



HAL
open science

Hierarchical clustering of spectral images with spatial constraints for the rapid processing of large and heterogeneous datasets from ancient material studies

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 datasets from ancient material studies. 2021. hal-03104488v1

HAL Id: hal-03104488

<https://uvsq.hal.science/hal-03104488v1>

Preprint submitted on 8 Jan 2021 (v1), last revised 10 Apr 2022 (v4)

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.

Hierarchical clustering of spectral images with spatial constraints for the rapid processing of large and heterogeneous datasets from ancient material studies

Gilles Celeux^{*}, Serge X. Cohen[†], Agnès Grimaud[‡], and Pierre Gueriau[§]

Abstract. The study of very complex and heterogeneous materials, such as those encountered in the science of ancient materials, benefits from the wealth of information provided by the acquisition and exploitation of full spectrum images, *i.e.* spectral images. In order to obtain a high dynamic range in both the spatial dimensions and composition, great efforts have made it possible to considerably accelerate data collection and increase the average size of a single data set, each image reaching up to several tens of GB. Rapid processing is now required to allow feedback during data collection, within the short time available for instruments and samples. Here we propose an approach combining hierarchical clustering and spatial constraint. Spatial constraints allow both a significant reduction in the computational cost of segmentation and a certain level of robustness with respect to the signal-to-noise ratio: the *prior* knowledge injected by the spatial constraint partially compensates for the increase in noise level; hierarchical clustering provides a statistically sound and known framework that allows accurate reporting of the instrument noise model. 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. It can be foreseen how such an approach could simultaneously lead to an increase in the spatial definition of the collected spectral image and to a reduction in the potentially harmful radiation dose density to which the samples are subjected.

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

1. Introduction

Ancient materials, studied by archaeology, paleontology or as part of the cultural heritage research, are very diverse but share the particularity of being composite and heterogeneous on several scales [6]. Moreover, they are 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]. In this context, spectral imaging, *i.e.* images for which each pixel is characterized by a full spectrum (see *e.g.* Figure 1), is a tool of

^{*}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.

32 choice for simultaneously obtaining physico-chemical information (*e.g.* elemental,
33 chemical or mineralogical composition), and the morphological information essen-
34 tial for understanding the behavior of the material over long periods of time. Such
35 datasets make several GB or even tens of GB, depending on the type of detection
36 used. Indeed, while 1D detectors typically record thousands of values (*i.e.* a few KB)
37 per pixel, 2D array detectors record images of several MB per pixel (*e.g.* [1, 3, 16]).
38 As such, these datasets are too massive to be timely exploited with standard algo-
39 rithms and we need to develop algorithms able to analyze such images in a time-
40 frame compatible with the data collection time to provide feedback possibilities on
41 the measurements (*e.g.* [1, 3]). One should also consider that on those measure-
42 ments involving a probe, increasing the signal to noise ratio (SNR) comes at a cost:
43 increasing probe/material interaction indeed most often leads to longer measure-
44 ment times and always to a higher radiation dose deposited in the material. In such
45 a framework one has to find a balance between SNR and dose/time, so that the
46 experiment is conclusive without producing alteration of the samples during the
47 analysis (*e.g.* [14]).

48 In this article, we focus on the question of image segmentation when the dataset
49 comes from X-ray fluorescence (XRF) mapping, a technique by which each individual
50 pixel is characterized by its XRF spectrum, providing elemental composition infor-
51 mation on that pixel (Figure 1). The classical approach to plot quickly or even *live*
52 elemental distributions recorded by XRF mapping consists of integrating the signal
53 (*i.e.* photons counted by the detector) in spectral regions of interest (ROI) correspond-
54 ing to targeted element peaks. This does not, however, hold true elemental distribu-
55 tion images as such ROI integrations additionally include significant contributions
56 from other elements or phenomena (namely scattering, and sum and escape peaks);
57 these overlapping biases can only be circumvented by applying slower approaches
58 allowing a spectral decomposition of the dataset (*e.g.* [15, 1]). Here, we propose
59 a *hierarchical segmentation* algorithm combining the characteristics of hierarchical
60 clustering with the imaging properties of a composite material. In other words, we
61 aim at proposing a hierarchical classification procedure of spectral dissimilarities
62 allowing to take into account the spatial proximities between the pixels.

63 It is important to understand the nature of the signal measured in such experi-
64 ment. In XRF, we measure the energy of the photons emitted by the material when
65 it is subjected to monochromatic incident radiation. Because this re-emission phe-
66 nomenon is a stochastic process, the measured spectrum is an empirical sampling

67 of the law of this process. Instead of analyzing the signal using *generic tools* for
 68 Euclidean spaces, such as the ℓ_2 distance, it is therefore more relevant to use tools
 69 adapted to the comparison of population samples. On this respect, the algorithm
 70 we propose is based on the χ^2 as a tool to assess homogeneity between two samples,
 71 in the present case two pixels of different compositions.

72 After defining the terms and notations used throughout this article, we expose
 73 the general framework of our dissimilarity measure (using a Ward criteria based on
 74 χ^2 , [subsection 2.2](#)), and then propose an approach to impose spatial constraint upon
 75 the agglomerative process of the hierarchical clustering in [subsection 2.3](#). Then, in
 76 [section 3](#), we concentrate on the proper steps at which the spatial constraint should
 77 be released to properly account for non connex domains made of the same material.
 78 We further consider the appropriate number of classes at which the agglomeration
 79 process should be stopped in [section 4](#). To illustrate our approach, we apply the
 80 proposed algorithm on a true dataset corresponding to the XRF mapping of a fossil
 81 teleost fish, including both the analysis of the experimental dataset in [section 5](#),
 82 and, in [section 6](#), the analysis of a synthetic dataset resembling the experimental
 83 one but providing the possibility to simulate various signal to noise ratio and giving
 84 insight into the robustness of the proposed algorithm to the noise level.

85 2. The proposed hierarchical clustering method

86 2.1. Notations and definitions

87 A spectral image of N pixels is considered. Each pixel $i \in \{1, \dots, N\}$ is character-
 88 ized by a spectrum $\mathcal{S}_i = (s_i(p))_{p \in \{1, \dots, P\}}$, where $s_i(p)$ is the number of photon counts
 89 for pixel i in energy canal p .

90 For $i \in \{1, \dots, N\}$ and $p \in \{1, \dots, P\}$, let

$$91 \quad f_{i,p} = \frac{s_i(p)}{s_{i\cdot}} \quad \text{and} \quad t_p^i = \frac{s_i(p)}{s_{i\cdot}} \quad \text{where} \quad s_{i\cdot} = \sum_{p=1}^P s_i(p) \quad \text{and} \quad s_{\cdot\cdot} = \sum_{i=1}^N s_{i\cdot}.$$

92 The aim is to propose a hierarchical classification procedure of spectra $(\mathcal{S}_i)_{i \in \{1, \dots, N\}}$
 93 using the conditional distributions (or profiles) of pixels $((t_p^i)_{p \in \{1, \dots, P\}}, i \in \{1, \dots, N\})$,
 94 the pixel i being weighted by $f_{i\cdot} = s_{i\cdot}/s_{\cdot\cdot}$ ($i \in \{1, \dots, N\}$). Since these profiles are prob-
 95 ability distributions, they are compared using the χ^2 distance.

