_i = \mathbf{B} \mathbf{D}_i$ il vient,

$$\mathbf{X}_i = \mathbf{S}_i^t \mathbf{C} \quad (2.32)$$

Le modèle trinéaire peut donc s'interpréter comme une série de décompositions bilinéaires sur la même base. De plus, les coordonnées de toutes ces décompositions s'écrivent comme une pondération des colonnes d'une matrice commune. En s'appuyant sur ces deux contraintes, inhérentes à l'aspect trinéaire des données, on montre que sous certaines conditions la décomposition par PARAFAC admet une solution unique à deux degrés de liberté près. Si $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ est une solution alors toute solution $(\mathbf{A}_m, \mathbf{B}_m, \mathbf{C}_m)$ vérifie :

$$\mathbf{A}_m = \mathbf{A} \mathbf{D}_{a,m} \mathbf{P}_m \quad (2.33)$$

$$\mathbf{B}_m = \mathbf{B} \mathbf{D}_{b,m} \mathbf{P}_m \quad (2.34)$$

$$\mathbf{C}_m = \mathbf{C} \mathbf{D}_{c,m} \mathbf{P}_m \quad (2.35)$$

avec \mathbf{P}_m une matrice de permutation et $\mathbf{D}_{a,m}$, $\mathbf{D}_{b,m}$, $\mathbf{D}_{c,m}$, trois matrices diagonales telles que $\mathbf{D}_{a,m} \mathbf{D}_{b,m} \mathbf{D}_{c,m} = \mathbf{I}_n$. Les seules indéterminations concernent donc l'ordre des loadings dans les matrices et leur échelle. Une démonstration de ce résultat essentiel est donnée dans [31]. A l'instar de tout système algébrique les conditions d'unicité sont liées aux nombres d'équations indépendantes et de variables à estimer. L'étude théorique des modèles multilinéaires a permis d'établir des conditions suffisantes plus ou moins contraignantes. Citons en particulier la règle de Kruskal [45].

Dans le cas des MEEF, ce résultat implique qu'aucune contrainte extérieure au modèle n'est nécessaire pour retrouver les spectres réels à condition d'avoir suffisamment d'échantillons indépendants. L'indétermination sur l'échelle des loadings empêche de connaître les concentrations respectives de chaque fluorophore dans un échantillon donné. En revanche la colonne n de la matrice \mathbf{A} fournit les concentrations du fluorophore n relatives à chaque échantillon.

Cette représentation en lamelles n'est pas la plus pratique mathématiquement. Les données peuvent être regroupées dans une unique matrice en dépliant le cube suivant une direction privilégiée à la manière de la figure 2.6. On note $\mathbf{X}_{J,KI}$ la matrice (J, KI) représentant le cube déplié dans la direction j . Nous cherchons alors la relation matricielle existant entre $\mathbf{X}_{J,KI}$ et les trois matrices à déterminer. Cette relation doit regrouper les I équations correspondant à 2.32. Pour cela, on introduit un produit tensoriel particulier appelé « produit de Khatri-Rao » [2], noté \odot et tel que pour deux matrices \mathbf{A} et \mathbf{C} ayant le même nombre de colonnes N , on a :

$$\mathbf{A} \odot \mathbf{C} = [\mathbf{a}_1 \otimes \mathbf{c}_1 \quad \mathbf{a}_2 \otimes \mathbf{c}_2 \quad \dots \quad \mathbf{a}_N \otimes \mathbf{c}_N] \quad (2.36)$$

\otimes désigne le produit de Kronecker. $\mathbf{X}_{J,KI}$ peut alors s'écrire :

$$\mathbf{X}_{J,KI} = \mathbf{B}^t (\mathbf{A} \odot \mathbf{C}) \quad (2.37)$$

Plus de détails sur le produit Khatri-Rao ainsi que de la démonstration de cette relation sont donnés en annexe B de ce mémoire. De même, on pourrait choisir de

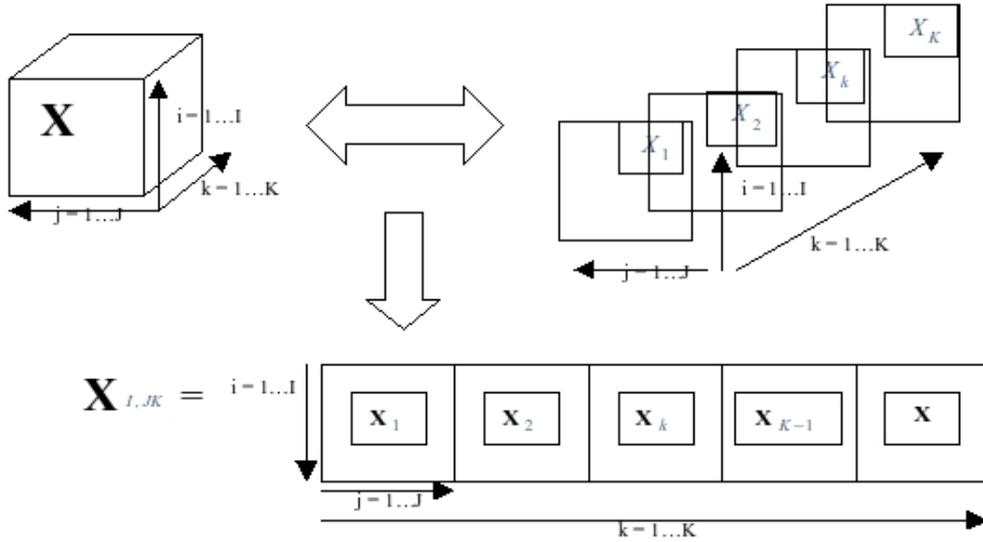


FIG. 2.6 – Exemple de représentation matricielle du cube de données
L'algorithme de décomposition est basé sur trois représentations matricielles possibles d'un même cube de données

déplier \mathbf{X} suivant i en une matrice $\mathbf{X}_{I,JK}$ ou encore suivant k en une matrice $\mathbf{X}_{K,IJ}$.
Et l'on aurait suivant le même raisonnement :

$$\mathbf{X}_{I,JK} = \mathbf{A}^t(\mathbf{C} \odot \mathbf{B}) \quad (2.38)$$

$$\mathbf{X}_{K,IJ} = \mathbf{C}^t(\mathbf{B} \odot \mathbf{A}) \quad (2.39)$$

Ces trois relations vont nous permettre d'estimer \mathbf{A} , \mathbf{B} et \mathbf{C} .

Il existe plusieurs algorithmes permettant d'estimer les loadings d'un modèle triliénaire [25]. Nous décrivons ici, l'algorithme le plus employé et parmi les plus robustes.

Moindres carrés récursifs. Estimer \mathbf{A} , \mathbf{B} et \mathbf{C} à partir de 2.37, 2.38 et 2.39 est relativement simple puisqu'il s'agit d'une régression linéaire récursive jusqu'à convergence du modèle. L'algorithme est similaire à celui employé dans le cas bilinéaire.

Reprenons les relations précédentes de manière récursive. Commençons par poser :

$$\mathbf{Z}_A = \mathbf{C} \odot \mathbf{B} \quad (2.40)$$

$$\mathbf{Z}_B = \mathbf{A} \odot \mathbf{C} \quad (2.41)$$

$$\mathbf{Z}_C = \mathbf{B} \odot \mathbf{A} \quad (2.42)$$

L'algorithme de résolution par les moindres carrés récursifs est le suivant :

1. Initialisation de \mathbf{B} et \mathbf{C}
2. Estimation de \mathbf{A} , \mathbf{B} et \mathbf{C} et calcul de \mathbf{Z}_A . On a :

$$\mathbf{X}_{I,JK} = \mathbf{A}^t \mathbf{Z}_A + \mathbf{E}_{I,JK} \quad (2.43)$$

ou encore :

$${}^t \mathbf{X}_{I,JK} = \mathbf{Z}_A {}^t \mathbf{A} + {}^t \mathbf{E}_{I,JK} \quad (2.44)$$

d'après 2.9, l'estimé de \mathbf{A} sera :

$$\widehat{\mathbf{A}} = \mathbf{X}_{I,JK} \mathbf{Z}_A ({}^t \mathbf{Z}_A \mathbf{Z}_A)^{-1} \quad (2.45)$$

On calcule \mathbf{Z}_B en utilisant $\widehat{\mathbf{A}}$ et \mathbf{C} et on estime \mathbf{B} :

$$\widehat{\mathbf{B}} = \mathbf{X}_{J,KI} \mathbf{Z}_B ({}^t \mathbf{Z}_B \mathbf{Z}_B)^{-1} \quad (2.46)$$

On calcule \mathbf{Z}_C en utilisant $\widehat{\mathbf{A}}$ et $\widehat{\mathbf{B}}$ et on estime \mathbf{C} :

$$\widehat{\mathbf{C}} = \mathbf{X}_{K,IJ} \mathbf{Z}_C ({}^t \mathbf{Z}_C \mathbf{Z}_C)^{-1} \quad (2.47)$$

3. Calcul de :

$$\mathbf{E}_{I,JK} = \| \mathbf{X}_{I,JK} - \widehat{\mathbf{A}}^t (\widehat{\mathbf{C}} \odot \widehat{\mathbf{B}}) \| \quad (2.48)$$

4. On recommence à partir de 2 jusqu'à la convergence de l'algorithme.

Il faut noter que même si la solution est unique, cela n'exclut pas le problème des minima locaux au cours de la résolution. Il se peut donc que l'algorithme converge vers un minimum local différent de la solution globale.

2.3.3.3 Utilisation pratique

Initialisation. Les matrices de loadings peuvent être initialisées aléatoirement ou par un pré-traitement des données. En théorie, du fait de l'unicité de la solution, le résultat de la décomposition ne devrait pas dépendre des valeurs choisies pour l'initialisation de l'algorithme. Cependant, une première décomposition bilinéaire de la matrice dépliée (par décomposition en valeurs singulières ou ACP par exemple) accélère la convergence de l'algorithme et peut éviter les minima locaux. Cependant, avec les puissances de calcul numérique actuelles, le critère de vitesse n'est pas fondamental pour notre utilisation. En outre, afin d'éviter les minima locaux, une solution plus sûre consiste à relancer plusieurs fois le programme à partir de matrices différentes jusqu'à obtenir la meilleure convergence. Nous choisissons cette dernière approche.

Contraintes. Nous avons vu que dans le cas de décompositions bilinéaires, il est nécessaire d'imposer des contraintes extérieures au modèle afin de choisir la solution appropriée. L'analyse trilineaire nous libère, en théorie, de cet impératif. Néanmoins, il peut être utile de forcer le modèle à respecter certaines contraintes afin d'éviter les minima locaux ou encore une solution aberrante qui serait due au bruit. Parmi les contraintes facilement applicables (orthogonalité, non négativité, unimodalité...), seule la contrainte de non négativité des loadings a un sens physique dans le cas des MEEF. Cela permet également d'imposer un résultat physiquement acceptable, sur les zones non linéaires des données. Un exemple est donné dans [13]

Rang du modèle. Les loadings sont estimés simultanément par PARAFAC et non successivement comme dans le cas d'autres méthodes. Par conséquent, le modèle estimé de rang N peut n'avoir aucun loading commun avec les modèles de rang inférieur à N . C'est pourquoi, le problème du choix du nombre de fluorophores requis est une étape décisive du traitement. Cette étape conditionne toute la chaîne de traitement. Par construction, quel que soit n un modèle de rang n convergera mieux qu'un modèle de rang inférieur à n . C'est pourquoi, en pratique, le seul critère de convergence ne permet pas toujours de dire si un composant supplémentaire améliore la modélisation du signal de fluorescence ou celle du bruit. On dispose de plusieurs méthodes de choix dont aucune n'est une panacée. Bro recommande d'ailleurs d'en comparer les résultats avant de conclure [5]. Ces méthodes sont principalement :

- Split half analysis [32] : On découpe les données en plusieurs jeux indépendants suivant l'un des modes (en général celui de plus grande dimension). On calcule alors un modèle pour chaque jeu. On choisira alors le rang le plus haut pour lequel les différents modèles sont suffisamment proches (aux permutations et aux facteurs d'échelle près).
- Core Consistency Diagnostic (CORCONDIA) [15] : Cette méthode permet de juger si pour un rang donné un modèle de Tucker, moins restrictif que PARAFAC sur sa formulation, modélise beaucoup mieux le jeu de données. Si c'est le cas, alors soit le rang est trop élevé soit les données ne sont pas trilineaires.

Citons également deux approches plus récentes :

- LTMC [40]
- ADD-ONE-UP truncating and fitting method [16]

La comparaison des différents résultats obtenus constitue une aide à la décision souvent efficace mais ce n'est pas toujours le cas. Lorsque le modèle trilineaire n'est pas une simple décomposition mathématique mais repose également sur une relation physique comme 1.26, il est légitime d'utiliser des connaissances *a priori* sur la solution pour guider le choix. En effet, un modèle de rang mal estimé produit parfois des résultats physiquement aberrants faciles à écarter.

Pré-traitement. Lorsque le jeu de données est de taille conséquente, ce qui est généralement le cas de l'analyse des MEEF, il n'est pas nécessaire de connaître toutes les valeurs de $\underline{\mathbf{X}}$. Les valeurs inconnues ou peu fiables pourront être considérées comme des données manquantes et seront remplacées par leurs estimations au fur et à mesure de l'algorithme. On a donc tout intérêt à traiter les valeurs dont le caractère trilineaire

est douteux comme des valeurs manquantes.

Il peut également être intéressant de normer les données suivant certains modes. Le critère de convergence est un critère quadratique portant sur l'ensemble des données. Ce critère favorise la modélisation des composants les plus fluorescents et d'attribuer aux autres un poids statistique équivalent au bruit. Lorsqu'on ne connaît pas le nombre réel de composants fluorescents, le risque est d'utiliser un modèle de rang sous estimé. On peut alors envisager différentes manières de normer les MEEF du jeu de données pour limiter ce problème.

Post-traitement. L'indétermination sur la norme des loadings nous oblige à la fixer arbitrairement. L'approche classique dans le traitement des MEEF consiste à diviser chaque spectre d'excitation et d'émission estimés (colonnes de \mathbf{B} et \mathbf{C}) par leur norme euclidienne. Afin de préserver la convergence du modèle, on multiplie les vecteurs concentrations relatives (colonnes de \mathbf{A}) correspondants par le produit des deux normes. Les loadings forment ainsi une base spectrale normée alors que les scores de la décomposition constituent une mesure du « facteur de fluorescence » de chaque composant. Ce facteur correspondrait au produit du coefficient d'absorption, de la concentration et du rendement quantique de fluorescence.

Exemple d'application. Nous renvoyons le lecteur au chapitre 5 où nous donnons plusieurs exemples détaillés de l'application de PARAFAC à l'analyse des MEEF.

2.3.4 Trilinear EFA

Récemment, Wang proposa une extension de l'EFA, TEFA [79] pour l'étude de données chromatographiques à 3 voies (chromatogrammes, spectre d'excitation, concentrations relatives). Il est possible d'étendre ce raisonnement à l'analyse trilineaire des spectres 3D même si aucune application ne semble avoir été publiée pour le moment. Nous conservons ici les notations utilisées pour PARAFAC.

2.3.4.1 Principe

TEFA exploite les trois voies du jeu de données pour construire une carte de rang plus simple à interpréter que dans le cas bilinéaire. On note \mathbf{X}_k la matrice des données correspondant à la longueur d'onde d'excitation k et on calcule le rang chimique de \mathbf{X}_k pour $k = 1..K$ par DVS. La carte de rang sera alors constituée par les courbes des valeurs singulières en fonction de la longueur d'onde d'excitation. On détermine alors la zone du spectre, W_1 , où un seul composant intervient (on parle de zone sélective). On choisit généralement le premier ou le dernier composant apparaissant sur la carte de rang.

$$\forall k \in W_1, \mathbf{X}_k = c_{k1} \mathbf{a}_{.1} {}^t \mathbf{b}_{.1} \quad (2.49)$$

On déduit des équations 2.49 les concentrations relatives $\mathbf{a}_{.1}$ et le spectre d'émission $\mathbf{b}_{.1}$ du premier composant sélectionné. La méthode RAFA (2.3.2) permet de retrouver la concentration relative d'un composant donné dans une MEEF inconnue à partir

de ses spectres d'excitation et d'émission. La méthode Coupled Vectors Resolution 1 (COVER1) [41] étend cette méthode à la résolution des concentrations relatives dans plusieurs échantillons inconnus. En appliquant le même raisonnement non plus au mode des concentration mais à celui des spectres d'excitation, COVER1 permet de retrouver le spectre d'excitation, \mathbf{c}_1 à partir de \mathbf{a}_1 , \mathbf{b}_1 et $\underline{\mathbf{X}}$. On connaît alors la contribution exacte de ce composant que l'on soustrait à $\underline{\mathbf{X}}$. On détermine une nouvelle carte de rang et on recommence les étapes précédentes pour un second composant et ainsi de suite.

2.3.4.2 Avantages et limites

Par rapport à L'EFA, la méthode permet de résoudre le problème des spectres imbriqués. En ce qui concerne la comparaison avec PARAFAC, les auteurs insistent surtout sur le côté non itératif de la méthode, prévenant ainsi tout problème de convergence. Par ailleurs, l'estimation successive des différents composants est censée limiter le problème critique du choix du nombre de composant du modèle. Cependant, la carte de rang n'est pas toujours simple à interpréter. En particulier, lorsque des composants sont noyés dans le bruit ou présentent des spectres d'excitation très corrélés entre eux. L'utilisation de PARAFAC nécessite une certaine prudence quant au choix du nombre de composants. On pourrait en dire autant pour le choix des zones sélectives qui conditionne tout autant le résultat et requiert la même pratique. Enfin, TEFA présente de nombreuses étapes de calcul (et donc autant de risques de propagation d'erreur) avant d'arriver au résultat.

2.3.5 Lagrange Operator Algorithm

A l'instar de PARAFAC, cette méthode publiée en 2003 [48] minimise de manière itérative un critère d'erreur quadratique. Il s'agit ici de minimiser l'erreur de reconstruction sur l'un des modes, choisi à l'avance, tout en respectant une contrainte sur les deux autres modes. Cette optimisation sous contraintes s'écrit alors sous la forme d'une seule fonction de pénalité que l'on minimise en utilisant l'opérateur de Lagrange. Cette approche est censée résoudre les problèmes de convergence de PARAFAC et surtout, elle doit réduire l'influence d'un mauvais choix du nombre de composants sur la solution obtenue.

Cependant, il nous est difficile d'être plus précis sur ce point car aucune justification mathématique n'est donnée. De même, l'utilisation pratique de l'algorithme est très peu abordée (initialisation des variables, influence de certains paramètres...). En outre, certaines formules comportent des erreurs. Il ne nous a pas été possible de retrouver les résultats prometteurs mis en avant par les auteurs et aucun article complémentaire n'a, semble-t-il, été publié depuis.

2.4 Approches non linéaires et effet d'écran

Du peu de méthodes d'analyse non linéaires. Le modèle linéaire de fluorescence est à l'origine de nombreuses méthodes de décomposition dont nous venons

de passer en revue les plus significatives. Lorsque ce modèle n'est plus valable, ces différents algorithmes ne sont pas directement applicables et de nouvelles approches doivent être considérées.

Nous avons distingué trois causes principales aux distortions non linéaires des données : La diffusion, le bruit de mesure et les effets d'écran. Chen et Yu exposent dans [17] une extension de PARAFAC, capable de prendre en compte les déviations non linéaires du jeu de données. Cependant, l'algorithme d'analyse proposé ne tient pas compte des phénomènes physiques à l'origine des non linéarités. De fait, celui-ci est beaucoup plus pensé pour corriger les déviations non linéaires d'un jeu de données globalement trilineaire plutôt qu'un jeu de données fondamentalement non linéaire comme cela peut être le cas en présence de fort effet d'écran. Nous avons donc pu vérifier que cette méthode ne donne aucun résultat concluant dans l'analyse de données avec effet d'écran (en fait, nous avons obtenu sensiblement les mêmes résultats qu'avec PARAFAC). Il n'existe pas à notre connaissance d'autre méthode d'analyse globale des mélanges non linéaires. L'effet d'écran doit donc être traité séparément et spécifiquement. Par ailleurs, nous ne reviendrons pas sur la diffusion qui a déjà fait l'objet d'une partie du chapitre 1 (1.4) ni sur le traitement du bruit de mesure.

Traitement classique de l'effet d'écran. L'effet d'écran est un phénomène bien connu [6][7][24]. Il est recommandé depuis longtemps de travailler à de faibles absorbances afin de mesurer des spectres fidèles, ce qui implique souvent une dilution importante des solutions. L'inconvénient majeur de cette procédure est une forte diminution du rapport signal sur bruit. Malgré la bonne sensibilité des spectrofluoromètres, certains cas sont problématiques. Par exemple, si le mélange original contient des composants fortement absorbants, à l'origine des effets d'écran, mais également un ou plusieurs fluorophore(s) peu concentré(s), la dilution risque de noyer le signal de ce(s) fluorophore(s) dans le bruit. Et cela, sans qu'il soit possible de le savoir. La correction de l'effet d'écran est donc souvent indispensable. La modélisation et la correction de l'effet d'écran a fait l'objet de plusieurs articles pratiques [54][78] ou théoriques [38] [81][49][20] [6][8][64]. Ces différents travaux apparaissent généralement comme autant de variations autour d'une même approche appliquée à différents cas de figure et constituant finalement un corpus assez flou. En pratique, un modèle particulier [60] est régulièrement utilisé [54][78][50][58]. Ce modèle permet donc de corriger l'effet d'écran en mesurant l'absorbance du mélange. Comme pour la diffusion, le traitement de l'effet d'écran s'effectue donc en deux temps : une procédure de correction permet de reconstituer un jeu de données bilinéaire ou trilineaire ensuite, la décomposition est réalisée à l'aide d'une des méthodes d'analyse linéaire. Les nouvelles approches que nous proposons aux chapitres 3 et 4 reposent sur le même principe. C'est pourquoi nous avons accordé dans ce mémoire une large part à la description des méthodes d'analyses linéaires.

Chapitre 3

Modélisation physique des effets d'écran

3.1 De l'intérêt de corriger l'effet d'écran

Nous avons montré, au chapitre 1 (1.5) que le modèle trinéaire ne pouvait pas s'appliquer à des solutions trop concentrées du fait des effets d'écran. En conséquence, les différentes méthodes présentées au chapitre 2 ne sont pas applicables directement au MEEF correspondant à de tels mélanges. Afin d'illustrer cette affirmation nous avons testé PARAFAC sur trois séries de mélanges de concentrations faibles, élevées et très élevées, de sulfate de quinine et de fluorescéine données par le tableau 3.1. Ces solutions ont servi de mélanges tests tout au long des chapitres 3 et 4. En effet, ces deux molécules présentent un fort signal de fluorescence et d'importants chevauchements spectraux favorisant les effets d'écran. On note que le sulfate de quinine a un pH acide qui influence le signal de fluorescence du mélange, aussi les spectres de référence de la fluorescéine ont été mesurés à des pH équivalents à celui du mélange. Les résultats obtenus (figure 3.1) pour les différentes séries mettent en lumière l'influence de la concentration. Par ailleurs, les résultats ne sont clairement pas acceptables pour la série la plus concentrée. Toutefois, on constate que l'erreur de modélisation n'affecte pas de la même manière l'estimation des différents loadings. Le spectre d'émission du sulfate de quinine est bien restitué à l'inverse des autres loadings et

| Mélange | Solution | [F] (ppm) | [SQ] (ppm) |
|---------|----------|-----------|------------|
| | 1 | 0 | 108.75 |
| 1 | 2 | 8.43 | 97.73 |
| 2 | 3 | 24.92 | 76.15 |
| 3 | 4 | 41.58 | 54.38 |
| 4 | 5 | 58.23 | 32.6 |
| 5 | 6 | 74.72 | 11.02 |
| | 7 | 83.15 | 0 |

TAB. 3.1 – Concentrations des solutions de fluoresceine (F) et sulfate de quinine (SQ)

en particulier les spectres d'excitation. Par ailleurs, le modèle trinéaire reconstruit explique près de 95% de la variance totale. De prime abord, cette valeur élevée est susceptible de conduire à une erreur d'interprétation. Dans la pratique, il s'avère qu'un tel pourcentage est un résultat moyen. Il faut enfin souligner que la convergence de l'algorithme est beaucoup plus instable que lorsque le modèle trinéaire est justifié.

Par ailleurs, nous avons vu que les quelques extensions non linéaires des algorithmes d'analyse, basés sur des approches statistiques, ne sont pas adaptées au traitement de distortions aussi importantes. Il est donc indispensable de disposer d'un modèle mathématique de l'effet d'écran afin de le corriger. Par ailleurs, nous n'avons pas retrouvé dans la bibliographie de démonstration précise et globale du modèle existant. Nous proposons donc dans ce chapitre de repartir à zéro avec une nouvelle mise en équation, claire et détaillée, adaptée à l'étude des MEEF d'un mélange et aboutissant à une équation similaire. La seconde partie de ce chapitre sera consacrée à la validation expérimentale du modèle.

3.2 Modélisation de l'effet d'écran

3.2.1 Approximations et notations

Nous nous intéressons à la mesure $I_{3D}(\lambda_{ex}, \lambda_{em}, \mathbf{c})$ du spectre de fluorescence 3D effectuée par un spectrofluorimètre standard, dont le fonctionnement a été décrit au premier chapitre de ce mémoire. Pour rappel, il s'agit d'une mesure du signal de fluorescence à 90° du faisceau excitateur. Lorsque la solution à mesurer est correctement préparée, la cuve contenant la solution peut se diviser en une infinité de volumes élémentaires identiques présentant un profil de concentration équivalent à celui de la solution. Ces volumes de solution étant infiniment petits, le modèle linéaire s'y applique sans erreur. En fonction de sa position, chaque volume absorbe une quantité différente de lumière excitatrice, puis émet une certaine intensité de fluorescence à différentes longueurs d'onde, susceptible d'être réabsorbée à son tour par chaque volume élémentaire et ainsi de suite. En outre, la lumière (d'excitation ou de fluorescence) arrivant sur un volume élémentaire n'est pas forcément absorbée ou transmise, une partie est diffusée de manière anisotrope vers le reste de la cuve. Pour un couple de longueur d'onde fixé, un modèle parfait du système de mesure devrait donc intégrer les phénomènes de réabsorption en cascade et de diffusion multiple afin d'établir précisément les contributions de chaque volume élémentaire à l'intensité lumineuse détectée. De notre point de vue, essentiellement pratique, il s'agit cependant d'une modélisation excessive de phénomènes marginaux et très difficiles à quantifier. Puisque le modèle ne constitue pas une fin en soi mais une étape de la procédure de correction, nous nous contenterons d'un modèle simplifié permettant toutefois de rendre compte efficacement des effets de la concentration sur la fluorescence et de les corriger.

Nous verrons que certains phénomènes physiques peuvent être négligés. Par ailleurs, sans retomber sur le cas extrême du modèle linéaire, le dispositif de mesure permettra également plusieurs simplifications.

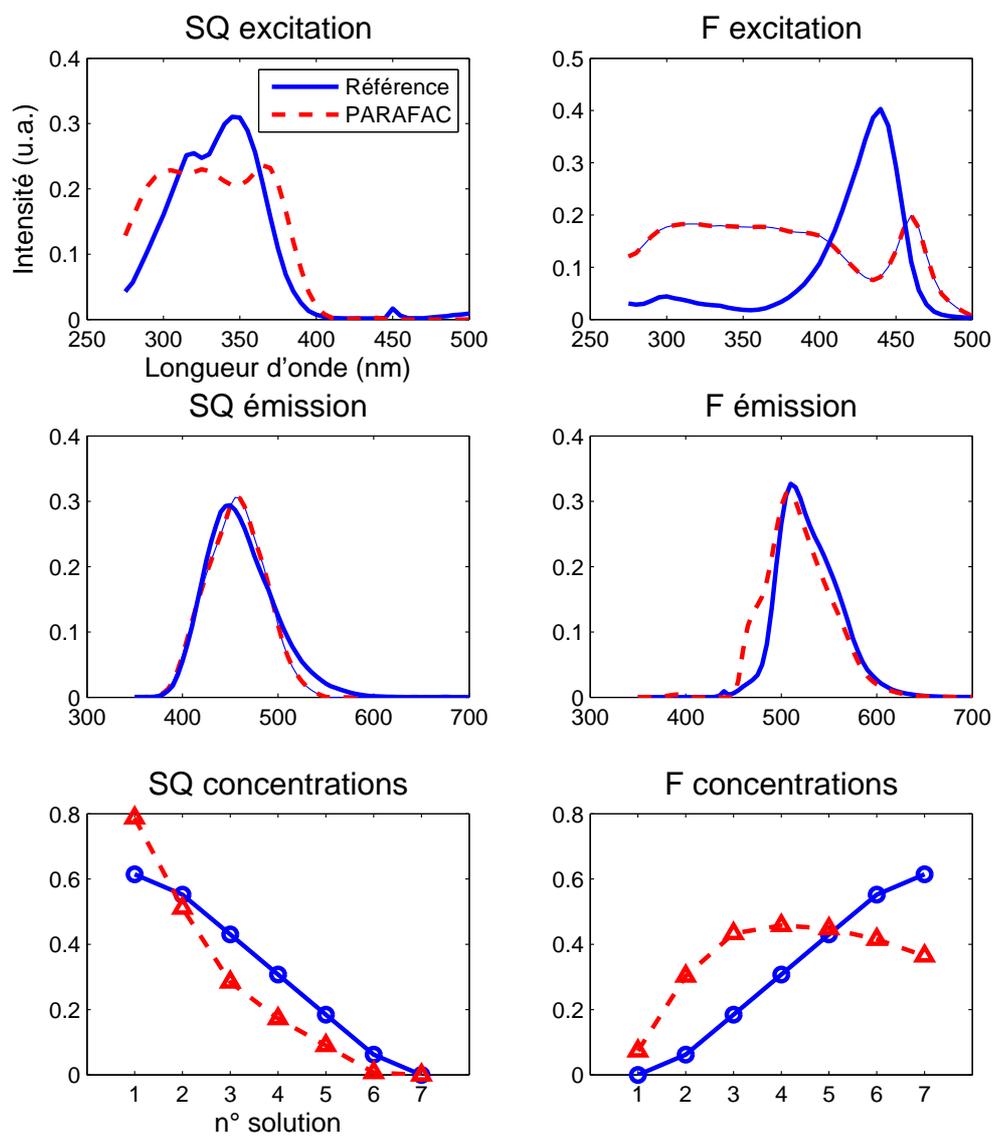


FIG. 3.1 – Résultats de PARAFAC sur les solutions les plus concentrées
 Comparaison entre les mesures des spectres individuels des solutions 1-7 du tableau 3.1 et les résultats obtenus par PARAFAC pour les spectres d'excitation, d'émission et les rapports de concentration.

Simplifications physiques. Les deux formes principales d'interaction lumière-matière sont l'absorption et la diffusion. Ces deux phénomènes concernent aussi bien le solvant que les composants de la solution, mais dans des proportions très différentes. La diffusion est principalement due au solvant (même dans le cas de solutions très concentrées, le nombre de molécules des solutés est négligeable devant le nombre de molécules de solvant). En revanche, celui-ci est généralement choisi pour que l'absorption soit la plus faible possible. Nous pouvons donc négliger l'absorption et la fluorescence éventuelles du solvant. Nous ne tiendrons également pas compte de la diffusion qui se superpose au signal de fluorescence et qui pourra être traitée ultérieurement (cf chapitre 1). Par ailleurs, nous négligerons les effets dus à la réflexion de la lumière sur les faces de la cuve ainsi que la possible réémission du signal de fluorescence absorbé par la solution. Finalement, nous ne tiendrons compte que des effets d'écrans primaires et secondaires dus à l'absorption progressive de la lumière d'excitation et d'émission.

Simplifications géométriques. La figure 3.2 représente le dispositif de mesure dans l'espace (xyz) . Nous raisonnerons dans un premier temps sur une section de la cuve, formant un carré de côté de l , orienté dans les directions (Ox, Oy) et divisé en surfaces élémentaires de dimension (dx, dy) , repérées par une abscisse et une ordonnée dans (xOy) . On note alors Δ_y l'épaisseur du faisceau excitateur d'intensité I_0 et Δ_x la largeur effective du système de détection. Le modèle sera ensuite construit à partir du schéma de principe, communément admis, de la figure 3.3. Nous ne raisonnerons donc plus sur des volumes mais sur des surfaces élémentaires. Les systèmes d'excitation et de détection du spectrofluorimètre sont conçus pour favoriser les surfaces élémentaires formant une « zone d'émission » (notée Z_e sur 3.3), au centre de la cuve, de dimension (Δ_x, Δ_y) . La simplification classique consiste à se limiter à ces surfaces élémentaires dans le calcul et surtout, à considérer uniquement les chemins optiques élémentaires illustrés par la figure 3.4. Nous supposons également que l'intensité de la lumière excitatrice est uniforme dans la direction (Oy) , tout comme la sensibilité du détecteur dans la direction (Ox) . Finalement, le système de mesure étant parfaitement symétrique par rapport au plan (xOy) , nous négligerons l'influence de la composante (z) du dispositif. Les seules variables utilisées sont les spectres d'excitation, d'émission et les concentrations de chaque composant. En conséquence, les paramètres indépendants de la longueurs d'onde d'excitation, de la longueur d'onde d'émission et d'un composant particulier seront traités comme des constantes. En particulier, tous les paramètres de mesure influençant de manière égale l'amplitude de l'intensité de la lumière détectée, I_{3D} , quelle que soit la longueur d'onde d'excitation, d'émission ou les concentrations du mélange, (par exemple la distance entre la cuve et le détecteur) peuvent être regroupés dans une constante mutiplicative, que l'on intègre dès le départ à I_0 par souci de simplicité.

La solution type est un mélange de fluorophores, de chromatophores et d'éléments non absorbants. Il est important de noter que ces derniers ont un impact nul sur le signal observé et ne sont donc pas considérés, ici, comme des composants de la solution. D'un point de vu mathématique, un chromatophore peut être vu comme un fluorophore de

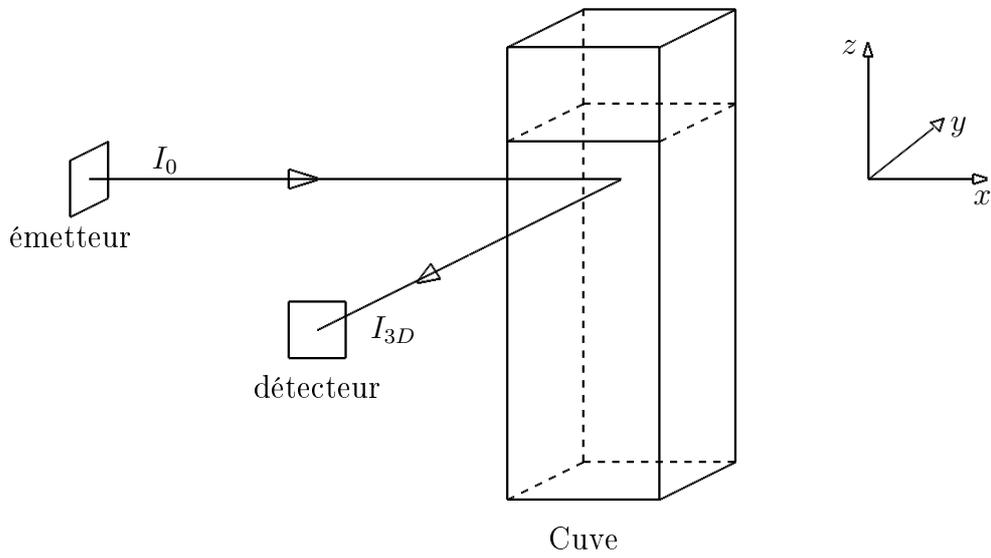


FIG. 3.2 – Représentation en perspective du dispositif de mesure

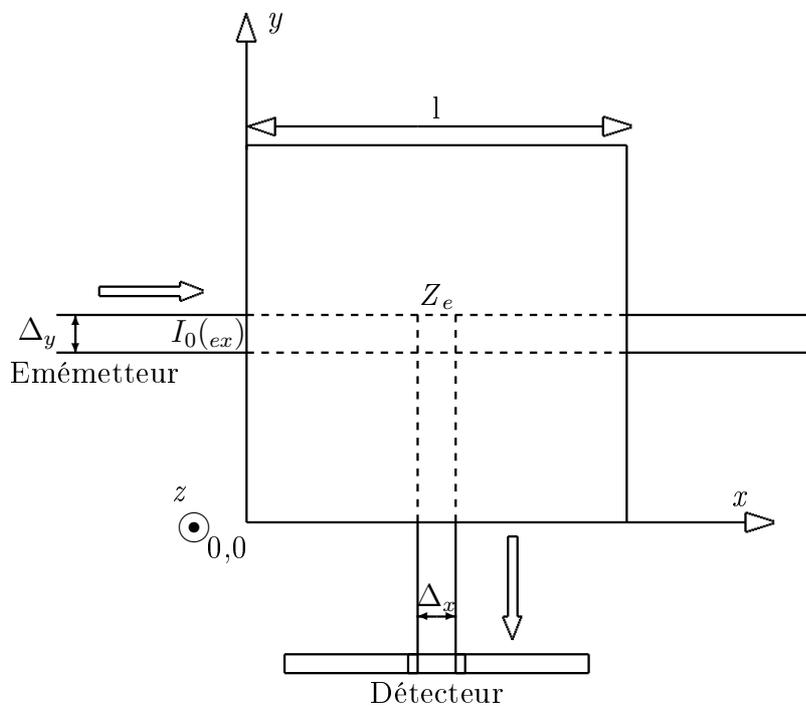
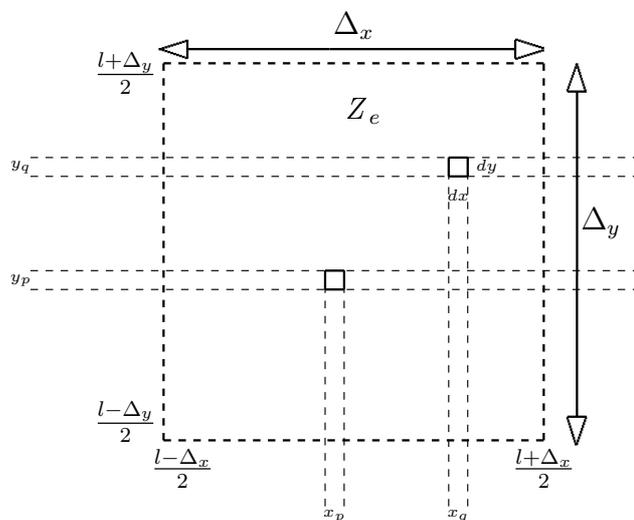


FIG. 3.3 – Schéma de principe du modèle

FIG. 3.4 – Zoom sur la zone d'émission Z_e

rendement quantique nul. Afin de simplifier les notations, sans restreindre le modèle à un cas particulier, nous assimilerons la solution à un mélange de N fluorophores dont le rendement quantique peut être nul. Enfin, à l'exception des quelques grandeurs introduites dans les paragraphes précédents, toutes les grandeurs utilisées ici ont été définies au chapitre 1, dont nous reprenons les notations.

3.2.2 Modèle théorique

La cuve peut se diviser en bandes élémentaires horizontales de dimensions (dy, l) ou en bandes élémentaires verticales (l, dx) . Pour une abscisse donnée, l'intensité locale du faisceau excitateur est supposée identique quelle que soit l'ordonnée, chaque bande horizontale reçoit en $x = 0$ une même fraction, $\frac{dyI_0}{\Delta y}$, de l'intensité totale du faisceau. La loi de Beer-Lambert permet alors de quantifier l'intensité transmise jusqu'aux surfaces élémentaires d'abscisse x . Dans chacune d'entre elles, le composant n absorbe alors une fraction, $\alpha_n dx$, de l'intensité transmise. Finalement, l'intensité totale I_{An} absorbée par le fluorophore n sur chaque bande d'émission et sur la longueur de la zone d'émission est donnée par :

$$I_{An}(\lambda_{ex}) = \frac{dyI_0}{\Delta y} \alpha_n(\lambda_{ex}) \int_{\frac{l-\Delta x}{2}}^{\frac{l+\Delta x}{2}} e^{-\alpha(\lambda_{ex})x} dx \quad (3.1)$$

$$I_{An}(\lambda_{ex}) = \frac{dyI_0}{\Delta y \alpha(\lambda_{ex})} \alpha_n(\lambda_{ex}) e^{-\frac{\alpha(\lambda_{ex})l}{2}} \left(e^{\frac{\alpha(\lambda_{ex})\Delta x}{2}} - e^{-\frac{\alpha(\lambda_{ex})\Delta x}{2}} \right) \quad (3.2)$$

$$I_{An}(\lambda_{ex}) = 2 \frac{dyI_0}{\Delta y \alpha(\lambda_{ex})} \alpha_n(\lambda_{ex}) e^{-\frac{\alpha(\lambda_{ex})l}{2}} \sinh \left(\frac{\alpha(\lambda_{ex})\Delta x}{2} \right) \quad (3.3)$$

Le signal de fluorescence émis par une bande horizontale à la longueur d'onde λ_{em} est égal à $\sum_n \Phi_n \gamma_n(\lambda_{em}) I_{An}$. La loi de Beer-Lambert permet cette fois de déterminer la fraction de signal transmis au détecteur, en fonction de la position y de la bande. En intégrant sur toutes les bandes élémentaires horizontales contenues dans la zone d'émission, on obtient l'intensité de fluorescence transmise au détecteur :

$$I_{3D}(\lambda_{ex}, \lambda_{em}) = \int_{\frac{l-\Delta_y}{2}}^{\frac{l+\Delta_y}{2}} \sum_n \Phi_n \gamma_n(\lambda_{em}) I_{An} e^{-\alpha(\lambda_{em})y} \quad (3.4)$$

$$I_{3D}(\lambda_{ex}, \lambda_{em}) = \frac{4I_0 e^{-\frac{\alpha(\lambda_{ex})l}{2}} \sinh\left(\frac{\alpha(\lambda_{ex})\Delta_x}{2}\right) e^{-\frac{\alpha(\lambda_{em})l}{2}} \sinh\left(\frac{\alpha(\lambda_{em})\Delta_y}{2}\right)}{\Delta_y \alpha(\lambda_{ex}) \alpha(\lambda_{em})} \sum_n \alpha_n(\lambda_{ex}) \Phi_n \gamma_n(\lambda_{em}) \quad (3.5)$$

$\frac{\alpha(\lambda_{ex})\Delta_x}{2}$ et $\frac{\alpha(\lambda_{em})\Delta_y}{2}$ sont supposés suffisamment petits pour autoriser l'approximation suivante :

$$I_{3D}(\lambda_{ex}, \lambda_{em}) = I_0 \Delta_x \left(\sum_n \alpha_n(\lambda_{ex}) \Phi_n \gamma_n(\lambda_{em}) \right) e^{-\frac{\alpha(\lambda_{ex})l}{2}} e^{-\frac{\alpha(\lambda_{em})l}{2}} \quad (3.6)$$

Finalement, en posant : $g = \frac{l}{2}$ et $G_n = I_0 \Delta_x \Phi_n$ on peut regrouper tous les paramètres constants ou inconnus de la correction de l'effet d'écran. Cela conduit à l'expression finale du modèle. Dans la suite, afin d'alléger les expressions, ε ne sera plus le coefficient d'absorption molaire mais le coefficient d'absorption molaire normé.

$$I_{3D}(\lambda_{ex}, \lambda_{em}) = \left(\sum_{n=1}^N G_n c_n \varepsilon_n(\lambda_{ex}) \gamma_n(\lambda_{em}) \right) \prod_{n=1}^N e^{-g(c_n \varepsilon_n(\lambda_{ex}) + c_n \varepsilon_n(\lambda_{em}))} \quad (3.7)$$

Le modèle trilineaire est donc enrichi de deux termes non linéaires multiplicatifs, traduisant l'influence des effets d'écran primaire et secondaire. On retombe finalement sur une reformulation du modèle de référence [60]. Par ailleurs, l'approximation au premier ordre de cette équation conduit bien au modèle trilineaire classique. Dans le cas de l'analyse d'une seule MEEF, la relation précédente est une relation bilinéaire entre les pseudo-spectres d'excitation (ε'_n) et d'émission (γ'_n) des différents composants définis par :

$$\varepsilon'_n(\lambda_{ex}) = G_n c_n \varepsilon_n(\lambda_{ex}) e^{-\sum_{p=1}^N c_p g \varepsilon_p(\lambda_{ex})} \quad (3.8)$$

$$\gamma'_n(\lambda_{em}) = \gamma_n(\lambda_{em}) e^{-\sum_{p=1}^N c_p g \varepsilon_p(\lambda_{em})} \quad (3.9)$$

En revanche, un jeu de MEEF ne satisfait plus aucune relation trilineaire de trois variables.

3.3 Validation expérimentale

Dans cette section, nous évaluons la capacité du modèle précédent à décrire expérimentalement les mesures de fluorescence obtenues à partir de différentes solutions

caractérisées par un fort effet d'écran. Notons que nous ne connaissons pas précisément tous les paramètres du modèle, en particulier les rendements quantiques et l'intensité des coefficients d'absorption molaires. La validation se fera donc en utilisant une méthode d'optimisation non linéaire avec les constantes inconnues de l'équation 3.7 comme paramètres. L'algorithme choisi est l'algorithme du simplexe [56].

3.3.1 Solution pure

La restriction du modèle au spectre d'émission d'une solution pure est obtenue en remplaçant tous les termes indépendants de la concentration et de la longueur d'onde d'émission par une constante multiplicative G' :

$$I_{3D}(\lambda_{em}, c) = G' c \gamma(\lambda_{em}) e^{-g c \varepsilon(\lambda_{em})} \quad (3.10)$$

Si on ne s'intéresse qu'à l'influence de la concentration, on fixe λ_{em} et l'on a :

$$I_{3D}(c) = G'' c e^{-g' c} \quad (3.11)$$

Cette équation devrait modéliser correctement l'évolution d'un point du spectre d'émission de la fluorescéine en fonction de la concentration. Après optimisation de G'' et g' , la courbe théorique obtenue (figure 3.5) est en bonne adéquation avec les données expérimentales (variance résiduelle relative : $3.68 \cdot 10^{-4}$).

