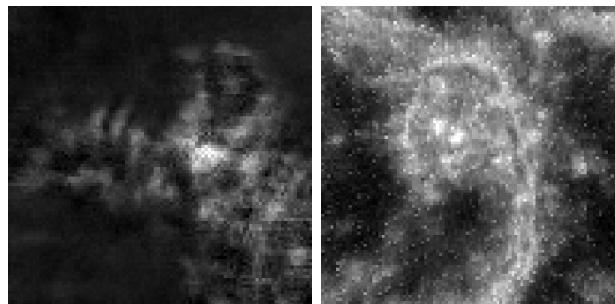


Clara CARLIER
M1 Mathématiques Appliquées
Université Paris-Sud, *Orsay*

Débruitage pour l'imagerie hyperspectrale à l'échelle atomique

Rapport de stage

du 13 Mai 2019 au 30 Août 2019



Tutrice : Nathalie BRUN
Équipe STEM
Laboratoire de Physique des Solides
CNRS, *Orsay*

Avec l'aide de : Thomas OBERLIN
Équipe Signal et Communications
Institut de Recherche en Informatique
Toulouse

Remerciements

Je tiens tout d'abord à remercier Nathalie BRUN qui m'a encadré durant ce stage. Elle a été très à l'écoute et attentive tout du long, ce qui m'a permis d'évoluer aisément dans un milieu qui ne m'était pas familier, que ce soit en physique ou même en traitement d'images. Je la remercie également de m'avoir encouragée à être curieuse et tester les idées que j'avais, un trait qui me semble être primordial dans le milieu de la recherche.

Je remercie également les "toulousains" qui ont apporté leur expertise et leurs nombreuses connaissances sur le débruitage d'images. Je remercie Thomas OBERLIN qui, par sa grande expérience, a su me conseiller et m'orienter dans mes tests et mes essais. Je remercie Étienne MONIER pour sa disponibilité et son accueil chaleureux lors de mon séjour de deux semaines à Toulouse.

Je tiens à remercier chaleureusement l'ensemble des membres du groupe STEM pour cette bonne humeur permanente qui a rendu mon stage des plus agréables et très enrichissant. Je remercie Odile STÉPHAN pour m'avoir accueilli dans son équipe. Je remercie Laura BOCHER, Luiz TIZEI, Michael WALLS, Marta DE FRUTOS, Adrien TEURTRIE et Mathieu KOCIAK pour le temps qu'ils ont pris à m'expliquer le fonctionnement des microscopes ainsi que les recherches qu'ils étaient en train de faire.

Enfin, je remercie Christian COLLIEX qui m'a accueilli chaleureusement dans son bureau et avec qui chaque conversation fut une source de connaissances précieuses.

Table des matières

Glossaire	1
Introduction	3
1 Contexte du stage et présentation du sujet	4
1.1 Présentation du laboratoire	4
1.2 Les microscopes électroniques de l'équipe STEM	5
1.3 L'imagerie hyperspectrale	7
1.4 Les différentes natures de bruit	8
1.5 État de l'art	8
1.5.1 Débruitage standard	8
1.5.2 Débruitage par sous-espace	9
1.5.3 Méthodes séquentielles et jointes	9
2 Modèles et pré-requis	11
2.1 Modèle du bruit	11
2.2 Méthodes utilisées	12
2.2.1 Analyse en composantes principales	12
2.2.2 BM3D	15
2.2.3 Estimation de l'écart-type du bruit	17
2.2.4 Transformée d'Anscombe	17
3 Implémentation de la méthode séquentielle A-BM3D	19
4 Protocoles expérimentaux	21
4.1 Création d'images synthétiques	21
4.2 Calculs d'erreurs	24
4.3 Validation de l'algorithme	25
4.3.1 Simulation de bruit gaussien additif	25
4.3.2 Simulation de bruit poissonnien	28
4.3.3 Inefficacité de la transformée d'Anscombe classique et généralisée	30

5 Résultats sur les données réelles	32
5.1 Construction de carte chimique	32
5.2 Échantillon biologique	35
5.3 Échantillon SmCeO ₂	42
Conclusion	49
Table des figures	52
Bibliographie	54
A Analyse en composantes principales normée	55
B Tableaux d'erreurs calculées sur les données synthétiques	56
C Simulation de bruit mixte	59
D Autre exemple de débruitage sur les données réelles	62

Glossaire

Abréviations

ACP	Analyse en Composantes Principales
BM3D	Block Matching and 3D filtering : algorithme de débruitage spatial
CCD	Charge-Coupled Device : type de caméra
CNRS	Centre National de la Recherche Scientifique
DM	Digital Micrograph : logiciel de contrôle et de traitement des données des microscopes
EELS	Electron Energy Loss Spectroscopy = spectroscopie par perte d'énergie des électrons
EQM	Erreur Quadratique Moyenne
IHS	Image(s) HyperSpectrale(s)
LPS	Laboratoire de Physique des Solides
MAD	Median Absolute Deviation
SNR	Signal-Noise Ratio
Spim	Spectre Image = hypercube = IHS en spectromicroscopie
SSIM	Structural Similarity Index
STEM	Scanning Transmission Electron Microscope
TEM	Transmission Electron Microscope
VST	Variance-Stabilizing Transformation

Notations

- ▶ Dimensions hypercube :
 - L : nombre de lignes
 - C : nombre de colonnes
 - $P = L \times C$: nombre de pixels
 - B : nombre de bandes
- ▶ Pour A matrice de taille $P \times B$:
 - A_p : spectre de la p -ième ligne
 - A^b : image vectorisée contenue dans la b -ième colonne
 - A_p^b : la valeur de p -ième ligne et b -ième colonne
 - \bar{A}_b : moyenne sur les colonnes
- ▶ Données du problème :
 - X : matrice des données non bruitées
 - Y : matrice des données bruitées
 - ε : bruit blanc gaussien
- ▶ Paramètres ACP :
 - K : nombre de composantes sélectionnées
 - **seuil** : valeur minimale de sélection de composantes sur le pourcentage de variance
- ▶ Paramètres BM3D :
 - **sigma** : estimation de l'écart-type du bruit gaussien
 - $r = \text{ratio}$: valeur par laquelle **sigma** est multiplié
- ▶ Paramètres transformée d'Anscombe généralisée :
 - **sigma_ANS** : estimation de l'écart-type du bruit
 - $\alpha = \text{gain}$: scalaire strictement positif qui multiplie la loi de Poisson

Introduction

J'ai réalisé un stage au sein de l'équipe de microscopie électronique du Laboratoire de Physique des Solides du 13 mai 2019 au 30 août 2019. J'ai été encadrée par Nathalie BRUN. Ce laboratoire fait parti de l'Université Paris-Sud et du CNRS. Il se situe à Orsay. L'objectif du stage était de débruiter les images et données de l'équipe.

Cette équipe possède plusieurs microscopes électroniques qui ont chacun leurs particularités. Ils leur permettent ainsi d'étudier différents types d'échantillons à des fins diverses et variées. Ils analysent la structure et la composition chimique d'échantillons par la détection des éléments présents et/ou de leurs liaisons chimiques. L'objectif est alors d'identifier la nature chimique des éléments et ainsi de réaliser par exemple leur carte chimique.

L'acquisition des données permet de créer, entre autres, des images dites hyperspectrales qui sont alors bruitées. Ce bruit pose un grand problème pour l'analyse et l'interprétation de leurs résultats. C'est pour cela qu'il est important de développer plusieurs outils permettant leur débruitage. Jusqu'à présent, l'analyse en composantes principales est la méthode la plus utilisée par le groupe. Cependant, elle présente des limites.

Pour ces raisons, l'équipe STEM collabore avec l'équipe de Signal et Communications de l'Institut de Recherche en Informatique à Toulouse et plus particulièrement avec Thomas OBERLIN, Nicolas DOBIGEON et Étienne MONIER. Ils les aident alors à développer des algorithmes performants adaptés à leurs types de données et qui soit facilement utilisables par des microscopistes.

Ma démarche s'est articulée autour de ces quatre étapes principales :

- 1) Familiarisation avec le contexte en testant des algorithmes déjà implémentés
- 2) Construction de notre méthode en implémentant les fonctions nécessaires
- 3) Réalisation de tests sur des données synthétiques afin de valider notre démarche
- 4) Application de notre méthode aux données réelles obtenues par les microscopistes

Le rapport se décompose en cinq parties : le contexte et la présentation du sujet, les pré-requis nécessaires, l'implémentation, les protocoles expérimentaux et enfin la présentation des résultats sur les données de l'équipe STEM.

Chapitre 1

Contexte du stage et présentation du sujet

1.1 Présentation du laboratoire

Le CNRS est le plus grand organisme public français de recherche scientifique. Il est né de la fusion d'une agence de moyens et d'une grande institution de laboratoires et de chercheurs en 1939. La fusion a été préparé par Jean ZAY et adopté par un décret-loi signé, entre autre, par le président de la République de l'époque : Albert LEBRUN. Il emploie environ 32 000 personnes pour un budget de 3.3 milliards d'euros par an. À sa tête, Antoine PETIT, chercheur français agrégé de mathématiques.

Le CNRS se découpe en dix instituts dont l'Institut de physique (INP) auquel est affilié le LPS. Ce dernier est une unité mixte de recherche de l'Université Paris-Sud et du CNRS (UMR 8502). Il a été fondé en 1959 par un trio de chercheurs : Raimond CASTAING, Jacques FRIEDEL et André GUINIER. Il est actuellement dirigé par Sylvain RAVY et emploie environ 200 personnes.

Il existe trois axes thématiques au LPS dont celui de phénomènes physiques aux dimensions réduites dans lequel s'inscrit l'équipe STEM. Elle est dirigée par Odile STÉPHAN. Cette équipe est spécialisée dans la microscopie électronique et plus particulièrement dans la spectroscopie par perte d'énergie des électrons (EELS). Leurs intérêts scientifiques sont vastes et englobent différents domaines de la physique et des sciences des matériaux : nanoparticules, interfaces, nanophotonique, matériaux biologiques, ... Ils sont sans cesse en évolution et possèdent des microscopes uniques qui sont à la pointe de la technologie.

1.2 Les microscopes électroniques de l'équipe STEM

La microscopie électronique regroupe plusieurs techniques expérimentales qui reposent sur la détection des signaux issus de l'interaction d'un faisceau électronique avec la matière contenue dans un échantillon. Ces signaux donnent des informations structurales ou spectroscopiques. Dans la figure 1.1, nous représentons les signaux produits lors de cette interaction.

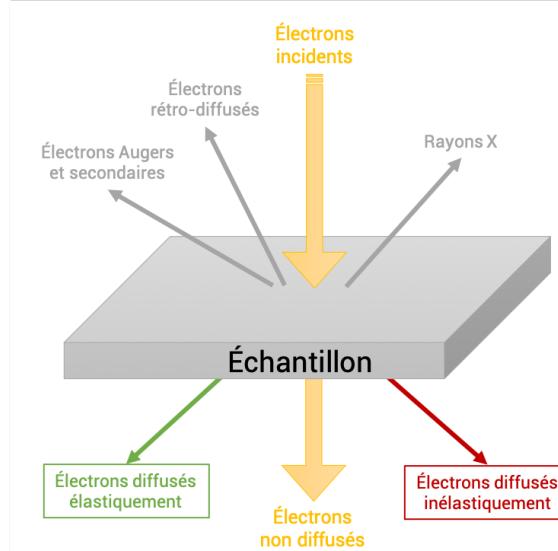


FIGURE 1.1 – Schéma représentant les signaux émis lors de l'interaction d'un faisceau électronique avec la matière

Les signaux qui nous intéressent sont les électrons diffusés inélastiquement qui sont la base de la spectroscopie EELS. Leur analyse repose sur la mesure de la quantité d'énergie perdue par les électrons incidents sur l'échantillon.

Le faisceau électronique réalise un balayage d'une zone choisie de l'échantillon. Pour chaque position, l'énergie perdue par les électrons est analysée par un spectromètre puis détectée par une caméra. Cette analyse nous donne un *spectre* dont des exemples sont donnés en figure 1.2.

Chaque spectre correspond donc à une courbe de perte d'énergie. À travers ces courbes, il est possible d'identifier les pics caractéristiques d'éléments chimiques. Sur la figure 1.2, les quatre grands pics du milieu correspondent aux pics caractéristiques du titane. Grâce à la mesure de leur intensité, nous pouvons obtenir sa carte chimique.

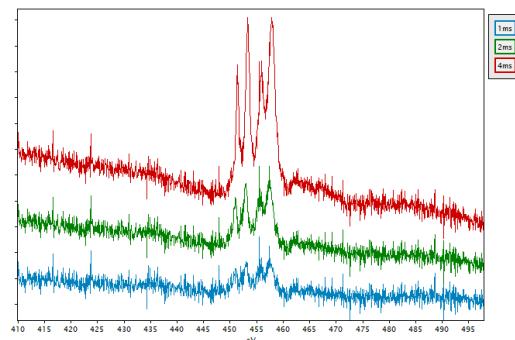


FIGURE 1.2 – Exemples de spectres du seuil $\text{Ti-L}_{2,3}$ pour différents temps d’acquisition

L’équipe possède trois microscopes de type STEM (Scanning Transmission Electron Microscope). Ils permettent une analyse structurale et chimique grâce à la spectroscopie EELS, la résolution spatiale dépend de la taille de la sonde (de 1 nm à moins de 0.1 nm). Ils ont chacun des résolutions spatiales et spectrales données ainsi que des caméras d’acquisition différentes. Chacun produit donc des données ayant différentes caractéristiques, que ce soit au niveau du signal ou du bruit.

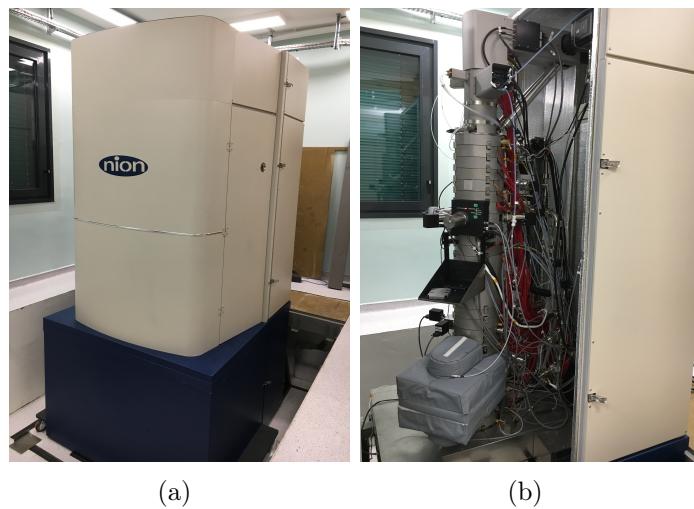


FIGURE 1.3 – Photos du ChromaTEM, un des microscopes STEM de l’équipe

Un élément important à prendre en compte est l’endommagement des échantillons généré par le faisceau électronique. En effet, il est possible d’augmenter le nombre d’électrons incidents, en augmentant soit le temps d’acquisition soit le courant du faisceau, cependant cela peut abîmer les échantillons et altérer la qualité des données acquises en perdant les informations. Plus l’échantillon est sensible, plus il faut diminuer la quantité d’électrons et plus les données finales sont alors bruitées. Il est important de trouver un juste équilibre entre le nombre d’électrons incidents et le niveau de bruit final.

1.3 L'imagerie hyperspectrale

Pour chaque acquisition, nous obtenons donc ce qui est appelée une image hyperspectrale (IHS) qui correspond à un cube de données, plus communément appelé hypercube. Le terme *Spim* (Spectre Image) est aussi utilisé. Dans la figure 1.4, nous en donnons un schéma.

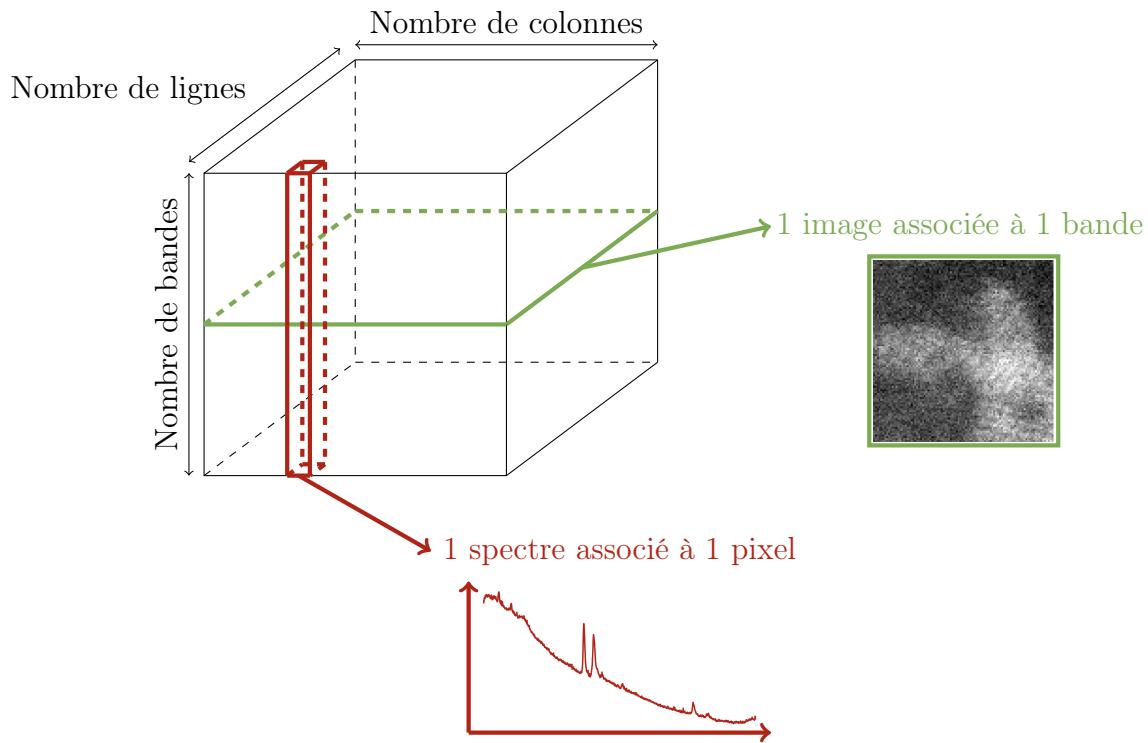


FIGURE 1.4 – Schéma d'un hypercube

Nous notons les dimensions de l'hypercube :

- L le nombre de lignes,
- C le nombre de colonnes,
- $P = L \times C$ le nombre de pixels,
- B le nombre de bandes.

Le balayage est réalisé sur le dessus du cube. Le nombre de lignes multiplié par le nombre de colonnes correspond donc au nombre de pixels. À chaque pixel est associé un vecteur de longueur le nombre de bandes. Il correspond à la courbe des pertes d'énergie : le spectre. De plus, notons qu'une image est associée à chaque bande.

Cet hypercube en trois dimensions contient donc deux types d'informations qu'il est important de débruiter : les informations spatiales mais aussi celles spectrales.

1.4 Les différentes natures de bruit

Il existe différents bruits qui entrent en jeu lors de l'acquisition des données. Chaque modèle de caméra possède un bruit caractéristique et dans notre cas, nous utilisons une caméra CCD. Nous allons présenter les deux natures principales du bruit.

Tout d'abord, il y a le bruit de lecture ou bruit de capteur. Il est dû à l'électronique embarquée et est lié à la conversion analogique-numérique du signal. Généralement, il est supposé additif, de moyenne nulle, spatialement décorrélé et gaussien.

Puis, il y a le bruit de photons ou bruit de comptage qui est dû aux variations statistiques naturelles du flux des photons incidents au capteur de la CCD. Il est supposé poissonnien et dépendant des données non bruitées.

Nous en concluons que nos données sont polluées par du bruit poissonnien et gaussien, usuellement appelé bruit mixte.

1.5 État de l'art

1.5.1 Débruitage standard

Il existe de nombreuses méthodes de débruitage adaptées à tout type d'image et de bruit. Il s'agit de méthodes qui peuvent être monodimensionnelles (applicables spectre-par-spectre) ou bien bidimensionnelle (applicables bande-par-bande).