For two pixels i and j , let

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

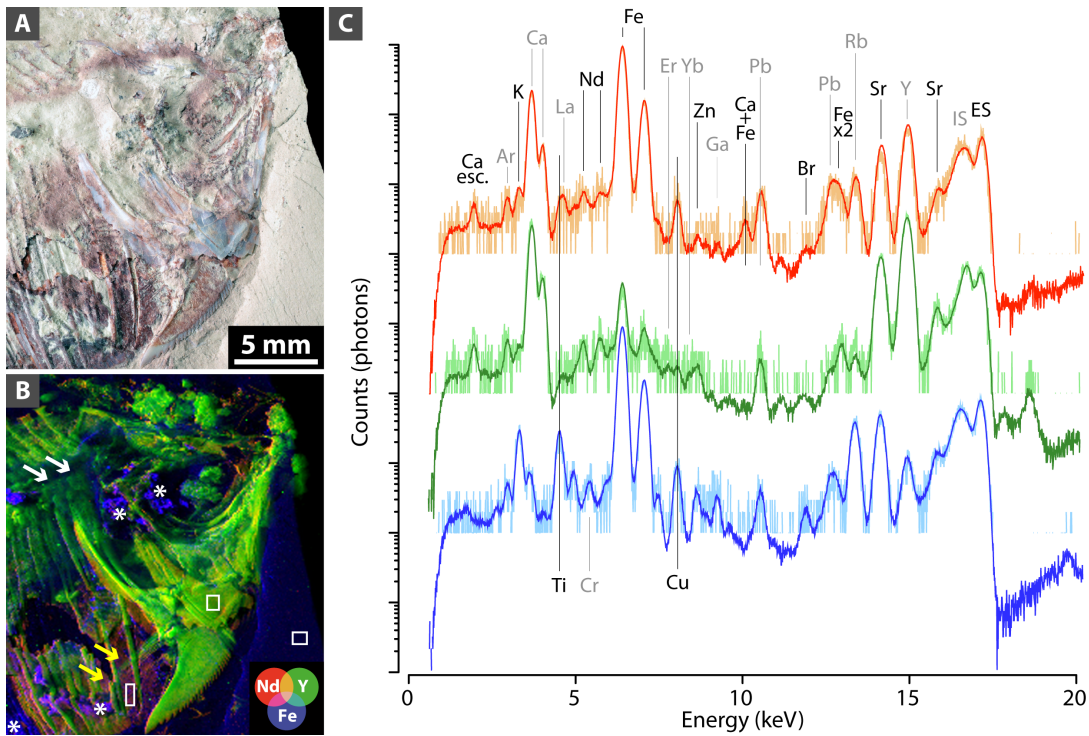


Figure 1. Synchrotron XRF mapping of major-to-trace elements of the anterior part (skull on the right) of the yet undescribed fish MHNM-KK-OT 03a from the Jbel Oum Tkout Lagerstätte (Upper Cretaceous, 100 Myr, Morocco). (A): optical photograph. (B): false color overlay of the distributions of two rare earth elements, neodymium (red) and yttrium (green), and of iron (blue), reconstructed from a full spectral decomposition of the data (modified from [15]). Acquisition parameters: $100 \times 100 \mu\text{m}^2$ 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 sedimentary matrix (dark and light blue), respectively. Spectra are shown using a logarithmic 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.

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

97 **Remarks:**

- 98 • It is assumed that $f_{.p} \neq 0$ for all p . If there is a canal p such that $f_{.p} = 0$, then
 99 $s_i(p) = 0$ for all $i \in \{1, \dots, N\}$, hence $t_p^i = t_p^j = 0$ for all $(i, j) \in \{1, \dots, N\}^2$. Thus,
 100 such canals are removed beforehand.
- 101 • It is assumed that $s_{i.} \neq 0$ for all pixel i (otherwise it would mean that the
 102 detector did not received any photon for the corresponding pixel).

2.2. The Ward criterion

Using the χ^2 distance as the proximity measure between the spectra of pixels, the hierarchical clustering is designed with the agglomerative Ward criterion δ_{χ^2} [22], which consists of minimizing the increase of the within-cluster inertia at each step. This agglomerative criterion for two clusters C and C' is:

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

where $\mu_C = \sum_{i \in C} f_i$ is the weight of cluster C and S_{g_C} the gravity center of cluster C ;

$$S_{g_C} = (g_c(p))_{p \in \{1, \dots, P\}} \text{ with } g_c(p) = \frac{1}{\mu_C} \sum_{i \in C} f_i \cdot t_p^i.$$

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 special occurrence of the general Lance and Williams formula, see [11, 17] 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:

$$\delta_{\chi^2}(C_k, C_i \cup C_j) = \frac{(\mu_{C_k} + \mu_{C_i})\delta_{\chi^2}(C_k, C_i) + (\mu_{C_k} + \mu_{C_j})\delta_{\chi^2}(C_k, C_j) - \mu_{C_k}\delta_{\chi^2}(C_i, C_j)}{\mu_{C_i} + \mu_{C_j} + \mu_{C_k}}.$$

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), hence it is computationally too expensive ($\mathcal{O}(N^2P) \approx 5 \times 10^{12}$ operations to design directly a hierarchical clustering with the Ward criterion described above). Moreover, it is desirable that the clusters form unions of spatially connected sub-clusters.

For these two reasons, following [18], 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 (Figure 2). It implies that most of the pixels have eight neighbors (Figure 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

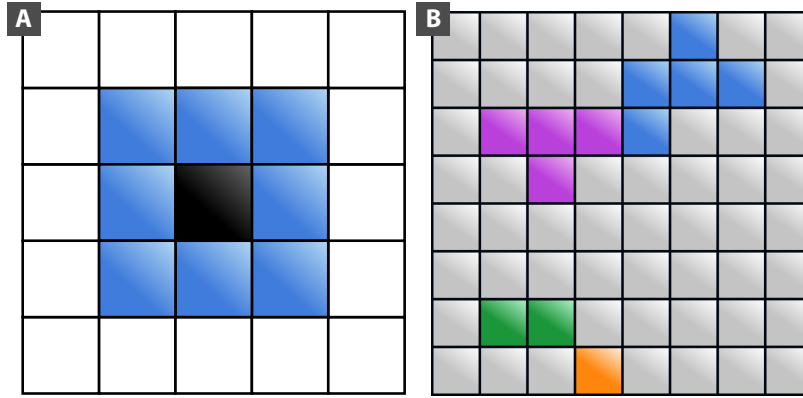


Figure 2. Schematic representation of the second-order neighborhoods approach. (A): neighbors for a pixel that is not located on an edge or at a corner. (B): example of clusters spatially neighboring; on the top, the blue and purple clusters are spatially neighboring, while on the bottom left the green 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.

127 dissimilarities have to be computed. Nevertheless, for this reason, it is not possi-
 128 ble to use the Lance and Williams formula [11, 17] to update the dissimilarities.
 129 Therefore, equation (2.1) is used to compute the dissimilarities needed to design the
 130 hierarchy.

131

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

Algorithm 2.1 Hierarchical spatial clustering

Initialization : computes the χ^2 distances between two spectra for neighboring pixels

Define $J := 1$

while $J < N$ **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).

$J := J + 1$

end while

134 If this algorithm is run until it remains only two clusters, we get a hierarchy
 135 where at each step the clusters are spatially connected. However, it is not desirable
 136 to impose such clusters connexion during the final steps. Indeed, from the point
 137 of view of the application domain scientist/specialist, the relevant clusters, while

138 connected at fine scale, have no reason to be spatially connected at large scale.

139 As a consequence, the proposed algorithm taking the spatial constraint into ac-
 140 count is run for J steps leading to spatially connected clusters, J being large (the
 141 choice of the switching step J will be discussed hereafter in [section 3](#)). It leads
 142 to $(N - J)$ spatially connected clusters, hereafter called *patches*. Then from these
 143 $(N - J)$ patches, unconstrained agglomerative hierarchical clustering algorithm with
 144 the Ward criterion is used. Thus, the proposed final clusters are union of the J spa-
 145 tially connected patches. Obviously, a relevant number of final clusters is to be
 146 chosen; this point is discussed in [section 4](#).

