E. Sepúlveda, P. A. Dowd, C. Xu

Fuzzy clustering with spatial correction and its application to geometallurgical domaining

Mathematical Geosciences, 2018; 50(8):895-928

Fuzzy clustering with spatial correction and its application to geometallurgical domaining

E. Sepúlveda[1,2,*], P.A. Dowd[1], and C. Xu[1]

[1]School of Civil, Environmental and Mining Engineering, The University of Adelaide, Australia
[2]School of Mining Engineering, Universidad de Talca, Chile
[*]Corresponding author, email: exsepulveda@utalca.cl

Accepted for publication at Mathematical Geosciences on 22/05/2018

Abstract

This paper describes a proposed method for clustering attributes on the basis of their spatial variability and the uncertainty of cluster membership. The method is applied to geometallurgical domaining in mining applications. The main objective of geometallurgical clustering is to ensure consistent feed to a processing plant by minimising transitions between different types of feed coming from different domains (clusters). For this purpose, clusters should contain not only similar geometallurgical characteristics but also be located in as few contiguous and compact spatial locations as possible so as to maximise the homogeneity of ore delivered to the plant. Most existing clustering methods applied to geometallurgy have two problems. Firstly, they are unable to differentiate subsets of attributes at the cluster level and therefore cluster membership can only be assigned on the basis of exactly identical attributes, which may not be the case in practice. Secondly, as they do not take account of the spatial relationships they can produce clusters which may be spatially dispersed and/or overlapped. In the work described in this paper a new clustering method is introduced that integrates three distinct steps to ensure quality clustering. In the first step, fuzzy membership information is used to minimise compactness and maximise separation. In the second step, the best subsets of attributes are defined and applied for domaining purposes. These two steps are iterated to convergence. In the final step a graph-based labelling method, which takes spatial constraints into account, is used to produce the final clusters. Three examples are presented to illustrate the application of the proposed method. These examples demonstrate that the proposed method can reveal useful relationships among geometallurgical attributes within a clear and compact spatial structure. The resulting clusters can be used directly in mine planning to optimise the ore feed to be delivered to the processing plant.

1 Introduction

Geometallurgy provides new opportunities for mine planning by integrating primary and response properties to enhance the value of information for decision-making processes (Coward et al. 2009, 2013; Coward and Dowd 2015). Even though significant progress has been made in sensing and collecting geometallurgical data, there is still a significant gap in achieving effective use of these data in practical applications. Discriminating among geometallurgical characteristics is a step towards an effective use of geometallurgy for mine planning, as similar geometallurgical characteristics will have similar responses in mineral processing. From this perspective, a proper clustering of *in-situ* resources based on geometallurgical characteristics is essential to optimising operations across the entire value chain from mining to mineral processing.

Traditional geological clustering, commonly known as rock type domaining, is important in understanding the nature of the deposit but it does not necessarily reflect the responses of the ore to the various processing stages. Geometallurgical clustering (or domaining) is similar to geological clustering but focusses on the geometallurgical characteristics of the orebody to provide a basis for integrated optimisation from mining to processing (Hoal et al. 2013).

Clustering is an important problem in machine learning and, for unsupervised problems, it is one of the hardest to formulate and solve. Regression and classification are supervised because the response is known, whereas clustering partitions data on the basis of similar characteristics and, at the same time, maximises the separation of those partitions.

The classic cut-off grade approach to ore selection clearly does not consider the mineral complexity and its responses to processing, such as the energy consumption due to different hardness and grindability, concentration of deleterious elements, different recovery rates due to different geometallurgical attributes. Clustering based on geometallurgical attributes has been an active research topic over the past decade. Having more material classes (clusters) may improve the ability to select the best processing route for each parcel of mined ore so that the overall operation is optimised (Dunham and Vann 2007; Hunt et al. 2013). However, the risk of misclassification increases as more clusters are defined and this risk must be considered in any geometallurgical clustering.

Geological domains are not necessarily useful in defining processing domains that are required to reflect characteristics such as the Bond ball mill work index (BMWi), which relates to the energy used in a ball mill (Bond, 1961), or the A×b comminution index (Napier-Munn et al. 1996), which is a measure of the ore impact breakage resistance. To remediate this problem, Keeney and Walters (2011) used principal component analysis (PCA) (Wold et al. 1987) to project variables onto a two-dimensional space representing geometallurgical attributes such as mineralogy and grindability indices. Different classes were then manually defined by drawing polygons around spatially contiguous pro-

jected points on the basis of mineralogical association. These classes were used to build predictive models and propagate them into the block model using standard geostatistical indicator approaches. The same method was used in a similar case study to define four geometallurgical domains at drill-hole scale which were then scaled up to block scale by four different methods: sectional interpretation and wireframe modelling, nearest neighbour assignment, indicator kriging, and stochastic trend analysis (Newton and Graham 2011). The two comminution parameters, A×b and BMWi, were then populated into the block model by applying specific regression models in each geometallurgical domain.

Leichliter and Larson (2013) developed a geometallurgical model to cluster a deposit into two classes for two different recovery circuits: flotation circuit for less oxidized ore and heap leaching for oxidized ore. They used the variables of assays, geological mapping, mineralogy, hardness, gravity and floatability attributes to define the classes.

Hunt et al. (2014) manually clustered copper recovery domains on the basis of Al and Fe content (Low Al - High Fe and High Al - Low Fe). They pre-clustered 24 archetypes using chemical and mineralogical information. For each recovery domain, they built linear regression models using Al, Cu, Fe and grinding index from drill hole data and batch flotation tests. These models were scaled up for the block model using standard geostatistical methods.

Nguyen and Keeney (2014) built a geometallurgical domaining system by hierarchical clustering at sample scale using assay values, geotechnical logging and petrophysical attributes to model and estimate grindability response indices. Goodfellow and Dimitrakopoulos (2017) performed clustering using grades and material types to define different ore destination policies, which were used to optimise scheduling.

Garrido et al. (2017) used clay content as a measure to define the concept of geometallurgical dilution in a manner similar to mining dilution. They defined geometallurgical dilution as the ratio between the most common clay cluster and all other clusters. This dilution concept can be used in scheduling optimisation to avoid excessive changes in clay content in the ore to be sent to the processing plant.

The research discussed above demonstrates the use and ability of geometallurgical domaining in improving processing decisions and optimising scheduling to processing plants. However, most of this research uses standard clustering methods and the resulting clusters are then up-scaled to the block model using geostatistical approaches. There is no explicit imposition of spatial contiguity and compactness in the determination of clusters. In addition, the uncertainty of the clustering is not assessed.

For the explicit use of the spatial component, Oliver and Webster (1989) incorporated into the dissimilarity measure a spatial variogram model, using an isotropic exponential structure with parameters of nugget effect, sill and range. Bourgault et al. (1992) generalised the Oliver and Webster (1989) method by using a multivariate (co)variogram to account for both spatial and attributes correlations in clustering. Allard and Guillot (2000) modelled the hard clustering problem as a mosaic of independent stationary normal random functions

for the univariate case. Three different optimisation approaches were tested. One approach was based on minimising the ratio between the variance whithin a cluster and the variance between clusters. The second optimisation method used negative log-likelihood to estimate the parameters. Finally, in the third approach they used the expectation-maximisation (EM) algorithm. The spatial structure is accounted for by the kriged (estimated) mean of the random function and the associated kriging variance. Guillot et al. (2006) assumed that the spatial component is characterised by a second order stationary random field. The inference of the parameters that define the covariance function and the clusters are found by a Markov chain Monte Carlo algorithm. This method uses quantitative and categorical multivariate data. For hard clustering in the univariate case, Carlo et al. (2017) incorporated the spatial component as a non-stationary Markov random field conditioned to the k-nearest neighbourhood structure. The optimisation was performed by the EM algorithm. The method allows each cluster to have different spatial interaction modulated by a spatial covariate. Fouedjio (2016) incorporated the spatial component in the definition of the dissimilarity measure in the agglomerative hierarchical clustering method. The dissimilarity/similarity between two observations is not a simple Euclidean measure but rather a function of their spatial correlation. It is not clear what effect negative correlations in cross-variograms have on this dissimilarity measure and its performance when the spatial correlation is low, but the method is consistent with geostatistical approaches.

Based on Gaussian mixture models, Ambroise et al. (1996) proposed a method that adds a regularisation component, derived from the spatial structure, to the clustering optimisation formulation. This method takes into account the membership of all neighbours of any observation for clustering. Romary et al. (2015) incorporated the spatial component into the distance metric using a hierarchical clustering method. The distance function takes into account the spatial connectivity introduced by a moving neighbourhood. Weights for each attribute can be defined by the user and incorporated into the distance function. The coordinates are also included as attributes.

In addition to clustering methods that incorporate a spatial component, cleaning realisations of lithofacies in a regular grid, or image, helps to preserve spatial continuity. Schnetzler (1994) used two image processing pixel-base methods of dilatation and erosion, to produce cleaner images. The resulting grid does not necessarily reproduce the original statistics of the lithofacies. To overcome this issue, a post-process changes the categorical value of the pixels to match the original statistics. The probability of accepting changes is defined as the ratio of the kriging variance to the total variance. This method is only applicable to a regular grid as it was designed to correct 'noisy' grids for visualisation purposes. Deutsch (1998) improved the method of Schnetzler (1994) by using the quantile transformation to correct proportions and produce less-noisy realisations. Locations in the borders between regions are candidates for relocation. The maximum a posteriori selection algorithm replaces each location by the most probable value according to the local neighbourhood structure, based on three aspects: closeness, conditioning data, and target proportions.

In this paper, a new adapted method is proposed to cluster diverse attributes to build geometallurgical domains. The method, spatial weighted fuzzy clustering (SWFC), is based on traditional fuzzy clustering (Dunn 1973) with a novel adaptation to support mixed attributes together with the capacity to include expert knowledge and spatial structures. The formulation of the clustering algorithm based on fuzzy clustering, flexible distance metrics and feature selection is given in the methodology section. The mathematics of the proposed SWFC method is then described in the following section. Three case studies are presented to illustrate the application of SWFC, starting with a very simple illustrative example, followed by a two-dimensional case, and finally, a comprehensive three-dimensional synthetic geometallurgical block model.

2 Methodology

A dataset is defined as a set of observations or samples. Each sample is a k-dimensional vector, where each dimension represents a feature or an attribute. Each attribute can be a continuous or a categorical variable (ordinal or nominal). The goal of clustering is to partition the dataset into P sets where samples within a partition are similar and partitions are well separated. The concept of similarity within a cluster (defined as compactness, see below) and separation distance between clusters are key aspects of clustering.

2.1 Definition of Symbols and Indices

Symbols:

$\mathcal{P}$	is a set of partitions or clusters.
P	is the number of partitions or clusters.
K	is the number of dimensions of a multivariate sample.
N	is the total number of samples.
S_j	is the number of samples in the jth cluster.
v_j	is the centroid of the jth cluster.
m	is the fuzzier used in the fuzzy clustering algorithm.
u	is the membership matrix with N rows and P columns.
w	are the weights of attributes.