Tout d'abord, il y a le débruitage par seuillage. Les premiers travaux sont basés sur le filtre de Wiener qui s'agit en fait d'un filtrage linéaire appliqué à l'image bruitée. Pour améliorer ce filtre, le seuillage par ondelettes est proposé. À travers la transformée en ondelettes discrète, une base de représentation possiblement orthogonale est fournie. Donoho et Johnstone [1] ont fortement contribué en montrant comment réaliser un débruitage efficace avec un simple seuillage de la transformée en ondelettes. À nouveau, pour améliorer les performances, le principe d'apprentissage de dictionnaire a été introduit. Il permet de construire des représentations adaptées à des classes génériques d'images voire adaptées à l'image elle-même.

Ensuite, il y a les méthodes géométriques qui favorisent une propriété de régularité dans l'image débruitée. Pour cela, il faut minimiser une fonctionnelle. La plus connue est la variation totale [2]. Cette dernière présentant des limites, deux alternatives ont été proposées : la variation totale généralisée [3] et la variation totale vectorielle [4].

Enfin, il y a les méthodes non-locales (à patchs). Elles reprennent les concepts de parcimonie et de dictionnaires en favorisant une propriété d'auto-similarité des images. Ce sont ces dernières qui nous intéressent. La méthode fondatrice est NL-Means [5]. De nombreuses variantes ont été

inventées comme NL-Bayes [6] ou BM3D [7]. Dans notre cas, nous appliquerons ce type de méthode bande-par-bande. Ces méthodes nous permettront d'exploiter la corrélation spatiale mais la corrélation spectrale sera négligée. Au premier abord, elles ne sont donc pas adaptées aux IHS.

1.5.2 Débruitage par sous-espace

Une façon de débruiter spectralement nos données est de réaliser un filtrage du signal. Pour cela, nous utilisons une approche par sous-espaces. Ce type d'approche permet d'exploiter le caractère redondant des IHS. Elles reposent principalement sur des analyses en valeurs singulières de matrice et il faut systématiquement choisir la dimension du sous-espace. Elles permettent ainsi d'améliorer les performances des algorithmes, de réduire les besoins en capacité de calcul et de stockage et de répondre à un soucis de localisation de l'information pertinente.

La plus connue et la plus utilisée par les microscopistes est l'ACP (analyse en composantes principales, cf. sous-section 2.2.1). De nombreuses méthodes ont été développées que ce soit par la communauté statistique ou celle de la télédétection hyperspectrale. Parmi elles, il y a le *maximum noise fraction* (MNF) [8] et la méthode *hyperspectral signal identification by minimum error* (HySime) [9].

Durant ce stage, nous avons choisi de travailler uniquement avec l'ACP car c'est la méthode qui semble la plus adaptée aux IHS.

1.5.3 Méthodes séquentielles et jointes

Que ce soit dans le cadre du débruitage standard ou du filtrage du signal, certaines informations ne sont pas prises en compte. C'est pour cette raison que des nouvelles méthodes ont été développées pour les IHS.

Il y a les approches séquentielles qui consistent tout simplement à combiner deux méthodes, une de débruitage spatial standard et une autre de filtrage du signal.

Puis il y a les approches jointes qui permettent d'opérer directement et conjointement sur l'hypercube. Elles évitent les approximations obtenues lors de l'utilisation des méthodes par sous-espace. Cependant, elles ont souvent une complexité très élevée. Il s'agit généralement d'extensions des méthodes standards en trois dimensions. Par exemple, il y a les ondelettes 3D [10] qui combinent les ondelettes 1D et 2D. Il existe les versions 3D de NL-Means et de BM3D qui sont nommées 3D NL-Means [11] et BM4D [12].

Durant ce stage, nous avons choisi de tester principalement les approches séquentielles en alliant l'ACP à une méthode de débruitage spatial standard et plus précisément une méthode à patchs. Puis, nous nous sommes lancés dans quelques tests de méthodes jointes.

Le schéma suivant résume les différentes approches qui existent en débruitage.

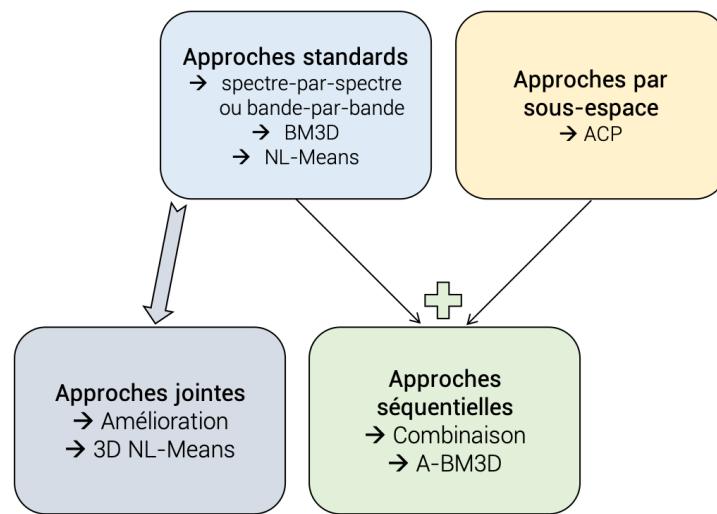


FIGURE 1.5 – Schéma récapitulatif des différentes approches de débruitages

Chapitre 2

Modèles et pré-requis

Durant tout ce chapitre, nous notons Y la matrice des données bruitées. Il s'agit de l'hypercube transformé en dimension 2 qui est alors de taille $P \times B$ en reprenant les notations précédentes. La matrice X correspond aux données non bruitées et est de même taille que la matrice Y .

Quelque soit le modèle choisi, l'objectif est de trouver l'image non bruitée X . Pour cela, nous l'estimons à l'aide de l'estimateur suivant.

$$\hat{X} = f_\theta(Y) \quad (2.1)$$

où f_θ correspond à la fonction appliquée par les algorithmes et θ aux paramètres nécessaires à l'algorithme.

2.1 Modèle du bruit

Tout d'abord, il est important de caractériser la nature du bruit. Il en existe plusieurs. C'est un critère à définir et à traiter en conséquence qui va influer sur les résultats obtenus. En effet, tous les algorithmes de débruitage ne sont pas adaptés à tous les types de bruit.

Les trois types de bruit que nous allons traiter sont donnés dans la liste ci-après.

- Bruit blanc gaussien, additif et iid :

$$Y = X + \varepsilon \quad \text{où} \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad (2.2)$$

- Bruit poissonnien, les données bruitées suivent une loi de Poisson dont le paramètre dépend des données non bruitées :

$$Y = \mathcal{P}(X) \quad (2.3)$$

- Bruit mixte, à la fois gaussien et poissonnien :

$$Y = \alpha \mathcal{P}(X) + \varepsilon \quad \text{où} \quad \alpha > 0 \quad \text{et} \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad (2.4)$$

Lorsque la matrice X est assez grande, le bruit poissonnien peut être assimilé à du bruit gaussien multiplicatif iid dont le modèle est donné par la formule suivante.

$$Y = \varepsilon X \quad \text{où} \quad \varepsilon \sim \mathcal{N}(1, \lambda^2) \quad (2.5)$$

De même, le bruit mixte peut être assimilé à du bruit gaussien additif et multiplicatif iid dont le modèle est donné par la formule suivante.

$$Y = X + \varepsilon \quad \text{où} \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 + \lambda X) \quad (2.6)$$

où λ dépend du α de la formule (2.4).

2.2 Méthodes utilisées

Durant la majorité du stage nous avons testé une approche séquentielle qui consiste à allier un filtrage du signal à un débruitage standard. Nous avons utilisé l'ACP alliée à la méthode à patchs BM3D. Nous avons également traité les trois types de bruit en transformant préalablement nos données à l'aide de la transformée d'Anscombe.

Par la suite, nous testerons éventuellement des approches jointes telle que 3D NL-Means par exemple.

2.2.1 Analyse en composantes principales

Dans cette partie, nous allons détailler l'ACP. Cette méthode permet d'identifier une base de représentation dans laquelle une décomposition du signal en composantes principales est réalisée.

Dans ce contexte, l'ACP est utile car les composantes ayant un faible pourcentage de variance sont considérées comme du bruit et ne contiennent, à priori, pas de signal. Cela nous permet alors de conserver l'information spectrale importante tout en retirant le bruit. C'est pour cela qu'initialement l'équipe STEM utilise l'ACP comme unique outil de débruitage. Plus le nombre de composantes conservées est faible, plus l'IHS paraît débruitée. Cependant, très souvent, les composantes non retenues contiennent toujours un peu de signal qui n'est alors pas conservé. Cela introduit un bon nombre d'artefacts et fait que les microscopistes sont à la recherche de nouvelles méthodes plus adaptées qui ne font pas apparaître ces faux résultats.

Pour pallier ce problème de perte d'information, nous avons décidé de réaliser une ACP centrée non normée en conservant suffisamment de composantes principales afin de ne perdre aucune information, aussi petite soit-elle. En fonctionnant ainsi, nous conservons beaucoup de bruit mais nous finirons par le supprimer en appliquant un débruitage spatial standard. De plus, parfois, appliquer une ACP centrée et normée permet de réduire le nombre de composantes sélectionnées sans pour autant créer d'artefact.

Plus concrètement, pour illustrer ce qu'est un artefact, nous montrons dans la figure 2.1 la carte chimique du samarium après ACP normée ou non en gardant le même nombre de composantes.

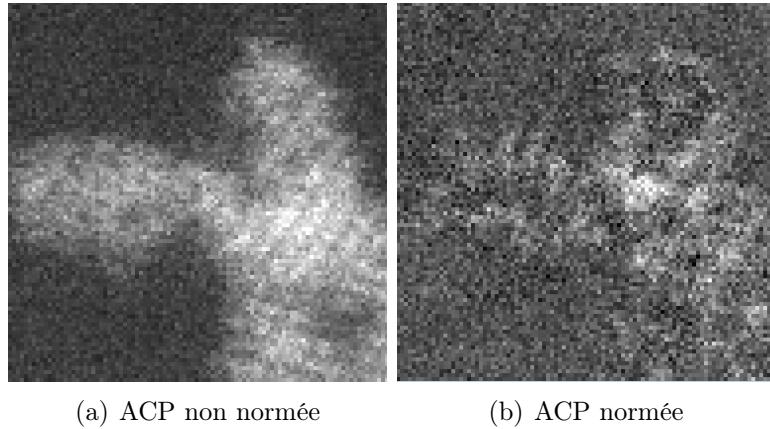


FIGURE 2.1 – Comparaison de l'ACP normée et de la non normée avec la carte du samarium et en gardant 3 composantes

Dans la sous-figure (a), nous avons représenté la carte chimique du samarium après ACP centrée non normée en gardant 3 composantes. Cependant, cette carte correspond à la carte chimique du cérium et non à celle du samarium. Ce résultat illustre parfaitement les problèmes qui sont rencontrés lorsque nous faisons face à un artefact. Dans la sous-figure (b), nous réalisons une ACP normée en gardant également 3 composantes. Nous retrouvons alors une meilleure carte du samarium qui ne semble pas présenter de structures du cérium.

L'échantillon étudié contient principalement du cérium qui a été dopé au samarium. Ce dernier est alors présent en petite quantité et son signal se perd dans le bruit. C'est ce qui explique que nous obtenions un artefact.

L'ACP que nous réalisons est classique : les composantes principales sont identifiées en décomposant en valeurs propres la matrice de corrélation. Pour obtenir cette matrice, il faut d'abord centrer nos données Y en calculant la moyenne sur les colonnes avec la formule suivante.

$$\bar{Y}_b = \frac{1}{P} \sum_{p=1}^P Y_p^b \quad (2.7)$$

Ainsi, nous obtenons les données centrées :

$$Y_c = Y - \bar{Y}_b = \left(Y_p^b - \bar{Y}_b \right)_{1 \leq p \leq P, 1 \leq b \leq B} \quad (2.8)$$

À cette étape, nous pouvons également normer nos données afin de réaliser une ACP normée (cf. Annexe A).

Enfin, la matrice de corrélation est donnée par :

$$C = \frac{1}{B} Y_c^T Y_c \quad (2.9)$$

Nous obtenons alors :

- un vecteur λ de taille B contenant les valeurs propres de C ,
- une matrice V de taille $B \times B$ qui contient les vecteurs propres $(v_b)_{1 \leq b \leq B}$ associés à chaque valeurs $(\lambda_b)_{1 \leq b \leq B}$ et qui correspond à la variable `coeff` dans le code.

Nous classons les deux par ordre décroissant des valeurs propres.

Nous sélectionnons alors le nombre K de composantes principales souhaitées qui est un entier compris entre 1 et B . Puis nous appliquons une ACP inverse afin de reconstruire nos données. Pour cela nous appliquons la formule suivante.

$$\hat{Y} = Y_c \cdot VV^T + \bar{Y}_b \quad (2.10)$$

$$= (Y - \bar{Y}_b) \cdot VV^T + \bar{Y}_b \quad (2.11)$$

Lorsque K est égal à B , aucun débruitage n'est effectué car nous gardons toute l'information. Le choix de ce paramètre K est très important pour la suite des étapes.

Pour choisir ce paramètre K , nous avons la possibilité de raisonner avec une valeur seuil. Dans ce cas-là, nous conservons les composantes admettant un pourcentage de variance supérieur ou égal à cette valeur seuil.

2.2.2 BM3D

La méthode BM3D est dite à patchs, elle exploite la redondance spatiale de l'image en regroupant des zones similaires afin de renforcer l'estimation des données non bruitées.

Dans cette partie, nous commençons par présenter les principales étapes de l'algorithme réalisé par BM3D. Pour plus de détails, nous nous référons à l'article [7]. L'algorithme se décompose alors ainsi :

1. **PREMIÈRE ÉTAPE : ESTIMATION DE BASE**
 - (a) **Estimations par patchs** : pour chaque patch de l'image bruitée, les étapes suivantes sont réalisées.
 - i. *Regroupement* : le patch sélectionné ait pris comme référence, il faut alors chercher tous ceux qui lui sont similaires. Puis ils sont tous empilés dans un tableau 3D nommé groupe.
 - ii. *Seuillage collaboratif* (hard-thresholding) : une transformée 3D est appliquée au groupe afin d'atténuer le bruit. Puis inversion de la transformée 3D afin de réaliser une estimation des patchs du groupe. Renvoie des estimations à leur position initiale.
 - (b) **Agrégation** : pondération des estimations obtenues qui se chevauchent.
2. **DEUXIÈME ÉTAPE : ESTIMATION FINALE**. En utilisant l'estimation de base, amélioration du regroupement et application d'un filtrage de Wiener collaboratif.
 - (a) **Estimations par patchs** : pour chaque patch de l'image bruitée, les étapes suivantes sont réalisées.
 - i. *Regroupement* : Utilisation de Block-Matching pour trouver les patchs similaires à celui de référence afin de former deux groupes : un d'image bruitée et l'autre d'estimations de base.
 - ii. *Filtre de Wiener collaboratif* : application d'une transformée 3D aux deux groupes. Application d'un filtre de Wiener au groupe de l'image bruitée en utilisant le groupe des estimations de base. Puis inversion de la transformée 3D afin de réaliser une estimation des patchs. Renvoie des estimations à leur position initiale.
 - (b) **Agrégation** : calcul de l'estimation finale de l'image en réalisant une moyenne pondérée de toutes les estimations qui se chevauchent.

Un schéma descriptif de l'algorithme est donné dans la figure suivante.

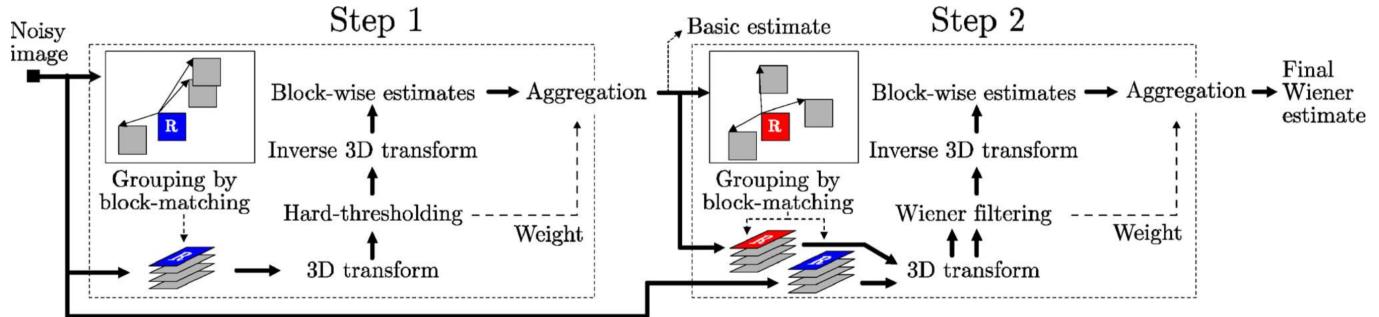


FIGURE 2.2 – Schéma de l'algorithme BM3D

Une illustration des étapes de regroupement des patchs est donnée dans la figure suivante.

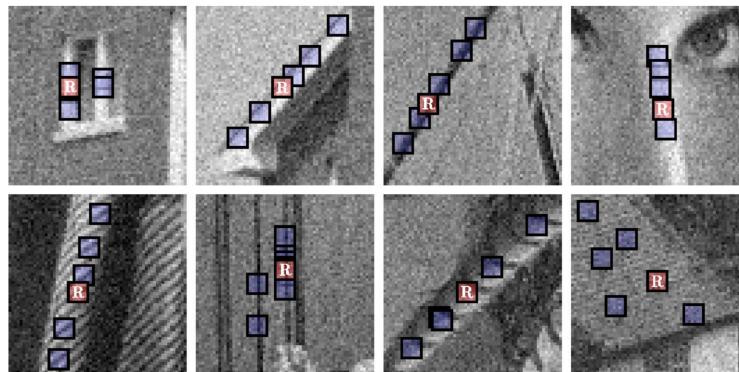


FIGURE 2.3 – Illustration des étapes de regroupement de BM3D, étapes 1.a.i. et 2.a.i

L'algorithme BM3D nécessite deux paramètres : l'image à débruiter et l'écart-type du bruit de cette image qui est alors supposé gaussien additif. BM3D n'a pas besoin que nous lui indiquions la taille des patchs à réaliser car il choisit lui-même. Cela facilite grandement son utilisation. C'est entre autres pour cela que nous l'avons choisi.

Les valeurs de l'image doivent être incluses dans $[0, 1]$ ou dans $[0, 255]$. Il est alors important de réduire les données si nécessaire. Dans notre implémentation, nous réalisons cette étape et réduisons nos données à l'échelle $[0, 255]$.

Pour l'écart-type du bruit, nous en réalisons une estimation. La méthode utilisée est présentée dans la sous-section 2.2.3. Il est important de noter que si nous réduisons les données à l'échelle $[0, 255]$, il faut également réduire cette valeur.

2.2.3 Estimation de l'écart-type du bruit

Pour réaliser une estimation de l'écart-type du bruit, nous nous plaçons dans le cas d'un hypercube redimensionné en matrice Y . Elle est alors supposée bruitée par un bruit blanc gaussien additif.

L'estimation repose sur la median absolute deviation (MAD) qui est une statistique robuste en présence de données aberrantes. Elle doit alors être calculée sur une bande b choisie préalablement par la formule suivante.

$$\text{MAD}(Y^b) = \text{med}(\{|Y_p^b - \text{med}(Y^b)|, p \in [1, P]\}) \quad (2.12)$$

où med correspond à la médiane. Puis l'estimation de l'écart-type du bruit est donnée par la formule suivante.

$$\hat{\sigma}_\varepsilon = 1.4826 \times \text{MAD}(Y^b) = \text{MAD}(Y^b)/0.6745 \quad (2.13)$$

Dans [13], il est proposé de réaliser préalablement une transformée en ondelettes puis d'utiliser MAD afin d'estimer l'écart-type sur l'ensemble de l'IHS. Nous avons décidé de retenir cette méthode qui semble renvoyer des résultats plus fiables et qui ne dépendent pas du choix d'une bande spécifique.