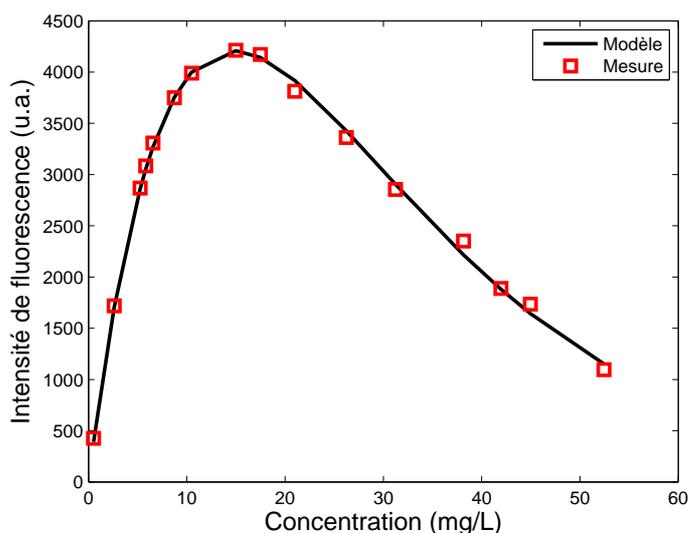


FIG. 3.5 – Modélisation de l'effet d'écran sur la fluorescéine seule

3.3.2 Mélanges à deux composants

Nous comparons ici les spectres 3D des cinq mélanges de fluorescéine et de sulfate de quinine du tableau 3.1 avec le modèle suivant :

$$I_{3D}(\lambda_{ex}, \lambda_{em}) = (G_1 c_1 \varepsilon_1(\lambda_{ex}) \gamma_1(\lambda_{em}) + G_2 c_2 \varepsilon_2(\lambda_{ex}) \gamma_2(\lambda_{em})) \times e^{-g_1(c_1 \varepsilon_1(\lambda_{ex}) + c_1 \varepsilon_1(\lambda_{em}))} e^{-g_2(c_2 \varepsilon_2(\lambda_{ex}) + c_2 \varepsilon_2(\lambda_{em}))} \quad (3.12)$$

Les indices 1 et 2 distinguent les termes relatifs aux deux composants. Les valeurs optimales de G_1 , G_2 , g_1 , g_2 sont déterminées par l'algorithme du simplexe. ε_1 , ε_2 , γ_1 , γ_2 sont estimés par les spectres d'excitation et d'émission des deux composants mesurés sur des solutions pures de faible absorbance et divisés par leur aire. Le critère de convergence retenu pour la minimisation du simplexe est l'écart absolu entre le spectre mesuré et le spectre modélisé (L_1). En effet, le critère quadratique L_2 se justifie essentiellement dans le cas d'un résidu gaussien dont il minimise la variance, or ce n'est pas le cas ici. La figure 3.6 illustre les résultats obtenus. Pour les cinq mélanges, l'erreur relative entre les spectres modélisés et les spectres mesurés est inférieure à 10% sur une large partie du domaine spectral, correspondant au signal d'intérêt. On note toutefois une erreur plus importante autour du maximum de la fluorescéine (massif secondaire en bas des spectres 3D). Nous avons testé une variante du modèle 3.7 permettant de tenir compte d'une possible réémission de la fluorescéine sans amélioration notable. Bien que ce modèle ne soit pas parfait, il apporte une amélioration conséquente du modèle linéaire tout en ayant une formulation relativement simple. Par ailleurs, nous verrons dans le prochain chapitre que la description du signal mesuré est suffisamment précise pour effectuer une correction satisfaisante de l'effet d'écran.

A l'inverse, il n'est pas judicieux d'utiliser un modèle plus simple. En effet, il est possible d'évaluer les importances respectives des deux termes non linéaires, et donc de l'EEP et de l'EES. D'une manière générale, si l'on néglige le terme lié à l'EES, l'erreur résiduelle globale est multipliée par un facteur compris entre 2 et 3. Si l'on ne tient pas compte de l'EEP, ce facteur est supérieur à 10.

Enfin, on peut comparer pour chaque mélange, la valeur des paramètres optimaux avec ceux obtenus pour les mêmes mélanges dilués d'un facteur deux. Logiquement ceux-ci devraient être divisés par deux. En pratique, on obtient pour le mélange n°3 les rapports suivants : 2.03, 1.95 1.85 1.92, soit une moyenne de 1.94 avec un écart type inférieur à 0.08.

3.3.3 Mélanges à trois composants

Nous avons utilisé ici quatre mélanges de phénylalanine, tyrosine et tryptophane dont les concentrations respectives sont données par le tableau 3.2. Ces trois composants répondent bien à nos critères. En outre, ils possèdent tous deux pics d'excitations très distincts, favorisant les effets de réabsorption. Cependant, lors de la mesure, le second pic de fluorescence (situés plus haut dans le domaine d'excitation) de la phénylalanine et de la tyrosine chevauche fortement les spectres de diffusion de Raman. La méthode de correction numérique employée donne de bons résultats sur les séries de solutions concentrées. En revanche ce n'est pas le cas sur les MEEF des solutions

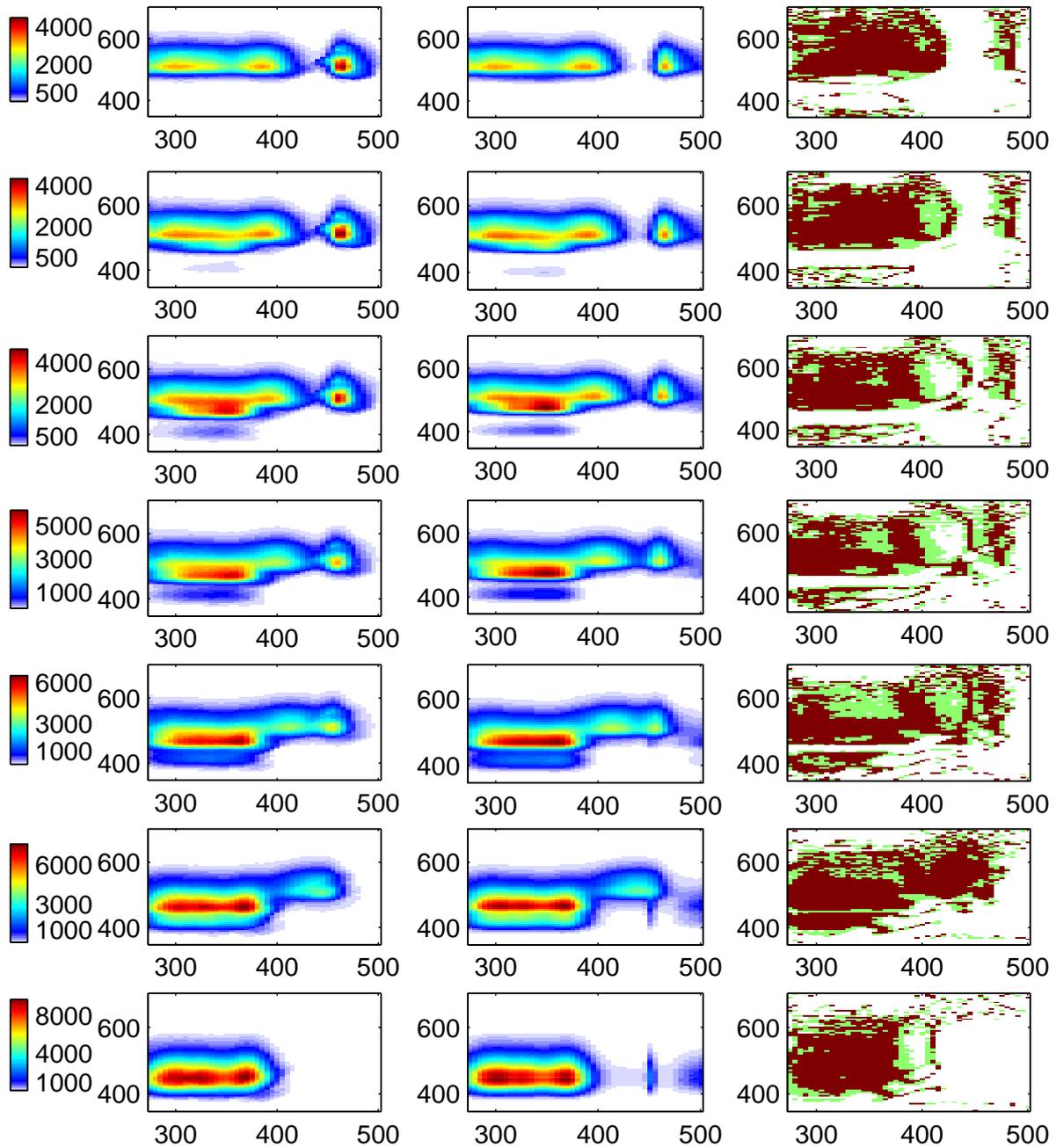


FIG. 3.6 – Mélange à deux composants, comparaison entre les MEEF mesurées et modélisées

Les figures de droite représentent une distribution grossière de l'erreur relative, en bleu, celle-ci est supérieure à 20%, en vert elle est comprise entre 10% et 20%, en rouge, elle est inférieure à 10 %

| Mélange | [PHE] ($10^{-6}mol.L^{-1}$) | [TYR] ($10^{-6}mol.L^{-1}$) | [TRP] $10^{-6}mol.L^{-1}$ |
|---------|-------------------------------|-------------------------------|---------------------------|
| 1 | 83 | 16.7 | 6 |
| 2 | 1250 | 12.5 | 4.5 |
| 3 | 625 | 25 | 4.5 |
| 4 | 625 | 12.5 | 9 |

TAB. 3.2 – Concentrations des mélanges de phénylalanine (PHE), tyrosine (TYR) et tryptophane (TRP)

diluées devant servir de référence et pour lesquelles le niveau du signal de diffusion de Raman est aussi important que celui du signal de fluorescence. Afin d'éviter tout artefact dû à la correction numérique nous nous contenterons donc d'étudier le phénomène sur les premiers pics de fluorescence dont l'excitation se situe en deça de 240 nm.

Le protocole de test est identique au cas à 2 composants si ce n'est que nous utilisons cette fois le modèle de rang 3, dépendant donc de 6 paramètres : $G_1, G_2, G_3, g_1, g_2, g_3$. Les résultats sont donnés par la figure 3.7. Là encore, ceux-ci sont satisfaisants puisque les seules erreurs notables concernent des pics secondaires, en particulier le pic situé autour du point (216/410 nm) du second mélange qui disparaît. L'amplitude et l'étalement spectral de ces pics sont faibles. Ils ont donc une influence moindre sur le critère d'optimisation, ce qui pourrait expliquer ces erreurs.

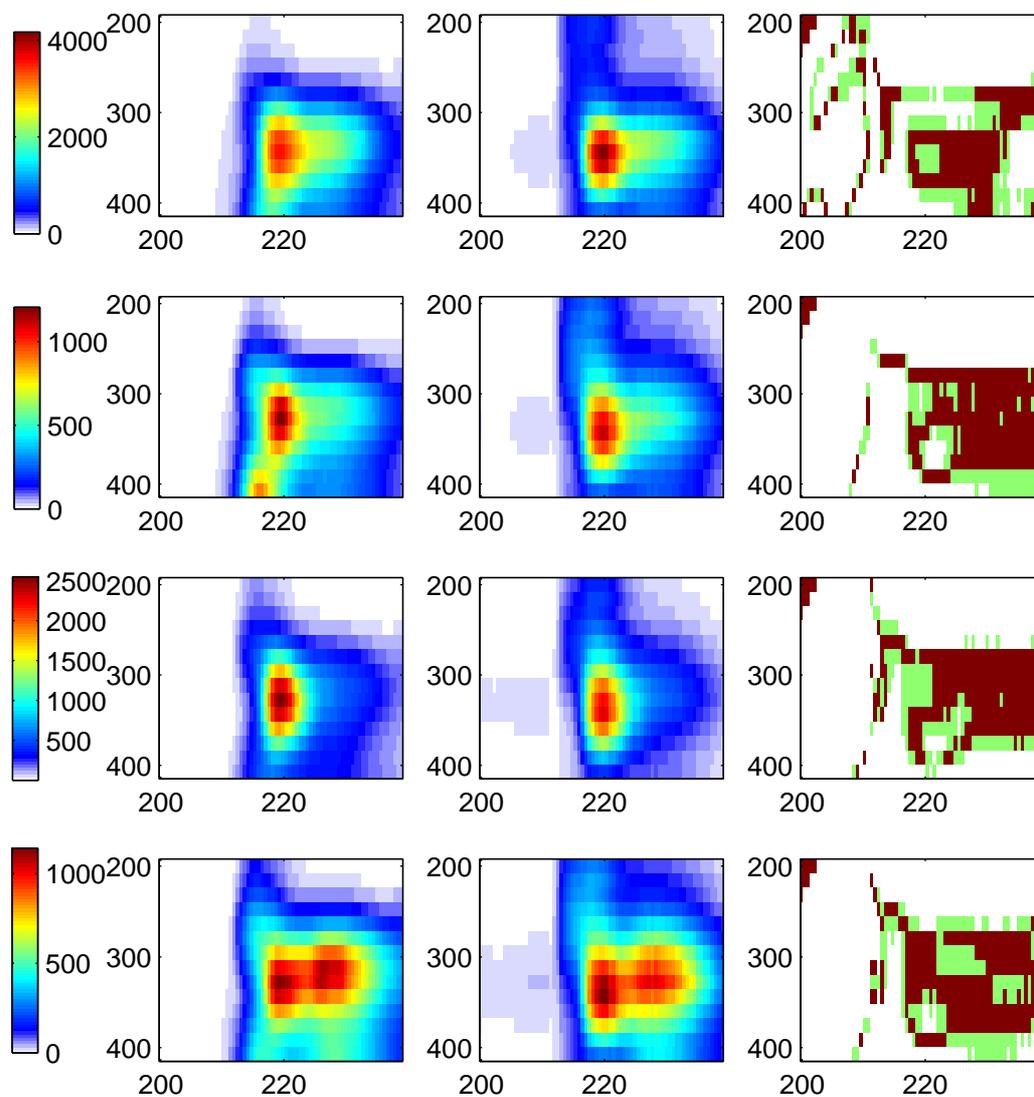


FIG. 3.7 – Mélange à trois composants : comparaison entre les MEEF mesurées et modélisées

Les figures de droite représentent une distribution grossière de l'erreur relative, en blanc, celle-ci est supérieure à 20%, en vert elle est comprise entre 10% et 20%, en rouge, elle est inférieure à 10 %

Chapitre 4

Correction des effets d'écran

4.1 Introduction

Nous disposons à présent d'un modèle mathématique général reliant l'intensité de fluorescence mesurée par un spectrofluorimètre standard aux trois grandeurs caractérisant les composants d'un mélange : concentration, spectre d'excitation et spectre d'émission. Ce modèle dont nous rappelons l'expression (équation 4.1) a été validé sur plusieurs mélanges connus de deux ou de trois fluorophores présentant d'importants effets d'écran.

$$I_{3D}(\lambda_{ex}, \lambda_{em}) = \left(\sum_{n=1}^N G_n c_n \varepsilon_n(\lambda_{ex}) \gamma_n(\lambda_{em}) \right) \prod_{n=1}^N e^{-g(c_n \varepsilon_n(\lambda_{ex}) + c_n \varepsilon_n(\lambda_{em}))} \quad (4.1)$$

Nous avons aussi vu que le traitement classique de tels mélanges nécessite deux étapes : dans un premier temps, on corrige l'effet d'écran (linéarisation) puis, on décompose le mélange sur une base proche des spectres individuels (analyse). Nous reprenons cette démarche à notre compte en ne nous intéressant ici qu'à la linéarisation des données.

Note : Dans toute la suite, les mélanges ou les MEEF présentant un effet d'écran non négligeable seront dit « non linéaires », par opposition aux mélanges dit « linéaires » respectant le modèle trilinéaire 1.24.

La correction se fait généralement en mesurant l'absorbance du mélange. Cependant, les techniques de mesure de la fluorescence et de l'absorbance d'une même solution sont sensiblement différentes (cf. chapitre 1) et ne peuvent être réalisées avec le même matériel. Certains spectrofluorimètres disposent de cellules spécifiques et permettent les deux types de mesures mais un traitement des spectres est nécessaire. Lorsque ce n'est pas le cas, on a recours à deux appareils différents. Dans tous les cas, l'utilisation de deux dispositifs de mesure différents, chacun avec leurs propres caractéristiques et tolérances a généralement pour conséquence d'introduire des biais expérimentaux. Afin d'éviter ce type de problème, nous proposons dans ce chapitre trois procédures originales de correction de l'effet d'écran ne nécessitant pas la mesure de l'absorbance.

Le chapitre 2 nous a permis de classer les principales méthodes de décomposition en fonction de leur champ d'application à des données du premier ou du second ordre. Il n'en va pas de même pour les méthodes de correction proposées. Les spectrofluorimètres actuels permettent d'obtenir des spectres 3D de bonne qualité. Ces spectres recellent beaucoup plus d'informations qu'un simple spectre 2D et sont aujourd'hui majoritairement utilisés pour les applications qui nous intéressent. Nous restreindrons donc le problème de la correction au traitement des MEEF. Les méthodes que nous proposons nécessitent toutes un jeu de données du second ordre formé d'au moins deux MEEF (suivant la méthode). Ceci pour une raison bien simple, il est inutile de vouloir essayer de corriger l'effet d'écran si l'on est pas capable de le détecter. Peut-on savoir, sans aucune autre connaissance sur l'échantillon que sa MEEF, si cette dernière est entachée ou non d'effet d'écran ? De plus, il peut être intéressant de savoir si un mélange est linéaire ou non à partir du plus petit nombre de données, même si celles-ci ne sont pas suffisantes pour la correction. Comme nous l'avons fait remarquer au chapitre précédent, la relation 4.1 est une relation bilinéaire. Non pas entre les spectres d'excitation et d'émission réels des fluorophores du mélange, mais entre ce que nous appelons les pseudospectres des composants :

$$\varepsilon'_n(\lambda_{ex}) = G_n c_n \varepsilon_n(\lambda_{ex}) e^{-\sum_{p=1}^N c_p g \varepsilon_p(\lambda_{ex})} \quad (4.2)$$

$$\gamma'_n(\lambda_{em}) = \gamma_n(\lambda_{em}) e^{-\sum_{p=1}^N c_p g \varepsilon_p(\lambda_{em})} \quad (4.3)$$

En pratique, ces spectres bien que généralement très différents des spectres réels, présentent toutes les caractéristiques de spectres 2D bien physiques. S'agit-il alors de la MEEF d'un mélange non linéaire ou d'un mélange linéaire entre d'autres groupements fluorescents ? Nous n'avons pas à l'heure actuelle développé d'outil d'analyse capable de répondre à cette question avec certitude. Pour ce faire, il nous est nécessaire de disposer de données supplémentaires. En particulier, la MEEF d'un autre mélange (linéaire ou non) à des concentrations différentes introduit une variation suffisante de l'argument des exponentielles pour supprimer, le cas échéant, le caractère linéaire. Dans ce cas, les spectres d'excitation ou d'émission des deux MEEF n'appartiennent pas aux mêmes sous-espaces vectoriels. La comparaison des deux bases principales permet alors de savoir si au moins l'un des deux mélanges est non linéaire. Cependant, les différentes procédures que nous avons imaginées, permettant d'affiner ce diagnostic à partir de telles données, reviennent de fait à corriger l'effet d'écran ! Etant maintenant entendu que ces deux problèmes, détection et correction, nécessitent les mêmes contraintes pour être résolus, nous ne nous attarderons pas plus sur ce problème de détection. Nous considérons donc à présent que nos données sont constituées de plusieurs spectres 3D.

La première méthode que nous présentons est à la fois la plus simple et la plus aboutie. De plus, chaque MEEF du jeu de données est corrigée indépendamment des autres. En revanche, la procédure de correction requiert la mesure d'un nouveau spectre 3D pour chaque mélange à corriger. Nous consacrons la seconde partie de ce chapitre à la description de la procédure de correction et à sa validation expérimentale sur des mélanges réels. Nous verrons que cette approche donne de bon résultats, cependant

nous avons cherché à nous affranchir de toute mesure supplémentaire. En effet, on peut imaginer une situation dans laquelle, on ne dispose plus des échantillons pour effectuer la seconde mesure. C'est pourquoi, nous proposons dans la troisième partie de ce chapitre deux autres méthodes qui ne nécessitent aucune manipulation supplémentaire. La première d'entre elle, basée sur des relations statistiques entre les données n'a pas encore été testée rigoureusement, nous n'en donnerons donc que les principes. L'autre approche est purement arithmétique et donne de premiers résultats prometteurs que nous exposerons en fin de chapitre. Par conséquent, nous n'accorderons qu'une moindre part à leur description

Protocole de test. Les résultats seront jugés sur deux plans : la qualité brute de la linéarisation et les répercussions de l'erreur de correction sur la procédure de décomposition.

Afin d'évaluer la qualité de la linéarisation, nous comparons pour un mélange donné, la MEEF corrigée avec la MEEF du même mélange fortement dilué. Le facteur de dilution est choisi de sorte que le mélange soit linéaire et que le rapport signal sur bruit soit suffisant après dilution. Pour cette étape nous nous intéressons principalement à la forme du spectre 3D. Toutes les MEEF que nous comparons auront donc été normées par leur volume. L'amplitude du signal ne sera considérée que dans la seconde partie de la validation. Des lors, plusieurs critères d'erreur peuvent être envisagés. Dans un premier temps, nous calculerons l'écart relatif entre la MEEF linéarisée et la MEEF linéaire en chaque point. Néanmoins, ce simple critère nous semble incomplet pour décrire complètement la qualité de la correction. C'est pourquoi nous utiliserons deux autres tests : a) Le rapport entre l'énergie de l'erreur et celle de l'image de référence permet d'avoir une idée plus globale du résultat. b) Afin d'affiner ce résultat tout en limitant l'influence des domaines de longueur d'onde où le signal est trop faible, nous comparons également les N premières composantes principales des MEEF de référence avec celles des MEEF originales et corrigées (dans le cas d'un mélange de N fluorophores). La comparaison s'effectue dans un premier temps sur les MEEF (comparaison en excitation) puis sur les MEEF transposées (comparaison en émission). Ce dernier test vient compléter l'analyse localisée fournie par l'erreur résiduelle. Cela permet de vérifier si les deux MEEF partagent la même structure. En réalité, ce second critère préfigure la seconde étape de validation puisqu'il s'agit également d'une procédure de décomposition. En revanche, celle-ci à l'avantage d'être beaucoup simple et par la même, plus robuste. En effet, nous devons ensuite vérifier dans quelle mesure l'erreur due à la correction influence la décomposition des mélanges. L'algorithme de décomposition utilisé est PARAFAC et nous comparerons pour chaque composant des mélanges, leurs spectres 2D et leurs concentrations relatives avec les spectres 2D et les concentrations relatives estimées par PARAFAC sur les mélanges linéaires (dilués), sur les mélanges non linéaires (originaux) et sur les mélanges linéarisés (originaux après correction). La décomposition s'effectue en deux temps : nous estimons d'abord les loadings spectraux en lançant PARAFAC sur les matrices normées puis, pour ne pas perdre l'information sur les concentrations relatives, leur mode est estimé par PARAFAC sur les matrices non normées mais en fixant les modes spectraux aux valeurs obtenues précédemment.

Notations. Nous conserverons le système d'indices utilisé dans la section sur PARAFAC (2.3.3) : les lettres j et k désignent respectivement les longueurs d'onde d'émission et d'excitation, i , l'indice du mélange ou de l'échantillon et f celui du fluorophore. Les majuscules représentent les valeurs maximales de chaque indice. Par ailleurs, nous noterons $L(i, j, k)$ le terme linéaire du modèle et $H(i, j, k)$ la déviation due à l'effet d'écran. On a alors,

$$I_{3D}(i, j, k) = L(i, j, k)H(i, j, k) \quad (4.4)$$

En posant :

$$h(i, m) = \prod_{n=1}^N e^{-g c_n(i) \varepsilon_n(m)} = e^{-\sum_{n=1}^N g c_n(i) \varepsilon_n(m)} \quad (4.5)$$

où m peut être une longueur d'onde d'excitation ou d'émission, on a

$$I_{3D}(i, j, k) = L(i, j, k)h(i, j)h(i, k) \quad (4.6)$$

Enfin, nous noterons $w = \ln(h)$. La procédure de linéarisation consistera donc à déterminer L , soit directement, soit en déterminant h au préalable.

4.2 Méthode de correction par faible dilution

La procédure de correction que nous proposons ici ne nécessite qu'un seul échantillon dont le spectre 3D est mesuré à deux facteurs de dilution différents. Pour cette description, l'indice i des échantillons est rendu muet. Nous noterons simplement $I_{3D}(j, k)$ l'échantillon d'intérêt et $I_{3Dp}(j, k)$ l'échantillon du même mélange dilué d'un facteur p .

4.2.1 principe

Le modèle non linéaire 4.1 de $I_{3Dp}(j, k)$ s'écrit :

$$I_{3Dp}(j, k) = \left(\sum_{n=1}^N G_n \frac{c_n}{p} \varepsilon_n(k) \gamma_n(j) \right) \prod_{n=1}^N e^{-g \left(\frac{c_n}{p} \varepsilon_n(j) + \frac{c_n}{p} \varepsilon_n(k) \right)} \quad (4.7)$$

On a donc,

$$I_{3D}(j, k) = \left(\sum_{n=1}^N G_n c_n \varepsilon_n(k) \gamma_n(j) \right) \prod_{n=1}^N e^{-g(c_n \varepsilon_n(j) + c_n \varepsilon_n(k))} \quad (4.8)$$

$$I_{3Dp}(j, k) = \frac{1}{p} \left(\sum_{n=1}^N G_n c_n \varepsilon_n(k) \gamma_n(j) \right) \left(\prod_{n=1}^N e^{-g(c_n \varepsilon_n(j) + c_n \varepsilon_n(k))} \right)^{\frac{1}{p}} \quad (4.9)$$

Ou plus simplement :

$$I_{3D}(j, k) = L(j, k)H(j, k) \quad (4.10)$$

$$I_{3Dp}(j, k) = \frac{1}{p} L(j, k)H^{\frac{1}{p}}(j, k) \quad (4.11)$$

$$(4.12)$$

Ce système d'équation peut se résoudre simplement, on a alors,

$$H(j, k) = \left(\frac{I_{3D}(j, k)}{pI_{3Dp}(j, k)} \right)^{\frac{p}{p-1}} \quad (4.13)$$

$$L(j, k) = \left(\frac{(pI_{3Dp}(j, k))^p}{I_{3D}(j, k)} \right)^{\frac{1}{p-1}} \quad (4.14)$$

$$(4.15)$$

Ainsi, l'équation 4.15 permet de déterminer la MEEF linéarisée en utilisant une simple dilution d'un facteur connu. Il faut noter que ce facteur peut être choisi suffisamment faible pour éviter de détériorer le rapport signal sur bruit de I_{3Dp} .

4.2.2 Bases de test

La constitution de bases de test pertinentes demande du temps et une certaine attention. Il convient ici de rendre hommage à l'expérimentateur en s'y attardant quelque peu. Les fluorophores choisis doivent répondre à plusieurs contraintes : Leurs gammes d'excitations et d'émissions doivent être proches afin de favoriser l'effet d'écran. Leurs spectres 2D doivent être connus ou facilement mesurables. Lors du mélange, il ne doivent présenter aucune interaction influençant leurs spectres où empêchant la mesure. On évitera enfin tout produit dangereux pour lequel le risque encouru serait jugé excessif devant l'intérêt présumé de la manipulation. Il restera alors à déterminer les deux gammes de concentrations : linéaires et non linéaires. Afin d'établir la gamme linéaire, on utilise pour le mélange et pour chaque fluorophore des courbes de dilution telles que celles de la figure 1.10. Lorsque tous les mélanges supposés linéaires sont constitués, la comparaison des spectres 2D mesurés pour chaque fluorophore avec ceux estimés par PARAFAC permet une seconde vérification du caractère linéaire des mélanges. En revanche, nous verrons qu'il n'est pas toujours simple de concilier linéarité et rapport signal sur bruit. Enfin, les concentrations relatives entre les fluorophores ne doivent pas masquer un des composants outre-mesure. Il va de soi que toutes ces contraintes ne sont pas liées aux procédures de correction de l'effet d'écran que nous proposons ici mais à la constitution de bases de test propres à valider ou non ces différentes approches.

Finalement, nous utiliserons trois jeux de données réelles différents pour tester cette méthode :

- Un jeu de 3×7 mélanges réels de deux fluorophores.
- Un jeu de 3×4 mélanges réels de trois fluorophores.
- Un jeu de 3×11 échantillons de matière organique.

Ces trois jeux sont composés de trois séries de mélanges. Les premières séries contiennent les mélanges originaux, suffisamment concentrés pour présenter un effet d'écran conséquent. Les secondes séries représentent les mêmes mélanges dilués d'un faible facteur (typiquement 2) qui nous serviront pour la correction. Les dernières séries sont également constituées par les mêmes mélanges, mais fortement dilués, de sorte que les distortions non linéaires soient négligeables.

Les MEEF que nous aurons à traiter possèdent un rapport signal sur bruit élevé.

| Solution | p | σ_p |
|----------|------|------------|
| 1 | 2.11 | 0.034 |
| 2 | 2.08 | 0.032 |
| 3 | 2.07 | 0.030 |
| 4 | 2.07 | 0.029 |
| 5 | 2.07 | 0.030 |
| 6 | 2.08 | 0.032 |
| 7 | 2.11 | 0.034 |

TAB. 4.1 – Facteur de dilution utilisé pour les sept solutions de fluorescéine et sulfate de quinine.

L'incertitude sur la valeur du facteur de dilution (p) est donné par son écart type (σ_p)

Cependant, la présence de bruit, même négligeable peut limiter les performances de nos procédures de correction, en particulier sur les deux derniers jeux de données, pour lesquels le bruit est plus important. Les MEEF seront donc filtrées. Nous avons vérifié au préalable que l'algorithme utilisé, mentionné en introduction du chapitre 2, ne détériore pas la structure du signal.

Mélanges de deux fluorophores. Nous réutilisons pour la première série les sept mélanges de fluorescéine et sulfate de quinine du chapitre 3 (tableau 3.1). Par commodité, nous avons choisi pour la seconde série un facteur de dilution proche de deux. En réalité ce nombre varie légèrement d'une solution à l'autre. Les valeurs sont données dans le tableau 4.1 ainsi que l'écart type dû à l'utilisation des pipettes. Les MEEF linéaires de référence sont mesurées sur des solutions 30 fois diluées. Enfin, nous avons mesuré le spectre d'absorption de chaque mélange afin de comparer nos méthodes de correction avec l'approche classique.

Mélanges de trois fluorophores. Nous utilisons dans ce cas trois séries de mélanges de phénylalanine, tyrosine et tryptophane. La première série est composé des quatre mélanges utilisés au chapitre 3 (tableau 3.2). A l'inverse du cas précédent, les facteurs de dilution utilisés entre les trois jeux varient selon le mélange : le premier mélange de la seconde série est obtenu par une dilution d'un facteur 5 du premier mélange de la première série et les trois autres après une dilution par 2 des mélanges correspondants. La troisième série consiste en une dilution par 17 du premier mélange original et par 50 des trois autres.

Matière organique. Ce jeu d'échantillons est composé de prélèvements de matière organique dans des sols. Celle-ci est très concentrée ce qui laissait présupposer d'un fort effet d'écran. Nous avons pu le vérifier en diluant les échantillons. La seconde série est diluée 2 fois par rapport à la première et la troisième 30 fois.

| Solution | r Orig. | r Corr.Abs. | r Corr.Dil. |
|----------|-----------|---------------|---------------|
| 1 | 117 | 31 | 11 |
| 2 | 109 | 38 | 5 |
| 3 | 81 | 59 | 2 |
| 4 | 46 | 33 | 1.8 |
| 5 | 29 | 14 | 0.3 |
| 6 | 20 | 7 | 0.06 |
| 7 | 13 | 8 | 0.03 |

TAB. 4.2 – Jeu 1 : Comparaison en énergie avec les MEEF de référence
Rapports entre l'énergie de l'erreur résiduelle et celle du signal de référence pour les MEEF originales (Orig.), corrigées par le spectre d'absorption (Corr.Abs.) et corrigées par notre approche (Corr.Dil.) (r). Les valeurs sont données en pourcentages

4.2.3 Résultats expérimentaux

4.2.3.1 Mélanges à 2 composants

Nous exposons ici les résultats obtenus sur la première série de mélanges : sulfate de quinine et fluorescéine. Les figures 4.1 à 4.7 montrent les résultats de la linéarisation des sept solutions de la série. Nous comparons les MEEF de référence avec les MEEF originales, les MEEF corrigées par le spectre d'absorption et les MEEF corrigées par la méthode de faible dilution que nous proposons. Le tableau 4.2 donne pour chaque solution et pour chaque type de MEEF le rapport r entre l'énergie de l'erreur et l'énergie du signal de référence. Plus la valeur du rapport est proche de 0, plus la MEEF est proche de la référence.

Quel que soit le mélange, une première comparaison des images originales, diluées et corrigées nous permet immédiatement deux remarques : a) Notre approche réduit considérablement la distortion non linéaire b) Et cela, bien mieux qu'en utilisant le spectre d'absorption. Ces deux présomptions sont confirmées et quantifiées par la mesure des erreurs résiduelles. Cependant, malgré cette nette amélioration et bien que contenue en deça de 20% sur la majorité de la zone d'intérêt, celle-ci reste localement élevée et vient tempérer nos premières constatations, en particulier sur les deux premiers mélanges. A ce propos il nous semble intéressant de noter que les erreurs de correction les plus importantes interviennent sur les deux solutions qui donnaient déjà de moins bons résultats lors du test de validation du modèle (cf. section 3.3.2). Ce sont également ces solutions qui présentent la plus forte distortion non linéaire. L'étude des composantes principales constitue un niveau d'analyse intermédiaire entre l'analyse purement qualitative et l'erreur point par point. De ce point de vue, les MEEF linéarisées par notre approche semble bien pouvoir se décomposer sur le même espace que les MEEF diluées. Enfin le critère énergétique résume correctement la tendance générale : tout comme la distortion non linéaire, l'erreur de correction décroît avec la diminution de la concentration en fluorescéine. Celle-ci est totalement négligeable pour les solutions 3 à 7 et reste très acceptable sur les deux premières solutions.

La deuxième phase de test consiste à comparer les résultats de la décomposition

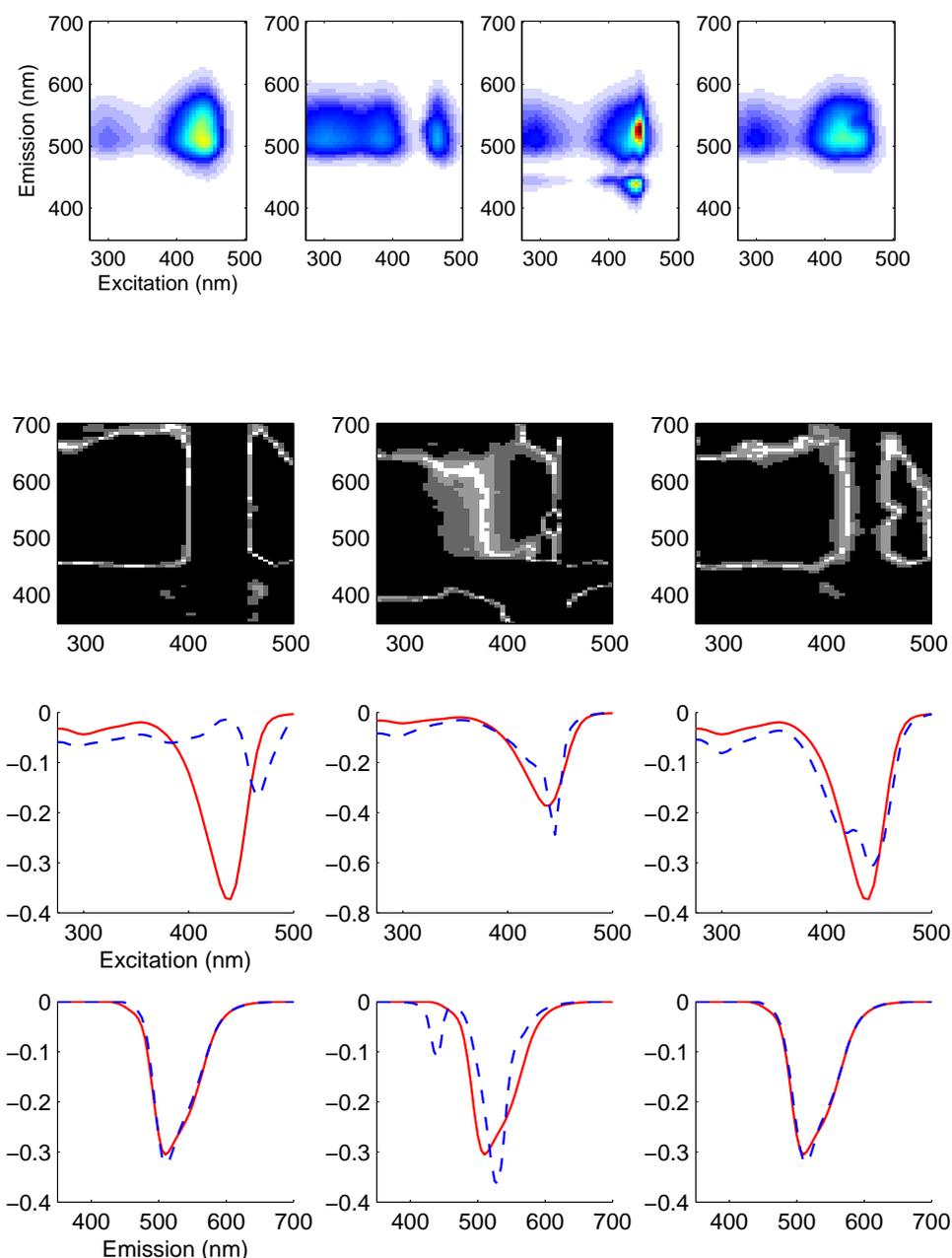


FIG. 4.1 – Jeu 1 solution 1 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées

Les figures de la ligne du haut représentent de gauche à droite la MEEF de référence, la MEEF originale, la MEEF corrigée par l'absorbance et la MEEF corrigée par la dilution. Les figures de la ligne du milieu représentent une distribution grossière de l'erreur relative entre la MEEF de référence et (de gauche à droite) la MEEF originale, la MEEF corrigée par l'absorbance et la MEEF corrigée par la dilution. Les quatre niveaux de gris, du blanc au noir, correspondent respectivement à une erreur inférieure à 5%, comprise entre 5% et 10%, comprise entre 10% et 20% et supérieure à 20%. Les deux lignes du bas montrent dans le même ordre les composantes principales significatives, en excitation et en émission. Celles de la MEEF de référence sont représentées en traits pleins

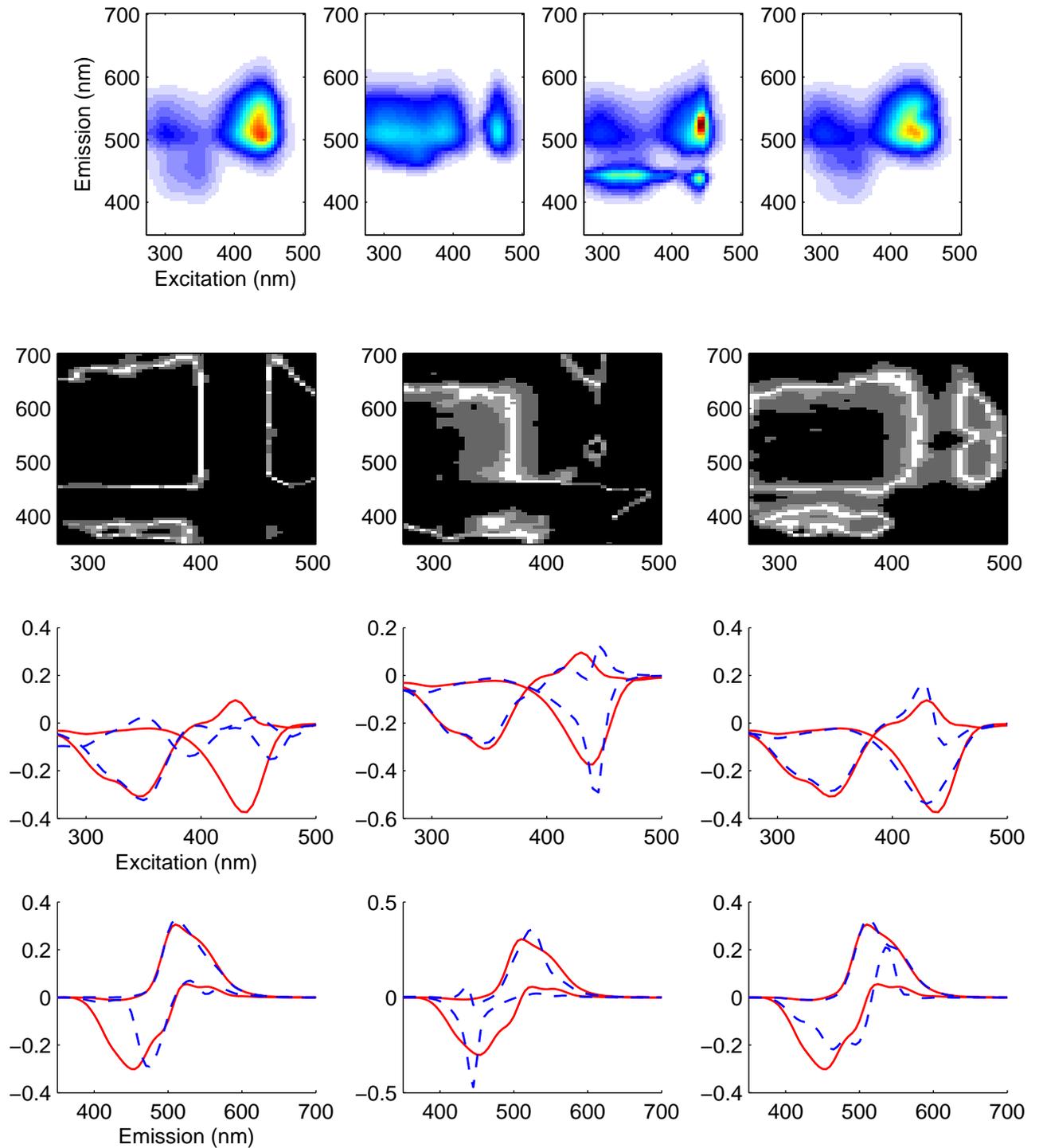


FIG. 4.2 – Jeu 1, solution 2 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées
Voir figure 4.1 pour la légende.

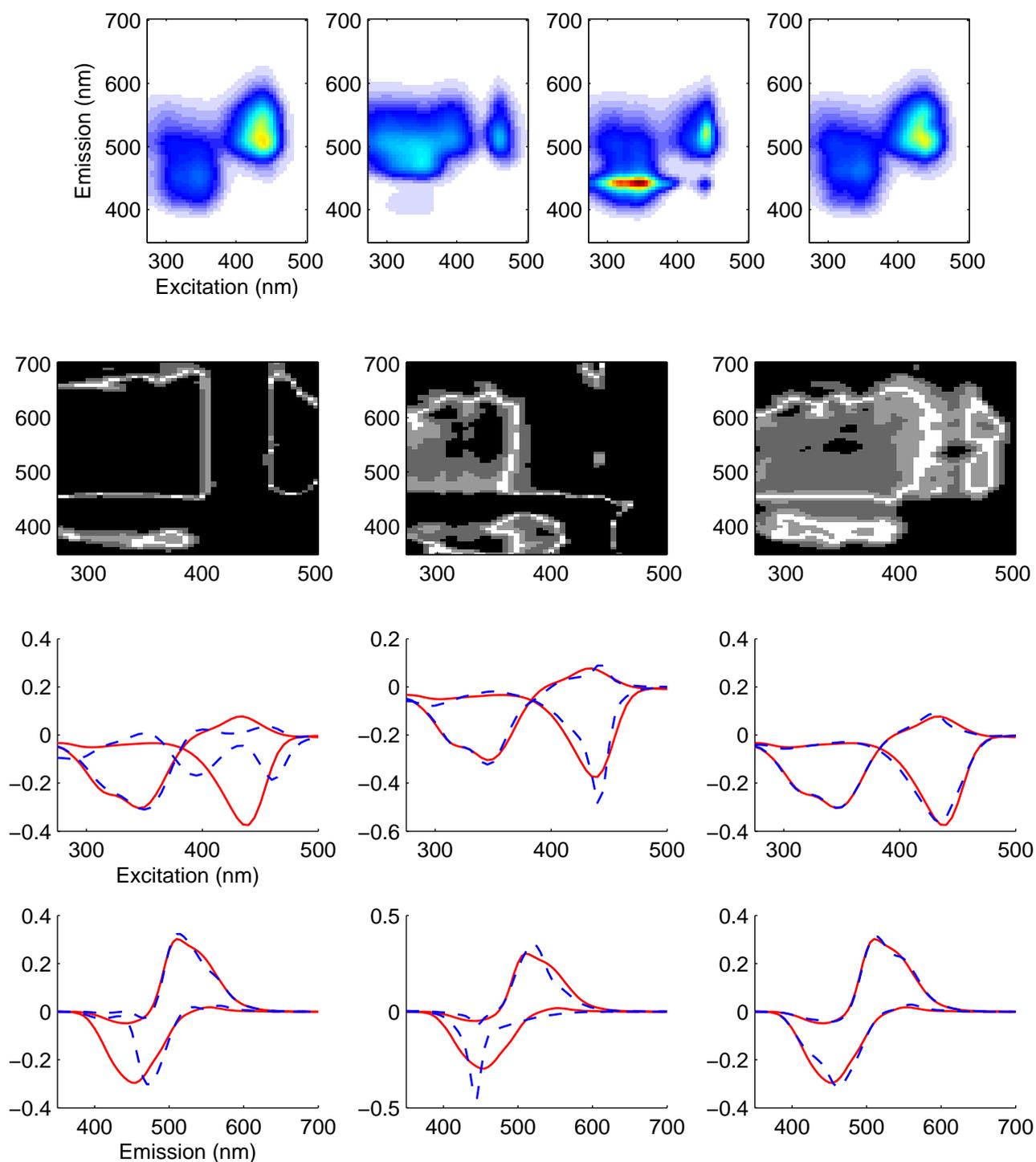


FIG. 4.3 – Jeu 1, solution 3 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées
Voir figure 4.1 pour la légende.