Indices:

i	indicates the ith sample, $1 <= i <= N$. For example x_i.
j	indicates the jth cluster, $1 <= j <= P$. For example v_j.
k	indicates the kth dimension, $1 <= k <= K$. For example w_k.

2.2 Hard Clustering

Hard clustering, or crisp clustering in the machine learning literature, seeks a non-overlapped, hard partition of a dataset and therefore the partitions $\mathcal{P}$ are

disjoint sets and each sample belongs only to one partition. One option for clustering is to find the centroids of clusters that minimise the overall distance of each sample to the centroid of its cluster, i.e.,

$$(v_1^*, \ldots, v_P^*) = \underset{v_1,\ldots,v_P}{\arg\min} \sum_{j=1}^{P} \sum_{i=1}^{S_j} D(x_i, v_j), \ \sum_{j=1}^{P} S_j = N, \tag{1}$$

where D is any distance metric, x_i is the ith sample belonging to the jth cluster.

There are many hard clustering methods and among the most used are K-Means, for continuous variables, K-Mode for categorical variables and several variants for mixed variables.

K-Means is probably the most used clustering method due to its simplicity. K-Means is solved by a two-stage iterative procedure to minimise the variance of the distances within clusters. In the first stage the centroids of clusters are assumed to be fixed and each sample is assigned to the closest centroid. In the second stage, the centroids are updated as the average of all samples within a cluster. The two stages are iterated until the overall variance of clusters is minimised. It is common to select initial centroids at random.

A very common variation is to perform an initial dimension reduction to compress the information into two or three dimensions by PCA (Ding and He 2004). After the dimensionality has been reduced, K-Means is applied to the compressed data.

For geometallurgical applications, it is important to quantify the uncertainty of belonging to a cluster but hard clustering cannot provide this assessment. The fuzzy clustering method assigns the grade of cluster membership to all samples. This grade can be used as a probability measure and, therefore, it can provide a simple way to quantify the uncertainty of clustering.

2.3 Fuzzy Clustering

Fuzzy clustering, as opposed to hard clustering, is a method that seeks to find the grade of membership of a sample with regard to each cluster (Ruspini 1969). The objective for optimisation, therefore, changes to

$$u^* = \underset{u}{\arg\min} \sum_{i=1}^{N} \sum_{j=1}^{P} (u_{ij})^m D(x_i, v_j), \ \sum_{j=1}^{P} u_{ij} = 1, \forall i = 1, \ldots, N \tag{2}$$

and

$$u_{ij}^{-1} = \sum_{j'=1}^{P} \left[\frac{D(x_i, v_j)}{D(x_i, v_{j'})} \right]^{2/(m-1)}, \tag{3}$$

where m is the fuzzier, which controls the degree of fuzziness. When m is close to 1, the fuzzy partition becomes a hard partition, that is, u_{ij} will be 0 or 1, and when m is large, u_{ij} will tend to be uniformly distributed, but always subject to $\sum_{j=1}^{P} u_{ij} = 1, \forall i$.

There are several methods to find the optimal membership, for example, fuzzy c-means, fuzzy k-modes and fuzzy k-prototypes. These methods are not designed to use mixed attributes and they do not perform feature selection. To take these features into account, a distance-based approach must be used.

2.4 Distance Metrics

Clustering essentially relies on similarity among observations and therefore the most critical aspect is the definition of a distance metric between observations. In general, the Euclidian distance (Eq. (4)) is the default selection when all attributes are continuous, however when there is a mix of continuous and categorical attributes the Euclidian distance

$$D_{\text{euclidean}}(x, y) = \sqrt{\sum_{k=1}^{K} (x_k - y_k)^2} \tag{4}$$

is not the best choice.

For two multivariate attributes x and y, the distance function can be formulated as the contribution of each dimension to the total distance (Friedman and Meulman 2004), which is given by

$$D^{(1)}(x, y) = \sum_{k=1}^{K} d_k(x_k, y_k). \tag{5}$$

This formulation gives a high degree of flexibility in the definition of specific distance functions for different kinds of attributes.

2.4.1 Continuous Attributes

For continuous attributes, such as grades, recovery rates and milling indices, the distance function is defined as

$$d_k(x, y) = |x - y|/s_k, \tag{6}$$

where s_k is any measure of dispersion, such as variance, standard deviation, interquartile range (Friedman and Meulman 2004). The importance of including dispersion is to avoid distortions with different scale values of the attributes. In this paper the standard deviation was used as dispersion measure.

2.4.2 Categorical Attributes

For categorical attributes, such as lithology, alteration types and mineralisation styles, the distance function is defined by a distance matrix, which is a symmetric square matrix of size $M \times M$, where M is number of unique values of that

attribute. For example, for a categorical attribute taking a set of possible values $h_1, h_2, \ldots, h_M$, the distance matrix is

$$\begin{bmatrix} 0 & \ldots & \theta_{1j} & \ldots & \theta_{1M} \\ \vdots & 0 & \vdots & \vdots & \vdots \\ \theta_{j1} & \ldots & 0 & \ldots & \theta_{jM} \\ \vdots & \vdots & \vdots & 0 & \vdots \\ \theta_{M1} & \ldots & \theta_{Mj} & \ldots & 0 \end{bmatrix},$$

where each θ_{ij} is a fixed value corresponding to the distance between the value h_i and h_j and $\theta_{ij} = \theta_{ji}, \forall i, j$ as the distance is symmetric.

If the categorical attribute has no preference among values, θ_{ij} can be a constant positive value for all i and $j \in 1, \ldots, M$. For example, when θ_{ij} is 1, the matrix becomes the traditional transformation to indicators, and is equivalent to

$$d_k(x, y) = \begin{cases} 0, & \text{if } x = y \\ 1, & \text{else} \end{cases}. \tag{7}$$

The flexibility of the matrix distance function allows for the definition of distance between categorical values, which is very useful for geometallurgical applications since there are, in general, categorical variables related to rock property attributes. For example, silication and silicification alterations are more similar compared to silication and argillic alteration. In this case, the distance between silication and silicification alterations should be smaller than that between silication and argillic alterations based on the definition above. The same can be considered in the case of metamorphic rocks, for example, phyllite and schist rocks are more similar compared to slate and gneiss.

The distance for the categorical attribute can be defined as

$$d_k(x, y) = \theta_{h(x)h(y)}/s_k, \tag{8}$$

where $h(x)$ denotes the value of the categorical variable x used in the definition of its distance matrix.

2.4.3 Targeted Attributes

Another flexibility of the proposed distance function is the option of including a target value in any distance function. There are situations when similarity needs to be defined as closeness to a target value; for example, the focus of interest could be on low, medium and high recoveries. Setting specific low, medium and high values of recoveries will tend to yield clusters according to those targets. Friedman and Meulman (2004) defined a distance function for one and two targets, t and u, as

$$g_k(x, y, t) = \max(d_k(x, t), d_k(y, t)) \tag{9}$$

and

$$g_k(x, y, t, u) = \min\left(g_k(x, y, t), g_k(x, y, u)\right). \tag{10}$$

These two metrics are not strictly distance metrics because they violate the identity property, but they work in practice. The problem arises when they are used to compare two values very close to each other. For example, both distances, $g_k(x, y, t)$ and $g_k(x + \epsilon, y, t)$ with $x + \epsilon < y$, are the same, due to the maximisation criterion in Eq. (9).

To correct this problem a new criterion for a single target t is defined as

$$g_k(x, y, t) = d_k(x, t) + d_k(y, t) \tag{11}$$

and its extension to multiple targets T is given by

$$g_k(x, y, T) = \min(g_k(x, y, t)), \forall t \in T. \tag{12}$$

Including the target is applicable to both continuous and categorical distance functions.

2.5 Feature Selection

In geometallurgy there are, in general, many attributes that can be used for clustering. The contributions of attributes to clustering may vary from very important to little or no importance. It is desirable that the clustering procedure considers the degree of importance of different attributes, which can sometimes be defined by expert knowledge. In this context, feature selection is an important procedure to determine the involvement of attributes and their degree of involvement as part of the clustering process. The most basic method is to consider all permutations of attributes and to select a set which performs the best. This approach obviously is computationally intensive and, as the number of attributes increases, the number of possible permutations increases exponentially. Note that the number of permutations of n attributes without repetitions is $2^n - 1$, which is equal to 1,048,575 permutations for a reasonable case of 20 attributes. On the other hand, forward and backward methods are greedy methods for feature selection. The forward method starts with one attribute and iteratively adds the attribute that most improves the distance metric. The backward method starts with all variables and removes the least useful attribute one at a time.

Another strategy is based on weights. Each attribute has, as an indicator of its degree of importance, a positive number within the range of $[0, 1]$ as its weight. This weight can then be included in the distance metric

$$D^{(2)}(x, y, w) = \sum_{k=1}^{K} w_k d_k(x_k, y_k), \tag{13}$$

where w_k is the weight of the k-feature, subject to $\sum_{k=1}^{K} w_k = 1$.

In the case of clustering, by default all attributes have the same weight in different clusters. In practice, it may be desired in some cases to impose different weights for attributes in different clusters, i.e.,

$$D^{(3)}(x, y, w, c) = \sum_{k=1}^{K} w_{ck} d_k(x_k, y_k), \tag{14}$$

where w_{ck} is the weight of the attribute k in the cluster c, subject to

$$\sum_{k=1}^{K} w_{ck} = 1, \forall c. \tag{15}$$

This weight-based feature selection mechanism can also be included in the optimisation process to determine the best weights for each attribute in each cluster.

As pointed out by Friedman and Meulman (2004), the best minimisation strategy is to assign all weight to the attribute with the lowest dispersion of observations in each cluster, which provides an incentive to spread the weights to more attributes, the distance is defined as

$$D^{(4)}(x, y, w, c, \lambda) = \sum_{k=1}^{K} d^{(4)}(x, y, w, c, \lambda) + \lambda \log(K), \tag{16}$$

where

$$d^{(4)}(x, y, w, c, \lambda) = w_{ck} d_k(x_k, y_k) + \lambda w_{ck} \log(w_{ck}). \tag{17}$$

The parameter λ controls how the weights are spread to other attributes. For larger λ, the weights will tend to be similar for all attributes whereas for smaller λ the weights will tend to be given one or a few attributes.

2.6 Spatial Correction

Another important characteristic is the spatial structure. Traditional clustering methods do not incorporate any spatial structure. In fact, if coordinates of samples are included as attributes, traditional methods are likely to produce erroneous results as samples in the same cluster are not necessarily spatially connected and these clustering procedures will tend to separate them into different clusters on the basis of their coordinates. Within the geometallurgical context, a cluster (ore parcels with similar geometallurgical characteristics) may, in many sectors, not be directly spatially connected across the deposit, and therefore a more advanced technique is required for taking the coordinates into account. As one of the goals of geometallurgical clustering is to generate clusters as spatially connected as possible, it is essential to apply a spatial correction to avoid compact zones that include a few observations that belong to a cluster different to that to which the majority belong.

Spatial correction is conceptually similar to image segmentation. In computer vision, image segmentation tries to simplify any image by assigning to each pixel a label (here equivalent to a cluster) from a small set of labels. Image segmentation has been successfully applied for applications such as cancer detection and automated driving (López and Malpica 2008; Tarabalka and Charpiat 2013; Tarabalka and Rana 2014; Wang et al. 2016).

