Hierarchical clustering of spectral images with spatial constraints for the rapid processing of large and heterogeneous data sets

Gilles Celeux, Serge X. Cohen, Agnès Grimaud, Pierre Gueriau

▶ **To cite this version:**

Gilles Celeux, Serge X. Cohen, Agnès Grimaud, Pierre Gueriau. Hierarchical clustering of spectral images with spatial constraints for the rapid processing of large and heterogeneous data sets. SN Computer Science, 2022, 3 (194), 10.1007/s42979-022-01074-4 . hal-03104488v4

HAL Id: hal-03104488
https://hal.uvsq.fr/hal-03104488v4

Submitted on 10 Apr 2022

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

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

Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Hierarchical clustering of spectral images with spatial constraints for the rapid processing of large and heterogeneous data sets

Gilles Celeux*, Serge X. Cohen†, Agnès Grimaud‡, and Pierre Gueriau§

Abstract.

When dealing with full spectrum images in which each pixel is characterized by a full spectrum, *i.e.* spectral images, standard segmentation methods, such as k-means or hierarchical clustering might be either inapplicable or inappropriate ; one aspect being the multi-GB size of such data set leading to very expensive computations. In the present contribution, we propose an approach to spectral image segmentation combining hierarchical clustering and spatial constraints. On the one hand spatial constraints allow to implement an algorithm with a reasonable computation time to obtain a segmentation and with a certain level of robustness with respect to the signal-to-noise ratio since the *prior* knowledge injected by the spatial constraint partially compensates for the increase in noise level. On the other hand hierarchical clustering provides a statistically sound and known framework that allows accurate reporting of the instrument noise model. In terms of applications, this segmentation problem is encountered particularly in the study of ancient materials that benefits from the wealth of information provided by the acquisition of spectral images. In the last few years, data collection has been considerably accelerated, enabling the characeterization of the sample with a high dynamic range in both the spatial dimensions and composition and leading to an average size of a single data set in the tens of GB range. Hence we also considered computational and memory complexity when developing the herein proposed algorithm. Taking on this application domain, we illustrate the proposed algorithm on a X-ray fluorescence spectral image collected on an *ca.* 100 Myr fossil fish, as well as on simulated data to assess the sensitivity of the results to the noise level. For such experiment, the lower sensitivity to noise simultaneously lead to an increase in the spatial definition of the collected spectral image, thanks to the faster acquisition time, and to a reduction in the potentially harmful radiation dose density to which the samples are subjected.

Key words. Spectral image segmentation , and Ward criterion , and spatial constraint , and ancient material , and X-ray fluorescence

1. Introduction

Spectral imaging, *i.e.* the collection of images for which each pixel is characterized by a full spectrum (see *e.g.* Fig. 1), is a tool of choice for simultaneously obtaining physico-chemical information (*e.g.* elemental, chemical or mineralogical composition), and the

*Inria Saclay-Île-de-France, IMO campus d'Orsay 91405 Orsay.

†IPANEMA, CNRS, ministère de la Culture, UVSQ, MNHN, USR3461, Université Paris-Saclay, 91192 Gif-sur-Yvette, France (serge.cohen@ipanema-remote.fr).

‡Université Paris-Saclay, UVSQ, CNRS, Laboratoire de Mathématiques de Versailles, 78000 Versailles, France.

§Institute of Earth Sciences, University of Lausanne, Géopolis, CH-1015 Lausanne, Switzerland

IPANEMA, CNRS, ministère de la Culture, UVSQ, MNHN, USR3461, Université Paris-Saclay, 91192 Gif-sur-Yvette, France.

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

morphological information essential for describing heterogeneous materials. Spectral images are particularly used to study ancient materials, such as encountered in archaeology and paleontology or as part of the cultural heritage research, which are very diverse but share the particularity of being composite and heterogeneous on several scales [6]. They are also the results of multiple processes at various time scales, inducing strong constraints in terms of handling and physico-chemical characterization whilst often having limited *a priori* certainties concerning them [5, 4, 7], making spectral imaging a unique approach to collect information on past states recorded in the materiality of these objects, and to understand their alteration through time.

Conversely, extracting exhaustively the information contained in spectral images is also challenging, and requires the use of multivariate analysis strategies that allow for the segmentation and/or the reduction of the dimensionality of the data set. The earliest and most commonly used methods are principal component analysis (PCA) and k-means clustering [31, 8, 32, 34, 28, 29, 14, 23]. Recently, the use of advanced statistical algorithms including Kullback-Leibler divergence [19] and t-distributed stochastic neighbourhood embedding (t-SNE) [15, 25, 23] have shown great promise to further discriminate and/or classify heterogeneities in spectral images. Yet, the application of these approaches is often limited by the fact that modern spectral imaging data sets make several GB or even tens of GB, depending on the type of detection used. Indeed, while 1D detectors typically record thousands of values (*i.e.* a few KB) per pixel, 2D array detectors record images of several MB per pixel (*e.g.* [1, 3, 20]). As such, these data sets are too massive to be timely exploited with the standard algorithms aforementioned, and we need to develop algorithms able to analyze such images and if possible in a timeframe compatible with the data collection time to provide feedback possibilities on the measurements (*e.g.* [1, 3]). One should also consider that on those measurements involving a probe, increasing the signal-to-noise ratio (SNR) comes at a cost: increasing probe/material interaction indeed most often leads to longer measurement times and always to a higher radiation dose deposited in the material. In such a framework one has to find a balance between SNR and dose/time, so that the experiment is conclusive without producing alteration of the samples during the analysis (*e.g.* [18]).

In this article, we focus on the question of image segmentation when the data set comes from X-ray fluorescence (XRF) mapping, a technique by which each individual pixel is characterized by its XRF spectrum, providing elemental composition information on that pixel (Fig. 1). The classical approach to plot quickly or even *live* elemental distributions recorded by XRF mapping consists of integrating the signal (*i.e.* photons counted

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

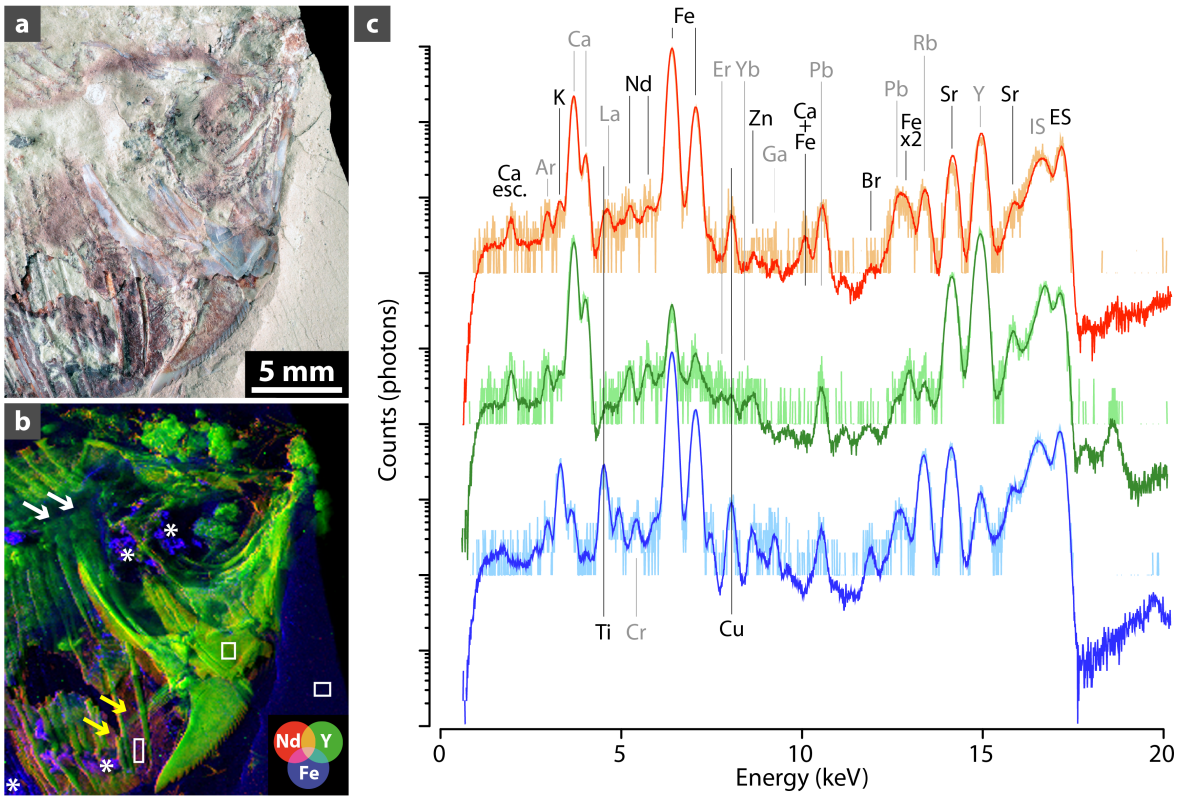

Figure 1. *Synchrotron XRF mapping of major-to-trace elements of the anterior part (skull on the right) of the yet undescribed fish MHNM-KK-OT 03a from the Jbel Oum Tkout Lagerstätte (Upper Cretaceous, 100 Myr, Morocco). (a): optical photograph. (b): false color overlay of the distributions of two rare earth elements, neodymium (red) and yttrium (green), and of iron (blue), reconstructed from a full spectral decomposition of the data (modified from [19]). Acquisition parameters: 100 x 100 μm^2 scan step, 50,851 pixels. Lighter tones indicate higher concentrations. Arrows and asterisks in b are discussed in the text. (c): Mean (dark colored; 90 pixels) and central individual (light colored) spectra from the boxes in b, corresponding to fossilized muscles (red and orange), bone (dark and light green) and the sedimentray matrix (dark and light blue), respectively. Spectra are shown using a logarythmic scale, vertically shifted for clarity. Main peaks are labelled. Abbreviations: esc., escape peak; ES, elastic scattering; IS, inelastic scattering; x2, sum (double) peak. Note that the Ar-peak does not arise from the sample but is due to excitation of Ar in the air (ca. 0.93 %) between the sample and the detector.*

by the detector) in spectral regions of interest (ROI) corresponding to targeted element peaks. This does not, however, hold true elemental distribution images as such ROI integrations additionally include significant contributions from other elements or phenomena (namely scattering, and sum and escape peaks); these overlapping biases can only be circumvented by applying slower approaches allowing a spectral decomposition of the data set (*e.g.* [19, 1]). Furthermore, while providing dimensional reduction, such processing does not provide a segmentation of the material and could only be an intermediate step towards the identification of specific constituent of the material and their

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

morphological features. An efficient model when it comes to analyse this type of samples is to consider that an image is made of a set of patches of uniform composition taken from a small, but unknown, set of compositions. This leads to two determining parameters for the model, the number of present compositions, *i.e.* the number of clusters in the segmentation, and the mean size of the patches which can also be measured as the patch density, that is the number of patches per unit surface of the image.

In the following, in Sect. 2, we propose a *hierarchical segmentation* algorithm combining the characteristics of hierarchical clustering with the imaging properties of a composite material. Compared to other methods, such as *k*-means, ascending hierarchical clustering provides a natural entry to apply spatial constraints. Furthermore, in the targeted imaging applications, the number of clusters (K) is not known *a priori*, and hierarchical clustering provides a structured way for the application domain scientist to assess the likely value(s) of K. In other words, we aim at proposing a hierarchical clustering procedure with spectral dissimilarities allowing to take into account the spatial proximities between the pixels. The herein proposed hierachical clustering with spatial constraint algorithm will be abbreviated by *HCSC*. Furthermore, the proposed method is able to estimate the patch density when it is not known *a priori*.

It is important to understand the nature of the signal measured in such experiment. In XRF, we measure the energy of the photons emitted by the material when it is subjected to monochromatic incident radiation. Because this re-emission phenomenon is a stochastic process, the measured spectrum is an empirical sampling of the law of this process. Instead of analyzing the signal using *generic tools* for Euclidean spaces, such as the ℓ_2 distance, it is therefore more relevant to use tools adapted to the comparison of population samples. On this respect, the algorithm we propose is based on the χ^2 as a tool to assess homogeneity between two samples, in the present case two pixels for which we want to test the potential similarity in composition.