2.2.4 Transformée d'Anscombe

Lorsque les algorithmes de débruitage standard sont adaptés au cas du bruit gaussien additif et que nos données possèdent un bruit poissonnien, il faut appliquer une transformée aux données bruitées afin de pouvoir utiliser correctement les algorithmes. Pour cela, il existe différentes solutions.

De manière générale, la démarche à suivre est la suivante.

- 1) modification des données bruitées en appliquant une VST (Variance-Stabilizing Transformation),
- 2) débruitage spatial et spectral avec un algorithme traitant le bruit gaussien additif,
- 3) application de la VST inverse sur les données débruitées afin de les récupérer à l'échelle originale.

Transformée d'Anscombe classique

Une première VST possible est la transformée d'Anscombe [14]. Elle est adaptée aux données ayant un bruit uniquement poissonnien. Nous la testerons tout en gardant en tête ses limites. En effet, bien souvent, nos données contiennent un bruit mixte. La transformée est alors donnée par :

$$\mathcal{T}^{Ans}(y) = 2\sqrt{y + \frac{3}{8}} \quad (2.14)$$

Comme la racine carrée n'est pas une fonction linéaire, appliquer son inverse introduirait un biais dans l'estimation. Une première possibilité est d'utiliser l'inverse asymptotiquement non biaisée, ce qui permettrait d'atténuer le biais.

Nous avons décidé de conserver la solution donnée par [15]. Il s'agit d'une approximation de l'inverse non biaisée de la transformée \mathcal{T}^{Ans} . Elle est donnée par la formule suivante.

$$\mathcal{I}^{Ans}(z) = \frac{1}{4}z^2 + \frac{1}{4}\sqrt{\frac{3}{2}}z^{-1} - \frac{11}{8}z^{-2} + \frac{5}{8}\sqrt{\frac{3}{2}}z^{-3} - \frac{1}{8} \quad (2.15)$$

Transformée d'Anscombe généralisée

Une autre VST possible est la transformée d'Anscombe généralisée [16]. Celle-ci est adaptée au bruit mixte. Elle est donnée par la fonction suivante.

$$\mathcal{T}_{\sigma_\varepsilon}^{GAns}(y) = \begin{cases} \frac{2}{\alpha}\sqrt{\alpha y + \frac{3}{8}\alpha^2 + \sigma_\varepsilon^2} & \text{si } y + \frac{3}{8}\alpha + \frac{1}{\alpha}\sigma_\varepsilon^2 > 0 \\ 0 & \text{sinon} \end{cases} \quad (2.16)$$

où α et σ_ε sont définis à la formule (2.4). La difficulté pour appliquer cette méthode réside alors dans l'estimation de ces deux paramètres.

Tout d'abord, dans [17], α est le gain du détecteur CCD. C'est une valeur dont il est possible de déterminer au moins l'ordre de grandeur pour chaque hypercube. Puis pour estimer σ_ε , nous utilisons la méthode décrite dans la sous-section 2.2.3.

Comme pour la transformée d'Anscombe classique, $\mathcal{T}_{\sigma_\varepsilon}^{GAns}$ n'étant pas affine, appliquer son inverse introduirait un biais dans l'estimation. Dans l'article [16], il est alors conseillé d'utiliser une approximation de l'inverse non biaisé de la transformée généralisée qui est donnée par l'égalité suivante.

$$\mathcal{I}_{\sigma_\varepsilon}^{GAns}(z) = \frac{1}{4}z^2 + \frac{1}{4}\sqrt{\frac{3}{2}}z^{-1} - \frac{11}{8}z^{-2} + \frac{5}{8}\sqrt{\frac{3}{2}}z^{-3} - \frac{1}{8} - \sigma_\varepsilon^2 \quad (2.17)$$

Chapitre 3

Implémentation de la méthode séquentielle **A-BM3D**

Comme la grande majorité des algorithmes de débruitage sont implémentés en Matlab, nous avons décidé d'utiliser ce langage. Dans un premier temps, nous allons décrire notre algorithme en donnant l'ordre d'application des différentes méthodes. Puis dans un second temps, nous préciserons les paramètres d'entrée et de sortie de chaque fonction.

Comme évoqué précédemment, nous nous dirigeons naturellement vers une approche séquentielle du débruitage de nos données. Ainsi nous exploitons toutes les informations existantes, qu'elles soient spatiales ou spectrales. De plus, comme l'algorithme de débruitage spatial standard doit être appliqué bande-par-bande indépendamment, une étape de pré-filtrage par **ACP** nous permet de décorréler les bandes de l'hypercube. De plus, allier l'ACP à une méthode de débruitage par patchs a déjà montré son efficacité dans des précédents résultats. C'est également une approche qui n'est pas trop compliquée à utiliser ni trop lourde, que ce soit du point de vu de la mémoire ou du temps d'exécution. Toutes ces raisons renforcent notre choix d'utiliser cette méthode.

Dans la liste suivante, nous donnons les différentes étapes de notre algorithme de débruitage que nous nommons **A-BM3D**. Nous indiquons également en gris les étapes optionnelles.

- 1) Transformée d'Anscombe classique ou généralisée : application aux données bruitées
 - ▷ Pour la transformée généralisée : estimation de l'écart-type du bruit
- 2) **ACP** : sélection de K composantes principales
- 3) **BM3D** : débruitage spatial
 - ▷ `madNoiseEst` : estimation de l'écart-type du bruit
- 4) **ACP inverse** : reconstruction du Spim
- 5) Transformée d'Anscombe inverse : application aux données débruitées

Lorsque nous traitons des données synthétiques, nous les bruitons avec la nature de bruit souhaitée puis nous appliquons **A-BM3D**. Nous pouvons ensuite mesurer l'erreur et ainsi ajuster les réglages des paramètres.

Lorsque nous traitons des données réelles, nous ne pouvons pas mesurer l'erreur car nous n'avons pas d'IHS non bruitée. Nous exportons donc nos données débruitées afin de les analyser avec DM et ainsi visualiser les cartes chimiques.

Paramètres d'entrée et de sortie

► Pour l'**ACP**, nous avons jusqu'à trois paramètres d'entrée et il y a jusqu'à quatre paramètres de sortie. La ligne de commande est donnée ci-après.

```
function [rep, coeff, score, ind] = my_ACP (Y, K, seuil)
```

Tout d'abord, **Y** correspond aux données bruitées sous forme matricielle. Le paramètre **K** est égal au nombre de composantes K que nous souhaitons garder. Le paramètre **seuil** a été présenté précédemment et permet de sélectionner les composantes ayant un pourcentage de variance supérieur ou égal à cette valeur. Uniquement un seul de ces deux derniers paramètres est utilisable, il faut choisir.

Puis, la variable **rep** correspond aux données débruitées et reconstruites après l'**ACP**. La sortie **coeff** est égale à la matrice des vecteurs propres de la matrice de corrélation classés par ordre décroissant des valeurs propres associées.

La sortie **score** correspond aux données bruitées projetées dans l'espace de l'**ACP** *ie* **score** = **Y** * **coeff**.

Enfin, la sortie **ind** correspond au nombre de composantes sélectionnées lorsque nous décidons d'utiliser le paramètre **seuil**, dans ce cas-là **ind** = K .

► Pour **BM3D**, nous avons trois paramètres d'entrée et deux paramètres de sortie. La ligne de commande est donnée ci-après.

```
function [rep, sig] = my_BM3D (score_r, sigM, r)
```

Le paramètre **score_r** correspond à **score** renvoyé par l'**ACP** mais réduit *ie* en gardant les K premières colonnes. Le paramètre **sigM** est égale à l'estimation de l'écart-type du bruit non réduit. Le paramètre **r** est égal à r le ratio par lequel est multiplié le paramètre **sigma** de BM3D.

La sortie **rep** correspond à **score_r** débruité après **BM3D**. La sortie **sig** est égale au paramètre **sigma** de **BM3D** qui est égal à la réduction de **sigM** à l'échelle [0, 255] et éventuellement multiplié par r .

Chapitre 4

Protocoles expérimentaux

Dans cette partie, nous nous attelons à la validation de notre algorithme **A-BM3D**. Pour cela, il faut construire des données synthétiques puis les bruitner. Ensuite, nous les débruitons avec différents paramètres et nous comparons les erreurs afin de déterminer lesquels permettent d'obtenir les meilleurs résultats. Cette étape de validation nous permet également de vérifier, par exemple, que l'estimation de l'écart-type du bruit est correctement réalisée.

Dans un premier temps, nous présentons l'IHS synthétique que nous allons utiliser. Puis, nous donnons les formules des erreurs que nous calculons. Enfin, nous présentons les résultats obtenus et quels paramètres nous avons retenu.

4.1 Création d'images synthétiques

Nous avons à notre disposition un Spim synthétique qui a été créé par Étienne MONIER (IRIT, Toulouse). L'hypercube est de taille $120 \times 70 \times 1435$.

Dans la figure 4.1, nous représentons la première bande du Spim ainsi que deux spectres : celui d'un pixel de la zone du dessus (en bleu) et celui d'un pixel de la zone du dessous (en vert). Nous remarquons que ce Spim ne contient pas de seuils d'éléments chimiques en petite quantité.

Dans la figure 4.2, nous présentons les pics caractéristiques des quatre éléments chimiques présents dans l'échantillon : l'oxygène (O), le lanthane (La), le nickel (Ni) et le néodyme (Nd).

Par la suite, nous pouvons simuler les trois natures de bruits que nous avons décidé de traiter : bruit blanc gaussien additif (formule 2.2), bruit poissonnien (formule 2.3) et bruit mixte (formule 2.4).

Dans la figure 4.3, nous présentons les premières bandes de ces trois cas de figure ainsi que la première bande des données non bruitées à titre de comparaison.

Dans la figure 4.4, nous présentons les spectres associés aux trois types de bruit en conservant le même code couleur des pixels du haut et du bas.

Dans cette quatre figures, les spectres bleus sont associés à un seul et même pixel. De même pour les spectres verts.

Notons qu'il est tout à fait possible d'augmenter l'intensité du bruit en prenant une variance plus grande pour la loi normale ou un paramètre α supérieur. Notons qu'en procédant ainsi nous réalisons un bruitage spatial et spectral.

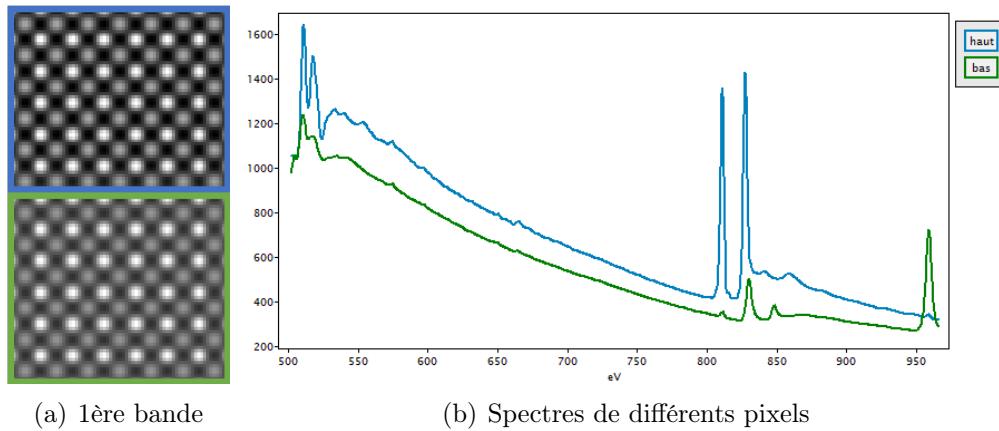


FIGURE 4.1 – Présentation du Spim synthétique : première bande et spectres

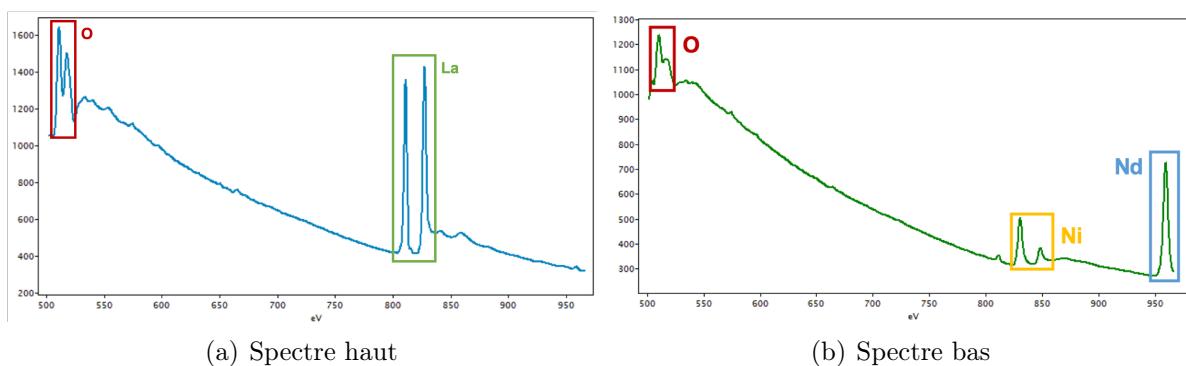


FIGURE 4.2 – Présentation des pics caractéristiques des éléments du Spim synthétique

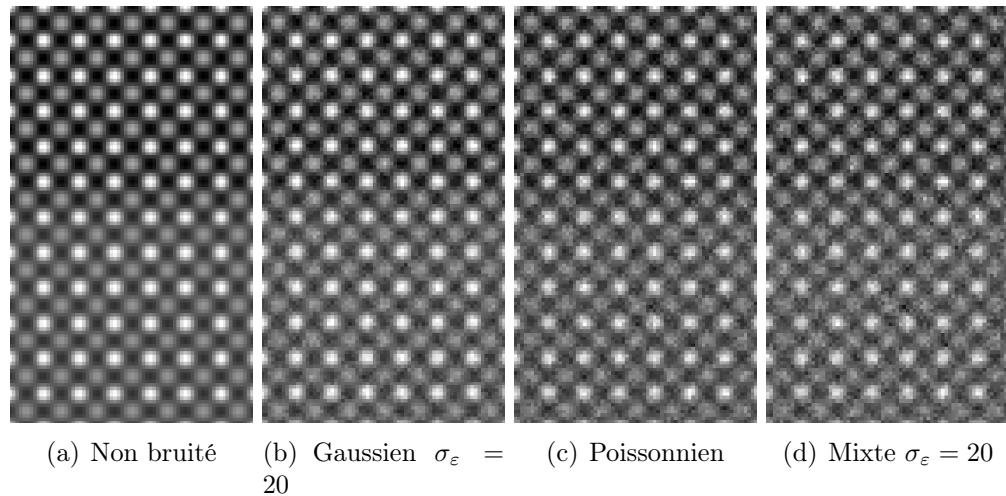


FIGURE 4.3 – Image des premières bandes des Spims synthétiques bruités

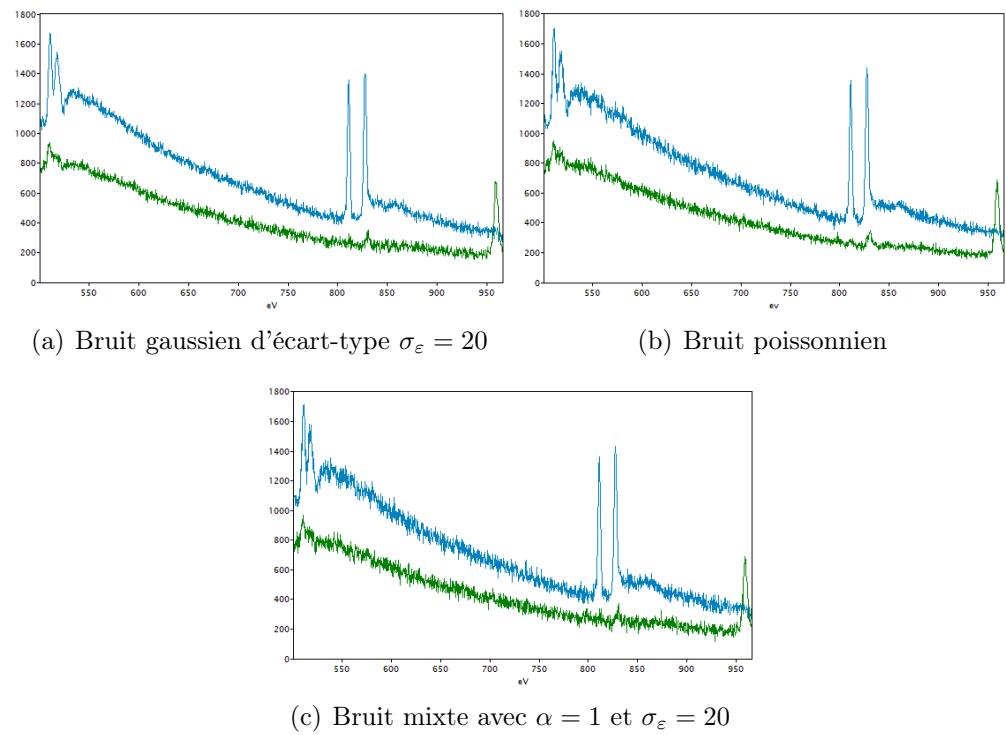


FIGURE 4.4 – Spectres de différents pixels des Spims synthétiques bruités

4.2 Calculs d'erreurs

Dans cette partie, nous donnons les formules des erreurs que nous allons calculer sur nos données synthétiques.

► Tout d'abord, nous allons utiliser l'EQM (Erreur Quadratique Moyenne). Pour $X, Y \in \mathcal{M}_{P,B}$, elle est donnée par la formule suivante.

$$\text{EQM}(X, Y) = \frac{1}{PB} \sum_{p=1}^P \sum_{b=1}^B (X_p^b - Y_p^b)^2 \quad (4.1)$$

► Ensuite, nous allons calculer la moyenne spatiale et la moyenne spectrale du SNR (Signal-to-Noise Ratio). Ce dernier est donné par la formule suivante où $U, V \in \mathbb{R}^N$ quelconque.

$$\text{SNR}(U, V) = 10 \log_{10} \left(\frac{\|U\|_2^2}{\|U - V\|_2^2} \right) \quad (4.2)$$

Puis les moyennes sont obtenues très aisément par les formules suivantes.

$$\text{MSNR}_P(X, Y) = \frac{1}{B} \sum_{b=1}^B \text{SNR}(X^b, Y^b) \quad (4.3)$$

$$\text{MSNR}_B(X, Y) = \frac{1}{P} \sum_{p=1}^P \text{SNR}(X_p, Y_p) \quad (4.4)$$

► Enfin, nous allons calculer la mesure globale MSSIM qui est obtenue à l'aide de la SSIM (Structural Similarity Index). Cette dernière est donnée par la formule suivante.

$$\text{SSIM}(X_p, Y_p) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \quad (4.5)$$

où :

- $\mu_X = \frac{1}{B} \sum_{b=1}^B X_p^b$ et $\mu_Y = \frac{1}{B} \sum_{b=1}^B Y_p^b$
- $\sigma_X^2 = \frac{1}{B-1} \sum_{b=1}^B (X_p^b - \mu_X)^2$ et $\sigma_Y^2 = \frac{1}{B-1} \sum_{b=1}^B (Y_p^b - \mu_Y)^2$
- $\sigma_{XY} = \frac{1}{B-1} \sum_{b=1}^B (X_p^b - \mu_X)(Y_p^b - \mu_Y)$
- $c_1 = (0.01d)^2$ et $c_2 = (0.03d)^2$ où d est la dynamique de l'image que nous avons pris égale à 1.

Cette valeur est calculée pour chaque pixel $p \in \llbracket 1, P \rrbracket$ puis nous prenons la moyenne pour finalement obtenir la MSSIM.

Au début, nous nous attendions à ce que les erreurs varient différemment. Cependant, cela n'a pas du tout été le cas : elles se comportaient de la même façon. C'est pour cette raison que nous avons uniquement utilisé l'EQM, bien que nous ayons calculé les quatre.

