



# Combined XRD-XRF cluster analysis for automatic chemical and crystallographic surface mappings

M. Bortolotti <sup>a)</sup>, L. Lutterotti, E. Borovin, and D. Martorelli 

Department of Industrial Engineering, University of Trento, IT, Italy

(Received 31 October 2018; accepted 20 February 2019)

X-ray diffraction-X-ray fluorescence (XRD-XRF) data sets obtained from surface scans of synthetic samples have been analysed by means of different data clustering algorithms, with the aim to propose a methodology for automatic crystallographic and chemical classification of surfaces. Three data clustering strategies have been evaluated, namely hierarchical, *k*-means, and density-based clustering; all of them have been applied to the distance matrix calculated from the single XRD and XRF data sets as well as the combined distance matrix. Classification performance is reported for each strategy both in numerical form as the corrected Rand index and as a visual reconstruction of the surface maps. Hierarchical and *k*-means clustering offered comparable results, depending on both sample complexity and data quality. When applied to XRF data collected on a two-phases test sample, both algorithms allowed to obtain Rand index values above 0.8, whereas XRD data collected on the same sample gave values around 0.5; application to the combined distance matrix improved the correlation to about 0.9. In the case of a more complex multi-phase sample, it has also been found that classification performance strongly depends on both data quality and signal contrast between different regions; again, the adoption of the combined dissimilarity matrix offered improved classification performance. © 2019 International Centre for Diffraction Data. [doi:10.1017/S0885715619000216]

Key words: combined XRD-XRF, cluster analysis, surface mappings

## I. INTRODUCTION

When dealing with the crystallographic and chemical characterization of inhomogeneous sample surfaces, it is often desirable to perform a preliminary qualitative screening to locate regions of particular interest. In the case of large areas and/or high-resolution mappings, acquisition of several thousands of patterns can be often required, and a complete Rietveld or Fundamental Parameters analysis of each data point becomes impractical. While in the case of X-ray fluorescence (XRF) data, a qualitative or semi-quantitative screening may be automatically performed in a reasonable amount of time, X-ray diffraction (XRD) data interpretation is in general much more challenging, requiring almost always support from experienced analysts. In the latter case, indeed, traditional qualitative approaches based on automated peak position extraction and comparison (a.k.a. *search match*), although comparatively fast and convenient, can pose serious problems in terms of accuracy, especially when dealing with complex samples and related effects such as multi-featured microstructure and the presence of preferred orientation. Additionally, the data collected on large areas scans typically presents quality issues such as low signal to noise ratio, high background, effects because of small sampling volume, etc. In such circumstances, the adoption of non-parametric statistics, in which no a-priori assumption is made about the underlying model, can offer an alternative, robust approach for data classification, allowing to sort the patterns in related classes, some of

which can eventually be selected for further quantitative analysis (Barr *et al.*, 2004), (Gilmore *et al.*, 2004).

In the previous paper (Bortolotti *et al.*, 2017), we have shown how XRD and XRF can be combined inside a Rietveld/Fundamental Parameters modeling framework to improve the accuracy in both chemical and crystallographic quantitative analysis of homogeneous samples; in this work, we propose to combine XRD and XRF surface mappings for automatic classification based on cluster analysis, with the goal to evaluate if the classification performance and error tolerance of the single techniques can be improved. To achieve this goal, significant statistical features obtained from the standalone XRD and XRF data sets (in particular, the data set distance matrix) are combined together and then used as a single source for cluster analysis to classify the data; the obtained cluster classification can be used to reconstruct a sample surface map highlighting chemically and crystallographically related regions.

## II. EXPERIMENTAL

To perform the XRD/XRF mappings, an INEL Equinox 3500 combined instrument equipped with a single X-ray source with two detectors for XRD and XRF operating simultaneously was adopted (Lutterotti *et al.*, 2016) (Bortolotti *et al.*, 2017) (Figure 1).