147 3. Selecting the switching step J

148 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:

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

149 where $\alpha \in \mathbb{R}_+$, $B(J)$ is the between-cluster inertia of a partition of the pixels into J
 150 patches and $G(J)$ is a measure of the spatial homogeneity of this partition. Following
 151 [\[2\]](#), we consider

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

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

154

155 The important point is to choose the scalar α to get an equilibrium between $B(J)$
 156 and $G(J)$. In the extreme situation of a partition into N patches, we have $G(N) = 0$
 157 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 sim-
 158 plicity, we consider here improperly that each pixel has 8 neighbors.)

159

160 Assuming an equilibrium $H(N) = H(1)$ between these two extreme situations
 161 leads to $\alpha = \frac{T}{4N}$, T being the total inertia of the whole set of pixels. Thus, the

162 criterion to be maximized is

$$163 \quad H(J) = B(J) + \frac{T}{4N}G(J).$$

164

165

166 However, as it will be apparent in the case study in [section 5](#), this choice of
 167 α leads to the selection of a too large number of patches J_{\max} . In order to select
 168 a more relevant number of patches, J , from which to release spatial constraints
 169 in the clustering, we propose to make use of the “one standard deviation” proce-
 170 dure proposed in [7] to cut a decision tree. This procedure consists of computing
 171 $H(K)$ for $K = N, \dots, 1$, then to compute the standard error $\text{sd}(H)$ of the resulting
 172 $(H(K))_{K=N, \dots, 1}$ and choosing the smallest \hat{J} such that

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

173 The rationale for this procedure is to determine the value of \hat{J} that corresponds
 174 to a large number of clusters and provides a good compromise between the between-
 175 cluster inertia and the spatial homogeneity. Note that, while it is expected that
 176 $H(J)$ increases from $H(N)$ to $H(J_{\max})$ and then decreasing back to $H(1)$, there is no
 177 guarantee for such a behaviour. Such an unexpected behaviour of the H criterion
 178 is obtained for low signal to noise ratio image in [Figure 6E](#) and [G](#). In order to also
 179 address these types of behaviour of H we express the choice of \hat{J} in a different way :

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

180 3.2. In practice

181 In practice, depending on the image size, it can be too long or impossible to
 182 compute $H(K)$ for all $K \in \{N, \dots, 1\}$. In this case, the following heuristic approach is
 183 proposed to determine \hat{J} :

184 As a first step, the idea is to compute $(H(\ell \times by))_l$, with $by \geq 1$ chosen to have a
 185 reasonable computing time and $\ell \in \mathbb{N}^*$ such that $\ell \times by \leq N$. However, after testing
 186 on the studied dataset described in [section 5](#), we noticed that the obtained \hat{J} can
 187 change significantly according to the value by . Hence, by must be small enough to
 188 obtain a correct value for \hat{J} (but we do not know its order of magnitude).

189 On the other hand we noticed that the obtained values for the ratios $\beta_k = \frac{\hat{J}_k}{N_k}$ (\hat{J}_k is

190 computed with the criterion described in **subsection 3.1**) are similar for sub-images
 191 $(I_k)_k$ of size $(N_k)_k$ having a similar type of morphology.

192 Therefore we propose to evaluate the “constant” $\beta = \frac{\hat{J}}{N}$, which seems linked to the
 193 morphology of the studied image, by cutting the image into q sub-images $(I_k)_{k \in \{1, \dots, q\}}$
 194 with size $(N_k)_{k \in \{1, \dots, q\}}$. Then, for each sub-image, \hat{J}_k is computed with either the
 195 criterion described in **subsection 3.1** or the introduction of a small value for by , and
 196 the terms $\beta_k = \hat{J}_k/N_k$ are computed for all $k \in \{1, \dots, q\}$.

197 Finally, \hat{J} is chosen as the nearest integer to $\bar{\beta} \times N$, with $\bar{\beta} = \frac{1}{q} \sum_{k=1}^q \beta_k$.

198 4. Selecting the number of clusters

199 4.1. Statistical heuristics

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

204 An another natural and popular criterion for choosing a relevant number of clus-
 205 ters K in a hierarchy designed with the Ward criterion is to use the value of K cor-
 206 responding to the maximum value of the Calinski and Harabasz criterion (CHC, [8])

$$207 \quad \text{CHC}(K) = \frac{\text{Tr}(B_K)}{(K-1)} / \frac{\text{Tr}(W_K)}{(N-K)},$$

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

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

and

$$\text{Tr}(W_K) = \sum_{k=1}^K \sum_{i \in C_k} f_i d_{\chi^2}^2(S_{g_{C_k}}, S_i)$$

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

210 This criterion has been shown to perform well in practical situations (see [19]).

4.2. Particular considerations in the case of spectral images from Ancient materials science

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 in the case of ancient material studies. Instead, we here prefer to use the following strategy:

- Preselect several number of clusters using the jump heuristic and 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 [section 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.

5. Application to a real world dataset

5.1. Data description

The proposed algorithm has been applied to a spectral image dataset collected on a yet undescribed *ca.* 100-million-year-old new teleost fish from Morocco ([Figure 1](#), [15]). The information embedded in this dataset is a synchrotron micro-X-ray fluorescence (μ XRF) major-to-trace-elemental map, where a full XRF spectrum has been recorded for each pixel, over a $22.5 \times 22.5 \text{ mm}^2$ area using a scan step of $100 \times 100 \mu\text{m}^2$ and a 500 ms counting time ($211 \times 241 = 50,851$ pixels in total; [Figure 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 $10 \times 7 \mu\text{m}^2$.

Very interestingly, the distribution of strontium and yttrium $K\alpha$ lines, which substitute for calcium in calcium phosphates such as bone apatite [15, 13] 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 ([Figure 1B](#), [15]). They particularly unveil the morphology of the first vertebrae (white arrows in [Figure 1B](#)), the neurocranium that extends into a sharp

245 supraoccipital at the top of the skull, the metapterygoid, and the hyomandibular
246 that appears dorsally flared. These new information help deciphering the affinities
247 of this new fossil species (in preparation). The other main outcome of this work was
248 that a false color overlay of the distribution of different rare earth elements (REEs;
249 e.g. neodymium and yttrium, red and green distributions in [Figure 1B](#), respectively)
250 discriminates phosphatized muscles (yellow arrows in [Figure 1B](#)) and bone [[15](#)].

251

252 5.2. Resulting hierarchical spatial clustering

253 In the following, the proposed algorithm has been implemented with R [[20](#)] on
254 this image of $N = 211 \times 241 = 50,851$ pixels, for which at each pixel i the spectrum S_i
255 has $P = 1780$ values. The size of the file containing the dataset is 1.7 GB.

256 For such a dataset, it is too long to compute the criterion H described in [section 3](#)
257 for all $K \in \{N, \dots, 1\}$ in order to determine the switching step \hat{J} . Hence, the image has
258 been cut into 4 sub-images with size 100×115 and \hat{J} has been computed as described
259 in [subsection 3.2](#). As explained in [subsection 3.2](#), to limit the computational cost
260 of the complete algorithm, the $(\beta_k)_k$ values are estimated on the four sub-images
261 with H computed every 5 agglomerative steps ($by = 5$). We obtained $(\beta_k)_k$ values of
262 0.84087, 0.93043, 0.85565 and 0.85522 respectively for top right, bottom right, bottom
263 left and top left sub-images leading to $\bar{\beta}_4 \simeq 0.87054$ and consequently $\hat{J} = 44268$.
264 Note that the $(\beta_k)_k$ values are close for the three sub-images having a similar type
265 of morphology, whilst the bottom-right sub-image consist mostly of sediment and is
266 more homogeneous than other sub-images.