4.3 Validation de l'algorithme

Dans cette partie, nous commençons la validation de notre algorithme. Pour cela, nous allons utiliser les données synthétiques que nous bruitons. Nous pourrons comparer les données débruitées à celles initiales en calculant notamment les erreurs introduites en section 4.2. Nous pourrons ainsi voir quels paramètres sont optimaux. Néanmoins, il faut faire attention à ces résultats qui seront alors adaptés à ce type de données. En effet, nous verrons dans le chapitre suivant que les échantillons ont chacun leurs spécificités et qu'elles influent fortement sur l'efficacité du débruitage et sur la façon d'utiliser les méthodes.

Dans un premier temps, nous allons simuler du bruit blanc gaussien additif (cf. formule (2.2)), nous réduirons la dimensionnalité avec l'ACP puis nous appliquerons BM3D bande-par-bande sur les données projetées dans le sous-espace sélectionné.

Dans un second temps, nous simulerons du bruit poissonnien (cf. formule (2.3)). Nous réaliserons la même démarche que pour le bruit gaussien additif qui sera, cette fois, précédée d'une transformée d'Anscombe classique (parfois notée AT).

Dans un troisième et dernier temps, nous simulerons du bruit mixte (cf. formule (2.4)). Nous réaliserons la même démarche que pour le bruit poissonnien mais avec la forme généralisée de la transformée d'Anscombe (parfois notée GAT).

Dans tous les cas, nous testerons différents K et différents **sigma** puis nous comparerons les erreurs obtenues. Nous n'utiliserons pas la valeur seuil de l'ACP, nous choisirons le nombre de composantes à la main. Ici, nous présenterons uniquement l'EQM pour la lisibilité du document. Les autres erreurs se trouvent en Annexe B.

Nos données synthétiques non bruitées sont mises sous forme matricielle. Nous notons cette matrice X afin d'être cohérents avec les notations précédemment introduites. De même, les données synthétiques bruitées seront notées Y .

4.3.1 Simulation de bruit gaussien additif

À chaque valeur de nos données, nous ajoutons une variable aléatoire qui suit une loi normale. Ainsi, nous obtenons :

$$Y_p^b = X_p^b + \varepsilon_p^b \quad \text{pour tout } p \in \llbracket 1, P \rrbracket \text{ et } b \in \llbracket 1, B \rrbracket \quad (4.6)$$

où ε_p^b suit une loi $\mathcal{N}(0, \sigma_\varepsilon^2 = 20^2)$ pour tout p et b .

Dans cette partie l'estimation de σ_ε vaut 19.4. Nous pouvons en conclure que notre méthode fonctionne correctement. Après réduction de $\widehat{\sigma}_\varepsilon$ (cf. dernier paragraphe de la sous-section 2.2.2), nous obtenons comme paramètre **sigma** = 0.41. Si nous appliquons la même réduction à 20, nous obtenons 0.425. Ces deux valeurs sont suffisamment proches pour ne pas influer sur l'efficacité de BM3D.

Dans la figure suivante, nous représentons l'EQM en fonction du paramètre `sigma` et du nombre de composantes K .

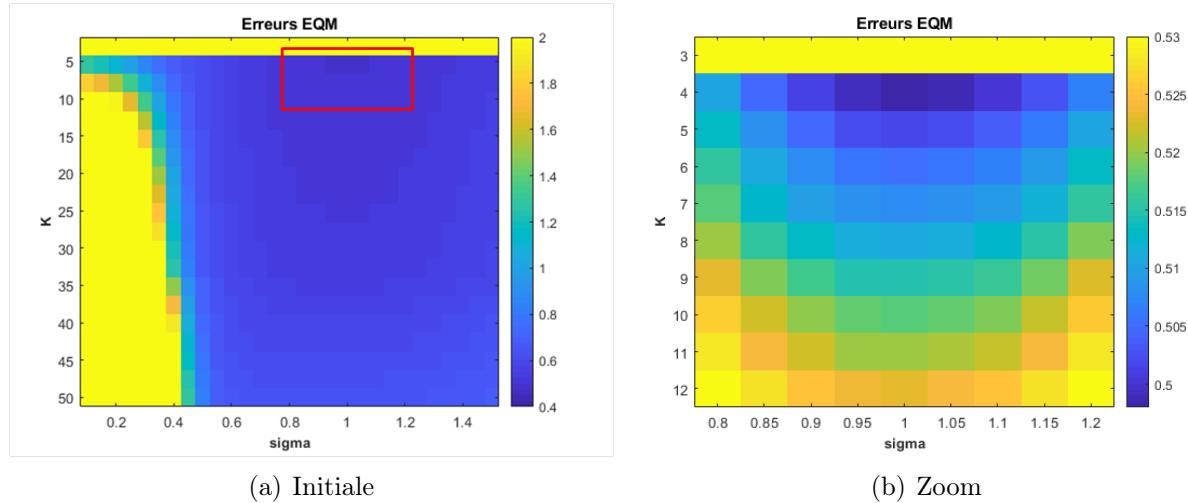


FIGURE 4.5 – Erreur quadratique moyenne pour différents paramètres `sigma` et K dans le cas du bruit blanc gaussien additif

La sous-figure (a) correspond à l'EQM pour K appartenant à $\{3, 4, \dots, 15\} \cup \{20, 25, \dots, 50\}$ et pour `sigma` variant dans $\{0.1, 0.15, \dots, 1.5\}$.

La sous-figure (b) correspond au zoom de l'encadré rouge de la sous-figure (a). Il s'agit de l'EQM pour K appartenant à $\{3, 4, \dots, 12\}$ et pour `sigma` variant dans $\{0.8, 0.85, \dots, 1.2\}$.

Nous remarquons que l'erreur minimale semble être obtenue pour $K = 4$ composantes et pour `sigma` compris entre 0.95 et 1.05. Nous avons donc tester à la main d'autres valeurs de `sigma` comprises dans cet intervalle et nous en avons conclu que 1.008 nous permettrait sûrement d'obtenir des erreurs plus basses.

Ensuite, nous calculons l'EQM pour différentes valeurs de paramètres :

- ▶ `sigma` = 0.41 : valeur calculée automatiquement,
- ▶ `sigma` = 1.008 : valeur sélectionnée à l'aide des graphiques de la figure 4.5.

Nous calculons également les erreurs obtenues après ACP seule, *i.e* sans BM3D.

<i>méthode</i>	ACP	A-BM3D	
sigma	×	0.41	1.008
<i>K</i>	<i>3 composantes</i>		
EQM	18.565	18.141	17.973
<i>K</i>	<i>4 composantes</i>		
EQM	1.342	0.680	0.498
<i>K</i>	<i>5 composantes</i>		
EQM	1.896	0.725	0.502
<i>K</i>	<i>32 composantes</i>		
EQM	16.328	1.737	0.583
<i>K</i>	<i>100 composantes</i>		
EQM	50.006	3.390	0.773
<i>K</i>	<i>221 composantes</i>		
EQM	103.693	4.919	1.071

TABLE 4.1 – Erreurs EQM pour du bruit gaussien additif

Quels que soient la méthode et le paramètre **sigma** utilisés, les meilleures erreurs sont obtenues pour $K = 4$ composantes. Cette valeur est égale au nombre de composantes utilisées lors de la synthèse de ces données.

Nous notons que la meilleure erreur est obtenue pour **sigma** = 1.008 et non avec la valeur calculée automatiquement. Cela n'est pas cohérent avec l'utilisation théorique de BM3D. En effet, l'algorithme est censé renvoyer ses meilleurs résultats avec l'écart-type estimé et réduit du bruit *ie* **sigma** = 0.42.

Utiliser la valeur 1.008 reviendrait à multiplier l'estimation obtenue automatiquement par un ratio de $r = 2.46$.

Nous remarquons qu'en dessous de 4 composantes, les résultats sont plutôt mauvais. Cela signifie que nous ne récupérons pas assez d'informations après l'ACP. Puis, au-delà de 4 composantes, les erreurs ne cessent de croître. Cela signifie que nous gardons trop de bruit. Cependant, BM3D semble suffisamment efficace et permet une nette réduction des erreurs quel que soit K .

Nous pouvons conclure que même si l'erreur 0.680 n'est pas minimale, elle est suffisamment petite et proche de **0.498**. Par la suite, nous continuerons tout de même de tester l'application d'un ratio sur **sigma**, car il est possible que son influence varie en fonction du type de données traitées.

4.3.2 Simulation de bruit poissonnien

Cette fois-ci, pour construire nos données bruitées, nous simulons $P \times B$ variables de Poisson de paramètres X_p^b selon la formule suivante.

$$Y_p^b = \mathcal{P}(X_p^b) \quad \text{pour tout } p \in \llbracket 1, P \rrbracket \text{ et } b \in \llbracket 1, B \rrbracket \quad (4.7)$$

Dans cette partie nous allons comparer les erreurs avec et sans transformée d'Anscombe. Ainsi, nous pourrons voir quelle amélioration elle apporte.

Si nous estimons l'écart-type du bruit gaussien sans appliquer de transformée d'Anscombe, nous obtenons une valeur de 20.5 qui une fois réduit vaut 0.43. Après application de la transformée d'Anscombe, ces deux valeurs valent respectivement 0.97 et 0.53.

Dans la suite, les résultats présentés sont obtenus lorsque la transformée d'Anscombe est appliquée, sauf indication contraire. Dans la figure suivante, nous représentons l'EQM en fonction du paramètre `sigma` et du nombre de composantes K .

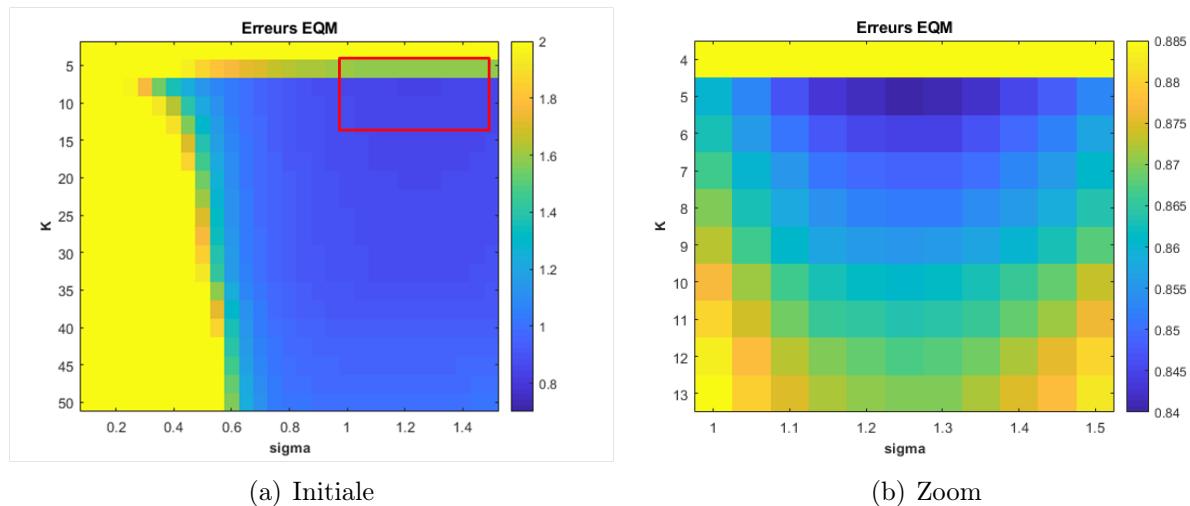


FIGURE 4.6 – Erreur quadratique moyenne pour différents paramètres `sigma` et K dans le cas du bruit poissonnien

La sous-figure (a) correspond à l'EQM pour K appartenant à $\{3, 4, \dots, 15\} \cup \{20, 25, \dots, 50\}$ et pour `sigma` variant dans $\{0.1, 0.15, \dots, 1.5\}$.

La sous-figure (b) correspond au zoom de l'encadré rouge de la sous-figure (a). Il s'agit de l'EQM pour K appartenant à $\{4, 5, \dots, 13\}$ et pour `sigma` variant dans $\{1, 1.05, \dots, 1.5\}$.

Cette fois-ci l'erreur minimale semble être obtenue pour $K = 5$ composantes et pour `sigma` compris entre 1.2 et 1.3. À nouveau, en modifiant `sigma` à la main, nous avons trouvé une valeur permettant d'obtenir de meilleurs résultats. Cette fois, elle vaut 1.27.

Ensuite, nous calculons les erreurs dans ces quatre cas de figure :

- ▶ après ACP sans transformée d'Anscombe,
- ▶ après ACP et BM3D sans transformée d'Anscombe et `sigma` calculé automatiquement,
- ▶ après ACP et BM3D avec transformée d'Anscombe et `sigma` calculé automatiquement,
- ▶ après ACP et BM3D avec transformée d'Anscombe et `sigma` = 1.27.

qui sont des valeurs de paramètres qui nous paraissent pertinentes grâce à la figure 4.6.
Elles sont alors résumées dans le tableau suivant.

<i>méthode</i>	ACP	A-BM3D		
Anscombe	×	×	✓	✓
<code>sigma</code>	×	0.43	0.53	1.27
<i>K</i>	<i>3 composantes</i>			
EQM	19.157	18.580	21.045	20.767
<i>K</i>	<i>4 composantes</i>			
EQM	2.002	1.150	1.840	1.577
<i>K</i>	<i>5 composantes</i>			
EQM	2.973	1.446	1.103	0.840
<i>K</i>	<i>6 composantes</i>			
EQM	3.935	1.750	1.159	0.845
<i>K</i>	<i>7 composantes</i>			
EQM	4.884	2.029	1.214	0.848
<i>K</i>	<i>41 composantes</i>			
EQM	34.961	9.763	2.749	0.965

TABLE 4.2 – Erreurs EQM pour du bruit poissonnien

Nous remarquons que sans transformée d'Anscombe, les meilleurs résultats sont obtenus pour $K = 4$ composantes. Tandis qu'avec transformée d'Anscombe, ils sont obtenus pour $K = 5$ composantes. Cela démontre bien que cette valeur est liée au nombre d'éléments chimiques présents dans l'échantillon mais n'est pas forcément égale.

L'erreur la plus basse est obtenue pour `sigma` = 1.27. Cela reviendrait alors à multiplier l'estimation automatique `sigma` = 0.53 par un ratio de $r = 2.40$.

Si nous appliquons une ACP et BM3D sans transformée d'Anscombe avec ce même paramètre, nous obtenons une erreur encore plus petite qui vaut **0.764** (résultat non présenté ici, voir

Annexe B.2). Mais l'erreur **0.840** étant assez proche et suffisamment petite, cela n'introduira pas de différence notable visuellement.

Notons que pour un nombre de composantes trop bas, la transformée d'Anscombe ne permet pas d'améliorer les erreurs obtenues, au contraire, elle semble les détériorer. Il devient intéressant de l'utiliser à partir de $K = 5$ composantes.

Sans transformée d'Anscombe et en prenant `sigma = 1.27`, les erreurs semblent systématiquement meilleures quelque soit le nombre de composantes K . Nous concluons qu'il n'est pas forcément pertinent de transformer nos données avant débruitage. De plus, cela nous mène à penser que la forme généralisée de la transformée d'Anscombe n'apportera pas non plus de nette amélioration. C'est pour cela que nous avons décidé de mettre la partie sur la simulation de bruit mixte en Annexe C, les résultats sont assez similaires et concluent par l'inefficacité de la forme généralisée de la transformée d'Anscombe.

4.3.3 Inefficacité de la transformée d'Anscombe classique et généralisée

Dans cette partie, nous cherchons à trouver une explication à l'inefficacité de la transformée d'Anscombe, qu'elle soit classique ou généralisée. Pour cela, nous allons étudier le comportement du pourcentage de variance des composantes lors de l'ACP.

Dans le cas du bruit poissonnier, nous comparons avec et sans transformée d'Anscombe classique. Dans la figure 4.7, nous représentons les courbes de pourcentage de variance en fonction des composantes à l'échelle logarithmique.

Dans le cas du bruit mixte, nous comparons avec et sans transformée d'Anscombe généralisée. Dans la figure 4.8, nous représentons les courbes de pourcentage de variance en fonction des composantes à l'échelle logarithmique.

Dans les deux cas, il n'y a pas de différence flagrante. Pour le bruit de Poisson, la règle du coude nous indique de conserver 4 composantes sans transformée et 5 composantes avec transformée d'Anscombe. Ce qui va dans le sens des résultats obtenus avec le bruit poissonnier. Pour le bruit mixte, les courbes sont encore plus semblables.

Ces similarités montrent bien que la transformée d'Anscombe n'influe pas sur le filtrage du signal par ACP.

Si nous comparons les spectres d'un même pixel des Spims débruités sans transformée, avec transformée d'Anscombe classique et avec transformée d'Anscombe généralisée, il n'y a pas non plus de différence notable.

Comme la transformée d'Anscombe n'apporte pas de nette amélioration que ce soit au niveau

des erreurs ou des spectres ou du pourcentage de variance, nous en déduisons qu'elle n'améliorera pas nos futurs résultats et nous n'insisterons donc pas sur son utilisation.

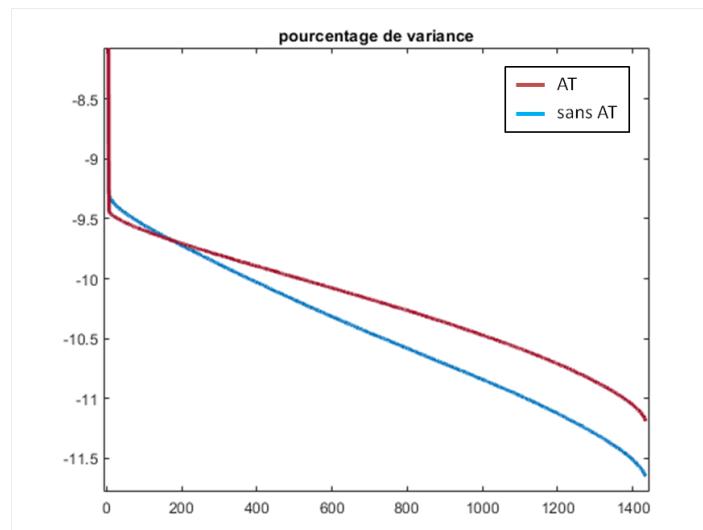


FIGURE 4.7 – Pourcentage de variance des composantes de l'ACP
dans le cas du bruit poissonnien

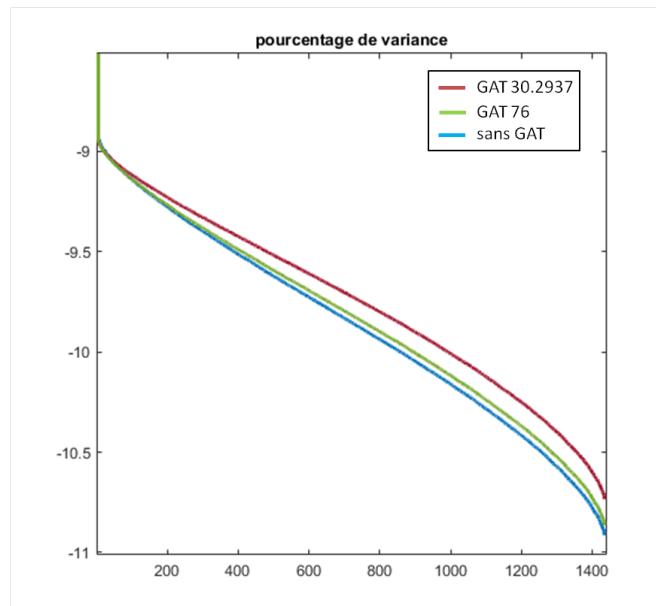


FIGURE 4.8 – Pourcentage de variance des composantes de l'ACP
dans le cas du bruit mixte

Chapitre 5

Résultats sur les données réelles

Dans ce chapitre, nous allons appliquer notre algorithme aux données réelles de l'équipe STEM. Ici, nous ne pourrons pas calculer d'erreurs car nous n'avons pas accès aux données non bruitées. Nous réaliserons donc des comparaisons visuelles uniquement à l'aide, entre autres, des cartes chimiques des éléments présents dans les échantillons. Parfois, nous pourrons également étudier l'allure des spectres.