The machine was equipped with a 50W microfocus source with Mo target (Figure 1(a)) coupled with a multilayer elliptic mirror (AXO Dresden GmbH) to suppress *k*- $\beta$  radiation and provide a quasi-parallel beam; the single source setup ensures that both XRD and XRF signals are recorded

<sup>a)</sup> Author to whom correspondence should be addressed. Electronic mail: [mauro.bortolotti@unitn.it](mailto:mauro.bortolotti@unitn.it)

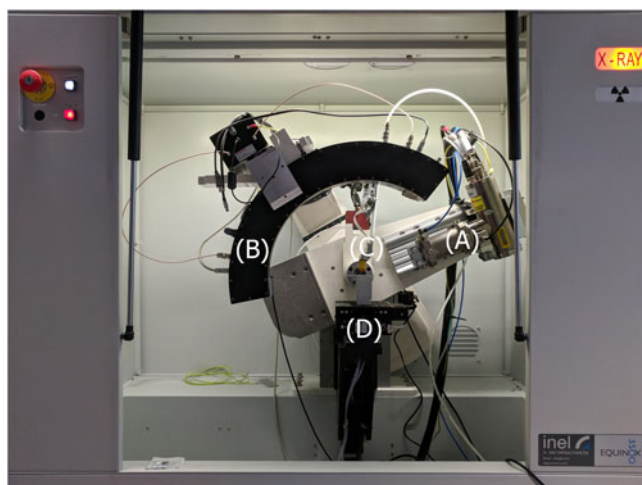


Figure 1. The combined XRD/XRF instrument for automated 2D data collection. (a) Microfocus Mo source (b) CPS120 XRD Curved Position Detector (c) Si Pin XRF detector (d) XYZ motorized sample stage.

from exactly the same point on the surface. Beam size on the sample was reduced by means of manual cross-slits to about  $1 \times 0.5 \text{ mm}^2$ . A curved position sensitive detector (Figure 1 (b)) was used to collect diffraction data (INEL CPS120); the detector operated over  $120^\circ$  in  $2\theta$  with a 4096 channel resolution. Fluorescence spectra were collected by means of a Si-PIN diode (Figure 1(c)) coupled to a 4096 channel analyzer (Moxtek X-PIN50). Finally, a 3-axis motorized sample holder (Oriental Motors) was used to automate the  $x$ - $y$  mappings with a scan step resolution of 0.5 mm (Figure 1(d)).

Both XRD and XRF detector operated in a fixed position to ensure the fastest data collection possible; acquisition times of 60 s were chosen as a compromise between data collection speed and signal quality.

For the validation of the different clustering strategies and their relative performances, synthetic samples were prepared with ad-hoc designed chemical and crystallographic surface features. The first sample (Figure 2(a)) was prepared by embedding a steel alloy cube (6 mm edge) in a metallographic resin, with the goal to provide the simplest classification problem possible (two classes). The second sample (Figure 2(b)) was designed to offer a more challenging benchmark for assessing classification performance: five different metallic rods based on different alloy compositions (Brass, Copper, Aluminium, Titanium) with diameters ranging from 3 to 7 mm were embedded in an amorphous metallographic resin. Both samples were cut and polished to obtain an ideally flat surface for the data collection.

### III. ANALYTICAL METHODOLOGY

Data clustering is one of the many available tools among data mining methodologies, whose main goal consists in finding subgroups (clusters) of similar observations inside a data set. Data clustering can be implemented by a large variety of algorithms (Kaufman *et al.*, 2005), (Tan *et al.*, 2005); for the aim of this work, three different clustering algorithms have been selected among the most widely adopted in various application fields, namely *hierarchical clustering* (HC), *k-means clustering* (KM), and *density-based clustering* (DB).

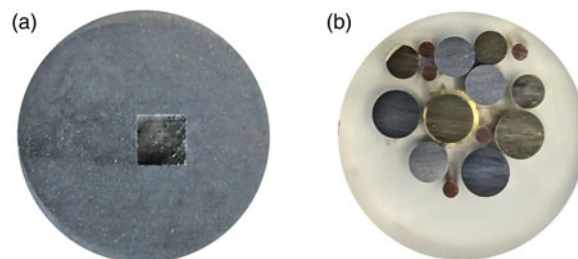


Figure 2. Synthetic samples providing a test suite for evaluating the performance of the different data clustering strategies. (a) Steel alloy cube embedded in resin; (b) metallic rods embedded in amorphous resin.