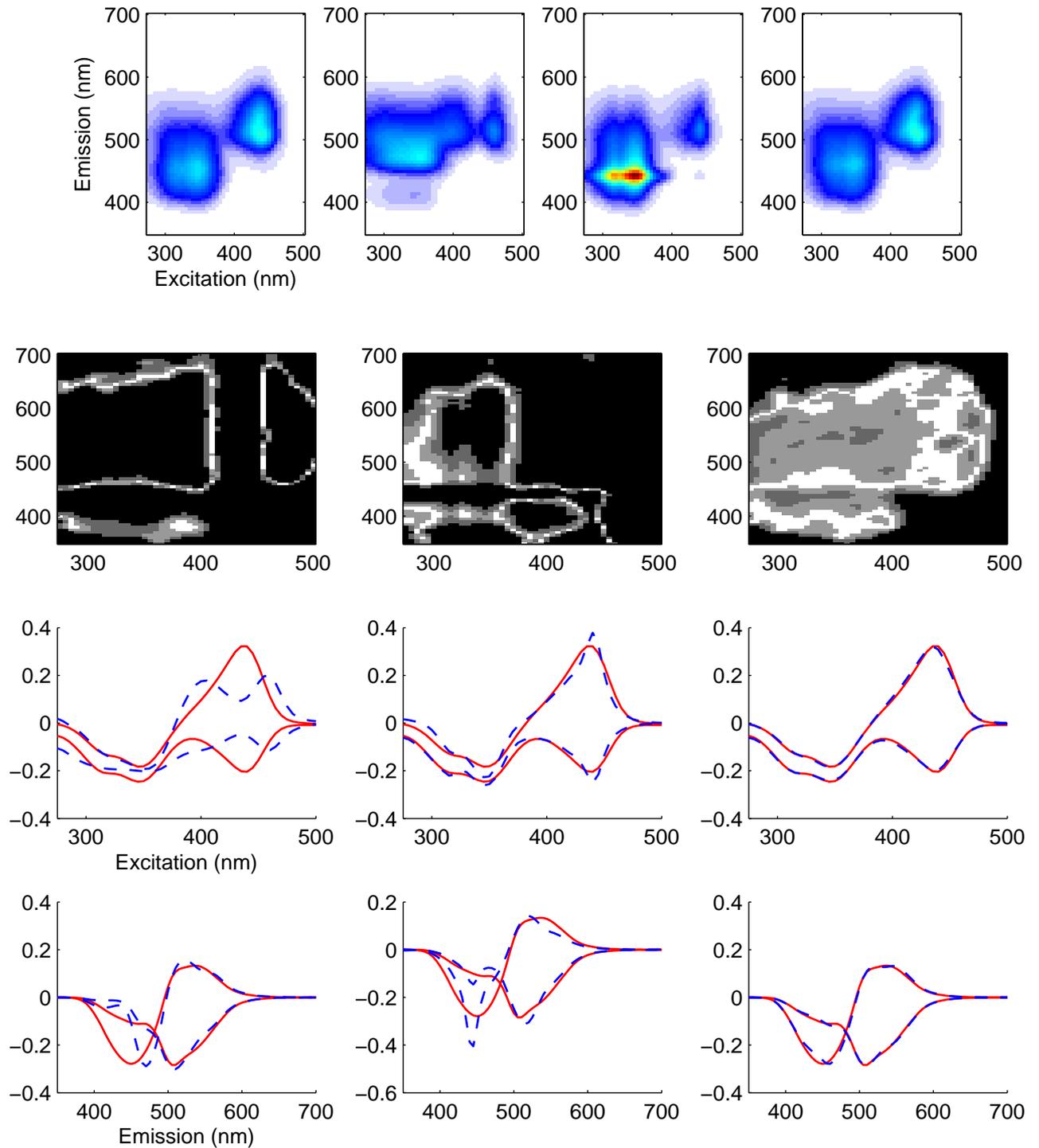


FIG. 4.4 – Jeu 1, solution 4 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées
 Voir figure 4.1 pour la légende.

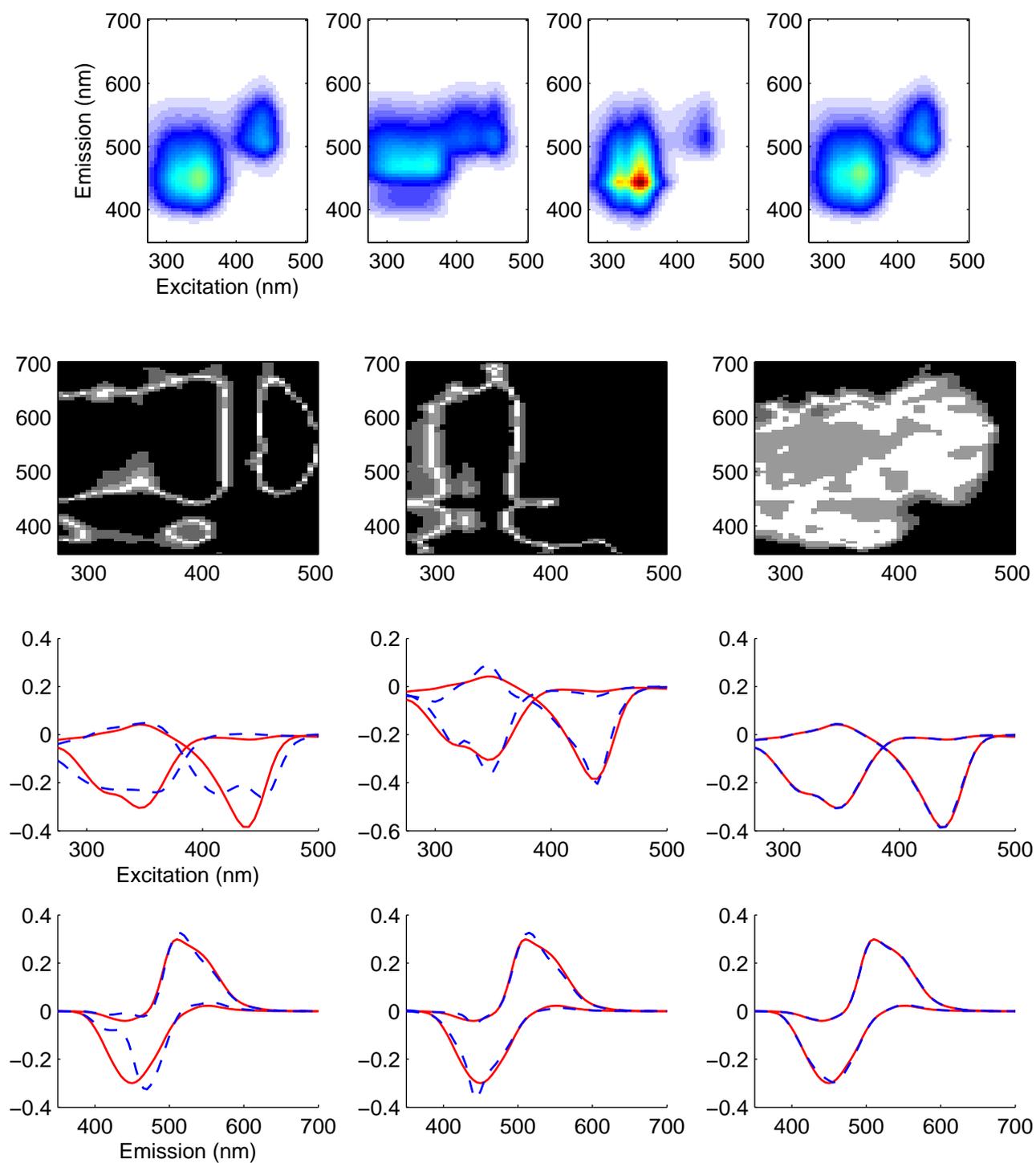


FIG. 4.5 – Jeu 1, solution 5 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées
Voir figure 4.1 pour la légende.

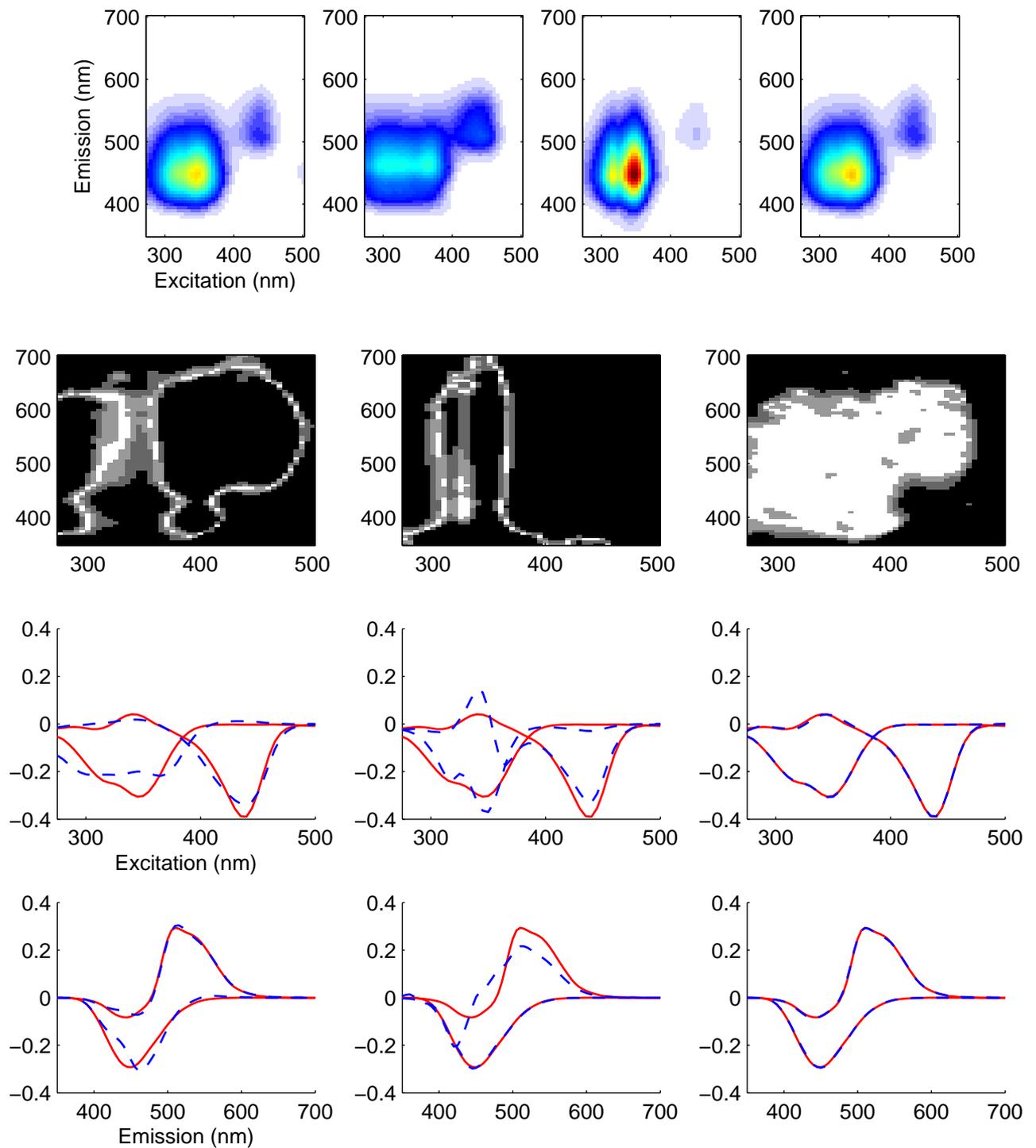


FIG. 4.6 – Jeu 1, solution 6 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées
Voir figure 4.1 pour la légende.

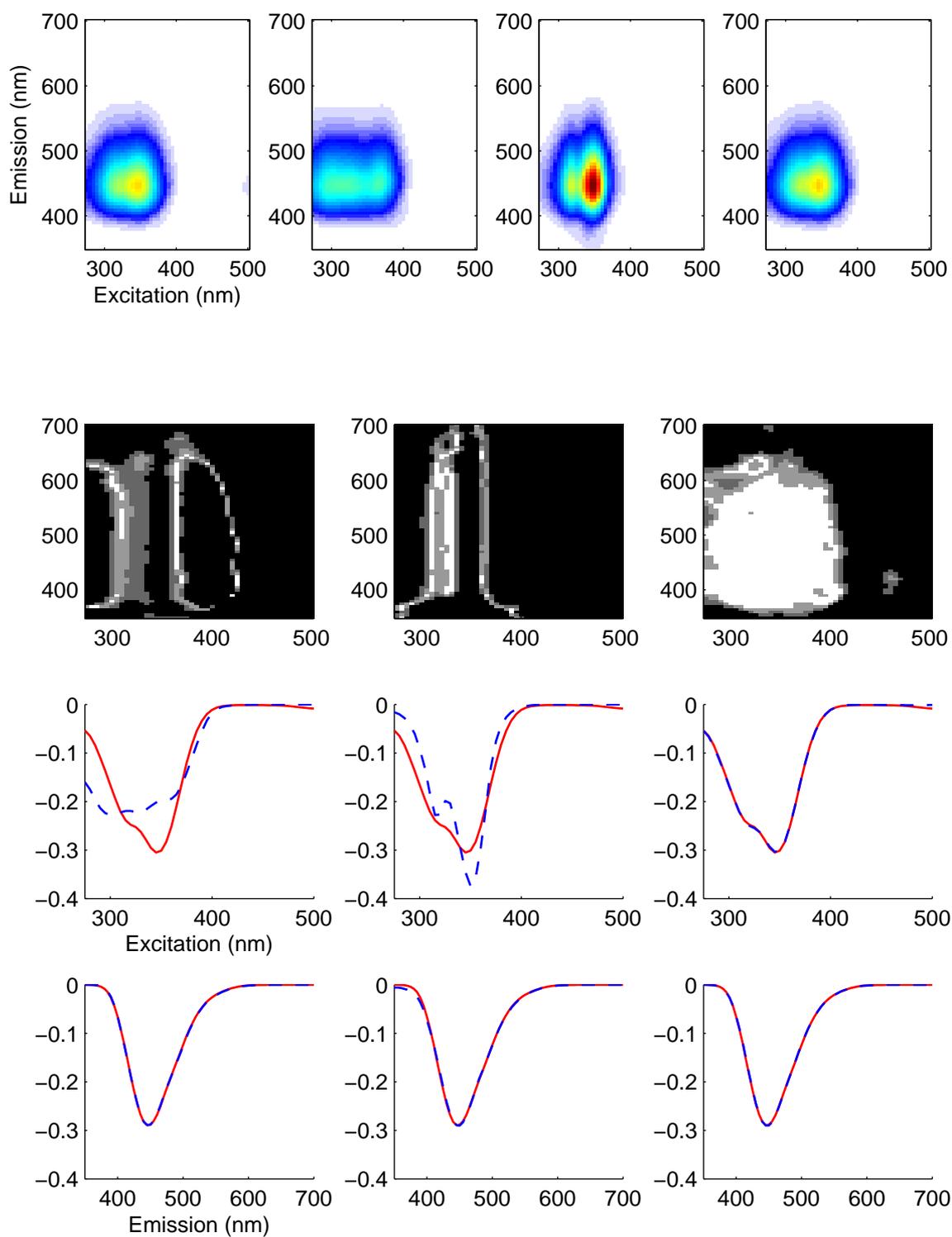


FIG. 4.7 – Jeu 1, solution 7 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées
Voir figure 4.1 pour la légende.

trilinéaire obtenue par PARAFAC sur les quatre bases de MEEF précédentes. Nous présentons les résultats obtenus dans les trois modes (concentrations, émission et excitation) sur les figures 4.8, 4.9 et 4.10. Par souci de lisibilité, nous ne représentons sous forme de figures que les loadings obtenus sur les MEEF originales et celles corrigées par notre méthode comparés aux valeurs de référence. En revanche, les tableaux 4.3, 4.4 et 4.5 récapitulent les résultats chiffrés de la comparaison avec les données de référence sur chacun des trois modes et pour les quatre bases de MEEF. L'erreur sur les concentrations relatives est quantifiée par le rapport r entre l'énergie de l'écart résiduel avec les concentrations réelles et l'énergie du vecteur des concentrations relatives réelles (4.3). Rappelons qu'il s'agit de comparer les concentrations d'un même fluorophore dans les différentes solutions et que la comparaison inverse, entre les valeurs obtenues pour chaque fluorophore dans un mélange donné, ne donne pas les rapports de concentrations entre les deux fluorophores (cf. 2.3.3). Concernant les spectres d'excitation (4.4) et d'émission (4.5), nous rajoutons quatre autres critères de comparaison : l'erreur relative sur la valeur maximale du spectre (en %), l'erreur sur la position de ce maximum (en nm), l'erreur relative sur la surface du spectre (en %) ainsi que l'erreur moyenne.

Avant toute chose, nous constatons que les résultats obtenus avec la base de solu-

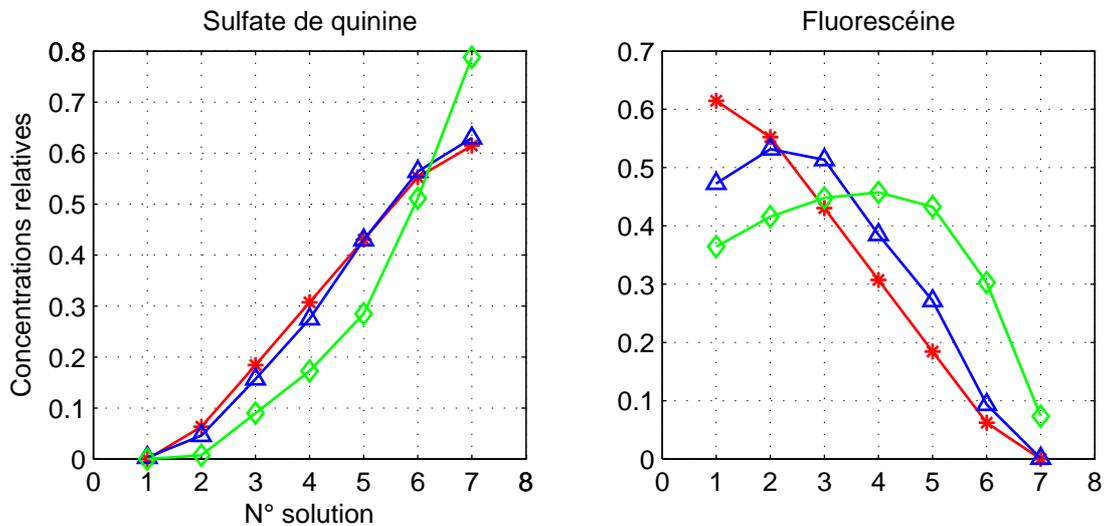


FIG. 4.8 – Jeu 1 : Comparaison des résultats de PARAFAC pour le mode des concentrations relatives

Les trois courbes représentent les résultats obtenus avec les MEEF de référence (en rouge, repérées par des étoiles), les originales (en vert, repérées par des losanges) et celles corrigées par notre méthode (en bleu, repérées par des triangles).

tions diluées sont très proches des données réelles à estimer. Par conséquent, les erreurs commises avec les autres bases ne peuvent être imputées à l'application de PARAFAC.

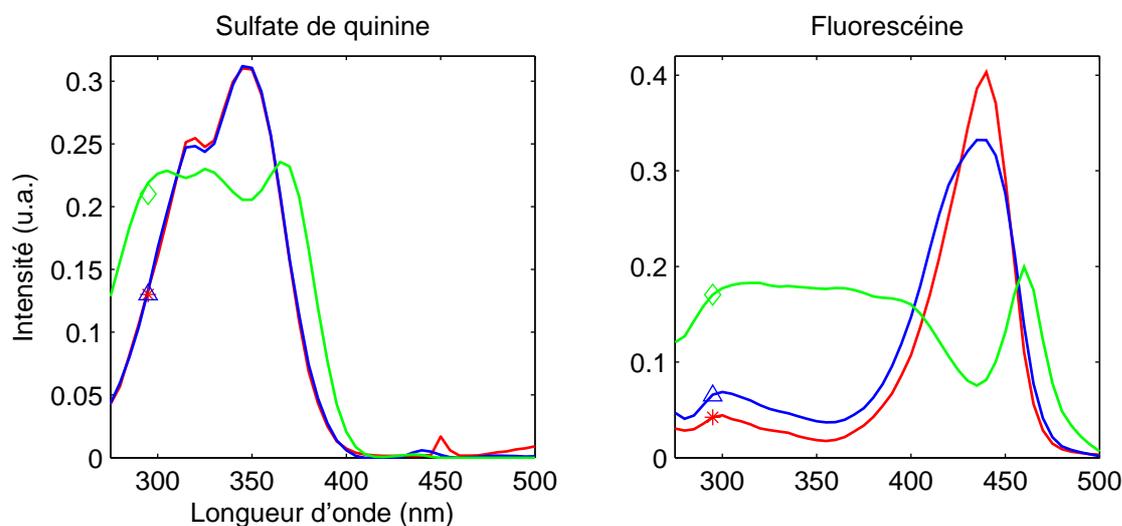


FIG. 4.9 – Jeu 1 : Comparaison des résultats de PARAFAC pour le mode des excitations

Les trois courbes représentent les résultats obtenus avec les MEEF de référence (en rouge, repérées par une étoile), les originales (en vert, repérées par un losange) et celles corrigés par notre méthode (en bleu, repérées par un triangle).

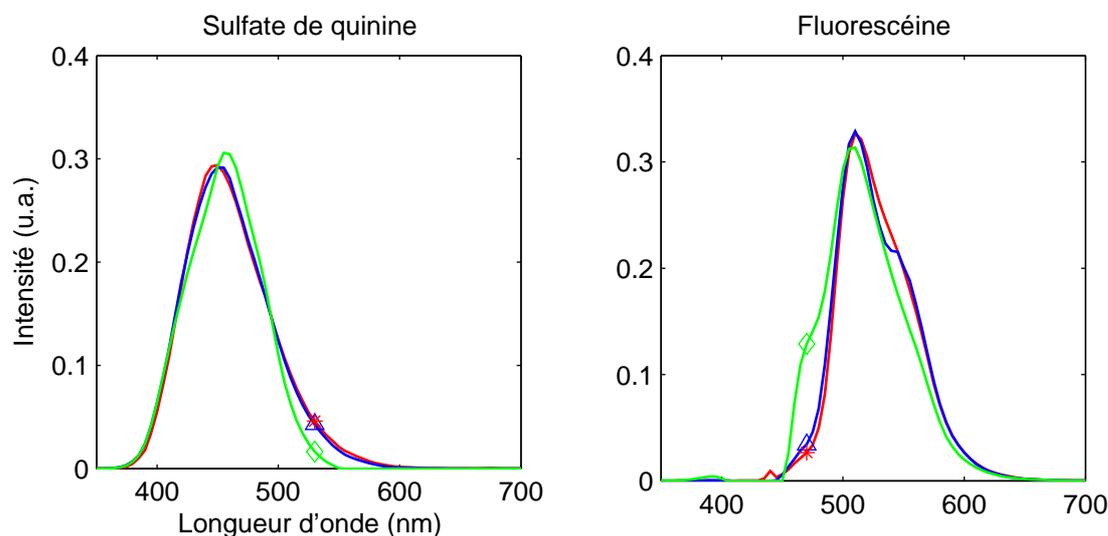


FIG. 4.10 – Jeu 1 : Comparaison des résultats de PARAFAC pour le mode des émissions

Les trois courbes représentent les résultats obtenus avec les MEEF de référence (en rouge, repérées par une étoile), les originales (en vert, repérées par un losange) et celles corrigés par notre méthode (en bleu, repérées par un triangle).

| Fluorophore | r Dil. | r Orig. | r Corr.Abs. | r Corr.Dil. |
|--------------------|----------|-----------|---------------|---------------|
| Sulfate de Quinine | 0.1 | 8.3 | 0.7 | 0.2 |
| Fluorescéine | 0.08 | 23 | 0.4 | 4.2 |

TAB. 4.3 – Jeu 1, Mode des concentrations : Comparaison en énergie pour les concentrations relatives de référence

Rapports r entre l'énergie de l'erreur résiduelle et celle du signal de référence pour les concentrations relatives estimées à partir des données diluées (Dil.), originales (Orig.), corrigées par le spectre d'absorption (Corr.Abs.) et corrigées par notre approche (Corr.Dil.). Les valeurs sont données en pourcentages.

| Fluorophore | Test | Dil. | Orig. | Corr.Abs. | Corr.Dil. |
|--------------------|------------------|-------|-------|-----------|-----------|
| Sulfate de quinine | Valeur max. | 0.4 | 24 | 23 | 0.5 |
| | Pos. valeur max. | 0 | 20 | 0 | 0 |
| | Surface | 0.15 | 9.6 | 11 | 0.8 |
| | Rap. énergie | 0.06 | 12.9 | 4.1 | 0.08 |
| | err. moy. | 0.002 | 0.04 | 0.02 | 0.003 |
| Fluorescéine | Valeur max | 2.5 | 50.5 | 15 | 17 |
| | Pos. valeur max. | 0 | 20 | 5 | 5 |
| | Surface | 2.1 | 48.7 | 15 | 14 |
| | Rap. énergie | 0.07 | 95.15 | 9.3 | 3.4 |
| | err. moy. | 0.003 | 0.12 | 0.04 | 0.02 |

TAB. 4.4 – Jeu 1, Mode des excitations : Comparaison des spectres d'excitation de référence

Résultat des cinq différents tests sur les spectres d'excitation estimés à partir des données diluées (Dil.), originales (Orig.), corrigées par le spectre d'absorption (Corr.Abs.) et corrigées par notre approche (Corr.Dil.) par rapport aux données de référence : écart sur la valeur du maximum (%), sur la position du maximum (nm), sur la surface (%), rapport entre l'énergie de l'erreur résiduelle et de la référence (%) et erreur résiduelle moyenne

| Fluorophore | Test | Dil. | Orig. | Corr.Abs. | Corr.Dil. |
|--------------------|------------------|--------|-------|-----------|-----------|
| Sulfate de quinine | Valeur max. | 1.4 | 4.1 | 4.1 | 0.7 |
| | Pos. valeur max. | 0 | 5 | 5 | 0 |
| | Surface | 0.7 | 5.8 | 2.14 | 0.23 |
| | r | 0.06 | 1.8 | 0.17 | 0.1 |
| | err. moy. | 0.0016 | 0.011 | 0.0036 | 0.0026 |
| Fluorescéine | Valeur max | 3.9 | 4 | 4.7 | 0.6 |
| | Pos. valeur max. | 0 | 0 | 15 | 0 |
| | Surface | 1.7 | 5.3 | 1.4 | 7.8 |
| | r | 0.4 | 7.6 | 15 | 0.5 |
| | err. moy. | 0.004 | 0.018 | 0.02 | 0.004 |

TAB. 4.5 – Jeu 1, Mode des émissions : Comparaison avec les spectres d'émission de référence

Résultat des cinq différents tests sur les spectres d'émission estimés à partir des données diluées (Dil.), originales (Orig.), corrigées par le spectre d'absorption (Corr.Abs.) et corrigées par notre approche (Corr.Dil.) par rapport aux données de référence : écart sur la valeur du maximum (%), sur la position du maximum (nm), sur la surface (%), rapport entre l'énergie de l'erreur résiduelle et de la référence (%) et erreur résiduelle moyenne

Mode des concentrations. Les concentrations relatives du sulfate de quinine sont correctement estimées après correction (quelle que soit la méthode). Par ailleurs, l'effet d'écran ne nuit pas à l'allure générale du profil de concentration. En revanche, le profil de la fluorescéine s'en trouvait nettement perturbé. La correction proposée améliore fortement les résultats mais l'erreur relative avec les rapports de concentration réels reste parfois élevée (30%), en particulier sur la première solution contenant la fluorescéine pure. On notera que contrairement à ce que l'on observe pour tous les autres loadings, la correction en absorbance est plus efficace que notre approche pour l'estimation des concentrations relatives de la fluorescéine.

Mode des excitations. Les deux spectres d'excitation sont fortement perturbés par l'effet d'écran. Nous verrons d'ailleurs sur les autres exemples qu'il s'agit là d'une tendance générale. Après correction par dilution, le spectre du sulfate de quinine est parfaitement déterminé ($r < 0.1\%$ au lieu de 13%) ce qui n'est pas le cas en utilisant l'absorbance ($r = 4\%$). Le résultat est encore plus spectaculaire avec la fluorescéine qui voyait son spectre totalement déformé. Malgré un certain tassement du spectre, la correction par dilution permet de retrouver son allure et sa position. (le rapport r passe de 95% à 3% et le décalage du maximum de 20 à 5 nm). Les résultats sont meilleurs que ceux obtenus avec l'absorbance.

Mode des émissions. Nous passons rapidement sur le spectre d'émission du sulfate de quinine qui n'était quasiment pas déformé par l'effet d'écran ($r < 2\%$) si ce n'est au niveau de son pied et du maximum (décalé de 5 nm). Après correction, quelle que soit la méthode, ces imperfections ont totalement disparu. De même, la légère dilatation du spectre de la fluorescéine vers les faibles longueurs d'onde a également été gommée. En revanche on observe une légère déformation du sommet du pic. Cette déformation

en forme de creux est très comparable à celle remarquée sur la MEEF corrigée de la solution de fluorescéine (4.1).

Conclusion. D'un point de vue général notre approche combinée à PARAFAC permet de retrouver des spectres très proches des valeurs réelles et nettement meilleurs (quel que soit le critère) que ceux estimés à partir des MEEF non corrigées originales ou des MEEF corrigées par le spectre d'absorption.

L'erreur de linéarisation la plus importante apparaît sur la solution de fluorescéine pure. Après la décomposition par PARAFAC, cela se traduit principalement sur le mode des concentrations pour lequel on constate une erreur importante sur l'estimation de la concentration relative à cette solution. Dans une moindre mesure, les spectres de la fluorescéine sont plus affectés que ceux du sulfate de quinine. Cette solution perturbe fortement les résultats de PARAFAC. Nous n'en montrons pas ici les résultats, mais on améliore nettement les résultats en retirant cette solution du jeu de données.

En conclusion, les spectres estimés à partir de notre approche sont suffisamment proches des spectres de référence pour que les constituants soient identifiés sans aucune ambiguïté alors que ce n'est pas le cas des mélanges non corrigés. Quant aux spectres obtenus après correction en absorbance, leur interprétation peut encore s'avérer problématique. Sur douze rapports de concentration estimés après notre correction, un seul est vraiment erroné. Si l'on excepte cette mauvaise estimation du premier rapport de concentration de la fluorescéine, les deux profils de concentration sont correctement retranscrits. Ces premiers résultats sont donc probants.

4.2.3.2 Mélanges à 3 Composants

Nous exposons ici les résultats obtenus sur le second jeu de mélanges : phénylalanine, tyrosine et tryptophane. Nous n'avons effectué pour ce mélange la correction des MEEF par les spectres d'absorption. Enfin, le protocole de test est identique à celui de la section précédente. Les figures 4.11 à 4.14 montrent les résultats de la linéarisation des quatre solutions de la série. Le tableau 4.6 donne pour chaque solution et pour chaque type de MEEF le rapport entre l'énergie de l'erreur et l'énergie du signal de référence.

Ces données constituent *a priori* un cas plus difficile puisque les spectres se chevauchent énormément. Il faut noter cependant que les modifications des MEEF dues à l'effet d'écran sont moins spectaculaires que dans le cas précédent. On observe surtout une importante modification des rapports d'intensité entre les différents pics. En particulier, un déplacement du pic localisé autour du point (205 nm, 390 nm) vers le pic de la tyrosine (225 nm, 300nm). Ce phénomène est bien corrigé par notre algorithme, avec une réserve pour le second mélange pour lequel les trois critères sont nettement en deçà des autres résultats. Globalement cependant, les valeurs des rapports d'énergie sont dans la moyenne de ceux du jeu précédent, aussi bien pour les images originales que pour les images corrigées. Malgré une erreur résiduelle localement importante, la similitude entre les trois premières composantes principales des MEEF corrigées et diluées nous montre que celles-ci ont la même structure. On note

| Solution | r Orig. | r Corr.Dil. |
|----------|-----------|---------------|
| 1 | 65 | 2.3 |
| 2 | 10.4 | 17 |
| 3 | 51 | 3.7 |
| 4 | 52 | 3.4 |

TAB. 4.6 – Jeu 2 : Comparaison en énergie avec les MEEF de référence
Rapports r entre l'énergie de l'erreur résiduelle et celle du signal de référence pour les MEEF originales (Orig.) et corrigées par notre approche (Corr.Dil.). Les valeurs sont données en pourcentages

| Fluorophore | r Orig. | r Corr.Dil. |
|---------------|-----------|---------------|
| phénylalanine | 0.08 | 0.49 |
| Tyrosine | 0.2 | 0.36 |
| tryptophane | 0.26 | 0.18 |

TAB. 4.7 – Jeu 2, Mode des concentrations : Comparaison en énergie avec les concentrations relatives de référence

Rapports entre l'énergie de l'erreur résiduelle et celle du signal de référence pour les concentrations relatives estimées à partir des données originales (Orig.), et corrigées par notre approche (Corr.Dil.). Les valeurs sont données en pourcentages

également avec ce test que les améliorations les plus notables ont lieu sur l'excitation.

Nous comparons à présent les résultats de PARAFAC sur les deux bases de MEEF avec les valeurs de référence. La comparaison est effectuée de la même manière qu'avec le jeu de données précédent. En revanche, nous ne disposons pas des spectres d'absorption pour effectuer ce type de correction et nous ne faisons plus figurer les résultats obtenus avec la base diluée qui n'a maintenant que peu d'intérêt.

Les premiers résultats obtenus étaient assez décevants, même avec la base diluée. Ceci s'explique par la conjugaison de trois facteurs : le faible nombre d'échantillons (4) devant le nombre de fluorophores à estimer (3), le petit nombre de longueurs d'onde d'excitation et l'importante corrélation entre les spectres, tout cela diminuant d'autant le nombre d'équations indépendantes du système. Ne disposant d'autre mélange concentré indépendant de l'un des quatre autres, nous avons donc rajouté un cinquième échantillon linéaire (tyrosine pure très diluée) dans chaque base. Cet ajout permet à l'algorithme de converger avec la base linéaire mais n'améliore pas fondamentalement les résultats de la base non linéaire. Par conséquent il ne biaise pas l'interprétation des résultats. Les loadings sont illustrés sur les figures 4.15, 4.16 et 4.17 et leur comparaison par les valeurs des tableaux 4.7, 4.8 et 4.9.

Mode des concentrations La tendance générale des profils de concentrations est conservée. En revanche, l'écart résiduel est important sur plusieurs points, en particulier pour le premier mélange. Par ailleurs, nous n'améliorons pas les résultats obtenus avec les spectres non corrigés. Les résultats sur ce mode ne sont donc pas satisfaisants. Cependant, il est clair que l'erreur commise sur le premier mélange influence le reste des résultats. L'amplitude du signal de ce mélange est surévaluée par la procédure

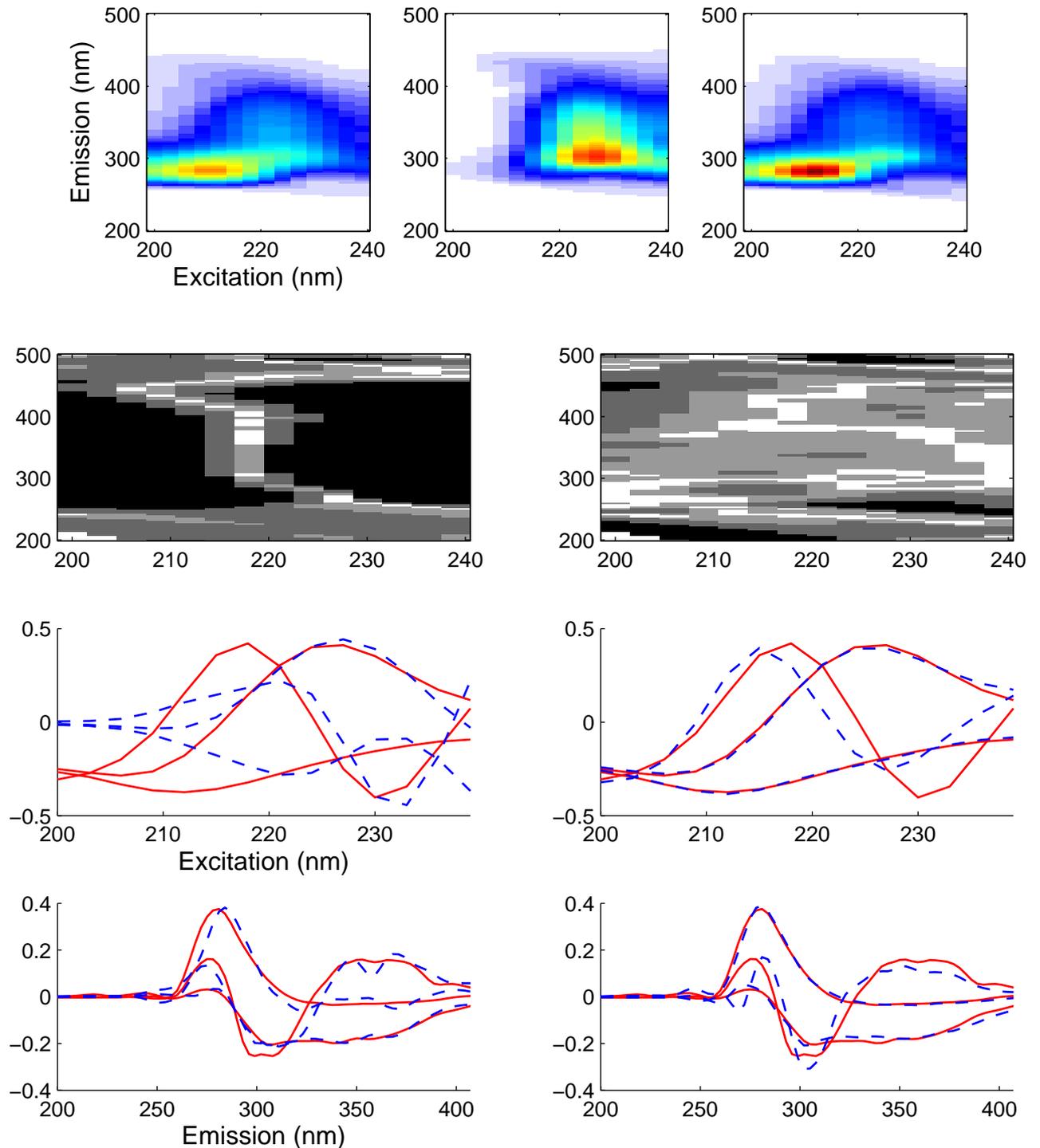


FIG. 4.11 – Série 2, solution 1 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées

Les figures de la ligne du haut représentent de gauche à droite la MEEF de référence, la MEEF originale et la MEEF corrigée par la dilution. Les figures de la seconde ligne représentent une distribution grossière de l'erreur relative entre la MEEF de référence et (de gauche à droite) la MEEF originale et la MEEF corrigée par la dilution. Les quatre niveaux de gris, du blanc au noir, correspondent respectivement à une erreur inférieure à 5%, comprise entre 5% et 15%, comprise entre 15% et 30% et supérieure à 30%. Les deux lignes du bas montrent dans le même ordre les composantes principales significatives, en excitation et en émission. Celles de la MEEF de référence sont représentées en traits pleins.

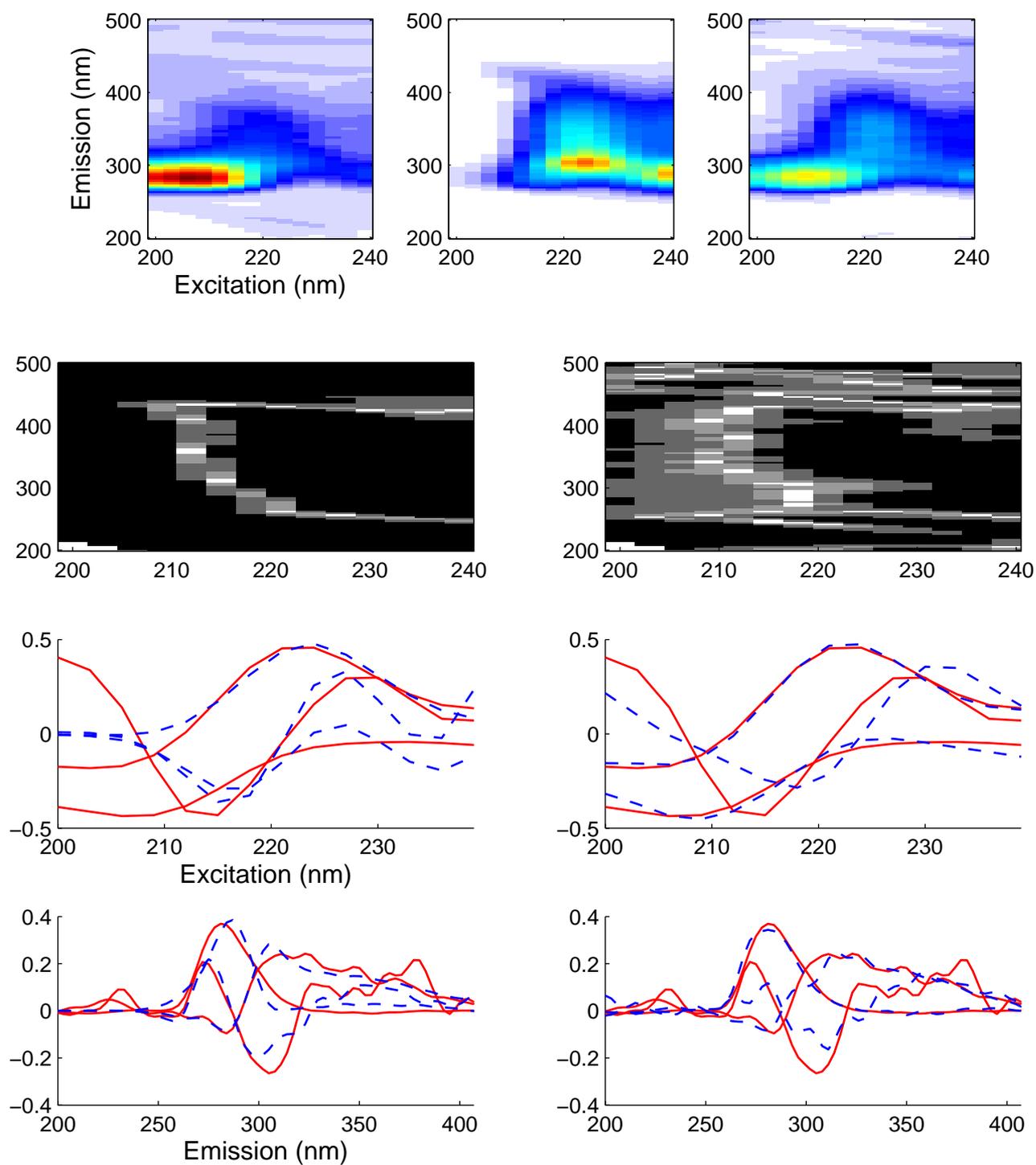


FIG. 4.12 – Jeu 2, solution 2 : comparaison entre la MEEFde référence, la MEEF originale et la MEEF corrigée
Voir figure 4.11 pour la légende.

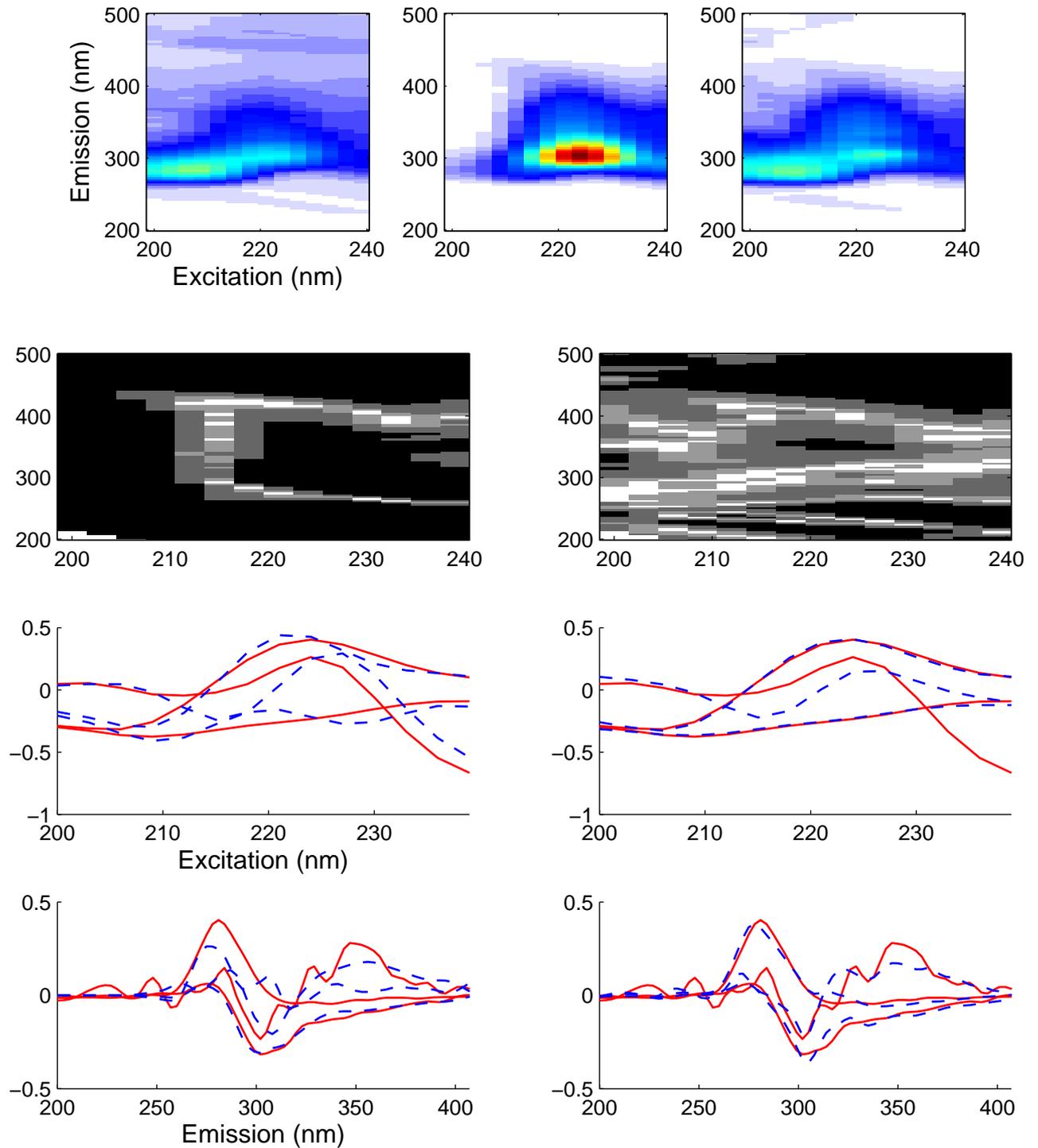


FIG. 4.13 – Jeu 2, solution 3 : comparaison entre la MEEFde référence, la MEEF originale et la MEEF corrigée
 Voir figure 4.11 pour la légende.

