

Nouvelles images et matériaux anciens : une intersection non-singulière

Serge X. COHEN¹

¹IPANEMA USR3461 CNRS/MCC
L'Orme des Merisiers, Saint-Aubin - BP48, 91192 Gif-sur-Yvette cedex, FRANCE
serge.cohen@synchrotron-soleil.fr

Résumé – Bien que très divers, les matériaux anciens, qu'ils soient étudiés dans le domaine du patrimoine culturel, de l'archéologie ou de la paléontologie, possèdent des caractéristiques communes quand à leur étude à travers les techniques de caractérisation physico-chimique. Leur nature hétérogène multi-échelle, le fait qu'ils soient le produit de plusieurs processus avec des temps caractéristiques très différents, la faible connaissance *a priori* que nous en avons initialement en font des sujets idéaux pour les techniques d'imagerie spectrale, surfacique et/ou volumique. Ces matériaux étant non reproductibles en laboratoire il devient impossible de les optimiser pour simplifier l'exploitation des mesures, et le développement de méthode d'analyse des données devient le levier principal pour extraire autant d'information que possible des échantillons. À travers de quelques exemples nous essaierons de formaliser les problèmes posés et dégager des pistes pour y répondre.

Abstract – Despite a strong diversity, ancient materials whether they are studied in cultural heritage, archaeology or palaeontology, share common features in terms of their study through physico-chemical characterisation techniques. Their multi-scale heterogeneity, the fact they are issued from multiple processes with very diverse time scales, the poor prior knowledge we have on them initially make them ideal subjects for many spectral imaging techniques probing the surface or the full volume. As these materials are not reproducible in the laboratory it is not possible to optimise the materials themselves to simplify measurements exploitation, and the development of data-analysis methodology becomes the main leverage left to extract as much information as possible from the samples. With the help of a small set of examples we will try to produce a formalisation of the problems at hand and to draw some leads on the possible answers.

L'intitulé *matériaux anciens* rassemble des matériaux extrêmement diversifiés, abordés dans une approche diachronique. Cette appellation générique regroupe les matériaux :

de la conservation et du patrimoine culturel incluant à la fois les matériaux conservés dans les musées (peinture, sculpture, arts du feu, textiles, etc.), ceux des sites et monuments historiques (notamment le patrimoine bâti), et les matériaux servant à l'archivage du patrimoine matériel et immatériel, matériaux de la restauration, etc. ;

de l'archéologie relevant des artefacts archéologiques : industrie lithique, matériaux céramiques, métaux, verres, textiles, matériaux de l'art rupestre, micro- et macro-restes organiques, etc. ;

de la paléontologie notamment osseux, dentaires ou exosquelettiques, mais également les autres phanères et tissus mous qui peuvent être préservés dans les cas de *conservation exceptionnelle*, etc. ;

paléo-environnementaux où l'on considérera notamment les marqueurs d'informations archéologiques ou paléontologiques sur les milieux d'enfouissement et les processus taphonomiques fondamentaux (milieux terrestres : micro- et macro-restes végétaux, pollens ; milieux marins

et lacustres), et les matériaux d'intérêt paléo-climatique (spéléothèmes, séquences sédimentaires, anneaux de croissance des arbres, coraux profonds), etc.

Malgré la grande diversité de ces matériaux, ils possèdent des caractéristiques communes dans le cadre de leur étude physico-chimique. Ils sont souvent formés et transformés à partir d'ingrédients d'origine naturelle, tant organiques que minérales. Ils peuvent être fortement altérés du fait de leur vieillissement à plus ou moins long terme. Ils possèdent de ce fait quatre particularités essentielles : ils sont hétérogènes / composites à toutes les échelles, le processus de formation n'est jamais reproductible empiriquement, les objets d'étude n'ont souvent pas de sens considérés hors d'un corpus d'objet et ils imposent des conditions particulières de conservation et de manipulation.

Les caractéristiques d'intérêt lors de l'analyse de ces matériaux sont le résultat de l'interaction de l'ensemble de constituants suivant la morphologie des interfaces caractéristiques aux différentes échelles. L'imagerie spectrale, c'est à dire la mesure d'un spectre complet en chaque pixel de l'image, devient un outil de choix pour atteindre simultanément les informations chimiques et spatiales des matériaux étudiés. La faible connaissance *a priori* des objets d'étude impose une approche statistique à l'extraction d'informations à partir des données collectées, le plus souvent suivant une approche *non supervisée*. De

plus l'absence de vérité terrain appelle à rechercher des preuves mathématiques au moins autant qu'une validation empirique du processus de traitement de données; validation qui n'est le plus souvent pas atteignable. La richesse de l'ensemble de ces contraintes motive la construction de nouvelles méthodologies de traitement des données d'imagerie, dont l'efficacité sera rigoureusement prouver et qui seront optimisées suivant la finalité des analyses réalisées.