After defining the terms and notations used throughout this article, we expose the general framework of our dissimilarity measure (using a Ward criteria based on χ^2, Sect. 2.2), and then propose an approach to impose spatial constraint upon the ascending process of the hierarchical clustering in Sect. 2.3, effectively segmenting the image into patches. Then, in Sect. 3, we concentrate on the proper steps at which the spatial constraint should be released to properly account for non connex domains made of the same material. We further consider the appropriate number of clusters at which the agglomeration process should be stopped in Sect. 4. To illustrate our approach, we apply the proposed algorithm on a true data set corresponding to the XRF mapping of a fossil teleost fish,

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

including the analysis of the experimental data set in Sect. 5. In Sect. 6 we apply the proposed algorithm to a purely synthetic data set for which, by construction, the *ground truth* is both known and abides by the uniform patches model exposed *supra*. This allows us to compare the results of our proposed algorithm to both k-means and unconstrained hierarchical ascending clustering in presence of various SNR. Whilst the uniform patches model is useful in the context of spectral image processing, it is still wrong, hence we perform, in Sect. 7, the analysis of a synthetic data set closer to the experimental one but providing the possibility to simulate various signal to noise ratio and giving insight into the robustness of the proposed algorithm to the noise level.

2. The proposed hierarchical clustering method

2.1. Notations and definitions

A spectral image of N pixels is considered. Each pixel of the set $\mathcal{I}$ is indexed by $i \in \{1, ..., N\}$ and is characterized by a spectrum $\mathcal{S}_i = (s_i(p))_{p\in\{1,...,P\}}$, where $s_i(p)$ is the number of photon counts for pixel i in energy channel p.

For $i \in \{1, ..., N\}$ and $p \in \{1, ..., P\}$, let $f_{i,p} = \dfrac{s_i(p)}{s_{\cdot\cdot}}$ and $t_p^i = \dfrac{s_i(p)}{s_{i\cdot}}$ where $s_{i\cdot} = \displaystyle\sum_{p=1}^{P} s_i(p)$ and $s_{\cdot\cdot} = \displaystyle\sum_{i=1}^{N} s_{i\cdot}$.

The aim is to propose a hierarchical classification procedure of spectra $(\mathcal{S}_i)_{i\in\{1,...,N\}}$ using the conditional distributions (or profiles) of pixels $((t_p^i)_{p\in\{1,...,P\},i\in\{1,...,N\}})$, the pixel i being weighted by $f_{i\cdot} = s_{i\cdot}/s_{\cdot\cdot}$ $(i \in \{1, ..., N\})$.

Since these profiles are probability distributions and the aim is to assess homogeneity between two pixels from their spectra (*i.e.* their potential similarity in composition), the comparaison of two profiles is made using the χ^2 euclidean distance.

So for two pixels i and j, let

$$d^2_{\chi^2}(\mathcal{S}_i, \mathcal{S}_j) = \sum_{p=1}^{P} \frac{(t_p^i - t_p^j)^2}{f_{\cdot p}}$$

with $f_{\cdot p} = \displaystyle\sum_{i=1}^{N} f_{i,p} = \frac{1}{s_{\cdot\cdot}} \sum_{i=1}^{N} s_i(p)$.

Remarks:

- It is assumed that $f_{\cdot p} \neq 0$ for all p. If there is a channel p such that $f_{\cdot p} = 0$, then $s_i(p) = 0$ for all $i \in \{1, ..., N\}$, hence $t_p^i = t_p^j = 0$ for all $(i, j) \in \{1, ..., N\}^2$. Thus, such channels are removed beforehand.

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

- It is assumed that $s_{i\cdot} \neq 0$ for all pixel i (otherwise it would mean that the detector did not received any photon for the corresponding pixel).

2.2. The Ward criterion

Using the χ^2 euclidean distance as the proximity measure between the spectra of pixels, the hierarchical clustering is designed with the agglomerative Ward criterion δ_{χ^2} [33], which consists of minimizing the increase of the within-cluster inertia at each step. This agglomerative criterion for two clusters C and C' is:

$$\delta_{\chi^2}(C, C') = \frac{\mu_C \mu_{C'}}{\mu_C + \mu_{C'}} d^2_{\chi^2}(S_{g_C}, S_{g_{C'}}) \tag{2.1}$$

where $\mu_C = \sum_{i \in C} f_{i\cdot}$ is the weight of cluster C and S_{g_C} the gravity center of cluster C; $S_{g_C} = (g_c(p))_{p \in \{1,\dots,P\}}$ with $g_c(p) = \frac{1}{\mu_C} \sum_{i \in C} f_{i\cdot} t^i_p$.

Note that the gravity center of the union of two clusters is $S_{g_{C \cup C'}} = \frac{\mu_C S_{g_C} + \mu_{C'} S_{g_{C'}}}{\mu_C + \mu_{C'}}$.

Usually, the dissimilarity matrix between clusters is updated with a specific occurrence of the general Lance and Williams formula, see [13, 21] for example. The dissimilarity between the possible aggregation $C_i \cup C_j$ of two clusters C_i and C_j and any other cluster C_k can be expressed by:

$$\begin{aligned} \delta_{\chi^2}(C_k, C_i \cup C_j) = & \frac{1}{\mu_{C_i} + \mu_{C_j} + \mu_{C_k}} \times \\ & \big((\mu_{C_k} + \mu_{C_i}) \delta_{\chi^2}(C_k, C_i) \\ & + (\mu_{C_k} + \mu_{C_j}) \delta_{\chi^2}(C_k, C_j) \\ & - \mu_{C_k} \delta_{\chi^2}(C_i, C_j) \big) \end{aligned} \tag{2.2}$$

2.3. Taking the spatial constraint into account

With spectral images, the dimension of the dissimilarity matrix at the start is too large ($\mathcal{O}(N^2) \approx 20$ GB) and it is computationally too expensive ($\mathcal{O}(N^2 P) \approx 5 \times 10^{12}$ operations to design directly a hierarchical clustering with the Ward criterion described above). Moreover, as mentioned in the introduction, an image is made of a set of patches of uniform composition. Hence it is desirable that the clusters form unions of patches, that is spatially connected sub-clusters.

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

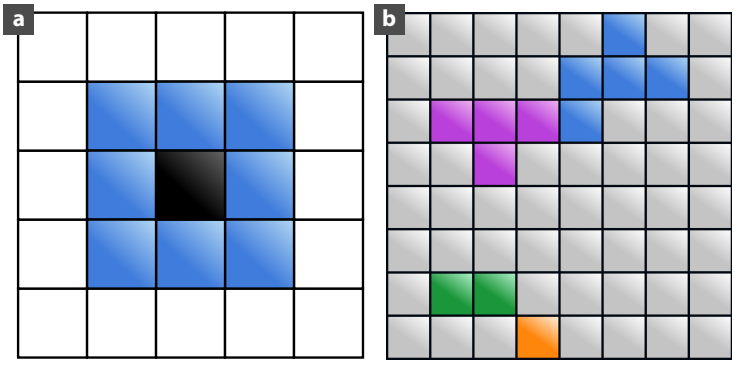

Figure 2. *Schematic representation of the second-order neighborhoods approach. (a): neighbors for a pixel that is not located on an edge or at a corner. (b): example of clusters spatially neighboring; on the top, the blue and purple clusters are spatially neighboring, while on the bottom left the green and orange clusters are spatially neighboring. All are spatially neighboring to the grey cluster. On another hand, for example, the purple and orange clusters are not spatially neighboring.*

For these two reasons, following [22], we propose a first hierarchical clustering algorithm that only aggregates two spatially neighboring clusters. More precisely, two clusters C and $C^{'}$ are spatially neighboring if there exists $(i, i^{'}) \in C \times C^{'}$ such that i and $i^{'}$ are neighboring pixels.

In our application, we will consider second-order neighborhoods (Fig. 2). It implies that most of the pixels have eight neighbors (Fig. 2a), while the pixels on an edge or at the corners have only five and three neighbors, respectively.

The advantage of this algorithm is, at each step, that for each cluster only a few dissimilarities have to be computed. Nevertheless, for this reason, it is not possible to use the Lance and Williams formula [13, 21] to update the dissimilarities. Therefore, equation 2.1 is used to compute the dissimilarities needed to design the hierarchy.

The hierarchical algorithm with the spatial constraint operates following the steps below:

If this algorithm is run until it remains only two clusters, we get a hierarchy where at each step the clusters are spatially connected. However, it is not desirable to impose such clusters connexion during the final steps. Indeed, from the point of view of the application domain scientist/specialist, the relevant clusters, while connected at fine scale, have no reason to be spatially connected at large scale. For example, when imaging a fossil, several bones will have a similar composition without touching each other.

As a consequence, the proposed algorithm taking the spatial constraint into account is run for J steps leading to spatially connected clusters, J being large (the choice of

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

Algorithm 2.1 Hierarchical spatial clustering

Initialization : computes the χ^2 distances between two spectra for neighboring pixels
Define $L := 1$
while $L < N$ **do**
  Aggregates the two neighboring clusters with the smallest Ward criterion value (or χ^2 distances at the first step)
  Updates the neighborhoods of clusters.
  Updates the dissimilarity matrix (for spatially neighboring clusters).
  $L := L + 1$
end while

the switching step J will be discussed hereafter in Sect. 3). It leads to $(N - J)$ spatially connected clusters or patches as called before. Then from these $(N - J)$ patches, unconstrained ascending hierarchical clustering algorithm with the Ward criterion is used. Thus, the proposed final clusters are union of the $(N - J)$ patches. Obviously, a relevant number of final clusters is to be chosen; this point is discussed in Sect. 4.

3. Selecting the switching step J

In the following, the patch density (for a unit surface of one pixel) is noted $\delta_p (\in]0, 1[)$. When the patch density of the sample is known *a priori*, the switching step is set so that the number of patches is an integer close to $\delta_p \times N$. This leads to release the spatial constraint at the step being the closest integer to $(1 - \delta_p) \times N$. Still, in most cases, the patch density is unknown and we herein propose a method to estimate this morphological characteristic of the sample and to select J.

3.1. The proposed criterion

In order to select the switching step J in the proposed hierarchical algorithm, a criterion balancing the between-cluster inertia with a regularization term measuring the spatial homogeneity of the clusters is proposed. This criterion to be maximized has the form:

$$H(J) = B(J) + \alpha G(J),$$

where $\alpha \in \mathbb{R}_+$, $B(J)$ is the between-cluster inertia of a partition of the pixels into $(N - J)$ patches and $G(J)$ is a measure of the spatial homogeneity of this partition. Following [2], we consider

$$G(J) = \frac{1}{2} \sum_{k=1}^{J} \sum_{i=1}^{N} \sum_{j=1}^{N} c_{ik} c_{jk} v_{ij}$$

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

where $v_{ij} = 1$ if i and j are neighbors, and 0 otherwise (with $v_{ii} = 0$ by convention), and $c_{ik} = 1$ if $i \in C_k$ and 0 otherwise.

To weight the $B(J)$ and $G(J)$ terms of the criterion, we can choose the scalar α to get a perfect balance between the two extreme cases : N clusters (*i.e.* a nul intra-cluster inertia) and one cluster (*i.e.* a perfect spatial homogeneity).

In the extreme situation of a partition into N clusters, we have $G(N) = 0$ and in the opposite extreme situation $G(1) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} v_{ij} \approx \frac{1}{2}\sum_{j=1}^{N} 8 = 4N$. (For simplicity, we consider here improperly that each pixel has 8 neighbors.)

Assuming a balance between these two extreme situations $H(N) = H(1)$ leads to $\alpha = \frac{T}{4N}$, with $T = B(N)$ being the total inertia of the whole set of pixels. Thus, the criterion to be maximized is

$$H(J) = B(J) + \frac{T}{4N} G(J).$$

However, as it will be apparent in the case study in Sect. 5, the choice of this criterion leads to the selection of a too large number of agglomerative steps $J_{\max}$, in other words of $(N - J_{\max})$ patches that are too small.