267 To select the number of clusters, we plotted the jump heuristic and the CHC
268 against the number of clusters (starting with two clusters) ([Figure 3](#)). These criteria
269 are here to complement the knowledge of the application domain specialist, in the
270 case of the present example a paleontologist (PG). The jump heuristic leads to pro-
271 pose 6 clusters, whilst the CHC leads to propose either 6 or 9 clusters, corresponding
272 to the two local maxima of the curve.

273 Looking more precisely at the differences between 6 and 9 clusters, it can be seen
274 that only the light green cluster of the 6-cluster solution is modified ([Figure 3C-E](#)).
275 The difference image ([Figure 3E](#)) illustrates how the light green cluster incorporates
276 three smaller clusters: one with 5 black isolated pixels, one green cluster with 50
277 isolated pixels, and a light-red cluster with 3014 pixels spread out it several patches.
278 If the additional clusters with 5 and 50 isolated pixels seem of little interest, the light-

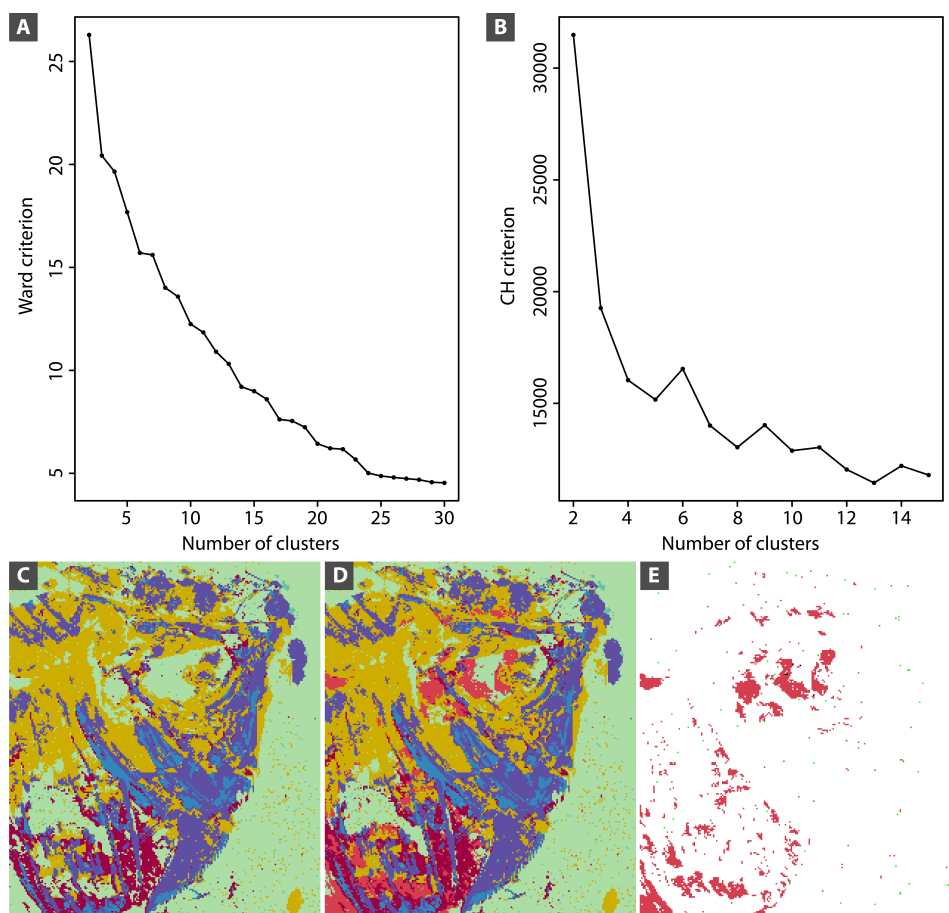


Figure 3. Defining the number of clusters used for hierarchical segmentation. (A, B): Ward (A) and Calinski and Harabasz (B) criteria against the number of clusters (starting with 2 clusters). (C–E): False color distributions obtained for 6 (C) and 9 (D) clusters, and difference (E).

279 red cluster appears to be interesting, as it highlights areas richer in iron (asterisks in
 280 **Figure 1B**) that are not clearly obvious in the μ XRF elemental maps obtained using
 281 ROI integration or spectral decomposition.

282 From a paleontological point of view, the segmentation offered by the selected
 283 clustering (**Figure 4A**) does not improve the visualization of hidden anatomical de-
 284 tails, but provides new insights into the chemical composition of the different tissues
 285 and materials present in the sample through the mean spectra of the clusters (**Fig-
 286 ure 4B**). While individual elemental distributions show no strong contrast in the
 287 incorporation of light REEs between bone and muscles (**Figure 4C**), following the
 288 distribution of calcium, which they substitute and that originates from a compara-
 289 ble depth (**Figure 4D**), the yttrium distribution shows strong enrichment in the bone

290 as compared to the muscles (Figure 4E). In fact, in the muscles area (yellow arrows
291 in Figure 1B), rather than following the type of tissue the yttrium distribution largely
292 follows the thickness of the material as shown by X-ray microtomography where most
293 of the muscles region appear to be very thin or not discernible (Figure 4F). Conse-
294 quently, thickness and information depth were likely responsible for the apparent
295 REE contrast. Nevertheless, the selected clustering clearly discriminates bone from
296 phosphatized muscles (blue/purple and dark red clusters in Figure 4A, respectively)
297 on the basis on the full μ XRF spectra. The muscles dark red cluster appears richer in
298 Fe and Pb (Figure 4B), which come from a reddish fossil biofilm made of iron hydrox-
299 ides and covering the phosphatized muscles [14, 10, 12] rather than the phosphatic
300 material itself. In turn, the bone blue and purple clusters contain much higher con-
301 tents in heavier REEs ($L\beta 1$ emission lines from erbium and ytterbium particularly
302 stand out in Figure 4B, as they do not fall in the same energy domain as major ele-
303 ments [15]). This is most likely again an effect of information depths and thickness
304 of the tissues.

305 On another hand, the selected clustering isolates well the large, highly absorbing
306 iron grains situated posteriorly to the orbit (asterisks in Figure 1B; light red cluster
307 in Figure 4A, B) that prevent segmentation of the first vertebrae and posterior part
308 of the head from the X-ray tomography data (asterisks in Figure 4F). These grains
309 are particularly rich in Fe, Ti, Cu and Ga, but not so much in Pb (Figure 4B) and are
310 therefore, besides their larger size, a different material than the reddish thin film of
311 iron hydroxides covering most of the fossil.

312 By providing a global discrimination of the different materials composing the
313 fossil much faster than a full spectral decomposition (approximately two hours
314 here for a computer with specifications i5-4590 @ 3.30GHz, 4 Core, four days using
315 the freeware PyMCA [21]), the proposed clustering methodology provides a robust
316 and quick way to extract, “live” at the beamline, chemical information not hampered
317 by local heterogeneity or contamination for further higher resolution mapping of
318 areas of interest, or point analyzes using, e.g., X-ray absorption spectroscopy.