5.1 Construction de carte chimique

Pour créer des cartes chimiques, nous utilisons un logiciel de traitement des données de microscopie appelé **Digital Micrograph** (DM). Il s'agit également du logiciel qui permet de contrôler les microscopes lors de l'acquisition des données.

Plus concrètement, nous donnons un exemple de la démarche à suivre pour construire une carte chimique. Dans la figure 5.1, nous présentons un spectre contenant les pics caractéristiques du samarium qui ont été isolés (cf. section 5.3 : Pics du samarium isolés).

Tout d'abord, il faut sélectionner ce qui est appelé le fond. Sur la figure 5.1, le fond correspond à la courbe rouge. Cette courbe est obtenue en réalisant un ajustement de loi puissance qui s'écrit sous la forme $y = ax^k$. Il faut alors estimer les paramètres a et k et il existe différentes méthodes qui dépendent, notamment, du nombre de rectangle(s) sélectionné(s) en rouge. Pour plus de détails, se référer à [18].

Ensuite, il faut définir la zone d'intérêt afin de sélectionner les seuils caractéristiques. Elle est définie par le rectangle vert nommé **signal**.

La carte chimique est finalement obtenue en sommant les intensités de chaque pixel d'une bande donnée puis en soustrayant les valeurs du fond correspondantes.

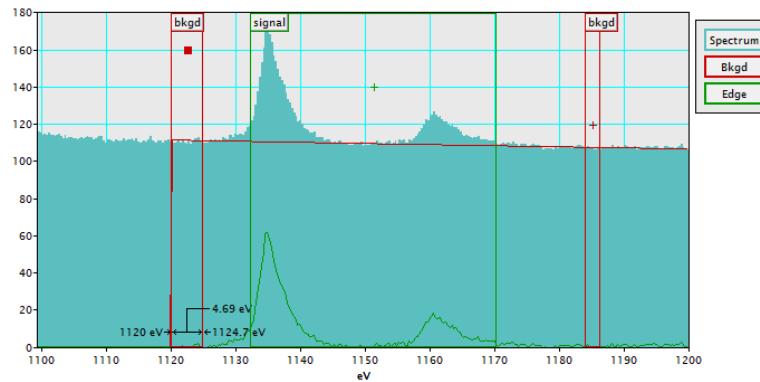


FIGURE 5.1 – Exemple de sélection de zones pour créer une carte chimique du Sm

Lorsque les données sont bruitées, identifier le fond fait partie des difficultés. En effet, d'un pixel à l'autre, il varie énormément. Pour en donner un exemple, nous présentons des spectres contenant les pics caractéristiques du titane dans la figure 5.2.

Les sous-figures (a), (b) et (c) indiquent les trois sélections que nous avons réalisés puis comparés.

Sur la sous-figure (d), nous avons superposé les fonds des trois sélections. Sur la sous-figure (e), nous avons superposé leur signal. Ainsi, nous voyons clairement l'influence que peut avoir la sélection du fond. Bien évidemment, plus les données sont bruitées, plus elles sont sujettes à ces variations.

Le débruitage permet un lissage du fond et donc une stabilisation de sa sélection. Une possibilité est de mettre deux fenêtres de fond. Ceci permet parfois d'améliorer la stabilisation du fond et ainsi d'acquérir de meilleurs résultats.

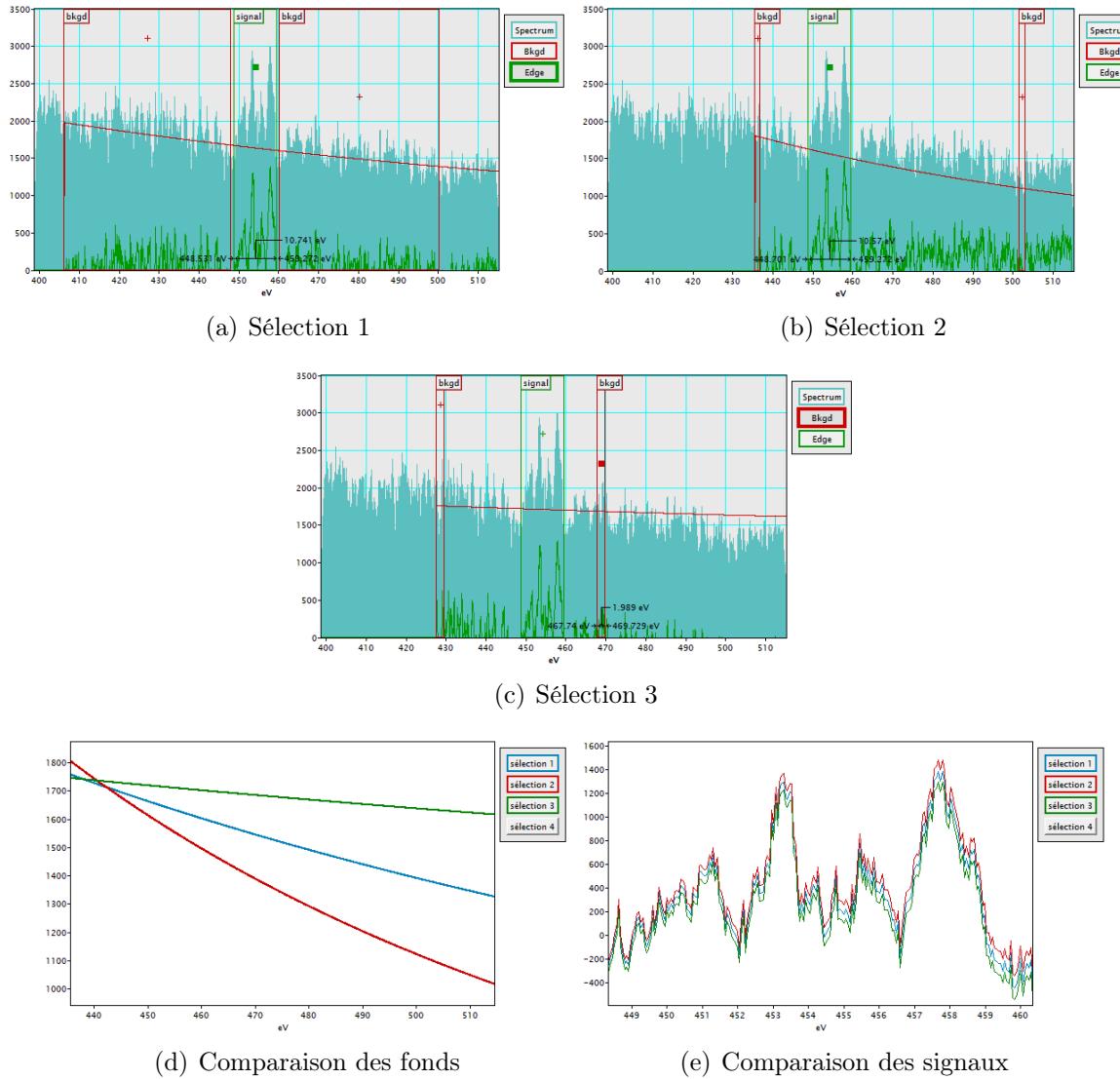


FIGURE 5.2 – Exemple de variation de sélection du fond d'un pixel sur l'échantillon de titane

5.2 Échantillon biologique

Dans cette partie nous étudions le Spim 11. Il est acquis avec une faible résolution spatiale (2 nm par pixel). Il s'agit de données portant sur un échantillon contenant de l'azote (N), du carbone (C) et de l'oxygène (O). Leur pic caractéristique se situe respectivement à 401 eV, 284 eV et 532 eV. Dans la figure 5.3, nous avons tracé la courbe de la somme de tous les spectres de l'hypercube bruité puis nous avons identifié les pics correspondant aux éléments chimiques.

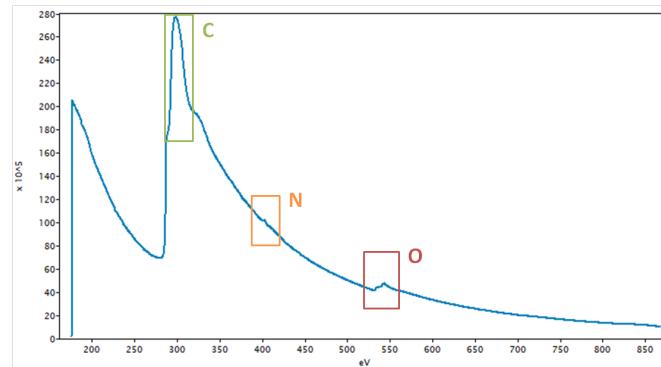


FIGURE 5.3 – Somme de tous les spectres du Spim 11, données bruitées

Dans cet échantillon, nous cherchons à identifier l'azote. Nous allons donc comparer les données initiales bruitées avec les données débruitées par **A-BM3D**.

Sans transformée d'Anscombe

Nous commençons par appliquer notre méthode sans utiliser de transformée d'Anscombe. L'écart-type estimé et réduit du bruit vaut 0.33. C'est donc cette valeur qui est donnée comme paramètre `sigma` à BM3D. Nous commençons par comparer les cartes chimiques de l'azote pour différents paramètres K .

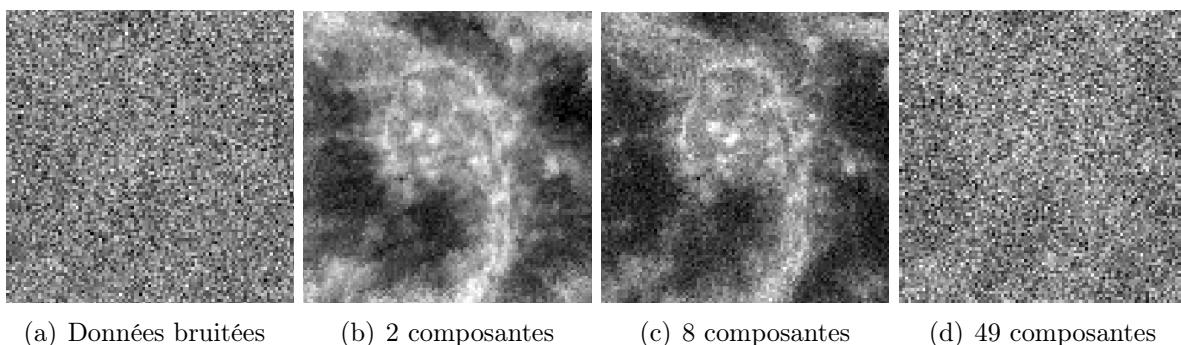


FIGURE 5.4 – Comparaison des cartes chimiques de l'azote pour différents K

Nous remarquons qu'en conservant 49 composantes, il y a trop de bruit. Nous avons donc choisi de ne pas s'intéresser au cas où $K \geq 50$. La carte chimique pour 8 composantes semblent être la plus adéquate. Cependant, nous notons qu'il reste tout de même un effet de bruit : comme un grésillement sur toute la carte.

Une possibilité pour palier à cet effet de grésillement est de multiplier **sigma** par le ratio r précédemment introduit. En multipliant par un ratio de $r = 8$, le débruitage est plus fort et nous obtenons une figure plus lisse. Dans ce cas-là, le paramètre **sigma** vaut $8 \times 0.33 = 2.64$.

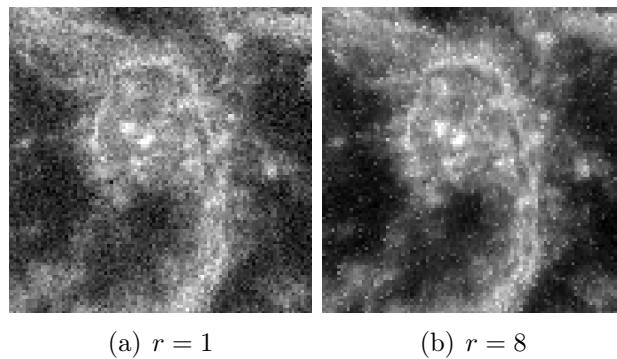


FIGURE 5.5 – Comparaison des cartes chimiques de l'azote pour $K = 8$ et $r \in \{1, 8\}$

Nous remarquons l'effet sur le fond noir de l'image qui paraît bien plus lisse sur l'image de droite. De plus, cette augmentation du paramètre ne semble pas influer négativement. Cependant, nous notons l'apparition de pixels blancs (cf. figure 5.6) lorsque nous prenons $r = 8$, ils doivent être dus au débruitage par BM3D.

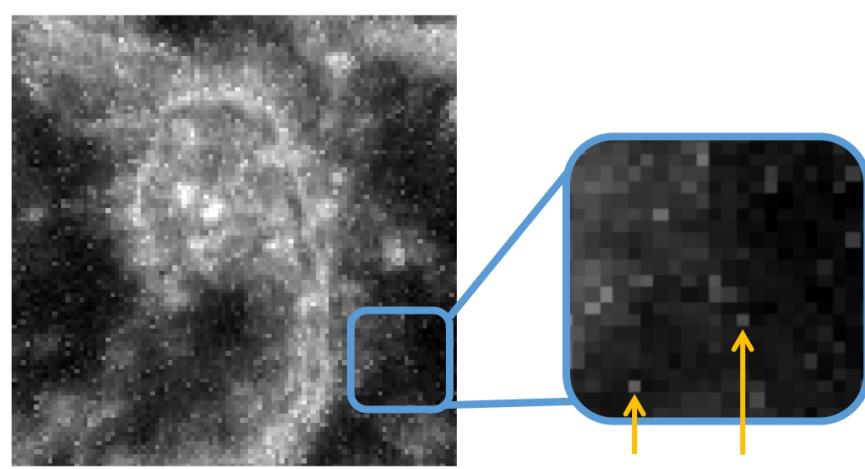


FIGURE 5.6 – Présentation des pixels blancs qui apparaissent après débruitage

Avec transformée d'Anscombe classique

Nous appliquons maintenant une transformée d'Anscombe avant de débruiter nos données. Pour le paramètre `sigma` de BM3D, l'écart-type estimé et réduit du bruit vaut 0.50. Lorsque nous le multiplions par 7, il vaut $7 \times 0.50 = 3.50$.

Nous allons comparer les cartes chimiques de l'azote pour K appartenant à $\{2, 4, 8\}$ et pour r appartenant à $\{1, 7\}$. Elles sont données dans la figure suivante.

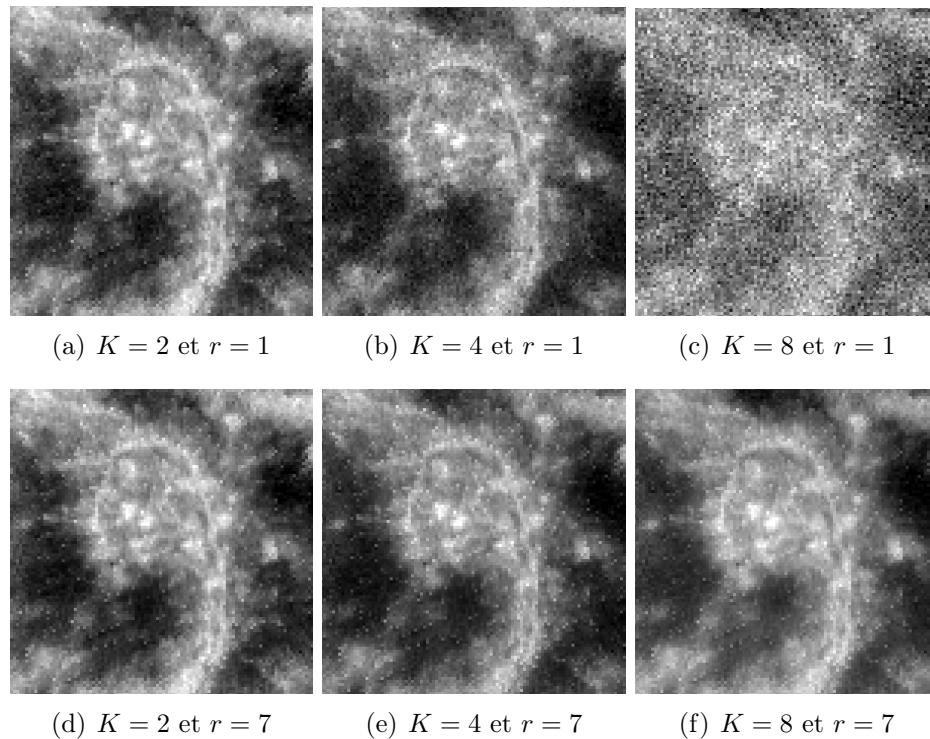


FIGURE 5.7 – Comparaison des cartes chimiques de l'azote pour différents K et r avec transformée d'Anscombe

La meilleure carte chimique est obtenue pour $K = 4$ composantes et $r = 7$. De plus, bien que nous conservions trop de composantes et donc trop de bruit, nous remarquons qu'en augmentant le ratio r , le débruitage est tout de même satisfaisant. En effet, la carte (f) est assez semblable aux cartes (d) et (e). À nouveau, nous remarquons l'apparition des mêmes pixels blancs que précédemment lorsque $r = 7$.

Nous allons maintenant comparer les cartes retenues dans ces deux dernières sections : avec et sans transformée d'Anscombe.

Dans la figure 5.8, la sous-figure (a) a été obtenue avec les paramètres $K = 4$ et $r = 7$. La sous-figure (b) avec $K = 8$ et $r = 8$. Les deux cartes sont assez similaires. Nous pouvons

tout de même noter qu'avec transformée d'Anscombe, la carte semble plus lisse, ce qui n'est pas forcément un critère de qualité. Sans transformée, nous distinguons mieux les détails indiquées par les flèches jaunes.

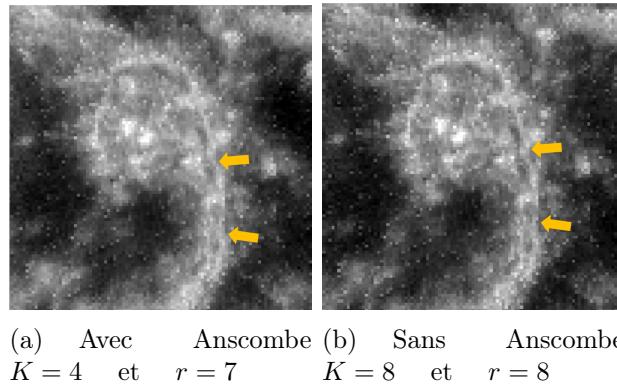


FIGURE 5.8 – Comparaison des cartes chimiques de l'azote avec et sans transformée d'Anscombe

Avec transformée d'Anscombe généralisée

Nous appliquons maintenant une transformée d'Anscombe généralisée avant de débruiter nos données. L'estimation de l'écart-type du bruit utilisée pour la transformée vaut 46.1. Pour le paramètre `sigma` de BM3D, l'écart-type estimé et réduit du bruit vaut 0.37. Lorsque nous le multiplions par 9.5, il vaut $9.5 \times 0.37 = 3.515$.

Nous allons comparer les cartes chimiques de l'azote pour les mêmes K que précédemment et pour r appartenant à $\{1, 9.5\}$. Elles sont données dans la figure 5.9.

Ces cartes chimiques sont assez similaires à celles obtenues avec transformée d'Anscombe classique. La transformée d'Anscombe généralisée n'apporte donc rien de plus. Elle ne permet pas non plus de réduire l'apparition des pixels blancs.