In order to select a more relevant number of patches, from which to release spatial constraints in the clustering, we propose to make use of the "one standard deviation" procedure proposed in [9] to cut a decision tree. This procedure consists of computing $H(J)$ for $J = N, \ldots, 1$, then to compute the standard error $\mathrm{sd}(H)$ of the resulting $(H(J))_{J=N,\ldots,1}$ and choosing the smallest $\hat{J}$ such that

$$H(\hat{J}) \geq H(J_{\max}) - \mathrm{sd}(H).$$

The rationale for this procedure is to determine the value of $\hat{J}$ that corresponds to a balanced number of patches that provides a good compromise between the between-cluster inertia and the spatial homogeneity. Note that, while it is expected that $H(J)$ increases from $H(N)$ to $H(J_{\max})$ and then decreasing back to $H(1)$, there is no guarantee for such a behaviour. Such an unexpected behaviour of the H criterion is obtained in particular for low signal to noise ratio image in Fig. 8e and g. In order to also address these types of behaviour of H we express the choice of $\hat{J}$ in a different way :

$$\hat{J} = \max\{J | (J \leq J_{\max}) \text{ and } (H(J) \leq H(J_{\max}) - \mathrm{sd}(H))\}$$

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

And this leads to estimate the patch density δ_p by $\hat{\delta}_p = \dfrac{N - \hat{J}}{N}$.

3.2. Heuristic to obtain $\hat{J}$ in practice

With the advent of fast detectors and rapid data collection capabilities larger spectral images may be collected to access higher spatial definition of the sample. When the image size becomes too large, it can be too long to compute $H(J)$ for all $J \in \{N, ..., 1\}$. In this case, two following heuristic approaches are proposed to determine $\hat{J}$ faster:

At first, the idea is to compute $(H(\ell \times by))_l$, with $by \geq 1$ chosen to have a reasonable computing time and $\ell \in \mathbb{N}^*$ such that $\ell \times by \leq N$. After testing on the studied data set described in Sect. 5, we noticed a significant speed gain (see Tab. 1 showing a 35% gain between $by = 1$ and $by = 5$). Notice that the obtained $\hat{J}$ can change significantly according to the value by. Hence, by must be small enough to obtain a correct value for $\hat{J}$ (but we do not know its order of magnitude). Anyway, increasing further the by parameter is not leading to much speed increase (data not shown).

If the gain from this first approach is not sufficient to compute the value of the switching step $\hat{J}$, we can tap on the relation seen above between $\hat{J}$ and the proposed estimation of the patch density δ_p. We propose to evaluate the "constant" δ_p, which is linked to the morphology of the studied image, by cutting the image into q sub-images $(I_k)_{k\in\{1,...,q\}}$ with size $(N_k)_{k\in\{1,...,q\}}$. Then, for each sub-image I_k, $\hat{J}_k$ is computed with the criterion described in Sect. 3.1 leading to an evaluation of δ_p: $\hat{\delta}_{p,k} = \dfrac{N_k - \hat{J}_k}{N_k}$.

Finally, we take as estimation of δ_p: $\bar{\delta_p} = \dfrac{1}{q}\sum_{k=1}^{q} \hat{\delta}_{p,k}$ and $\hat{J}$ is chosen as the closest integer to $(1 - \bar{\delta_p}) \times N$.

Remark: q must be chosen small enough so that the sub-images are large enough to reflect the studied image in terms of patch density. And q is also chosen to obtain an estimation of δ_p in a "reasonable" computation time.

4. Selecting the number of clusters

4.1. Statistical heuristics

A first and simple way to properly assess the number of clusters from a dendrogram is to select the numbers of clusters producing the greater jumps in the plot of the cluster criterion values (*i.e.* here the Ward criterion), against the number of clusters. We refer to this strategy as the *jump heuristic*.

An another natural and popular criterion for choosing a relevant number of clusters K in a hierarchy designed with the Ward criterion is to use the value of K corresponding

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

to the maximum value of the Calinski and Harabasz criterion (CHC, [10])

$$\text{CHC}(K) = \frac{\text{Tr}(B_K)}{(K-1)} / \frac{\text{Tr}(W_K)}{(N-K)},$$

where B_K and W_K are respectively the between-cluster matrix and the within-cluster matrix of the partition $C_1, \ldots, C_K$. In the present context, we have

$$\text{Tr}(B_K) = \sum_{k=1}^{K} \mu_{C_k} d^2_{\chi^2}(S_{g_{C_k}}, S_g)$$

and

$$\text{Tr}(W_K) = \sum_{k=1}^{K} \sum_{i \in C_k} f_{i.} d^2_{\chi^2}(S_{g_{C_k}}, \mathcal{S}_i)$$

where S_g is the gravity center of the N pixels, and for $k = 1, \ldots, K$, $S_{g_{C_k}}$ is the gravity center of cluster C_k.

This criterion has been shown to perform well in practical situations (see [24]).

4.2. Practical considerations in the case of segmentation of spectral images

Although we listed above several ways to statistically determine the number of clusters to retain, it is however not recommended to choose a unique number of clusters with a formal technique when dealing with data coming from an applied scientifc domain, as it is the case of ancient material studies. Instead, we here prefer to use the following strategy:

- Preselect several number of clusters using the jump heuristic based on the Ward criterion or local maximum of the CHC.
- Analyze the preselected clusterings with the help of a specialist of the application domain. Having this purpose in mind, it is desirable to provide the specialist with the mean spectra of the preselected clusters, which represent complementary information to those obtained from usual spectral image processing (*e.g.* ROI integration and full spectral decomposition, see Sect. 1) and are, as such, critical to assess the robustness and benefits of the approach.
- Select with this person the clustering(s) to be interpreted.

The present paper exemplifies in the following section this way of assessing ancient material clusterings.

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

4.3. Matching colors when representing multiple segmentations

Visual comparison of different segmentation results might be a difficult task since visual perception of a segmentation may be strongly affected by the mere permutation of the false color palette used to differentiate between clusters. When the segmentations to be compared belong to the same hierarchy, as is the case in Fig. 3, one may use the hierarchy itself to propose an optimal match of the palettes used to represent both segmentations. At each ascending step, the newly agglomerated cluster takes the color of its larger ascendant, hence minimizing the perceptual difference between the ascendant and the descendant segmentation.

False color palette matching is more tricky when the segmentation do not have a hierarchical relationship, as is the case of most representation proposed in this article. We based our appraoch onto the use of the Rand index [27] which provides a global measure of similarity between two different segmentations of the same image. Given $(X_l)_{l\in\{1,..,r\}}$ and $(Y_m)_{m\in\{1,..,s\}}$ two segmentations of the same image, in other words they are two partitions of the set $\mathcal{I}$, their similarity is measured by :

$$R((X_l),(Y_m)) = \frac{a+b}{\binom{N}{2}}$$

with a being the number of pairs of pixels, $\{i,j\} \subset \mathcal{I}$, which are elements of the same cluster both in segmentation (X_l) and (Y_m) and b being the number of pairs of pixels which are elements of different clusters both in segmentation (X_l) and (Y_m) :

$$a = |\{\{i,j\} \subset \mathcal{I} \mid \exists (l,m) \text{ such } \{i,j\} \subset X_l \text{ and } \{i,j\} \subset Y_m\}|$$
$$b = |\{\{i,j\} \subset \mathcal{I} \mid \exists (l,m) \text{ such } i \in X_l, j \notin X_l, i \in Y_m, j \notin Y_m\}|$$

Both these cardinals may be decomposed as a double summation on indices of both segmentation, l and m. Precisely :

$$a = \sum_{l=1}^{r}\sum_{m=1}^{s} |\{\{i,j\} \subset X_l \cap Y_m\}|$$

while b needs a $\frac{1}{2}$ factor to ensure that each pair is not counted twice :

$$b = \frac{1}{2}\sum_{l=1}^{r}\sum_{m=1}^{s} |\{\{i,j\} \mid i \in X_l \cap Y_m, j \in X_l^c \cap Y_m^c\}|$$

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

where X_l^c and Y_m^c are respectively the complements of X_l and Y_m in $\mathcal{I}$. Using these decomposition, we can now express the Rand index as a sum over $\{1,..,r\} \times \{1,..,s\}$ of *partial Rand indices* measuring how well a given cluster X_l match to cluster Y_m :

$$\begin{aligned} R((X_l),(Y_m)) &= \frac{1}{\binom{N}{2}} \sum_{l=1}^{r} \sum_{m=1}^{s} (a_{l,m} + b_{l,m}) \\ \text{with } a_{l,m} &= |\{\{i,j\} \subset X_l \cap Y_m\}| \\ b_{l,m} &= \frac{1}{2} |\{\{i,j\} \mid i \in X_l \cap Y_m, j \in X_l^c \cap Y_m^c\}| \end{aligned}$$

To match colors of segmentation (Y_m) onto the one used for segmentation (X_l), we use the index matching injection $\{1,..,r\} \to \{1,..,s\}, l \mapsto h(l)$ that maximizes the sum of corresponding *partial Rand indices* :

$$\sum_{l=1}^{r} \left(a_{l,h(l)} + b_{l,h(l)}\right)$$

considering that, when $r > s$, $a_{l,h(l)} = b_{l,h(l)} = 0$ when $h(l)$ is not defined. Conversely, when $r < s$ all indices m not in the image of injection h have to be colored by a color not used in the representation of (X_l).

5. Application to a real world data set

5.1. Data description

The proposed algorithm has been applied to a spectral image data set collected on a yet undescribed *ca.* 100-million-year-old new teleost fish from Morocco (Fig. 1, [19]). The information embedded in this data set is a synchrotron micro-X-ray fluorescence (μXRF) major-to-trace-elemental map, where a full XRF spectrum has been recorded for each pixel, over a $22.5 \times 22.5\ mm^2$ area using a scan step of $100 \times 100\ \mu m^2$ and a 500 ms counting time ($211 \times 241 = 50,851$ pixels in total; Fig. 1b,c). The experiment was performed at the DiffAbs beamline (SOLEIL synchrotron, Gif-sur-Yvette, France) using a 17.2 keV incident beam focused down to a diameter of $10 \times 7\ \mu m^2$.

Very interestingly, the distribution of strontium and yttrium Kα lines, which substitute for calcium in calcium phosphates such as bone apatite [19, 17] and whose information depths under hard X-rays reach 200-300 μm in pure apatite with the used geometry, revealed previously indiscernible anatomical features in this peculiar new fish (Fig. 1b,

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

[19]). They particularly unveil the morphology of the first vertebrae (white arrows in Fig. 1b), the neurocranium that extends into a sharp supraoccipital at the top of the skull, the metapterygoid, and the hyomandibular that appears dorsally flared. These new information help deciphering the affinities of this new fossil species (in preparation). The other main outcome of this work was that a false color overlay of the distribution of different rare earth elements (REEs; *e.g.* neodymium and yttrium, red and green distributions in Fig. 1b, respectively) discriminates phosphatized muscles (yellow arrows in Fig. 1b and bone [19].

5.2. Resulting hierarchical spatial clustering

In the following, the proposed algorithm has been implemented with R [26] on this image of $N = 211 \times 241 = 50,851$ pixels, for which at each pixel i the spectrum $\mathcal{S}_i$ has $P = 1780$ values. The size of the file containing the data set is 1.7 GB.

For this data set, it is possible to compute the criterion H described in Sect. 3 for all $J \in \{N, ..., 1\}$ in order to determine the switching step $\hat{J}$. We obtain $\hat{J} = 44175$ (and an evaluation of the patch density $\hat{\delta}_p \simeq 0.13139$).

To select the number of clusters, we plotted the Ward criterion and the CHC against the number of clusters (starting with two clusters) (Fig. 3). These criteria are here to complement the knowledge of the application domain specialist, in the case of the present example a paleontologist (PG). The jump heuristic leads to propose 5 clusters, whilst the CHC leads to propose 10 clusters (or 5), corresponding to the local maxima of the curve.