The classification process is not an automatic task: clustering parameters (dissimilarity metric, number of expected clusters, minimum observations number to define a cluster, etc.) depend on both the algorithm chosen and, of course, on the data domain under consideration; indeed, it is often the case, that the best clustering algorithm as well as its operating parameters need to be chosen experimentally for a specific problem solving. In particular, the choice of the dissimilarity criterion is often critical as it defines how two observations in a data set can be considered distant from each other in the data set space. Several distance metrics have been defined for different application domains; some commonly used metrics in cluster analysis are Euclidean distance, squared Euclidean distance and Manhattan distance, as defined below:

$$d(i, j)_{\text{euclidean}} = \sqrt{\sum_m (i_m - j_m)^2} \quad (1)$$

$$d(i, j)_{\text{euclidean}^2} = \sum_m (i_m - j_m)^2 \quad (2)$$

$$d(i, j)_{\text{manhattan}} = \sum_m |i_m - j_m| \quad (3)$$

In our case, the standard Euclidean distance was chosen as the default metric for all the clustering algorithms, as it turned out to be the best performing in all cases.

In the following, we briefly describe the basic operation of the different clustering algorithms.

HC (Lance and Williams, 1967) classifies objects by performing a subdivision of the data set into a hierarchy of different subgroups based on their relative distances; classification can proceed in an agglomerative fashion (all observations start as distinct clusters, which are then merged based on their similarity) or a divisive one (all the observations starts inside a single cluster, which is then subdivided in successive iterations). In addition to the distance between each observation, another important decision needed by HC is the linkage criterion between clusters, that is the criterion by which two clusters  $C_i$  and  $C_j$  are classified based on the pairwise distances between the  $n_i$ ,  $n_j$  observations they contain. Some commonly used strategies are *single linkage*, based on the minimum distance found between observations in the two clusters (eq. (4)), *complete linkage*, based on the maximum distance (eq. (5)), and *average linkage*, defined as the average distance between each point in one cluster to every point in the other cluster (eq. (6)); this latter criterion is also the one adopted in the current

work.

$$\max \{d(i, j) : i \in C_i, j \in C_j\} \quad (4)$$

$$\min \{d(i, j) : i \in C_i, j \in C_j\} \quad (5)$$

$$\frac{1}{|C_i||C_j|} \sum_{i \in C_i} \sum_{j \in C_j} d(i, j) \quad (6)$$

HC does not provide a unique partitioning of the data set; the principal output of the algorithm is a *dendrogram*, a hierarchical plot, where the y-axis represents the distance at which different clusters merge; the user still needs to decide at which distance to cut the tree, obtaining a partitioning at a selected precision and a corresponding number of clusters.

The standard algorithm has a time complexity of  $O(n^3)$  and is thus not suitable for highly complex problems; for the scope of the current work, this is a negligible shortcoming but could become problematic for large data sets.

KM clustering (Macqueen, 1967) is a kind of centroid-based clustering algorithm in which cluster centers are represented by a vector in the data set  $p$ -space; center points can belong to the cluster itself or not. The original algorithm was originally designed for signal processing (Lloyd, 1982) and aims to partition an  $n$ -point data set into  $k$  cluster centers (with  $k < n$ ), grouping the  $n$  observations in the different clusters in such a way to minimize the *within-cluster* sum of squared distances:

$$\min_C \sum_{i=1}^k \sum_{x \in C_i} |x - \mu_i|^2 \quad (7)$$

where  $\mu_i$  is the center of cluster  $C_i$  (e.g. the mean of points contained in cluster  $C_i$ ) and both  $x$  and  $\mu_i$  are  $p$ -dimensional vectors. The general solution to the problem is NP-hard; when the number of clusters  $k$  and the dimensionality of the space  $p$  are fixed in advance, the time complexity can be reduced to  $O(n^{pk+1})$ , still computationally significant for all but the most trivial applications. Thus, only approximate solutions (*local optima*) can be found, by using for example the Lloyd's heuristic algorithm (Lloyd, 1982), also adopted in this work; in cases like this, it is common practice to run the iteration several times and pick the best solution found among the different runs.

The main disadvantage of the algorithm is that it requires the knowledge of the number of clusters in advance; also, KM clustering tends to prefer clusters of the same size (since it is based on the distance from the centroid), often producing incorrectly cut cluster borders.

A final clustering strategy examined was DB clustering. DB clustering is based on the distinction between *high-density* areas in the data set, which can be classified as actual clusters, and *low-density* areas which are treated as noise or cluster borders. The most famous algorithm for performing DB clustering is DBSCAN (Ester *et al.*, 1996), which works by grouping points within certain distance thresholds. The main advantage of this algorithm is that, differently from HC and KM, it can find arbitrarily shaped clusters in the  $p$ -space; additionally, the number of clusters is not required as an input, and it is quasi-deterministic (core points classification is deterministic whereas border points classification can sometimes be

non-deterministic). DBSCAN has a fast average converging time of  $O(n \log n)$ . As the negative features, this algorithm assumes clusters of similar densities and requires an additional range parameter as input. An extension of the method, OPTICS, which does not suffer from these issues was recently proposed (Ankerst *et al.*, 1999) and has been adopted in this work as a reference implementation for DB methods.