319 5.3. Regarding the chosen switching step J

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

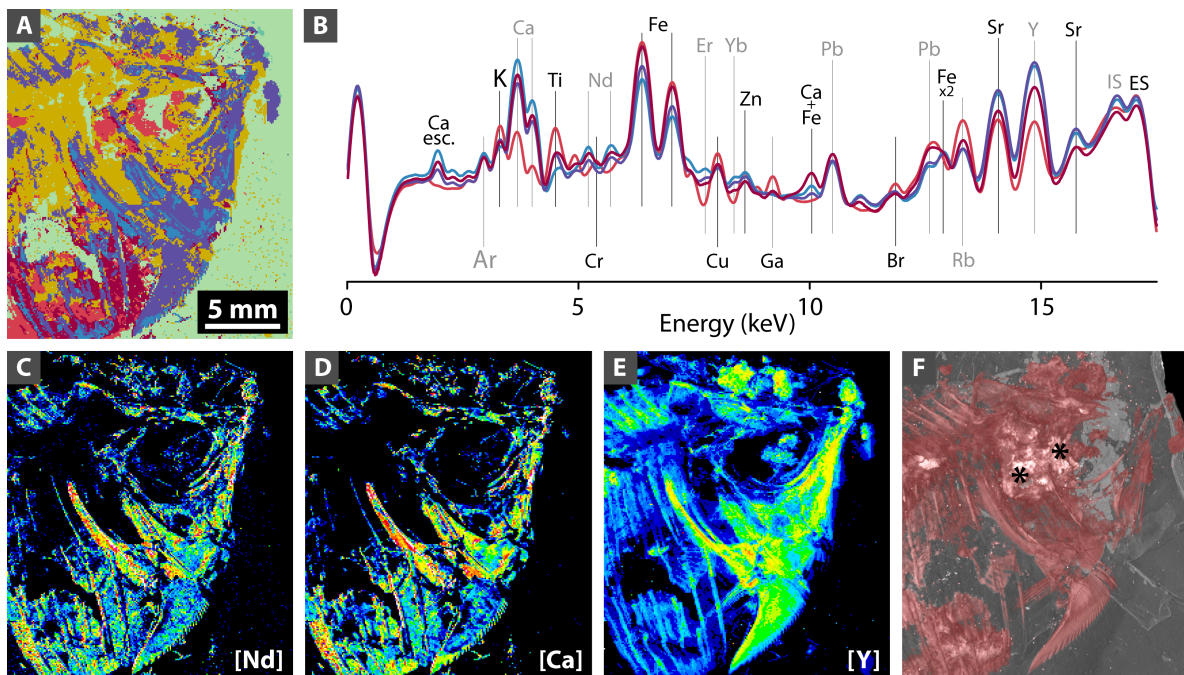


Figure 4. Hierarchical segmentation of the synchrotron μ XRF spectral image dataset of the yet undescribed fish (MHNM-KK-OT 03a) from the Jbel Oum Tkout Lagerstätte (Upper Cretaceous, 100 Myr, Morocco). (A): Segmentation results when 9 classes are selected with the proposed algorithm, disabling spatial constraint at agglomerative step $\hat{J} = 44268$. (B): mean spectra from 4 of the 9 classes 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: $(24.7 \text{ mm})^3$.

324 heuristic and the CHC for $J = 43000$ and $J = 46000$, respectively.

325 For $J = 43000$, the jump heuristic plot leads to propose 6 or 10 clusters while the
 326 CHC leads to 3, 10 or 12 clusters (Figure 5A,B). For $J = 46000$, the jump heuristic
 327 plot leads to propose 5 clusters or maybe 9, and the CHC leads to propose 9
 328 clusters (first local maximum) or more (Figure 5E,F). These results show that the
 329 value of the switching step has an influence on the result of the hierarchical cluster-
 330 ing. Comparisons of the graphic representations for $J = 43000$, 44268 and 46000
 331 (Figure 5 C,D,G) clearly identify the segmentation resulting from the latter as abso-
 332 lutely unsatisfactory as many fossil areas are found mixed up with the surrounding
 333 sediment (Figure 5G). Graphic representations for $J = 43000$ and 44268 appear in
 334 turn very similar. Nevertheless, representation for $J = 44268$ (the computed \hat{J} value,
 335 see subsection 5.2) more accurately reflects elemental distributions (Figure 1B), par-

336 ticularly regarding the iron-rich phase located around the fish orbit.

337 6. Assessing robustness of the segmentation to signal to noise ratio

338 To assess the robustness of the proposed segmentation method in regard of the
339 signal-to-noise ratio (SNR) we prepared simulated data having features close to the
340 one of the experimental dataset used in the previous section. Starting with a single
341 realistic model, we generated a family of simulated observation with a decreasing
342 SNR. Performing the segmentation on this family of simulated data, which are all
343 originating from the same generative model, enabled us to assess the effect of SNR
344 levels on the proposed segmentation results. To achieve consistency we generated
345 this simulated dataset in two steps: (i) we constructed a *zero noise model* that would
346 correspond to a likely observed object; (ii) from this *zero noise model* we generated
347 simulated observation by applying a noise generation process that mimics the phys-
348 ical observation process while providing control on the noise level of the simulated
349 data. We will present the two steps of this procedure, then the results in terms of
350 segmentation.

351 6.1. Building a *zero noise model* and simulating data with controlled SNR

352 We based our *zero noise model* on the above studied experimental dataset that
353 we regularized using local polynomial regression smoothing, through the `loess` func-
354 tion in R [20, 9]. To account for the nature and the dynamic of the signal on the
355 observed X-ray fluorescence data, the weight was set to the reciprocal square root
356 of the observation when the observation is not 0, and to 1 otherwise. The second
357 important parameter was the *span* of the filter that we set to 0.02 in order to account
358 for the approximate width of the fluorescence bands on such spectra. Finally, a
359 thresholding was performed on the regularized form so that its value is never lower
360 than 0.001.

361 While this procedure is producing a realistic *zero noise* X-ray fluorescence spectra
362 in each pixel of our image, it has to be noted that this should not be considered as
363 a *ground truth* version of the observation. Indeed, since each spectrum is dealt
364 with independently from its neighbors, there is no spatial regularization and the
365 *estimations* performed are far from optimal for detector channels that have measured
366 a low level of photons.

367 The noise present in the observation is mostly due to the counting statistic of
368 each channel of the detector. Hence, we can generate a simulated observed spectra
369 with the same SNR as the raw observation, by simply replacing the value of the *zero*