Looking more precisely at the differences between 10 and 5 clusters, two major clusters merging (in terms of number of pixels involved) are observed: they involve the purple and gold clusters (6729 and 8123 pixels, respectively), and the light-red and light-green clusters (3204 and 21945 pixels) (Fig. 3c,d). In addition, the smaller jade and pale yellow clusters (87 and 435 pixels; mostly distributed on the top right corner of the image) also merge. The remaining mergers (the dark- and light-orange clusters–50 and 5 pixels– that merge into the light-green one), can be better pinpointed on the difference image (Fig. 3e). Of particular interest is the light-red cluster, as it highlights areas richer in iron (asterisks in Fig. 1b) that are not clearly obvious in the μXRF elemental maps obtained using ROI integration or spectral decomposition. On another hand, merging of the purple and gold clusters makes it possible to see again the neural spines behind the skull (top right corner of the image), well visible on the specimen but not discriminated in the 5-classes clustering analysis.

From a paleontological point of view, the segmentation offered by the selected cluster-

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

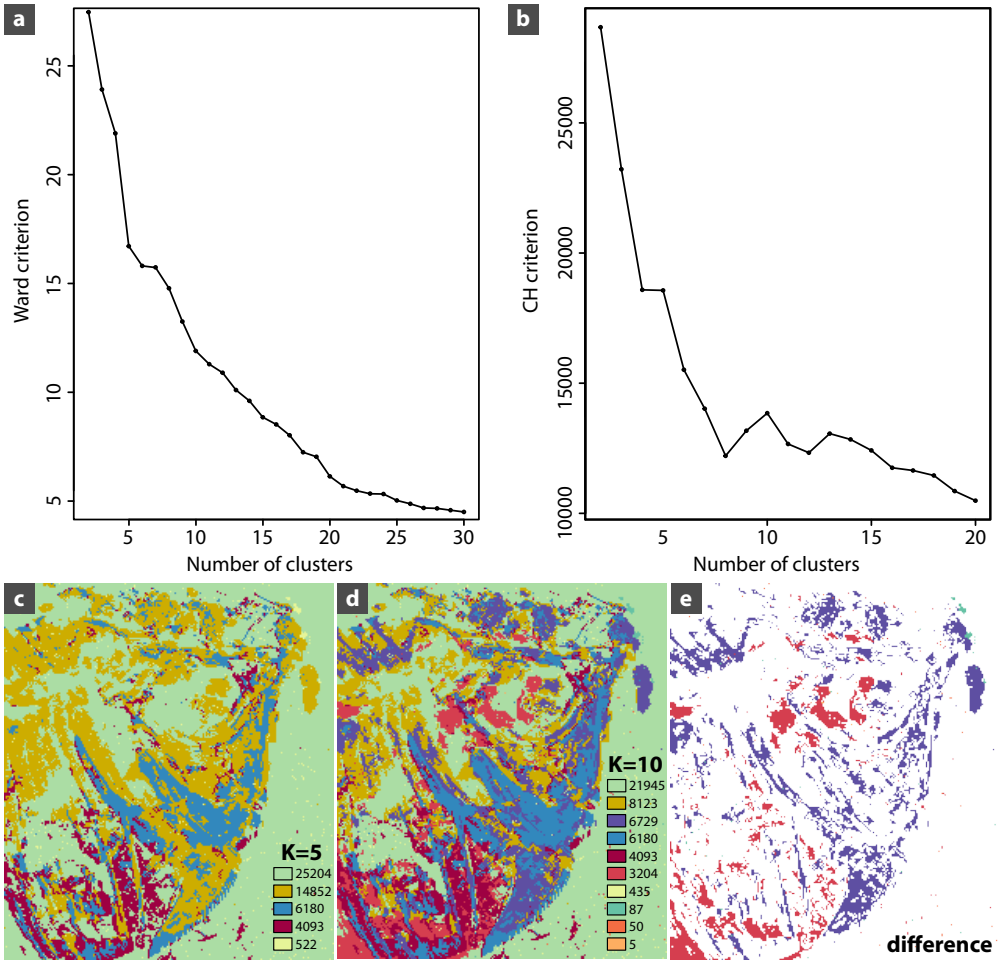

Figure 3. *Defining the number of clusters used for hierarchical segmentation (a, b): Ward (a) and Calinski and Harabasz (b) criteria against the number of clusters (starting with 2 clusters). (c–e): False color distributions obtained for 5 (c) and 10 (d) clusters, and difference (e).*

ing (Fig. 4a) does not improve the visualization of hidden anatomical details, but provides new insights into the chemical composition of the different tissues and materials present in the sample through the mean spectra of the clusters (Fig. 4b). While individual elemental distributions show no strong contrast in the incorporation of light REEs between bone and muscles (Fig. 4c), following the distribution of calcium, which they substitute and that originates from a comparable depth (Fig. 4d), the yttrium distribution shows strong enrichment in the bone as compared to the muscles (Fig. 4e). In fact, in the muscles area (yellow arrows in Fig. 1b), rather than following the type of tissue the yttrium distribution largely follows the thickness of the material as shown by X-ray microtomography where most of the muscles region appear to be very thin or not discernible (Fig. 4f).

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

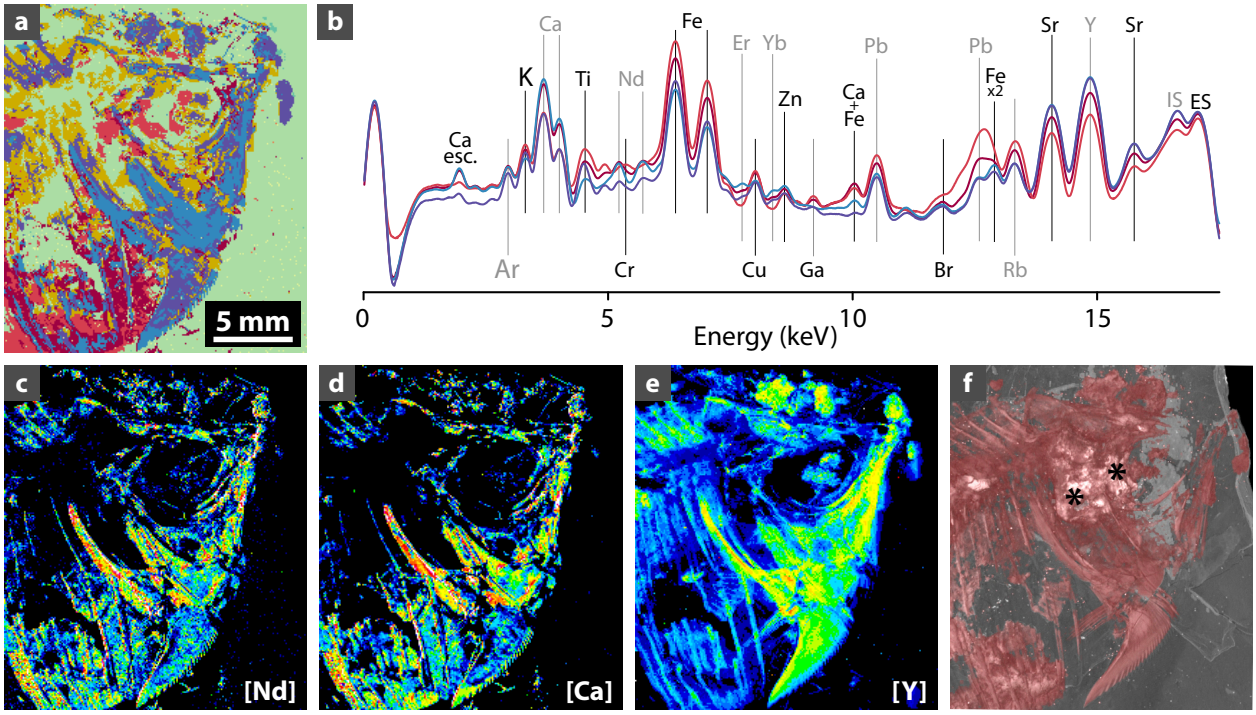

Figure 4. *Hierarchical segmentation of the synchrotron μXRF spectral image data set of the yet undescribed fish (MHNM-KK-OT 03a) from the Jbel Oum Tkout Lagerstätte (Upper Cretaceous, 100 Myr, Morocco). (a): Segmentation results when 10 clusters are selected with the proposed algorithm, disabling spatial constraint at step $\hat{J} = 44175$. (b): mean spectra from 4 of the 10 clusters visible in (a). (c–e): concentration maps of neodymium (c), calcium (d) and yttrium (e). The color scale goes from dark blue (for low concentration) to red (high concentration) going through green and yellow. (f): micro-computed tomography 3D rendering of the fossil within the sedimentary matrix after rapid segmentation. Voxel size: (24.7 mm)^3.*

Consequently, thickness and information depth were likely responsible for the apparent REE contrast. Nevertheless, our clustering clearly discriminates bone from phosphatized muscles (blue/purple and dark red clusters in Fig. 4a, respectively) on the basis on the full μXRF spectra. The muscles dark-red cluster appears richer in Fe and Pb (Fig. 4b), which come from a reddish fossil biofilm made of iron hydroxides and covering the phosphatized muscles [18, 12, 16] rather than the phosphatic material itself. In turn, the bone blue and purple clusters contain much higher contents in Sr, Y and heavy REEs (Lβ1 emission lines from erbium and ytterbium particularly stand out in Fig. 4b, as they do not fall in the same energy domain as major elements [19]). This is most likely again an effect of information depths and thickness of the tissues.

On another hand, the obtained clustering isolates well the large, highly absorbing iron grains situated posteriorly to the orbit (asterisks in Fig. 1b; light-red cluster in Fig. 4a,b)

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

that prevent segmentation of the first vertebrae and posterior part of the head from the X-ray tomography data (asterisks in Fig. 4f). These grains are particularly richer in Ti, Cr, Fe and Pb (Fig. 4b) and are therefore, besides their larger size, a different material than the reddish thin film of iron hydroxides covering most of the fossil.

Concerning the computational time, the proposed algorithm, as well as all other algorithm used for comparison herein, has been run on a computer with an AMD EPYC-3 CPU (EPYC 7542) with 32 cores running at $3.4\,\mathrm{GHz}$ and having 512GB of shared RAM. Tab. 1 provides the timing of the computation decomposed in the spatially constrained phase (spat. const. clust) and, for the final round of unconstrained hierarchical clustering, the full distance matrix computation (dist.mat.comp) and the clustering (clustering) itself which uses Lance-William formulae.

Full segmentation of the image took three to four minutes when estimating $\hat{J}$ and using the estimated value to release spatial constraints. This time could be decreased to slightly above two minutes if $\hat{J}$ is estimated using an aggressive technique (using a $by = 5$ value for computing $H(J)$). If the patch density δ_p is known *a priori* fifteen extra seconds could be spared on the computation. While this is not significant on the tested data set, we expect this difference to increase with larger images.

We also tried to determine the switching step $\hat{J}$ by cutting the image into $q = 4$ sub-images with size 100×115 which took 15s by sub-image. While using sub-images to estimate $H(J)$ does not brings any time benefit on this image, we foresee that it will help keeping low computational time on larger spectral images. So δ_p is evaluated from four $(\hat{\delta}_{p,k})_k$ values obtained on the four sub-images with H computed every 5 steps ($by = 5$). Moreover we obtained: 0.15913, 0.06957, 0.14435 and 0.14478 respectively for top right, bottom right, bottom left and top left sub-images leading to $\bar{\delta_p} \simeq 0.12946$ and consequently $\hat{J} = 44268$ (compared to $\hat{J} = 44175$ when considering the full image). Note that the $(\hat{\delta}_{p,k})_k$ values are close for the three sub-images having a similar type of morphology, whilst the bottom-right sub-image consist mostly of sediment and is more homogeneous than other sub-images.