Relative performances of different clustering methods on the same data set can be evaluated with several numerical criteria (Theodoridis and Koutroumbas, 2009), which can be broadly distinguished into two main strategies: internal and external evaluation. Internal evaluation is based on the same data used for clustering; in general, the quality scores are biased towards high intra-cluster and low inter-clusters similarity. As such, internal evaluation criteria can suggest how different algorithms perform but not necessarily the quality of the clustering itself. One of the most used internal validation criteria is the Dunn index (Dunn, 1974):

$$\text{DUNN} = \frac{\min_{1 \leq i < j \leq n} d(i, j)}{\max_{1 \leq k \leq n} d'(k)} \quad (8)$$

where  $d(i, j)$  represents the value of the adopted distance metric calculated between clusters  $i$  and  $j$ , whereas  $d'(k)$  measures the intra-cluster distance of cluster  $k$ .

In the external evaluation, clustering performances are compared against some reference benchmark of pre-classified data. A large selection of numerical criteria can be found in the literature; in this work we have chosen the widely adopted Rand index (Rand, 1971), which is a measure of the similarity between the trial clustering classification and the benchmark:

$$\text{RAND} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (9)$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. It is also possible to define a *chance-corrected* adjusted Rand index (Rand, 1971), which takes into account the weight difference between false positives and false negatives. In both cases, the Rand index may take values from  $-1$  (meaning the classifications are completely uncorrelated) to  $1$  (perfect correlation).

In the present work, the clustering evaluation procedure begins starting from the calculation of the Euclidean distance matrix of the XRD and XRF data sets. Each matrix is  $n \times n$  dimension, where  $n$  is the total number of experimental data points, which are typically (but not necessarily) collected over a square  $k \times k$  points grid. Each matrix element  $d_{ij}$  corresponds to the Euclidean distance between  $i$  and  $j$  points; the matrix is thus symmetric and the diagonal elements are equal to one. Each observation is associated with  $p$  features, which in our case are represented by distinct measurements of integrated X-ray counts collected over a discrete interval (Bragg angle or Energy). Additionally, the combined distance matrix is calculated, with each element  $d_{\text{comb},ij}$  equal to the sum of the XRD and XRF Euclidean distances:

$$d_{\text{comb},ij} = \sqrt{d_{\text{xd},ij}^2 + d_{\text{xf},ij}^2} \quad (10)$$

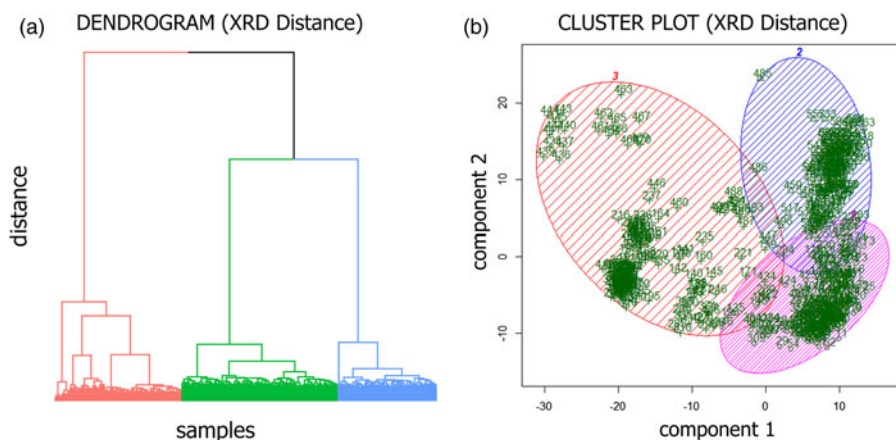


Figure 3. (a) Dendrogram plot obtained from the hierarchical clustering classification on the XRD data set; (b) cluster plot on the same data after the dendrogram cut corresponding to three clusters.

Clustering algorithms are then executed on the different distance matrices, and their results numerically compared by the means of adjusted Rand index; finally, a contour map of the sample surface is reconstructed based on the cluster assignment of each measurement point.