1 Nouvelles imageries spectrales pour les matériaux anciens

Que ce soit en spectro-microscopie ou d'une façon plus générale lors de l'analyse statistique de données multivariées recueillies sur un corpus de matériaux anciens, les approches statistiques classiques n'ont qu'une utilité limitée du fait du rapport défavorable entre dimension du problème et quantité de données. Il faut prendre en compte ce nombre limité d'échantillons pour produire une estimation aussi réaliste que possible des modèles statistiques ainsi que des segmentations et/ou classifications en découlant, en exploitant une approche non-asymptotique.

Dans la suite, nous noterons \mathcal{S} une image spectrale sur l'ensemble des pixels \mathbf{N} de cardinal $\|\mathbf{N}\| = N = N_1 \times N_2$, et pour tout $n = (n_1, n_2)$, indice de la position d'un pixel, le spectre mesuré est $\mathcal{S}[n]$, une fonction de λ échantillonnée sur $(\lambda_i)_{i \in [1, P]}$. Typiquement N est de l'ordre de quelques centaines à plusieurs dizaines de milliers, et P est compris entre plusieurs centaines et quelques milliers.

Suivant l'échelle d'observation deux modèles «extrémaux» peuvent être envisagés:

1. Quand la taille du pixel est un peu plus grande que la dimension caractéristique des hétérogénéités, chaque pixel est un mélange de ces constituants. L'échantillon est représenté mathématiquement par une fonction \mathcal{A} de $[0, L] \times [0, l] \rightarrow [0, 1]^K$, où L et l sont les dimensions de l'objet et K le nombre de constituants purs et observables par la modalité spectroscopique utilisée. Cette fonction est à variation totale bornée et vérifie ($\forall \mathbf{r} \in [0, L] \times [0, l]$, $\|\mathcal{A}(\mathbf{r})\|_1 \leq 1$), des constituants peuvent être invisibles pour certaines modalités spectroscopiques.
2. Quand la taille du pixel est plus faible que la dimension caractéristique des hétérogénéités, le nombre de classes de spectres observées est alors fini et le matériaux est observé comme l'union de zones quasi-uniformes, représenté par une fonction \mathcal{S} de $[0, L] \times [0, l] \rightarrow [1, K]$ où cette fois K est le nombre de classes de spectres observés.

Même si le premier modèle correspond plutôt à un problème de factorisation de matrice de bas rang alors que le second modèle correspond plus à un problème de classification non-supervisée, dans les deux cas la morphologie de l'image est la même : des zones régulières (voir constantes) avec des discontinuités aux frontières ayant un support rectifiable, et porte

donc une part importante de l'information. La recherche en traitement d'image a montré à quel point, dans le cas d'un faible nombre de canaux, cette contrainte de variation totale bornée est efficace. Il faut maintenant exploiter cette contrainte pour améliorer l'analyse d'images ayant un nombre de canaux très supérieur à ce qui se fait dans l'état de l'art, même en imagerie hyperspectrale, tout en prenant en compte le fait que chaque point de données correspond à l'échantillonnage *fin* d'une courbe, un spectre, très régulière.

2 Représentation : approximation ou discrimination optimale

L'imagerie spectrale évite de choisir en amont les zones spectrales pertinentes. La contrepartie est l'obtention de données dans un espace de très grande dimensions (plusieurs centaines à quelques milliers). Si le bruit de mesure remplit *a priori* relativement uniformément l'espace de mesure, le signal appartient à une variété de bien plus petite dimension. La caractérisation du sous-espace minimum contenant cette variété est un objectif important tant du point de vue théorique, permettant ainsi d'estimer la dimension intrinsèque du problème, que du point de vue pratique, permettant d'avoir un coût de calcul raisonnable. Là où les travaux de réduction de dimension visent généralement une reconstruction optimale au sens de la norme ℓ_2 ou même ℓ_1 , nous visons plus souvent à une optimisation de la structuration de l'ensemble des données avec des critères plus proches de ceux utilisés pour la sélection de variables.

Pour choisir les variables pertinentes en grande dimension quand l'objectif est la discrimination plus que la fidélité de représentation, des algorithmes ont été proposés mais ils butent sur des problèmes de complexité algorithmique lorsque la dimension est trop grande (de l'ordre de quelques dizaines). Ils nécessitent en effet une exploration quasi exhaustive des sous-ensembles possibles de variables. Abordé de cette façon le problème est encore plus complexe lorsqu'on s'autorise un changement de base et n'a pas de solution lorsque le nombre d'échantillons est trop petit devant la dimension. La théorie de l'approximation permet de comprendre comment réduire la dimension sans trop dégrader la représentation. De nombreuses méthodes, analyse en composantes principales, représentations creuses, *K-Singular Value Decomposition* ont été proposées pour trouver des représentations de faible dimension dans lesquelles les données sont représentées efficacement. Il est naturel d'essayer de coupler ces méthodes avec des algorithmes de classification non supervisée.