There are many techniques for image segmentation, but the graph cut method (Boykov and Veksler 2006) is of special interest in this work because it can be integrated easily with fuzzy clustering.

The image segmentation, or labelling, problem can be formulated as an energy minimisation problem in a graph, given by

$$E(L) = \sum_{p \in I} D_p(L_p) + \sum_{pq \in \mathcal{N}} V_{pq}(L_p, L_q), \tag{18}$$

where L_p represents the label of a pixel p of the image I, D_p is the data penalty function, V_{pq} is the interaction potential, or the spatial relationship, and $\mathcal{N}$ is the neighbourhood (spatial connectivity).

Clearly there are some similarities between the image segmentation problem and our proposed clustering method. An image corresponds to the entire deposit whereas a pixel corresponds to an observation and the pixel value corresponds to an attribute of a sample. The fuzzy membership information (Eq. (3)) of each observation can be interpreted as the data penalty function. This means that each observation has a probability of belonging to a cluster and using $D_p(L_p) = -\log(u_{pL_p})$ assigns a lower data cost when the membership probability is higher and *vice versa*. The interaction potential V_{pq} corresponds to the spatial relationship among observations.

The neighbourhood can be determined by the k-nearest neighbours in the case of unstructured locations, or the surrounding cells in the case of a regular grid, which defines the connections of data in the form of a graph. Complex interaction potential functions can be formulated in the form of geostatistical (co)variograms or correlograms as defined in Bourgault et al. (1992), but a simpler one is the Potts model, which focuses on discontinuities. The Potts model is defined as

$$V_{pq}(L_p, L_q) = K_{pq} * \begin{cases} 1, & \text{if } L_p = L_q \\ 0, & \text{else} \end{cases}, \tag{19}$$

where K_{pq} may be a constant value or the cost of the spatial relationship between p and q, for example, the distance between p and q. The Potts model favours a clearer segmentation among clusters, opposite to smooth transitions, which is desired for domaining.

3 Proposed Method

The proposed method, SWFC, combines two components: (i) an adapted version of fuzzy clustering, termed weighted fuzzy clustering (WFC), and (ii) the spatial correction using the graph cut method. Both components are formulated as optimisation problems.

3.1 Optimisation Formulations

For the proposed fuzzy clustering, the concepts of compactness and separation are combined in a single objective formulation, and they are defined below.

3.1.1 Compactness

Compactness is defined as

$$COMP(m, u, v, w, \lambda) = \sum_{j=1}^{P} \sum_{i=1}^{N} u_{ij}^{m} D^{(4)}(x_i, v_j, w, j, \lambda), \tag{20}$$

where λ controls the weight values among attributes. The u matrix is given by

$$u_{ij}^{-1} = \sum_{j'=1}^{P} \sum_{k=1}^{K} \left[\frac{d_k^{(4)}(x_i, v_j, w, j, \lambda)}{d_k^{(4)}(x_i, v_{j'}, w, j', \lambda)} \right]^{2/(m-1)}. \tag{21}$$

3.1.2 Separation

Separation is defined as

$$SEP(m, v, w, \lambda) = \sum_{j=1}^{P} \sum_{j'=1, i \neq j'}^{P} \sum_{k=1}^{K} d_k^{(4)}(v_{ik}, v_{j'k}, \max(w_j, w_{j'}), j, \lambda). \tag{22}$$

The maximum criterion in Eq. ((22)) is required because different clusters may not share the same weights, in which case the maximum weight is used for the kth dimension.

3.1.3 Objective

The proposed clustering seeks to minimise compactness of clusters and, at the same time, to maximise separation between clusters. A single objective formulation that incorporates both aims is defined as

$$(v^*, w^*) = \arg\min_{v,w} (COMP(m, u, v, w, \lambda) + \frac{C}{SEP(m, v, w, \lambda)}), \tag{23}$$

where C is a constant which scales the importance of separation as a criterion in the optimisation formulation. The lower the value of C, the less important is any increment in separation. Our experiments indicate that a value of $C = 15$ is appropriate to give more relative importance to compactness over separation but a complete assessment of the impact of different values is recommended. Other expressions that combine compactness and separation in a single objective may also be explored.

There are two main obstacles to solving this optimisation problem. The first is the non-convexity of the problem, meaning that the global minimum is hard to find. The second is the difficulty of finding the cluster centroids and defining a proper set of weights that can be used.

The first obstacle can be dealt with by the use of metaheuristics, which is a simple technique to solve optimisation problems with many local optima. The second obstacle is solved by a two-stage procedure in the proposed method. In the first stage, the optimal centroids and membership are found for a given set of fixed weights. In the second stage, the weights are optimised given the clusters and memberships found in the first stage. These two steps are iterated until convergence.

3.2 Implementations

Genetic algorithm (GA) metaheuristic is used not only because of its simplicity, flexibility and good performance, but also because GA has been successfully used as an optimisation method for clustering (Luchi et al. 2016; Maulik and Bandyopadhyay 2000; Nanda and Panda 2014).

3.2.1 Genetic Algorithm

A genetic algorithm is a metaheuristic optimisation method that emulates the process of evolution. There are three main concepts involved in GA: selection, crossover and mutation.

The selection operation imitates the natural selection process in which better individuals have more chances to pass their genes to the next generation. A fitness value is assigned to each individual, which corresponds to the evaluation function to be optimised. Crossover produces new individuals combining the genes of the parents. Mutation produces a new individual by mutating a small part of the gene of an individual.

These three operations are executed for many generations to ensure that the best final individual of the population is a good local optimum. A complete tutorial on GA can be found in Whitley (1994).

The hyper-parameters of GA are the number of individuals in the population, the number of generations, the operations of selection, crossover and mutation, and the probabilities of crossover and mutation. Given these hyper-parameters, which depend on the problem to be solved, the GA procedure for minimisation is given by algorithm 1.

Algorithm 1: Minimisation by GA

```
Result: best_individual, best_fitness
Data: npop: size of the population, ngen: number of generations
population ← set of random individuals of size npop;
foreach ind ∈ population do
    fitness (ind) ← evaluation(ind);
end
fitness (ind) ← set of random individuals of size npop;
for iteration ← 1 to ngen do
    offspring ← selection(population);
    foreach ind ∈ offspring do
        if rand() < prbcx then ind ← crossover(ind);
        if rand() < prbmut then ind ← mutation(ind);
        fitness (ind) ← evaluation(ind);
        if fitness (ind) < best_fitness then
            best_fitness ← fitness (ind);
            best_individual ← ind;
        end
    end
end
```

The most important design aspect of any GA is the solution codification (genome), which is problem dependent. For a given problem codification, its corresponding crossover and mutation operations must also be defined. In our implementations, the crossover function is the standard uniform crossover. Selection is performed by tournament selection.

3.2.2 GA for Optimising Centroids

For the first stage discussed above, the problem reduces to finding the cluster centroids that minimise the objective function given in Eq (23). Thus, the genome in this case represents the centroids of each cluster.

The initial centroids are selected from samples at random. The mutation operation perturbs one dimension of one centroid: for continuous variables, the perturbation corresponds to a random value drawn from a normal distribution $\mathcal{N}(\mu = 0, \sigma = 0.1)$, whereas for categorical variables, the perturbation simply selects a different value of their categories at random.

The evaluation function for optimising centroids is given by algorithm 2 and the mutation operator is given by algorithm 3. The function $\dim(A)$ returns the number of the rows and columns of a matrix A.

Algorithm 2: Evaluate clustering criteria for optimising centroids

Result: $\mathsf{compactness} + \frac{C}{\mathsf{separation}}$
Data: V: centroids
Parameters: m: the fuzzier, λ: weight strenght, w: weights, C: constant for the contribution of separation
$U \leftarrow$ Equation (21);
$\mathsf{compactness} \leftarrow$ Equation (20);
$\mathsf{separation} \leftarrow$ Equation (22);

Algorithm 3: Mutation for finding centroids

Result: mutated V
Data: V: centroids
Parameters: C: set of categorical values for the attribute k
$P, K \leftarrow dim(V)$;
$j \leftarrow randint(1, P)$;
$k \leftarrow randint(1, K)$;
if *attribute k is continuous* **then** $V[j,k] \leftarrow V[j,k] + randnorm(0, 0.1)$;
else $V[j,k] \leftarrow$ select at random from $C - V[j,k]$;

3.2.3 GA for Optimising Weights

In the second stage, the weights are optimised with fixed centroids, and the problem reduces to finding the weights that minimise the objective function given in Eq. (23).

As the weights are within the range $[0, 1]$, the mutation adds a small number drawn from the normal distribution, $\mathcal{N}(\mu = 0.0, \sigma = 0.01)$. Two integer numbers are selected at random, one for a cluster and the other for an attribute to modify.

Note that the weights must sum to one. In addition, expert knowledge can be used to set a specific weight to an attribute and the perturbation can preserve these values.

The optimisation of weights is given by algorithm 4 and the mutation operator by algorithm 5.

Algorithm 4: Evaluate clustering criteria for optimising weights

Result: $\mathsf{compactness} + \frac{C}{\mathsf{separation}}$
Data: W: weights
Parameters: m: the fuzzier, λ: weight strenght, V: centroids, C: constant for the contribution of separation
$U \leftarrow$ Equation (21);
$\mathsf{compactness} \leftarrow$ Equation (20);
$\mathsf{separation} \leftarrow$ Equation (22);

Algorithm 5: Mutation for optimising weights

Result: mutated W
Data: W: weights
$P, K \leftarrow dim(W)$;
$j \leftarrow randint(1, P)$;
$k \leftarrow randint(1, K)$;
$W[j, k] \leftarrow W[j, k] + randnorm(0, 0.01)$;
//Normalisation
$W[j] = W[j] / \sum_{k=1}^{K} W[j, k], \forall j$;

3.2.4 Proposed Clustering Method (SWFC)

The final proposed method is shown in algorithm 6 and the spatial correction is given by algorithm 7.

Algorithm 6: Spatial weighted fuzzy clustering

Result: *clusters*: clusters assigned to each sample, U: membership matrix, V: centroids, W: weights,
Data: *locations*: coordinates of observations, *samples*: multivariate attributes of observations, P: number of clusters
Parameters: m: the fuzzier, λ: weight strength, V: centroids, C: constant for the contribution of separation,
//N samples of K attributes
$N, K \leftarrow dim(samples)$;
//start with uniform weights on all attributes
$CurrentW \leftarrow ones(P, K)$;
$CurrentW \leftarrow CurrentW/K$;
repeat
  //stage1
  $V \leftarrow$ OptimiseCentroids$(CurrentW, samples, P, C)$;
  //stage2
  $W \leftarrow$ OptimiseWeights$(CurrentW, V, m, C)$;
  $CurrentW \leftarrow W$;
until $\sum |W - CurrentW| < \epsilon$;
//Spatial correction
$clusters \leftarrow$ SpatialCorrection$(locations, U)$;

The function BuildNeighbourhood(*locations*) returns the edges of the spatial structure of the *locations*. The spatial structure can be defined using Delaunay tessellation, k-nearest neighbour, or the surrounding blocks in a structured block model.