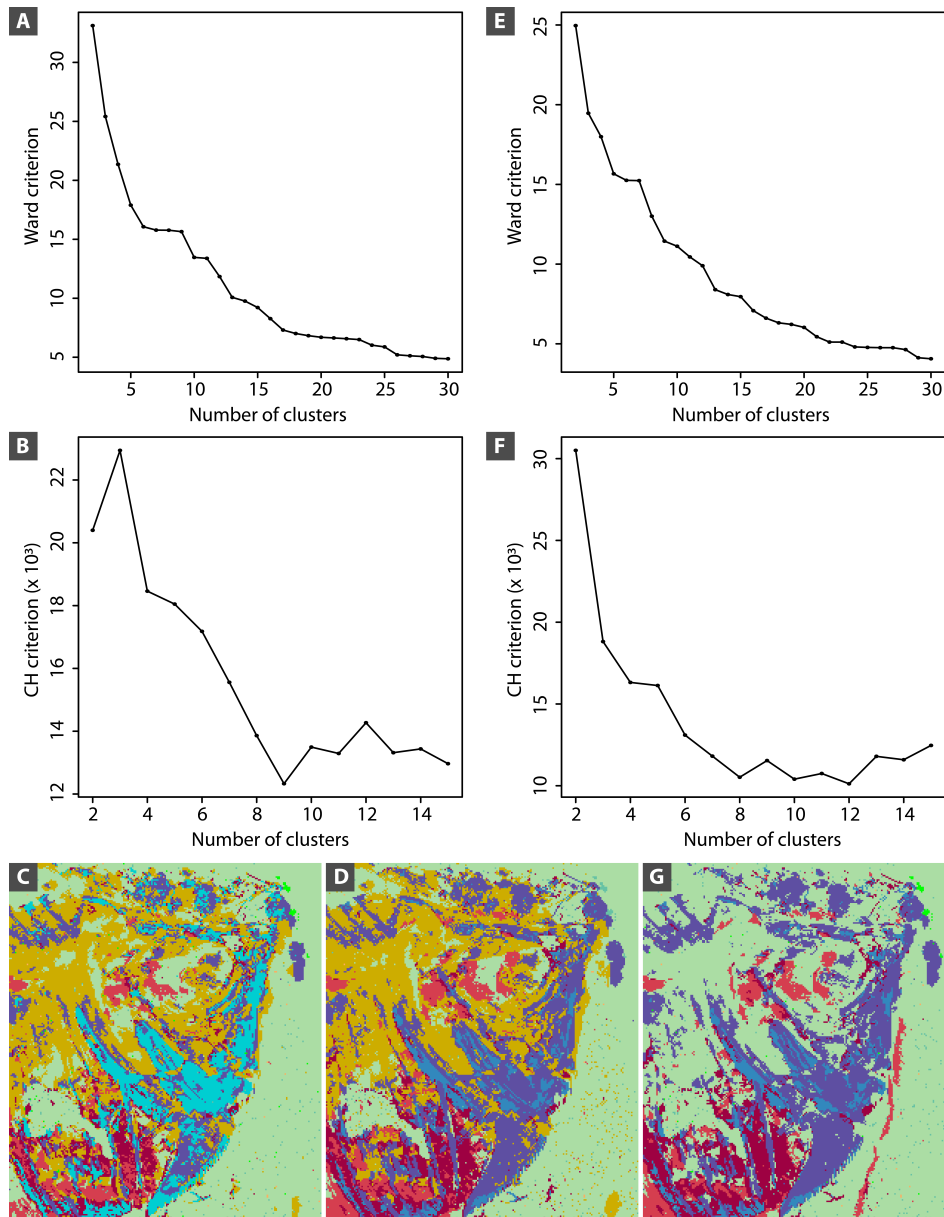


Figure 5. Hierarchical segmentation for different choices of switching step $J = 43000$ and $J = 46000$. (A, B): Ward (A) and Calinski and Harabasz (B) criteria against the number of clusters (starting with 2 clusters) for $J = 43000$. (C): False color distributions obtained for $J = 43000$ (10 clusters). (D): False color distributions obtained for $J = 44268$ (9 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).

370 *noise spectra* by a single realization of a Poisson random process with its parameter
371 being the *zero noise spectra*'s value. We generated such a dataset, for which we have,
372 by construction, the *ground truth* and a SNR equal to the one of the raw dataset. This
373 simulated dataset is later on referred as a *plus 0db dataset* (p0db in short).

374 Starting from the same *zero noise model*, we also generated simulated observation
375 with lower SNR. Since each *theoretical value* is replaced by a Poisson realization,
376 dividing the model by a factor of 2 would decrease the SNR by a factor of $\sqrt{2}$, which
377 correspond to removing 3db to the SNR. This simulated dataset is later on referred
378 as a *minus 3db dataset* (m3db in short). Repeating this procedure two more times
379 enabled us to generate a *minus 6db dataset* (m6db) and finally a *minus 9db dataset*
380 (m9db).

381 Each of these datasets resembles what could have been measured if the expo-
382 sure time was divided by two incrementally. In other words, obtaining for the m3db
383 dataset a spatial clustering similar to that obtained for the p0db dataset would lead
384 to the conclusion that the experiment could have been done twice faster without sig-
385 nificant loss in term of the explained morphology of the fossil. A shorter exposure
386 time also means a lower radiation dose for the sample and correspondingly lower
387 risk of alteration during and due to the measurements.

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

395 6.2. Impact of noise on hierarchical spatial clustering results

396 Following the same process as in [subsection 5.2](#), we use the criterion H to de-
397 termine the switching step \hat{J} . In [Figure 6](#), we can see that the curve of H has
398 the expected shape for datasets p0db ([Figure 6A](#)) and m3db ([Figure 6C](#)), while the
399 shape begins to change for dataset m6db ([Figure 6E](#)) and is significantly different in
400 m9db ([Figure 6G](#)) (the curve correspond to the top left quadrant sub-image but the
401 same behaviour is observed for the three other sub-images). The values obtained
402 for these dataset are: $\hat{J} = 44063$ for p0db, $\hat{J} = 45871$ for m3db, $\hat{J} = 48988$ for m6db.
403 Such choice is not possible for dataset m9db for the exact reason explained at the

404 end of [subsection 3.1](#), hence for each of the four sub-image we took the smallest \hat{J}_k
405 on the right (greater than 8000) such that $H(\hat{J}_k) \geq H(J_{\max}) - sd(H)$ (where here J_{\max}
406 and $sd(H)$ are values computed for the associated sub-image I_k). This led to choose
407 $\hat{J} = 50387$ for m9db. The higher the noise, the higher the \hat{J} , getting closer to the total
408 number of pixels N in the image.

409 According to the plot of the jump heuristic, to the CHC and to clusters appearing
410 to be interesting from a paleontological point of view, the selected number of clusters
411 is 11 for p0db ([Figure 6B](#)), 10 for m3db ([Figure 6D](#)) and 9 for m6db ([Figure 6F](#)).
412 Concerning the m9db dataset, no fossil morphology can be seen when the selected
413 number of clusters is 10 or lower, hence we have decided to represent the 11-cluster
414 segmentation for this dataset ([Figure 6H](#)).

415 As expected, similarity of the graphic representations as compared to the original
416 and simulated datasets quickly degrades with increasing noise, and most morpho-
417 logical information is lost for dataset with a SNR greater than or equal to 6db from
418 the original data. Increasing further the level of noise leads to totally unexploitable
419 data, with which the morphology of the sample could hardly be observed, as demon-
420 strated on the m9db simulation ([Figure 6H](#)). Note that, in the m3bd representation
421 ([Figure 6D](#)), the pale yellow triangular area that “appears” on the top right of the im-
422 age and clusters with some of the fossil corresponds to air (there is no sample there,
423 see [Figure 1A](#)); it otherwise clusters with the sediment in the other noise models,
424 which can be explained by the geometry used during the experiment where the beam
425 came from the right of the sample with a 45° angle, leading for pixels in that area to
426 record the X-rays-sediment interaction below the fossil surface.

427 7. Discussion

428 In this article we propose a spatially constrained hierarchical clustering method
429 to be applied on spectral images, in particular on energy resolved X-ray fluorescence
430 images. The first aspect of the method is to choose an agglomerative criteria based
431 on a dissimilarity measure that is consistent with the noise model of the measured
432 spectra. Then, the main aspect of this method is to apply constraints during the
433 agglomerative process such that only spectra belonging to neighboring pixels could
434 be clustered together. While this constraint is meaningful as long as the classes
435 form small clusters on the image, it is obvious that when the number of classes is
436 small this spatial constraint should not be applied anymore, bringing the problem of
437 the *proper step* at which the spatial constraint should be released. To address this