By providing a global discrimination of the different materials composing the fossil much faster than a full spectral decomposition (less than four minutes here (Tab. 1) and several hours to a few days using the freeware PyMCA [30]), the proposed clustering methodology provides a robust and quick way to extract, "live" at the beamline, chemical information not hampered by local heterogeneity or contamination for further higher resolution mapping of areas of interest, or point analyzes using, e.g., X-ray absorption

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

Algorithm	Step	Parameters	Processing time
HCSC	spat. const. clust.	211×241 px ; not est. H	102 s
		211×241 px ; $by = 1$	181 s
		211×241 px ; $by = 5$	118 s
		100×115 px ; $by = 5$†	13 s
	dist. mat. comp.	$J = 44175/6676$ patches	127 s
		$J = 46000/4851$ patches	111 s
		$J = 44175/6676$ patches ; `OpenCL`‡	12 s
		$J = 46000/4851$ patches ; `OpenCL`‡	7 s
	clustering	$J = 44175/6676$ patches	3 s
		$J = 46000/4851$ patches	1 s
k-means		$k = 10$, $n = 100$	64 min
		$k = 15$, $n = 100$	78 min
		$k = 20$, $n = 100$	104 min

Table 1

Computational cost of the herein implemented or tested spectral image segmentation algorithms. All computations were performed on the exact same computer, based on a AMD EPYC-3 CPU (EPYC 7542) with 32 cores running at 3.4 GHz *and having 512GB of shared RAM. All computations were done using single threaded code except* † *for which each quadrant is computed in a separate thread (4 threads in total) and* ‡ *exploiting all the 32 cores of the CPU through the use of* `OpenCL` *on CPU (*`POCL`*). Independent runs of each of these processes show small timing variation, in the range of up to 20%.*

spectroscopy.

5.3. Regarding the chosen switching step J

One may wonder if the value of the switching step, J, has an influence on the results for the choice of the number of clusters and for the clusters shape. In this section we tackle this question by applying the algorithm using switching steps equal to $J = 43000$ and $J = 46000$. In Fig. 5 are the graphic representations of the jump heuristic and the CHC for $J = 43000$ and $J = 46000$, respectively.

For $J = 43000$, the jump heuristic plot leads to propose 6 or 10 clusters while the CHC leads to 3, 10 or 12 clusters (Fig. 5a,b). For $J = 46000$, the jump heuristic plot leads to propose 5 clusters or maybe 9, and the CHC leads to propose 9 clusters (first local maximum) or more (Fig. 5e,f). These results show that the value of the switching step has an influence on the result of the hierarchical clustering. Comparison of the graphic representations for $J = 43000$, 44175 and 46000 (Fig. 5c,d,g) clearly identifies the segmentation resulting from the latter as absolutely unsatisfactory as many fossil areas are found mixed up with the surrounding sediment (Fig. 5g). Graphic representations for $J = 43000$ and 44175 appear in turn very similar. Nevertheless, representation for $J = 44175$ (the computed $\hat{J}$ value, see Sect. 5.2) more accurately reflects elemental dis-

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

tributions (Fig. 1b), particularly regarding the iron-rich phase located around the fish orbit.

5.4. Comparison with *k*-means based segmentation

In this paragraph we compare the results obtained with the proposed HCSC algorithm with *k*-means clustering. To be close to what was done above, we apply the *k*-means algorithm to the data $(\mathcal{D}_i)_{i\in\{1,\dots,N\}}$ with

$$\mathcal{D}_i = \left(\frac{t_p^i}{\sqrt{f_{\cdot p}}} \right)_{p\in\{1,\dots,P\}}$$

i.e. the conditional distributions of pixels, the pth coordinate being divided by $\sqrt{f_{\cdot p}}$ (see Sect. 2.1).

The number of clusters K is *a priori* unknown so here it must be arbitrarily determined in advance with the advice of the application domain specialist or the *k*-means algorithm must be run for several values. In general it is safer to overestimate the number of clusters rather than risking underestimating it, hence we used $K = 15$ and $K = 20$ clusters. Tests were performed using the function `kmeans` from the `stats` package of `R` and requesting 100 random initializations (parameter n in the suite of the text) to limit the risk of finding a sub optimal local minimum due to the stochastic nature of the algorithm. On our test computer, these runs took respectively $1h18$ and $1h44$. Obviously, estimating the value of K through analysis of results of multiple run of the *k*-means with different K would then lead to a much higher computational time, even-though we may decide to run less random initializations, practically trading off quality of the results to decrease computational time as exemplified and illustrated in Sect. 6. For sake of full comparison, we also performed *k*-means using $K = 10$ (completed in $1h04$) to meet the best conditions for comparing the segmentation resulting from *k*-means with the best result of the proposed HCSC algorithm. Segmentations obtained by these runs of *k*-means are visible in Fig. 6.

From the point of view of data interpretation, all three segmentations obtained using *k*-means clustering (using $K = 10$, $K = 15$ and $K = 20$) pinpoint the muscles (dark-red cluster) and highly absorbing iron grains (light-red cluster), with growing differences as the number of classes increases. The results obtained with our HCSC approach and *k*-means with $K = 10$ are highly comparable, except for two main features that appear on the *k*-means segmentation but are not discriminated using HCSC, namely (i) a nearly vertical light-red line to the right of the fossil, and (ii) a pink and gray triangle in the top-right corner of the image (Fig. 6a,b). The latter results from the discrimination between

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

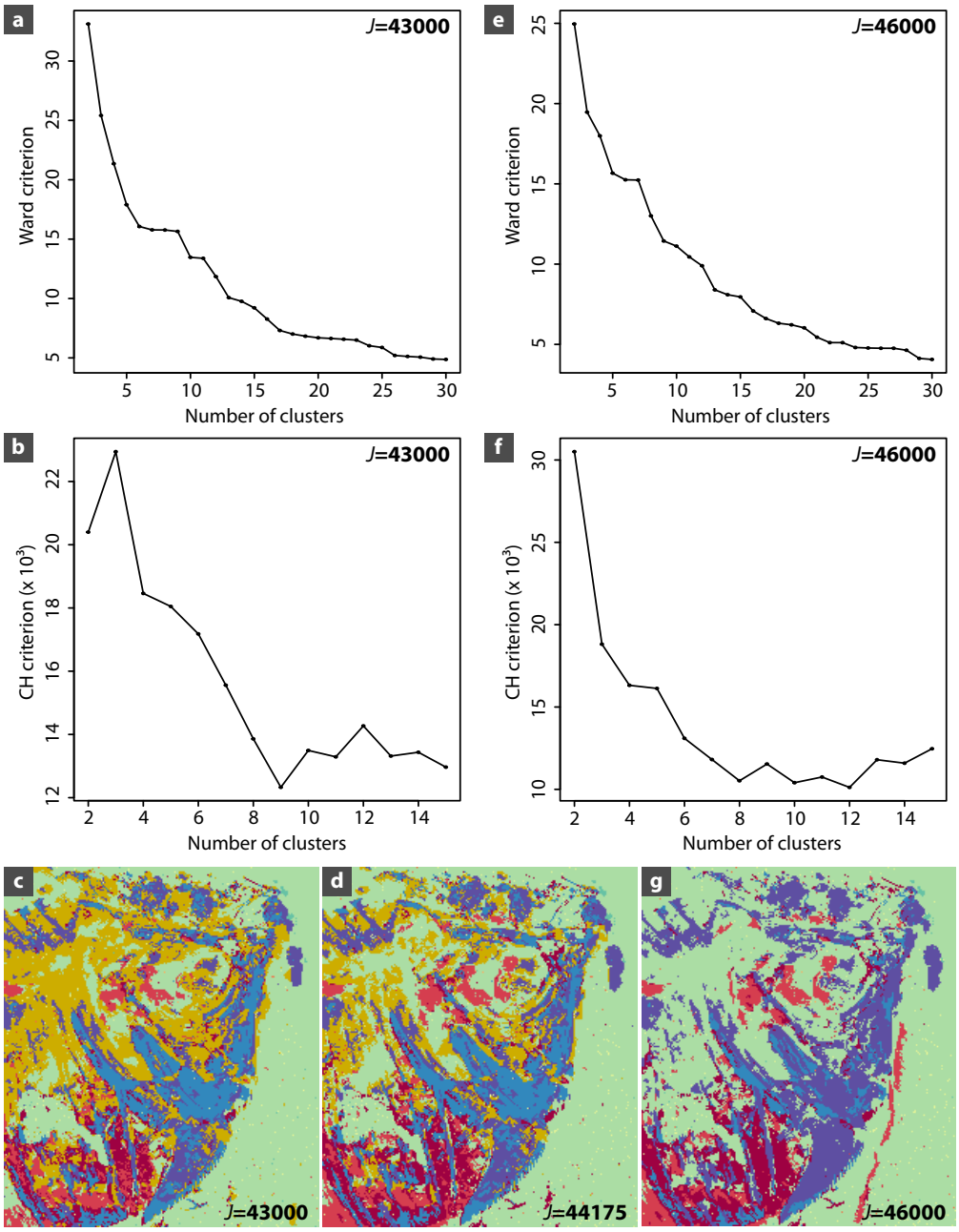

Figure 5. *Hierarchical segmentation for different choices of switching step* $J = 43000$ *and* $J = 46000$. *(a, b): Ward (a) and Calinski and Harabasz (b) criteria against the number of clusters (starting with 2 clusters) for* $J = 43000$. *(c): False color distributions obtained for* $J = 43000$ *(10 clusters). (d): False color distributions obtained for* $J = 44175$ *(10 clusters). (e, f): Ward (e) and Calinski and Harabasz (f) criteria against the number of clusters (starting with 2 clusters) for* $J = 46000$. *(g): False color distributions obtained for* $J = 46000$ *(9 clusters). Colors of (c) and (g) were matched to the colors of (d) using the approach presented in Sect. 4.3*

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

air around the sample (see Fig. 1a) and the sediment, which clearly represents an improvement as compared to the HCSC segmentation because air should be discriminated from sediment as both have very different compositions. Nonetheless, such a lack of discrimination in the HCSC segmentation can easily be explained by the geometry used during the experiment. Indeed, as the X-ray beam came from the right of the sample with a 45° angle, pixels corresponding to air in that area recorded the X-rays-sediment interaction below the fossil surface, which albeit at much lower concentrations produced similar spectra that clustered together when applying spatial constraints. The other feature that appears on the *k*-means segmentation but not in the HCSC one (i.e. the nearly vertical light-red line to the right of the fossil) is, on the other hand, rather problematic. Indeed, it clusters with the highly absorbing iron grains whereas it is purely artefactual and surely does not have the same elemental composition; it only corresponds to sample topography (see Fig. 1a,b). Furthermore, we can also notice that there is much more isolated small groups of pixels in the *k*-means segmentations than in the HCSC one, illustrating the main interest of using spatial constraints. Finally, another major difference is that *k*-means clustering takes a significantly longer time than the proposed HCSC algorithm (64 minutes for *k*-means using $K = 10$ and 100 random initializations vs. 2–3 minutes for HCSC; Tab. 1) for a segmentation that usually looks like the HCSC results but is less readable, and in the present case mixes up topographical and chemical information.