All calculations of this work have been performed using the package R (R Development Core Team, 2016), with the aim of the additional packages “Cluster” (Maechler *et al.*, 2015), “Factoextra” (Kassambara and Mundt, 2017) and “DBSCAN” (Hahsler and Piekenbrock, 2017) for cluster analysis; a source code example for testing the methodologies described is provided in the online supplementary material.

#### IV. RESULTS AND DISCUSSION

We report here qualitative and quantitative results obtained from cluster classification applied to the samples described in the experimental section.

Sample 1 was used as a simple test case to perform a preliminary evaluation of the relative strengths and weaknesses of different clustering strategies. Given the sample basic geometry, a coarser mapping was performed using only a  $21 \times 21$  point matrix over a  $12 \times 12 \text{ mm}^2$  area; Figure 4(a) reports the ideal cluster classification as obtained from the optical image color discretization of the area under consideration. As reported in Table I, all of the clustering algorithms offered good classification performance on the XRF data set, with values of the Corrected Rand index above 0.8; on the other hand, both HC and KM provided poor performance on the XRD data set, with DB clustering completely failing. When using the combined distance matrix to perform the classification, both HC and KM algorithms were able to improve the performance with respect to the single XRD and XRF data set; DB was able to correctly perform the classification, albeit with a lower Rand value with respect to the XRF data alone. As an example, we also report here qualitative results obtained from HC clustering on the XRD data set. In Figure 3(a), the dendrogram plot of the Euclidean distance matrix calculated on the XRD data is represented; the partitioning shows tree principal clusters, corresponding to the steel sample, the matrix and a “noise” cluster of unidentified points. Figure 3(b) reports the corresponding cluster plot representing the grouping of each

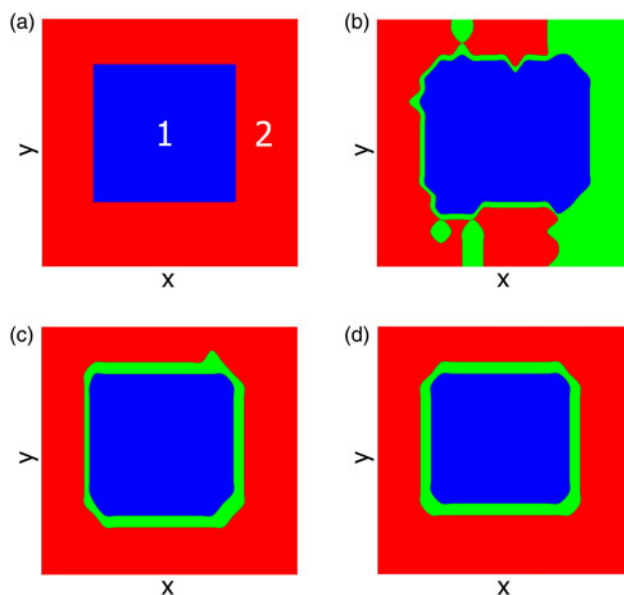


Figure 4. Sample 1 area reconstruction based on the Hierarchical Clustering classification performed on the distance matrix calculated on the XRD data set (b), XRF data set (c) and combined data set (d). (a) represents the reference classification obtained from the color discretization of the optical image of the area under consideration.

data point in the data set, after the dendrogram cut performed at a number of clusters equal to 3.

Figure 4 reports the area reconstructions of the sample surface based on the different classifications. Figure 4(a) shows the ideal classification, as obtained from the optical image of the sample; Figures 4(b) and (c) show the reconstruction performed starting from the XRD and XRF data sets, respectively, evidencing the poor performance of the XRD data set as compared to the XRF one. Finally, the combined data set shows a slight improvement in the sample/matrix interface reconstruction with respect to the XRF case (Figure 4(d)).

The second sample was used as a test to provide more realistic quantitative performance results from the different classification algorithms. The same algorithmic strategy was chosen, testing three different clustering algorithms on the Euclidean distance matrix calculated from the single data

TABLE I. Corrected Rand index values obtained from the comparison between the data clustering runs performed on sample 1 and sample 2 and their relative reference classifications.