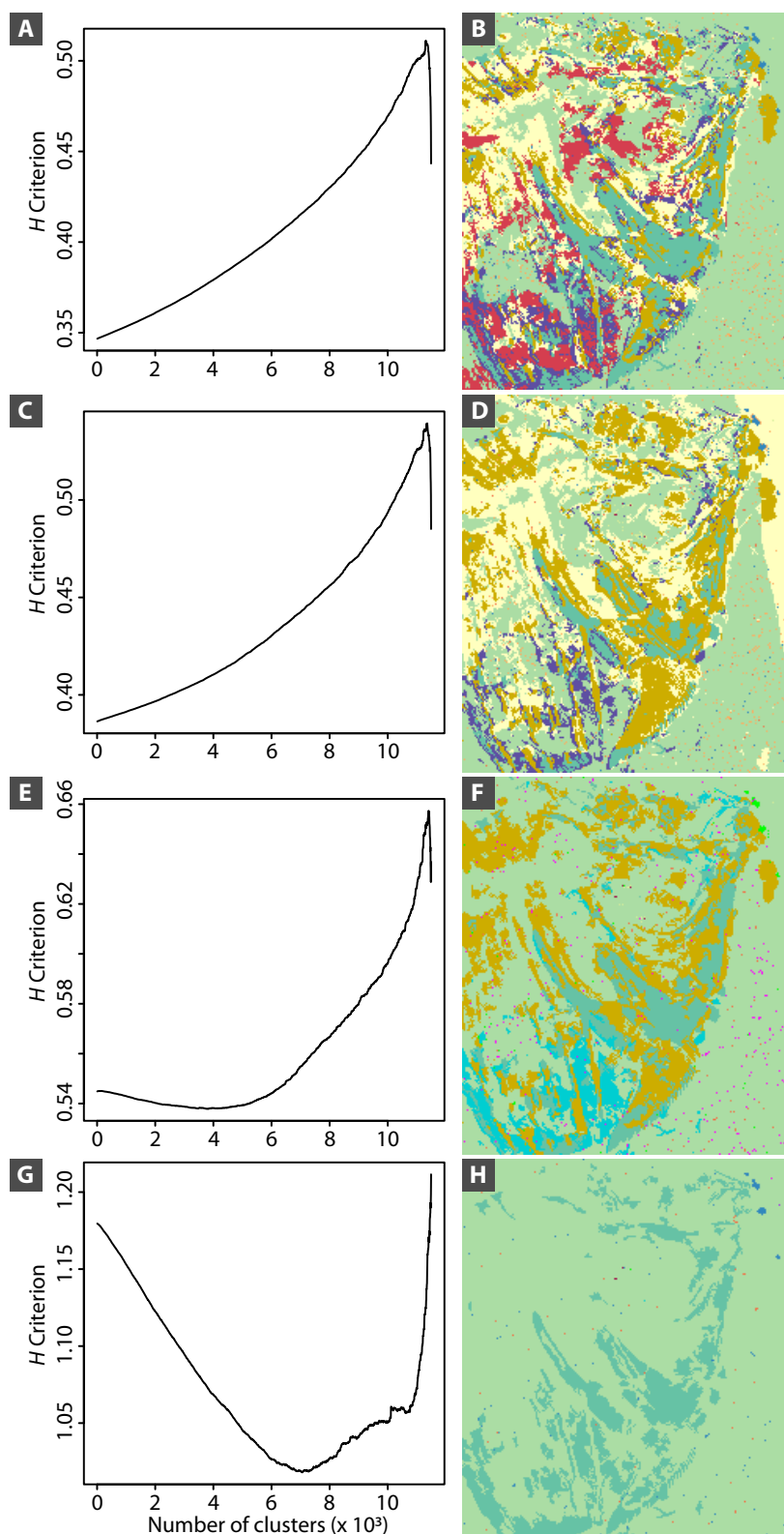


Figure 6. Behavior of the H criterion for selecting the switching step J while adding noise to the zero noise model simulated dataset. The H criterion computed in the top left quadrant sub-image (as explained in [subsection 3.2](#) and [subsection 5.2](#)) and resulting hierarchical spatial clustering image for the zero noise model (adding 0db) (A,B), and after removing 3db (C,D), 6db (E,F) and 9db (G,H).

438 problem, we proposed a heuristic that balances the spatial coherence of the proposed
439 segmentation, as measured through the G penalization, and the *between-cluster*
440 *inertia* deriving from the Ward agglomerative criteria. The outcome of this algorithm
441 is a hierarchy of possible segmentations that the practitioner should choose from.
442 To aid this final selection step, the Ward and Calinski and Harabasz criteria are both
443 computed to determine the most significant segmentation within the full hierarchy.

444 The advantages of such a *simple minded* algorithm is two-fold: first the general
445 principles of the method do not require deep knowledge of statistical methods and as
446 such can be grasped by the application domain specialist, the paleontologist in the
447 presented example. Second, the computational cost of the segmentation is relatively
448 low, even for a rather large dataset, and the processing time is on par with the typical
449 measurement time for such spectral images. Hence, this method can be applied to
450 the data while the experiment is still ongoing and used for a rapid diagnostic and
451 experimental feedback within the global data acquisition strategy.

452 As a diagnostic tool, this method helps at finding a balance between a higher
453 signal to noise ratio of individual spectra and the measurement time and radiation
454 dose to which the sample is subjected. In such μ XRF imaging modality, the SNR is
455 inversely proportional to the square root of the radiation dose. Increasing the SNR
456 increases the risk of producing radiation-induced damages to the sample during
457 the experiment, but also often leads to increased measurement time and fewer (or
458 smaller) samples being characterized in the allocated time slot. In such a situation
459 it is therefore important to quickly and properly assess the optimal exposure param-
460 eters (mostly time, but possibly also beam intensity), which need to be sufficient to
461 produce exploitable spectra while avoiding any risk of radiation-induced damages to
462 the sample and enabling large maps to be collected. Using simulated data, we have
463 shown that the algorithm is robust to an increased level of measurement noise and
464 as such is not only helpful in asserting an optimal measurement time but also in
465 reducing it and lowering the radiation dose.

466 In our SNR test application, it seems indeed that the behavior of H is a good early
467 indicator of the quality of the observed data, providing insight into the discrimination
468 power of the collected spectra. Indeed the curve in [Figure 6G](#) illustrates a behavior
469 significantly different from the ones of [Figure 6A,C,E](#), which we link to the fact that
470 the segmentation obtained in [Figure 6H](#) is not very informative. In other words,
471 the behavior of the H criterion as classes get aggregated is a good predictor of the
472 usefulness of the segmentation that will be attained with the data.

473 Note that while the simulated data tested herein demonstrate that the behavior
474 of the H criterion depends on the SNR, it seems equally likely that this behavior also
475 depends on the type of morphology of the sample being imaged. This is somehow
476 evidenced in our real data test when comparing the H criterion found in the four
477 quadrants of the image, three of which have a very similar morphology and H crite-
478 rion curve, while the fourth bottom-right quadrant, with mostly sediment and very
479 little fossil features, produces a slightly different H criterion curve.

480 As a continuation of the present work, one could assess how the H criterion
481 depends on the morphology of the image. From our currently limited experience, it
482 seems likely that if the studied dataset exhibits a similar type of morphology in all the
483 image, a possibility is to choose a sub-image of size N_0 representative of the image
484 morphology. The parameter β can be then evaluated by $\beta_0 = \hat{J}_0/N_0$ (and \hat{J} is chosen
485 as the nearest integer to $\beta_0 \times N$), leading to a drastic reduction of the computational
486 cost of the evaluation of this parameter. Furthermore, this would promote β to be a
487 scalar descriptor of the image's morphology.

488 Last but not least, this method provides to the practitioner a *complete* view of
489 the information contained in a given spectral image dataset. When such data are
490 collected, the *prior knowledge* on the chemistry of the sample often leads to the
491 selection of very specific features of the spectra to be analyzed. Moreover, although
492 entire μ XRF spectra mostly contain XRF elemental information they also include ad-
493 ditional, non-elemental signal including escape and sum peaks, as well as inelastic
494 and elastic scattering and peaks from elements present in the air between the sam-
495 ple and the detector such as Ar (Figure 1C). Depending on the sample, some of these
496 peaks can carry interesting signal and one could need to keep them in the analysis.
497 However, it is often preferred to remove them from the analysis and crop the spectra
498 to the "true" elemental signal only, or only a few peaks, prior to the analysis. This
499 can simply be done at the practitioner's discretion prior to applying the algorithm.