6. A fully synthetic spectral image data set

Unfortunately, the study of the above mentioned data set can not address the question of how well the proposed algorithm is able to produce the true model from the data. This is because we do not have access to the *ground truth* of the observed sample. To our knowledge there is no publicly accessible data set for which both the data and a validated ground truth model are widely available, and such data set are very hard to produce on true sample since synchrotron-based XRF mapping is considered to be one of the most informative methods. Hence we propose to use a fully synthetic (and generative) model for which we can simulate likely measured data. The synthesis of the data is performed in two successive steps in order to be able to control the SNR of the generated data. First, we use our model to generate, for each pixel, a noise-less spectrum that leads to what we call the *zero noise model*. Then, the spectrum of each pixel is replaced by a spectrum which noise resembles that of the physical XRF measurements. Note that we decided to generate a smaller, $N = 100 \times 100$, image so that we can rapidly test many different options

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

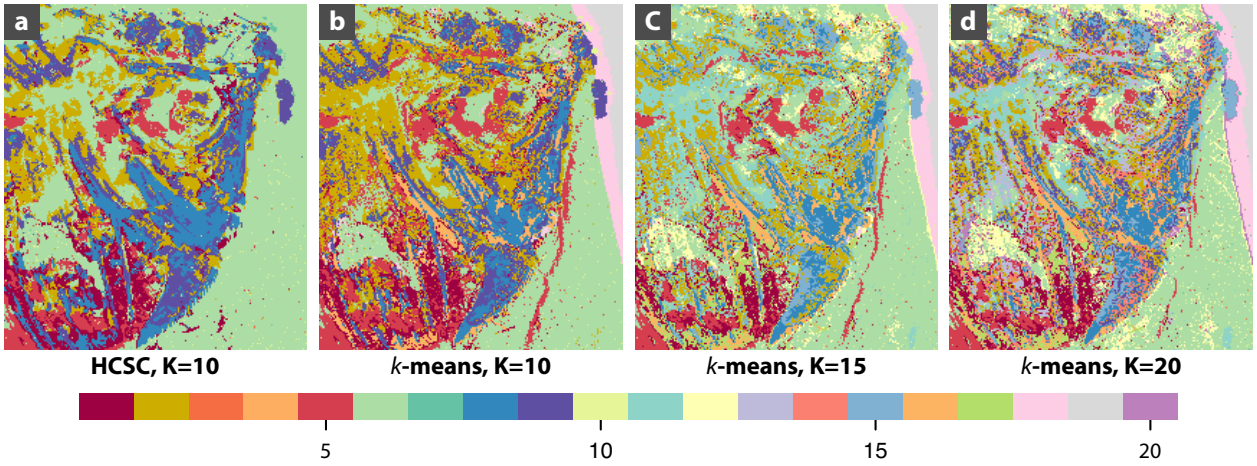

Figure 6. *Comparison of the graphic representations obtained using our hierarchical clustering with spatial constraints (a) and* k*-means clustering for 10 (b), 15 (c) and 20 (d) clusters. Correspondence between colors and clusters is shown at the bottom, with only the first 10 colors used in (a) and all 20 in (d). Using the approach proposed in Sec. 4.3, colors of (d) were matched to those of (a), then colors of (b) and (c) were matched to those of (d)*

of the HCSC algorithm but also, more importantly, of other algorithms for comparison.

6.1. Building a fully synthetic *zero noise model*

To be able to propose a noiseless spectrum for each pixel of the image, we start from the *uniform patches model* exposed in the introduction. We drew a fossil composed of six classes, namely sediment, bone, bone coating, muscles, eye and iron-rich grains. Each of those classes is assigned a reference spectrum taken from the mean spectra of the clusters identified above using HCSC (these mean spectra are obtained from a large number of pixels and hence exhibit the strong regularity of noiseless spectra, as can be seen on Fig. 4b). We also want our model to account for the overall amplitude variations that are due to small fluctuations of the incident beam as well as local sample density variations, since these amplitude variations affect the SNR of the spectra measured for each pixel. Each pixel is then assigned the reference spectrum of its class multiplied by an amplitude factor being the exponential of a zero-mean Gaussian random field of appropriate variance and spatial regularity. While being very regular, as would be noiseless spectra, the generated pixel spectra exhibit levels in correspondence with the noisy ones observed in our real data set. This synthetic model provides *ground truth* both in terms of class and spectrum for each pixel.

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

6.2. Simulating data with controlled SNR from the *zero noise model*

The noise present in the observation is mostly due to the counting statistic of each channel of the detector. Hence, we can generate a simulated observed spectra with the same SNR as the raw observation, by simply replacing the value of the *zero noise spectra* by a single realization of a Poisson random process with its parameter being the *zero noise spectra*'s value. We generated such a data set, for which we have, by construction, the *ground truth* and a SNR equal to the one of the experimental data set. This simulated data set is later on referred as a *plus 0dB data set* (p0dB in short).

Starting from the same *zero noise model*, we also generated simulated observation with lower SNR. Since each *theoretical value* is replaced by a Poisson realization, dividing the model by a factor of 2 would decrease the SNR by a factor of $\sqrt{2}$, which corresponds to removing 3dB to the SNR. This simulated data set is later on referred as a *minus 3dB data set* (m3dB in short). Repeating this procedure two more times enabled us to generate a *minus 6dB data set* (m6dB) and finally a *minus 9dB data set* (m9dB).

Each of these data sets resembles what could have been measured if the data acquisition exposure time was divided by two incrementally. In other words, obtaining for the m3dB data set a spatial clustering similar to that obtained for the p0dB data set would lead to the conclusion that the experiment could have been done twice faster without significant loss in term of the explained morphology of the fossil. A shorter exposure time also means a lower radiation dose for the sample and correspondingly lower risk of alteration during and due to the measurements.

6.3. Tentative segmentation of the data set, comparing HCSC, *k*-means and unconstrained hierarchical clustering

The four data sets, p0dB, m3dB, m6dB and m9dB, were all subjected to segmentation using three different algorithms, namely HCSC, *k*-means and standard hierarchical clustering (HC). Unconstrained hierarchical clustering was possible on this data set owing to its limited size. It was performed using the `hclust` function from the `stats` package of `R` together with the `OpenCL` accelerated version of the `dist` function implemented for use by HCSC.

The obtained segmentations are presented in Fig. 7, including the parameter used (the switching step $\hat{J}$ for HCSC, or number of initializations n for *k*-means). The figure also reports the computational time taken by each algorithm. As expected, the execution time of HC is nearly constant and unaffected by the SNR of the data sets while *k*-means execution is strongly affected by SNR as higher noise levels impact the convergence speed

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

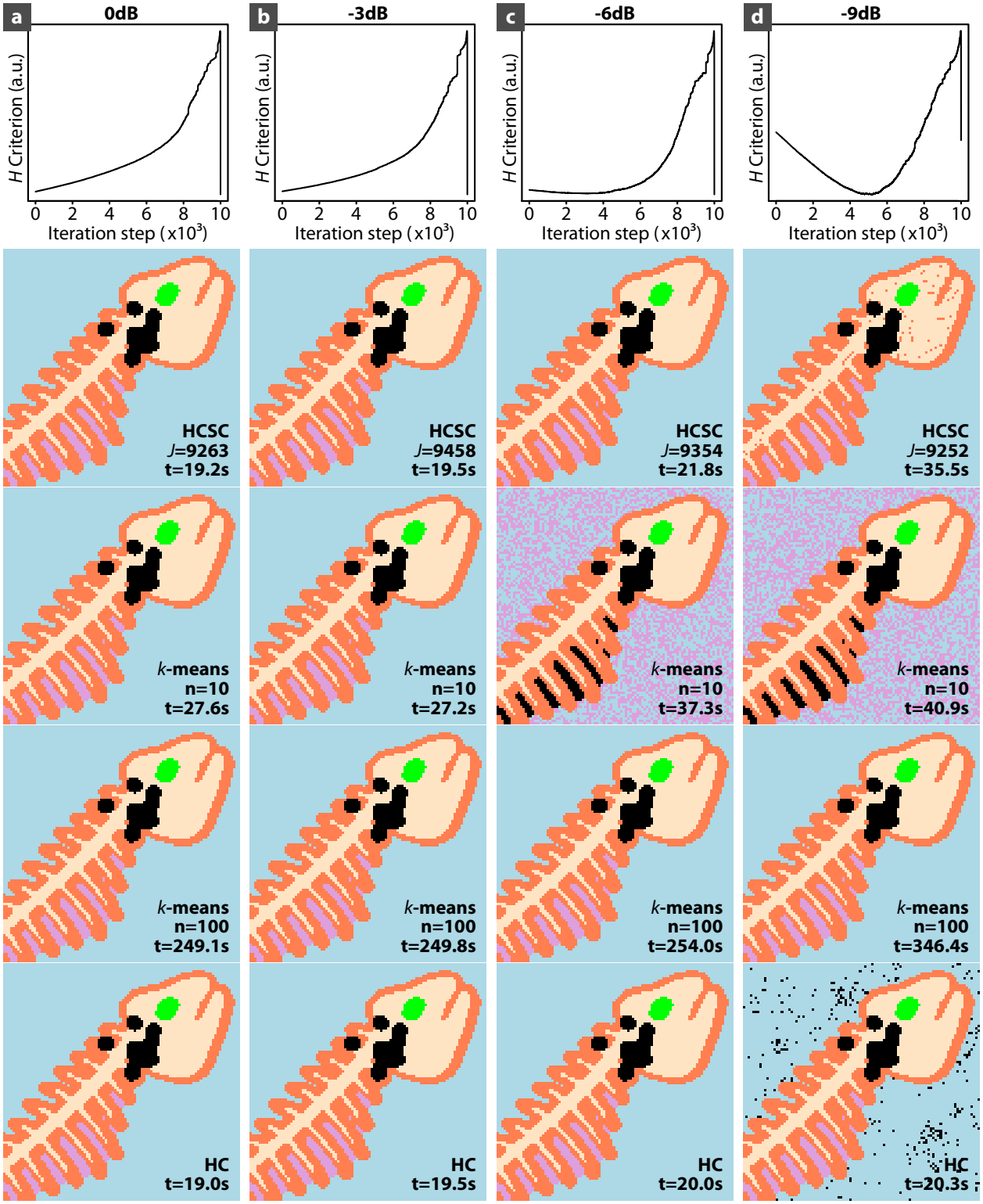

Figure 7. *Effectiveness and efficiency of our hierarchical clustering with spatial constraints as compared to that of* k*-means (with both 10 and 100 initializations) and standard hierarchical clustering (HC) on the purely synthetic data set, when removing 0dB (a), 3dB (b), 6dB (c) and 9dB (d).to the SNR. Note that HC was performed using the* `OpenCL` *accelerated version of the* `dist` *function to compute the complete distance matrix, using the default function would have led to much higher execution times as exposed in Tab. 1.*

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

of the algorithm. HCSC is in an intermediate position, with computational speed being affected by the SNR, but at a lower level than *k*-means.

In terms of effectiveness of classification, all three tested algorithms performed perfectly for the reference SNR (p0dB data set), showing not a single miss-classification. When the SNR decreases by 9dB, that is for the m9dB data set, both HC and HCSC start producing miss-classified pixels, with HC producing significantly more of those compared to HCSC. While both effectiveness and efficiency of HC and HCSC seems rather similar, even on the m9dB data set, one has to remember that HC has a $\mathcal{O}(N^2)$ memory complexity and $\mathcal{O}(N^2P)$ computing complexity as compared to only $\mathcal{O}(N)$ and $\mathcal{O}(NP)$ complexities for HCSC. In other words, the cost of HC will very rapidly increase to intractable levels when the image size increases, while HCSC will incur much lower costs with increasing image size. For larger images, effectiveness of HC and HCSC would likely remain similar but HCSC would be much more efficient than HC thanks to its complexity advantage. Indeed, this complexity difference is the reason why we could not run HC on the experimental data set used in Sect. 5.

Due to its stochastic nature, *k*-means has a fairly different behaviour than HC or HCSC. The user has to choose for a balance between quality of the result and computation time. For data sets m6dB and m9dB, with lower SNR, one may favour fast computation ($n = 10$ in Fig. 7) at the cost of poor effectiveness with a high level of miss-classified pixels, or quality of the results (no miss-classified pixels) at the cost of very low efficiency ($n = 100$, computational time of $346\,\mathrm{s}$). Note that the *k*-means results we are showing are overestimating the efficiency since the reported time corresponds only to processing the data set with a single value of K, known *a priori*, which you could typically not do on a real data set.

In this comparison, HCSC seems to provide a good balance between effectiveness and efficiency, and provides a rich set of information to the application domain specialist by proposing segmentations at different scales of complexity. Unlike HC, HCSC keeps this property for even rather large spectral images. Moreover for HCSC, we can notice on Fig. 7 that the curve of the criterion $H(J)$ used to determine the switching step $\hat{J}$ has an expected shape for data sets p0dB (Fig. 7a) and m3dB (Fig. 7b), while the shape begins to change for data set m6dB (Fig. 7c) and is significantly different for m9dB (Fig. 7d). This change in shape of the $H(J)$ might be used to propose an early detection of a SNR which starts to be too low for the data set to be exploitable. This change of behaviour when the SNR is decreasing will be further exemplified in Sect. 7, see Fig. 8.