Corr. Rand	S1 – XRD	S1 – XRF	S1 – COMB	S2 – XRD	S2 – XRF	S2 – COMB
K-means	0.520	0.855	0.886	0.475	0.404	0.476
Hierarchical	0.474	0.864	0.898	0.407	0.290	0.474
Density-based	N/A	0.857	0.733	N/A	0.259	0.014

sets as well as the one obtained from the sum of the distances, for a total of nine classification runs. In this case, the corrected Rand index values are in general much lower (Table I), meaning a lower general agreement between the reference classification, and the one obtained from the clustering. For this test sample, classification performed on XRD data set alone performed better than on the XRF data set, with the exception of the DB strategy which failed. Classification performed on the combined distance matrix, on the other hand, gave slightly improved results compared to the XRD data set in the case of KM and HC; the DB algorithm was able to run correctly but gave a Rand Index of almost zero (no correlation).

Examination of the reconstructed two-dimensional (2D) maps (Figure 5) may explain different algorithm behaviors. Compared to the reference map (Figure 5(a)), classification performed on the XRD data set (Figure 5(b)) was able to perfectly detect cluster 1A (brass alloy n.1), but not to distinguish it from cluster 1B (brass alloy n.2) and cluster 2 (Copper); this is reasonable, considering the similarity of the corresponding diffraction patterns and the low-quality level of the collected XRD data. Cluster 3 (Aluminum alloy) was reconstructed only loosely, whereas cluster 4 (Titanium) was completely undetected, probably because of the low signal to noise ratio of the corresponding diffraction pattern.

Classification performed on the XRF data set (Figure 5(c)), on the other hand, offered a good reconstruction of the areas relative to the Ti and Cu alloys, whereas Al alloy was completely undetected (because of the very low intensity of Al fluorescence lines when measuring in air); again, the two Brass alloys were detected but not distinguished.

Finally, the classification performed on the combined distance matrix (Figure 5(d)) was able to retain complementary information from the two techniques; in particular, both the Ti and the Al areas were reconstructed and the Cu regions distinguished from the Brass alloys regions. Interestingly, the amorphous matrix was separated fairly well in all cases.

## V. CONCLUSIONS

In this work we have evaluated different data clustering strategies applied to XRD/XRF surface mapping data collected on flat surfaces of ad-hoc designed samples. All the clustering algorithms were applied to the Euclidean distance matrix calculated from the standalone data sets as well as the combined distance matrix. In general, the latter approach seemed to offer better performances both in terms of the corrected Rand Index (measuring the agreement between the experimental clustering and the reference classification) and the visual reconstruction of the sample surface in terms of chemically and crystallographically homogenous areas. Hierarchical and KM clustering strategies offered typically similar results; DB clustering, on the other hand, provided in general poor performances and failed a couple of times.

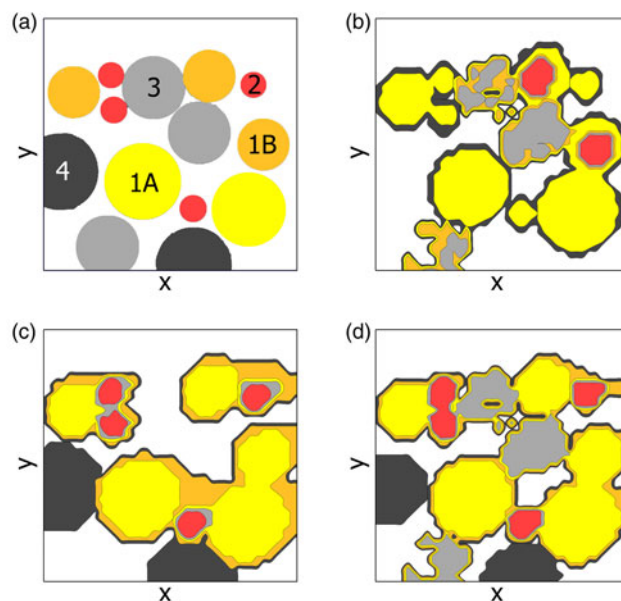


Figure 5. Sample 2 area reconstruction based on the Hierarchical Clustering classification performed on the distance matrix calculated on the XRD data set (b), XRF data set (c) and combined data set (d). (a) represents the reference classification obtained from the color discretization of the optical image of the area under consideration.