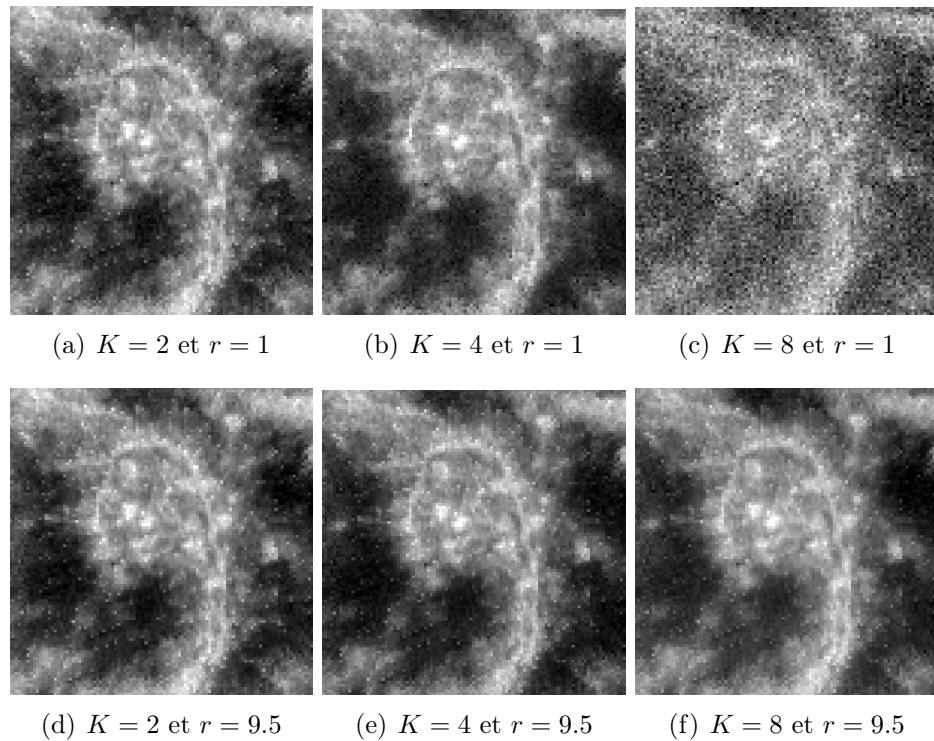


FIGURE 5.9 – Comparaison des cartes chimiques de l'azote pour différents K et r
avec transformée d'Anscombe généralisée

Visualisation des spectres

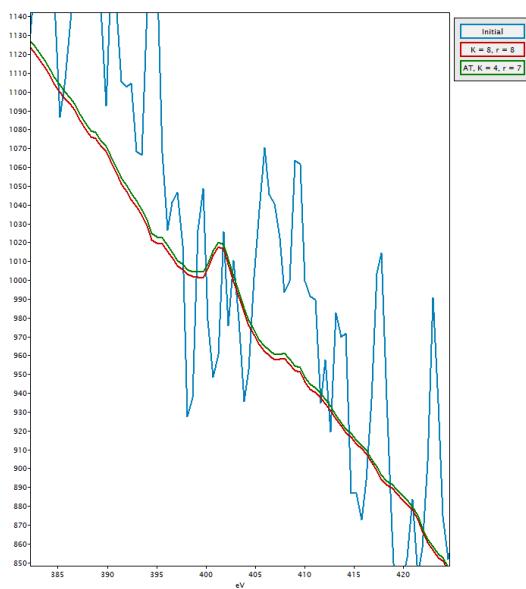


FIGURE 5.10 – Comparaison des spectres d'un pixel avant et après débruitage

Dans la figure 5.10, nous avons représenté les spectres d'un même pixel issu de trois Spims différents :

- ▶ un des données bruitées : courbe bleue,
- ▶ un des données débruitées sans transformée d'Anscombe, $K = 8$ et $r = 8$: courbe rouge,
- ▶ un des données débruitées avec transformée d'Anscombe classique, $K = 4$ et $r = 7$: courbe verte.

Nous avons zoomé sur la zone où se trouve le seuil caractéristique de l'azote.

Tout d'abord, en comparant les courbes des Spims débruités à celle des données bruitées, nous notons que le débruitage spectral semble efficace. En effet, sur la courbe bleue nous ne distinguons même pas de seuil caractéristique.

Puis, comme pour les cartes chimiques, il n'y a pas de différence notable entre le débruitage avec transformée d'Anscombe et celui sans transformée.

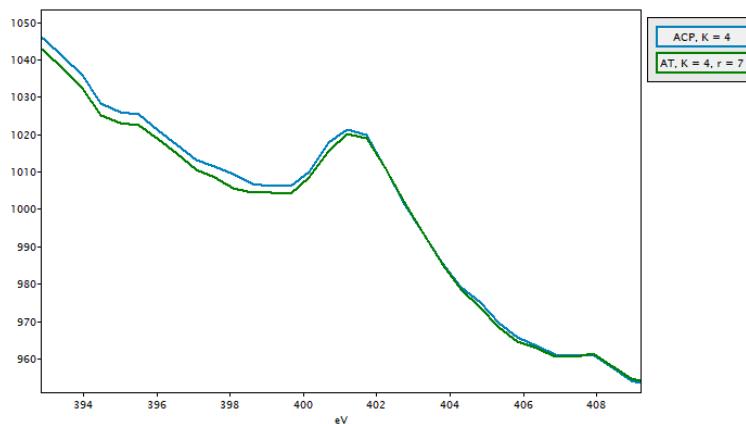


FIGURE 5.11 – Comparaison des spectres d'un pixel avec un débruitage par ACP seule et avec un débruitage par A-BM3D

Dans la figure 5.11, nous avons représenté les spectres du même pixel que précédemment issu de deux Spims différents :

- ▶ un des données débruitées avec une ACP seule (sans BM3D) et en gardant 4 composantes : courbe bleue,
- ▶ un des données débruitées avec transformée d'Anscombe classique, $K = 4$ et $r = 7$: courbe verte.

Nous avons zoomé sur la zone où se trouve le seuil caractéristique de l'azote.

Grâce à cette figure, nous comprenons que le débruitage spectral dépend uniquement de l'ACP. En appliquant BM3D à la suite, nous obtenons une nette amélioration des cartes chimiques et donc un meilleur débruitage spatial. Cependant, le débruitage spectral ne s'améliore pas plus.

5.3 Échantillon SmCeO₂

Dans cette partie, nous étudions un échantillon qui contient du cérium (Ce) et du samarium (Sm). Il est acquis avec une résolution atomique (0.06 nm par pixel). Leur pic caractéristique se situe respectivement à 883 eV et 1080 eV. Le samarium est présent en faible quantité, il est donc difficile de retrouver sa carte chimique. Comme pour le Spim 11, nous traçons la courbe de la somme des spectres de l'hypercube bruité et nous y identifions les pics caractéristiques.

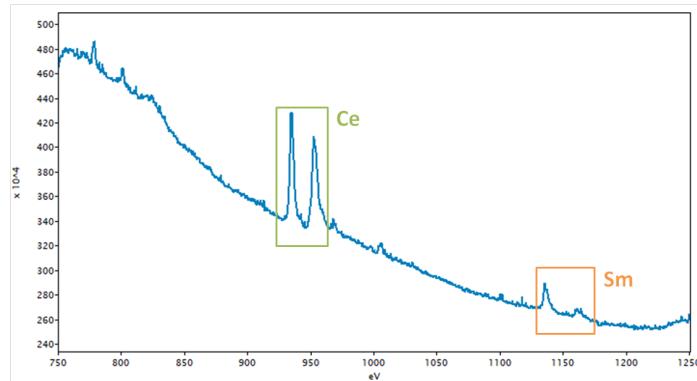


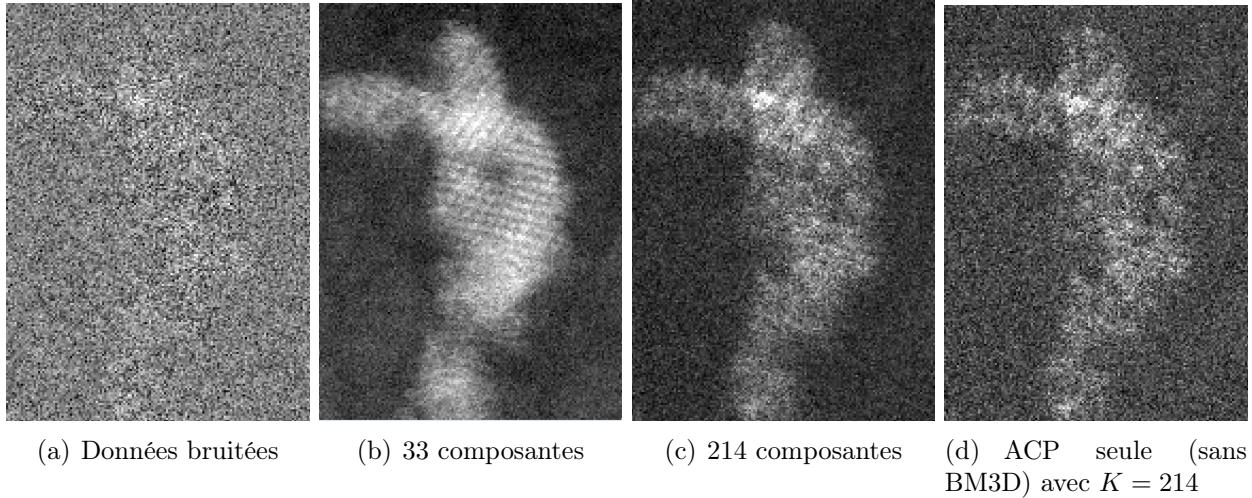
FIGURE 5.12 – Somme de tous les spectres de l'échantillon SmCeO₂, données bruitées

Dans cet échantillon, comme il est facile d'obtenir la carte du cérium, nous réaliserons alors notre sélection de paramètres sur la base de l'analyse des cartes du samarium.

Sans transformée d'Anscombe

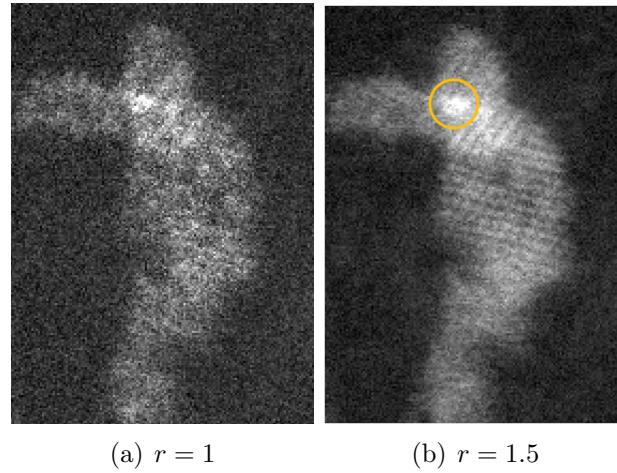
Cette fois, pour le paramètre `sigma` de BM3D, l'écart-type estimé et réduit du bruit vaut 3.32.

Dans la figure 5.13, nous comparons les cartes chimiques du samarium pour différentes valeurs de K . À titre de comparaison, nous avons représenté dans la sous-figure (d) la carte chimique du samarium après application d'une ACP uniquement et en gardant 214 composantes. Cela permet de voir le débruitage effectué par BM3D.

FIGURE 5.13 – Comparaison des cartes chimiques du samarium pour différents K

Lorsque nous prenons $K = 33$ composantes, nous remarquons l'apparition d'un artefact. En effet, il s'agit de la carte chimique du cérium et non du samarium. Cependant, à partir de $K = 214$ composantes, les informations du samarium apparaissent. L'image (c) ne nous satisfait pas complètement. En effet, il y a des parties qui ressemblent à la carte du cérium et l'image reste assez bruitée.

Voyons comment varient les résultats en utilisant le ratio précédemment introduit. Nous prenons $r = 1.5$. Dans ce cas-là, le paramètre **sigma** vaut 4.98.

FIGURE 5.14 – Comparaison des cartes chimiques du samarium pour $K = 214$ et $r \in \{1, 1.5\}$

Dans la sous-figure (b), la petite tache blanche entourée en jaune correspond au samarium mais nous voyons désormais très nettement qu'il s'agit toujours d'un artefact, même après un

débruitage plus fort des données. Augmenter le nombre de composantes ne semble pas atténuer le problème.

Nous proposons alors une première solution qui est d'isoler les deux pics caractéristiques du samarium et d'appliquer le débruitage en se limitant à cet intervalle de bandes.

Pics du samarium isolés

Lorsque nous isolons les pics, le paramètre `sigma` de BM3D change et vaut désormais 6.23. Nous avons construit les cartes chimiques du samarium pour différents paramètres K . Elles sont données dans la figure suivante.

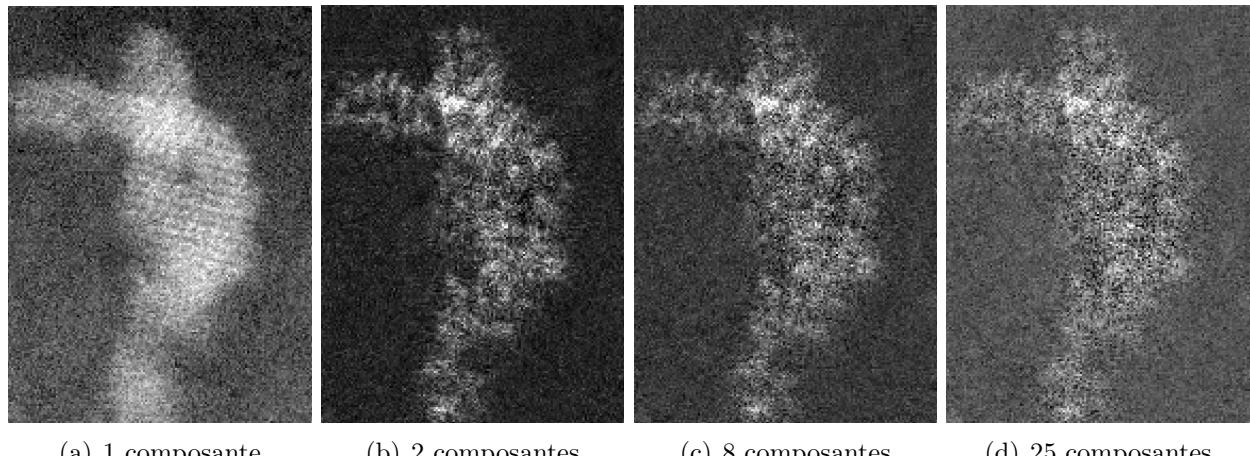


FIGURE 5.15 – Comparaison des cartes chimiques du samarium pour différents K après débruitage réalisé sur les pics isolés

Prendre plus de 2 composantes ne semble apporter aucune information importante en plus. Nous décidons donc de conserver $K = 2$ composantes et d'appliquer un ratio r au paramètre `sigma`.

Lorsque nous le multiplions par 1.5, il vaut 9.345 et pour un ratio de 2, il vaut 12.46. Nous obtenons donc les nouvelles cartes chimiques suivantes.

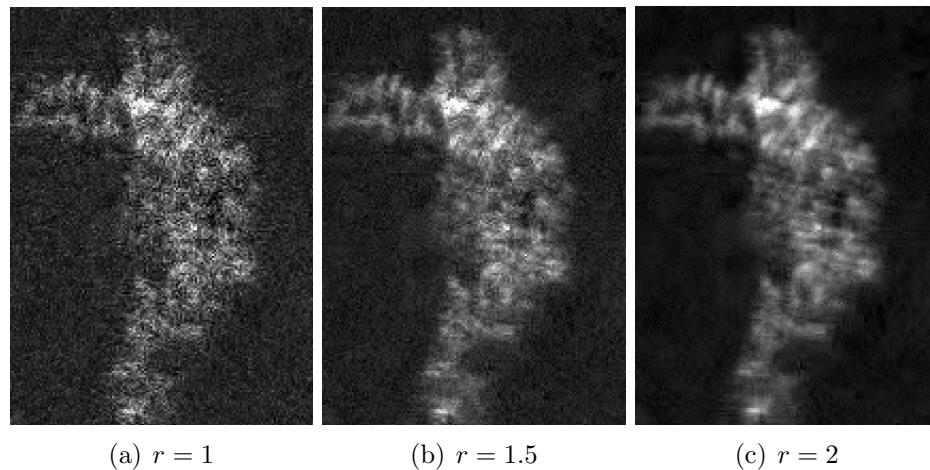


FIGURE 5.16 – Comparaison des cartes chimiques du samarium pour $K = 2$ et $r \in \{1, 1.5, 2\}$ après débruitage réalisé sur les pics isolés

Malgré l'apparition de quelques plans atomiques qui sont propres à la carte chimique du cérium, nous obtenons finalement des résultats plutôt satisfaisants. Le paramètre $r = 1.5$ semble renvoyer la meilleure carte.

Ces résultats nous mènent à penser que la qualité des cartes chimiques d'éléments en faible quantité dépend encore plus qu'habituellement de la sélection du *fond* (voir section 5.1). En effet, en isolant les pics, le fond semble plus facile à identifier et donc à débruiter, ce qui permet de mieux conserver la précieuse information du samarium.

Carte du samarium isolée

Dans cette partie, nous avons testé une nouvelle approche : nous avons construit la carte chimique du samarium à partir des données bruitées. Puis nous avons débruité l'image obtenue avec BM3D. Nous avons réalisé ce débruitage pour différentes valeurs de `sigma` et les résultats obtenus sont donnés dans la figure suivante. L'écart-type estimé automatiquement et réduit vaut 22.34.

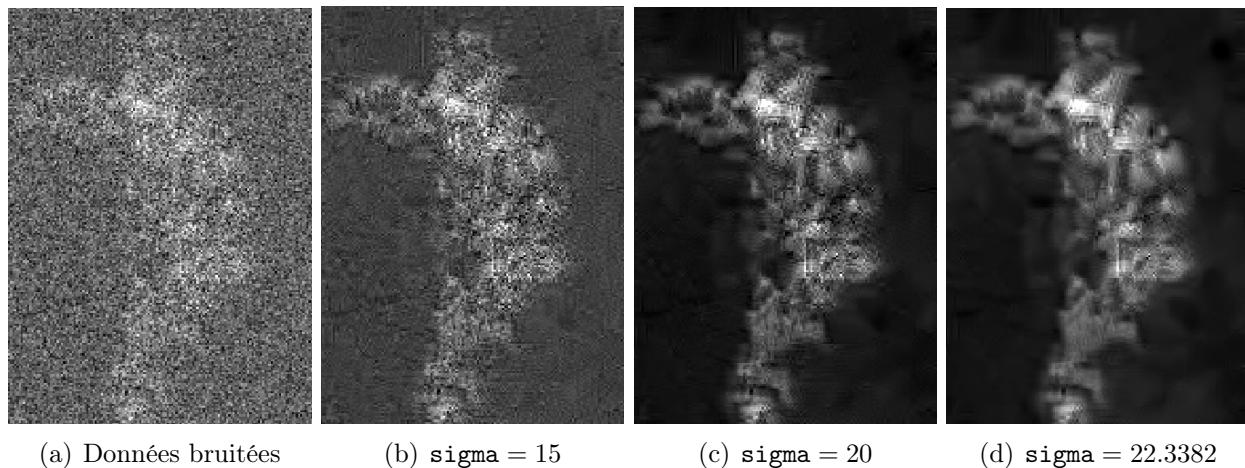


FIGURE 5.17 – Comparaison des cartes chimiques du samarium pour différents σ après débruitage réalisé sur la carte isolée

Ces résultats semblent beaucoup dépendre de la carte chimique initiale qui est alors très variable car le fond est encore trop bruité lorsqu'elle est construite. Par ailleurs, nous ne retrouvons pas les plans atomiques de la carte du cérium mais les cartes sont moins nettes que précédemment.

Il semblerait que ce n'est pas lors de la construction de la carte chimique que l'artefact apparaît, il semble introduit par le processus de débruitage et plus précisément lors du débruitage du fond. En effet, nous pensons que c'est ce dernier qui contient l'information rémanente du cérium qui vient polluer la carte du samarium.

Dans la figure 5.18, nous construisons la carte chimique du fond en sélectionnant la zone qui précède les pics du samarium, nous indiquons cette zone à l'aide de la sous-figure (c). En sous-figure (a), nous représentons la carte chimique du cérium afin d'identifier les ressemblances.