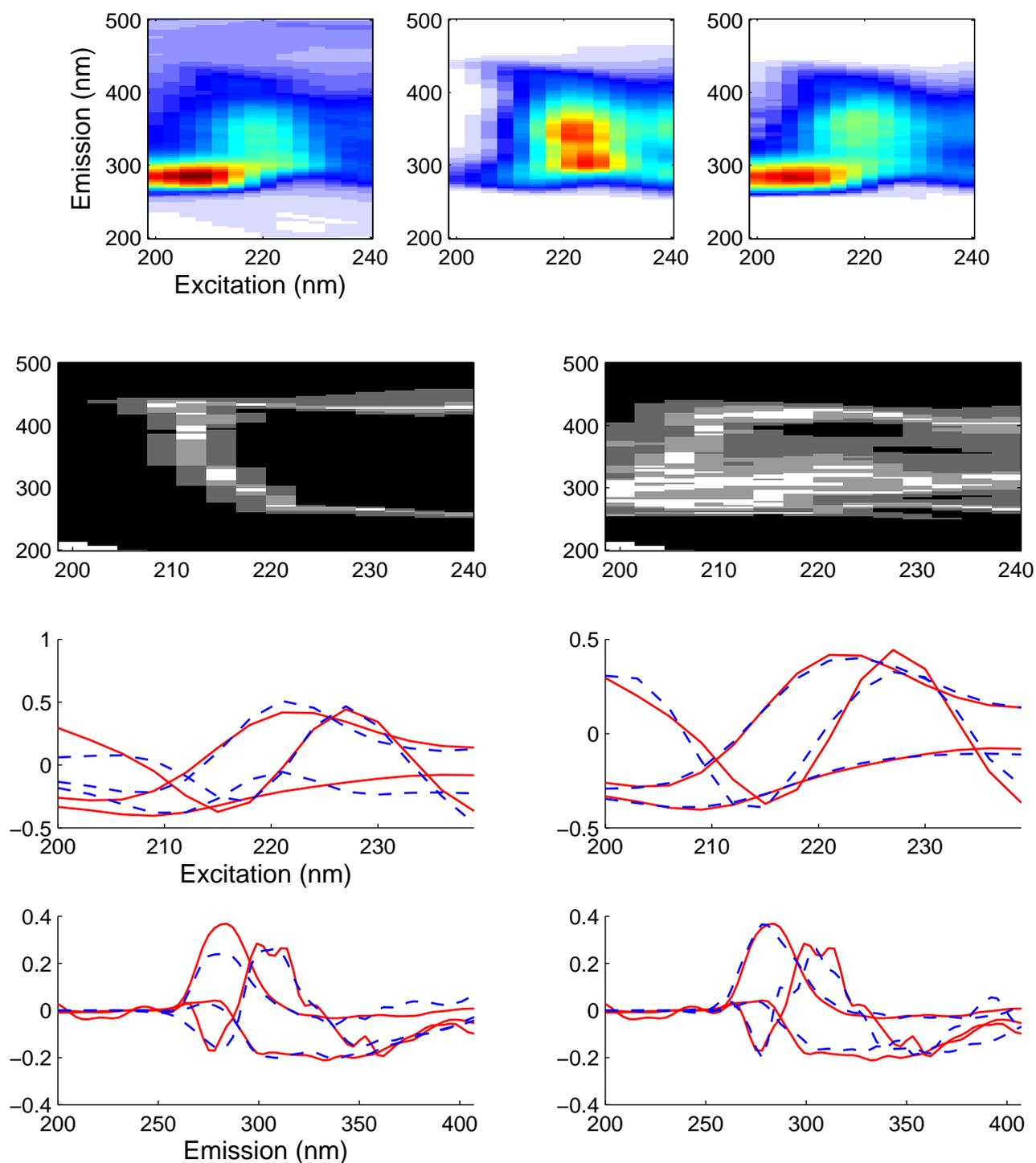


FIG. 4.14 – Jeu 2, solution 4 : comparaison entre la MEEFde référence, la MEEF originale et la MEEF corrigée
Voir figure 4.11 pour la légende.

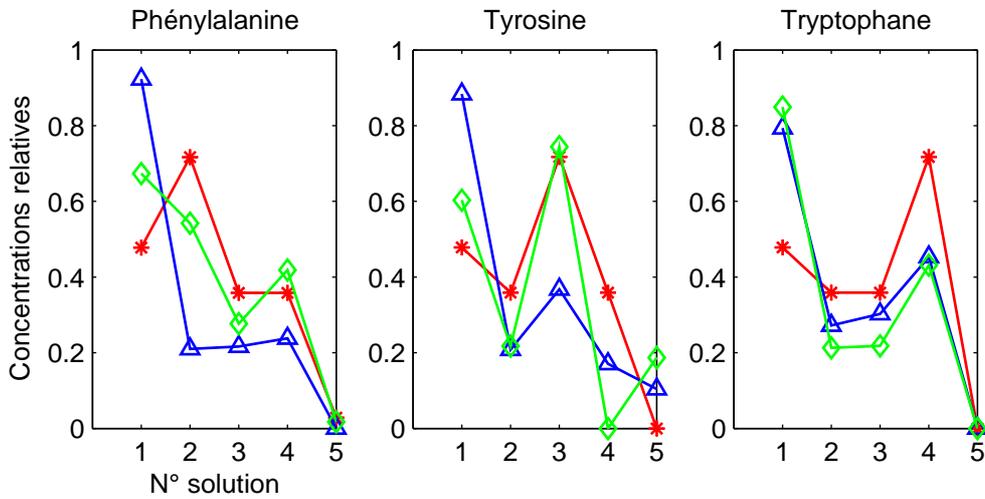


FIG. 4.15 – Jeu 2 : Comparaison des résultats de PARAFAC pour le mode des concentrations relatives

Les trois courbes représentent les résultats obtenus avec les MEEF de référence (en rouge, repérées par des étoiles), les originales (en vert, repérées par des losanges) et celles corrigées par notre méthode (en bleu, repérées par des triangles).

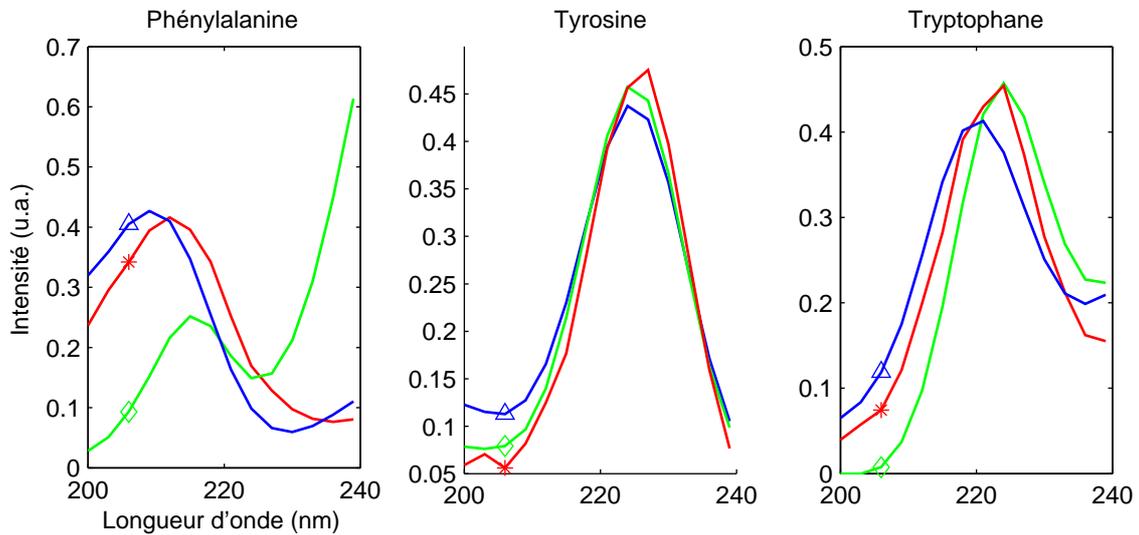


FIG. 4.16 – Jeu 2 : Comparaison des résultats de PARAFAC pour le mode des excitations

Les trois courbes représentent les résultats obtenus avec les MEEF de référence (en rouge, repérées par une étoile), les originales (en vert, repérées par un losange) et celles corrigées par notre méthode (en bleu, repérées par un triangle).

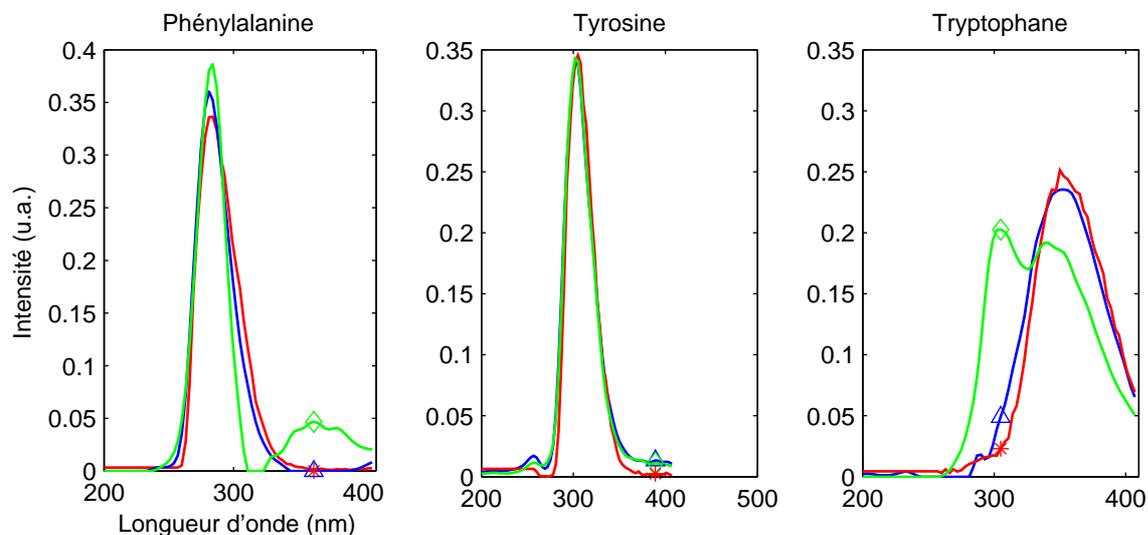


FIG. 4.17 – Jeu 2 : Comparaison des résultats de PARAFAC pour le mode des émissions

Les trois courbes représentent les résultats obtenus avec les MEEF de référence (en rouge, repérées par une étoile), les originales (en vert, repérées par un losange) et celles corrigées par notre méthode (en bleu, repérées par un triangle).

| Fluorophore | Test | Orig. | Corr.Dil. |
|---------------|------------------|-------|-----------|
| phénylalanine | Valeur max. | 47 | 2.5 |
| | Pos. valeur max. | 27 | 3 |
| | Surface | 6 | 4 |
| | r | 79 | 5 |
| | err. moy. | 0.2 | 0.05 |
| Tyrosine | Valeur max. | 3.8 | 8 |
| | Pos. valeur max. | 3 | 3 |
| | Surface | 3.4 | 8.2 |
| | r | 0.7 | 2.2 |
| | err. moy. | 0.02 | 0.04 |
| tryptophane | Valeur max. | 0.7 | 9 |
| | Pos. valeur max. | 0 | 3 |
| | Surface | 7 | 6 |
| | Rap. énergie | 6 | 3 |
| | err. moy. | 0.06 | 0.04 |

TAB. 4.8 – Jeu 2, Mode des excitations : Comparaison avec les spectres d'excitation de référence

Résultat des cinq différents tests sur les spectres d'excitation estimés à partir des données originales (Orig.) et corrigées par notre approche (Corr.Dil.) par rapport aux données de référence : écart sur la valeur du maximum (%), sur la position du maximum (nm), sur la surface (%), rapport entre l'énergie de l'erreur résiduelle et de la référence (%) et erreur résiduelle moyenne

| Fluorophore | Test | Orig. | Corr.Dil. |
|---------------|------------------|-------|-----------|
| phénylalanine | Valeur max. | 14.8 | 7 |
| | Pos. valeur max. | 0 | 3 |
| | Surface | 3.4 | 3.9 |
| | r | 13 | 2 |
| | err. moy. | 0.03 | 0.01 |
| Tyrosine | Valeur max. | 0.7 | 1.8 |
| | Pos. valeur max. | 3 | 3 |
| | Surface | 7.6 | 11 |
| | r | 1.5 | 1.8 |
| | err. moy. | 0.01 | 0.01 |
| tryptophane | Valeur max. | 19.3 | 6 |
| | Pos. valeur max. | 45 | 3 |
| | Surface | 11.5 | 1.7 |
| | r | 34 | 1.5 |
| | err. moy. | 0.048 | 0.01 |

TAB. 4.9 – Jeu 2, Mode des émissions : Comparaison des spectres d'émission de référence

Résultat des cinq différents tests sur les spectres d'excitation estimés à partir des données originales (Orig.) et corrigées par notre approche (Corr.Dil.) par rapport aux données de référence : écart sur la valeur du maximum (%), sur la position du maximum (nm), sur la surface (%), rapport entre l'énergie de l'erreur résiduelle et de la référence (%) et erreur résiduelle moyenne

de correction alors que le spectre est correctement estimé (cf. figure 4.11). Une telle différence ne peut pas être expliquée par une erreur sur le facteur de dilution. Il est possible que les concentrations indiquées pour cette solution ne soient pas fiables. Ces résultats devront donc être confirmés ou améliorés avec un jeu de données plus important.

Mode des excitations. Seul le spectre d'excitation de la phénylalanine est vraiment modifié par l'effet d'écran. En revanche la déformation est très importante puisque l'on note l'apparition d'un double pic dont le maximum est décalé de 27 nm par rapport à l'original (figure 4.16). Ceci explique en partie la disparition du pic constaté sur les MEEF originales. Tous les tests mettent d'ailleurs en évidence cet écart important. L'erreur après correction reste relativement faible ($r = 4.5\%$), malgré un décalage de 3 nm de l'ensemble du spectre, la forme restant quant à elle identique. L'amélioration apportée par la correction sur les deux autres spectres est minime puisque ceux-ci ne sont que peu déformés.

Mode des émissions. Le spectre de la tyrosine n'est pas modifié par l'effet d'écran. De fait, la correction n'apporte aucune amélioration. Il en irait presque de même du spectre de la phénylalanine si l'on ne constatait l'apparition d'un second pic résiduel vers 360 nm que notre procédure permet d'éliminer ; ce qui explique les meilleurs résultats obtenus à nos tests après correction. Le tryptophane voit en revanche sont

spectre être assez largement modifié (apparition d'un double pic, décalage de 45 nm) et très bien corrigé (r passant de 34% à 1.5%).

Conclusion. Ces résultats montrent que notre procédure peut s'appliquer avec succès à des mélanges de trois fluorophores puisqu'après correction, les spectres sont clairement identifiables. Qui plus est, dans un cas assez difficile où les chevauchements de spectres sont importants. En revanche, les résultats obtenus sur le mode des concentrations présentent des erreurs importantes. Une étude sur un plus grand nombre de spectres doit être menée afin de vérifier si ces erreurs sont imputables ou non à notre procédure.

4.2.3.3 Matière organique

Nous exposons ici les résultats obtenus sur la troisième série de mélanges constituée des MEEF mesurées sur de la matière organique extraite des sols. La mise à l'épreuve de notre démarche sur ce type de données constitue une étape décisive. En effet, comme nous le verrons à l'occasion du dernier chapitre de ce mémoire, l'étude des signaux de fluorescence de matière organique est largement employée en chimie de l'environnement, constituant par là même, un champ d'application privilégié des méthodes de décomposition linéaires et en particulier de PARAFAC.

Nous n'avons pas effectué pour ce mélange la correction des MEEF par les spectres d'absorption. La première partie du protocole de tests ne change pas. Les figures 4.18 à 4.29 montrent donc les résultats de la linéarisation des onze solutions de la série. Le tableau 4.10 donne pour chaque solution et pour chaque type de MEEF le rapport entre l'énergie de l'erreur et l'énergie du signal de référence.

Sur ces MEEF, l'effet d'écran se traduit principalement par une dominance du pic situé autour du point (480 nm, 520 nm) au détriment d'un large massif situé vers 240 nm en excitation et occupant la quasi totalité de la bande d'émission. Celui-ci semble être formé de différents pics, mais il est difficile d'effectuer une analyse plus fine à partir de ces simples images. Sur la majorité de nos échantillons, la procédure de correction permet de reconstituer assez fidèlement l'allure générale de la MEEF linéaire. Cette observation est confirmée par l'étude des trois premières composantes principales. Il faut préciser à ce propos que dans le cas de mélanges inconnus, le nombre de composantes principales significatives n'est pas toujours facile à établir. Ici, il oscille clairement entre deux et trois. Par ailleurs, l'analyse par PARAFAC montre que trois fluorophores peuvent être identifiés sur l'ensemble des échantillons et qu'ils sont présents dans chacun d'entre eux. Nous avons donc choisi de garder trois composantes. Encore une fois l'analyse de l'erreur résiduelle montre une nette amélioration après la correction mais reste localement importante. Enfin, les valeurs de r , autour de 40% pour les images originales et de 3% pour les images corrigées sont là aussi dans la moyenne des valeurs obtenues avec les précédents jeux de données, laissant présager de bons résultats pour la suite.

Pour la seconde partie du test, nous nous contenterons ici de comparer les résultats de PARAFAC sur les trois types de mélanges (originaux, dilués et corrigés). En effet,

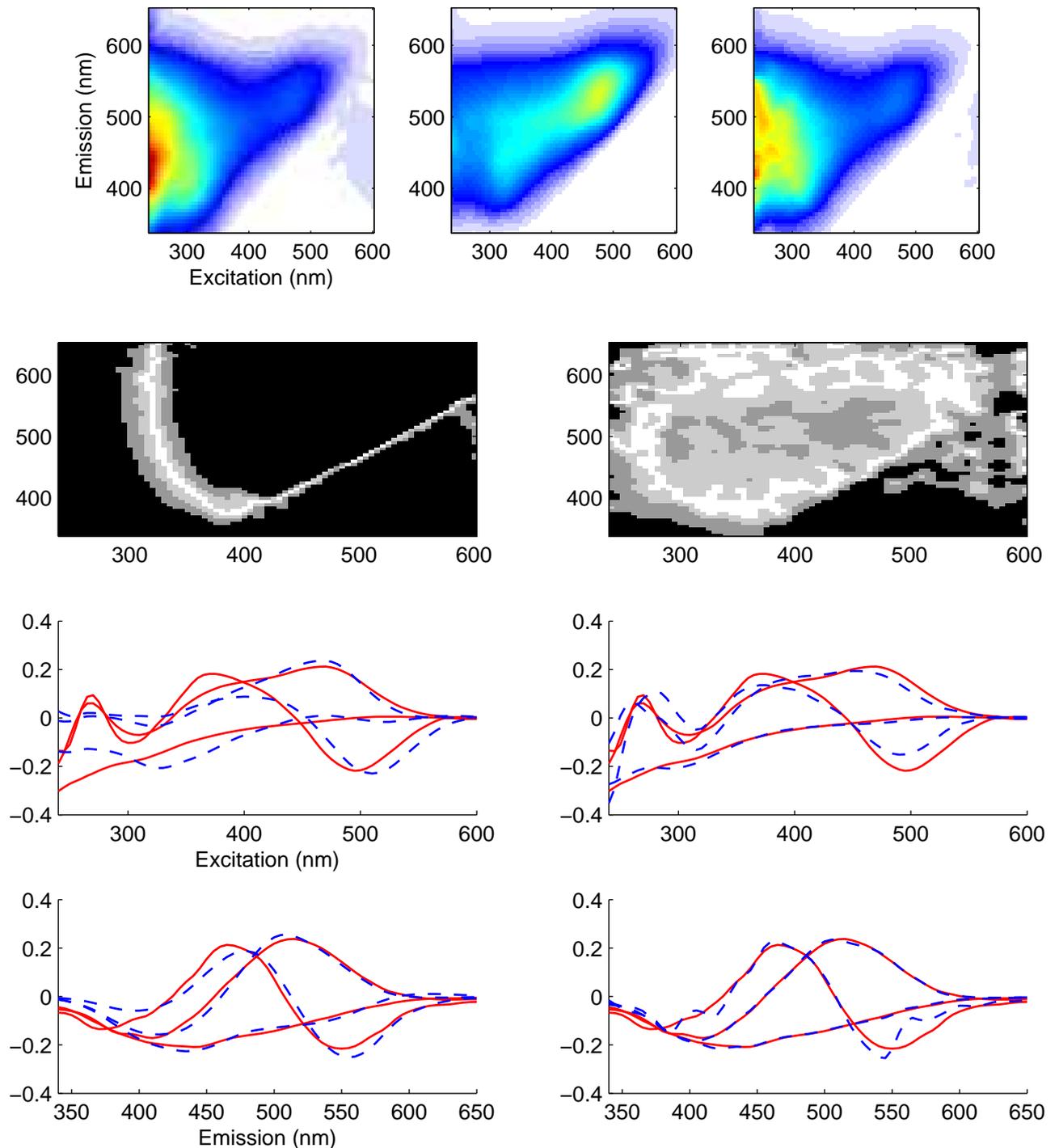


FIG. 4.18 – Jeu 3, échantillon 1 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées

Les figures de la ligne du haut représentent de gauche à droite la MEEF de référence, la MEEF originale et la MEEF corrigée par la dilution. Les figures de la deuxième ligne représentent une distribution grossière de l'erreur relative entre la MEEF de référence et (de gauche à droite) la MEEF originale et la MEEF corrigée par la dilution. Les quatre niveaux de gris, du blanc au noir, correspondent respectivement à une erreur inférieure à 5%, comprise entre 5% et 15%, comprise entre 15% et 30% et supérieure à 30%. Les deux lignes du bas montrent dans le même ordre les composantes principales significatives, en excitation et en émission. Celles de la MEEF de référence sont représentées en traits pleins

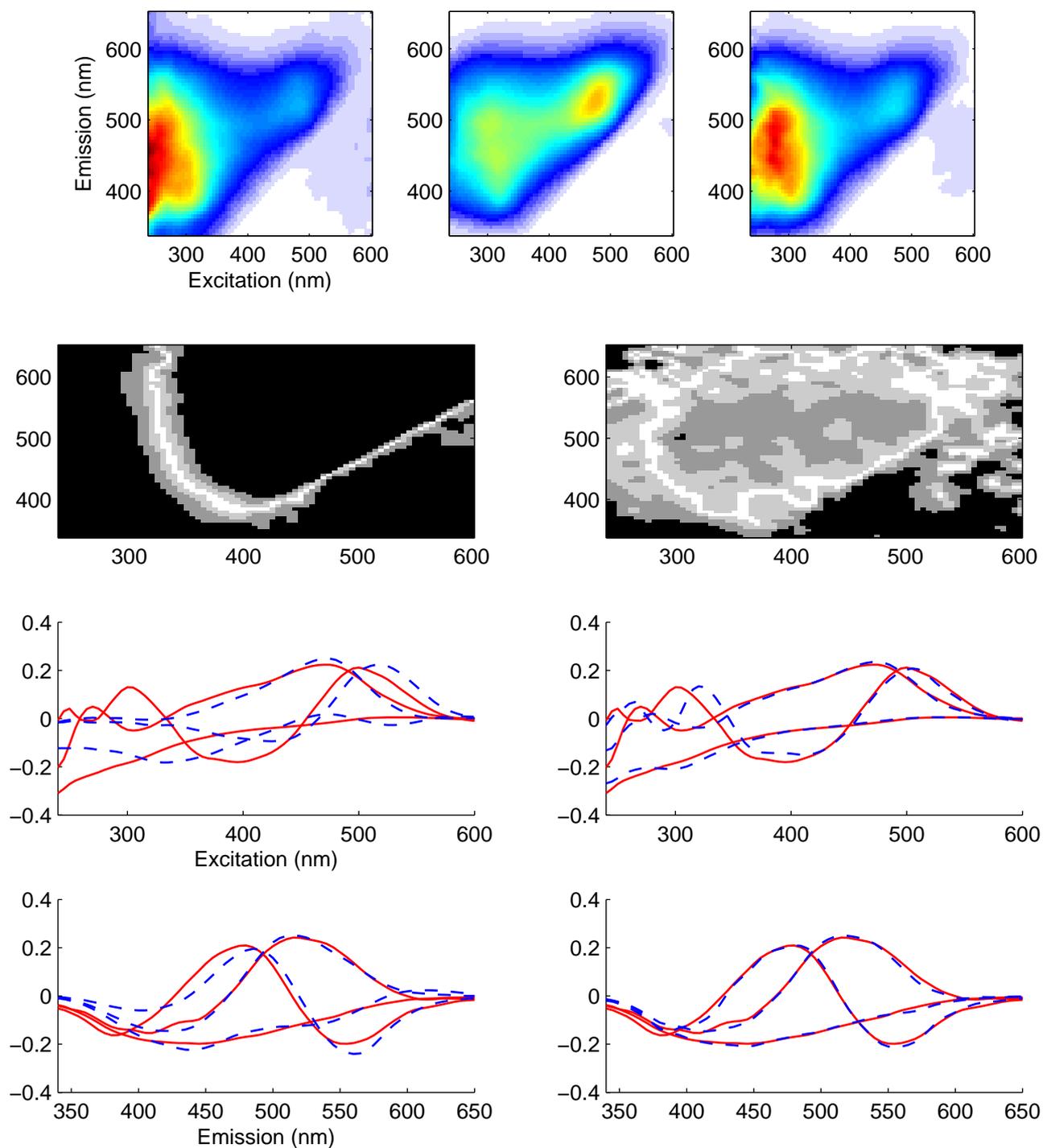


FIG. 4.19 – Jeu 3, échantillon 2 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée
Voir figure 4.18 pour la légende.

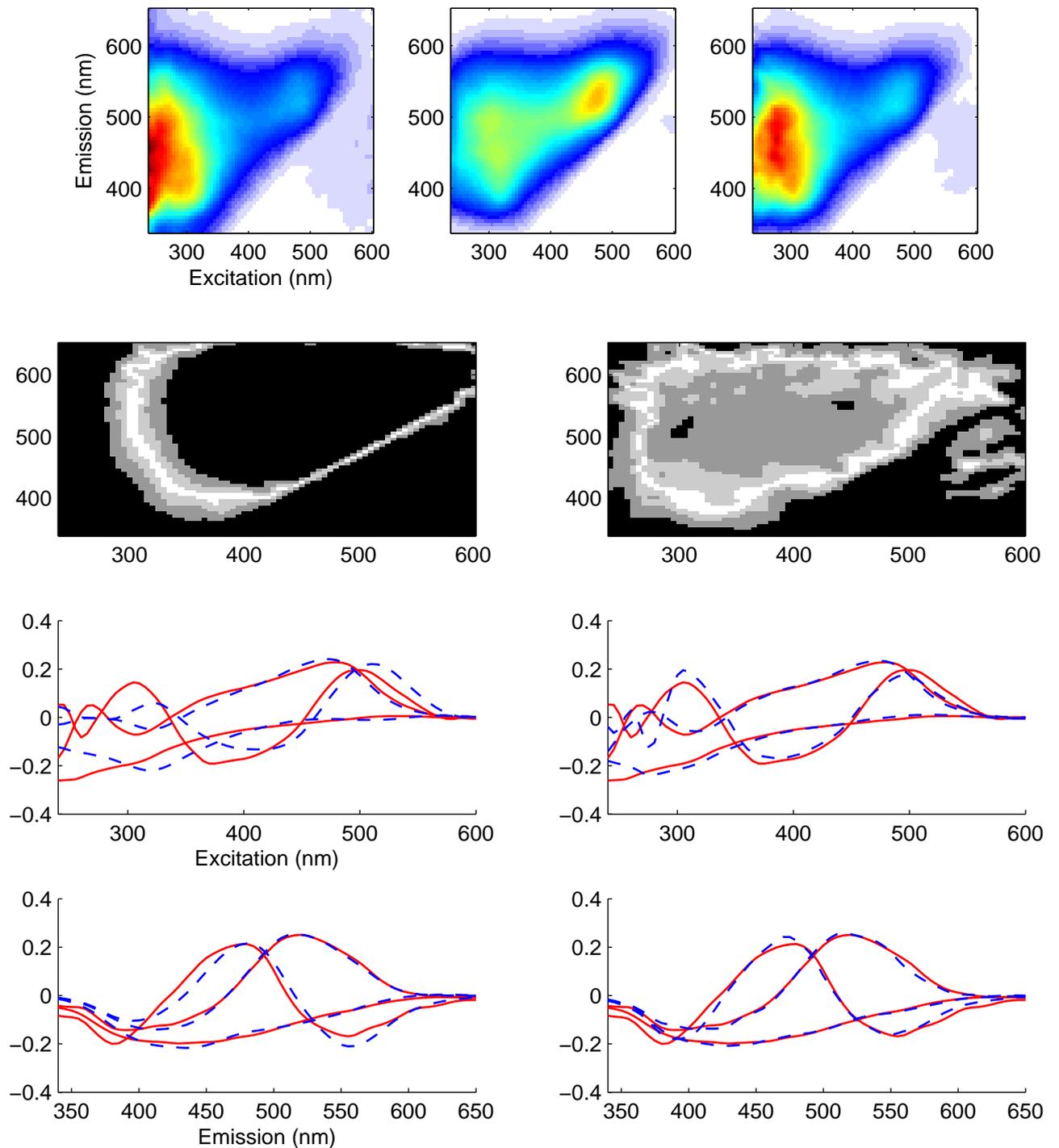


FIG. 4.20 – Jeu 3, échantillon 3 : comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée
Voir figure 4.18 pour la légende.

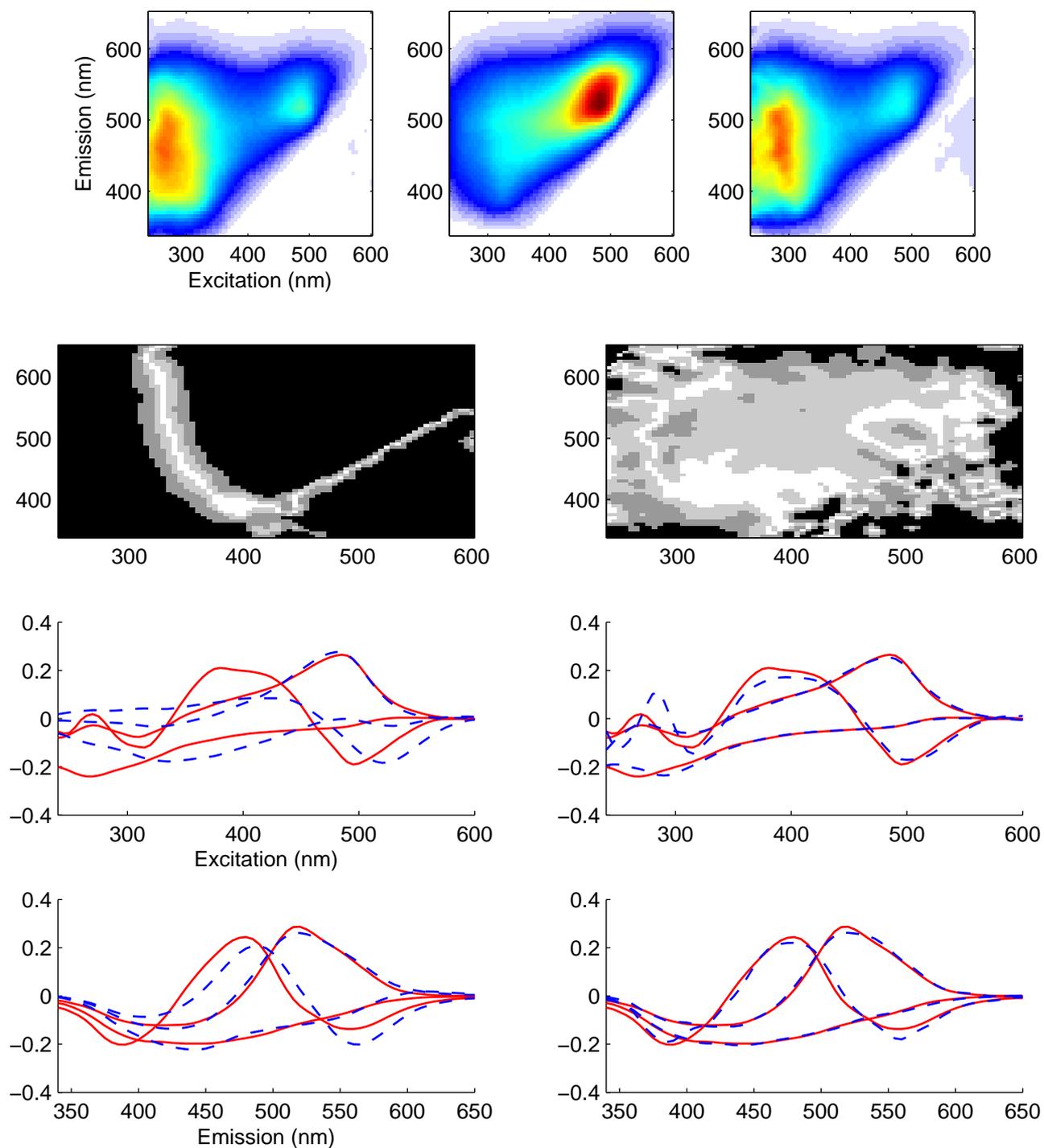


FIG. 4.21 – Jeu 3, échantillon 4 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée
Voir figure 4.18 pour la légende.

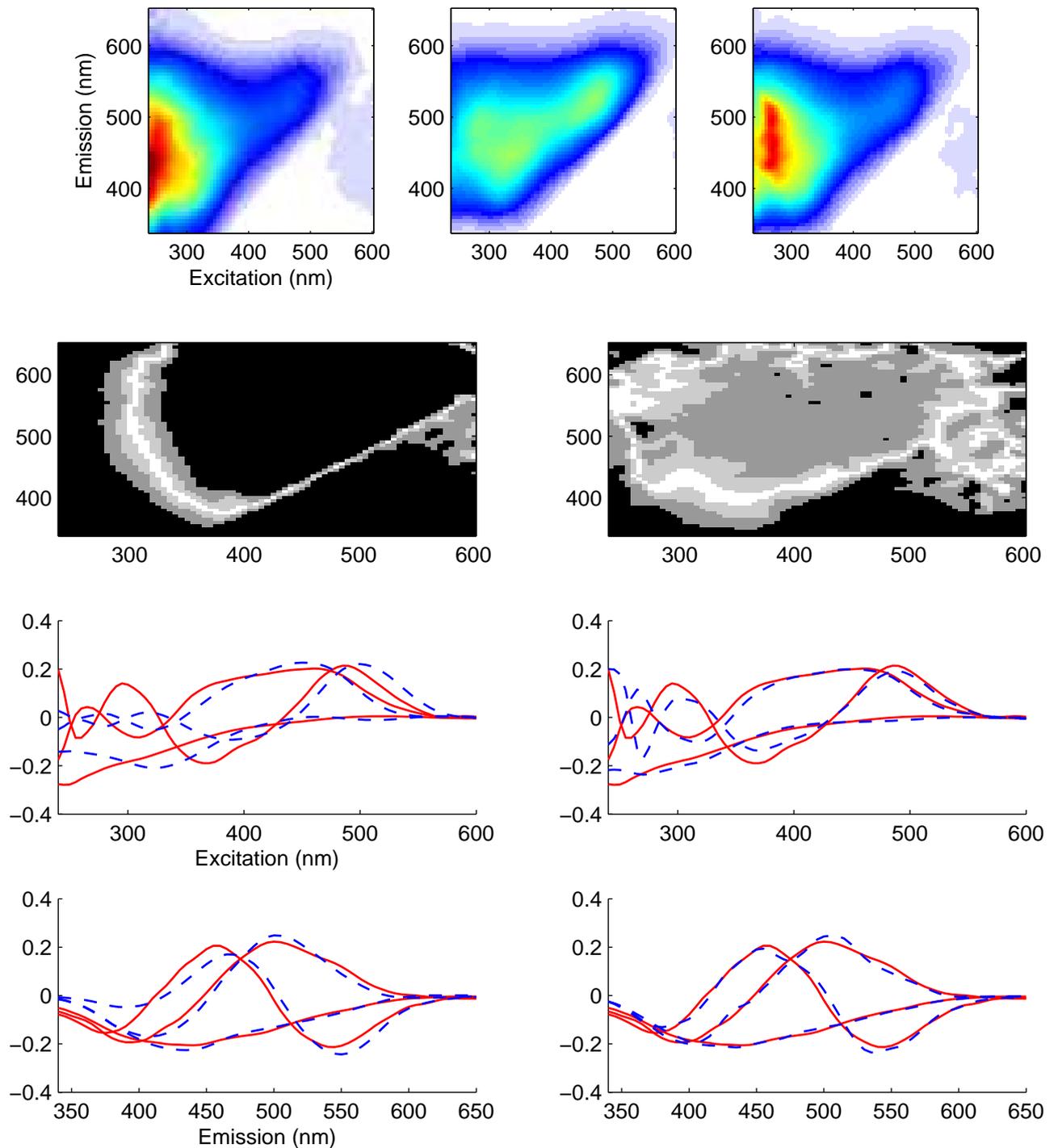


FIG. 4.22 – Jeu 3, échantillon 5 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée
Voir figure 4.18 pour la légende.

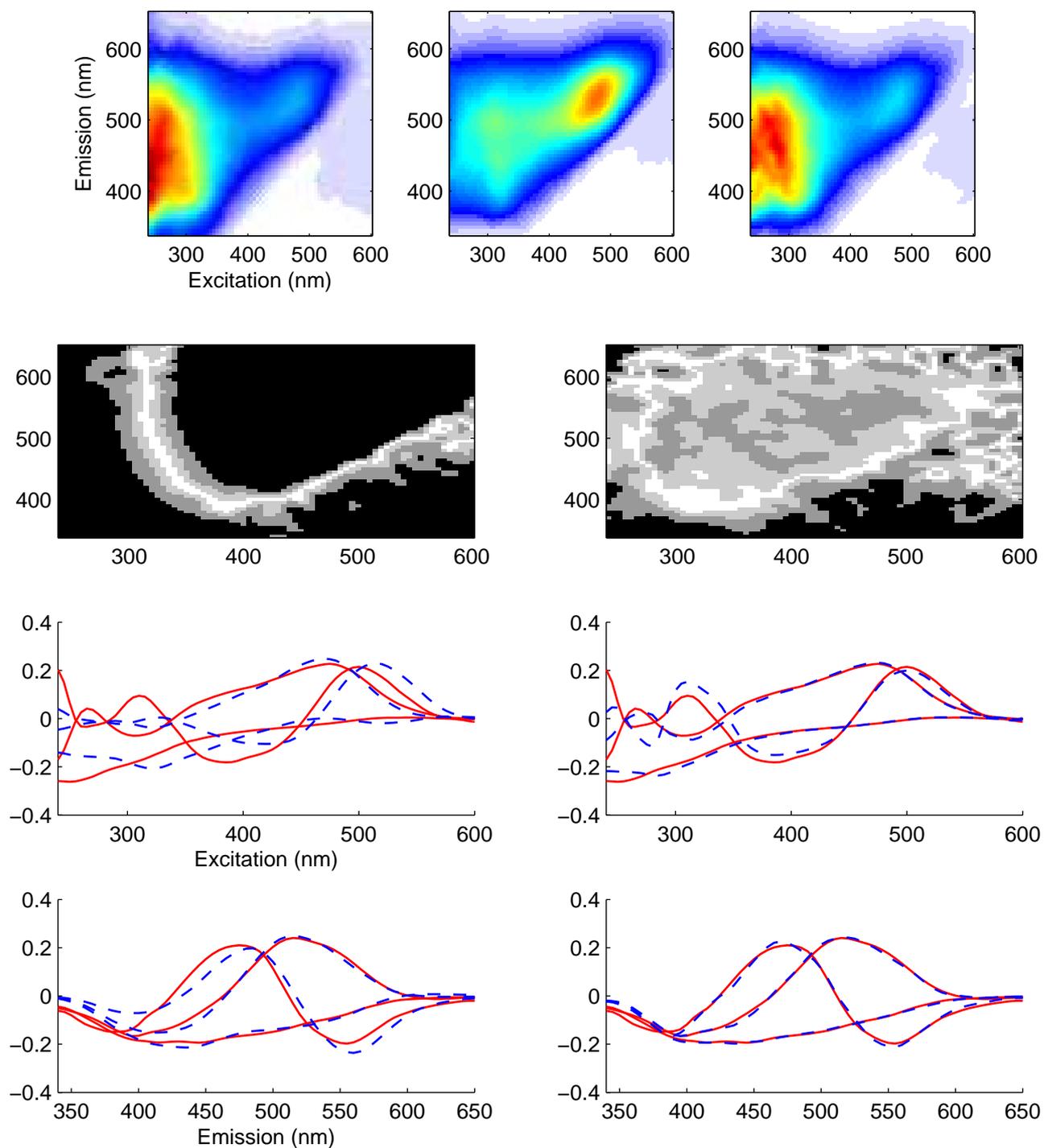


FIG. 4.23 – Jeu 3, échantillon 6 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée
Voir figure 4.18 pour la légende.

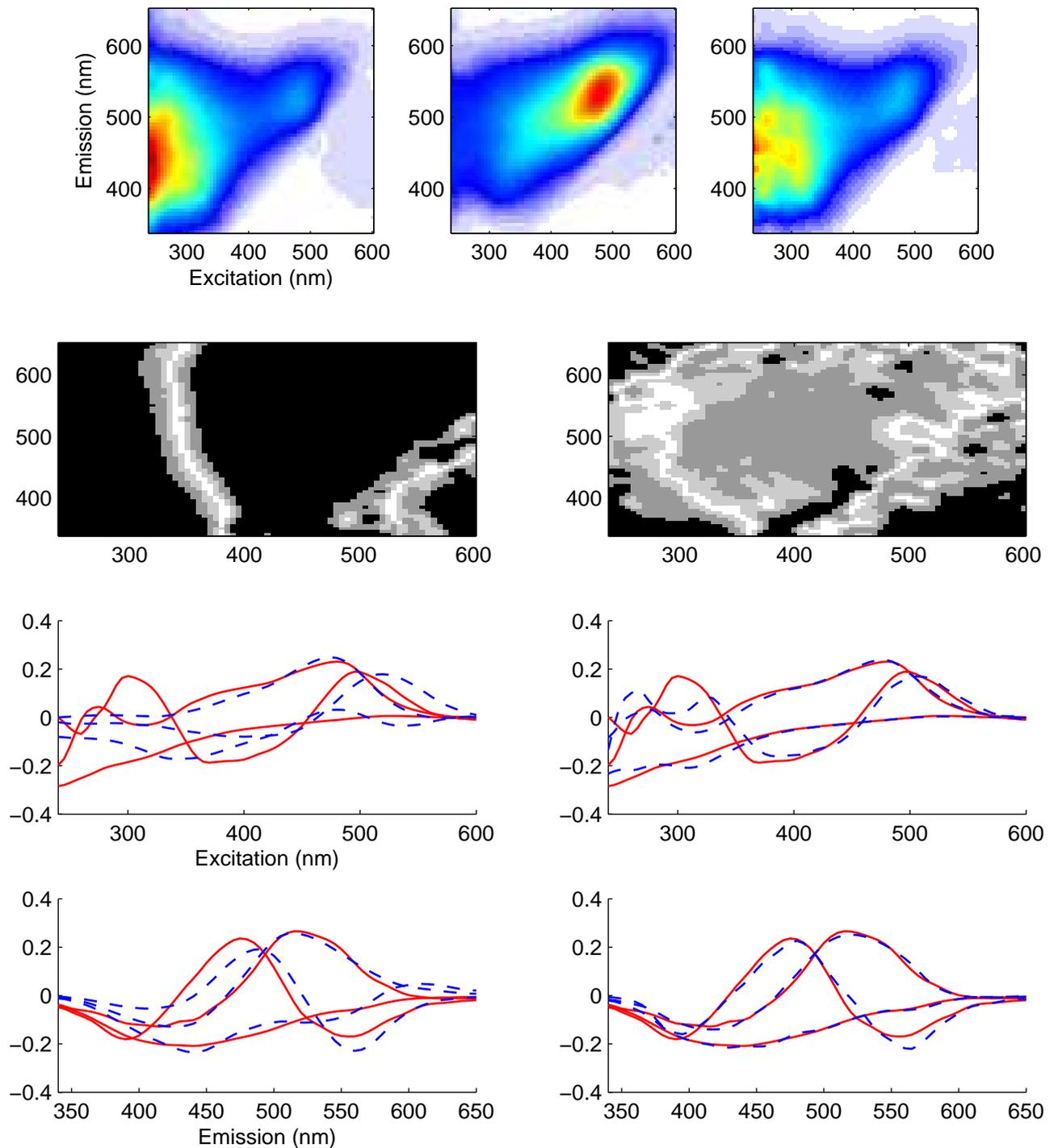


FIG. 4.24 – Jeu 3, échantillon 7 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée
Voir figure 4.18 pour la légende.

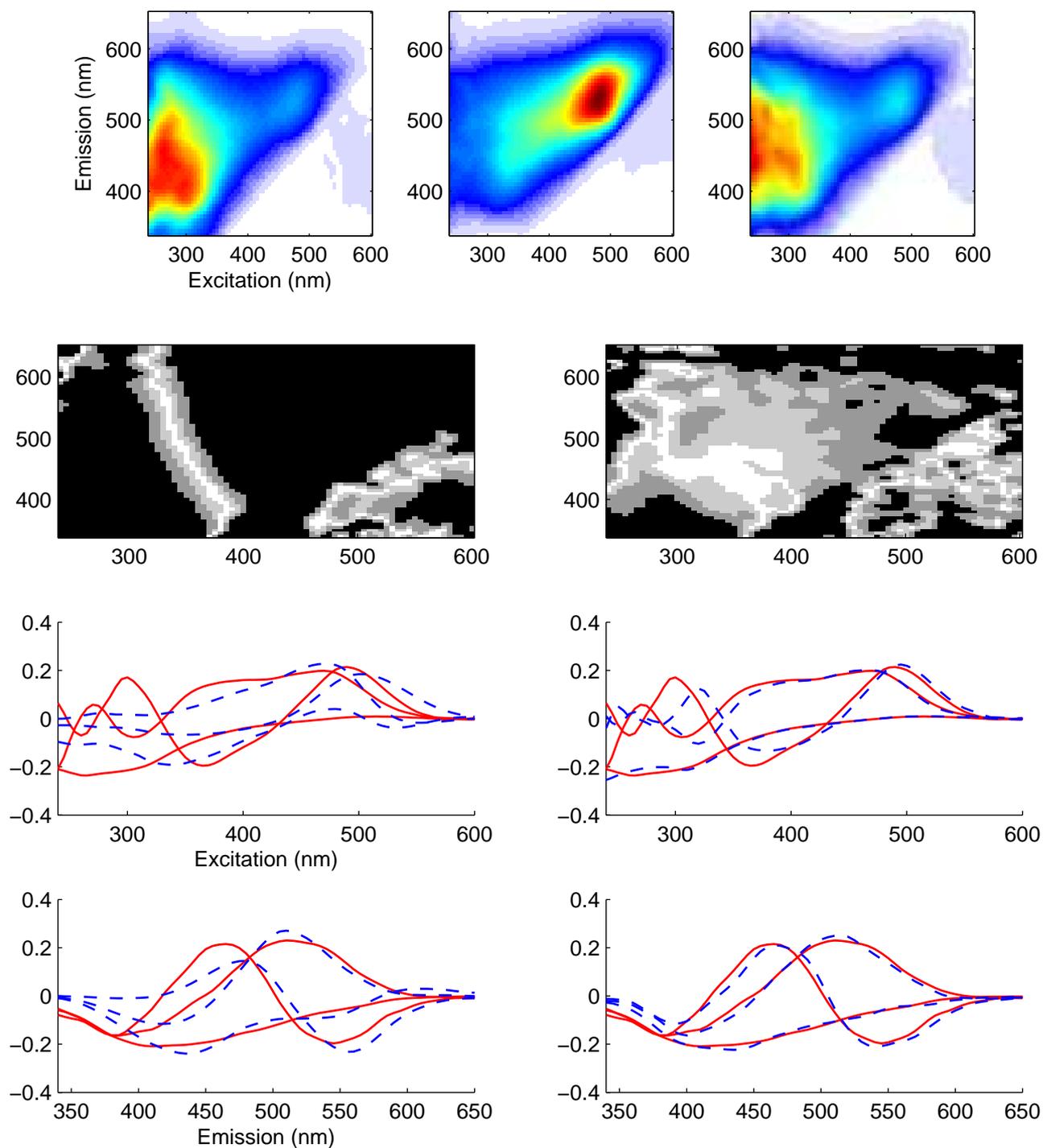


FIG. 4.25 – Jeu 3, échantillon 8 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée
Voir figure 4.18 pour la légende.