In conclusion, the methodology has proven promising for the goal of quick classification of flat sample surfaces, when a particular region of interest of the sample is not known in advance; in this scenario, the true chemical and crystallographic characterization is not important as this can be assessed later if necessary. The examples presented are limited to synthetic samples with an examined area of around 1 cm<sup>2</sup> and a scan resolution of 0.5 mm; when dealing with high brilliance sources setups (e.g. rotating anode sources or synchrotron facilities) it is possible to imagine that the methodology can be extended in terms of larger areas, higher resolution or/and better data quality. A further potential application could be represented by hyperspectral data obtained from recent full-field radiographic imaging techniques (Egan *et al.*, 2014), as well as their extension to 3D imaging. On the other hand, some of the algorithms presented here scale very poorly with the number of data points, so that data analysis could represent the real bottleneck with respect to data acquisition.

Finally, it has to be emphasized that the scope of this work was mainly about automatic classification; on the other hand, it is also possible to imagine applying a similar methodology to the detection of specific chemical and crystallographic patterns in a data set, e.g. when looking for specific area of interests in a sample surface; this could be the topic of further research.

## SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at <https://doi.org/10.1017/S0885715619000216>

## ACKNOWLEDGEMENT

The authors wish to acknowledge the support of the European Project H2020 Solsa, Grant Agreement n. 689868.

- Ankerst, M., Breunig, M. M., Kriegel, H. P., and Sander, J. (1999). In Proceedings ACM SIGMOD'99 International Conference on Management of Data. doi:10.1145/304182.304187.
- Barr, G., Dong, W., and Gilmore, C. J. (2004). "High-throughput powder diffraction. II. Applications of clustering methods and multivariate data analysis," J. Appl. Crystallogr. **37**, 243–252.
- Bortolotti, M., Lutterotti, L., and Peponi, G. (2017). "Combining XRD and XRF analysis in one Rietveld-like fitting," Powder Diffr. **32**(S1), 1–6.
- Dunn, J. C. (1974). J. Cybern., **4**, 95–104.
- Egan, C. K., Jacques, S. D. M., Connolley, T., Wilson, M. D., Veale, M. C., Seller, P., and Cernik, R. J. (2014). "Dark-field hyperspectral X-ray imaging," Proc. Royal Soc. A: Math. Phys. Eng. Sci. **470**, 20130629–20130629.
- Ester, M., Kriegel, H. P., Sander, J., and Xu, X. (1996). "A density-based algorithm for discovering clusters in a large spatial databases with noise," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining* (AAAI Press, Menlo Park, CA), pp. 226–231.
- Gilmore, C. J., Barr, G., and Paisley, J. (2004). "High-throughput powder diffraction. I. A new approach to qualitative and quantitative powder diffraction pattern analysis using full pattern profiles," J. Appl. Crystallogr. **37**, 231–242.
- Hahsler, M., and Piekenbrock, M. (2017). <https://CRAN.R-project.org/package=dbscan>.
- Kassambara, A., and Mundt, F. (2017). <http://www.sthda.com/english/rpkgs/factoextra>.
- Kaufman, L., Kaufman, L., Rousseeuw, P. J., and Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis* (John Wiley & Sons, Hoboken, NJ).
- Lance, G. N., and Williams, W. T. (1967). Comput. J., **9**(4), 373–380.
- Lloyd, S. P. (1982). "Least squares quantization in PCM," IEEE Trans. Inf. Theory. **28**, 129–137.
- Lutterotti, L., Dell'Amore, F., Angelucci, D. E., Carrer, F., and Gialanella, S. (2016). "Combined X-ray diffraction and fluorescence analysis in the cultural heritage field," Microchem. J. **126**, 423–430.
- Macqueen, J. (1967). "Some methods for classification and analysis of multivariate observations," Proc. Fifth Berkeley Symp. on Math. Statist. and Prob. **1**, 281–297.
- Maechler, M., Struyf, A., Hubert, M., Hornik, K., Studer, M., and Roudier, P. (2015). <https://cran.r-project.org/web/packages/cluster/index.html>.
- Rand, W. M. (1971). "Objective criteria for the evaluation of clustering methods," J. Am. Stat. Assoc. **66**, 846–850.
- R Development Core Team, R. (2016). R Foundation for Statistical Computing. doi:10.1007/978-3-540-74686-7.
- Tan, P.-N., Steinbach, M., and Kumar, V. (2005). Introduction to Data Mining. doi:10.1016/0022-4405(81)90007-8.
- Theodoridis, S., and Koutroumbas, K. (2009). In Pattern Recognition (Fourth Edition). doi:<http://dx.doi.org/10.1016/B978-1-59749-272-0.50018-9>.