Il est maintenant clair que l'artefact apparaît car le fond contient des informations du cérium. La suppression du fond ne semble pas être suffisante pour s'en débarrasser. Ces informations indésirables seraient alors mélangées aux pics du samarium, ce qui expliquerait l'inefficacité de la suppression du fond. De plus, comme l'artefact ne semble pas apparaître dans le cas du débruitage de la carte isolée, nous nous demandons si l'ACP ne mélangerait pas les informations et reconstruirait des pics du samarium pollués par des bouts de cérium.

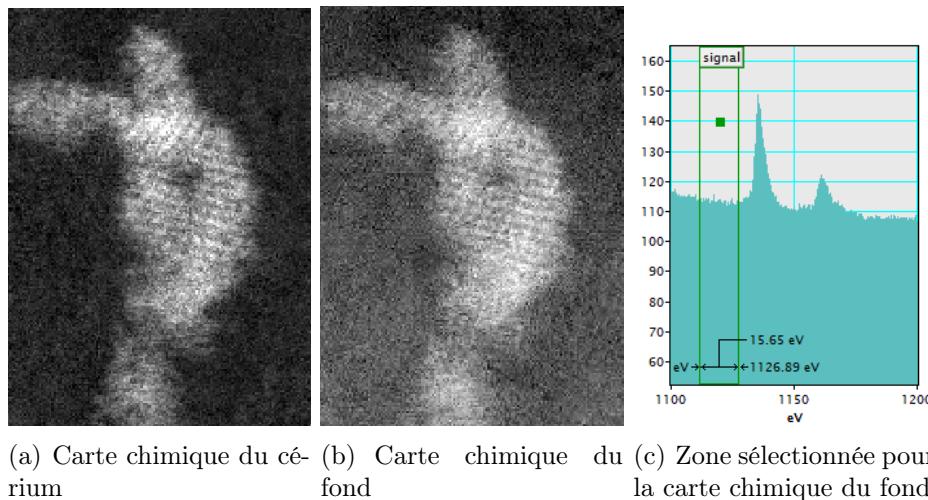


FIGURE 5.18 – Comparaison des cartes chimiques du samarium pour différents sigma après débruitage réalisé sur la carte isolée

Transformée d'Anscombe classique et généralisée

Nous avons réalisé plusieurs tests avec transformée d'Anscombe classique et généralisée mais les résultats sont très similaires voir identiques à ceux obtenus sans transformée. Nous avons donc décidé de ne pas les faire apparaître dans ce rapport car ils n'apportent aucune amélioration significative.

Visualisation des spectres

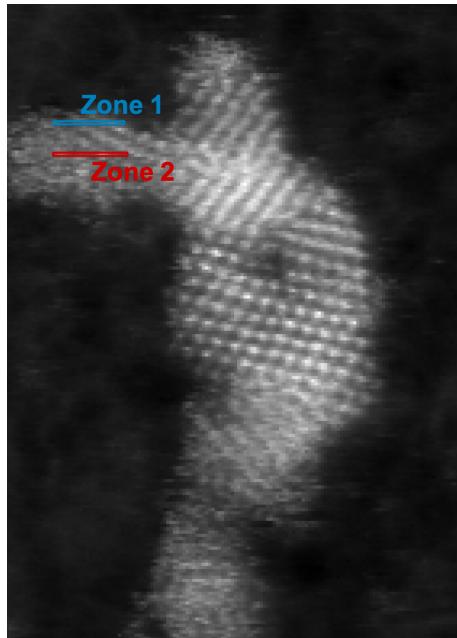
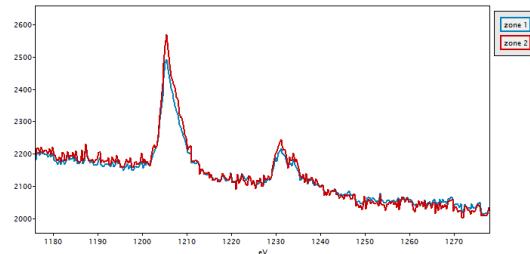


FIGURE 5.19 – Identification des zones sélectionnées pour les spectres

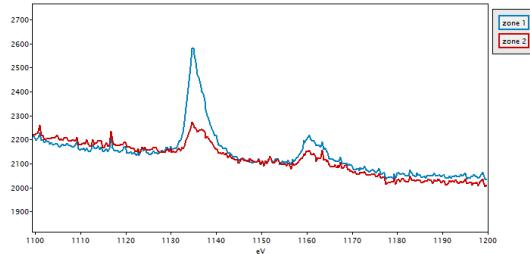
Dans les figures précédentes nous cherchons à visualiser les effets spectraux du débruitage de A-BM3D. Dans la figure 5.20, la sous-figure (a) correspond au Spim débruité dans son intégralité avec $K = 214$ et $r = 1.5$ (voir figure 5.14). La sous-figure (b) correspond au Spim débruité lorsque nous avons isolé les pics du samarium avec $K = 2$ et $r = 1.5$ (voir figure 5.16).

D'après [19], dans cet échantillon, nous sommes censés trouver une plus grande quantité de samarium sur la surface de la particule. À travers l'étude des spectres, nous pouvons conclure la même chose qu'avec l'analyse des cartes chimiques : le débruitage classique ne permet pas de retrouver la carte du samarium car nous n'observons aucune différence d'intensité entre les deux zones. Cependant, lorsque nous isolons les pics du samarium, nous obtenons un net écart d'intensité : elle est élevée dans la zone 1 et plus faible dans la zone 2.

Par ailleurs, nous notons la présence de bruit qui n'a pas été supprimé, notamment le long du fond. Cela nous mène à penser que le débruitage spectral est insuffisant.



(a) Débruitage classique : sur le Spim entier sans transformée d'Anscombe



(b) Débruitage sur les pics du Sm isolés

FIGURE 5.20 – Comparaison des spectres des zones 1 et 2 avec deux types de débruitage

Conclusion

Le principal outil de débruitage d'IHS initialement utilisé par les microscopistes était l'ACP. Bien que très efficace pour le débruitage spectral, cette méthode présente des limites comme parfois l'introduction d'artefacts.

L'association de l'ACP et de BM3D s'est faite naturellement. En effet, d'après des résultats obtenus par l'équipe de Nicolas DOBIGEON sur le débruitage d'IHS obtenues par télédétection, l'utilisation d'une approche séquentielle représentait le choix le plus intéressant.

Nous avons également tenté d'adapter notre approche aux données de l'équipe STEM en introduisant la transformée d'Anscombe afin de traiter le bruit poissonnien et le bruit mixte. Comme elle n'a apporté aucune amélioration, nous avons fini par ne plus l'utiliser.

Finalement, notre algorithme **A-BM3D** semble être efficace et plus spécifiquement pour le débruitage spatial. Les cartes chimiques construites sont de meilleure qualité. Cependant, le débruitage des spectres n'est pas optimal. En effet, ils sont principalement débruités par l'ACP ce qui est limitant car il arrive parfois que du bruit soit contenu dans les premières composantes principales que nous souhaitons conserver. Ce bruit sera alors toujours présent et BM3D ne le supprimera pas. Cela nous paraît cohérent car BM3D est une approche standard de débruitage spatial.

Un débruitage trop léger des spectres peut poser problème car certains microscopistes souhaitent étudier en détails leur allure. Une possibilité serait d'utiliser deux algorithmes : un premier comme **A-BM3D** qui réaliseraient un débruitage global mais pas parfait, puis, lorsque c'est nécessaire, un deuxième afin de débruiter les spectres un-à-un ce qui permettrait alors une analyse plus fine des seuils caractéristiques.

L'utilisation de l'ACP alliée à BM3D reste assez classique : dans la majorité des cas, la sélection du nombre de composantes est réalisée avec la "règle du coude" sur les pourcentages de variance tout en prenant en compte le nombre d'éléments chimiques présents dans l'échantillon. Cependant, dans certains cas spécifiques, il est important de conserver un grand nombre de composantes afin de ne pas perdre les petites informations, ce qui ne garantie pas non plus un résultat forcément positif comme pour la carte du samarium.

Par la suite, nous avons réalisé de nouveaux tests avec une méthode jointe : 3D NL-Means. Comme BM3D, elle repose sur la construction de patchs mais, cette fois-ci, en trois dimensions.

Les résultats obtenus ne sont pas significativement meilleurs, c'est pour cela que nous avons décidé de ne pas les intégrer à ce rapport afin de ne pas le surcharger.

Dans le tableau suivant, nous réalisons un récapitulatif des points forts et points faibles des méthodes que nous avons évoqué dans le rapport.

<i>méthode</i>		ACP	A-BM3D	3D NL-Means
performance	spatiale	—	++	+
	spectrale	+	+	+
mémoire		—	—	—
temps		++	+	+
robustesse (artefacts)		+	+	—

TABLE 5.1 – Bilan des avantages et inconvénients des méthodes utilisées ou évoquées

Dans des travaux futurs, nous pouvons imaginer l'utilisation d'une autre approche standard de débruitage spatial 2D comme NL-Bayes ou NL-Means à la place de BM3D. Cela nous indiquerait si une de ces méthodes se détache des autres dans le cadre des Spims.

Une autre possibilité serait de remplacer l'ACP avec une autre approche par sous-espaces. En effet, c'est l'étape qui semble limiter le plus l'amélioration des résultats. Par exemple, dans [20], ils suggèrent d'utiliser une décomposition de tenseur.

De nombreuses méthodes basées sur le Machine Learning ou le Deep Learning se développent également et il serait intéressant de les tester sur les données de l'équipe STEM.

Table des figures

1.1	Schéma représentant les signaux émis lors de l'interaction d'un faisceau électro-nique avec la matière	5
1.2	Exemples de spectres du seuil Ti-L _{2,3} pour différents temps d'acquisition	6
1.3	Photos du ChromaTEM, un des microscopes STEM de l'équipe	6
1.4	Schéma d'un hypercube	7
1.5	Schéma récapitulatif des différentes approches de débruitages	10
2.1	Comparaison de l'ACP normée et de la non normée avec la carte du samarium et en gardant 3 composantes	13
2.2	Schéma de l'algorithme BM3D	16
2.3	Illustration des étapes de regroupement de BM3D, étapes 1.a.i. et 2.a.i	16
4.1	Présentation du Spim synthétique : première bande et spectres	22
4.2	Présentation des pics caractéristiques des éléments du Spim synthétique	22
4.3	Image des premières bandes des Spims synthétiques bruités	23
4.4	Spectres de différents pixels des Spims synthétiques bruités	23
4.5	Erreur quadratique moyenne pour différents paramètres <i>sigma</i> et <i>K</i> dans le cas du bruit blanc gaussien additif	26
4.6	Erreur quadratique moyenne pour différents paramètres <i>sigma</i> et <i>K</i> dans le cas du bruit poissonnien	28
4.7	Pourcentage de variance des composantes de l'ACP dans le cas du bruit poissonnien	31
4.8	Pourcentage de variance des composantes de l'ACP dans le cas du bruit mixte	31
5.1	Exemple de sélection de zones pour créer une carte chimique du Sm	33
5.2	Exemple de variation de sélection du fond d'un pixel sur l'échantillon de titane	34
5.3	Somme de tous les spectres du Spim 11, données bruitées	35
5.4	Comparaison des cartes chimiques de l'azote pour différents <i>K</i>	35
5.5	Comparaison des cartes chimiques de l'azote pour <i>K</i> = 8 et <i>r</i> ∈ {1, 8}	36
5.6	Présentation des pixels blancs qui apparaissent après débruitage	36
5.7	Comparaison des cartes chimiques de l'azote pour différents <i>K</i> et <i>r</i> avec transformée d'Anscombe	37
5.8	Comparaison des cartes chimiques de l'azote avec et sans transformée d'Anscombe	38

5.9	Comparaison des cartes chimiques de l'azote pour différents K et r avec transformée d'Anscombe généralisée	39
5.10	Comparaison des spectres d'un pixel avant et après débruitage	40
5.11	Comparaison des spectres d'un pixel avec un débruitage par ACP seule et avec un débruitage par A-BM3D	40
5.12	Somme de tous les spectres de l'échantillon SmCeO ₂ , données bruitées	42
5.13	Comparaison des cartes chimiques du samarium pour différents K	43
5.14	Comparaison des cartes chimiques du samarium pour $K = 214$ et $r \in \{1, 1.5\}$. .	43
5.15	Comparaison des cartes chimiques du samarium pour différents K après débruitage réalisé sur les pics isolés	44
5.16	Comparaison des cartes chimiques du samarium pour $K = 2$ et $r \in \{1, 1.5, 2\}$ après débruitage réalisé sur les pics isolés	45
5.17	Comparaison des cartes chimiques du samarium pour différents <code>sigma</code> après débruitage réalisé sur la carte isolée	46
5.18	Comparaison des cartes chimiques du samarium pour différents <code>sigma</code> après débruitage réalisé sur la carte isolée	47
5.19	Identification des zones sélectionnées pour les spectres	48
5.20	Comparaison des spectres des zones 1 et 2 avec deux types de débruitage	48
B.1	Erreurs calculées sur le Spim synthétique dans le cas de bruit gaussien additif .	56
B.2	Erreurs calculées sur le Spim synthétique dans le cas de bruit poissonnien	57
B.3	Erreurs calculées sur le Spim synthétique dans le cas de bruit mixte	58
C.1	Erreur quadratique moyenne pour différents paramètres <code>sigma</code> et K dans le cas du bruit mixte	60
D.1	Somme de tous les spectres de l'échantillon STO Kuro Camera, données bruitées	62
D.2	Comparaison des cartes chimiques du titane pour différents temps d'acquisition avec $K \in \{1, 2\}$ et $r \in \{1, 5, 7\}$	64

Bibliographie

- [1] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3) :425–455, 1994.
- [2] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D : Nonlinear Phenomena*, 60(1) :259–268, 1992.
- [3] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3) :492–526, 2009.
- [4] P. Blomgren and T. F. Chan. Color TV : total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3) :304–309, 1998.
- [5] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2 :60–65, 2005.
- [6] M. Lebrun, A. Buades, and J.-M. Morel. A nonlocal bayesian image denoising algorithm. *SIAM Journal on Imaging Sciences*, 6(3) :1665–1688, 2013.
- [7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8) :2080–2095, 2007.
- [8] A.A. Green, M. Berman, P. Switzer, and M.D. Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1) :65–74, 1988.
- [9] J.M. Bioucas-Dias and J.M.P. Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8) :2435–2445, 2008.
- [10] B. Rasti, J.R. Sveinsson, M.O. Ulfarsson, and J.A. Benediktsson. Hyperspectral image denoising using 3D wavelets. *IEEE International Geoscience and Remote Sensing Symposium*, pages 1349–1352, 2012.
- [11] Y. Qian, Y. Shen, M. Ye, and Q. Wang. 3-d nonlocal means filter with noise estimation for hyperspectral imagery denoising. *IEEE International Geoscience and Remote Sensing Symposium*, pages 1345–1348, 2012.
- [12] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Transactions on Image Processing*, 22(1) :119–133, 2013.

- [13] S. Mallat. Une exploration des signaux en ondelettes. *Les Éditions de l'École Polytechnique*, pages 456–457, 2000.
- [14] F.J. Anscombe. The transformation of poisson, binomial and negative-binomial data. *Bio-metrika*, 35(3-4) :246–254, 1948.
- [15] M. Mäkitalo and A. Foi. A closed-form approximation of the exact unbiased inverse of the anscombe variance-stabilizing transformation. *IEEE Transactions on Image Processing*, 20(9) :2697–2698, 2011.
- [16] M. Mäkitalo and A. Foi. Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise. *IEEE Transactions on Image Processing*, 22(1) :91–103, 2013.
- [17] J.L. Starck, F. Murtagh, and J. Fadili. Sparse image and signal processing : Wavelets and related geometric multiscale analysis. *Cambridge University Press*, 2015.
- [18] R.F. Egerton. Electron energy-loss spectroscopy in the electron microscope. *Springer Science & Business Media*, pages 269–272, 1996.
- [19] S.-Y. Chen, R.-J. Chen, W. Lee, C.-L. Dong, and A. Gloter. Spectromicroscopic evidence of interstitial and substitutional dopants in association with oxygen vacancies in Sm-doped ceria nanoparticles. *Phys. Chem. Chem. Phys.*, 6(3) :3274–3281, 2014.
- [20] J. Spiegelberg, J. Rusz, and K. Pelckmans. Tensor decompositions for the analysis of atomic resolution electron energy loss spectra. *Ultramicroscopy*, 175 :36–45, 2017.

Annexe A

Analyse en composantes principales normée

Pour réaliser une ACP normée, il suffit de modifier notre matrice Y de la forme suivante.

$$Y_{\text{cr}} = \frac{Y - \bar{Y}_b}{\sigma_b} = \left(\frac{Y_p^b - \bar{Y}_b}{\sigma_b} \right)_{1 \leq p \leq P, 1 \leq b \leq B} \quad (\text{A.1})$$

où la moyenne sur les colonnes est donnée par :

$$\bar{Y}_b = \frac{1}{P} \sum_{p=1}^P Y_p^b \quad (\text{A.2})$$

et la variance sur les colonnes par :

$$\sigma_b^2 = \frac{1}{P} \sum_{p=1}^P (Y_p^b - \bar{Y}_b)^2 \quad (\text{A.3})$$

Ensuite les calculs sont les mêmes que pour l'ACP classique en prenant Y_{cr} au lieu de Y_c .