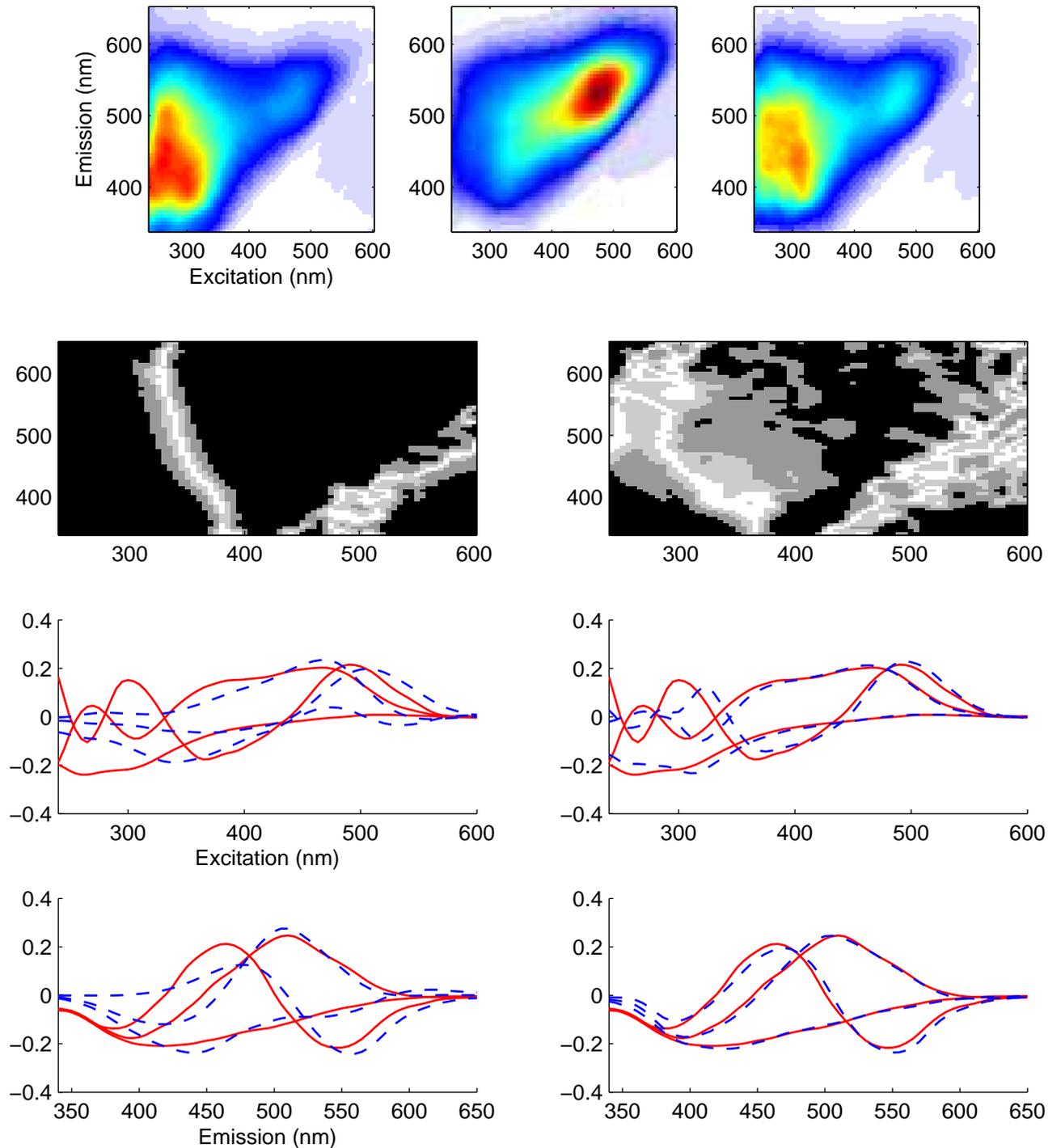


FIG. 4.26 – Jeu 3, échantillon 9 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée
 Voir figure 4.18 pour la légende.

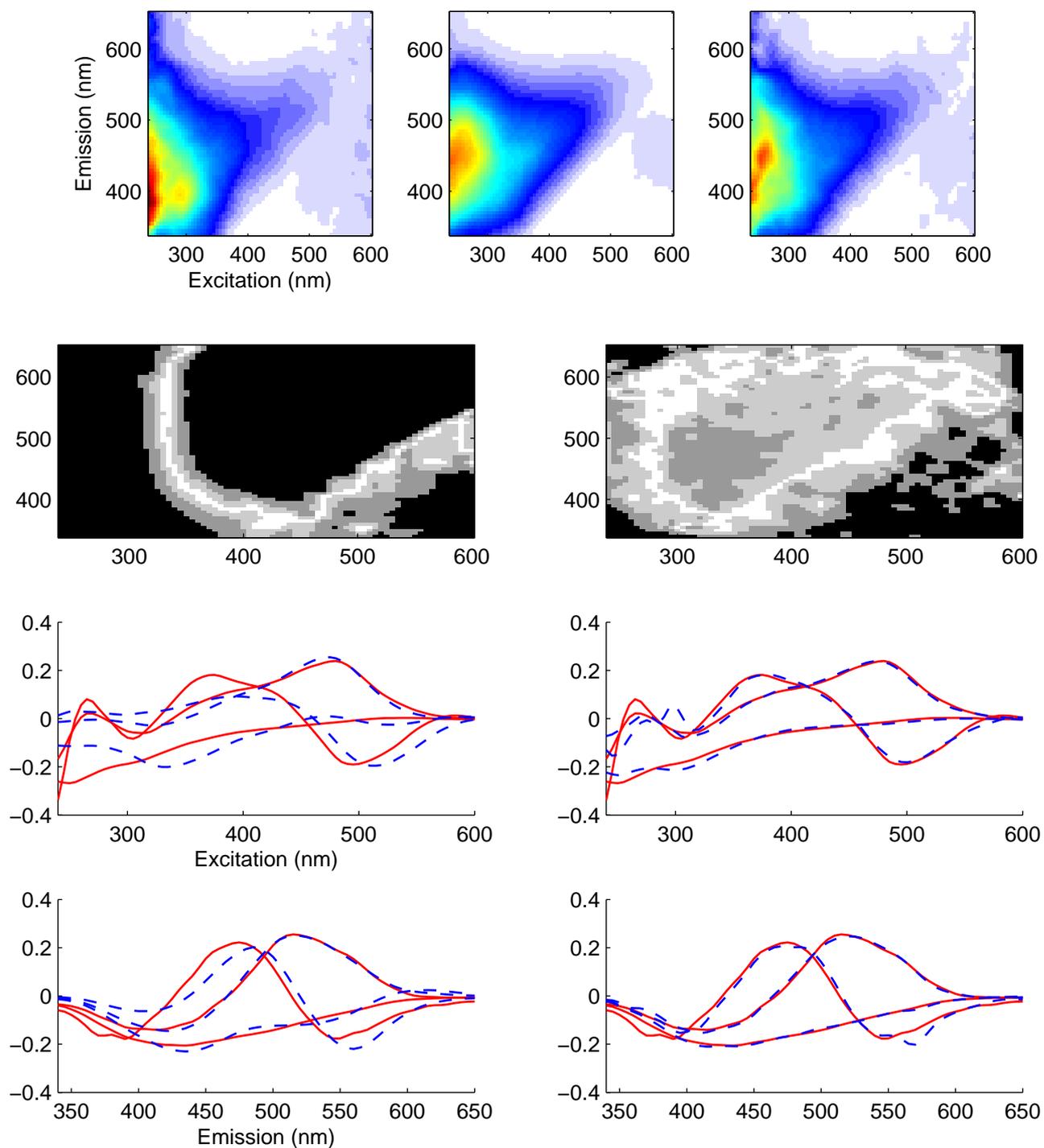


FIG. 4.27 – Jeu 3, échantillon 10 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée
Voir figure 4.18 pour la légende.

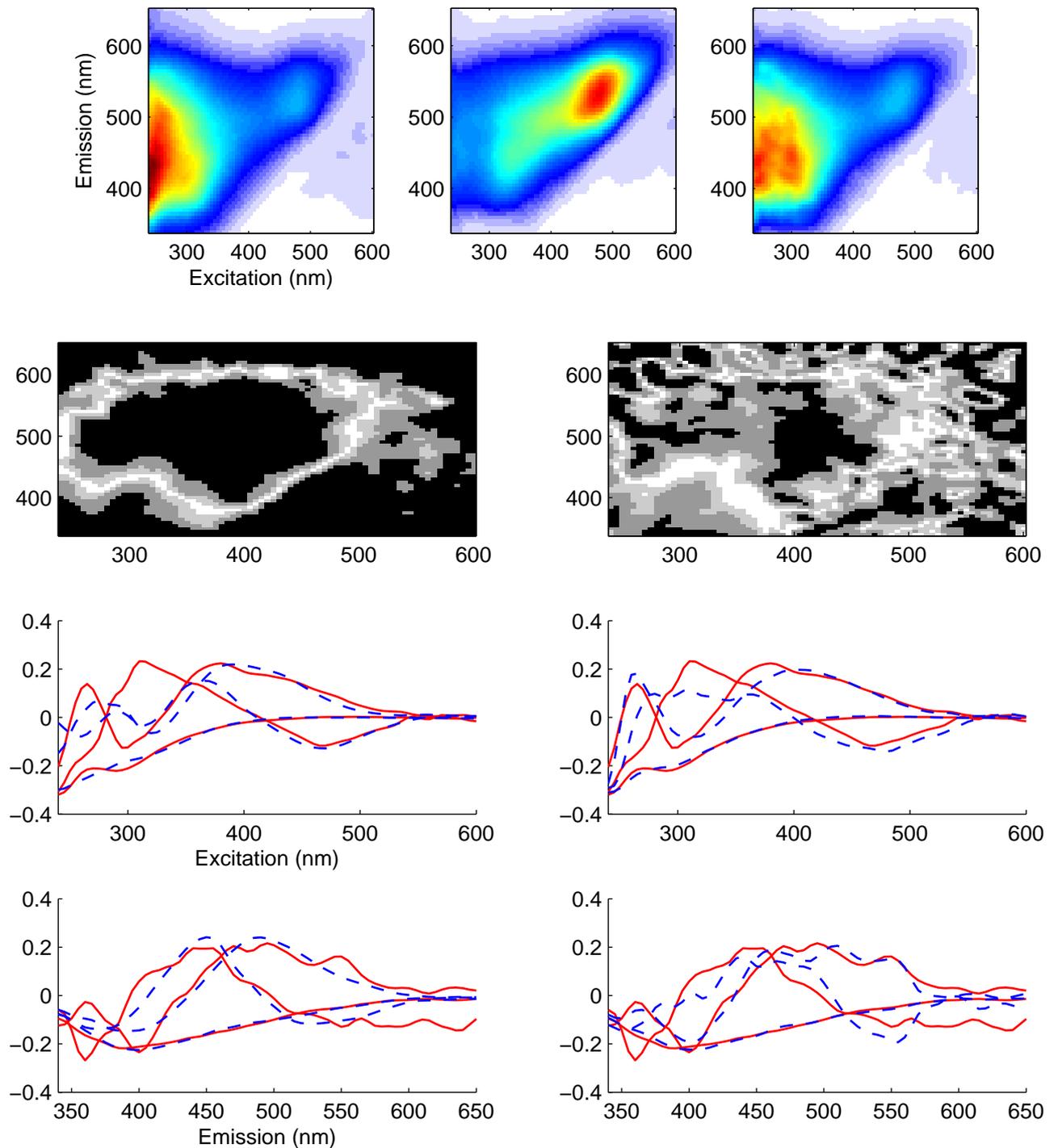


FIG. 4.28 – Jeu 3, échantillon 11 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée
 Voir figure 4.18 pour la légende.

| Echantillon | r Orig. | r Corr.Dil. |
|-------------|-----------|---------------|
| 1 | 51 | 3 |
| 2 | 66 | 3 |
| 3 | 35 | 5 |
| 4 | 59 | 1.4 |
| 5 | 35 | 5 |
| 6 | 44 | 3 |
| 7 | 83 | 4 |
| 8 | 76 | 5 |
| 9 | 83 | 6 |
| 10 | 10 | 5 |
| 11 | 60 | 2 |

TAB. 4.10 – Jeu 3 : Comparaison en énergie avec les MEEF de référence
Rapports r entre l'énergie de l'erreur résiduelle et celle du signal de référence pour les MEEF originales (Orig.) et corrigées par notre approche (Corr.Dil.). Les valeurs sont données en pourcentages

| Fluorophore | r Orig. | r Corr.Dil. |
|-------------|-----------|---------------|
| 1 | 6.6 | 5.6 |
| 2 | 26 | 4 |
| 3 | 7.3 | 8.2 |

TAB. 4.11 – Jeu 3, Mode des concentrations : Comparaison en énergie avec les concentrations relatives de référence

Rapports entre l'énergie de l'erreur résiduelle et celle du signal de référence pour les concentrations relatives estimées à partir des données originales (Orig.), et corrigées par notre approche (Corr.Dil.). Les valeurs sont données en pourcentages

la matière organique est composée de groupements fluorescents inconnus dont nous ne disposons bien évidemment pas des spectres 2D de référence ni des concentrations relatives.

La référence des résultats de PARAFAC sera donc constituée par les loadings obtenus sur la base des MEEF diluées. Les résultats de PARAFAC sur les trois bases de MEEF sont illustrés sur les figures 4.30, 4.31 et 4.33. Les tableaux 4.11, 4.12 et 4.13 récapitulent les résultats chiffrés de la comparaison avec les données de références pour chacun des trois modes.

Mode des concentrations. Dans ce mode, le fluorophore 2 est de loin le plus affecté par l'effet d'écran puisque le profil de concentration obtenu à partir des MEEF originales est radicalement différent de celui obtenu à partir des MEEF diluées ($r=26\%$). Ce chiffre est ramené à 4% après correction. A l'inverse du comportement des jeux précédents, le meilleur résultat après correction (fluorophore 2) correspond au plus mauvais score avant correction. Les courbes nous montrent que cela est dû aux deux échantillons les plus mal corrigés (8 et 9). En effet sans cela, la correction donnerait d'excellent résultats sur les deux autres fluorophores alors que ceux-ci ne sont finalement qu'acceptables.

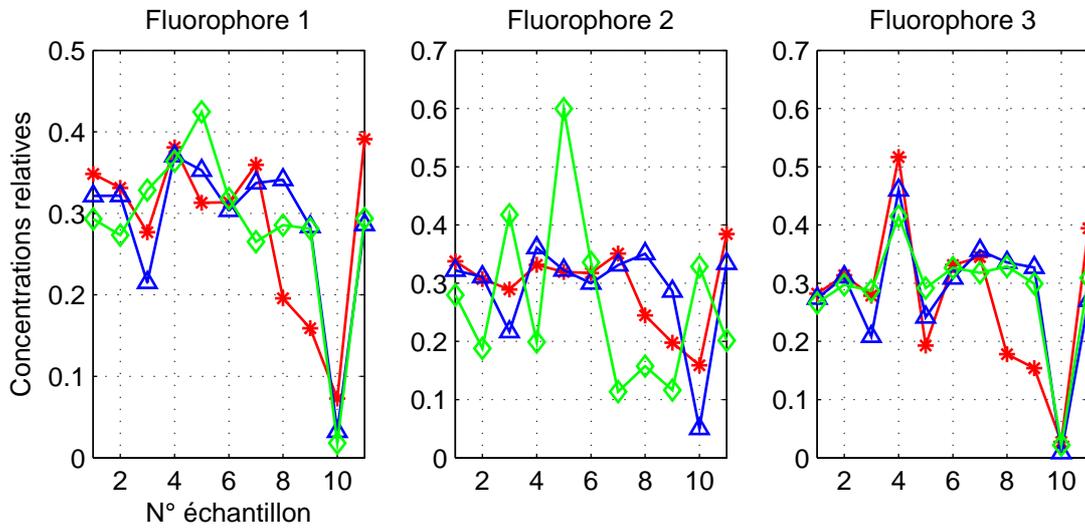


FIG. 4.29 – Jeu 3 : Comparaison des résultats de PARAFAC pour le mode des concentrations relatives

Les trois courbes numérotées représentent les résultats obtenus avec les MEEF de référence (en rouge), les originales (en vert) et celles corrigés par notre méthode (en bleu).

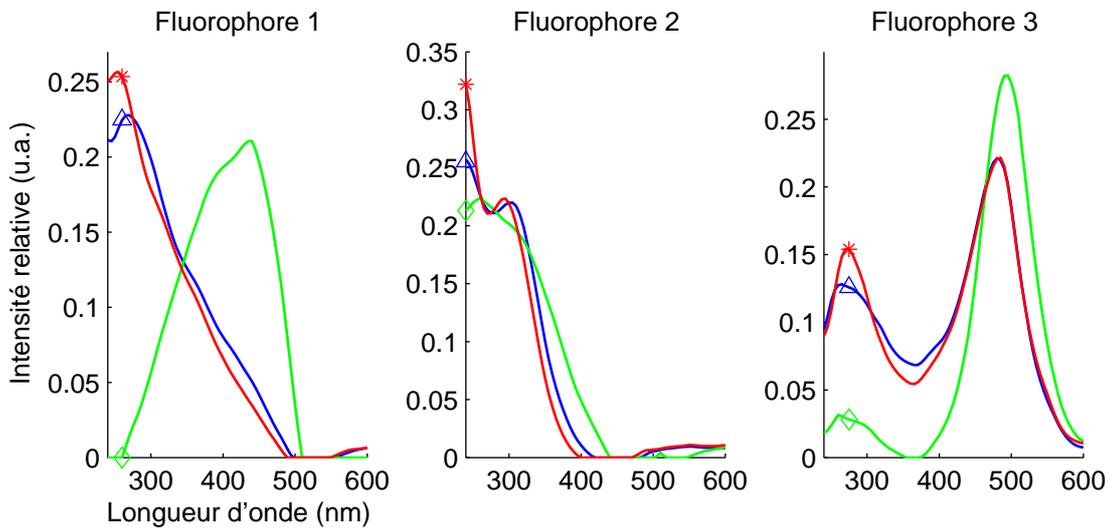


FIG. 4.30 – Jeu 3 : Comparaison des résultats de PARAFAC pour le mode des excitations

Les trois courbes numérotées représentent les résultats obtenus avec les MEEF de référence (en rouge), les originales (en vert) et celles corrigés par notre méthode (en bleu).

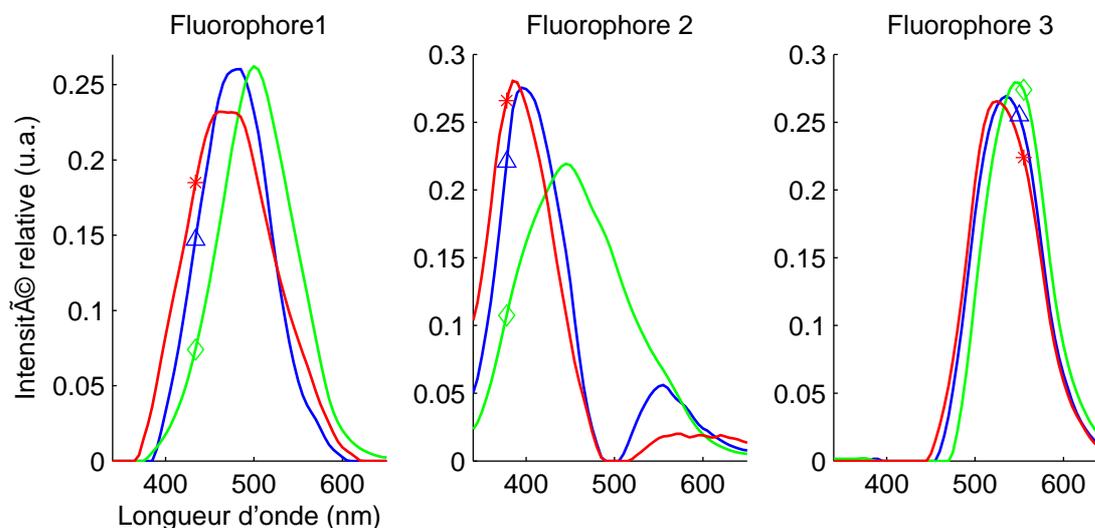


FIG. 4.31 – Jeu 3 : Comparaison des résultats de PARAFAC pour le mode des émissions

Les trois courbes numérotées représentent les résultats obtenus avec les MEEF de référence (en rouge), les originales (en vert) et celles corrigées par notre méthode (en bleu).

| Fluorophore | Test | Orig. | Corr.Dil. |
|-------------|------------------|--------|-----------|
| 1 | Valeur max. | 18 | 11 |
| | Pos. valeur max. | 185 | 15 |
| | Surface | 6 | 5 |
| | r | 110 | 1.5 |
| | err. moy. | 0.09 | 0.01 |
| 2 | Valeur max. | 30 | 20 |
| | Pos. valeur max. | 20 | 0 |
| | Surface | 13.7 | 5.5 |
| | r | 9.6 | 2.1 |
| | err. moy. | 0.02 | 0.01 |
| 3 | Valeur max. | 127.5 | 0.2 |
| | Pos. valeur max. | 10 | 5 |
| | Surface | 25.6 | 1 |
| | r | 35 | 0.6 |
| | err. moy. | 0.0615 | 0.007 |

TAB. 4.12 – Jeu 3, Mode des excitations : Comparaison des spectres d'excitation de référence

Résultat des cinq différents tests sur les spectres d'excitation estimées à partir des données originales (Orig.) et corrigées par notre approche (Corr.Dil.) par rapport aux données de référence : écart sur la valeur du maximum (%), sur la position du maximum (nm), sur la surface (%), rapport entre l'énergie de l'erreur résiduelle et de la référence (%) et erreur résiduelle moyenne

| Fluorophore | Test | Orig. | Corr.Dil. |
|-------------|------------------|-------|-----------|
| 1 | Valeur max. | 13 | 12 |
| | Pos. valeur max. | 40 | 25 |
| | Surface | 4.6 | 9.6 |
| | r | 21 | 4 |
| | err. moy. | 0.04 | 0.02 |
| 2 | Valeur max. | 21 | 1.8 |
| | Pos. valeur max. | 60 | 10 |
| | Surface | 25 | 6 |
| | r | 63 | 6 |
| | err. moy. | 0.08 | 0.02 |
| 3 | Valeur max. | 5.2 | 1.4 |
| | Pos. valeur max. | 20 | 10 |
| | Surface | 3.6 | 1.7 |
| | r | 8.8 | 0.9 |
| | err. moy. | 0.026 | 0.008 |

TAB. 4.13 – Jeu 3, Mode des émissions : Comparaisons les spectres d'émission de référence

Résultat des cinq différents tests sur les spectres d'émission estimées à partir des données originales (Orig.) et corrigées par notre approche (Corr.Dil.) par rapport aux données de référence : écart sur la valeur du maximum (%), sur la position du maximum (nm), sur la surface (%), rapport entre l'énergie de l'erreur résiduelle et de la référence (%) et erreur résiduelle moyenne

Mode des excitations. Les pseudospectres des fluorophores 1 et 3 illustrent parfaitement les deux types d'erreur engendrés par l'effet d'écran. Le premier est complètement modifié par rapport au spectre attendu ($r=110\%$, décalage de 185 nm) tandis que le second voit l'amplitude de son premier pic diminuer au profit du second sans que l'allure générale du spectre ne soit par trop modifiée. La procédure de correction permet de reconstituer très fidèlement les spectres linéaires avec des valeurs de r inférieures ou égales à 2%. On notera que deux des trois fluorophores (2 et 3) présentent des doubles pics qui sont bien retranscrits après linéarisation.

Mode des émissions. Cette fois, seul le fluorophore 2 est perturbé par une forte distortion non linéaire ($r=63\%$, décalage de 60 nm) que notre correction permet de supprimer efficacement. Les décalages des fluorophores 1 (40 nm) et 2 (20 nm) sont également fortement réduits (25 nm et 10 nm). Ces résultats sont donc très acceptables bien que sensiblement moins bons que ceux observés sur les spectres d'excitation.

Conclusion Les résultats sont très concluants puisqu'on peut supposer qu'une étude concrète menée sur ces échantillons donnerait (à quelques erreurs près sur les rapports de concentration de deux échantillons) les résultats attendus alors qu'une absence de correction ou une mauvaise correction des MEEF biaiserait l'ensemble de l'analyse. C'était bien là tout le sens de notre travail.

4.3 Méthode de correction par optimisation

4.3.1 Objectif et principe de la méthode

Les algorithmes décrits dans ce chapitre sont encore en cours de développement ou de test. Aussi, les démarches que nous présentons ici sont moins abouties que la précédente et nous n'accorderons qu'une moindre part à leur description. Nous insisterons principalement sur les grandes lignes puisque les modalités d'application sont susceptibles d'évoluer.

Objectif de la méthode. La méthode de linéarisation des MEEF par faible dilution que nous avons décrite et testée dans la section précédente a un avantage manifeste : elle est applicable à un jeu de donné limité à un seul échantillon. En contre partie, elle nécessite une procédure expérimentale (dilution) et une seconde mesure spectrale par échantillon. Ainsi, lorsque l'on dispose d'un grand nombre d'échantillons (pour appliquer ensuite PARAFAC, par exemple) il peut être intéressant d'utiliser ces données supplémentaires pour s'affranchir de cette manipulation et de cette mesure supplémentaire. L'objectif de cette seconde méthode sera donc de linéariser les MEEF d'un jeu d'échantillon sans manipulation supplémentaire.

Principe de la méthode. La démarche générale est la suivante. Il s'agit dans un premier temps, de déterminer une base quelconque de l'espace vectoriel engendré par les spectres d'excitation $\{\varepsilon_n[k]\}_n$ des fluorophores. Cet espace est noté E et sa dimension est égale au nombre de fluorophores présents dans les mélanges, N .

Rappel : pour un échantillon i et une longueur d'onde m , on note $w(i, m) = \ln(h(i, m))$ (cf. section 4.1).

Nous travaillons ici sur des grandeurs numériques, liées à la longueur d'onde d'excitation k et qui seront vues comme des vecteurs colonnes de \mathbb{R}^K dont les coordonnées sont indicées par k .

Les MEEF \mathbf{I}_{3D_i} sont des matrices de taille (J, K) .

Par définition $\forall(i, j)$, $\mathbf{L}_{i,j}[k]\mathbf{h}_{i,j}$ et $\mathbf{w}_i[k]$ appartiennent à E puisqu'il s'agit de combinaisons linéaires des $\{\varepsilon_n[k]\}_n$. Par conséquent, si l'on note $\{\mathbf{u}_n[k]\}_n$ une base quelconque de E alors il existe deux séries de paramètres $\{a_{nij}\}_n$ et $\{b_{ni}\}_n$ telles que :

$$\mathbf{L}_{i,j}[k]\mathbf{h}_{i,j} = \sum_{n=1}^N a_{nij}\mathbf{u}_n[k] \quad (4.16)$$

et

$$\mathbf{w}_i[k] = \sum_{n=1}^N b_{ni}\mathbf{u}_n[k] \quad (4.17)$$

On peut donc écrire $\mathbf{I}_{3D_{i,j}}[k]$ en fonction de $\{\mathbf{u}_n\}_n$, $\{a_{nij}\}_n$ et $\{b_{ni}\}_n$:

$$\mathbf{I}_{3D_{i,j}}[k] = \mathbf{L}_{i,j}[k]\mathbf{h}_{i,j}e^{\mathbf{w}_i[k]} = \sum_{n=1}^N a_{nij}\mathbf{u}_n[k]e^{\sum_{n=1}^N b_{ni}\mathbf{u}_n[k]} \quad (4.18)$$

Dès lors, si l'on arrive à déterminer une base $\{\mathbf{u}_n\}_n$ de E , les paramètres $\{a_{nij}\}_n$ et $\{b_{ni}\}_n$ pourront alors être estimés pour chaque MEEF, ($i = 1 \cdots I$), par optimisation non linéaire d'un critère. Les paramètres a dépendent de j mais pas les paramètres b . Or ce sont eux qui nous intéressent pour déterminer \mathbf{h} . Il est ainsi inutile d'effectuer une optimisation pour chaque valeur de j . Par conséquent, le critère d'optimisation utilisé ne doit pas dépendre de j . On cherche donc un unique vecteur représentatif de toutes les lignes de \mathbf{I}_{3D_i} . On note alors \mathbf{s}_i ce vecteur désignant la somme des N premières composantes principales en excitation de la MEEF \mathbf{I}_{3D_i} et nous proposons finalement le critère M_i suivant :

$$M_i(\{a_{ni}\}_n, \{b_{ni}\}_n) = \min \left(\left\| \mathbf{s}_i - \sum_{n=1}^N a_{ni} \mathbf{u}_n[k] e^{\sum_{n=1}^N b_{ni} \mathbf{u}_n[k]} \right\| \right) \quad (4.19)$$

Cependant, nous insistons sur le fait que ce critère devra probablement être amélioré car nous n'avons pas pu démontrer rigoureusement l'unicité de la solution. Néanmoins, nous verrons qu'il donne des premiers résultats encourageants.

Pour chaque échantillon i , on déduit alors immédiatement $\mathbf{h}_i[k]$ de 4.17 puis $H(i, j, k)$ et donc la MEEF linéarisée $L(i, j, k)$.

Nous proposons deux procédures différentes pour estimer une base de E . Celles-ci nécessitent un certain nombre d'échantillons (cf section 4.3.4.1) dont les rapports de concentration sont linéairement indépendants. Afin de simplifier le problème, nous considérons que les échantillons possèdent tous le même nombre de fluorophores identiques N et ce nombre est supposé connu.

4.3.2 Estimation statistique d'une base de E

Cette démarche est basée sur des approximations statistiques entre les signaux des différents échantillons du jeu de données. En effet, celui-ci contient une très grande quantité d'information, justifiant ainsi l'emploi d'une telle approche.

Nous savons que les vecteurs propres de la matrice de variance-covariance de $\mathbf{w}_i[k]$ constituent une base de E . La première idée est donc d'essayer d'estimer cette matrice à partir de nos données. Les équations précédentes peuvent s'écrire de manière identique en inversant le rôle des longueurs d'onde d'excitation (indice k) et d'émission (indice j). Nous utiliserons donc plutôt les spectres d'émission en supposant que ceux-ci sont moins corrélés avec \mathbf{w} . Posons alors $\mathbf{Y}_{i,k}[j] = \ln(\mathbf{I}_{3D_i,k}[j])$ et $\mathbf{X}_{i,k}[j] = \ln(\mathbf{L}_{i,k}[j]h_{i,k})$ on a alors :

$$\mathbf{Y}_{i,k}[j] = \mathbf{X}_{i,k}[j] + \mathbf{w}_i[j] \quad (4.20)$$

Les $\mathbf{Y}_{i,k}[j]$ pourront donc être considérés comme les observations de la somme de deux signaux aléatoires, $\mathbf{X}[j]$ et $\mathbf{w}[j]$. Nous noterons \mathbf{C}_y , \mathbf{C}_x et \mathbf{C}_w les matrices de variance-covariance de ces signaux et \mathbf{C}_{xw} la matrice de covariance entre \mathbf{X} et \mathbf{w} . On a alors :

$$\mathbf{C}_y = \mathbf{C}_x + \mathbf{C}_w + \mathbf{C}_{xw} + {}^t\mathbf{C}_{xw} \quad (4.21)$$

Les observations permettent d'estimer \mathbf{C}_y efficacement. Par ailleurs, il faut noter que pour différentes séries d'observation connues (en fait, pour chaque échantillon i fixé),

$\mathbf{h}[j]$ a un comportement parfaitement déterministe. Par conséquent, la matrice de covariance de \mathbf{Y} estimée uniquement à partir de l'une de ces séries pourra constituer une estimation satisfaisante de \mathbf{C}_x . Malheureusement, \mathbf{C}_{xw} reste totalement inaccessible et ne peut généralement pas être négligée devant \mathbf{C}_x . Cette première approche n'est donc pas viable.

Une seconde possibilité serait d'utiliser directement \mathbf{C}_y à la place de \mathbf{C}_w pour la décomposition. En effet, en pratique, \mathbf{C}_y est souvent proche de \mathbf{C}_w (cette remarque n'a qu'une valeur empirique). Dans ce cas, l'espace engendré par les N_y ($N_y > N$) premiers vecteurs propres de \mathbf{C}_y contient E . Cependant, rien ne garantit plus alors que les vecteurs issus de l'étape d'optimisation appartiennent à E . En particulier, l'équation 4.17 n'est plus nécessairement vérifiée. En l'état, et malgré quelques bons résultats, cette démarche n'est pas encore assez fiable.

Nous avons alors envisagé une troisième approche plus simple. Si l'on considère deux MEEF du jeu de données $\mathbf{I}_{3D_{i_1}}$ et $\mathbf{I}_{3D_{i_2}}$ telles que :

$$\mathbf{I}_{3D_{i_1}}[j, k] = \mathbf{L}_{i_1}[j, k] \mathbf{h}_{i_1}[j] \mathbf{h}_{i_1}[k] \quad (4.22)$$

et,

$$\mathbf{I}_{3D_{i_2}}[j, k] = \mathbf{L}_{i_2}[j, k] \mathbf{h}_{i_2}[j] \mathbf{h}_{i_2}[k] \quad (4.23)$$

Chaque spectre d'excitation étant alors vu comme une observation d'un signal aléatoire. Si l'on note \mathbf{C}_{i_1} et \mathbf{C}_{i_2} les matrices de variance covariance de $\mathbf{I}_{3D_{i_1}}[k]$ et $\mathbf{I}_{3D_{i_2}}[k]$, $\mathbf{C}_{L_{i_1}}$ et $\mathbf{C}_{L_{i_2}}$ celles de $\mathbf{L}_{i_1}[j, k] \mathbf{h}_{i_1}[j]$ et $\mathbf{L}_{i_2}[j, k] \mathbf{h}_{i_2}[j]$ on a alors :

$$\frac{\mathbf{C}_{i_1}[k_1, k_2]}{\mathbf{C}_{i_2}[k_1, k_2]} = \frac{\mathbf{C}_{L_{i_1}}}{\mathbf{C}_{L_{i_2}}} \cdot \frac{\mathbf{h}_{i_1}[k_1] \mathbf{h}_{i_1}[k_2]}{\mathbf{h}_{i_2}[k_1] \mathbf{h}_{i_2}[k_2]} \quad (4.24)$$

soit :

$$\ln \left(\frac{\mathbf{C}_{i_1}[k_1, k_2]}{\mathbf{C}_{i_2}[k_1, k_2]} \right) = \ln \left(\frac{\mathbf{C}_{L_{i_1}}[k_1, k_2]}{\mathbf{C}_{L_{i_2}}[k_1, k_2]} \right) + \mathbf{w}_{i_1}[k_1] + \mathbf{w}_{i_1}[k_2] - \mathbf{w}_{i_2}[k_1] - \mathbf{w}_{i_2}[k_2] \quad (4.25)$$

En fixant une ligne (ou une colonne), il vient :

$$\ln \left(\frac{\mathbf{C}_{i_1}[k]}{\mathbf{C}_{i_2}[k]} \right) = \ln \left(\frac{\mathbf{C}_{L_{i_1}}[k]}{\mathbf{C}_{L_{i_2}}[k]} \right) + \mathbf{w}_{i_1}[k] - \mathbf{w}_{i_2}[k] + K_0 \quad (4.26)$$

où K_0 est une constante. Les lignes des matrices L_{i_1} et L_{i_2} peuvent être considérées comme autant d'observations d'un même signal aléatoire. Les matrices de covariance estimées sur ces deux matrices seront donc proches. Le terme $\ln \left(\frac{\mathbf{C}_{i_1}[k]}{\mathbf{C}_{i_2}[k]} \right)$ peut donc constituer une bonne approximation d'un élément de E à une constante additive près. Dès lors, si l'on dispose d'assez de mélanges, présentant des rapports de concentration distincts, il est possible en répétant N fois la même démarche de constituer une base de E aux constantes additives près. Enfin, en rajoutant un paramètre modélisant le terme constant, la procédure d'optimisation décrite plus haut permet d'identifier le terme correctif et d'achever la linéarisation.

Les premiers résultats obtenus avec cette dernière approche, non présentés ici, sont prometteurs mais des études complémentaires restent à mener avant de conclure définitivement sur la pertinence de cette démarche.

4.3.3 Estimation d'une base de E par optimisation

Cette seconde procédure ne nécessite plus le recours des statistiques. Deux MEEF indépendantes (en termes de rapport de concentration) sont là aussi nécessaires pour déterminer un élément de E .

Nous repartons donc des deux MEEF précédentes, modélisées par les équations 4.22 et 4.23. Nous nommerons ici pseudospectres d'excitation de F_i , tout élément de l'espace vectoriel F_i engendré par $\{\varepsilon_n[k]h_i[k]\}_n$. Typiquement, les lignes de \mathbf{I}_{3D_i} sont des pseudospectres d'excitation de F_i . Pour chacune des deux matrices $\mathbf{I}_{3D_{i_1}}$ et $\mathbf{I}_{3D_{i_2}}$, l'analyse en composantes principales nous permet de construire deux bases de F_{i_1} et F_{i_2} respectivement. Nous noterons \mathbf{v}_{i_1n} et \mathbf{v}_{i_2n} ces deux bases et V_{i_1} et V_{i_2} les matrices (K, N) associées comportant respectivement \mathbf{v}_{i_1n} et \mathbf{v}_{i_2n} sur leurs colonnes.

De la même manière, nous appellerons U la matrice (K, N) associée à une base de E . Pour chaque mélange i , on construit la matrice \mathbf{G}_i de dimension (K, N) de la manière suivante :

$$\mathbf{G}_i = [\mathbf{h}_i[k] \quad \mathbf{h}_i[k] \quad \cdots \quad \mathbf{h}_i[k]] \quad (4.27)$$

il existe alors deux matrices carrées \mathbf{R}_{i_1} et \mathbf{R}_{i_2} de dimension N telles que

$$\mathbf{V}_{i_1} = \mathbf{G}_{i_1} \bullet (\mathbf{UR}_{i_1}) \quad (4.28)$$

et

$$\mathbf{V}_{i_2} = \mathbf{G}_{i_2} \bullet (\mathbf{UR}_{i_2}) \quad (4.29)$$

où \bullet dénote le produit de Hadamard (produit matriciel terme à terme). Soient \mathbf{y}_{i_1} et \mathbf{y}_{i_2} deux éléments respectifs de F_{i_1} et F_{i_2} (typiquement, une ligne de $\mathbf{I}_{3D_{i_1}}$ et une ligne de $\mathbf{I}_{3D_{i_2}}$) et \mathbf{p}_{i_1} et \mathbf{p}_{i_2} , deux vecteurs de dimension N tels que :

$$\mathbf{y}_{i_1} = \mathbf{V}_{i_1}\mathbf{p}_{i_1} = (\mathbf{G}_{i_1} \bullet (\mathbf{UR}_{i_1}))\mathbf{p}_{i_1} = \mathbf{h}_1 \bullet (\mathbf{UR}_{i_1}\mathbf{p}_{i_1}) \quad (4.30)$$

et

$$\mathbf{y}_{i_2} = \mathbf{V}_{i_2}\mathbf{p}_{i_2} = (\mathbf{G}_{i_2} \bullet (\mathbf{UR}_{i_2}))\mathbf{p}_{i_2} = \mathbf{h}_{i_2} \bullet (\mathbf{UR}_{i_2}\mathbf{p}_{i_2}) \quad (4.31)$$

On cherche ensuite à déterminer deux autres vecteurs de F_{i_1} et F_{i_2} notés \mathbf{z}_{i_1} et \mathbf{z}_{i_1} tels que :

$$\mathbf{z}_{i_1} = \mathbf{V}_{i_1}\mathbf{q}_{i_1} = \mathbf{h}_{i_1} \bullet (\mathbf{UR}_{i_2}\mathbf{p}_{i_2}) \quad (4.32)$$

et

$$\mathbf{z}_{i_2} = \mathbf{V}_{i_2}\mathbf{q}_{i_2} = \mathbf{h}_{i_2} \bullet (\mathbf{UR}_{i_1}\mathbf{p}_{i_1}) \quad (4.33)$$

Pour cela, on remarque que :

$$\forall k \quad \frac{\mathbf{y}_{i_2}[k]}{\mathbf{z}_{i_1}[k]} = \frac{\mathbf{z}_{i_2}[k]}{\mathbf{y}_{i_1}[k]} \quad (4.34)$$

L'estimation de \mathbf{z}_{i_1} par \mathbf{z}_{i_2} se fait alors par minimisation de l'écart absolu entre ces deux termes. Cela revient alors à déterminer les valeurs de \mathbf{q}_{i_1} et \mathbf{q}_{i_2} minimisant la grandeur S :

$$S = \sum_{k=1}^K \left| \frac{\mathbf{y}_{i_2}[k]}{\mathbf{z}_{i_1}[k]} - \frac{\mathbf{z}_{i_2}[k]}{\mathbf{y}_{i_1}[k]} \right| \quad (4.35)$$

La faible dimension de ces deux vecteurs permet alors d'utiliser dans de bonnes conditions l'algorithme itératif du simplexe. En pratique, après cette étape d'optimisation non linéaire, nous obtenons \mathbf{z}_{i_1}/C et $C\mathbf{z}_{i_2}$, où C est une constante inconnue, indépendante de k . Par ailleurs, il conviendra de démontrer que le critère d'optimisation rend la solution unique à la constante près. On remarque alors que :

$$\forall k \quad C \frac{\mathbf{y}_{i_2}[k]}{\mathbf{z}_{i_1}[k]} = C \frac{\mathbf{z}_{i_2}[k]}{\mathbf{y}_{i_1}[k]} = C \frac{\mathbf{h}_{i_1}[k]}{\mathbf{h}_{i_2}[k]} \quad (4.36)$$

Or $\ln\left(C \frac{\mathbf{h}_{i_1}[k]}{\mathbf{h}_{i_2}[k]}\right)$ appartient à E à une constante $\ln(C)$ près. Par conséquent, $\ln\left(C \frac{\mathbf{y}_{i_2}}{\mathbf{z}_{i_1}}\right)$ est également un élément de E à cette constante près. Le problème de la constante est réglé en ajoutant un paramètre à l'optimisation finale (équation 4.19).

Dès lors, si l'on dispose d'assez de mélanges, présentant des rapports de concentration distincts, il est possible en répétant la même démarche de constituer une base de E . On en déduit alors \mathbf{h} par optimisation de 4.19 puis la matrice linéarisée.

4.3.4 Utilisation pratique et premiers résultats

4.3.4.1 Nombre de mélanges nécessaires à l'estimation de la base

Nous avons proposé deux approches pour déterminer une base de E . Dans les deux cas, chaque vecteur de base nécessite deux mélanges indépendants pour être déterminé. Nous pouvons donc connaître le nombre minimum I_0 de mélanges indépendants nécessaires au traitement de N fluorophores.

Un simple calcul combinatoire montre alors que

$$I_0 \geq \frac{1 + \sqrt{1 + 8N}}{2} \quad (4.37)$$

Cette limitation constitue pour nous une contrainte acceptable. En particulier lorsqu'il s'agit de prétraiter les mélanges avant l'application de PARAFAC comme ce sera le cas dans ce chapitre et pour les applications présentées au chapitre 5. Notons simplement à titre de comparaison que 5 échantillons indépendants suffisent pour linéariser un mélange de 10 composants et que dans [75], Stedmon identifie avec PARAFAC huit composants à partir de plus de 1200 échantillons. Au contraire, il convient bien plus de se poser le problème inverse : quelles MEEF utiliser pour la linéarisation lorsque leur nombre est supérieur à N ? Et lorsque ce premier choix est fait, quel couple de pseudospectres choisir? Nous ne traiterons pas de ces questions dans ce mémoire et nous nous contenterons pour l'heure de vérifier que lorsque le choix est correctement effectué la méthode donne de bons résultats. Nous noterons simplement que le choix de ces paramètres conditionne fortement la qualité des résultats. Les différents résultats exposés ici ne sont donc pas suffisants pour valider complètement la méthode, mais constituent un préalable concluant et par la même, la limite actuelle de nos investigations.

4.3.4.2 Résultats sur les mélanges synthétiques

Tous les résultats que nous présentons ici ont été obtenus en utilisant la seconde procédure de détermination de la base de E (section 4.3.3). Nous donnons dans un premier temps les résultats obtenus sur deux séries de mélanges synthétisés numériquement sur le modèle de l'équation 4.1. La première série est constituée de trois MEEF formées à partir des spectres 2D de la fluorescéine et du sulfate de quinine. Les paramètres liés aux rapports de concentration ont été choisis aléatoirement mais suffisamment élevés pour que l'effet d'écran modélisé ne soit pas négligeable. Pour la seconde série, nous avons rajouté à ces mélanges un composant purement artificiel dont les spectres s'intercalent entre ceux du sulfate de quinine et de la fluorescéine. De cette façon, les chevauchements spectraux sont très larges, comme illustré à la figure 4.32.

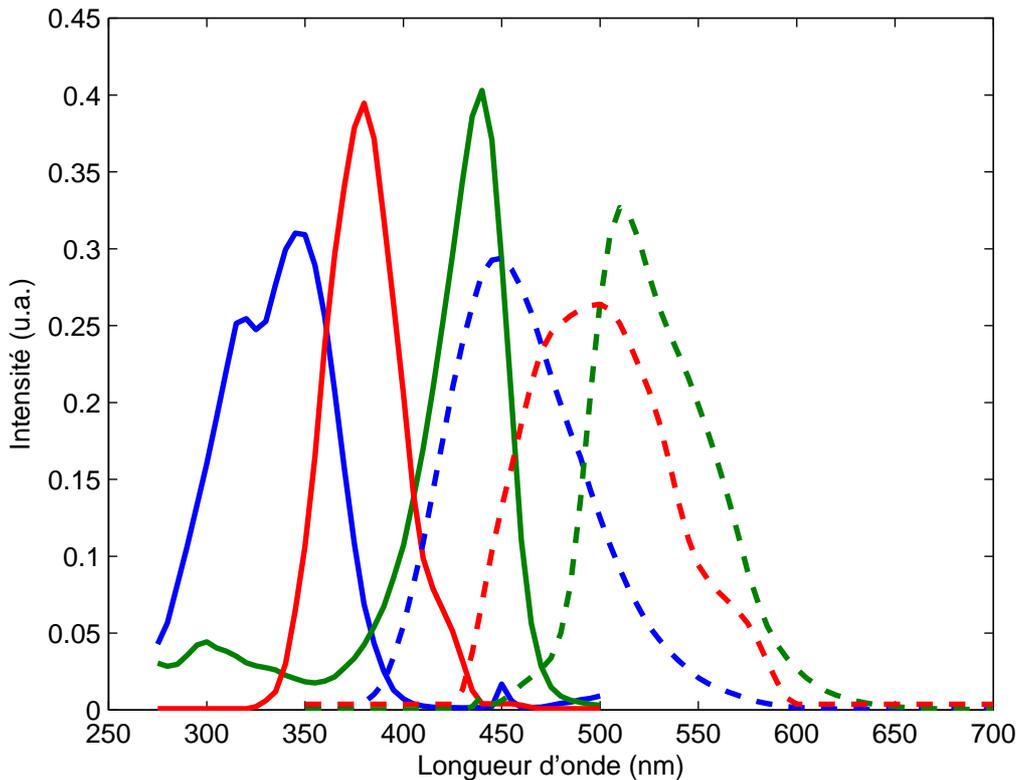


FIG. 4.32 – Spectres 2D utilisés pour la réalisation des mélanges synthétiques à 3 composants

Les spectres d'excitation sont en traits pleins et les spectres d'émission en pointillés

Nous présentons pour chaque série la base de E (au terme constant près) estimée permettant de reconstruire l'image non linéaire par optimisation (figures 4.33 et 4.37), puis les images linéarisées (figures 4.34 à 4.36 et 4.38 à 4.42).