Finally, comparing the results obtained on the real data set and the fully synthetic

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

data set as exposed respectively in Fig. 3 or Fig. 6 and Fig. 7, one may feel that the segmentation of the synthetic data set behaves more nicely than that of the real data set. The difference between the two data set might lead to the conclusion that the *uniform patches model*, used to generate the synthetic data, while being useful is still sufficiently wrong to significantly underestimate the variation existing between the spectra collected on the pixels of the image. Hence we propose, *infra*, a second synthetic *zero noise model* which produces spectra with variation closer to the one experienced in the real data set.

7. A more realistic synthetic model to assess robustness of the segmentation to signal to noise ratio

To assess the robustness of the proposed segmentation method in regard of the signal-to-noise ratio (SNR) we prepared a second simulated data set having features closer to the one of the experimental (real) data set used in Sect. 5, *i.e.* a spectral image with $N = 211 \times 241 = 50851$ pixels and at each pixel is associated a spectrum of size $P = 1780$ channels. As in Sect. 6, starting with a single zero noise model, we generated a family of simulated observations with a decreasing SNR. Performing the segmentation on this family of simulated data, which are all originating from the same generative model, enabled us to assess the effect of SNR levels on the proposed segmentation results. As compared to the model built in in Sect. 6, this one differs only in the way the zero noise model is prepared. While proposing a more realistic *ground truth* for the spectra of each pixel, the model proposed in the current section lacks *ground truth* for the class of each pixel. As a result, the segmentation resulting from HCSC can not be compared to some *true segmentation*. With this aspect in mind, we will mention here, as we did in Sect. 5, the *cluster*, rather than the *class*, of a pixel.

7.1. Building the *zero noise model*

We based our *zero noise model* on the above studied experimental data set that we regularized using local polynomial regression smoothing, through the `loess` function in `R` [26, 11]. To account for the nature and the dynamic of the signal on the observed X-ray fluorescence data, the weight was set to the reciprocal square root ($1/\sqrt{\cdot}$) of the observation when the observation is not 0, and to 1 otherwise. The second important parameter was the *span* of the filter that we set to 0.02 in order to account for the approximate width of the fluorescence bands on such spectra. Finally, a thresholding was performed on the regularized form so that its value is never lower than 0.001.

While this procedure is producing a realistic *zero noise* X-ray fluorescence spectra in each pixel of our image, it has to be noted that this should not be considered as a *ground*

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

truth version of the observation. Indeed, since each spectrum is dealt with independently from its neighbors, there is no spatial regularization and the *estimations* performed are far from optimal for detector channels that have measured a low level of photons.

One has to note that the protocol we use here to smooth the data is not valid as a denoising algorithm since it has some advert effects on the concentration of the trace elements, and in particular the REEs. Still, while the obtained spectra are not properly estimating the *ground truth* of this particular fossil, they have all the features making them likely to be present in a fossil. Hence, the generated data set should be considered as the XRF spectral image of a purely *phantom fossil*, enabling us to test the proposed hierarchical clustering algorithm on totally controlled data.

From this zero noise model, data sets with various SNR were generated in much the same ways as that exposed in Sect. 6.2, leading again to p0dB, m3dB, m6dB and m9dB data sets.

7.2. Impact of noise on hierarchical spatial clustering results

Following the same process as in Sect. 5.2, we use the criterion H to determine the switching step $\hat{J}$. In Fig. 8, we can see, as in Sect. 6.3, that the curve of H has the expected shape for data sets p0dB (Fig. 8a) and m3dB (Fig. 8c), while the shape begins to change for data set m6dB (Fig. 8e) and is significantly different for m9dB (Fig. 8g). The values obtained for these data set are: $\hat{J} = 44063$ for p0dB, $\hat{J} = 45871$ for m3dB, $\hat{J} = 48988$ for m6dB and $\hat{J} = 50387$ for m9dB. Here the higher the noise, the higher the $\hat{J}$, getting closer to the total number of pixels N in the image; the decrease in SNR leads to giving more relative weight to the spatial homogeneity, leading to the later release of the spatial constraint in the process.

According to the plot of the jump heuristic and CHC (curves not shown), and for the obtained clusters to be interesting from a paleontological point of view, the selected number of clusters is 11 for p0dB (Fig. 8b), 11 for m3dB (Fig. 8d) and 12 for m6dB (Fig. 8f). Concerning the m9dB data set, no fossil morphology can be seen when the selected number of clusters is 11 or lower, hence we have decided to represent the 12-cluster segmentation for this data set (Fig. 8h).

From the paleontological point of view, none of the obtained representations provides as much information as that of the original experimental data set. This is absolutely normal as the generated synthetic model is different from the original data. Instead, there are pros and cons to favoring the p0dB or m3dB segmentation (Fig. 8b,d; the former better resolves some bones, whereas only the latter features the muscles and highly absorbing

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

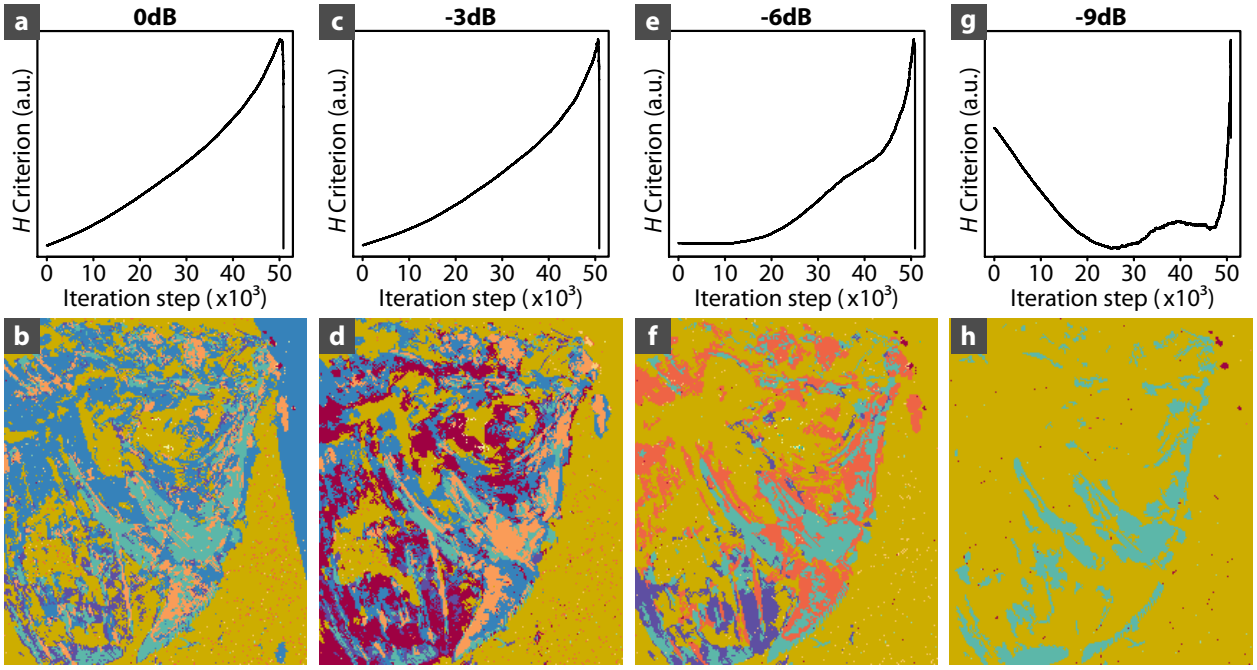

Figure 8. *Behavior of the H criterion for selecting the switching step J , and resulting HCSC images for the* zero noise model *(adding 0dB) (a,b), and after removing 3dB (c,d), 6dB (e,f) and 9dB (g,h) to the SNR.*

iron-rich grains, though without discriminating between them). Note that, in the p0dB representation (Fig. 8b), the blue triangular area that "appears" on the top right of the image and clusters with some of the fossil corresponds to air (there is no sample there, see Fig. 1a); it otherwise clusters with the sediment in the other noise models, which can be explained by the geometry used during the experiment where the beam came from the right of the sample with a 45° angle, leading for pixels in that area to record the X-rays-sediment interaction below the fossil surface.

Nevertheless, from m3dB and on, and as expected, the graphic representations quickly degrade with increasing noise, and most morphological information is lost for data set with a SNR greater than or equal to 6dB from the original data. Increasing further the level of noise leads to totally unexploitable data, with which the morphology of the sample could hardly be observed, as demonstrated on the m9dB simulation (Fig. 8h).

8. Discussion

In this article we propose a spatially constrained hierarchical clustering method to be applied on spectral images, in particular on energy resolved X-ray fluorescence images. The first aspect of the method is to choose an agglomerative criteria based on a dissim-

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

ilarity measure that is consistent with the noise model of the measured spectra. Then, the main aspect of this method is to apply constraints during the agglomerative process such that only spectra belonging to neighboring pixels could be clustered together. While this constraint is meaningful as long as the clusters form small patches on the image, it is obvious that when the number of cluster is small this spatial constraint should not be applied anymore, bringing the problem of the *proper step* at which the spatial constraint should be released. To address this problem, we proposed a heuristic criterion that balances the spatial coherence of the proposed segmentation, as measured through the G penalization, and the *between-cluster inertia* deriving from the Ward agglomerative criteria. The outcome of this algorithm is a hierarchy of possible segmentations that the practitioner should choose from. To aid this final selection step, the Ward and Calinski and Harabasz criteria are both computed to determine the most significant segmentation within the full hierarchy.

The advantages of such a *simple minded* algorithm is two-fold: first the general principles of the method do not require deep knowledge of statistical methods and as such can be grasped by the application domain specialist, the paleontologist in the presented example. Second, the computational cost of the segmentation is very low, even for a rather large data set, and the processing time, a few minutes, is significantly smaller than the typical measurement time for such spectral images, at least few hours. Hence, this method can be applied to the data while the experiment is still ongoing and used for a rapid diagnostic and experimental feedback within the global data acquisition strategy. As illustrated in Sect. 6, this is in sharp contrast with the other two simple methods we have tested, namely *k*-means and unconstrained hierarchical clustering (HC). HC has the required properties in term of processing speed for small spectral images, it is inapplicable as soon as the image size starts to increase; it seems also to be more sensitive to noise level than HCSC. The other tested method, *k*-means, while applicable to even large spectral images, is time consuming and, in its simple form, is not providing estimation of the number, K, of clusters present in the sample.

As a diagnostic tool, this method helps at finding a balance between a higher signal to noise ratio of individual spectra and the measurement time and radiation dose to which the sample is subjected. In such μXRF imaging modality, the SNR is inversely proportional to the square root of the radiation dose. Increasing the SNR increases the risk of producing radiation-induced damages to the sample during the experiment, but also often leads to increased measurement time and fewer (or smaller) samples being characterized in the allocated time slot. In such a situation it is therefore important to

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

quickly and properly assess the optimal exposure parameters (mostly time, but possibly also beam intensity), which need to be sufficient to produce exploitable spectra while avoiding any risk of radiation-induced damages to the sample and enabling large maps to be collected. Using simulated data, we have shown that the algorithm is robust to an increased level of measurement noise and as such is not only helpful in asserting an optimal measurement time but also in reducing it and lowering the radiation dose.

In the fully synthetic data set and in our SNR test application, it seems indeed that the behavior of H is a good early indicator of the quality of the observed data, providing insight into the discrimination power of the collected spectra. Indeed, the curves in Fig. 7d and Fig. 8g illustrate a behavior significantly different from the ones of Fig. 7a,b,c and of Fig. 8a,c,e. With respect to the SNR, we link this to the fact that the segmentation obtained in Fig. 8h is not very informative. In other words, the behavior of the H criterion as clusters get aggregated is a good predictor of the usefulness of the segmentation that will be attained for the data set.