Annexe B

Tableaux d'erreurs calculées sur les données synthétiques

méthodes	ACP	ACP + BM3D	ACP + BM3D	ACP	ACP + BM3D	ACP + BM3D
<i>estimation bruit</i>	X	19.4348	X	X	19.4348	X
<i>paramètre BM3D</i>		0.4096	1.008		0.4096	1.008
nb. de composantes	3 composantes			4 composantes		
EQM	18.565	18.1408	17.9731	1.3426	0.68047	0.4982
MSNR _B	47.2137	47.8758	48.2805	54.3886	57.2541	58.5757
MSNR _P	46.4946	46.9506	47.2861	54.4155	56.681	57.7413
MSSIM	0.99915	0.99915	0.99915	0.99929	0.9993	0.9993
nb. de composantes	5 composantes			6 composantes		
EQM	1.8961	0.72458	0.50175	2.4473	0.77007	0.50496
MSNR _B	52.8449	56.9192	58.5346	51.6683	56.6114	58.4987
MSNR _P	52.4983	56.286	57.6955	51.061	55.91	57.654
MSSIM	0.99929	0.9993	0.9993	0.99928	0.9993	0.9993
nb. de composantes	10 composantes			32 composantes		
EQM	4.6366	0.93443	0.51727	16.328	1.7373	0.58272
MSNR _B	48.6566	55.63	58.3641	42.8976	52.6834	57.7283
MSNR _P	47.6464	54.7914	57.5024	41.665	51.5929	56.7858
MSSIM	0.99926	0.99929	0.9993	0.99914	0.99929	0.9993
nb. de composantes	100 composantes			221 composantes		
EQM	50.0062	3.3902	0.77257	103.6927	4.9191	1.071
MSNR _B	37.9437	49.6697	56.3214	34.7537	48.026	54.7996
MSNR _P	36.6754	48.4547	55.2568	33.4795	46.7747	53.6444
MSSIM	0.99879	0.99927	0.9993	0.99823	0.99925	0.99929

FIGURE B.1 – Erreurs calculées sur le Spim synthétique dans le cas de bruit gaussien additif

méthodes	ACP	ACP + BM3D	ACP + BM3D	ACP + BM3D + AT	ACP + BM3D + AT
estimation bruit	×	20.5267	×	0.97163	×
paramètre BM3D		0.43145	1.27	0.52671	1.27
<i>nb. de composantes</i>		2 composantes			
EQM	114.0282	113.7166	113.3723	114.5447	114.2879
MSNR _B	38.7945	38.8909	39.0061	38.8904	38.9662
MSNR _P	40.6751	40.7684	40.8738	38.6947	38.7233
MSSIM	0.99836	0.99836	0.99837	0.99834	0.99834
<i>nb. de composantes</i>		3 composantes			
EQM	19.1574	18.58	18.2017	21.0445	20.7671
MSNR _B	46.5016	47.1384	47.7978	45.768	46.0599
MSNR _P	45.7541	46.206	46.7385	46.7938	47.0762
MSSIM	0.99915	0.99915	0.99915	0.99912	0.99912
<i>nb. de composantes</i>		4 composantes			
EQM	2.0024	1.1497	0.76103	1.8399	1.5764
MSNR _B	52.8356	55.3103	56.9807	53.9951	55.0239
MSNR _P	52.8967	54.8777	56.2345	53.6836	54.4639
MSSIM	0.99929	0.99929	0.9993	0.99929	0.99929
<i>nb. de composantes</i>		5 composantes			
EQM	2.973	1.4462	0.76445	1.1032	0.84040
MSNR _B	51.1018	54.2255	56.9512	55.295	56.5612
MSNR _P	52.0035	54.3973	56.2213	54.8737	55.8277
MSSIM	0.99928	0.99929	0.9993	0.99929	0.9993
<i>nb. de composantes</i>		6 composantes			
EQM	3.935	1.7497	0.76834	1.1592	0.84451
MSNR _B	49.7785	53.2994	56.9194	55.015	56.5315
MSNR _P	51.2428	53.9461	56.2061	54.5754	55.7976
MSSIM	0.99927	0.99929	0.9993	0.99929	0.9993
<i>nb. de composantes</i>		7 composantes			
EQM	4.8844	2.0285	0.77282	1.2144	0.84828
MSNR _B	48.7411	52.5821	56.8837	54.7726	56.5054
MSNR _P	50.589	53.5542	56.1886	54.3019	55.7705
MSSIM	0.99926	0.99929	0.9993	0.99929	0.9993
<i>nb. de composantes</i>		41 composantes			
EQM	34.9606	9.7627	0.93449	2.7493	0.96531
MSNR _B	39.6355	45.2597	55.7913	50.7456	55.7619
MSNR _P	43.0268	48.1845	55.5918	49.9676	54.9907
MSSIM	0.99896	0.99921	0.9993	0.99928	0.99929
<i>nb. de composantes</i>		314 composantes			
EQM	215.7358	31.264	1.9321	7.3411	1.8211
MSNR _B	31.6167	40.0973	52.2608	46.332	52.6023
MSNR _P	33.6151	42.4303	52.8061	45.419	51.7295
MSSIM	0.99715	0.99899	0.99928	0.99923	0.99929

FIGURE B.2 – Erreurs calculées sur le Spim synthétique dans le cas de bruit poissonnien

méthodes	ACP	ACP + BM3D	ACP + BM3D	ACP + BM3D + GAT	ACP + BM3D + GAT	ACP + BM3D + GAT
gain	×	×	×	1	1	1
paramètre GAT				30.2937	76	76
estimation bruit		28.7485		0.76983	0.36456	×
paramètre BM3D		0.60353	1.52	0.65763	0.61858	1.52
nb. de composantes	2 composantes					
EQM	114.726	114.1075	113.6565	111.4179	113.0489	112.6237
MSNR _B	38.5948	38.7623	38.9057	38.9188	38.8109	38.9428
MSNR _P	40.4572	40.6216	40.7481	39.8004	40.3898	40.4967
MSSIM	0.99836	0.99836	0.99836	0.99837	0.99837	0.99837
nb. de composantes	3 composantes					
EQM	20.1861	19.0557	18.5729	18.8424	18.8975	18.4696
MSNR _B	45.613	46.5429	47.1928	46.3399	46.5381	47.1243
MSNR _P	45.0638	45.7159	46.2539	46.0815	45.8056	46.3041
MSSIM	0.99914	0.99915	0.99915	0.99914	0.99915	0.99915
nb. de composantes	4 composantes					
EQM	3.365	1.7024	1.2048	1.8815	1.7333	1.2761
MSNR _B	50.4726	53.462	54.8763	53.0415	53.3506	54.581
MSNR _P	50.5169	52.8775	54.0469	52.487	52.7198	53.741
MSSIM	0.99927	0.99929	0.99929	0.99929	0.99929	0.99929
nb. de composantes	5 composantes					
EQM	4.8106	1.9102	1.2127	2.0177	1.8889	1.2843
MSNR _B	48.8866	52.8781	54.8378	52.6496	52.9034	54.5441
MSNR _P	49.2663	52.4803	54.0211	52.106	52.3868	53.7131
MSSIM	0.99926	0.99929	0.99929	0.99929	0.99929	0.99929
nb. de composantes	6 composantes					
EQM	6.2426	2.1308	1.2202	2.1194	2.0668	1.2912
MSNR _B	47.6731	52.3388	54.8015	52.3657	52.4366	54.5138
MSNR _P	48.2427	52.0847	53.9959	51.8126	52.0276	53.6887
MSSIM	0.99925	0.99929	0.99929	0.99928	0.99929	0.99929
nb. de composantes	7 composantes					
EQM	7.6723	2.3564	1.228	2.2313	2.2607	1.2989
MSNR _B	46.6995	51.8352	54.7643	52.0911	51.9895	54.479
MSNR _P	47.3866	51.7095	53.9687	51.5419	51.6728	53.662
MSSIM	0.99923	0.99928	0.99929	0.99928	0.99928	0.99929
nb. de composantes	40 composantes					
EQM	52.2308	7.8743	1.483	5.2764	6.8207	1.5375
MSNR _B	37.8578	46.1616	53.7162	47.9071	46.7886	53.5487
MSNR _P	39.0667	46.9392	53.1947	47.276	47.1328	52.9169
MSSIM	0.99878	0.99923	0.99929	0.99925	0.99924	0.99929
nb. de composantes	148 composantes					
EQM	180.2615	17.3683	2.177	10.7522	14.7903	2.2394
MSNR _B	32.3886	42.6378	51.8173	44.6914	43.3281	51.69
MSNR _P	33.0815	43.2446	51.5348	43.9608	43.5563	51.2335
MSSIM	0.99749	0.99913	0.99928	0.9992	0.99916	0.99928

FIGURE B.3 – Erreurs calculées sur le Spim synthétique dans le cas de bruit mixte

Annexe C

Simulation de bruit mixte

Pour construire nos données bruitées dans le cas du bruit mixte, nous combinons les deux formules (4.6) et (4.7) utilisées précédemment. Ainsi nous obtenons comme formule finale :

$$Y_p^b = \alpha \mathcal{P}(X_p^b) + \varepsilon_p^b \quad \text{pour tout } p \in \llbracket 1, P \rrbracket \text{ et } b \in \llbracket 1, B \rrbracket \quad (\text{C.1})$$

où ε_p^b suit une loi $\mathcal{N}(0, \sigma_\varepsilon^2 = 20^2)$ pour tout p et b et α est un réel strictement positif. Nous l'avons pris égal à 1.

Dans cette partie, il faut réaliser deux estimations de la variance du bruit. Dans un premier temps, nous estimons l'écart-type du bruit afin d'appliquer la transformée d'Anscombe généralisée. Ce paramètre vaut 30.3. Dans un second temps, nous estimons l'écart-type du bruit afin de le donner en paramètre à BM3D. Sans transformée d'Anscombe généralisée, il vaut 28.7 et réduit 0.60. Avec transformée d'Anscombe généralisée de paramètre 30.3, il vaut 0.77 et réduit 0.66.

Dans la suite, les résultats présentés sont obtenus lorsque la transformée d'Anscombe généralisée de paramètre 30.3 est appliquée, sauf indication contraire. Dans la figure suivante, nous représentons l'EQM en fonction du paramètre `sigma` et du nombre de composantes K .

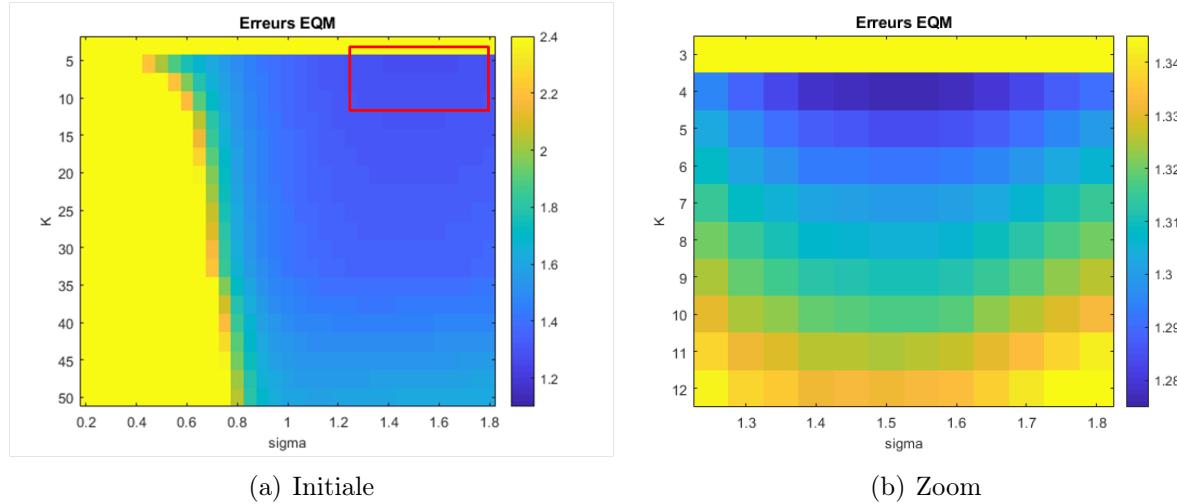


FIGURE C.1 – Erreur quadratique moyenne pour différents paramètres `sigma` et K dans le cas du bruit mixte

La sous-figure (a) correspond à l'EQM pour K appartenant à $\{3, 4, \dots, 15\} \cup \{20, 25, \dots, 50\}$ et pour `sigma` variant dans $\{0.2, 0.25, \dots, 1.8\}$.

La sous-figure (b) correspond au zoom de l'encadré rouge de la sous-figure (a). Il s'agit de l'EQM pour K appartenant à $\{3, 4, \dots, 12\}$ et pour `sigma` variant dans $\{1.25, 1.3, \dots, 1.8\}$.

L'erreur minimale est obtenue pour $K = 4$ composantes et pour `sigma` compris entre 1.45 et 1.6. En modifiant `sigma` à la main, nous avons trouvé une valeur permettant d'obtenir de meilleurs résultats. Elle vaut 1.52.

Ensuite, nous calculons les erreurs dans ces cinq cas de figure :

- ▶ après ACP sans transformée d'Anscombe généralisée,
- ▶ après ACP et BM3D sans transformée d'Anscombe généralisée,
- ▶ après ACP et BM3D avec transformée d'Anscombe généralisée et les estimations réalisées automatiquement,
- ▶ après ACP et BM3D avec transformée d'Anscombe généralisée où `sigma_ANS` = 76 et `sigma` calculé automatiquement,
- ▶ après ACP et BM3D avec transformée d'Anscombe généralisée où `sigma_ANS` = 76 et `sigma` = 1.52.

qui sont des valeurs de paramètres qui nous paraissent pertinentes grâce à la figure C.1. Elles sont alors résumées dans le tableau suivant.

<i>méthode</i>	ACP	A-BM3D			
Anscombe	×	×	✓	✓	✓
sigma_ANS	×	×	30.3	76	76
sigma	×	0.60	0.66	0.62	1.52
<i>K</i>		<i>3 composantes</i>			
EQM	20.186	19.056	18.842	18.898	18.470
<i>K</i>		<i>4 composantes</i>			
EQM	3.365	1.702	1.882	1.733	1.276
<i>K</i>		<i>5 composantes</i>			
EQM	4.811	1.910	2.018	1.889	1.284
<i>K</i>		<i>40 composantes</i>			
EQM	52.231	7.874	5.276	6.821	1.538
<i>K</i>		<i>148 composantes</i>			
EQM	180.262	17.368	10.752	14.790	2.239

TABLE C.1 – Erreurs EQM pour du bruit mixte

Nous remarquons qu'avec ou sans transformée d'Anscombe généralisée, les meilleurs résultats sont obtenus pour $K = 4$ composantes.

L'erreur la plus basse est obtenue pour **sigma_ANS** = 76 et **sigma** = 1.52. Cela reviendrait alors à multiplier l'estimation initiale **sigma** = 0.62 par un ratio de $r = 2.45$.

Si nous appliquons une ACP et BM3D sans transformée d'Anscombe avec ce même **sigma**, nous obtenons une erreur encore plus petite qui vaut **1.205** (résultat non présenté ici, voir Annexe B.3). L'erreur **1.276** en est assez proche et est suffisamment petite, nous ne devrions donc pas noter de différence visuelle.

Contrairement à la transformée d'Anscombe classique, notons que la forme généralisée permet une amélioration des erreurs quasi-systématique. En effet, lorsque $K \leq 4$, la transformée d'Anscombe classique semblait détériorer les résultats, tandis que pour $k \geq 5$, elle semblait les améliorer.

Cependant, l'amélioration apportée par la transformée d'Anscombe généralisée n'est pas considérable pour $K \leq 7$ (voir Annexe B.3).

Comme attendu, la transformée d'Anscombe n'apporte pas de nette amélioration. Par conséquent, nous ne l'utiliserons pas beaucoup dans la suite des tests.

Annexe D

Autre exemple de débruitage sur les données réelles

Dans cette partie, nous allons débruiter trois hypercubes issus d'un même échantillon de SrTiO_3 nommé STO Kuro Camera. Nous nous intéressons plus particulièrement au titane (Ti). Ces trois Spims ont été construit avec trois temps d'acquisition différents : 1 ms, 2 ms et 3 ms. La résolution est de l'ordre de 0.025 nm par pixel. Son pic caractéristique se situe à 456 eV. À nouveau, nous traçons la courbe de la somme des spectres de chacun des trois hypercubes bruités.

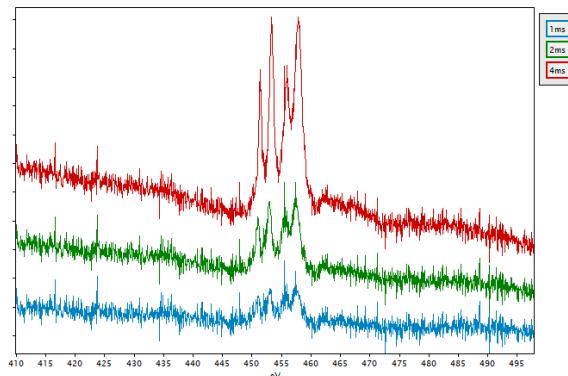


FIGURE D.1 – Somme de tous les spectres de l'échantillon STO Kuro Camera, données bruitées

Au vu des premiers résultats, nous avons conclu qu'il n'était pas nécessaire de garder beaucoup de composantes principales lors de l'ACP. En effet, au-delà de $K = 2$ composantes, les cartes chimiques sont assez similaires. Nous allons donc nous contenter d'étudier les cas $K = 1$ et $K = 2$. Nous n'appliquerons pas non plus de transformée d'Anscombe qu'elle soit classique ou généralisée au vu de son inefficacité.

Pour 1 ms, le paramètre `sigma` de BM3D vaut 5.94. Pour 2 ms, il vaut 3.68. Enfin, pour 4 ms, il vaut 2.52.

Au niveau des spectres, il n'y a pas de différence notable pour les trois temps d'acquisition.

Nous allons donc étudier les cartes chimiques du titane pour différents K et r . Les cartes obtenues sont données dans la figure D.2.

Nous remarquons que pour cet échantillon, la multiplication du paramètre de BM3D est nécessaire. Si nous le laissons tel quel, les cartes restent très (trop) bruitées et BM3D ne semble rien apporter de plus après l'ACP.

Pour le temps d'acquisition de 1 ms, nous ne retrouvons pas nettement les colonnes atomiques. De plus, quelques déformations de structures apparaissent. Cela signifie que le temps d'acquisition est trop bas et cela introduit trop de bruit que **A-BM3D** ne parvient pas à supprimer correctement.

En augmentant le temps d'acquisition, les colonnes atomiques sont plus nettes. Les cartes ne sont tout de même pas très claires et il est difficile de les analyser. Finalement, nous pouvons conclure que dans ce cas de figure il faut éviter de descendre en dessous d'un temps d'acquisition de 4 ms.

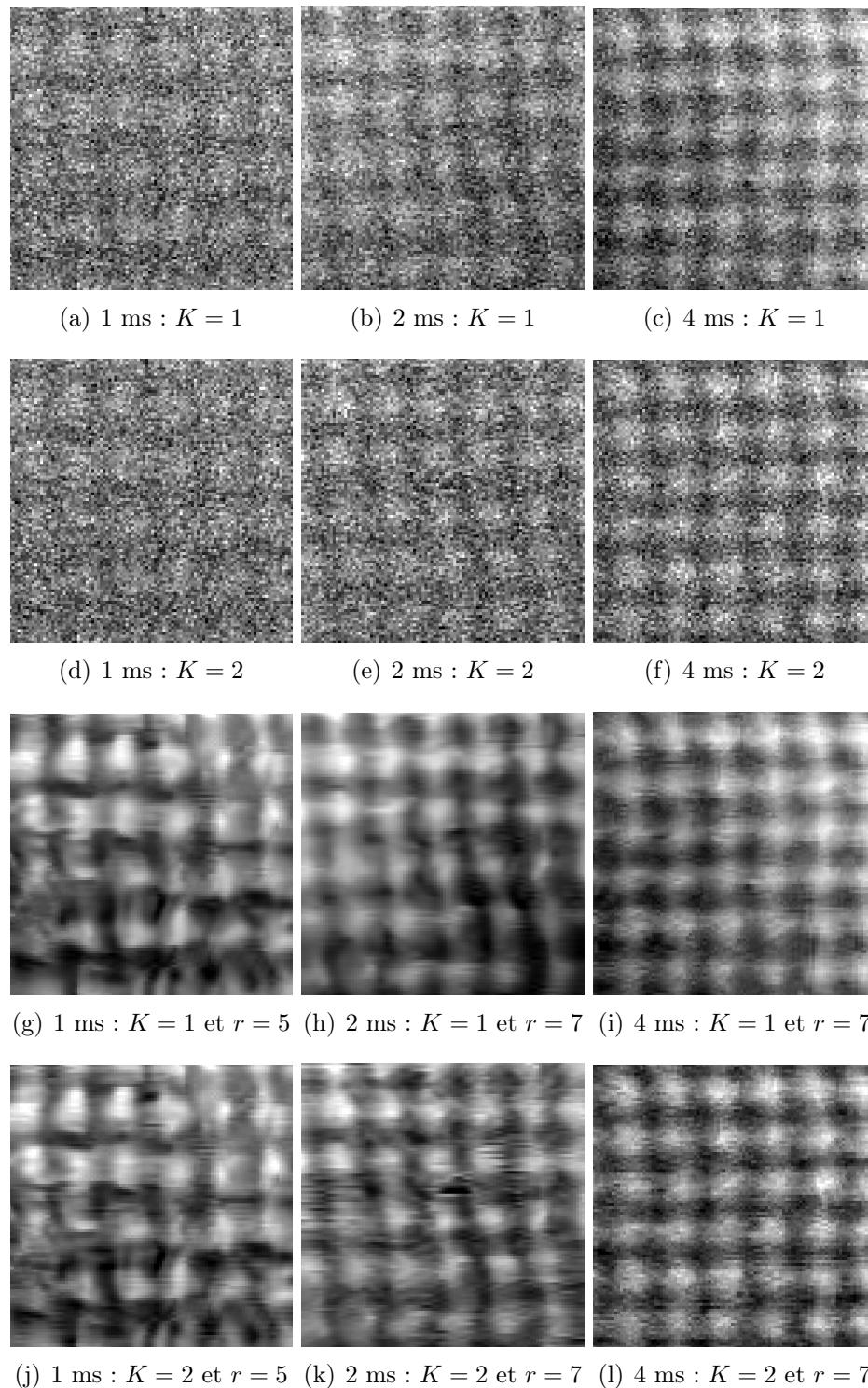


FIGURE D.2 – Comparaison des cartes chimiques du titane pour différents temps d’acquisition avec $K \in \{1, 2\}$ et $r \in \{1, 5, 7\}$