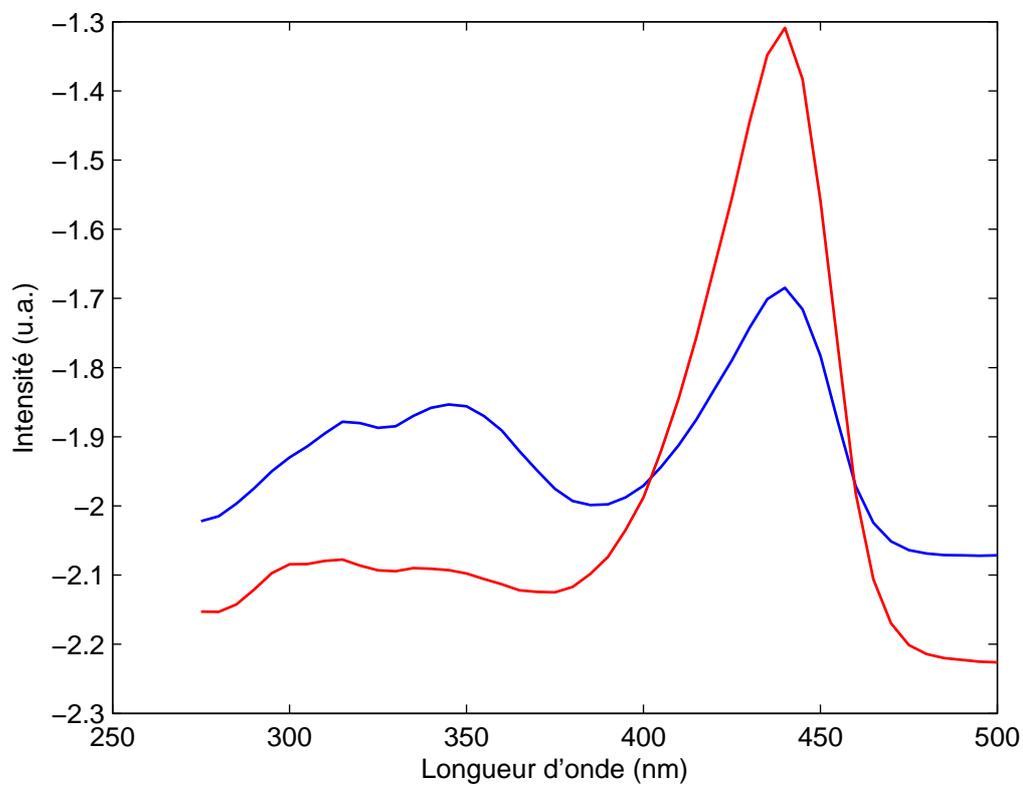


FIG. 4.33 – Base estimée de E sur les mélanges synthétiques à deux composants

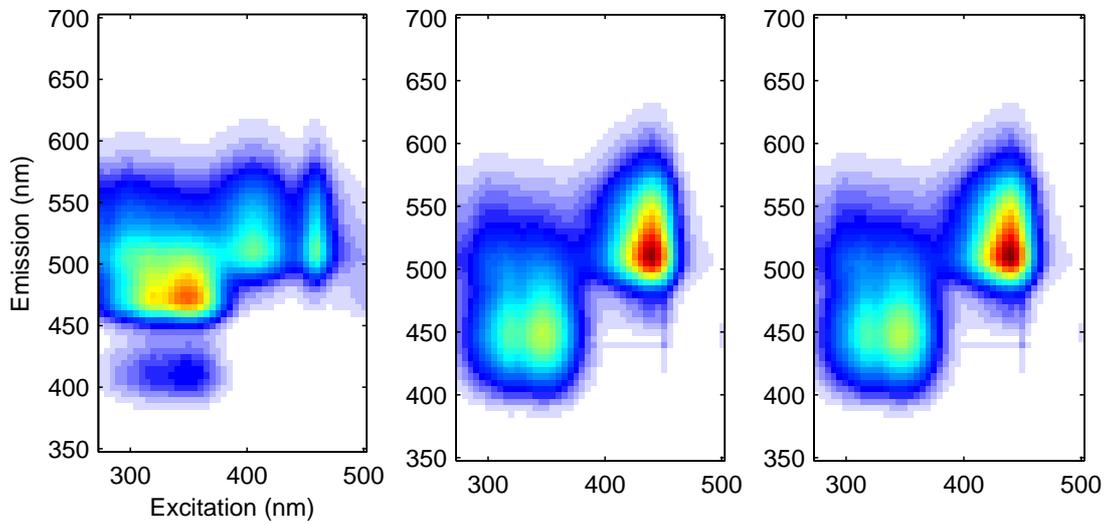


FIG. 4.34 – Mélanges synthétiques à deux composants : linéarisation de la première image

L'image de gauche représente le la MEEF synthétique non linéaire, l'image du centre, la MEEF synthétique linéaire correspondant et l'image de droite, la MEEF linéarisée

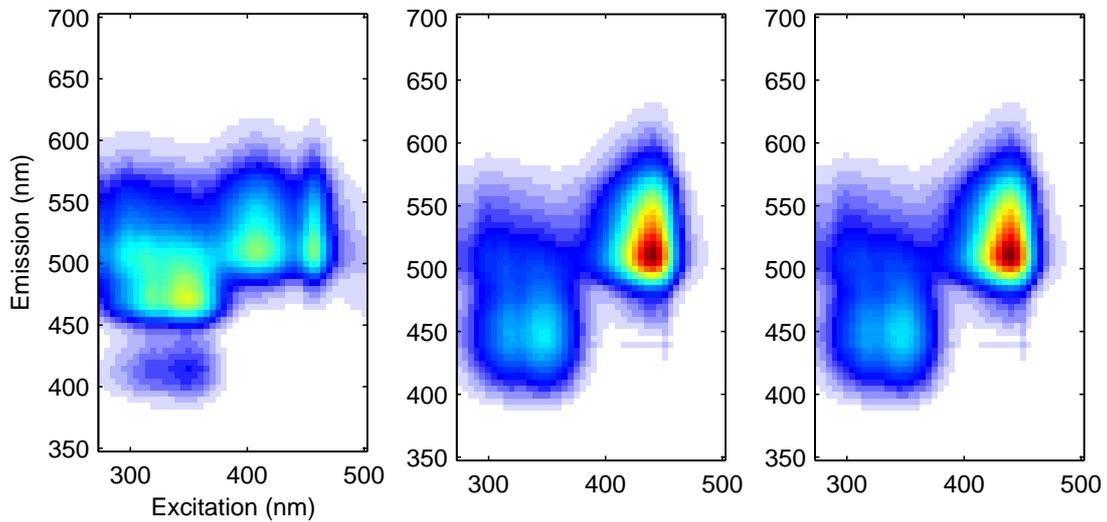


FIG. 4.35 – Mélanges synthétiques à deux composants : linéarisation de la seconde image

L'image de gauche représente le la MEEF synthétique non linéaire, l'image du centre, la MEEF synthétique linéaire correspondant et l'image de droite, la MEEF linéarisée

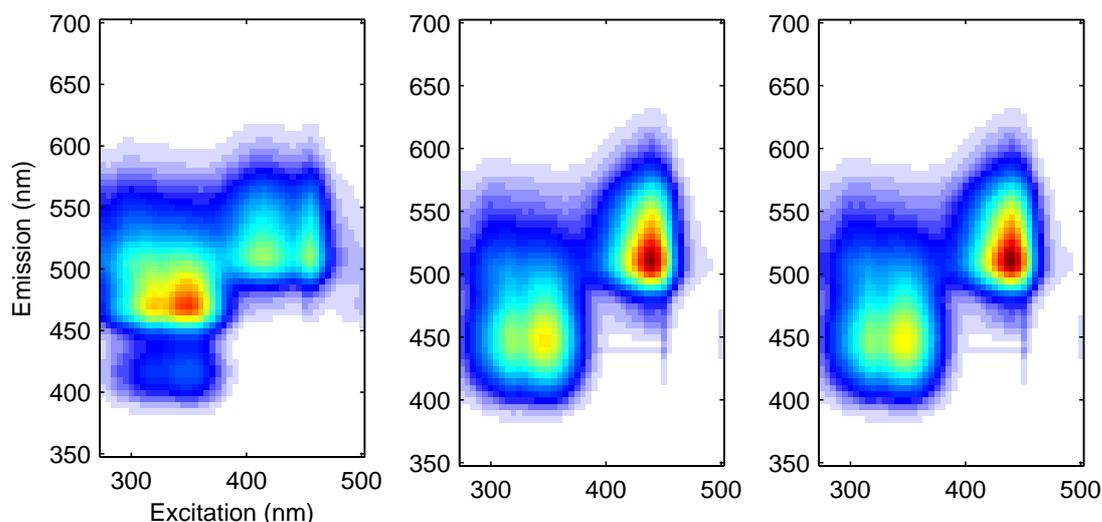


FIG. 4.36 – Mélanges synthétiques à deux composants : linéarisation de la troisième image

L'image de gauche représente le la MEEF synthétique non linéaire, l'image du centre, la MEEF synthétique linéaire correspondant et l'image de droite, la MEEF linéarisée

Pour tous les mélanges, l'erreur résiduelle relative est inférieure à 10^{-10} . Ces résultats n'ont d'autre intérêt que de montrer que la méthode est viable d'un point de vue numérique, en particulier les procédures d'optimisation. Dès lors avec ce type de méthode arithmétique, il n'est pas étonnant d'obtenir des résultats parfaits sur des données idéales.

4.3.4.3 Résultats sur les mélanges réels

Nous avons également testé la méthode sur les solutions 3, 4 et 5 de fluorescéine et sulfate de quinine (tableau 3.1). La figure 4.41 montre les deux vecteurs de base estimés à partir des MEEF correspondantes. Afin de vérifier leur pertinence, nous représentons également les combinaisons linéaires des spectres d'excitation de ces deux composants et d'un vecteur constant, les plus proches au sens des moindres carrés. Au vu de ces résultats, il est clair que nos deux vecteurs forment à un terme constant près une bonne approximation d'une base de l'espace engendré par les spectres d'excitation, à partir de laquelle nous pouvons déduire une bonne approximation des valeurs de \mathbf{h} pour toutes les longueurs d'onde d'excitation. Cependant, celle-ci ne nous a pas permis pour l'instant de linéariser correctement les MEEF. En effet, dans le cas idéal, les termes du vecteur $\mathbf{h}[j]$ valent 1 pour les valeurs de j supérieures à K (longueurs d'onde pour lesquelles les spectres d'excitation n'ont pas été mesurés car supposés trop faibles). Or, ce n'est pas tout à fait le cas en pratique et ce décalage introduit une erreur non négligeable dans la procédure de correction. En revanche, puisque \mathbf{h} est estimé pour toutes les longueurs d'onde d'excitation, nous pouvons linéariser tous les spectres d'excitation de la MEEF, seule leur amplitude respective reste inconnue.

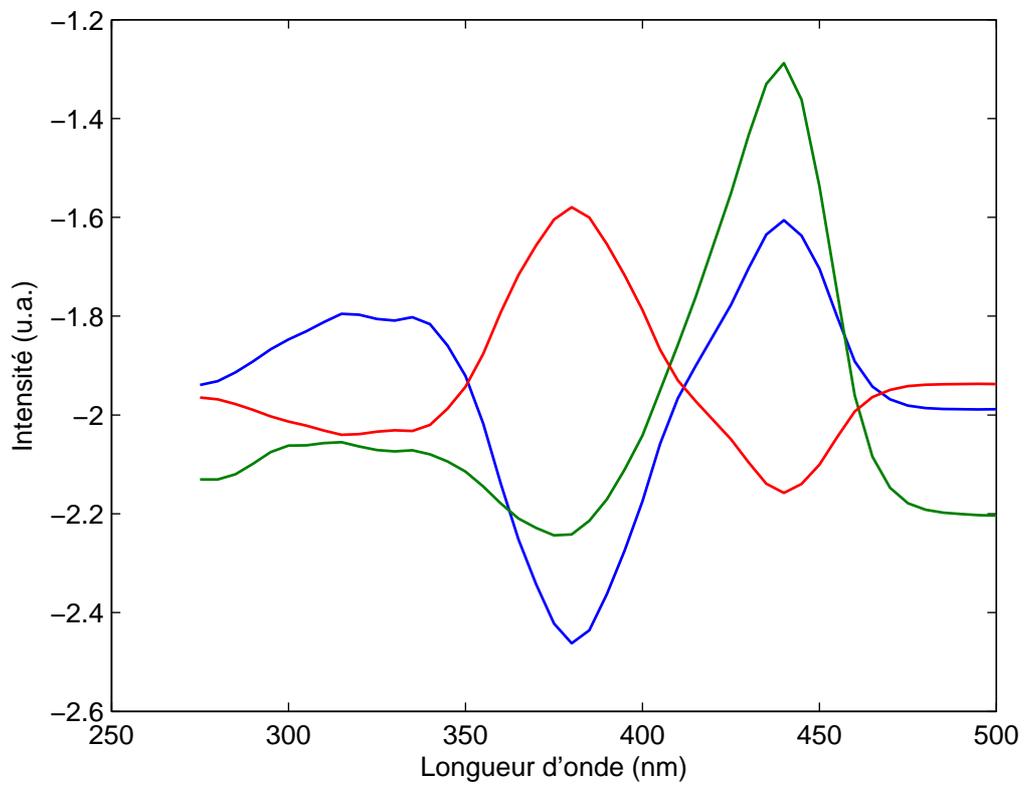


FIG. 4.37 – Base estimée de E sur les mélanges synthétiques à trois composants

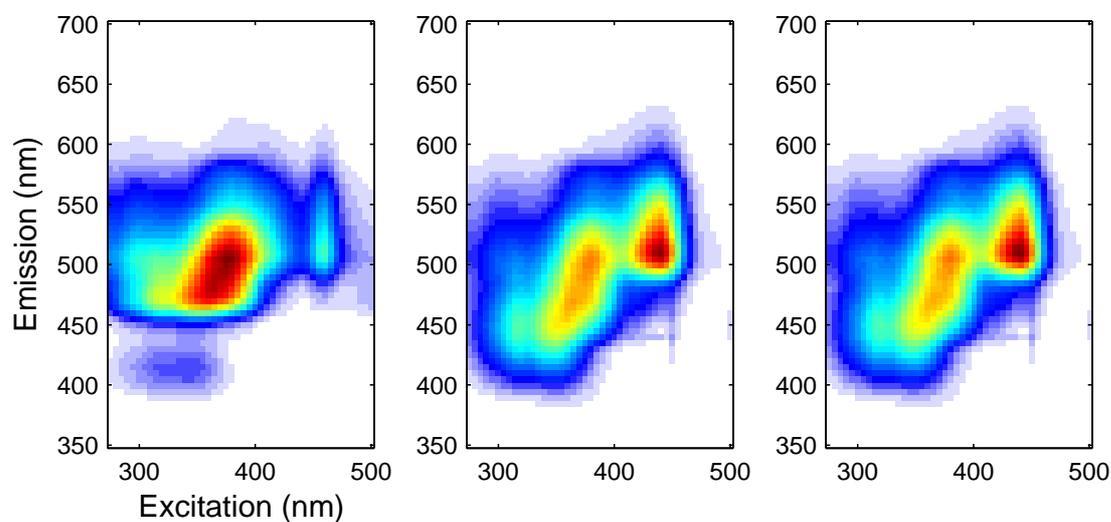


FIG. 4.38 – Mélanges synthétiques à trois composants : linéarisation de la première image

L'image de gauche représente le la MEEF synthétique non linéaire, l'image du centre, la MEEF synthétique linéaire correspondant et l'image de droite, la MEEF linéarisée

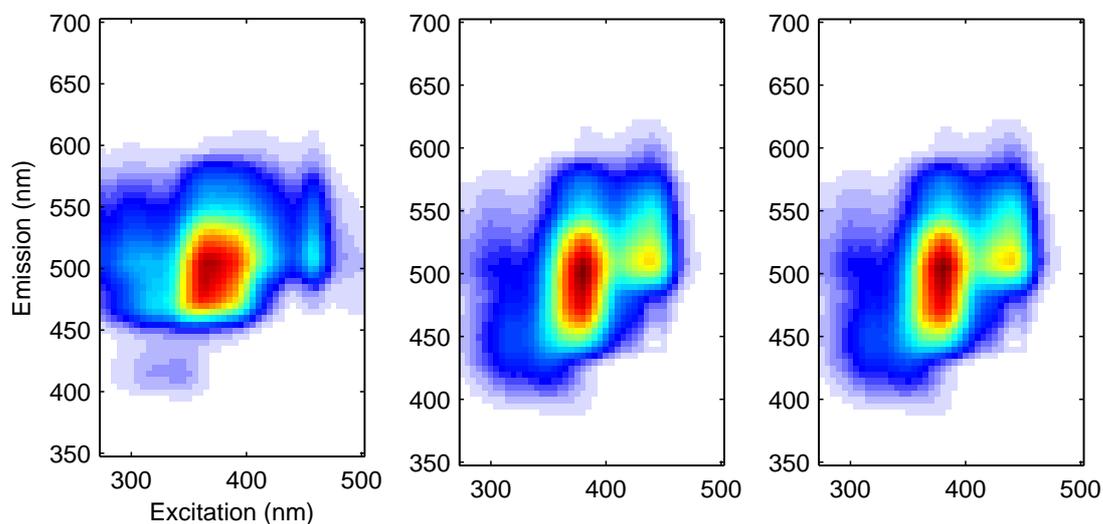


FIG. 4.39 – Mélanges synthétiques à trois composants : linéarisation de la seconde image

L'image de gauche représente le la MEEF synthétique non linéaire, l'image du centre, la MEEF synthétique linéaire correspondant et l'image de droite, la MEEF linéarisée

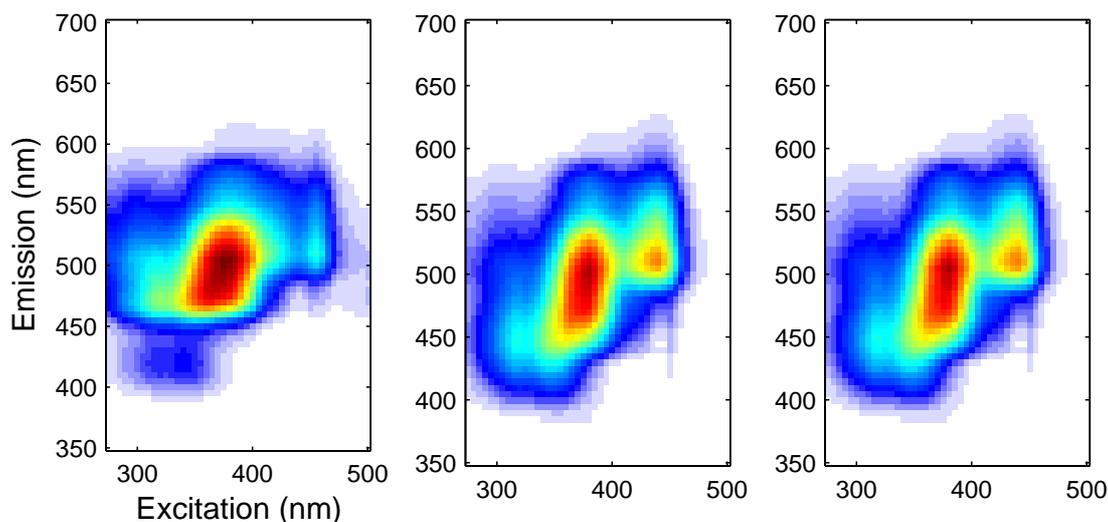


FIG. 4.40 – Mélanges synthétiques à trois composants : linéarisation de la troisième image

L'image de gauche représente le la MEEF synthétique non linéaire, l'image du centre, la MEEF synthétique linéaire correspondant et l'image de droite, la MEEF linéarisée

Nous montrons sur la figure 4.42 les résultats obtenus avec notre vecteur \mathbf{h} estimé pour trois spectres d'excitation caractéristiques de ces MEEF. Nous n'avons pas choisi ici les spectres donnant les meilleurs résultats mais trois des plus représentatifs. Ces résultats pratiques sont très encourageants car comme nous l'avons vu, l'impact sur la qualité de la décomposition d'une erreur relative importante reste limité à condition que celle-ci ne nuise pas à l'allure générale du spectre (au sens des composantes principales) ce qui est le cas ici. Cependant, il nous reste deux problèmes à résoudre avant de pouvoir valider cette approche : trouver une justification rigoureuse d'une solution unique à l'équation 4.19 ou une autre manière de déterminer \mathbf{h} à partir d'une base de E et déterminer le terme correctif à appliquer lorsque $j > K$.

4.4 Conclusion

Nous ne sommes pas en mesure de corriger ni même de détecter les effets d'écrans à partir d'une seule et unique MEEF d'un même mélange. Cependant, nous avons proposé dans ce chapitre plusieurs méthodes de traitement peu contraignantes tirant parti d'informations supplémentaires. Dans tous les cas, il s'agit d'utiliser des MEEF du même mélange dans des concentrations différentes.

La première méthode présentée, basée sur une faible dilution a été validée sur trois jeux de données très différents dont un jeu de matière organique. Les résultats de la linéarisation des spectres 3D sont très satisfaisants puisque tous les spectres corrigés sont très proches des spectres linéaires. Les résultats obtenus après décomposition par PARAFAC doivent être interprétés en deux temps. Sur les trois jeux de données,

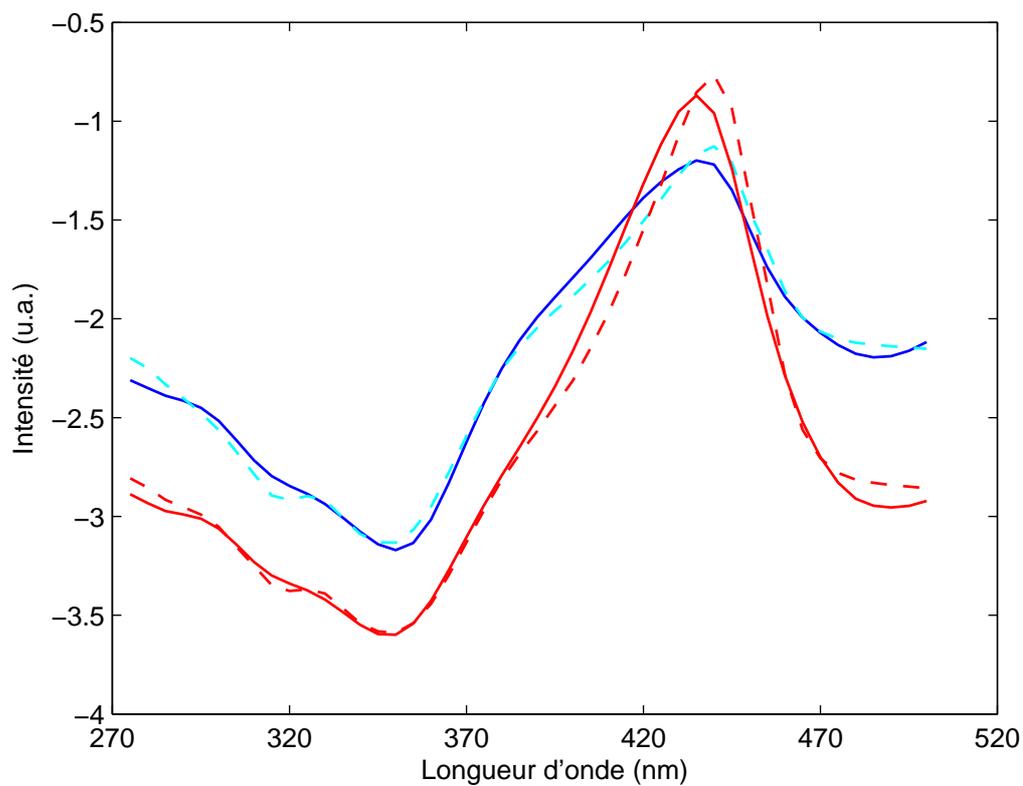


FIG. 4.41 – Base estimée de E sur les mélanges réels à deux composants
En pointillés, la meilleure approximation des vecteurs estimés par combinaison linéaire des spectres d'excitation réels

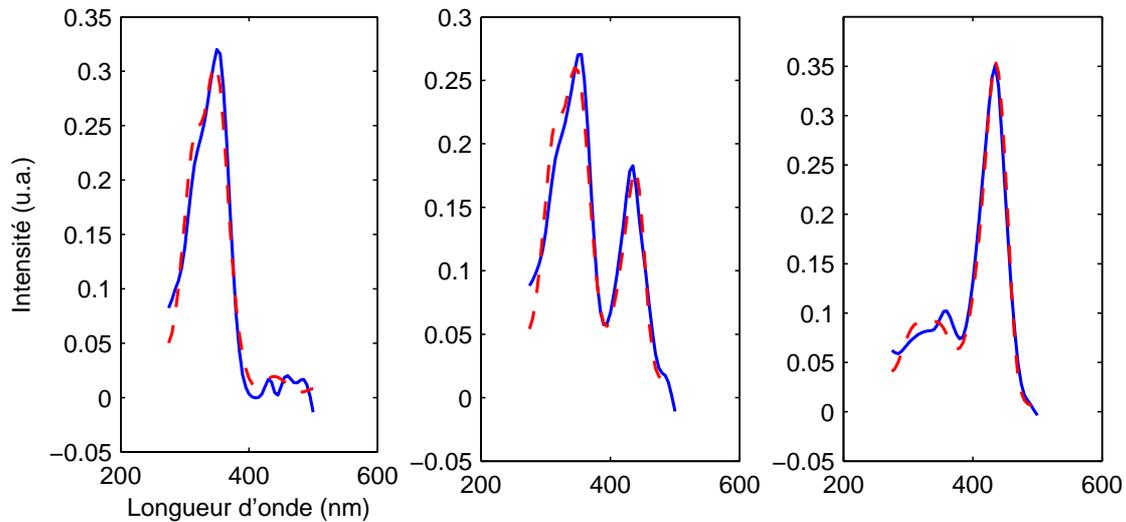


FIG. 4.42 – Linéarisation de spectres d'excitation d'un mélange réel à deux composants. Les spectres réels sont représentés en pointillés et les spectres estimés correspondants, en traits pleins.

tous les loadings spectraux sont correctement estimés et clairement identifiables, ce qui était inenvisageable avant la linéarisation et qui n'est pas toujours possible avec la correction classique. Il est clair que ces bons résultats sur les spectres 2D sont la conséquence directe de la remarque précédente. En revanche, les résultats constatés sur les profils de concentration sont corrects mais restent en deçà, avec localement des erreurs importantes. Nous envisageons deux explications à cela. Premièrement, ce mode est formé de peu de points, le profil global est donc très sensible à des erreurs importantes sur un ou deux mélanges ou échantillons. On constate d'ailleurs que pour les deux jeux de données à trois composants (jeux 2 et 3), les résultats sont bien meilleurs sur le second jeu qui comporte 11 échantillons contre 5 au premier. Par ailleurs, il faut rappeler que le modèle d'effets d'écran utilisé donne de bons résultats mais reste une approximation. Afin d'éclaircir ce point et les limites exactes de la méthode, une étude complémentaire doit donc être menée sur un jeu de MEEF connues bien plus conséquent. Enfin, ces résultats montrent que cette première méthode peut être dès maintenant appliquée avec profit au traitement des spectres 3D de mélanges concentrés. A ce titre, nous donnons un exemple au chapitre suivant.

Nous avons présenté également deux variantes d'une seconde méthode dont l'intérêt est de s'affranchir de l'étape de faible dilution. L'approche statistique que nous avons envisagée doit encore être testée sur des données réelles et constitue plus une piste de réflexion. En revanche, en utilisant la seconde procédure d'estimation de la base de l'espace des spectres d'excitation nous avons déjà obtenu des résultats prometteurs sur notre jeu de données réelles à deux composants. Et, si plusieurs aspects du traitement numérique doivent encore être affinés, nous pensons qu'il sera rapidement possible d'obtenir avec cette approche une alternative intéressante.

Enfin, une autre possibilité que nous avons un temps envisagée, sans parvenir à de

meilleurs résultats serait d'effectuer la décomposition linéaire des MEEF avant la correction de l'effet d'écran puisque la bilinéarité des MEEF est conservée. Dans ce cas cependant, il n'est plus envisageable d'utiliser PARAFAC.

Chapitre 5

Applications au suivi de la matière organique

Le suivi de la matière organique dissoute (MOD) occupe de nombreuses équipes de recherche en chimie de l'environnement [52] [55]. Les propriétés chimiques de la MOD font des spectres de fluorescence 3D un bon outil de caractérisation et un traceur efficace de la MOD dans les sols [53], les rivières [4], les estuaires [73], les eaux usées [27] ou l'eau de mer [9]. Coble a identifié [19] différents pics de fluorescence permettant de caractériser la matière organique en fonction de son origine. Aujourd'hui encore, ces résultats font référence en termes de caractérisation spectrale de la MOD. Depuis peu, l'utilisation de PARAFAC pour ce type d'étude tend à se généraliser permettant notamment d'interpréter les différences d'intensités relatives entre les différents pics de fluorescence identifiés par Coble [36] [74] [75]. Par ailleurs le traitement de l'effet d'écran est souvent évoqué. Cependant, comme nous l'avons vu à l'occasion des chapitres 3 et 4, les méthodes employées sont peu détaillées et leur efficacité est sujette à caution.

Nous présentons dans ce chapitre deux applications des différents traitements présentés dans ce mémoire au suivi de la MOD. A ce titre, le troisième jeu de données utilisé au chapitre précédent, constitué de MEEF de matière organique est en réalité un extrait d'un jeu de 88 échantillons prélevés dans différents sols camerounais. L'analyse de ces spectres doit mettre en évidence la relation entre le type de matière organique et la nature des différents sols. La matière organique prélevée dans les sols est très concentrée. Par conséquent la mesure des spectres 3D correspondants est déformée par un effet d'écran important, comme nous l'avons vu au chapitre précédent. Cette première application illustre notamment l'intérêt de notre procédure de correction lorsque celle-ci est couplée à l'analyse par PARAFAC.

Le second exemple proposé est une étude détaillée de l'influence des fleuves côtiers sur les eaux de la baie de Sépétiba (Brésil).

5.1 Exemple 1 : Sols du Cameroun

5.1.1 Prétraitement des données

Le jeu de données est formé à partir de 44 prélèvements de sol de différentes natures. Ces prélèvements peuvent être classés en fonction de la roche (basalte ou trachyte), de l'utilisation du sol (cultivé, en friche, pour l'écobuage ou à l'état naturel), de la durée d'utilisation en années (plus de dix ans, entre cinq et dix ans, moins de cinq ans et non utilisé) ainsi que de la profondeur du prélèvement. En outre, pour chaque prélèvement, les fractions d'acide humique et d'acide fulvique ont été extraites et séparées, constituant finalement un jeu de 88 échantillons.

Les 88 MEEF correspondantes ont été mesurées entre 240 nm et 600 nm en excitation et entre 340 nm et 650 nm en émission avec un pas de 5 nm dans les deux cas à l'aide d'un spectrofluorimètre Hitachi F4500 corrigé des fluctuations de la lampe en excitation et en émission. Une seconde série de 88 MEEF a été mesurée sur l'ensemble des échantillons dilués deux fois.

Le spectre de diffusion, peu intense, a pu être corrigé pour chaque MEEF par la méthode de Zepp [82]. En utilisant les deux séries de mesures, nous avons appliqué la méthode de correction de l'effet d'écran par faible dilution décrite au chapitre précédent et dont les onze résultats les plus représentatifs nous avaient servi d'illustrations. Avant d'appliquer la décomposition, il est nécessaire de repérer l'existence de points aberrants (outliers). Pour cela, une étude de l'influence (leverage) de chaque échantillon a été réalisée à l'aide de la procédure statistique « Jack-Knife » dont l'application à PARAFAC est détaillée dans [66]. Cette étude n'a révélé aucun outlier parmi les échantillons, confirmant les résultats d'une ACP réalisée préalablement sur le cube de données déplié. Le jeu de données final sera donc un cube de dimension $88 \times 63 \times 73$ formé de 88 spectres 3D corrigés de la diffusion et des effets d'écran.

5.1.2 Décomposition par PARAFAC

Le choix du nombre de composants de la décomposition constitue l'étape la plus importante. Nous avons comparé les résultats de quatre critères parmi les cinq proposés au paragraphe 2.3.3.3 : convergence du modèle, CORCONDIA, LTMC et split half analysis dont les résultats sont donnés au tableau 5.1. La procédure de split half a été menée de la manière suivante : les 88 MEEF du cube sont séparées en deux lots de 44 MEEF de manière aléatoire. Les deux lots issus de cette première séparation ne sont pas nécessairement représentatifs. Par exemple, si plusieurs échantillons du jeu de données original ont des spectres proches, ceux-ci ne doivent pas tous se trouver dans le même lot car cela biaiserait la comparaison. Afin de limiter ce risque, on comparera également deux autres lots de 44 MEEF. Le premier lot est formé des premières moitiés de chaque lot précédent et le second, des MEEF restantes. On applique alors PARAFAC à chacun des deux premiers lots et on compare les loadings spectraux obtenus pour chaque lot. On répète l'opération sur les deux autres lots. Si tous les loadings spectraux sont identiques pour l'un ou l'autre des deux couples de lots, le rang du modèle est validé.

| Critère | Rang du modèle | | | |
|-------------|----------------|--------|--------|--------|
| | 2 | 3 | 4 | 5 |
| Convergence | 97.97% | 98.7% | 99.24% | 99.39% |
| CORECONDIA | 99.6% | 90.88% | 62% | 34% |
| split half | 5 | 10 | 11 | 42 |
| LTMC | 0.0176 | 0.0183 | 0.047 | 0.024 |

TAB. 5.1 – Résultats des quatre critères de choix du rang du modèle.

La convergence du modèle est mesurée en terme de « variance expliquée » par le modèle, la valeur indiquée pour le split half analysis correspond en pourcentage au plus grand écart relatif moyen observé entre deux spectres correspondants issus des décompositions des lots du couple donnant les meilleurs résultats.

Ces différents tests montrent qu'au moins trois composants sont significatifs et qu'un modèle de rang supérieur à quatre serait inadapté. En effet, les résultats de CORECONDIA et LTMC suggèrent de se limiter à trois composants alors que le split half donne également de bons résultats sur des modèles de rang 4. La question du choix se pose donc réellement quant à l'utilité d'un quatrième composant. Avant de trancher en faveur de l'un ou l'autre des deux modèles, il convient de s'interroger sur les répercussions distinctes de cette alternative si le choix s'avérait erroné. Les différences entre les loadings des trois premiers composants du modèle de rang 4 et les composants du modèle de rang 3 sont minimales et n'ont pas d'incidence sur leur interprétation (rappelons cependant que d'une manière générale, les modèles de rangs N et $N + 1$ n'ont pas nécessairement N composants en commun). Par conséquent, si le nombre de composants significatifs est trois, choisir un modèle de rang supérieur nous amènera à interpréter un spectre qui n'a pas de sens physique. Dans le cas contraire on perdra une partie des informations contenues dans le jeu de données. Ce second cas de figure nous semble préférable. En outre, le spectre 3D de ce composant ne ressemble à rien de connu et son profil de concentration est complètement aléatoire à l'inverse des trois autres. Ces dernières observations ne constituent pas un critère absolu mais nous incitent finalement à privilégier le modèle de rang trois.

Les résultats de la décomposition sont donnés sur la figure 5.1. Les spectres 3D normés des trois composants correspondent au produit des spectres d'excitation par les spectres d'émission. Par ailleurs, nous regroupons sur la figure 5.2 les résultats obtenus par le même traitement sur les MEEF non corrigées de l'effet d'écran. La comparaison de ces deux séries de résultats met en évidence la nécessité du prétraitement.

5.1.3 Première analyse des résultats

L'analyse des résultats de la décomposition se fait en deux étapes. Dans un premier temps, nous cherchons à caractériser chaque composant à partir de leur spectre 3D. Dans le cas de la matière organique, cette étape ne peut être aussi précise que pour l'analyse d'un mélange de laboratoire entre des molécules clairement identifiées dont les propriétés spectrales sont connues par ailleurs. Au contraire, la matière organique comporte de nombreux groupements fluorescents dont les spectres sont inconnus. Ainsi, il s'agit généralement d'une analyse comparative avec les différents spectres

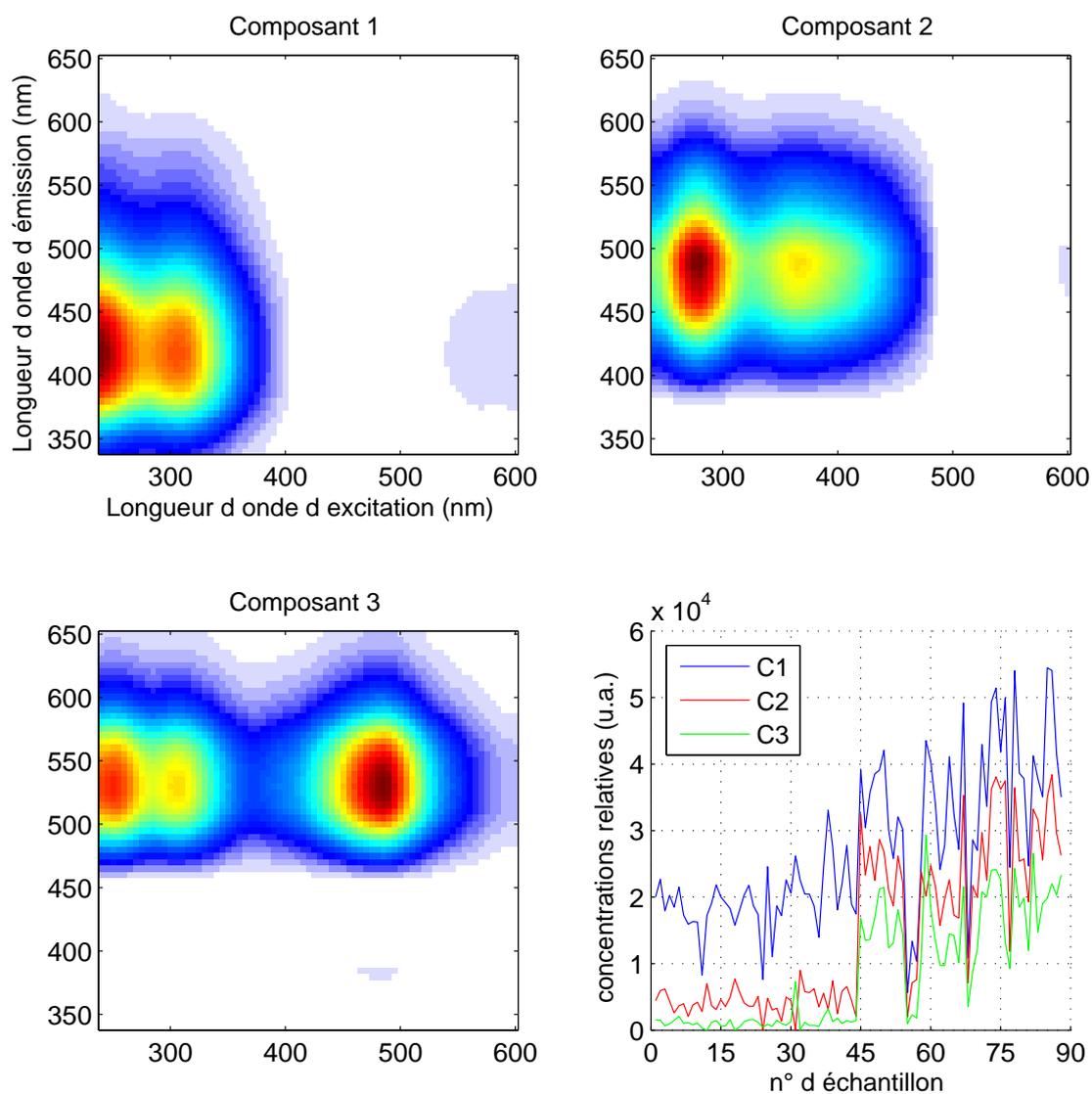


FIG. 5.1 – Résultat de la décomposition sur les MEEF corrigées

Les spectres 3D sont normés par leur norme euclidienne. Les trois courbes tracées sur le quadrant en bas à droite représentent les concentrations relatives du composant 1 (en bleu), 2 (en rouge) et 3 (en vert) à travers le jeu d'échantillons. Les acides fulviques correspondent aux 44 premiers points. Les acides humiques correspondants occupent, dans le même ordre, les places 45 à 88.

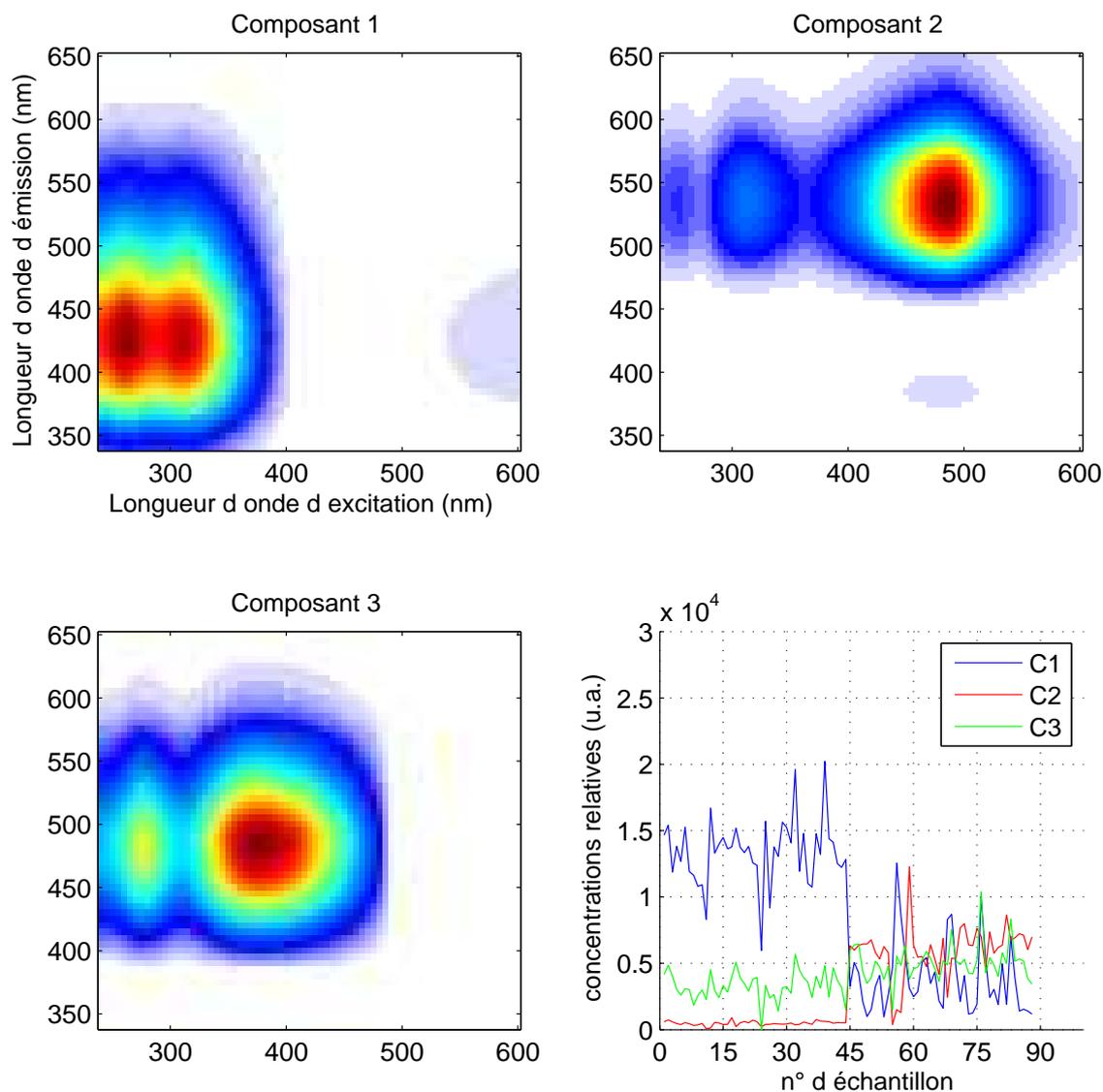


FIG. 5.2 – Résultats de la décomposition sur les MEEF non corrigées

Les spectres 3D sont normés par leur norme euclidienne. Les trois courbes tracées sur le quadrant en bas à droite représentent les concentrations relatives du composant 1 (en bleu), 2 (en rouge) et 3 (en vert) à travers le jeu d'échantillons. Les acides fulviques correspondent aux 44 premiers points. Les acides humiques correspondants occupent, dans le même ordre, les places 45 à 88.

3D rapportés dans la littérature à l'occasion d'études similaires. Le principal critère de comparaison porte sur la position des pics de fluorescence et éventuellement la surface du massif. Les pics de fluorescence A, B, C, M et T décrits par Coble dans [19] font références. A ce titre, il faut souligner l'importance de la correction de l'effet d'écran qui modifie fortement la position des pics et leur intensité relative. Nous analysons ensuite la répartition des différents composants dans la base d'échantillons. Cette étape permet un suivi de la matière organique dans l'environnement. Pour cette première application, nous testons la pertinence des spectres de fluorescence comme facteur discriminant de la nature des différents types de sols. Si le test est concluant, la synthèse des deux étapes, caractérisation des composants et répartition dans les différents sols doit permettre d'en tirer des conclusions géochimiques. Cette étude est encore en cours, toutefois, nous rapportons ici nos premiers résultats. Ceux-ci montrent plutôt une absence de lien statistique entre la nature des sols et les variations de leur spectre de fluorescence.

Caractérisation des composants. Le spectre 3D du premier composant est constitué par un double pic en excitation (240 et 300 nm) et un pic en émission centré sur 420 nm. Ce composant est proche du type A de Coble.