Algorithm 7: Spatial Correction by graph cut method

```
Result: clusters: clusters assigned to each sample
Data: U: membership matrix, locations: coordinates of observations,
//N samples and P clusters
N, P ← dim(U);
//Use K-Nearest-Neighbour or regular grid
edges ← BuildNeighbourhood(locations);
//Data penalty is lower for higher probabilities and higher for lower
 probabilities
D ← −log(U);
//Interaction potential by Potts model
for i ← 1 to P do
    for j ← 1 to P do
        if i = j then V[i, j] ← 0;
        else V[i, j] ← 1;
    end
end
clusters ← GraphCut(edges,D,V);
```

3.2.5 Efficiency and Scalability

The efficiency of SWFC relies mainly on its two components, the optimisation of the centroids and weights and the spatial correction by the graph-cut method.

When the centroids are optimised, GA calculates, for each individual, the membership matrix, separation and compactness. The complexity of the calculation of the membership matrix is $\mathbf{O}(N * P^2)$, separation is $\mathbf{O}(N * P * K)$, and compactness is $\mathbf{O}(K * P^2)$. As usually $N \gg K$ and assuming that $K > P$, an upper bound for the complexity of the evaluation of each individual is $\mathbf{O}(N * P * K)$.

The GA algorithm needs to evaluate *npop* individuals over *ngen* generations, therefore, the total complexity of the WFC algorithm is $\mathbf{O}(N * P * K * npop * ngen)$. Our results indicate that WFC converges in less than 20 iterations (main loop in algorithm 6).

The complexity of the graph-cut algorithm used is $\mathbf{O}(N * P^2)$ (Boykov and Veksler 2006).

The complexity of the fuzzy clustering is comparable to the K-Means algorithm at each iteration, but the difference is in the optimisation procedure, where SWFC is more computational intensive. Nevertheless, it overcomes two aspects: the use of GA helps in escaping from local minima and both fuzzy clustering and feature selection are jointly optimised.

The complexity of SWFC does not have a high impact on the number of samples N, since it scales linearly as a function of N. Also, P is usually small for practical reasons (no greater than 10) and K is, in general, less than 100.

GA is a stochastic optimisation method and the results may be affected by the initial random seed. Our experiments showed that different seeds produce

very stable results. All results reported are based on a single, representative run.

3.2.6 Assessing the Number of Clusters

There is no common choice for the number of clusters and this selection largely depends on the data and the application of the clusters. However, there are several indices that can be used to assess the cluster quality. The silhouette index (SI) (Eq. (24)) comprises compactness and separation (Rousseeuw 1987). This index is a real number in the range [-1,1]. An index close to -1 means there is little or no cluster structure and close to 1 indicates perfect compactness within clusters and clear separation between clusters. This index uses only the distance among observations and it is defined as

$$SI = \frac{1}{N}\sum_{k=1}^{N} SI_k \tag{24}$$

where

$$SI_k = \frac{1}{N}\sum_{i=1}^{N}((b_i - a_i))/(\max(b_i - a_i)), \tag{25}$$

N is the total number of points, a_i is the average distance between point i and all other points in its own cluster, and b_i is the minimum of the average distances between i and points in other clusters.

Another index for assessing the quality of clusters is the Davies-Bouldin index (DBI), which describes how well the clustering has been done as measured by the distance between observations and cluster centroids. Values of this index close to 0 suggest better cluster structures (Davies and Bouldin 1979). DBI is calculated using the following equations

$$DBI = \frac{1}{P}\sum_{i=1}^{P} D_i, \tag{26}$$

$$D_i = \max(R_{ij}),\ i = 1, \ldots, P, j = 1, \ldots, P, \tag{27}$$

$$R_{ij} = \frac{T_i + T_j}{M_{ij}}, \tag{28}$$

$$M_{ij} = \left(\sum_{k=1}^{K} |v_{ik} - v_{jk}|^p\right)^{1/p}, \tag{29}$$

$$T_i = \left(\frac{1}{S_i}\sum_{j=1}^{S_i} |x_j - v_i|^p\right)^{1/p}, \tag{30}$$

with $p = 2$ for the Euclidean norm.

Table 1: Parameters used in the algorithms

Parameter	Value	Observation
npop	100	Population size
ngen	100	Number of generations
prbcx	0.8	Probability of crossover
prbmut	0.4	Probability of mutation
Tournament size	9	9 individuals used in the tournament selection
λ	0.25	Weight strength
m	2.0	Fuzzier
C	15	Constant for the contribution of separation

4 Application

Three examples are presented in this paper to demonstrate the application of the proposed method. The first example is a simple synthetic case that illustrates the difficulties of traditional methods when clustering using different attributes. The second example is a cross-section of a simulated copper porphyry deposit (Garrido et al. 2017). The third example is a full synthetic geometallurgical block model (Lishchuk 2016).

The results of K-Means and PCA clustering methods are compared. The spatial correction also is applied to K-Means and PCA, using the membership matrix as the inverse of the squared distance between each sample to the centroids, given by

$$u_{ij} = \frac{1/\left\|x_i - v_j\right\|^2}{\sum_{j'=1}^{P} 1/\left\|x_i - v_{j'}\right\|^2}. \tag{31}$$

K-Means and PCA with spatial correction are denoted by SK-Means and SPCA respectively.

Table 1 shows the values of the parameters used in the algorithms for the three examples.

4.1 Illustrative Example

In this example, there are four attributes: grades of copper, gold and iron, and recovery of copper, denoted as Cu, Au, Fe and Rec respectively. Although these attributes do not usually follow normal distributions, for the sake of simplicity, they were taken from normal distributions, but their means and standard deviations are different so as to form four clusters, see Table 2.

All attributes follow their default normal distributions but for specific clusters, they follow normal distributions with different means and lower variances. Cluster 1 includes only Cu, Fe and Rec; cluster 2: Cu, Fe and Au; cluster 3: Cu, Fe, Rec; and cluster 4 only Fe and Au.

A spatial component was assigned to each cluster: half of cluster 1 is uniformly located in the region of $[(10.0 - 35.0), (10.0 - 35.0)]$ and the other half

Table 2: The design of four clusters based on combination of Cu, Fe, Au and Rec

Attribute	Default	C1 (red)	C2 (blue)	C3 (green)	C4 (yellow)
Cu	$\mathcal{N}(0.68, 0.3)$	$\mathcal{N}(0.8, 0.05)$	$\mathcal{N}(0.9, 0.05)$	$\mathcal{N}(0.4, 0.05)$	
Fe	$\mathcal{N}$(2.56,1.15)		$\mathcal{N}$(1.5,0.1)	$\mathcal{N}$(1.2,0.1)	$\mathcal{N}$(4.0,0.1)
Au	$\mathcal{N}$(21.5,11.08)	$\mathcal{N}$(30.0,1.0)	$\mathcal{N}$(15.0,1.0)		$\mathcal{N}$(40.0,1.0)
Rec	$\mathcal{N}$(81.43,6.04)	$\mathcal{N}$(88.0,1.0)		$\mathcal{N}$(70.0,2.0)	

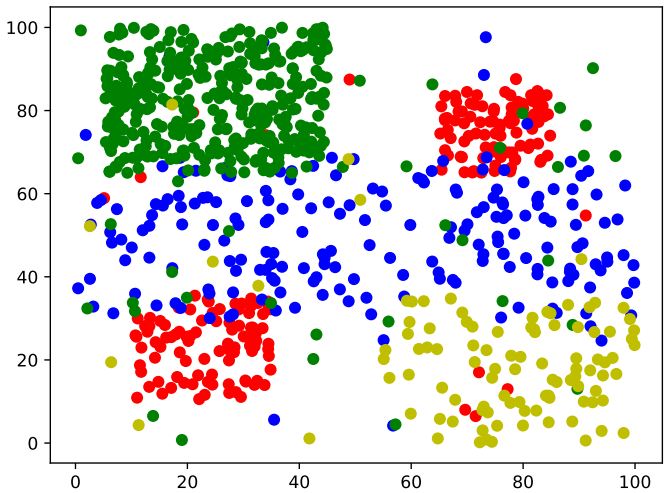

Figure 1: Scatter plot of true four clusters

in $[(65.0 - 85.0), (65.0 - 85.0)]$; all of cluster 2 is located uniformly in the region of $[(0.0 - 100.0), (30.0 - 70.0)]$; all of cluster 3 is located uniformly in the region of $[(5.0 - 45.0), (65.0 - 100.0)]$; and finally, cluster 4 is located uniformly in $[(55.0 - 100.0), (0.0 - 35.0)]$.

The illustrative example comprises 200 samples of cluster 1, 200 samples of cluster 2, 400 samples of cluster 3, and 100 samples of cluster 4. Of these 900 samples, the locations of 100 samples are randomised uniformly for the entire region of $[(0.0 - 100.0), (0.0 - 100.0)]$. Figure 1 shows the final locations of all samples.

In this example, the number of clusters is known and, therefore, the example can be used to assess the efficacy of different clustering methods. Both K-Means and PCA perform better with three clusters, although PCA has similar results with four clusters. The proposed method, with or without spatial correction, significantly outperforms K-Means and PCA in finding the correct number of clusters (Table 3).

Using four clusters, the performance of different clustering methods can be

Table 3: Davies-Bouldin and silhouette indices of K-Means, PCA, and WFC for different number of clusters

Clusters	K-Means		PCA		WFC	
	DBI	SI	DBI	SI	DBI	SI
2	0.852	0.504	1.026	0.496	0.234	0.802
3	0.819	0.537	0.801	0.537	0.306	0.776
4	0.974	0.427	0.818	0.520	0.185	0.866
5	0.907	0.452	0.856	0.442	1.771	0.577
6	0.923	0.439	0.931	0.426	1.231	0.600

Table 4: Explained variance of PCA components

Component	Explained variance (%)
1	53.55
2	30.14

further assessed. The imposed spatial structure and added noise make it difficult for the K-means clustering method to reproduce four clusters (Fig. 3a). The poor performance of the K-means method is clearly shown in the figure in which green and blue clusters are correctly identified but the other two clusters are not. In addition, the importance of different attributes is not identified. For example, the Cu attribute is well clustered (good separation and low variance) across all clusters (Fig. 2a) despite the fact that no Cu dependency is imposed in cluster 4 (Table 2).

PCA overcomes some of the problems of K-Means. Two components were used. Table 4 depicts the contribution of each component to the total variance. When the data are projected, and therefore compressed to only two dimensions, the cluster structure can be clearly seen by visual inspection (Fig. 5). Despite the obvious cluster structure, PCA clustering performs better than K-Means but clusters 1 and 4 are still misclassified to some extent (Fig. 4).