Les réductions de la dimension de type analyse en composantes principales donnent de bons résultats pratique quand le niveau de bruit est suffisamment bas. L'objectif de discrimination est cependant différent de l'objectif de représentation. Une variable peut être nécessaire pour bien représenter des objets mais inutile pour les discriminer, c'est le cas dans la figure 1 pour les composantes 7 et 15, qui bien qu'étant plus significative en terme d'approximation des spectres sont moins

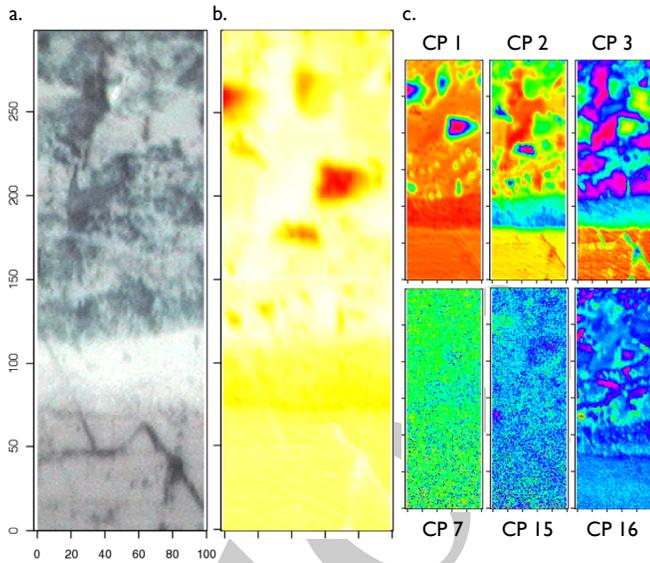


Figure 1: Imagerie spectrale d'absorption des infrarouge d'une stratigraphie issue d'une couche picturale (*UDI board*, National Gallery *Londres*). Pour réduire la dimension du problème une analyse en composante principale a été utilisée et seules les projections sur les 16 premières composantes ont été conservées. De gauche à droite : (a) image visible; (b) énergie totale (l_2) des spectres d'absorption infrarouge allant du blanc pour les faibles énergie au rouge pour les fortes énergie; (c) projections de l'image spectrale sur 6 des 16 premières composantes principales allant du rouge pour les faibles valeurs au violet pour les valeurs élevées de la projection. La cartographie est réalisée sur un échantillonnage uniforme et carré de 192×64 pixels, avant projection les spectres sont échantillonnés sur 829 canaux.

informatives en terme de morphologie que la composante 16. Nous cherchons à exploiter ce potentiel supplémentaire de réduction de dimension à l'aide d'un couplage entre la recherche de l'espace de faible dimension et le problème de classification. Les algorithmes d'approximation visent à s'adapter au mieux aux données observées et favorisent donc très fortement le biais par rapport à la variance. Les conséquences sont faibles dans le cadre de l'approximation mais préjudiciable dans le cadre de la classification ou de l'estimation de la loi de distribution. Ces deux dernières applications requiert en effet des résultats généralisables au delà de la seule réalisation observée.

3 Analyse superficielle et en profondeur

Dans la grande majorité des cas les courbes obtenues en chacun des pixels d'une image spectrale sont reliées à une caractéristique chimique ou physico-chimique moyennée pour un volume élémentaire correspondant à une couche mince à la surface de l'échantillon étudié. L'analyse en profondeur de l'objet requiert alors une préparation invasive, telle que la réalisation

d'un coup transverse pour faire une analyse stratigraphique, et cela même si le procédé de mesure est en lui même non-destructif. Pour les matériaux du patrimoine, tant culturel que naturel, il est en revanche particulièrement intéressant de mettre en place des méthodologie de caractérisation qui ne sont ni destructive ni invasive.

Ainsi il devient utile d'exploiter les relations qui existeraient entre le spectre mesuré en un pixel et la géométrie locale de l'objet, même si cette relation est complexe. Un exemple d'un tel processus est l'exploitation du signal de diffusion inélastique des rayons X, aussi appelé diffusion Compton ou diffusion incohérente, qui est régit par une lois liant la perte d'énergie du photon X à l'angle de diffusion. La mesure en chaque pixel d'un spectre de perte d'énergie doit permettre, dans des conditions qu'il reste à déterminer, de reconstruire l'ensemble du volume illuminé et cela sans même avoir à faire tourner l'échantillon comme cela serait le cas en tomographie. Le processus de formation d'image résultant de ce phénomène est sensiblement plus compliqué que ceux plus classiques du contrastes d'absorption exploiter en radiographie ou en tomographie de rayon X, si bien que la diffusion inélastique est généralement considérée comme une perturbation ou un bruit. En radiographie classique, le processus de formation d'image correspond à la transformée de Radon, ou une transformation dérivée, du volume illuminé si bien que la solution du problème inverse est bien caractérisée. Dans le cas de la diffusion inélastique le processus de formation d'image est déjà plus complexe, mettant en jeu des lignes brisées sur lesquels chaque changement de trajectoire correspond à une perte d'énergie et l'intégration le long de ces lignes des coefficients d'absorption et de diffusion élastique.