500 Conversely, we here propose to confront the result of such *focused* analysis with
501 an analysis based on the full spectra. Indeed, both the *focused* and *complete* analy-
502 sis could be performed using the same algorithm but selecting for each one either a
503 subset or the fullset of the spectral channels of the image. Using such an approach
504 the application scientist could both use the data in a *prior knowledge* directed ap-
505 proach, verifying pre-existing hypothesis on the nature of the signal to be detected
506 in the spectral image, as well as a *unsupervised discovery* approach where the full
507 spectral dataset is subjected to the segmentation with *a priori* on which channel is

508 of importance to exploit the image. Finally, this algorithm might even be used as
509 a *post-hoc* analysis to test *a posteriori* the importance of unexpected features of the
510 spectra as *discovered* discriminant features of the sample, as exemplified herein with
511 the iron-rich phase located around the fossil fish orbit for which the cluster mean
512 spectrum provided complementary and new information to decipher its chemistry.

513 Contribution of authors.

514 This work arose from discussions between GC, SXC and AG. SXC proposed the
515 exploitation of spatial constraints and the use of χ^2 as an adapted dissimilarity mea-
516 sure for XRF spectra. GC proposed the heuristic rule to stop applying spatial con-
517 straint on the segmentation. AG proposed a version of the χ^2 metric consistent
518 between the spatially constrained initial steps and the unconstrained agglomerative
519 steps, so that Lance and William formulae could be used in this latter part. SXC
520 and AG implemented the algorithm and its result representations in R. SXC pro-
521 posed and implemented the *zero noise model*. PG performed all the experimental
522 measurements and interpretations on the fossil, and oriented the algorithm design
523 to ensure results are valuable for the practitioner. All authors contributed to the
524 writing of this manuscript.

525

BIBLIOGRAPHY

- 526 [1] M. ALFELD AND K. JANSSENS, *Strategies for processing mega-pixel x-ray fluorescence hyperspec-*
527 *tral data: a case study on a version of caravaggio's painting supper at emmaus*, Journal of
528 analytical atomic spectrometry, 30 (2015), pp. 777–789.
- 529 [2] M. AMBROISE AND G. GOVAERT, *Convergence of an EM-type algorithm for spatial clustering*, Pattern
530 recognition letters, 19 (1998), pp. 919–327.
- 531 [3] A. BERGAMASCHI, K. MEDJOUBI, C. MESSAOUDI, S. MARCO, AND A. SOMOGYI, *Mmx-i: data-processing*
532 *software for multimodal x-ray imaging and tomography*, Journal of synchrotron radiation, 23
533 (2016), pp. 783–794.
- 534 [4] L. BERTRAND, M. COTTE, M. STAMPANONI, M. THOURY, F. MARONE, AND S. SCHÖDER, *Development*
535 *and trends in synchrotron studies of ancient and historical materials*, Physics Reports, 519
536 (2012), pp. 51–96, <https://doi.org/10.1016/j.physrep.2012.03.003>.
- 537 [5] L. BERTRAND, L. ROBINET, M. THOURY, K. JANSSENS, S. X. COHEN, AND S. SCHÖDER, *Cultural her-*
538 *itage and archaeology materials studied by synchrotron spectroscopy and imaging*, Applied
539 physics. A, Materials science & processing, 106 (2012), pp. 377–396, [https://doi.org/10.](https://doi.org/10.1007/s00339-011-6686-4)
540 [1007/s00339-011-6686-4](https://doi.org/10.1007/s00339-011-6686-4).
- 541 [6] L. BERTRAND, M. THOURY, AND E. ANHEIM, *Ancient materials specificities for their synchrotron ex-*
542 *amination and insights into their epistemological implications*, Journal of Cultural Heritage,
543 14(4) (2013), pp. 277–289.

- 544 [7] L. BREIMAN, J. FRIEDMAN, C. J. STONE, AND R. A. OLSHEN, *Classification and Regression Trees*,
545 Taylor & Francis, 1984.
- 546 [8] T. CALINSKI AND A. HARABASZ, *A dendrite method for cluster analysis*, *Communications in Statis-*
547 *tics*, 3 (1974), pp. 1–27.
- 548 [9] W. CLEVELAND, E. GROSSE, AND W. M. SHYU, *Statistical Models in S*, Wadsworth & Brooks/Cole,
549 New York, 1992, ch. Chapter 8 : Local Regression Models.
- 550 [10] D. DAVESNE, P. GUERIAU, D. DUTHEIL, AND L. BERTRAND, *Exceptional preservation of a cretaceous*
551 *intestine provides a glimpse of the early ecological diversity of spiny-rayed fishes (acanthomor-*
552 *pha, teleostei)*, *Scientific Reports*, 8 (2018), p. 8509.
- 553 [11] B. S. EVERITT, S. LANDAU, M. LEESE, AND D. STAHL, *Cluster Analysis, 5th edition*, Wiley, 2010.
- 554 [12] P. GUERIAU, S. BERNARD, F. FARGES, C. MOCUTA, D. B. DUTHEIL, T. ADATTE, B. BOMOU, M. GODET,
555 D. THIAUDIÈRE, S. CHARBONNIER, ET AL., *Oxidative conditions can lead to exceptional preserva-*
556 *tion through phosphatization*, *Geology*, (2020).
- 557 [13] P. GUERIAU, C. JAUVION, AND M. MOCUTA, *Show me your yttrium, and i will tell you who you are:*
558 *implications for fossil imaging*, *Palaeontology*, 61(6) (2018), pp. 981–990.
- 559 [14] P. GUERIAU, C. MOCUTA, AND L. BERTRAND, *Cerium anomaly at microscale in fossils*, *Analytical*
560 *Chemistry*, 87(17) (2015), pp. 8827–88367.
- 561 [15] P. GUERIAU, D. MOCUTA, C. AND DUTHEIL, S. COHEN, D. THIAUDIÈRE, THE OT1 CONSORTIUM, S. CHAR-
562 BONNIER, G. CLÉMENT, AND L. BERTRAND, *Trace elemental imaging of rare earth elements dis-*
563 *criminates tissues at microscale in flat fossils*, *PLoS One*, 9(1) (2014), p. e86946.
- 564 [16] P. GUERIAU, S. RÉGUER, N. LECLERCQ, C. CUPELLO, P. BRITO, C. JAUVION, S. MOREL, S. CHARBONNIER,
565 D. THIAUDIÈRE, AND C. MOCUTA, *Visualizing mineralization processes and fossil anatomy using*
566 *synchronous synchrotron X-ray fluorescence and X-ray diffraction mapping*, *Journal of the*
567 *Royal Society Interface*, 17 (2020), p. 20200216, <https://doi.org/10.1098/rsif.2020.0216>.
- 568 [17] G. N. LANCE AND W. T. WILLIAMS, *A general theory of classificatory sorting strategies: II. Clustering*
569 *systems*, *The Computer Journal*, 10 (1967), pp. 271–277, [https://doi.org/10.1093/comjnl/](https://doi.org/10.1093/comjnl/10.3.271)
570 [10.3.271](https://doi.org/10.3.271).
- 571 [18] L. LEBART, *Programme d'agrégation avec contrainte*, *Cahiers de L'analyse des Données*, 3 (1978),
572 pp. 275–287.
- 573 [19] G. MILLIGAN AND M. COOPER, *An examination of procedures for determining the number of clusters*
574 *in a data set*, *Psychometrika*, 50 (1985), pp. 159–179.
- 575 [20] R CORE TEAM, *R: A Language and Environment for Statistical Computing*, R Foundation for Statis-
576 tical Computing, Vienna, Austria, 2020, <https://www.R-project.org/>.
- 577 [21] V. A. SOLÉ, E. PAPILLON, M. COTTE, P. WALTER, AND J. SUSINI, *A multiplatform code for the analysis*
578 *of energy-dispersive x-ray fluorescence spectra*, *Spectrochimica Acta B*, 62 (2007), pp. 63–68.
- 579 [22] J. H. WARD, *Hierarchical grouping to optimize an objective function*, *Journal of the American Sta-*
580 *tistical Association*, 58 (1963), pp. 236–244.