There are two hyper-parameters that need to be defined to apply SWFC: the fuzzier m and the parameter λ for weights. Most cases reported in the literature suggest that a value of 2.0 for m is a reasonable choice to account for uncertainty (Pal and Bezdek 1995; Ren et al. 2016), and $m = 2.0$ is used in all applications of SWFC discussed in this paper. For parameter λ, there is no rule of thumb guidance in the literature. A small value close to 0 means that all weight will be assigned to one attribute, while a large value will tend to assign the same weights to all attributes. For this example, the influences of different λ values between 0.05 and 1.0 on the weights assigned to different attributes are shown in Fig. 6. This figure is useful for assessing the impact of λ on the number of attributes that are considered significant for finding the cluster structure so that an appropriate λ value can be selected. For this example, a value of 0.25 was selected for λ because it tends to give importance to two or three attributes for clustering, which matches the number used to create the clusters in the first

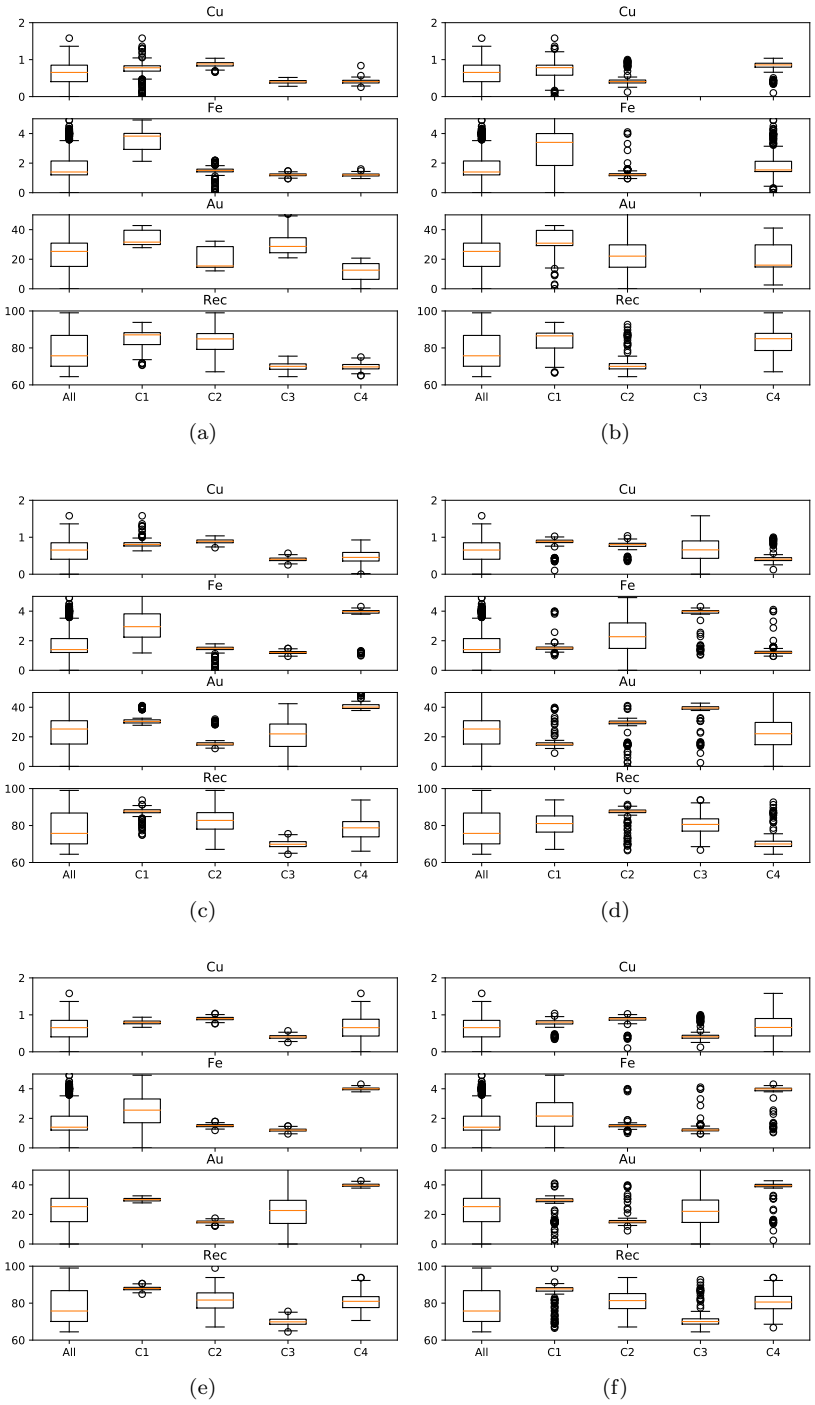

Figure 2: Boxplots of all attributes in clusters found by (a) K-Means, (b) SK-Means, (c) PCA, (d) SPCA, (e) WFC and (f) SWFC. Attributes from top to bottom are Cu, Fe, Au and Re

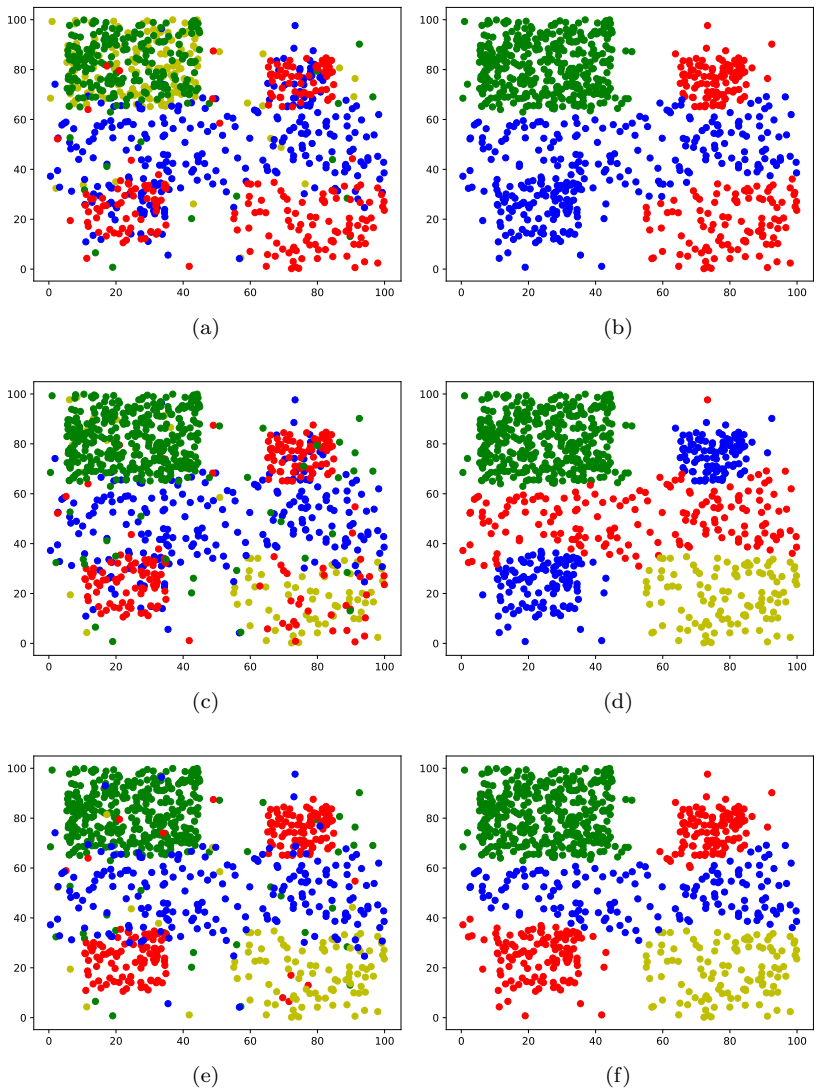

Figure 3: Scatter plot of clusters found by (a) K-Means, (b) SK-Means, (c) PCA, (d) SPCA, (e) WFC, and (f) SWFC

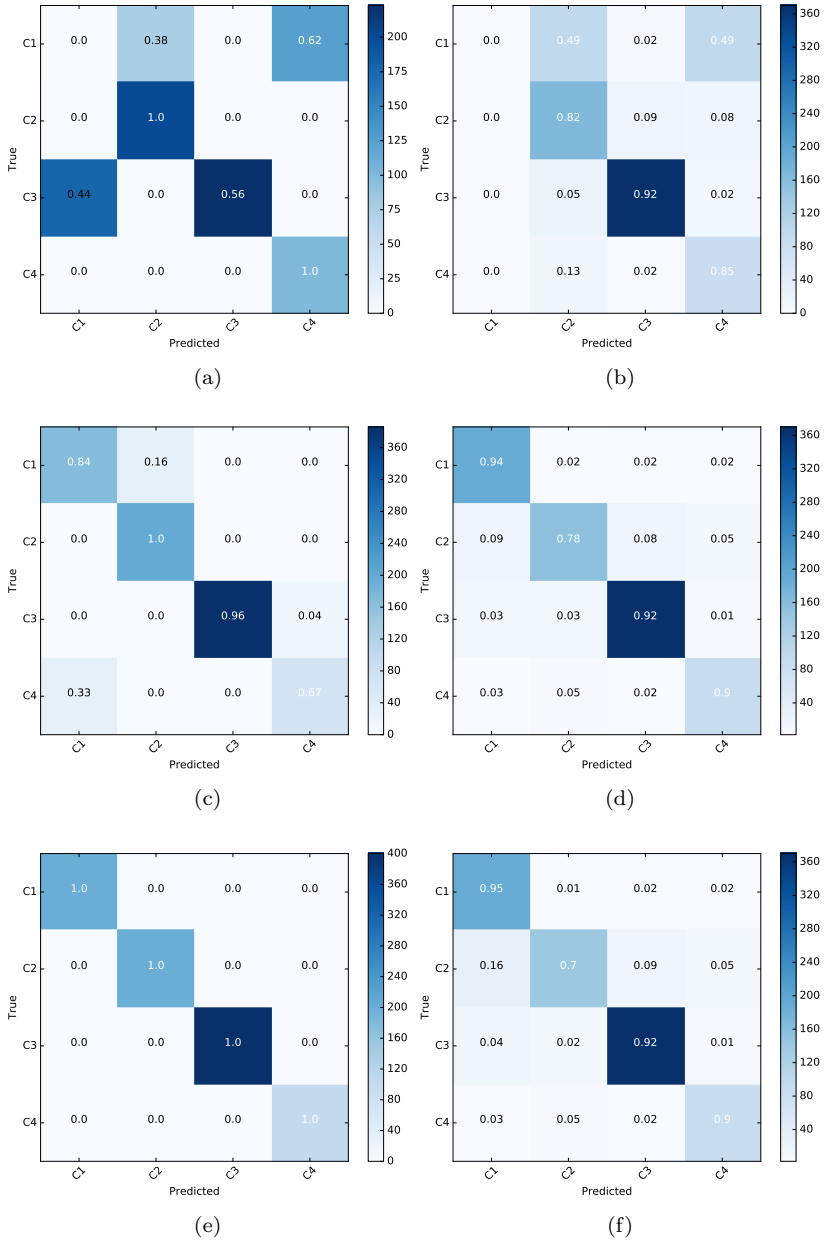

Figure 4: Confusion matrix of clusters found by (a) K-Means, (b) SK-Means, (c) PCA, (d) SPCA, (e) WFC, and (f) SWFC

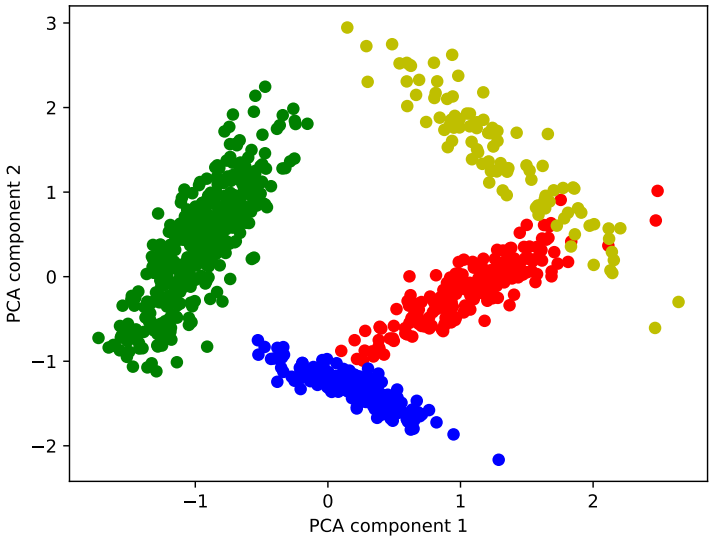

Figure 5: Projection on first and second principal components of PCA

place.