Note that while the simulated data tested herein demonstrate that the behavior of the H criterion depends on the SNR, it also depends on the type of morphology of the sample being imaged and in particular in the patch density. Morphology is particularly having an effect when applying the proposed algorithm to only part of the image, as evidenced by our attempt to cut our real data set into four sub-images (see Sect. 5.2), one of which having a different morphology than the other three. In this case, the three sub-images with a very similar morphology were found to also have a similar H criterion curve, while the fourth sub-image, with mostly sediment and very little fossil features, produces a slightly different H criterion curve.

As a continuation of the present work, one could assess how the H criterion depends on patch density of the sample. From our currently limited experience, it seems likely that if the studied data set exhibits a similar type of morphology in all the image, a possibility is to choose a sub-image of size N_0 representative of the image morphology. The parameter δ_p can be then evaluated by $\delta_{p,0} = \frac{N_0 - \hat{J}_0}{N_0}$ (and $\hat{J}$ is chosen as the nearest integer to $(1-\delta_{p,0}) \times N$), leading to a significant reduction of the computational cost of the evaluation of this parameter. Furthermore, this promotes δ_p as a scalar descriptor of the image's morphology. Here again, we see an advantage of HCSC over *k*-means since this latter is not providing any synthetic figure characterizing the morphology of the sample.

Last but not least, this method provides to the practitioner a *complete* view of the information contained in a given spectral image data set. When such data are collected, the *prior knowledge* on the chemistry of the sample often leads to the selection of very

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

specific features of the spectra to be analyzed. Moreover, although entire μXRF spectra mostly contain XRF elemental information they also include additional, non-elemental signal including escape and sum peaks, as well as inelastic and elastic scattering and peaks from elements present in the air between the sample and the detector such as Ar (Fig. 1c). Depending on the sample, some of these peaks can carry interesting signal and one could need to keep them in the analysis. However, it is often preferred to remove them from the analysis and crop the spectra to the "true" elemental signal only, or only a few peaks, prior to the analysis. This can simply be done at the practitioner's discretion prior to applying the algorithm.

Conversely, we here propose to confront the result of such *focused* analysis with an analysis based on the full spectra. Indeed, both the *focused* and *complete* analysis could be performed using the same algorithm but selecting for each one either a subset or the fullset of the spectral channels of the image. Using such an approach the application scientist could both use the data in a *prior knowledge* directed approach, verifying pre-existing hypotheses on the nature of the signal to be detected in the spectral image, as well as an *unsupervised discovery* approach where the full spectral data set is subjected to the segmentation without *a priori* on which channel is of importance to exploit the image. Finally, this algorithm might even be used as a *post-hoc* analysis to test *a posteriori* the importance of unexpected spectral features, as exemplified herein with the iron-rich phase located around the fossil fish orbit for which the cluster mean spectrum provided complementary and new information to decipher its chemistry.

Contribution of authors.

This work arose from discussions between GC, SXC and AG. SXC proposed the exploitation of spatial constraints and the use of χ^2 as an adapted dissimilarity measure for XRF spectra. GC proposed the heuristic rule to stop applying spatial constraint on the segmentation. AG proposed a version of the χ^2 metric consistent between the spatially constrained initial steps and the unconstrained agglomerative steps, so that Lance and William formulae could be used in this latter part. SXC and AG implemented the algorithm and its result representations in R. SXC proposed and implemented the color matching in segmentation representation and the *zero noise models*. PG performed all the experimental measurements and interpretations on the fossil, and oriented the algorithm design to ensure results are valuable for the practitioner. All authors contributed to the writing of this manuscript.

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

Acknowledgments.

We thank S. Charbonnier, G. Clément, N.-E. Jalil, Didier B. Dutheil (MNHN, Paris), A. Tourani (Cadi Ayyad University, Marrakesh), P.M. Brito (Rio de Janeiro State University, Rio de Janeiro), F. Khaldoune, H. Bourget and B. Khalloufi for organizing and/or participating in the field work that collected the fossil. This field expedition to Morocco was supported by the Muséum national d'Histoire naturelle through the "ATM Biodiversité actuelle et fossile" and by UMR 7207 CR2P. We acknowledge Synchrotron SOLEIL for provision of beamtime, and C. Mocuta and D. Thiaudière for assistance at the DiffAbs beamline. Authors would also like to thanks the peers that reviewed the manuscript for their constructive comments and advices that helped us enhancing the presentation of our results.

Conflict of interest statement.

On behalf of all authors, the corresponding author states that there is no conflict of interest.

BIBLIOGRAPHY

[1] Alfeld, M., Janssens, K.: Strategies for processing mega-pixel x-ray fluorescence hyperspectral data: a case study on a version of Caravaggio's painting Supper at Emmaus. Journal of Analytical Atomic Spectrometry **30**(3), 777–789 (2015)

[2] Ambroise, C., Govaert, G.: Convergence of an EM-type algorithm for spatial clustering. Pattern recognition letters **19**, 919–327 (1998)

[3] Bergamaschi, A., Medjoubi, K., Messaoudi, C., Marco, S., Somogyi, A.: Mmx-i: data-processing software for multimodal x-ray imaging and tomography. Journal of Synchrotron Radiation **23**(3), 783–794 (2016)

[4] Bertrand, L., Cotte, M., Stampanoni, M., Thoury, M., Marone, F., Schöder, S.: Development and trends in synchrotron studies of ancient and historical materials. Physics Reports **519**(2), 51–96 (2012). DOI 10.1016/j.physrep.2012.03.003

[5] Bertrand, L., Robinet, L., Thoury, M., Janssens, K., Cohen, S.X., Schöder, S.: Cultural heritage and archaeology materials studied by synchrotron spectroscopy and imaging. Applied Physics A, Materials science & processing **106**(2), 377–396 (2012). DOI 10.1007/s00339-011-6686-4

[6] Bertrand, L., Thoury, M., Anheim, E.: Ancient materials specificities for their synchrotron examination and insights into their epistemological implications. Journal of Cultural Heritage **14(4)**, 277–289 (2013)

[7] Bertrand, L., Thoury, M., Gueriau, P., Anheim, É., Cohen, S.: Deciphering the chemistry of cultural heritage: Targeting material properties by coupling spectral imaging with image analysis. Accounts of Chemical Research **Online first**, 10.1021/acs.accounts.1c00063 (2021)

[8] Bonnet, N., Herbin, M., Vautrot, P.: Multivariate image analysis and segmentation in microanalysis. Scanning Microsc **11**, 1–21 (1997)

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

[9] Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and Regression Trees. Taylor & Francis (1984)

[10] Calinski, T., Harabasz, A.: A dendrite method for cluster analysis. Communications in Statistics **3**, 1–27 (1974)

[11] Cleveland, W., Grosse, E., Shyu, W.M.: Statistical Models in S, chap. Chapter 8 : Local Regression Models. Wadsworth & Brooks/Cole, New York (1992)

[12] Davesne, D., Gueriau, P., Dutheil, D., Bertrand, L.: Exceptional preservation of a cretaceous intestine provides a glimpse of the early ecological diversity of spiny-rayed fishes (acanthomorpha, teleostei). Scientific Reports **8**, 8509 (2018)

[13] Everitt, B.S., Landau, S., Leese, M., Stahl, D.: Cluster Analysis, 5th edition. Wiley (2010)

[14] Fiske, L.D., Katsaggelos, A.K., Aalders, M.C.G., Alfeld, M., Walton, M., Cossairt, O.: A data fusion method for the delayering of x-ray fluorescence images of painted works of art. In: 2021 IEEE International Conference on Image Processing (ICIP), pp. 3458–3462 (2021). DOI 10.1109/ICIP42928.2021.9506300

[15] Grabowski, B., Masarczyk, W., Głomb, P., Mendys, A.: Automatic pigment identification from hyperspectral data. Journal of Cultural Heritage **31**, 1–12 (2018)

[16] Gueriau, P., Bernard, S., Farges, F., Mocuta, C., Dutheil, D.B., Adatte, T., Bomou, B., Godet, M., Thiaudière, D., Charbonnier, S., et al.: Oxidative conditions can lead to exceptional preservation through phosphatization. Geology (2020)

[17] Gueriau, P., Jauvion, C., Mocuta, M.: Show me your yttrium, and i will tell you who you are: implications for fossil imaging. Palaeontology **61(6)**, 981–990 (2018)

[18] Gueriau, P., Mocuta, C., Bertrand, L.: Cerium anomaly at microscale in fossils. Analytical Chemistry **87(17)**, 8827–88367 (2015)

[19] Gueriau, P., Mocuta C.and Dutheil, D., Cohen, S., Thiaudière, D., the OT1 Consortium, Charbonnier, S., Clément, G., Bertrand, L.: Trace elemental imaging of rare earth elements discriminates tissues at microscale in flat fossils. PLoS One **9(1)**, e86946 (2014)

[20] Gueriau, P., Réguer, S., Leclercq, N., Cupello, C., Brito, P., Jauvion, C., Morel, S., Charbonnier, S., Thiaudière, D., Mocuta, C.: Visualizing mineralization processes and fossil anatomy using synchronous synchrotron X-ray fluorescence and X-ray diffraction mapping. Journal of the Royal Society Interface **17**(169), 20200216 (2020). DOI 10.1098/rsif.2020.0216

[21] Lance, G.N., Williams, W.T.: A general theory of classificatory sorting strategies: II. Clustering systems. The Computer Journal **10**(3), 271–277 (1967). DOI 10.1093/comjnl/10.3.271

[22] Lebart, L.: Programme d'agrégation avec contrainte. Cahiers de L'analyse des Données **3**, 275–287 (1978)

[23] Mihalić, I.B., Fazinić, S., Barac, M., Karydas, A.G., Migliori, A., Doračić, D., Desnica, V., Mudronja, D., Krstić, D.: Multivariate analysis of pixe+ xrf and pixe spectral images. Journal of Analytical Atomic Spectrometry **36**(3), 654–667 (2021)

[24] Milligan, G., Cooper, M.: An examination of procedures for determining the number of clusters in a data set. Psychometrika **50**, 159–179 (1985)

[25] Pouyet, E., Rohani, N., Katsaggelos, A.K., Cossairt, O., Walton, M.: Innovative data reduction and visualization strategy for hyperspectral imaging datasets using t-sne approach. Pure and Applied Chemistry **90**(3), 493–506 (2018)

[26] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s

Computing, Vienna, Austria (2020). URL https://www.R-project.org/

[27] Rand, W.M.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association **66**(336), 846–850 (1971)

[28] Rodriguez, M.A., Kotula, P.G., Griego, J.J., Heath, J.E., Bauer, S.J., Wesolowski, D.E.: Multivariate statistical analysis of micro-x-ray fluorescence spectral images. Powder Diffraction **27**(2), 108–113 (2012)

[29] Sciutto, G., Oliveri, P., Prati, S., Quaranta, M., Bersani, S., Mazzeo, R.: An advanced multivariate approach for processing x-ray fluorescence spectral and hyperspectral data from non-invasive in situ analyses on painted surfaces. Analytica chimica acta **752**, 30–38 (2012)

[30] Solé, V.A., Papillon, E., Cotte, M., Walter, P., Susini, J.: A multiplatform code for the analysis of energy-dispersive x-ray fluorescence spectra. Spectrochimica Acta B **62**, 63–68 (2007)

[31] Vekemans, B., Janssens, K., Vincze, L., Aerts, A., Adams, F., Hertogen, J.: Automated segmentation of μ-xrf image sets. X-Ray Spectrometry **26**(6), 333–346 (1997)

[32] Vogt, S., Maser, J., Jacobsen, C.: Data analysis for x-ray fluorescence imaging. In: Journal de Physique IV (Proceedings), vol. 104, pp. 617–622. EDP sciences (2003)

[33] Ward, J.H.: Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association **58**, 236–244 (1963)

[34] Webb, S.: The microanalysis toolkit: X-ray fluorescence image processing software. In: AIP Conference Proceedings, vol. 1365, pp. 196–199. American Institute of Physics (2011)

Author's version, published in SN Computer Science; 10.1007/s42979-022-01074-4s