Le spectre du second composant est formé d'un pic principal centré sur le point (275/480 nm) et d'un pic secondaire (360/480 nm) s'apparentant au pic de type C.

Le troisième spectre se compose de deux massifs distincts centrés autour de 530 nm en émission. Le massif principal se situe autour de 480 nm en excitation alors que le massif secondaire a la forme d'un double pic d'excitation (250/300 nm). La présence de doubles pics dans les spectres est toujours sujette à caution. En effet il est difficile de savoir s'il s'agit du spectre réel d'un fluorophore ou du mélange de deux groupements fluorescents distincts mal séparés par l'algorithme. L'allure caractéristique de ce spectre 3D présentant un écart important (200 nm environ) entre les deux massifs laisse penser qu'il s'agit plutôt de deux groupements fluorescents distincts. Il faut préciser néanmoins que le modèle de rang 4 ne favorise pas plus cette séparation et seul l'ajout de nouveaux spectres dans la base de données permettrait de confirmer cette conjecture.

Le premier composant est généralement décrit comme de type fulvique et le second de type humique. Le troisième composant quant à lui n'est pas rapporté dans la littérature (on notera que ce composant est également présent lorsqu'on ne corrige pas l'effet d'écran, sa présence ne s'explique donc pas par notre traitement).

Analyse statistique des résultats. Les profils de concentrations relatives (loadings du premier mode de PARAFAC) sont tracés sur la figure 5.1. Deux constatations s'imposent d'emblée : a) La distinction entre acides humiques et fulviques apparaît clairement : Le composant 3 est pratiquement absent de la composition des acides fulviques et ceux-ci présentent globalement une intensité de fluorescence nettement inférieure. Ces observations sont quantifiées sur la figure 5.3. b) Les trois profils sont très corrélés (coefficients de corrélation : $r_{12} = 0.91$, $r_{13} = 0.89$, $r_{23} = 0.93$). Afin d'affiner ces observations nous pouvons représenter les échantillons dans le sous-espace des composantes principales significatives. Les deux premiers vecteurs propres expliquent

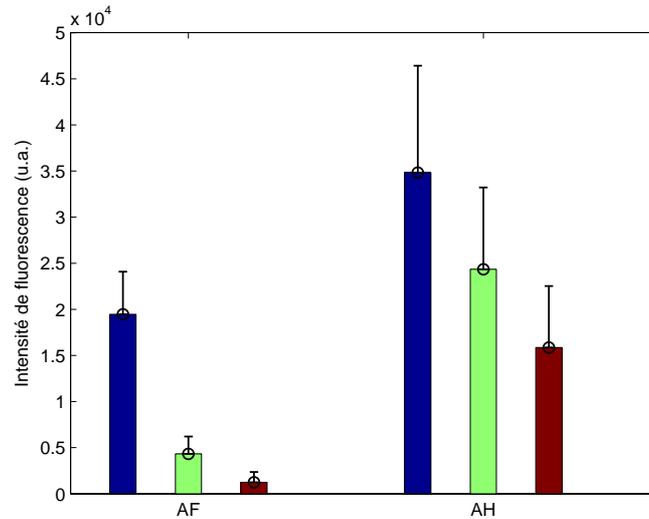


FIG. 5.3 – Moyennes des profils de concentration des trois composants sur les acides humiques et fulviques

Pour chaque groupe d'échantillons, les concentrations relatives moyennes des composants 1 à 3 sont représentées de gauche à droite, avec les écarts types en barres d'erreur

97% de la variance totale et représentent donc une bonne base d'approximation. La figure 5.4 montre la projection des échantillons et des axes originaux (les composants de PARAFAC) sur le plan principal. La distinction entre échantillons de types fulviques et humiques s'opère clairement selon l'axe du composant 2, confirmant en cela les résultats précédents. Nous noterons également que le composant 1 décrit fidèlement le jeu d'échantillons fulviques. Cependant, cette représentation apporte peu d'éléments nouveaux en l'état.

Nous reprenons la même démarche sur les échantillons humiques uniquement. Nous repérons sur la figure 5.5 les échantillons en fonction des quatre catégories les caractérisant : la roche, l'utilisation du sol, la durée d'utilisation et la profondeur du prélèvement. S'il existe un lien entre les spectres de fluorescence et ces catégories, l'ACP doit le mettre en évidence. Malheureusement, comme le laissait présupposer les fortes corrélations entre les profils de concentration, nous constatons sur la figure 5.5 que ce lien ne semble pas exister. Ce résultat conclut donc provisoirement notre étude.

Nous avons montré dans ce travail que, dans le cas de la matière organique extraite, les signaux de fluorescence peuvent caractériser le mode d'extraction en substances humiques ou fulviques. Ces deux caractères ne révélant pas cependant de différences en fonction du temps, de la roche mère ou du type de culture.

Il est à noter que le traitement de l'effet d'écran est primordial. Enfin, les fluorophores obtenus grâce à l'analyse PARAFAC ne permettent pas d'obtenir des spectres unimodaux autour d'un maximum bien défini.

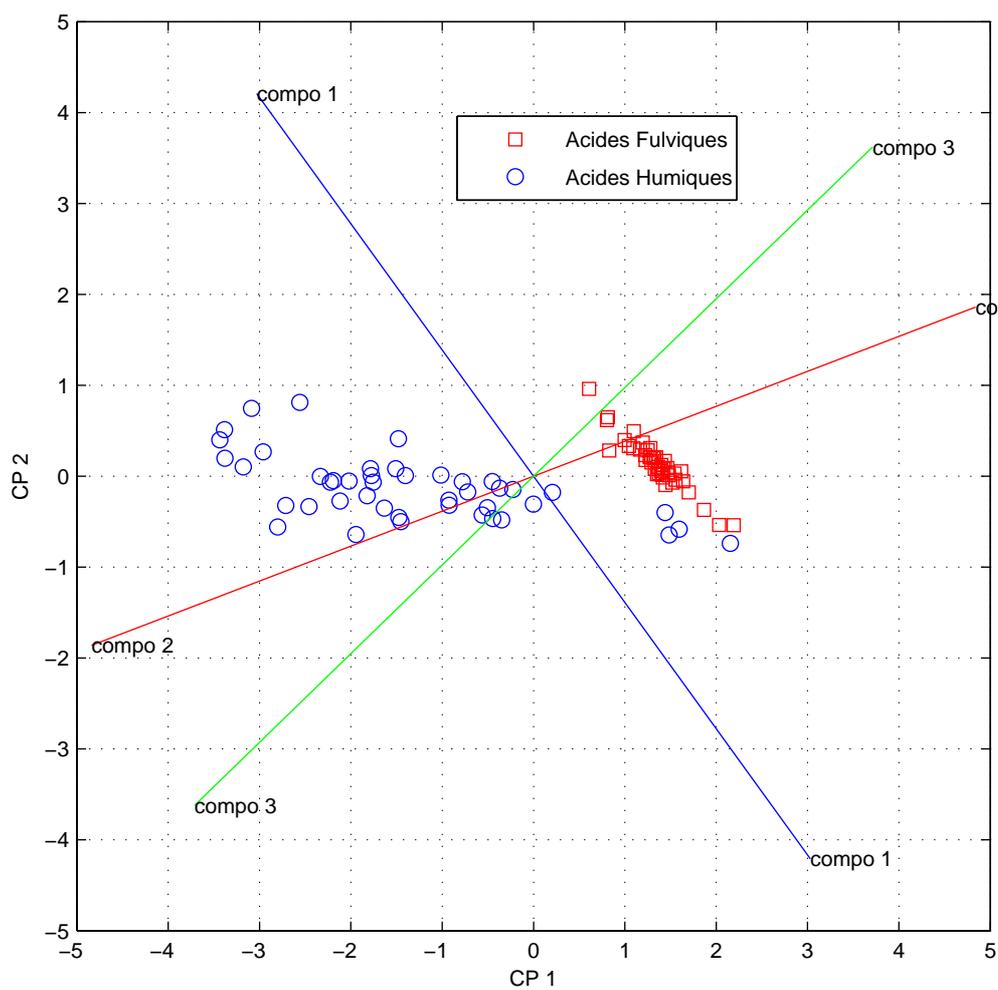


FIG. 5.4 – Analyse en composantes principales des concentrations relatives

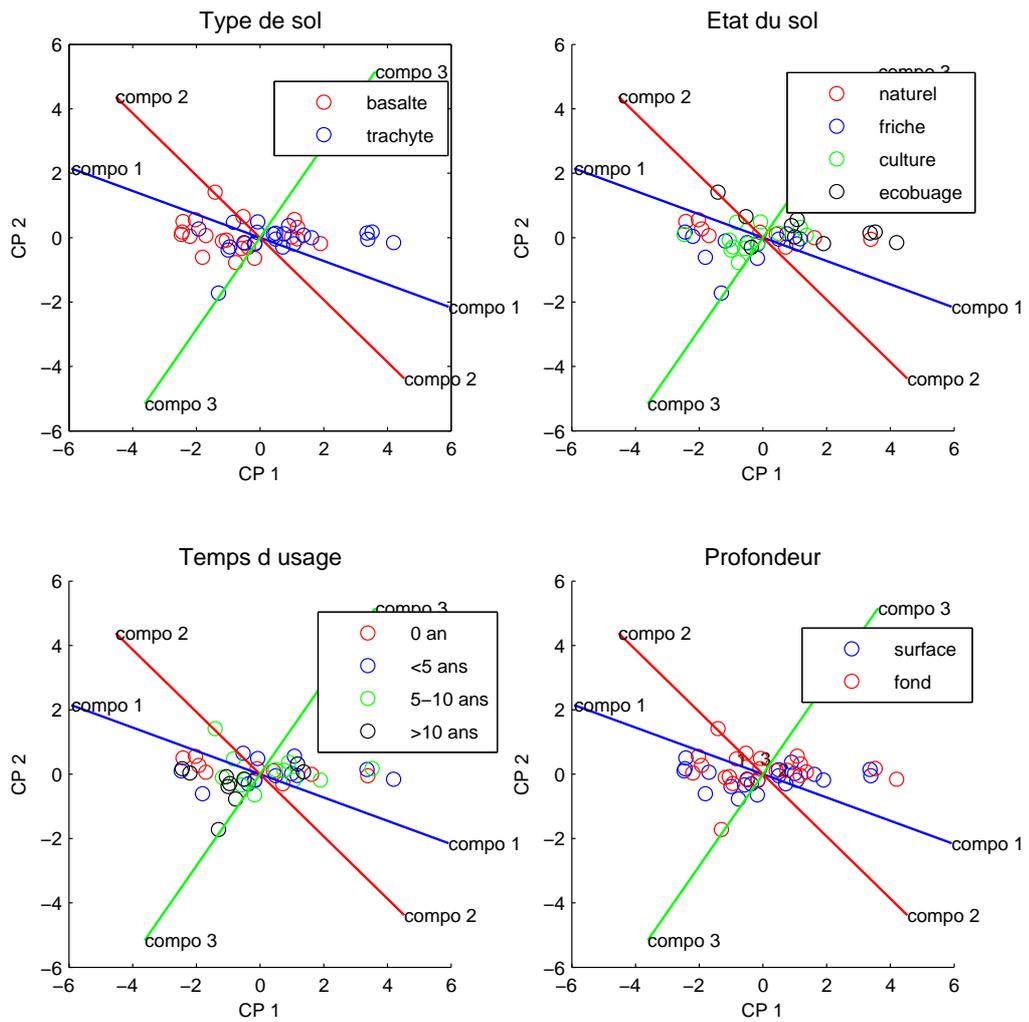


FIG. 5.5 – Répartition des différents types d'échantillons dans l'espace des deux premières composantes principales

5.2 Exemple 2 : Baie de Sepetiba

Nous nous intéressons ici à l'influence des fleuves côtiers sur les eaux de la baie de Sepetiba dont l'un des bassins versants est fortement urbanisé. La matière organique prélevée à cette occasion est naturellement diluée et les MEEF mesurées ne présentent pas d'effet d'écran. En revanche, ce travail montre que la combinaison de PARAFAC pour la caractérisation de la matière organique avec l'ACP pour son suivi peut être fructueuse. En effet, à la différence de l'exemple précédent, un lien pertinent a pu être établi entre la nature des échantillons et leur spectre de fluorescence.

La démarche employée est proche de celle utilisée précédemment pour l'étude des sols. Dans un premier temps, la décomposition des spectres par PARAFAC a permis d'identifier quatre fluorophores de base dont un pic de fluorescence original. L'étude des contributions relatives des différents composants montre qu'une fraction de la matière organique dissoute dans les eaux côtières est principalement d'origine terrestre, à l'exception d'une fraction « marine », identifiée par un pic de fluorescence centré sur 255/410 nm.

Ces résultats ont été récemment publiés. Nous renvoyons donc à la lecture de l'article placé en Annexe A de ce mémoire.

Conclusion

Principaux résultats. Le travail de recherche présenté dans ce mémoire visait à résoudre le problème posé par l'analyse des spectres de fluorescence des solutions fortement absorbantes.

Nous avons rappelé dans la première partie de ce mémoire les principes du mécanisme de fluorescence et les différentes méthodes d'analyse des spectres de fluorescence. Nous avons également mis en évidence sur plusieurs exemples significatifs l'inefficacité des algorithmes multilinéaires pour traiter le cas des solutions présentant d'importants effets d'écran ainsi que les limites des techniques de prétraitement classiques (forte dilution ou correction par le spectre d'absorption).

Nous avons alors détaillé dans une seconde partie plusieurs pistes de réflexion autour de ce problème. Les différentes approches proposées sont basées sur un même modèle physique des effets d'écran. Ce modèle repose sur certaines approximations que nous avons précisées dans ce mémoire. Néanmoins, celui-ci a donné de bons résultats sur les deux bases de test composées de mélanges connus. De même, le schéma d'analyse en deux étapes : a) correction des spectres mesurés, b) décomposition trilinéaire des spectres linéarisés, est commun à toutes nos approches. Nous avons particulièrement détaillé dans ce mémoire notre procédure de correction par faible dilution. Le principe de cette méthode est simple et sa mise en oeuvre ne nécessite qu'un spectre de fluorescence supplémentaire par échantillon. Celle-ci a été appliquée avec succès à trois bases de test dont un jeu d'échantillons de matière organique. Ces différents résultats ont à la fois montré que l'erreur de linéarisation était maîtrisée dans la plupart des cas et que cette erreur résiduelle ne nuisait pas à la décomposition en spectres élémentaires par PARAFAC. Par ailleurs, l'une de ces bases de test nous a permis de comparer les résultats de cette méthode avec ceux obtenus avec la procédure de correction classique, nécessitant le spectre d'absorption. Il en ressort que la méthode proposée est à la fois plus efficace et moins contraignante expérimentalement. Nous avons également décrit les principes théoriques de deux autres approches dont l'objectif était de s'affranchir de toute mesure supplémentaire. Ces approches, dont nous reparlerons dans les perspectives, sont respectivement basées sur les relations statistiques et arithmétiques existant entre les données. Les premiers résultats expérimentaux obtenus sont encourageants.

Enfin nous avons montré dans une dernière partie comment notre travail peut s'insérer dans une application concrète de la spectroscopie de fluorescence au suivi de la matière organique dissoute. En particulier, le premier exemple met en évidence l'utilité et l'efficacité du traitement de l'effet d'écran pour ce type d'étude.

Limites et perspectives. Ce mémoire nous a également permis de souligner les limites de notre approche. Nous avons d'emblée insisté sur le fait que le modèle d'effet d'écran utilisé reposait sur plusieurs approximations, aussi bien dans la description physique du phénomène que dans sa transcription mathématique. Cela se traduit par une erreur de modélisation qui n'est pas négligeable dans tous les cas. Nous avons vu que la procédure de correction par faible dilution que nous proposons répercute logiquement ces erreurs lors de l'étape de linéarisation. Nous noterons à cet égard que les erreurs les plus importantes interviennent sur les profils de concentration, les spectres étant quant à eux correctement estimés dans tous les cas. Par ailleurs, cette méthode qui repose sur le rapport de deux signaux est fatalement sensible au bruit. Nous remarquerons que le bruit de mesure observé sur les spectres de fluorescence est généralement assez faible et que ce problème concerne essentiellement les zones spectrales éloignées des pics de fluorescence. Néanmoins, nous avons pu vérifier que sur certains spectres, une étape préalable de débruitage améliore sensiblement les résultats. Par ailleurs, il convient de connaître avec précision la valeur du facteur de dilution employé pour pouvoir appliquer la procédure. Le degré de précision nécessaire reste à établir. Nous n'avons pas abordé ce sujet dans ce travail car la formulation du problème ne permet pas d'estimer simplement les répercussions d'une erreur commise sur ce paramètre. Nous avons simplement vérifié empiriquement que, sur nos données, cette erreur pouvait être égale à l'incertitude due à la procédure expérimentale de dilution sans fausser les résultats. Une étude rigoureuse reste donc à mener sur ce sujet. En outre, cela nous amène également à la contrainte de cette méthode, qui nécessite une dilution et deux mesures par échantillon. Même si cela est peu contraignant et que le risque d'erreur de mesure est limité, il peut être utile de disposer d'une méthode nous libérant de toute mesure supplémentaire. Les deux autres méthodes de correction que nous avons proposées répondent à ce critère mais doivent être finalisées et testées sur les mêmes données afin d'établir des comparaisons significatives. Enfin, même si cela nous paraît peu envisageable, il serait très intéressant de disposer d'un test permettant de savoir si un spectre donné, sans autre information extérieure, présente ou non un effet d'écran.

Nous avons précisé dès l'introduction que tout notre travail serait basé sur des mesures statiques de fluorescence. En effet, après correction, l'intensité de la lumière d'excitation peut être considérée comme invariante dans le temps jusqu'à la fin de l'acquisition du spectre. Le champ d'application couvert par ce type de dispositif est tellement large qu'il est difficile de voir en cela une limitation. Cependant, la fluorescence est un phénomène transitoire dont la durée de vie constitue souvent une information pertinente. Ainsi, de nombreuses études sont basées sur l'utilisation de LASER impulsionsnels comme sources d'excitation, le tube PM étant alors remplacé par une barrette de photodiodes intensifiée ou par une matrice CCD intensifiée (ICCD). En revanche, les effets d'écran sont toujours présents. Par ailleurs il existe un modèle analytique de la durée de vie de fluorescence. La prochaine étape envisagée est donc de coupler ce modèle avec nos résultats afin d'établir une procédure de traitement de l'effet d'écran sur des mesures dynamiques des spectres.

Annexe A

Exemple d'application 2

Tracing of dissolved organic matter from the SEPETIBA Bay (Brazil) by PARAFAC analysis of total luminescence matrices

X. Luciani ^{a,*}, S. Mounier ^b, H.H.M. Paraquetti ^{c,d}, R. Redon ^a, Y. Lucas ^b,
A. Bois ^a, L.D. Lacerda ^d, M. Raynaud ^a, M. Ripert ^a

^a *Laboratoire PROTEE-ISO, Université du Sud Toulon Var, BP 20132, 83957 La Garde Cedex, France*

^b *Laboratoire PROTEE-CAPTE, Université du Sud Toulon Var, BP 20132, 83957 La Garde Cedex, France*

^c *Departamento de Geoquímica, Universidade Federal Fluminense, Campus do Valonguinho, Niterói, 2020-007 RJ, Brazil*

^d *Instituto de Ciências do Mar, Universidade Federal do Ceará, Av. Abolição 3207, Fortaleza, 60165-081 CE, Brazil*

Received 4 April 2007; received in revised form 18 September 2007; accepted 20 September 2007

Abstract

Fluorescent excitation-emission matrices (FEEM) of the fluorescent dissolved organic matter (FDOM) are widely used for DOM characterization and tracing. In this work, a set of FEEM from sampling campaigns in the Sepetiba Bay (Brazil) was decomposed into independent components using the parallel factor analysis (PARAFAC) algorithm. Four independent components were extracted describing the total fluorescence of the FDOM. The well described peaks A, C, M, B and T were found, and a new peak, A', linked to the C peak, was detected. Relative contribution of each of four components to the total fluorescence confirms that the coastal water has DOM of terrestrial origin, except for the 275Ex/400–500 nm range (nm), which primarily occurs in marine waters.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Coastal waters; Organic matter; Luminescence; Mathematical methods; PARAFAC

1. Introduction

Sepetiba Bay (Fig. 1) is a Brazilian semi-enclosed water body located at latitude 23°S and longitude 44°W, about 60 km south of Rio de Janeiro city whose fast and unplanned development has resulted in high contamination. The site has been studied for many years (Karez et al., 1987; Ovalle et al., 1990; Barcellos and Lacerda, 1994; Marins et al., 1998; Mounier et al., 2001; Lacerda et al., 2001) particularly concerning metal pollution (Paraquetti et al., 2004). A description of the environment is well documented in Lacerda et al. (2001). Some studies were carried out on the role of organic matter (Mounier et al., 2001; Lacerda et al., 2001) but few on its tracing.

* Corresponding author. Tel.: +33 494142516; fax: +33 494142168.

E-mail address: lucianix@gmail.com (X. Luciani).



Fig. 1. Location of the Sepetiba Bay, Brazil.

Dissolved organic matter (DOM) is present in numerous environmental media. Its role in transport is the focus of many research teams (Marins et al., 1996; Benaim and Mounier, 1998; Mounier et al., 2001). DOM has fluorescent properties which permit monitoring in soils (Miano and Senesi, 1992; Trubetskaya et al., 2002), rivers (Ahmad et al., 2002), estuaries (Stedmon et al., 2003; Jaffé et al., 2004), waste waters (Galapate et al., 1998; Vassel and Praet, 2002) or sea waters (Baker and Spencer, 2004; Kowalczyk et al., 2005).

The first studies used fluorescence emission spectra in order to determine the origin of the fluorescent DOM (FDOM), but they were only discriminating in terms of the fluorescence efficiency and emission peak's position (Senesi et al., 1991). Ten years ago, technological advances meant that, it was possible to rapidly obtain synchronous spectra (Cabaniss and Shuman, 1987) and fluorescence excitation-emission matrices (FEEM) (Coble, 1996). Since the reference work of Coble (1996), the FDOM is described by fluorescence maximum localization, giving qualitative information for determining the origin of DOM. Some ubiquitous peaks were rapidly detected, such as peaks A, C, D and M referenced by their emission-excitation range (Mobed et al., 1996; Parlanti et al., 2000; Patel-Sorrentino et al., 2002; Kowalczyk et al., 2005). But, as the total luminescence is a mix of a large number of different fluorescent compounds, the interpretation of variations and comparisons between studies are at present difficult. For these reasons, mathematical tools such as principal component analysis (PCA) are commonly used to obtain this information (Persson and Wedborg, 2001; Boehme et al., 2004; Esteves da Silva et al., 2006).

Recently, the parallel factors analysis (PARAFAC) algorithm was successfully used to treat a large amount of FDOM samples analyzed by FEEM, leading to the decomposition of a set of fluorescent compounds (Stedmon et al., 2003; Stedmon and Markager, 2003, 2005; Holbrook et al., 2006). Due to its unique properties, PARAFAC modeling provides a singular spectral decomposition which is more readily interpretable than bilinear decompositions. This work has two main goals: (1) the characterization of the fluorescent dissolved organic matter (FDOM) in the bay by extraction of the fluorescing spectra, and (2) the emphasis of the river effects on the FDOM tracing through the bay. The spectral representation of the components obtained from PARAFAC analysis of a set of coastal catchments is compared with the Coble peaks. Then, the sample scores in this PARAFAC analysis are used for FDOM tracing.

2. Materials and methods

2.1. Data collection and pretreatment

Several results are available in the literature dealing with the geochemical role of organic matter in the Sepetiba Bay (Mounier et al., 2001; Paraquetti et al., 2004). Samples from coastal waters were taken during

several campaigns carried out in Sepetiba Bay: in 2000 the northern coastal and outer bay samples, influenced by mining processes (Mad campaign); in 2001 the southern bay samples (Tran campaign); and the last in 2005 for input river samples (Rivers campaign) including a mangrove sample taken during low tide. Samples characteristics and references are shown in Table 1 and the sampling location is given in Fig. 2.

All samples were measured by 3D-fluorescence spectroscopy after filtration on site through a GF-filter (pre-heated to 450 °C) to separate the particulate organic matter and the dissolved organic matter. GF-filters were chosen to avoid carbon contamination on further analysis. Before being analyzed, samples were fixed by an addition of 100 µL of 1 M NaN₃ (Aldrich) solution for each 250 mL of sample and conserved at 4 °C in the dark. Sampling flasks were cleaned by nitric acid 10%, rinse several time by milliQ water and stored dry before use. All analysis were completed within two months of sampling. Salinity and pH were measured on site whenever possible. The Dissolved Organic Carbon (DOC) measurements were done with a high temperature combustion Shimadzu TOC-5000, using a four point potassium hydrogen phthalate calibration curve.

Before analysis for DOC, samples (5 mL) were acidified by 100 µL of hydrochloric acid 30% (Aldrich) and bubbled by a CO₂-free nitrogen flux for 10 min to eliminate the inorganic carbon. Triplicates were analyzed until the relative standard deviation reached 2%. The FEEM were obtained from Hitachi F4500 with the following settings: speed scan 30,000 nm/min, excitation and emission bandwidth 5 nm, excitation and emission step 3 nm, response 0.04 s, excitation interval 200–500 nm, emission interval 250–600 nm. Response was

Table 1
Sampling campaigns, position, DOC concentration, on site conductivity and pH measurement

| Sample reference | Campaign | Date | Position | | DOC (mg L ⁻¹) | Cond (µS cm ⁻¹) | pH |
|------------------|----------------------|------------------|-----------|-----------|---------------------------|-----------------------------|-----|
| | | | N | W | | | |
| 1 | Tran | 27 April 2001 | 22°56.20' | 43°54.94' | 2.37 | – | 8.2 |
| 2 | Tran | 27 April 2001 | 22°56.70' | 43°55.68' | 2.37 | – | 8.3 |
| 3 | Tran | 27 April 2001 | 22°59.23' | 43°59.48' | 2.41 | – | 8.3 |
| 4 | Tran | 27 April 2001 | 23°00.95' | 44°02.07' | 2.17 | – | 8.2 |
| 5 | Tran | 27 April 2001 | 23°02.13' | 44°01.74' | 2.23 | – | 8.3 |
| 6 | Tran | 27 April 2001 | 23°02.13' | 44°00.94' | 3.46 | – | 8.2 |
| 7 | Tran | 27 April 2001 | 23°02.95' | 43°58.05' | 2.13 | – | 8.2 |
| 8 | Tran | 27 April 2001 | 23°01.54' | 43°57.61' | 2.19 | – | 8.2 |
| 9 | Tran | 27 April 2001 | 23°00.37' | 43°56.49' | 3.91 | – | 8.3 |
| 10 | Tran | 27 April 2001 | 22°59.16' | 43°55.67' | 2.48 | – | 8.2 |
| 11 | Mad | 21 November 2000 | 22°55.81' | 43°47.16' | 2.36 | – | – |
| 12 | Mad | 21 November 2000 | 22°57.66' | 43°46.12' | 1.22 | – | – |
| 13 | Mad | 27 November 2000 | 22°55.93' | 43°52.33' | 1.49 | – | 8.5 |
| 14 | Mad | 27 November 2000 | 22°55.52' | 43°49.61' | 1.57 | – | 8.5 |
| 15 | Mad | 21 November 2000 | 22°55.00' | 43°45.78' | 2.23 | – | – |
| 16 | Mad | 21 November 2000 | 22°55.00' | 43°45.78' | 2.23 | – | – |
| 17 | Mad | 27 November 2000 | 22°54.97' | 43°45.71' | 4.40 | – | 6.8 |
| 18 | Mad | 21 November 2000 | 22°56.85' | 43°45.78' | 1.32 | – | – |
| 19 | Mad | 27 November 2000 | 22°55.24' | 43°52.31' | 1.87 | – | 8.4 |
| 20 | Mad | 27 November 2000 | 22°55.52' | 43°49.05' | 1.50 | – | 8.5 |
| 21 | Mad | 29 November 2000 | 22°55.33' | 43°53.17' | 2.13 | – | 8.0 |
| 22 | Mad | 27 November 2000 | 22°54.51' | 43°52.59' | 1.34 | – | 8.4 |
| 23 | Mad | 27 November 2000 | 22°59.91' | 43°49.51' | 1.99 | – | 8.6 |
| 24 | Mad | 21 November 2000 | 22°56.16' | 43°46.12' | 2.48 | – | – |
| 25 | Mad | 27 November 2000 | 22°54.55' | 43°51.00' | 1.53 | – | 8.3 |
| 26 | Mad | 21 November 2000 | 22°57.07' | 43°44.82' | 2.67 | – | – |
| 27 | Mad | 27 November 2000 | 22°55.86' | 43°45.71' | 4.01 | – | 7.5 |
| 28 | Rios (Ita) | 13 June 2005 | 22°53.94' | 43°41.14' | 3.44 | 254 | 5.9 |
| 29 | Rios (Guandu) | 13 June 2005 | 22°53.84' | 43°41.59' | 3.39 | 161 | 6.4 |
| 30 | Rios (Sao Francisco) | 13 June 2005 | 22°55.28' | 43°43.06' | 4.2 | 62 | 7.2 |
| 31 | Rios (Guarda) | 13 June 2005 | 22°52.48' | 43°44.67' | 4.18 | 267 | 6.0 |
| 32 | Rios (Itimirim) | 13 June 2005 | 22°54.55' | 43°53.50' | 9.25 | 98 | 5.8 |
| 33 | Rios (Itinguçu) | 13 June 2005 | 22°54.68' | 43°53.27' | 7.7 | 45 | 6.8 |
| 34 | Rios (Mangrove) | 13 June 2005 | 22°54.95' | 43°52.72' | 4.95 | 295 | 6.9 |
| 35 | Rios (Piraque) | 13 June 2005 | 22°59.47' | 43°36.26' | 6.61 | 294 | 7.1 |

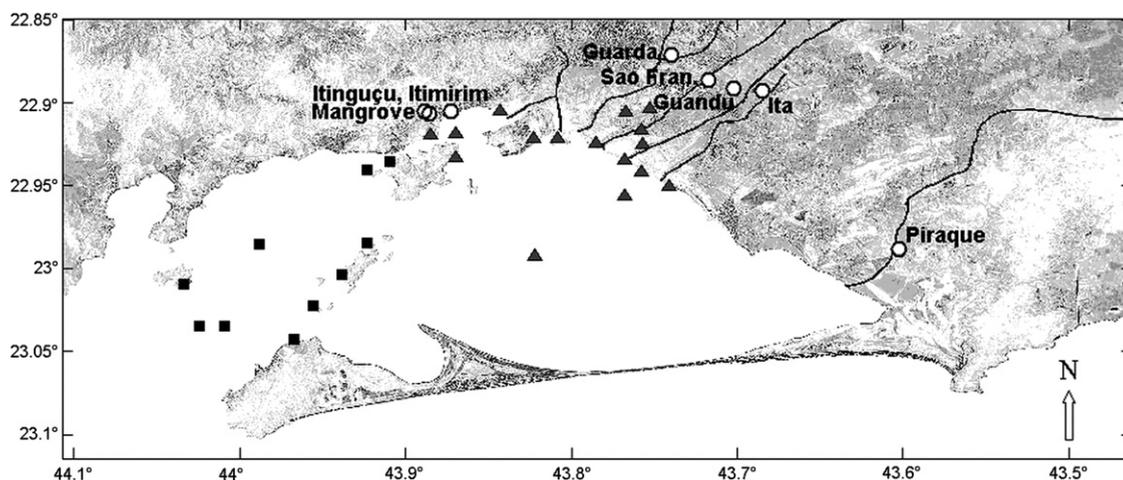


Fig. 2. Sampling map of the campaigns: Mad (▲ symbols), Trans (■) and Rivers (○) localization.

automatically corrected for the excitation light fluctuation. Samples were bubbled with nitrogen for ten minutes to remove dissolved oxygen. The resulting 35 FEEM were numerically corrected to eliminate the Rayleigh and Raman scattering peaks using the three-dimensional interpolation of the data (Delaunay triangulation method) proposed in Zepp et al. (2004). All the 35 corrected FEEM were gathered in a 3-way data array. Before processing it is necessary to remove from the data array, samples with high leverage and residual. For this purpose, the data array was first unfolded into a (35×4331) matrix by combining the two spectral modes. Then, a PCA was performed on the variable space in order to highlight the correlations between samples and to detect possible outliers. No outlier was identified with this primary procedure. A post-analysis will refine this result.

2.2. Parallel factor analysis (PARAFAC) of FEEM data

Since the theory was at first developed for psychology applications by Harshman (1970), PARAFAC is now widely used in various scientific works as a powerful decomposition algorithm. In this paragraph, the method is briefly described, but we refer to the existing literature for further descriptions of the PARAFAC theory and algorithms (Harshman, 1970; Harshman and Lundy, 1994; Bro, 1997; Faber et al., 2003). An interesting review about PARAFAC analysis of FEEM data can be found in Andersson and Bro (2003).

PARAFAC is a multi-way analysis method based on a multilinear decomposition of the data set. In fluorescence spectroscopy, one analyzes a set sample in which each FEEM is a different mixture of F underlying fluorescent components. Considering $x_{i,j,k}$ as the fluorescence intensity of a given sample i at a given couple (j,k) of excitation and emission wavelengths, a_{if} as the concentration of the component f in the sample i , b_{jf} as the amount of fluorescence intensity emitted by f at the emission wavelength λ_j and c_{kf} as the amount of excitation intensity absorbed by f at the excitation wavelength λ_k , the Beer–Lambert law gives in a first order approximation

$$\forall (i, j, k) \in ([1; I], [1; J], [1; K]), \quad x_{i,j,k} = \sum_{f=1}^F \mathbf{a}_{if} \mathbf{b}_{jf} \mathbf{c}_{kf} + \varepsilon_{i,j,k}, \quad (1)$$

where ε is the residual error term. This trilinear relationship between a measured variable $x_{i,j,k}$ and three other unknown variables is a 3-way PARAFAC model of rank F .

For each component, f , vectors \mathbf{a}_f , \mathbf{b}_f and \mathbf{c}_f represent, respectively its concentration profile through the samples set, its normalized emission spectrum and its normalized excitation spectrum. These vectors are called the loading vectors of the decomposition. As many other linear decomposition methods like PCA, PARAFAC minimizes the error term to find the optimal loading vectors in a least square sense. In order to ensure the convergence of the PARAFAC model, we used in this work the Alternating Least Square algorithm (ALS)

(Bro, 1997; Faber et al., 2003) of the ‘N-way toolbox for MATLAB’ (Andersson and Bro, 2002), with non-negativity constraints on the three modes. This algorithm provides a good and fast estimation of the loading vectors. The F matrix products, $\mathbf{b}_f \mathbf{c}_f^T$, provide a normalized FEEM “basis” of the set sample while vectors \mathbf{a}_f are stored in the F columns of the score matrix \mathbf{A} . Therefore, the i th row of \mathbf{A} contains the scores of the i th sample in the PARAFAC components basis. Thanks to the linear independence between the loadings and the trilinear nature of the data set, the solution of the PARAFAC decomposition is unique, in the exception of the trivial indetermination on loadings scale and order (Harshman, 1972; Sidiropoulos and Bro, 2000). This provides a very well defined outcome compared to bilinear methods which makes the PARAFAC analysis an appropriate technique for chemical interpretations.

However, the number of significant fluorescing species in the samples is unknown. Without *a priori* knowledge (de la Peña et al., 2006), several methods help to determine the “best number of components” or best model rank, which is a crucial step of the PARAFAC processing. Residual variance analysis, core consistency diagnostic (CORCONDIA) (Bro and Kiers, 2003) and split half analysis (Harshman and Lundy, 1994) were used in this study. The first runs of the algorithm were made without any other pre-treatment. During this preliminary step, it was pointed out that nine samples (15–17, 25–29 and 35) exhibited a high leverage level. This small number of highly fluorescing samples could have a disproportionate influence on the ALS convergence. Therefore, these could hide some components which could be present in the set sample but not in this influencing subset. However, the relative residual variance was less than 5% for every sample arguing there was no outlier. This assumption was confirmed by the Identity Match Plots from a jackknife procedure (Riu and Bro, 2002) (not shown). Consequently, all the samples were kept in the data set but, each 35 FEEM were weighed by the inverse Euclidian norm before new processing. With this weighed data set, the CORCONDIA was 92%, 66.73%, 40.15%, 33.33%, 17.68% for the two to six components models. However, the variance explained percentages were respectively, 94.17%, 96.21%, 98.23%, 98.55% and 98.84%, indicating that the third and fourth components have a reasonable usefulness while the fifth and the sixth components should describe the noise or non-trilinear deviations. Moreover, the spectral shape of the components obtained with the four component model was meaningful. In addition, a “split half analysis” was tried by splitting the sample mode in two partitions of independent halves. The best correlations (above 0.9) between the loadings of the two spectral modes for each spectrum were found for the four component model. According to these results the linear least square estimation of the sample scores was performed by fitting a four component PARAFAC model on the raw data set, with the spectral loadings obtained from the weighted data set. Finally this model gave a good general description of all the samples in the data set and the dissimilarities were mostly localized in zones with almost no fluorescing signal.

One should note that fluorescence excitation measures under 250 nm are usually avoided with standard spectrofluorimeter. This is due to the lamp limitations which degrade the signal to noise ratio at the lowest wavelengths. In order to evaluate the influence of the noisy 200–250 nm excitation range, all the FEEM were cut down under 250 nm in excitation. Then, the PARAFAC results between the original and truncated data sets were compared in the excitation range above 250 nm. The model rank was estimated to four in both cases and no significant modification of the loadings shape was found whether the 200–250 nm range was included or not. Actually, the noise in this region is modeled when more than four components are used. In addition, a large part of one excitation peak would be lost if the data were truncated. Eventually, the 200–250 nm excitation range was kept in the PARAFAC proceeding.

3. Results and discussion

3.1. PARAFAC spectral results

The spectral loadings and the corresponding four components based on FEEM are presented in Fig. 3. These components are similar to the Stedmon results (Stedmon et al., 2003, 2005). They are compared to the Coble (1996) and Parlanti (Parlanti et al., 2000) peaks in Table 2. Component 1 is composed of two non-separated peaks of different excitation. One with a wavelength domain about 350/400–450 which corresponds to the type C, humic like compound, and one ranging in the domain 275/400–500 (noted A'). This last peak is excited at a similar wavelength as the tyrosine type but with a red-shifted emission range of about 100 nm. Component 2

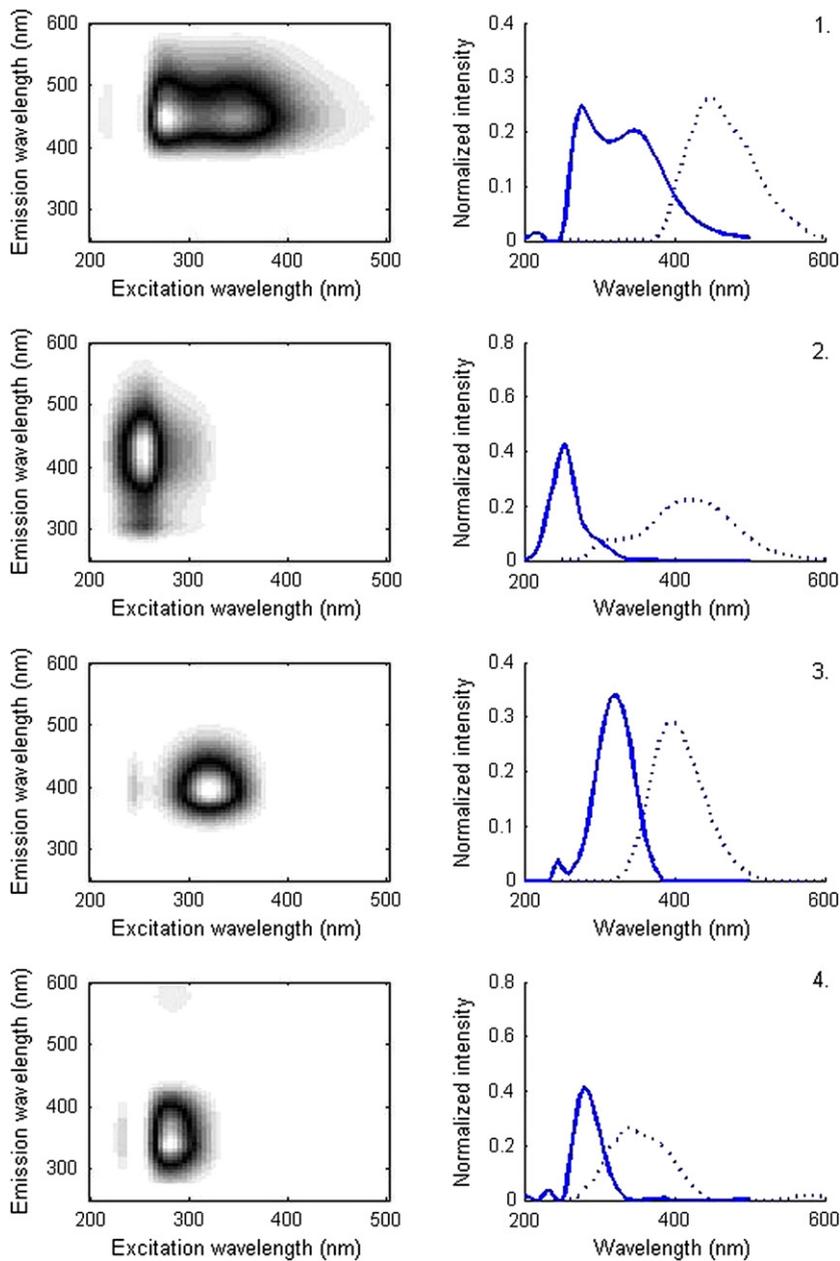


Fig. 3. Spectral representations of the normalized PARAFAC components (1–4): FEEM representation (left) and the corresponding individual excitation (solid lines) and emission (dotted lines) spectra (right).

(255/380–460) is typical of the A type, described by Coble (1996). Component 3 (320/380–420) matches the M type, initially described as marine humic-like matter. Finally, the component 4 is similar to the protein like compounds. It includes the B and T types as the emission domain is large (275/300–360). Note that the component A' is red shifted compared to the excitation wavelength of component 2 (type A).

3.2. FDOM tracing

Five geographical sample classes were chosen a priori, having undergone anthropic influence. These regroup, in ascending order of anthropic-influence, from the bay samples (class 1), presumably less influenced

Table 2

Comparison between the referenced fluorescence peaks and the four PARAFAC components of this study

| Referenced fluorescence peaks | | | | PARAFAC components | | | |
|-------------------------------|--------------|-------------|-------------|-------------------------------|------------------|-------------|-------------|
| Parlanti et al. (2000) | Coble (1996) | Ex max (nm) | Em max (nm) | Coble description | Component number | Ex max (nm) | Em max (nm) |
| α | C | 350 | 420–480 | Humic-like | 1 | 350 | 400–450 |
| | | | | A' | 1 | 275 | 400–500 |
| α' | A | 260 | 380–460 | Humic-like | 2 | 255 | 380–460 |
| β | M | 312 | 380–420 | Marine-humic-like | 3 | 320 | 380–420 |
| γ | T | 275 | 340 | Tryptophan-like, protein-like | 4 | 280 | 300–360 |
| δ | B | 275 | 310 | Tyrosine-like, protein-like | | | |

by the river inputs, the western (class 2a) and eastern (class 2b) coastal samples, the western river samples (class 3a) and the eastern river samples (class 3b). Although the condensed representation of the sample set in the four-dimensional PARAFAC analysis is much more informative than the original FEEM set, some basic statistical treatments of the sample scores are performed in order to achieve the sample classification.

Fig. 4. represents the mean proportions of the four components in each of the five classes. Two main kinds of samples are clearly discriminated. The three marine classes (1, 2a and 2b) show very similar and quite balanced profiles with a dominant contribution of component 2 and 3. On the other hand, component 2 almost disappears in the river classes (3a–b) contributions, where component 1 is highly dominant. The component 1 represents a terrigenous character as its contribution is twice as important in river compared to the marine samples (Fig. 4). This component is composed of two peaks, C and A' (Fig. 5). The type A' fluorophore was poorly described before, but the type C is usually described as a specific terrigenous signal (Coble, 1996; Parlanti et al., 2000). Component 2 is characteristically marine. On another hand, component 3 was initially described as marine in origin by Coble (1996) and Parlanti et al. (2000), but this is not the case here. This component is slightly more important in marine than in the river samples, but is not characteristically marine. The contradiction was already pointed out by Stedmon et al. (2003) when they observed this component in “terrestrially dominated end-member samples”. The terrigenous contribution is less and less important in the marine samples fluorescence but never vanishes, and stays almost as important as the component 2 signal. This means that a major part of the marine FDOM is from terrestrial origin. In this study, only the component 2 seems to be marine in origin. Moreover in marine FDOM, the protein-like signal is less important than in river FDOM, and its contribution to the fluorescence signal increase from coast to bay exit.

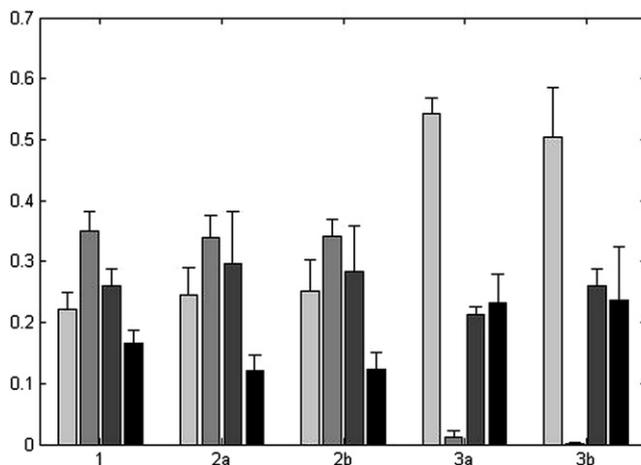


Fig. 4. Relative mean contributions of each PARAFAC component (1–4, from lighter to darker grey) in the five geographical sample classes from open bay (1) to riverine (3a, 3b). Error bars represent the standard deviations.

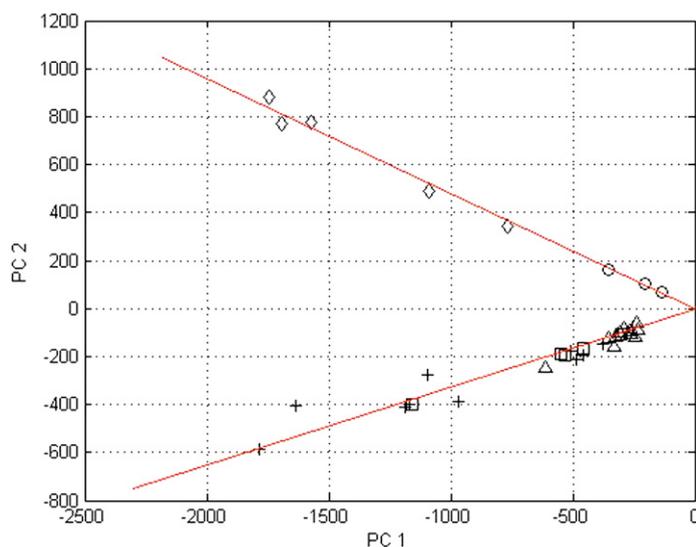


Fig. 5. Sample scores for the first two principal components (respectively 85% and 14% of variance explained) of the PARAFAC score matrix. Δ symbols are used for class 1, \square for class 2a, $+$ for class 2b, \circ for class 3a and \diamond for class 3b.