The spatial correction was also applied to the three clustering algorithms, K-Means, PCA and WFC, to compare the effect of this correction. The spatial correction applied to K-Means results in the loss of cluster 1 (Fig. 4b), due to its weak membership values, which are reassigned to cluster 4 (Fig. 3b). Although the correction improves PCA compared to K-Means, its performance remains the same as WFC (Fig. 3e). The spatial correction applied to PCA gives results that are similar to SWFC (Fig. 4 d and f). WFC is impressively exact, even identifying the correct clusters for the randomly located observations. The performance of SWFC slightly decreases but it still significantly outperforms SK-Means (Fig. 4 b and f). The spatial correction alters the final cluster membership of only a few observations in order to make the clusters more spatially compact (Fig. 4f). The boxplots for WFC and SWFC show no substantial difference in their performance statistics (Fig. 2 e and f).

This simple illustrative example clearly shows that traditional methods struggle to find the cluster structure correctly when those clusters are defined by different attributes. The proposed method significantly outperforms the traditional methods and can perfectly reveal the cluster structure in this case as well as producing compact clusters in terms of spatial connectivity.

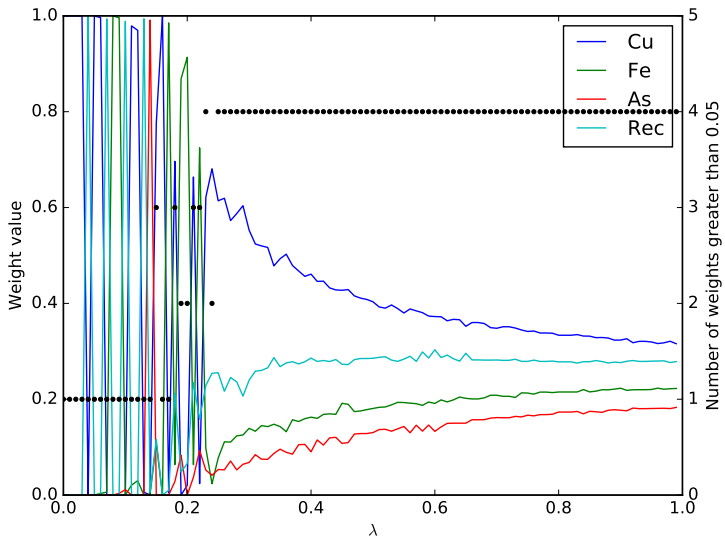

Figure 6: Weights (left y-axis) of the four attributes for different values of λ (x-axis). The black points indicate the number of weights greater than 0.05 (right y-axis) at each value of λ

Table 5: Explained variance of PCA components

Component	Explained variance (%)
1	53.15
2	17.13
3	14.50

4.2 Simulated Copper Porphyry Deposit Example

This example is a simulated deposit based on actual data from a copper porphyry deposit (Garrido et al. 2017). The orebody is mainly dominated by disseminated chalcopyrite and with four categories of large, moderate, small and minimum presence of clay. A cross-section, comprising 6,462 blocks is used to illustrate the results of SWFC in two dimensions. The first level clustering results in 4,268 blocks of waste and 2,194 blocks of ore. The ore cluster has two grade elements (copper and arsenic), two response attributes (copper recovery and bond index), and one categorical attribute (presence of clay in low, medium and high degree). SWFC is applied to the ore super-cluster to find four sub-clusters using these 5 attributes. For PCA clustering, three principal components were used. Table 5 depicts the explained variance of each component of the total variance.

Cluster 1 is characterised by high content of clay (category 2), low grade

Table 6: Centroids of the four clusters found by SWFC

Cluster	Clay	Copper	Arsenic	Recovery	Bond index
1	2	0.436	20.00	76.48	10.65
2	1	0.454	20.00	83.03	12.70
3	0	0.968	63.34	94.11	16.04
4	0	0.780	20.00	94.98	16.49

Table 7: Weights of the four clusters found by SWFC

Cluster	Clay	Copper	Arsenic	Recovery	Bond index
1	0.5275	0.0971	0.3152	0.007	0.0533
2	0.4017	0.1102	0.3877	0.0357	0.0647
3	0.5159	0.0241	0.0034	0.2786	0.1780
4	0.4595	0.0118	0.2403	0.2039	0.0846

values of copper and arsenic (detection limit of 20 for arsenic), and low recovery due to the clay content and low hardness. Cluster 2 is characterised by medium content of clay (category 1), low grade values of copper and arsenic, and slightly higher recovery. Interestingly, SWFC has identified two additional clusters for low content of clay (cluster 3 and 4), which are characterised by high recovery and high bond index but are well separated by arsenic content (low and high), see Table 6.

Table 7 lists the weights assigned to different attributes in SWFC, which effectively shows the degree of importance of each attribute for each cluster. Clay content is the most important attribute for all clusters, which is consistent with the copper recovery performance and hardness as high clay content is related to low copper recovery and softer rocks. The second most relevant attribute differs for different clusters. Arsenic content is more relevant for clusters 1, 2 and 4, whereas recovery is for cluster 3. One interpretation is that SWFC was capable of separating clusters 3 and 4 in terms of arsenic attribute although both have low clay content.

Figure 7 shows the statistics of the four most relevant attributes for K-Means, PCA and WFC. Clay content is well separated, but K-Means separates clays in a different way. The clusters found by PCA and WFC look very similar, except for the size of cluster 3. For all methods, copper grade is split into two main groups: low and high. High content of arsenic is very relevant for cluster 3 in both PCA and WFC, whereas recovery is well separated among clusters. In general, PCA and WFC perform similarly and both are superior to K-Means.

The spatial connectivity of the clusters is another important aspect for SWFC. The results in terms of spatial connectivity for SK-Means, SPCA, and SWFC are shown in Fig. 8. SK-Means does not preserve cluster 2 (Fig. 8 a and b) and SPCA does not preserve cluster 4 (Fig. 8 c and d) due to the poor connectivity of the clusters. For WFC, there are several blocks in cluster 4 (blue) that are spatially unconnected (Fig. 8e); the spatial correction generates

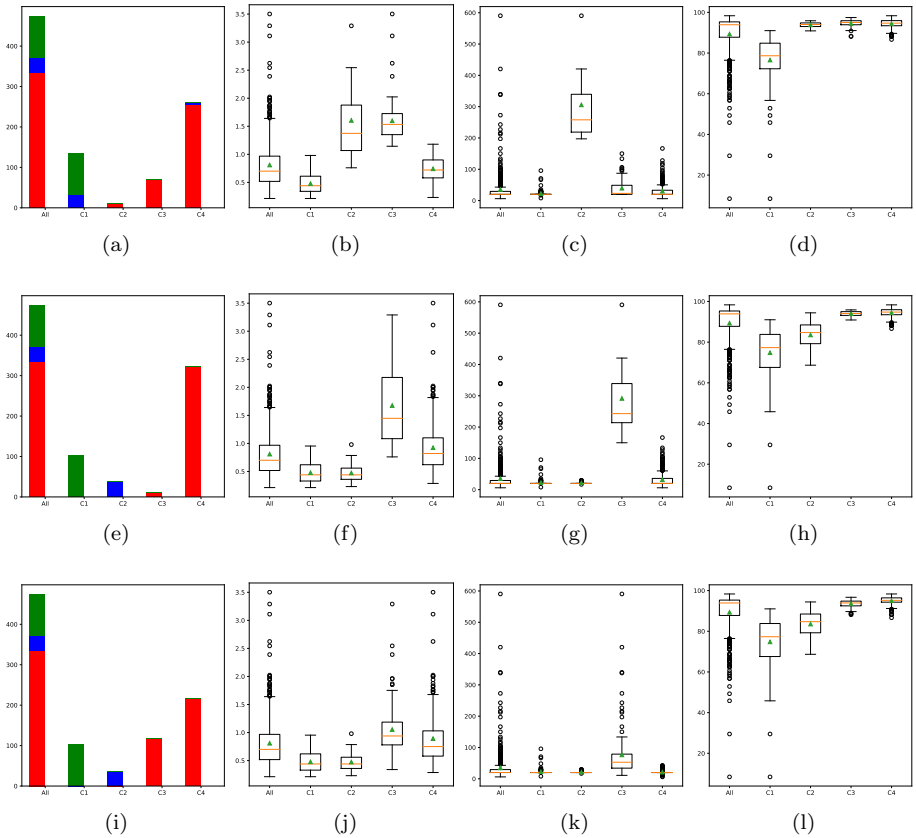

Figure 7: Statistics of the four most relevant attributes for all observations and for the four clusters by (a-d) K-Means, (e-h) PCA, and (i-l) WFC. Attributes are from left to right: clay content, copper, arsenic, and recovery

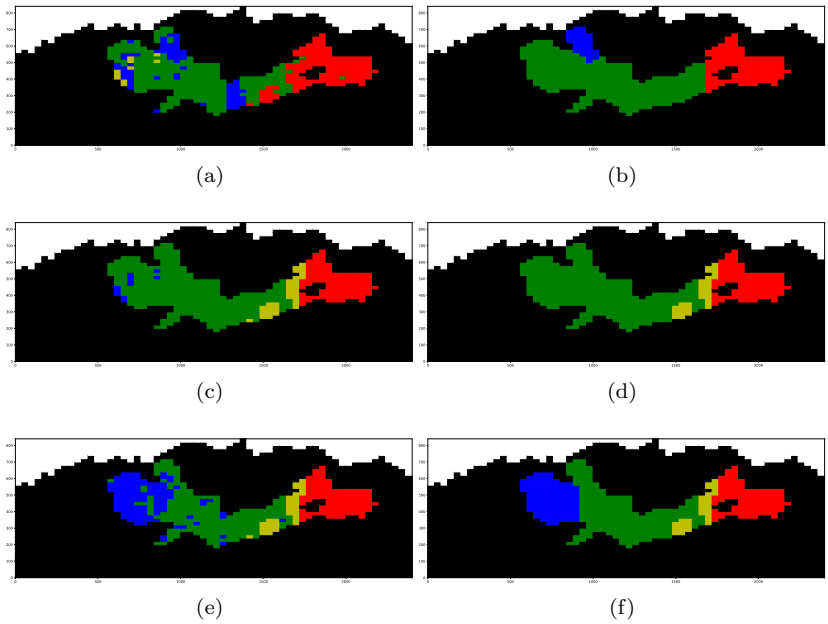

Figure 8: Map of (a) K-Means, (b) SK-Means, (c) PCA, (d) SPCA, (e) WFC, and (f) SWFC. Black represents waste rock. Red, Yellow, Green and Blue represent the four clusters

Table 8: Attribute descriptions of the geometallurgical block model

Type of Attribute	Attributes	Observation
Rock properties	Lithology, Specific gravity	Lithology codes: 1: Semi-massive ore. Feldspar rich dominated with albite. 2: Massive ore. Amphibole dominated with minor apatite and biotite. 3: Massive ore. Apatite dominated with minor Amphibole.
Mineral groups	Magnetite (Mgt), Hematite (Hem), Albite (Ab), Actinolite (Act), Apatite (Ap), Biotite (Bt)	Fe minerals: magnetite and hematite. Gangue: albite, actinolite, Apatite and biotite. Actinolite has some recoverable content of Fe.
Chemical elements	O, F, Na, Mg, Al, Si, P, Cl, K, Ca, Ti, V, Mn and Fe	
Processing	Iron recovery	Recovery in magnetic separation process

much more spatially compact clusters (Fig. 8f), which is a desirable property achieved by the proposed method.

This simple two-dimensional case study illustrates the power of SWFC to produce compact and well separated clusters while preserving their spatial connectivity. It does so by selecting the appropriate attributes relevant for the cluster structure using the optimisation technique discussed above. The resulting clusters can then be much more effectively used for scheduling. Taking into account the characteristics of each cluster, for example, the scheduler may avoid too many jumps between different clusters in order to derive sets of blocks with similar characteristics to be delivered to the plant for a particular time period.

4.3 Simulated Geometallurgical Block Model Example

This geometallurgical block model was built based on the Malmberget iron deposit in northern Sweden using simulation modules for geology, sampling, production and mining economics. The complete methodology used to build this geometallurgical block can be found in (Lishchuk 2016; Lund et al. 2015).

This geometallurgical model has $50 \times 50 \times 50$ number of blocks of size $5 \times 5 \times 5$m, where 21,710 of them are ore blocks. The 23 attributes used for clustering are: lithology, 6 mineral grades, 14 chemical element grades, specific gravity, and iron recovery (Table 8).

Table 9: Explained variance of PCA components

Component	Explained variance (%)
1	50.51
2	24.65
3	12.26

Table 10: DBI and SI of K-Means, PCA, and WFC for different number of clusters

Clusters	K-Means		PCA		WFC	
	DBI	SI	DBI	SI	DBI	SI
2	0.98	0.50	1.05	0.46	0.53	0.69
3	0.95	0.39	1.38	0.33	0.48	0.71
4	0.96	0.44	0.97	0.43	0.63	0.61
5	1.20	0.43	1.27	0.41	1.00	0.53
6	1.21	0.42	1.38	0.41	1.56	0.42
7	1.24	0.41	1.28	0.39	2.57	0.44
8	1.36	0.40	1.27	0.31	2.59	0.36
9	1.34	0.32	1.38	0.30	4.13	0.36
10	1.41	0.32	1.32	0.30	6.40	0.29

In this example, the focus is on building geometallurgical domains for iron recovery. Lithology should play an important role in clustering, but lithology alone in this case is not sufficient to discriminate iron recovery.Finding the other attributes that can contribute to a better identification of clusters is very important.

The flexibility of SWFC is illustrated by setting the objective as achieving a geometallurgical domaining for Fe recovery. To do so, the targeted distance is used for iron recovery with the values at 15%, 50% and 85% percentiles of its distribution, corresponding to recovery values of 82.41%, 88.98% and 91.22% respectively, and a weight of 15% was used for the recovery attribute. These conditions provide a guide for SWFC to find three clusters. The purpose of imposing the target and weight to Fe recovery is to find which secondary attributes would be useful for clustering the structure as it is expected that discovered clusters will tend to have Fe recovery values close to the defined targets.

The number of clusters is set to three according to both DBI and SI values (Table 10). For PCA clustering, three principal components were used. Table 9 shows the explained variance of each component of the total variance.

The results for apatite, magnetite, iron, iron recovery, and rock type are used to compare the performance of the clustering methods.

Some interesting observations can be made about the centroids of the three clusters (Table 11). The targeted distance applied to iron recovery is very well represented by the SWFC method, but not so by SK-Means and SPCA because

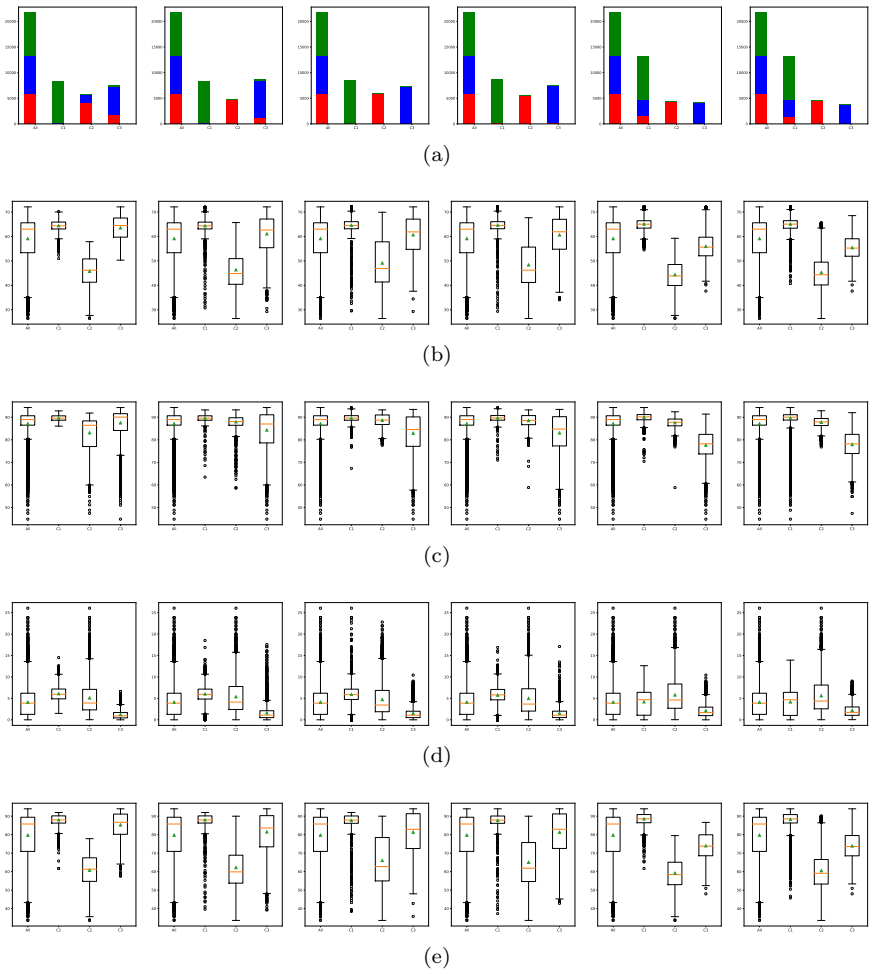

Figure 9: Distribution of (a) Lithology, (b) Fe, (c) Fe recovery, (d) Apatite, and (e) Magnetite. Clustering methods from left to right are: K-Means, SK-Means, PCA, SPCA, WFC and SWFC

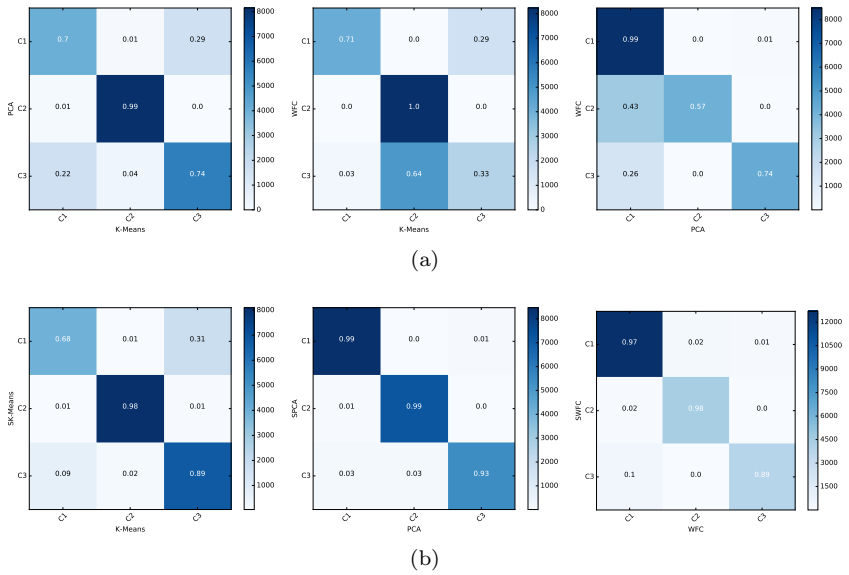

Figure 10: (a) Pairwise cluster discrepancy between K-Means, PCA and WFC. (b) Pairwise cluster comparison between each clustering method before and after spatial correction

they do not use targeted distance functions. Lithology is also well discriminated and the positive correlation between apatite and iron recovery is maintained. Apatite is separated only as low content for cluster 3 in SWFC, whereas SK-Means and SPCA do not separate apatite to the same extent in cluster 3 but are similar for clusters 1 and 2. The centroids of the three clusters for iron grade show a similar separation for the three methods.

Figure 9 shows the statistics of the six clustering methods. SPCA separates each lithology into each cluster as does SK-Means. SWFC clusters lithology in a different way, for example, cluster 1 contains the three lithologies, but clusters 2 and 3 contain only lithology 1 and 3 respectively. This difference may be explained by the fact that WFC and SWFC seek the imposed targets for iron recovery.

A summary of the differences of clustering among K-Means, PCA and WFC is given in Fig. 10a. K-Means shows some differences compared to PCA and WFC, while PCA and WFC are very similar in performance (Fig. 10b).

This case study of a complete three-dimensional geometallurgical block model demonstrates the flexibility of applying SWFC in practice. The objective was for the three clusters to be centred in specific values of iron recovery, which was fully achieved. The spatial correction step in the three clustering methods makes some changes in the final membership (Fig. 10b). Although these changes are small, they are worthwhile as they ensure that the derived clusters

Table 11: Centroids of the three clusters for lithology, apatite, magnetite, iron and iron recovery found by SK-Means, SPCA and SFWC.

Cluster	Lithology	Apatite	Magnetite	Iron	Iron recovery
SK-Means					
1	3	2.85	87.96	64.37	89.60
2	1	3.81	62.20	46.33	87.94
3	2	2.14	81.48	61.07	84.33
SPCA					
1	3	2.73	87.78	64.66	89.71
2	1	3.49	65.04	48.35	88.51
3	2	2.30	81.29	60.64	83.07
SWFC					
1	3	4.51	88.08	64.87	91.22
2	1	4.88	57.05	42.98	88.98
3	2	1.82	72.87	54.83	82.41

are as spatially compact as possible.