From a statistical point of view, the PARAFAC analysis should provide some redundant information and a more condensed representation may be beneficial for the classification. Consequently, a PCA of the PARAFAC score matrix is performed as proposed in Ohno and Bro (2006). The principal components are linear combinations of the PARAFAC components and have no physical meaning but the subspace spanned by only the first two principal components explains more than 98% of the samples variance. Therefore, the sample representation in this plan (Fig. 5) should reveal some new information and makes the classification simpler. The PCA confirms the discrimination between river and marine samples as these two kinds of samples are clearly distributed around two different axis. Furthermore, it highlights the degree of difference between samples. In the eastern part of the bay, Piraque, Ita, Guandu, Sao Francisco and Guarda rivers have a strong fluorescence signal compared to the Itimirim and Itinguçu rivers. This is due to the morphologic difference between these two types of catchment area. The urbanization is far more dense in the east. In comparison, Itimirim and Itinguçu are smaller and non urbanized rivers. The marine samples are also discriminated by their signal amplitude. There is a general tendency for the fluorescence signal of the marine samples to gradually decrease from terrestrial to open sea samples. This is confirmed for each of the four components and supports the observations in previous studies (Stedmon et al., 2003; Stedmon and Markager, 2003) as a result of FDOM sinking in estuary zone.

4. Conclusion

The data treatment of a subtropical FEEM series by PARAFAC analysis describes four fluorescent components within the fluorescent dissolved organic matter. One spectral contribution (255/380–460) to the fluorescent organic matter and its evolution in this environment gives a clear distinction between river and marine waters. All the previous fluorescent peaks A, M, C, T and B were found in this study. A new terrestrial fluorescent moiety, the A' (275/400–500) peak, linked to the C type component was present in these subtropical samples. Hence, each component can be described as a contribution to the total luminescence signal. Surprisingly in this study, the component number 3 whose domain corresponds to the M type, is not characteristic of marine waters. Only the A type (humic-like compound) is really a feature of marine samples. This study, using PARAFAC analysis on FEEM samples shows that the fluorescence response the marine FDOM is principally due to terrestrial FDOM except for the 255/380–460 domain. The terrestrial FDOM is diluted in the bay and is in similar proportion to the marine FDOM even for the more seaward-samples. More investigations are required to understand why only the A type component is characteristic of marine waters in these sub-tropical

samples. PARAFAC provides a useful tool to discriminate between different components contributing to the FDOM in the coastal environment.

Acknowledgements

Thanks to Dr. D. Février and Pr. J. Maddock from the federal university of Niteroi for the sampling campaign and on site help. Part of this work was supported by the CAPES-COFECUB fund.

References

- Ahmad, U.K., Ujang, Z., Yusop, Z., Fong, T.L., 2002. Fluorescence technique for the characterization of natural organic matter in river water. *Water Sci. Technol.* 46 (9), 117–125.
- Andersson, C.A., Bro, R., 2002. The N-way toolbox for Matlab. *Chemometr. Intell. Lab. Syst.* 52, 1–4, toolbox available at (<http://www.models.kvl.dk/source/>).
- Baker, A., Spencer, R.G.M., 2004. Characterisation of dissolved organic matter from source to sea using fluorescence and absorbance spectroscopy. *Sci. Total Environ.* 333 (1–3), 217–232.
- Barcellos, C., Lacerda, L.D., 1994. Cadmium and zinc source assessment in the Sepetiba Bay and basin region. *Environ. Monitor. Assess.* 29 (2), 183–199.
- Benaim, J.Y., Mounier, S., 1998. Metal transport by organic carbon in the Amazon Basin. *Croat. Chem. Acta* 71 (2), 405–419.
- Boehme, J., Coble, P., Conmy, R., Stovall-Leonard, A., 2004. Examining CDOM fluorescence variability using principal component analysis: seasonal and regional modeling of three-dimensional fluorescence in the Gulf of Mexico. *Mar. Chem.* 89, 3–14.
- Bro, R., 1997. PARAFAC. Tutorial and applications. *Chemometr. Intell. Lab. Syst.* 38, 149–171.
- Bro, R., Kiers, H.A.L., 2003. A new efficient method for determining the number of components in PARAFAC models. *J. Chemometr.* 17, 274–286.
- Cabaniss, S.E., Shuman, M.S., 1987. Synchronous fluorescence spectra of natural waters: sources of dissolved organic matter. *Mar. Chem.* 21 (1), 37–50.
- Coble, G.P., 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Mar. Chem.* 51, 325–346.
- de la Peña, A.M., Díez, N.M., Gil, D.B., Olivieri, A.C., Escandar, G.M., 2006. Simultaneous determination of flufenamic and meclofenamic acids in human urine samples by second-order multivariate parallel factor analysis (PARAFAC) calibration of micellar-enhanced excitation–emission fluorescence data. *Anal. Chim. Acta* 569 (1–2), 250–259.
- Esteves da Silva, J.C.G., Tavares, M.J.C.G., Tauler, R., 2006. Multivariate curve resolution of multidimensional excitation-emission quenching matrices of a Laurencian soil fulvic, acid. *Chemosphere* 64, 1939–1948.
- Faber, N.M., Bro, R., Hopke, P.K., 2003. Recent developments in CANDECOMP/PARAFAC algorithms: a critical review. *Chemometr. Intell. Lab. Syst.* 65, 119–137.
- Galapate, R.P., Baes, A.U., Ito, K., Shoto, E., Okada, M., 1998. Detection of domestic waste in Kurose river using synchronous fluorescence spectroscopy. *Water Res.* 32 (7), 2232–2239.
- Harshman, R.A., 1970. Foundations of the PARAFAC procedure: models and conditions for an “explanatory” multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16, 1–84. (UMI Serials in Microform, No. 10,085).
- Harshman, R.A., 1972. Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA Working Papers in Phonetics*, 22, 111–117. (UMI Serials in Microform, No. 10,085).
- Harshman, R.A., Lundy, M.E., 1994. PARAFAC: Parallel factor analysis. *Comput. Stat. Data Anal.* 18, 39–72.
- Holbrook, R.D., Yen, J.H., Grizzard, T.J., 2006. Characterizing natural organic material from the Occoquan Watershed (Northern Virginia, US) using fluorescence spectroscopy and PARAFAC. *Sci. Total Environ.* 361 (1–3), 249–266.
- Jaffé, R., Boyer, J.N., Lu, X., Maie, N., Yang, C., Scully, N.M., Mock, S., 2004. Source characterization of dissolved organic matter in a subtropical mangrove-dominated estuary by fluorescence analysis. *Mar. Chem.* 84, 195–210.
- Karez, C.S., Magalhaes, V.F., Pfeiffer, W.C., Amado, G.M., 1987. Trace metal accumulation by algae in Sepetiba Bay, Brazil. *Environ. Pollut.* 83 (3), 351–356.
- Kowalczyk, P., Stoń-Egiert, J., Cooper, W.J., Whitehead, R.F., Durako, M.J., 2005. Characterization of chromophoric dissolved organic matter (CDOM) in the Baltic Sea by excitation emission matrix fluorescence spectroscopy. *Mar. Chem.* 96 (3–4), 273–292.
- Lacerda, L.D., Marins, R.V., Paraquetti, H.H.M., Mounier, S., Benaim, J., Février, D., 2001. Mercury distribution and reactivity in waters of a subtropical coastal lagoon, Sepetiba Bay, SE Brasil. *J. Brazilian Chem. Soc.* 12 (1), 93–98.
- Marins, R.V., Silva Filho, E.V., Lacerda, L.D., 1996. Atmospheric deposition of Mercury over Sepetiba Bay, SE Brazil. *Sociedade Brasileira de Química* 7 (3), 177–180.
- Marins, R.V., Lacerda, L.D., Paraquetti, H.H.M., de Paiva, E.C., Villas, R.C., 1998. Geochemistry of mercury in sediments of a subtropical Coastal Lagoon, Sepetiba Bay, Southeastern Brazil. *Bull. Environ. Contaminat. Toxicol.* 61 (1), 57–64.
- Miano, T.M., Senesi, N., 1992. Synchronous excitation fluorescence spectroscopy applied to soil humic substances chemistry. *Sci. Total Environ.* 117, 41–51.
- Mobed, J.J., Hemmingsen, S.L., Autry, J.L., McGown, L.B., 1996. Fluorescence characterization of IHSS humic substances: total luminescence spectra with absorbance correction. *Environ. Sci. Technol.* 30 (10), 3061–3065.

- Mounier, S., Lacerda, L.D., Marins, R.V., Benaim, J., 2001. Copper and Mercury Complexing Capacity of Organic Matter From a Mangrove Mud Flat Environment, Sepetiba Bay, Brazil. *Environ. Contaminat. Toxicol.* 67, 519–525.
- Ohno, T., Bro, R., 2006. Dissolved Organic Matter Characterization Using Multiway Spectral Decomposition of Fluorescence Landscapes. *Soil Sci. Soc. Am. J.* 70, 2028–2037.
- Ovalle, A.R.C., Rezende, C.E., Lacerda, L.D., Silva, C.A.R., 1990. Factors affecting the hydrochemistry of a mangrove tidal creek, Sepetiba Bay, Brazil. *Estuar. Coast. Shelf Sci.* 31 (5), 639–650.
- Paraquetti, H.H.M., Ayres, G.A., de Almeida, M.D., Molisani, M.M., Lacerda, L.D., 2004. Mercury distribution, speciation and flux in the Sepetiba Bay tributaries. *SE Brazil. Water Res.* 38 (6), 1439–1448.
- Parlanti, E., Wörz, K., Geoffroy, L., Lamotte, M., 2000. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. *Organ. Geochem.* 31 (12), 1765–1781.
- Patel-Sorrentino, N., Mounier, S., Benaim, J., 2002. Excitation-emission fluorescence matrix to study pH influence on organic matter fluorescence in the Amazon basin. *Water Res.* 36 (10), 2571–2581.
- Persson, T., Wedborg, M., 2001. Multivariate evaluation of the fluorescence of aquatic organic matter. *Anal. Chim. Acta* 434, 179–192.
- Riu, J., Bro, R., 2002. Jack-knife technique for outlier detection and estimation of standard errors in PRAFAC models. *Chemometr. Intell. Lab.* 65, 35–49.
- Senesi, N., Miano, T.M., Provenzano, M.R., Brunetti, G., 1991. Characterization, differentiation, and classification of humic substances by fluorescence spectroscopy. *Soil Sci.* 152 (4), 259–271.
- Sidiropoulos, N.D., Bro, R., 2000. On the uniqueness of multilinear decomposition of N-way arrays. *J. Chemometr.* 14 (3), 229–239.
- Stedmon, C.A., Markager, S., Bro, R., 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Mar. Chem.* 82 (3-4), 239–254.
- Stedmon, C.A., Markager, S., 2003. Behaviour of the optical properties of coloured dissolved organic matter under conservative mixing. *Estuar. Coast. Shelf Sci.* 57, 1–7.
- Stedmon, C.A., Markager, S., 2005. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. *Limnol. Oceanogr.* 50 (2), 686–697.
- Trubetskaya, O., Trubetskoj, O., Guyot, G., Andreux, F., Richard, C., 2002. Fluorescence of soil humic acids and their fractions obtained by tandem size exclusion chromatography-polyacrylamide gel electrophoresis. *Organ. Geochem.* 33 (3), 213–220.
- Vasel, J.L., Praet, E., 2002. On the use of fluorescence measurements to characterize wastewater. *Water Sci. Technol.* 45 (4-5), 109–116.
- Zepp, R.G., Sheldon, W.M., Moran, M.A., 2004. Dissolved organic fluorophores in southeastern US coastal waters: correction method for eliminating Rayleigh and Raman scattering peaks in excitation–emission matrices. *Mar. Chem.* 89, 15–36.

Annexe B

Produit de Khatri-Rao

Le produit de Khatri-Rao est un produit tensoriel particulièrement adapté au modèle PARAFAC puisqu'il permet de réécrire le jeu de données sous la forme d'un produit matriciel classique.

B.1 Définitions et propriétés

produit de Kronecker Le produit de Kronecker \otimes entre deux matrices \mathbf{A} de taille (m, n) et \mathbf{B} de taille (p, r) est défini de la manière suivante :

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{bmatrix} \quad (\text{B.1})$$

Le résultat est donc une matrice de taille (mp, nr) .

Le produit de Kronecker est bilinéaire et associatif. On a également les deux propriétés suivantes :

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD} \quad (\text{B.2})$$

et

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \quad (\text{B.3})$$

produit de Khatri-Rao On définit le produit de Khatri-Rao \odot entre deux matrices \mathbf{A} de taille (I, N) et \mathbf{C} de taille (K, N) à partir du produit de Kronecker :

$$\mathbf{A} \odot \mathbf{C} = [\mathbf{a}_{.1} \otimes \mathbf{c}_{.1} \quad \mathbf{a}_{.2} \otimes \mathbf{c}_{.2} \quad \dots \quad \mathbf{a}_{.N} \otimes \mathbf{c}_{.N}] \quad (\text{B.4})$$

où $\mathbf{a}_{.1}$ désigne la première ligne de la matrice \mathbf{A} . Les deux matrices doivent avoir le même nombre de colonnes et le résultat est une matrice de taille (IK, N) .

Remplacer le produit de Kronecker par son expression permet ici une meilleure lecture des choses :

$$\mathbf{A} \odot \mathbf{C} = \begin{bmatrix} a_{11}\mathbf{c}_{.1} & a_{12}\mathbf{c}_{.2} & \dots & a_{1n}\mathbf{c}_{.N} \\ a_{21}\mathbf{c}_{.1} & a_{22}\mathbf{c}_{.2} & \dots & a_{2n}\mathbf{c}_{.N} \\ \vdots & \vdots & & \vdots \\ a_{I1}\mathbf{c}_{.1} & a_{I2}\mathbf{c}_{.2} & \dots & a_{In}\mathbf{c}_{.N} \end{bmatrix} \quad (\text{B.5})$$

Le produit de Khatri-Rao est associatif et distributif. Il possède également la propriété suivante :

$${}^t(\mathbf{A} \odot \mathbf{B})(\mathbf{A} \odot \mathbf{B}) = ({}^t\mathbf{AA}) \bullet ({}^t\mathbf{BB}) \quad (\text{B.6})$$

Où \bullet dénote le produit de Hadamard (produit matriciel terme à terme).

B.2 Lien avec PARAFAC

Le modèle trinéaire de rang N s'écrit :

$$\forall i, j, k \quad x_{i,j,k} = \sum_{n=1}^N a_{in} b_{jn} c_{kn} \quad (\text{B.7})$$

Soit $\underline{\mathbf{X}}$ le cube de données de dimension (I, J, K) correspondant. Celui-ci peut se déplier suivant une direction et se représenter sous forme matricielle comme nous l'avons vu au chapitre 2 : Par exemple :

$$\mathbf{X}_{J,KI} = [\mathbf{X}_1 \quad \dots \quad \mathbf{X}_i \quad \dots \quad \mathbf{X}_I] \quad (\text{B.8})$$

Les matrices \mathbf{X}_i sont de taille (J, K) et contiennent l'évolution des données pour une valeur fixée du paramètre i . Dans notre cas, \mathbf{X}_i est le spectre 3D de l'échantillon i . Nous avons alors écrit que

$$\mathbf{X}_{J,KI} = \mathbf{B}^t (\mathbf{A} \odot \mathbf{C}) \quad (\text{B.9})$$

$$\mathbf{X}_{I,JK} = \mathbf{A}^t (\mathbf{C} \odot \mathbf{B}) \quad (\text{B.10})$$

$$\mathbf{X}_{K,IJ} = \mathbf{C}^t (\mathbf{B} \odot \mathbf{A}) \quad (\text{B.11})$$

Ces relations sont à la base de l'algorithme de résolution. Elle peuvent se démontrer facilement :

Nous nous contenterons de démontrer la première équation, les deux autres devenant ensuite évidentes par symétrie.

Afin de démontrer ce résultat, nous allons simplement chercher à exprimer les termes génériques des matrices $\mathbf{X}_{J,KI}$ et $\mathbf{B}^t (\mathbf{A} \odot \mathbf{C})$.

On pose $P = KI$. On vérifie alors immédiatement que les deux matrices sont bien de même taille (J, P) . Nous noterons j et p les indices dans ces deux directions. Nous devons donc montrer que :

$$\forall \quad 1 < j < J \quad \text{et} \quad 1 < p < P \quad \mathbf{X}_{J,KI}(j, p) = \mathbf{B}^t (\mathbf{A} \odot \mathbf{C})(j, p) \quad (\text{B.12})$$

Commençons par déterminer $\mathbf{X}_{J,KI}(j, p)$ en fonction des éléments de \mathbf{A} , \mathbf{B} et \mathbf{C} . D'après l'équation B.7 on a :

$$\mathbf{X}_i(j, k) = \sum_{n=1}^N a_{in} b_{jn} c_{kn} \quad (\text{B.13})$$

L'élément (j, p) de $\mathbf{X}_{J,KI}$ appartient à l'une des matrices \mathbf{X} . On note u_p l'indice et v_p

la position de $\mathbf{X}_{J,KI}(j, p)$ dans \mathbf{X}_{u_p} . Par construction de $\mathbf{X}_{J,KI}$, nous pouvons exprimer ces deux valeurs en fonction de p :

$$\begin{cases} u_p = E \left[\frac{p-1}{K} \right] + 1 \\ v_p = p - KE \left[\frac{p-1}{K} \right] \end{cases} \quad (\text{B.14})$$

où $E[x]$ désigne la partie entière de x . On a donc finalement :

$$\forall 1 < j < J \quad \text{et} \quad 1 < p < P \quad \mathbf{X}_{J,KI}(j, p) = \mathbf{X}_{u_p}(j, v_p) = \sum_{n=1}^N a_{u_p n} b_{j n} c_{v_p n} \quad (\text{B.15})$$

Raisonnons de la même manière avec $\mathbf{B}^t(\mathbf{A} \odot \mathbf{C})(j, p)$. Soit \mathbf{Z} la matrice de taille (N, P) telle que :

$$\mathbf{Z} = {}^t(\mathbf{A} \odot \mathbf{C}) \quad (\text{B.16})$$

D'après B.5, $\mathbf{Z}(n, p)$ peut s'écrire comme le produit d'un élément de la ligne n de \mathbf{A} par un élément de la ligne n de \mathbf{C} . Là encore, par construction, les indices correspondant à ces deux éléments peuvent s'exprimer en fonction de p . On retrouve d'ailleurs :

$$\forall 1 < n < N \quad \text{et} \quad 1 < p < P \quad \mathbf{Z}(n, p) = \mathbf{A}(u_p, n) \mathbf{C}(v_p, n) \quad (\text{B.17})$$

Après multiplication par \mathbf{B} il vient :

$$\forall 1 < j < J \quad \text{et} \quad 1 < p < P \quad \mathbf{BZ}(j, p) = \sum_{n=1}^N b_{j n} z_{n p} = \sum_{n=1}^N b_{j n} a_{u_p n} c_{v_p n} \quad (\text{B.18})$$

On déduit alors immédiatement B.12 de B.15 et B.18.

Bibliographie

- [1] *Fluorescence and Phosphorescence Analysis. Principles and Applications*. New York, Interscience Publishers, 1966.
- [2] *Generalized inverse of matrices and its applications*. John Wiley Sons, 1972.
- [3] *Luminescence spectrometry in analytical chemistry*, vol. 38, p. 293–294. John Wiley Sons, 1972.
- [4] AHMAD (U.), UJANG (Z.), YUSOP (Z.) et FONG (T.), « Fluorescence technique for the characterization of natural organic matter in river water. », *Water Science and Technology*, vol. 46, n° 9, p. 117–125, 2002.
- [5] ANDERSEN (C.) et BRO (R.), « Parctical aspects of PARAFAC modeling of fluorescence excitation-emission data », *Journal of chemometrics*, vol. 17, p. 200–215, 2003.
- [6] ANDERSON (G. G.), BRIAN (K. D.) et BOOKSH (K. S.), « Weighted parallel factor analysis for calibration of HPLC-UV / Vis spectrometers in the presence of beer's law deviations », *Chemometrics and intelligent laboratory systems*, vol. 49, p. 195–213, 1999.
- [7] ANDRADE (J. M.), GOMEZ-CARRACEDO (M. P.), KRZANOWSKI (W.) et KUBISTA (M.), « Procruste rotation in analytical chemistry, a tutorial », *Chemometrics and intelligent laboratory systems*, vol. 72, p. 123–132, 2004.
- [8] ANDRADE-EIROA (A.), VAZQUEZ-BLANCO (E.), LOPEZ-MAHIA (P.) et MUNIATEGUI-LORENZO (S.), « Modeling of inner filter effect in synchronous spectrofluorimetry by using partial least squares », *Analysis*, vol. 28, p. 148–154, 2000.
- [9] BAKER (A.) et SPENCER (R.), « Characterisation of dissolved organic matter from source to sea using fluorescence and absorbance spectroscopy », *Science of the Total Environment*, vol. 333, n° 1-3, p. 217–232, 2004.
- [10] BECQUIREL (E.), *Annales de chimie et physique*, vol. 9, p. 257–322, 1842.
- [11] BOEHME (J.), COBLE (P.), CONMY (R.) et STOVALL-LEONARD (A.), « Examining CDOM fluorescence variability using principal component analysis : seasonal and regional modeling of three-dimensional fluorescence in the Gulf of Mexico », *Marine chemistry*, vol. 89, p. 3–14, 2004.
- [12] BOOKSH (K. S.) et KOWALSKI (B. R.), « Calibration method choice by comparison of model basis functions to the theoretical instrumental response function », *Analytica chimica acta*, vol. 348, p. 1–9, 1997.

- [13] BRO (R.), « PARAFAC. tutorial and applications », *Chemometrics and intelligent laboratory systems*, 1997.
- [14] BRO (R.). *Multi-way analysis in the food industry*. PhD thesis, Royal Veterinary and Agricultural University, 1998.
- [15] BRO (R.) et KIERS (H. A.), « A new efficient method for determining the number of components in PARAFAC models », *Journal of chemometrics*, vol. 17, p. 274–286, 2003.
- [16] CHEN (Z.-P.), LIU (Z.), CAO (Y.-Z.) et YU (R.-Q.), « Efficient way to estimate the optimum number of factors for trilinear decomposition », *Analytica chimica acta*, vol. 444, p. 295–307, 2001.
- [17] CHEN (Z.) et R.Q. (Y.), « Mitigating model deficiency in three-way data analysis by the combination of background constraining and iterative correcting techniques », *Analytica chimica acta*, vol. 487, n° 2, p. 171–180, 2003.
- [18] CHIANG (L. H.), LEARDI (R.), PELL (R. J.) et SEASHOLTZ (M. B.), « Industrial experiences with multivariate statistical analysis of batch process data », *Chemometrics and intelligent laboratory systems*, vol. 81, p. 109–119, 2006.
- [19] COBLE (G.), « Characterization of marine and terrestrial dom in seawater using excitationemission matrix spectroscopy », *Marine chemistry*, vol. 51, p. 325–346, 1996.
- [20] CREDI (A.) et LUCA (P.), « From observed to corrected luminescence intensity of solution systems : an easy-to-apply correction method for standard spectrofluorimeters », *Spectrochimica Acta Part A*, vol. 54, p. 159–170, 1998.
- [21] CUESTA SANCHEZ (F.), RUTAN (S.), GIL GARCIA (M.) et MASSART (D.), « Resolution of multicomponent overlapped peaks by the orthogonal projection approach, evolving factor analysis and window factor analysis », *Chemometrics and intelligent laboratory systems*, vol. 36, p. 153–164, 1997.
- [22] CUESTA SANCHEZ (F.), TOFT (J.), VAN DEN BOGAERT (B.) et MASSART (D.), « Orthogonal projection approach applied to peak purity assessment », *Analytical Chemistry*, vol. 68, n° 1, p. 79–85, 1996.
- [23] DE JUAN (A.) et TAULER (R.), « Comparison of three-way resolution methods for non-trilinear chemical data sets », *Journal of chemometrics*, vol. 15, p. 749–772, 2001.
- [24] DORRIT (B.). « Factors affecting 3-way modelling (PARAFAC) of fluorescence landscapes ». « Factors affecting 3-way modelling (PARAFAC) of fluorescence landscapes », 1999.
- [25] FABER (N. M.), BRO (R.) et HOPKE (P. K.), « Recent developments in CANDECOMP/PARAFAC algorithms : a critical review », *Chemometrics and intelligent laboratory systems*, vol. 65, p. 119–137, 2003.
- [26] FABER (N.), BUYDENS (L.) et KATEMAN (G.), « Generalized rank annihilation method. 1 : derivation of eigenvalue problems », *Journal of chemometrics*, 1994.
- [27] GALAPATE (R.), BAES (A.), ITO (K.) *et al.*, « Detection of domestic waste in kurose river using synchronous fluorescence spectroscopy », *Water Research*, vol. 32, n° 7, p. 2232–2239, 1998.

- [28] GARRIDO FRENICH (A.), PICON ZAMORA (D.), MARTINEZ VIDAL (J.) et MARTINEZ GALERA (M.), « Resolution (and quantitation) of mixtures with overlapped spectra by orthogonal projection approach and alternating least squares », *Analytica Chimica Acta*, vol. 449, p. 143–155, 2001.
- [29] GUIMET (F.), FERRÉ (J.), BOQUÉ (R.) et RIUS (F. X.), « Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy », *Analytica Chimica Acta*, vol. 515, n° 75-85, 2004.
- [30] HARSHMAN (R.), « Foundations of the parafac procedure : Model and conditions for an "explanatory multi-mode factor analysis" », *UCLA Working Papers in phonetics*, vol. 16, 1970.
- [31] HARSHMAN (R.), « Determination and proof of minimum uniqueness conditions for PARAFAC1 », *UCLA Working Papers in phonetics*, vol. 22, 1972.
- [32] HARSHMAN (R. A.) et DE SARBO (W. S.), « An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques ».
- [33] HARSHMAN (R. A.) et LUNDY (M. E.), « The PARAFAC model for three-way factor analysis and multidimensional scaling ».
- [34] HARSHMAN (R. A.) et LUNDY (M. E.), « PARAFAC :Parallel factor analysis », *Computational statistics Data analysis*, vol. 18, p. 39–72, 1994.
- [35] HENDERSON (G.), « The effects of absorption and self-absorption quenching on fluorescent intensities », *Journal of chemical education*, vol. 54, p. 57–59, 1977.
- [36] HOLBROOK (R. D.), DEROSE (P. C.), LEIGH (S. D.) *et al.*, « Excitation-emission matrix fluorescence spectroscopy for natural organic matter characterization : A quantitative evaluation of calibration and spectral correction procedures », *Applied spectroscopy*, vol. 60, n° 7, 2006.
- [37] HOLBROOK (R. D.), YEN (J. H.) et GRIZZARD (T. J.), « Characterizing natural organic material from the occoquan watershed (northern virginia, us) using fluorescence spectroscopy and PARAFAC », *Science of the total environment*, vol. 361, n° 249-266, 2006.
- [38] HOLLAND (J. F.), TEETS (R. E.), KELLY (P. M.) et TIMNICK (A.), « Correction of right-angle fluorescence measurements for the absorption of excitation radiation », *Analytical chemistry*, 1977.
- [39] HOLSCHNEIDER (M.), KRONLAND-MARTINET (R.), MORLET (J.) et TCHAMITCHIAN (P.), « A real-time algorithm for signal analysis with the help of the wavelet transform », *Wavelets, Time-Frequency Methods and Phase Space*, p. 289–297, 1989.
- [40] HU (L.-Q.), WU (H.-L.), JIANG (J.-H.) *et al.*, « Estimating the chemical rank of three-way data arrays by a simple linear transform incorporating monte carlo simulation », *Talanta*, 2006.
- [41] JIANG (J.-H.), WU (H.-L.), CHEN (Z.-P.) et YU (R.-Q.), « Coupled vectors resolution method for chemometric calibration with three-way data », *nalytical Chemistry*, vol. 71, p. 4254–4262, 1999.

- [42] JIJI (R. D.), ANDERSON (G. G.) et BOOKSH (K. S.), « Application of PARAFAC for calibration with excitation-emission matrix fluorescence spectra of three classes of environmental pollutants », *Journal of chemometrics*, vol. 14, p. 171–185, 2000.
- [43] KARJALAINEN (E.), « The spectrum reconstruction problem : Use of alternating regression for unexpected spectral components in two-dimensional spectroscopies », *Chemometrics and Intelligent Laboratory Systems*, vol. 7, p. 31–38, 1989.
- [44] KELLER (H.) et MASSART (D.), « Evolving factor analysis », *Chemometrics and Intelligent Laboratory Systems*, vol. 12, n° 3, p. 209–224, 1992.
- [45] KRUSKAL (J.), « Three-way arrays : Rank and uniqueness of trilinear decomposition, with application to arithmetic complexity and statistics », *Linear algebra and its applications*, vol. 18, p. 95, 1977.
- [46] LI (H.), HOU (J.), WANG (K.) et ZHANG (F.), « Resolution of multicomponent overlapped peaks. a comparison of several curve resolution methods », *Talanta*, vol. 70, p. 336–343.
- [47] LORBER (A.), « Quantifying chemical composition from two-dimensional data arrays », *Analytica chimica acta*, vol. 164, p. 293–297, 1984.
- [48] LU (J.-Z.), WU (H.-L.), JIANG (J.-H.) *et al.*, « An improved trilinear decomposition algorithm based on a lagrange operator », *Analytical sciences*, vol. 19, p. 1037–1043, 2003.
- [49] MACDONALD (B. C.), LVIN (S. J.) et PATTERSON (H.), « Correction of fluorescence inner filter effects and the partitioning of pyrene to dissolved organic carbon », *Analytica chimica acta*, vol. 338, p. 155–162, 1997.
- [50] MACKNIGHT (D. M.), BOYER (E. W.), WESTERHOFF (P. K.) *et al.*, « Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity », *Limnol. Oceanogr*, vol. 46, n° 1, p. 38–48, 2001.
- [51] MALINOWSKI (E.), « Window factor analysis : theoretical derivation and application to flow injection analysis data », *Journal of chemometrics*, vol. 6, p. 29–40, 1992.
- [52] MARINS (R.), SILVA FILHO (E.) et LACERDA (L.), « Atmospheric deposition of mercury over sepetiba bay, southeastern brazil. », *Journal of the Brazilian Chemical Society*, vol. 7, n° 3, p. 177–180, 1996.
- [53] MIANO (T.) et SENESI (N.), « Synchronous excitation fluorescence spectroscopy applied to soil humic substances chemistry. », *The Science of the Total Environment*, vol. 117, p. 41–51, 117.
- [54] MOBED (j. J.), HEMMINGSEN (S. L.), AUTRY (J. L.) et MACGOWN (L. B.), « Fluorescence characterization of IHSS humic substances : total luminescence spectra with absorbance correction », *Environmental science technology*, vol. 30, n° 10, p. 3061–3065, 1996.

- [55] MOUNIER (S.), LACERDA (L.), MARINS (R.) et BENAÏM (J.), « Copper and mercury complexing capacity of organic matter from a mangrove mud flat environment, sepetiba bay, brazil. », *Environmental Contamination and Toxicology*, vol. 67, p. 519–525, 2001.
- [56] NELDER (J. A.) et MEAD (R.), « A simplex method for function minimization », *Computer Journal*, vol. 7, p. 308–313, 1965.
- [57] NIKOLAJSEN (R. P.), BOOKSH (K. S.), HANSEN (A. M.) et BRO (R.), « Quantifying catecholamines using multi-way kinetic modelling », *analytica chimica acta*, vol. 475, p. 137–150, 2003.
- [58] OHNO (T.), « Fluorescence inner-filtering correction for determining the humification index of dissolved organic matter », *Environmental science technology*, vol. 36, n° 4, p. 742–746, 2002.
- [59] OHNO (t.) et BRO (R.), « Dissolved organic matter characterization using multiway spectral decomposition of fluorescence landscapes », *Soil science*, 2006.
- [60] PARKER (C.) et BARNES (W.), « Some experiments with spectrofluorimeters and filter fluorimeters », *The analyst*, vol. 82, p. 606–618, 1957.
- [61] PERSSON (T.) et WEDBORG (M.), « Multivariate evaluation of the fluorescence of aquatic organic matter », *Analytica chimica acta*, 2001.
- [62] POMPA (P.), CICCARELLA (G.), SPADAVECCHIA (J.) *et al.*, « Spectroscopic investigation of inner filter effects by phthalocyanine solutions », *Journal of photochemistry and photobiology A : Chemistry*, 2003.
- [63] REIS (M. M.), BILOTI (D. N.), FERREIRA (M. M.) *et al.*, « PARAFAC for spectral curve resolution : a case study using total luminescence in human dental tartar », *applied spectroscopy*, vol. 55, n° 7, p. 847–851, 2001.
- [64] RIESZ (J.), GILMORE (J.) et MEREDITH (P.), « Quantitative photoluminescence of broad band absorbing melanins : a procedure to correct for inner filter and re-absorption effects », *Spectrochimica acta part A*, 2004.
- [65] RINNAN (A.) et ANDERSEN (C. M.), « Handling of first-order Rayleigh scatter in parafac modelling of fluorescence excitation-emission data », *Chemometrics and intelligent laboratory systems*, 2005.
- [66] RIU (J.) et BRO (R.), « Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models », *Chemometrics and intelligent laboratory systems*, 2003.
- [67] RODRIGUEZ-CUESTA (M.), BOQUÉ (R.), RIUS (F.) *et al.*, « Determination of carbendazim, fuberidazole and thiabendazole by three-dimensional excitation-emission matrix fluorescence and parallel factor analysis », *Analytica chimica acta*, vol. 491, n° 47-56, 2003.
- [68] SANCHEZ (E.) et KOWALSKI (B.), « Generalized rank annihilation factor analysis », *Analytical chemistry*, vol. 58, n° 2, p. 496–499, 1986.
- [69] SANCHEZ (E.) et KOWALSKI (B.), « Tensorial resolution : A direct trilinear decomposition », *Journal of Chemometrics*, vol. 4, n° 1, 1990.

- [70] SARABIA (L.), ORTIZ (M.), LEARDI (R.) et DRAVA (G.), « A program for non-ort(ref) hogonal rotation in factor analysis », *Trends in analytical chemistry*, 1993.
- [71] SCARMINIO (I.) et KUBISTA (M.), « Analysis of correlated spectral data », *Analytical chemistry*, vol. 65, n° 4, p. 409–416, 1993.
- [72] SEQUARIS (J.-M.) et KOGLIN (E.), « Evolving factor analysis for the resolution of overlapping chromatographic peaks », *Analytical chemistry*, 1987.
- [73] STEDMON (C.), MARKAGER (S.) et BRO (R.), « Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. », *Marine Chemistry*, vol. 82, n° 3-4, p. 239–254, 2003.
- [74] STEDMON (C.) et MARKAGER (S.), « Behaviour of the optical properties of coloured dissolved organic matter under conservative mixing », *Estuarine Coastal and Shelf Science*, vol. 57, p. 973–979, 2003.
- [75] STEDMON (C.) et MARKAGER (S.), « Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using parafac analysis », *Limnology and Oceanography*, vol. 50, n° 2, 2005.
- [76] STOKES (G.), « On the refrangibility of light », *Philosophical transactions*, vol. 142, p. 463–562, 1852.
- [77] TREVISAN (M. G.) et POPPI (R. J.), « Determination of doxorubicin in human plasma by excitation-emission matrix fluorescence and multi-way analysis », *analytica chimica acta*, vol. 493, p. 69–81, 2003.
- [78] TUCKER (S. A.), AMSZI (V. L.) et E. (A. W.), « Primary and secondary inner filtering », *Journal of chemical education*, vol. 69, n° 1, 1992.
- [79] WANG (Z.-G.), JIANG (J.-H.), DING (Y.-J.) *et al.*, « Trilinear evolving factor analysis for the resolution of three-way multi-component chromatograms », *Analytica chimica acta*, 2006.
- [80] WINDIG (W.) et HECKLER (C.), « Self-modeling mixture analysis of categorized pyrolysis mass spectral data with the simplisma approach », *Chemometrics and Intelligent Laboratory Systems*, vol. 14, p. 195–207, 1992.
- [81] YAPPERT (C. M.) et INGLE (J.), « Correction of polychromatic luminescence signals for inner-filter effects », *Applied spectroscopy*, vol. 43, p. 759–767, 1989.
- [82] ZEPP (R. G.), SHELDON (W. M.) et MORAN (M. A.), « Dissolved organic fluorophores in southeastern us coastal waters : correction method for eliminating Rayleigh and Raman scattering peaks in excitation-emission matrices », *Marine chemistry*, vol. 89, p. 15–36, 2003.

Table des figures

| | | |
|------|---|----|
| 1.1 | Diagramme de Jablonski | 15 |
| 1.2 | Principe de mesure du spectre d'absorption | 16 |
| 1.3 | Mesures de la lumière transmise pour une solution de sulfate de quinine très concentrée (109 ppm) et pour le solvant seul | 18 |
| 1.4 | Spectre d'absorption du sulphate de quinine | 19 |
| 1.5 | Diagramme de Jablonski complet | 20 |
| 1.6 | Schema de principe du spectrofluorimètre | 22 |
| 1.7 | Exemple de MEEF de matière organique | 25 |
| 1.8 | Impacte de la diffusion sur une MEFF | 28 |
| 1.9 | Exemple de diffusion 2D | 29 |
| 1.10 | Evolution du maximum du spectre d'émission mesuré pour différentes solution de fluorescéine graduellement concentrées | 31 |
| 1.11 | Modélisation linéaire du spectre d'émission d'un mélange concentré | 31 |
| 1.12 | Influence de la concentration sur le spectre 3D du sulfate de Quinine | 32 |
| 1.13 | Déviation relative entre l'intensité absorbée et son approximation linéaire en fonction de l'absorbance | 32 |
| 1.14 | Spectres du sulfate de quinine et de la fluorescéine : mise en evidence du chevauchement spectral | 34 |
| 2.1 | Analyse en composantes principales d'une matrice d'excitation-émission | 40 |
| 2.2 | Courbes de dissemblance d'un mélange synthétique de trois fluorophores | 43 |
| 2.3 | Courbes EFA, analyse allée | 45 |
| 2.4 | Courbes EFA, analyse retour | 46 |
| 2.5 | Comparaison entre les spectres d'excitation des fluorophores composants le mélange et les courbes EFA croisées. | 47 |
| 2.6 | Exemple de représentation matricielle du cube de données | 54 |
| 3.1 | Résultats de PARAFAC sur les solutions les plus concentrées | 63 |
| 3.2 | Représentation en perspective du dispositif de mesure | 65 |
| 3.3 | Schéma de principe du modèle | 65 |
| 3.4 | Zoom sur la zone d'émission Z_e | 66 |
| 3.5 | Modélisation de l'effet d'écran sur la fluorescéine seule | 68 |
| 3.6 | Mélange à deux composants, comparaison entre les MEEF mesurées et modélisées | 70 |

| | | |
|------|--|-----|
| 3.7 | Mélange à trois composants : comparaison entre les MEEF mesurées et modélisées | 72 |
| 4.1 | Jeu 1 solution 1 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées | 80 |
| 4.2 | Jeu 1, solution 2 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées | 81 |
| 4.3 | Jeu 1, solution 3 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées | 82 |
| 4.4 | Jeu 1, solution 4 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées | 83 |
| 4.5 | Jeu 1, solution 5 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées | 84 |
| 4.6 | Jeu 1, solution 6 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées | 85 |
| 4.7 | Jeu 1, solution 7 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées | 86 |
| 4.8 | Jeu 1 : Comparaison des résultats de PARAFAC pour le mode des concentrations relatives | 87 |
| 4.9 | Jeu 1 : Comparaison des résultats de PARAFAC pour le mode des excitations | 88 |
| 4.10 | Jeu 1 : Comparaison des résultats de PARAFAC pour le mode des émissions | 88 |
| 4.11 | Série 2, solution 1 : Comparaison entre la MEEFde référence, la MEEF originale et les MEEF corrigées | 93 |
| 4.12 | Jeu 2, solution 2 : comparaison entre la MEEFde référence, la MEEF originale et la MEEF corrigée | 94 |
| 4.13 | Jeu 2, solution 3 : comparaison entre la MEEFde référence, la MEEF originale et la MEEF corrigée | 95 |
| 4.14 | Jeu 2, solution 4 : comparaison entre la MEEFde référence, la MEEF originale et la MEEF corrigée | 96 |
| 4.15 | Jeu 2 : Comparaison des résultats de PARAFAC pour le mode des concentrations relatives | 97 |
| 4.16 | Jeu 2 : Comparaison des résultats de PARAFAC pour le mode des excitations | 97 |
| 4.17 | Jeu 2 : Comparaison des résultats de PARAFAC pour le mode des émissions | 98 |
| 4.18 | Jeu 3, échantillon 1 : Comparaison entre la MEEF de référence, la MEEF originale et les MEEF corrigées | 101 |
| 4.19 | Jeu 3, échantillon 2 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée | 102 |
| 4.20 | Jeu 3, échantillon 3 : comparaison entre la MEEFde référence, la MEEF originale et la MEEF corrigée | 103 |
| 4.21 | Jeu 3, échantillon 4 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée | 104 |

| | | |
|------|---|-----|
| 4.22 | Jeu 3, échantillon 5 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée | 105 |
| 4.23 | Jeu 3, échantillon 6 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée | 106 |
| 4.24 | Jeu 3, échantillon 7 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée | 107 |
| 4.25 | Jeu 3, échantillon 8 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée | 108 |
| 4.26 | Jeu 3, échantillon 9 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée | 109 |
| 4.27 | Jeu 3, échantillon 10 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée | 110 |
| 4.28 | Jeu 3, échantillon 11 : Comparaison entre la MEEF de référence, la MEEF originale et la MEEF corrigée | 111 |
| 4.29 | Jeu 3 : Comparaison des résultats de PARAFAC pour le mode des concentrations relatives | 113 |
| 4.30 | Jeu 3 : Comparaison des résultats de PARAFAC pour le mode des excitations | 113 |
| 4.31 | Jeu 3 : Comparaison des résultats de PARAFAC pour le mode des émissions | 114 |
| 4.32 | Spectres 2D utilisés pour la réalisation des mélanges synthétiques à 3 composants | 121 |
| 4.33 | Base estimée de E sur les mélanges synthétiques à deux composants | 122 |
| 4.34 | Mélanges synthétiques à deux composants : linéarisation de la première image | 123 |
| 4.35 | Mélanges synthétiques à deux composants : linéarisation de la seconde image | 123 |
| 4.36 | Mélanges synthétiques à deux composants : linéarisation de la troisième image | 124 |
| 4.37 | Base estimée de E sur les mélanges synthétiques à trois composants | 125 |
| 4.38 | Mélanges synthétiques à trois composants : linéarisation de la première image | 126 |
| 4.39 | Mélanges synthétiques à trois composants : linéarisation de la seconde image | 126 |
| 4.40 | Mélanges synthétiques à trois composants : linéarisation de la troisième image | 127 |
| 4.41 | Base estimée de E sur les mélanges réels à deux composants | 128 |
| 4.42 | Linéarisation de spectres d'excitation d'un mélange réel à deux composants | 129 |
| 5.1 | Résultat de la décomposition sur les MEEF corrigées | 134 |
| 5.2 | Résultats de la décomposition sur les MEEF non corrigées | 135 |
| 5.3 | Moyennes des profils de concentration des trois composants sur les acides humiques et fulviques | 137 |
| 5.4 | Analyse en composantes principales des concentrations relatives | 138 |

5.5 Répartition des différents types d'échantillons dans l'espace des deux premières composantes principales 139

Résumé

Les techniques d'analyse multicomposantes, permettant de décomposer un signal en différentes composantes élémentaires, jouent un rôle important dans divers domaines scientifiques. En chimie analytique, combinées à la spectroscopie de fluorescence, il s'agit d'outils efficaces pour l'analyse de mélanges ou de composés organiques. En effet, un mélange de composants fluorescents peut être caractérisé par sa Matrice d'Excitation-Emission de Fluorescence (MEEF ou spectre 3D). On cherchera alors à déterminer les caractéristiques spectrales et les concentrations relatives de chaque constituant du mélange à partir de sa MEEF ou d'une série de MEEF du même mélange pour des rapports de concentration différents. Dans un premier temps, il conviendra de connaître précisément la nature des signaux physiques sur lesquels ces méthodes s'appliquent. Nous décrirons donc le phénomène de fluorescence, ces origines, les techniques de mesure utilisées pour l'acquisition des signaux (spectrofluorimètre en absorption, en émission), ainsi que les différentes lois physiques utiles à la modélisation du problème. Enfin, à partir de la loi de Beer-Lambert, nous insisterons particulièrement sur le modèle trilineaire classique du spectre de fluorescence 3D et ses limites. Nous ferons ensuite le point sur les méthodes d'analyses linéaires et multilinéaires utilisées en spectroscopie de fluorescence dans le cas de mélanges ou de composés organiques. Nous décrirons en particulier l'algorithme de décomposition trilineaire, PARAFAC qui fait aujourd'hui référence. Cependant, l'application de ces méthodes est restreinte aux solutions pour lesquelles l'approximation linéaire de la loi de Beer-Lambert est valable. C'est à dire les solutions peu concentrées ou tout du moins, celles dont l'absorbance reste faible. Dans le cas contraire, relativement fréquent, le chevauchement des différents spectres d'excitation et d'émission, crée des interactions entre les composants (effets d'écran) qui ne sont pas prises en compte par cette approximation. Nous illustrerons alors sur des données réelles, les limites intrinsèques de PARAFAC pour l'analyse de telles solutions. La technique usuelle de correction de l'effet d'écran est basée sur un vieux modèle, que nous redémontrons, et nécessite la mesure de l'absorbance des solutions. Ceci est, contraignant, source d'erreurs, voir dans certains cas impossible. Une des principales avancées de cette thèse concernera donc la mise en oeuvre d'un protocole de correction simple des effets d'écrans ne nécessitant pas la mesure de l'absorbance mais simplement une seconde mesure du spectre 3D. Nous proposerons également d'autres approches purement numériques. Notre propos sera illustré sur des mélanges synthétiques, des mélanges réalisés en laboratoire, et des échantillons de matière organique (MO) très concentrés. Enfin, nous donnerons deux exemples d'application des différentes techniques exposées au suivi de la matière organique dans l'environnement :

1. L'analyse d'échantillons de MO prélevés dans des sols. Les signaux de fluorescence de ces échantillons présentent un fort effet d'écran que nous corrigerons avant de les décomposer à l'aide de PARAFAC. Cette application met en avant l'importance et l'efficacité de la correction de l'effet d'écran dans un cas concret.
2. L'analyse d'échantillons de MO provenant de la baie de Sépétiba (Brésil). Nous insisterons plus ici sur la technique de suivi de la matière organique à l'aide de PARAFAC.