5 Conclusions and Future Work

Identifying geometallurgical clusters or domains in mining applications is very important not just to characterise geology and geochemistry, but also to assist in choosing optimal processing routes for parcels of ore with different properties. Geometallurgy is increasingly incorporating more information and more variables, which makes it more difficult to find useful cluster structures for mine planning purposes.

In this paper, the difficulty of traditional clustering methods is demonstrated when dealing with multivariate scenarios in which the cluster structures depend on different attributes, as is commonly the case in practice. A new clustering method is proposed which is based on fuzzy clustering but incorporates additional valuable characteristics such as feature selection, spatial correction and the flexibility of including expert knowledge. Expert knowledge in the proposed method can be incorporated through an appropriate distance definition (categorical or targeted distances) and forcing a specific weight to a particular attribute.

Three case studies were presented to illustrate the application of SWFC. The first case study was explicitly designed to construct clusters that depend on different subsets of attributes. While the traditional methods fail to discover the true clusters, WFC and SWFC can readily find the designed cluster structure and SWFC also constructs well-connected clusters by incorporating spatial information. The second case study was used to illustrate graphically the effectiveness of SWFC using a two-dimensional synthetic geometallurgical

model. The clusters found by SWFC are spatially well-connected and the most compact. Finally, SWFC was applied to a complete synthetic geometallurgical block model to demonstrate its capability and flexibility in building clusters, which in this case are geometallurgical domains for iron recovery. By imposing a targeted distance for iron recovery and weight, SWFC can find the relevant secondary attributes that control the cluster structures.

In summary, SWFC has been demonstrated to be capable of defining meaningful geometallurgical domains for different application scales, based either on samples or complete block models.

In future research, the geometallurgical uncertainty will also taken into account for the clustering method. Uncertainty can be introduced by generating many realisations of the block model. The SWFC method can then include the distances between realisations to account for uncertainty. Also, the interaction potential could be reformulated to incorporate some form of multivariate spatial correlation, such as semivariogram or correlogram, instead of the Potts model. It is also necessary to investigate an optimisation formulation that can include the minimisation of compactness, maximisation of separation, feature selection and spatial correction all within a single integrated step.

Acknowledgements

This work was supported with the supercomputing resources provided by the Phoenix HPC service at the University of Adelaide. The first author acknowledges the support of CONICYT in the form of a Becas Chile scholarship.

References

Allard, D. and Guillot, G. (2000), Clustering geostatistical data, *in* 'Proceedings of the sixth geostatistical conference'.

Ambroise, C., Dang, M. and Govaert, G. (1996), Clustering of Spatial Data by the EM Algorithm, *in* 'Geostatistics for environmental applications', Lisboa, pp. 493–504.

Bourgault, G., Marcotte, D. and Legendre, P. (1992), 'The multivariate (co) variogram as a spatial weighting function in classification methods', *Mathematical Geology* **24**(5), 463–478.

Boykov, Y. and Veksler, O. (2006), 'Graph cuts in vision and graphics: Theories and applications', *Handbook of Mathematical Models in Computer Vision* pp. 79–96.

Carlo, G., Paolo, G. and Roberto, P. (2017), 'Spatial clustering of curves with an application of satellite data', *Spatial statistics* **20**, 110–124.

Coward, S. and Dowd, P. (2015), Geometallurgical models for the quantification of uncertainty in mining project value chains, *in* '37th APCOM Conference', Alaska, pp. 360–369.

Coward, S., Dowd, P. A. and Vann, J. (2013), Value chain modelling to evaluate geometallurgical recovery factors, *in* '36th APCOM Conference', Fundação Luiz Englert, Brazil; ISBN 978-85-61155-02-5, Porto Alegre, Brazil, pp. 288–289.

Coward, S., Vann, J., Dunham, S. and Stewart, M. (2009), The Primary-Response Framework for Geometallurgical Variables, *in* 'Seventh International Mining Geology Conference', pp. 109–113.

Davies, D. L. and Bouldin, D. W. (1979), 'A cluster separation measure.', *IEEE transactions on pattern analysis and machine intelligence* **1**(2), 224–227.

Deutsch, C. V. (1998), 'Cleaning categorical variable (lithofacies) realizations with maximum a-posteriori selection', *Computers & Geosciences* **24**(6), 551–562.

Ding, C. and He, X. (2004), K-means clustering via principal component analysis, *in* 'Twenty first international conference on Machine learning', ACM, Banff, Canada, p. 29.

Dunham, S. and Vann, J. (2007), Geometallurgy, Geostatistics and Project Value - Does Your Block Model Tell You What You Need to Know?, *in* 'Project Evaluation Conference', Melbourne, pp. 189–196.

Dunn, J. C. (1973), 'A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters', *Cybernetics* **3**, 32–57.

Fouedjio, F. (2016), A Clustering Approach for Discovering Intrinsic Clusters in Multivariate Geostatistical Data, *in* 'Machine Learning and Data Mining in Pattern Recognition', Vol. 9729, pp. 491–500.

Friedman, J. H. and Meulman, J. J. (2004), 'Clustering objects on subsets of attributes', *Journal of the Royal Statistical Society. Series B: Statistical Methodology* **66**(4), 815–839.

Garrido, M., Sepúlveda, E. and Navarro, F. (2017), Optimization of planning and scheduling of ore body with open pit extraction considering homogeneity in clays as geometallurgical variables, *in* '5th International Seminar on Mine Planning', Santiago, pp. 294–300.

Goodfellow, R. and Dimitrakopoulos, R. (2017), 'Simultaneous stochastic optimization of mining complexes and mineral value chains', *Mathematical Geosciences* **49**(3), 341–360.

Guillot, G., Kan-King-Yu, D., Michelin, J. and Huet, P. (2006), 'Inference of a hidden spatial tessellation from multivariate data: application to the delineation of homogeneous regions in an agricultural field', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **55**(3), 407–430.

Hoal, K. O., Woodhead, J. and Smith, K. S. (2013), The Importance of Mineralogical Input into Geometallurgy Programs, *in* 'The second AUSIMM international geometallurgy conference', number October, Brisbane, QLD, pp. 17–25.

Hunt, J., Berry, R., Bradshaw, D., Triffett, B. and Walters, S. (2014), 'Development of recovery domains: Examples from the Prominent Hill IOCG deposit, Australia', *Minerals Engineering* **64**, 7–14.

Hunt, J., Kojovic, T. and Berry, R. (2013), Estimating Comminution Indices from Ore Mineralogy , Chemistry and Drill Core Logging, *in* 'The Second AusIMM International Geometallurgy Conference', number October, Brisbane, QLD, pp. 173–176.

Keeney, L. and Walters, S. G. (2011), A Methodology for Geometallurgical Mapping and Orebody Modelling, *in* 'The First AusIMM International Geometallurgy Conference', number September, Brisbane, QLD, pp. 217–225.

Leichliter, S. and Larson, D. (2013), 'Geometallurgy for two recovery process operations at Cripple Creek & Victor gold mine', *Mining Engineering* (January), 29–33.

Lishchuk, V. (2016), Geometallurgical Programs - Critical Evaluation of Applied Methods and Techniques, Licentiate thesis, Lulea University of Technology.

López, A. A. and Malpica, J. A. (2008), 'High resolution satellite classification with graph cut algorithms', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **5359 LNCS**(PART 2), 105–112.

Luchi, D., Rodrigues, A., Varejão, F. M. and Santos, W. (2016), A genetic algorithm approach for clustering large data sets, *in* 'Tools with Artificial Intelligence (ICTAI), 2016 IEEE 28th International Conference on', IEEE, pp. 570–576.

Lund, C., Lamberg, P. and Lindberg, T. (2015), 'Development of a geometallurgical framework to quantify mineral textures for process prediction', *Minerals Engineering* **82**, 61–77.

Maulik, U. and Bandyopadhyay, S. (2000), 'Genetic algorithm-based clustering technique', *Pattern Recognition* **33**, 1455–1465.

Nanda, S. J. and Panda, G. (2014), 'A survey on nature inspired metaheuristic algorithms for partitional clustering', *Swarm and Evolutionary Computation* **16**, 1–18.

Napier-Munn, T. J., Morrell, S., Morrison, R. D. and Kojovic, T. (1996), *Mineral comminution circuits: their operation and optimisation*, Vol. 2, Julius Kruttschnitt Mineral Research Centre, University of Queensland.

Newton, M. and Graham, J. (2011), Spatial modelling and optimisation of geometallurgical indices, *in* 'The First AusIMM International Geometallurgy Conference', Brisbane, pp. 247–261.

Nguyen, K. and Keeney, L. (2014), Spatial domaining of highly variable continuous geometallurgical data, *in* 'XXVII International Mineral Processing Congress - IMPC 2014', Gecamin Digital Publications, pp. 116–127.

Oliver, M. A. and Webster, R. (1989), 'A geostatistical basis for spatial weighting in multivariate classification', *Mathematical Geology* **21**(1), 15–35.

Pal, N. R. and Bezdek, J. C. (1995), 'On cluster validity for the fuzzy c-means model', *IEEE Transactions on Fuzzy systems* **3**(3), 370–379.

Ren, M., Liu, P., Wang, Z. and Yi, J. (2016), 'A Self-Adaptive Fuzzy c-Means Algorithm for Determining the Optimal Number of Clusters', *Computational Intelligence and Neuroscience* **2016**(1).

Romary, T., Ors, F., Rivoirard, J. and Deraisme, J. (2015), 'Unsupervised classification of multivariate geostatistical data: Two algorithms', *Computers & Geosciences* **85**, 96–103.

Rousseeuw, P. J. (1987), 'Silhouettes: A graphical aid to the interpretation and validation of cluster analysis', *Journal of Computational and Applied Mathematics* **20**(C), 53–65.

Ruspini, E. H. (1969), 'A new approach to clustering', *Information and control* **15**(1), 22–32.

Schnetzler, E. (1994), Visualization and cleaning of pixel-based images, PhD thesis, Stanford University.

Tarabalka, Y. and Charpiat, G. (2013), 'A graph-cut-based method for spatio-temporal segmentation of fire from satellite observations', *International Geoscience and Remote Sensing Symposium (IGARSS)* **2**, 3494–3497.

Tarabalka, Y. and Rana, A. (2014), Graph-Cut-Based Model for Spectral-Spatial Classification of Hyperspectral Images, *in* 'International Geoscience and Remote Sensing Symposium', Quebec, Canada.

Wang, Y., Song, H. and Zhang, Y. (2016), 'Spectral-Spatial Classification of Hyperspectral Images Using Joint Bilateral Filter and Graph Cut Based Model', *Remote Sensing* **8**(9), 748.

Whitley, D. (1994), 'A genetic algorithm tutorial', *Statistics and computing* **4**(2), 65–85.

Wold, S., Esbensen, K. and Geladi, P. (1987), 'Principal component analysis', *Chemometrics and intelligent laboratory systems* **2**(1-3), 37–52.