

# Cell Based Associations: A procedure for considering scarce and mixed mineral occurrences in predictive mapping

Bruno Tourlière, Evren Pakyuz-Charrier, Daniel Cassard, Luc Barbanson,

Charles Gumiaux

### ▶ To cite this version:

Bruno Tourlière, Evren Pakyuz-Charrier, Daniel Cassard, Luc Barbanson, Charles Gumiaux. Cell Based Associations: A procedure for considering scarce and mixed mineral occurrences in predictive mapping. Computers & Geosciences, 2015, 78, pp.53-62. 10.1016/j.cageo.2015.01.012 . insu-01164639

## HAL Id: insu-01164639 https://insu.hal.science/insu-01164639

Submitted on 18 Jun2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

#### 1 CELL BASED ASSOCIATIONS, A PROCEDURE FOR CONSIDERING SCARCE AND MIXED 2 MINERAL OCCURRENCES IN PREDICTIVE MAPPING

Bruno Tourlière\*<sup>1</sup>, Evren Pakyuz-Charrier<sup>2</sup>, Daniel Cassard<sup>1</sup>, Luc Barbanson<sup>3</sup> and Charles
Gumiaux<sup>3</sup>

5

6 <sup>1</sup> Bureau de Recherches Géologiques et Minières (BRGM), Orléans, France

- 7 <sup>2</sup>Center for Exploration Targeting (University of Western Australia), Perth, Australia
- 8 <sup>3</sup>Institut des Sciences de la Terre d'Orléans (Univ. Orléans, CNRS, BRGM), Orléans, France

9
\*(Corresponding Author: <u>b.tourliere@brgm.fr</u>) BRGM
3 avenue Claude Guillemin,
BP 36009
45060 Orléans Cedex 02
France.
Phone: +33 2 38 64 33 78
17

18

19

#### Abstract

20 Cell Based Association is an innovative mineral favorability procedure designed to answer special 21 needs of the mining industry in data wise critical situations where usual favorability methods may 22 not yield satisfactory results. Those situations relate to input data quality (e.g.: clustered points, 23 mixed and scarce data, approximate location) or some assumptions that are considered 24 unreasonable (e.g.: map areas relevance, conditional independence).

The principle of CBA consists in replacing polygons of geological units with a square cell grid (hence the 'cell-based'). Each cell contains a range of units ('association') that are binary coded in terms of their presence (1) or absence (0) within study area. The loss of resolution inherent to this procedure is compensated by the enriched information contained in each cell owing to the notion of (lithological) association.

Lithological associations are considered as binary spectra and as such are classified using
 Ascendant Hierarchical Clustering (AHC) thus obtaining a synthetic map of lithological
 associations. The prospectivity map shows as favorable the cells of the same AHC classes that the
 ones including mineral occurrences.

It was observed that CBA can distinguish between different ore deposit varieties from a blended
 mineral occurrences data set. CBA can theoretically include any spatialized data (e.g.: geophysics,

- structural data) as an extra variable to specify classification and narrow favourable areas. Doing so
  would make it an independent favorability mapping procedure and is still under development.
- 38 Cell size in a grid is a critical parameter of the procedure; it must be compatible with the 39 looked-for phenomena and should have a sufficient lithological variability.
- In addition to its use for producing favorability maps, a CBA-derived map could help in
  understanding the background information contained in geological maps. CBA can also be applied
  to other fields, such as agriculture and urban planning.
- 43
- 44 Key words: Favorability mapping, grid cell, classification, attribute association
- 45

#### 46 1. Introduction

47 Mineral favorability is a branch of predictive analytics that focuses on designing statistical methods
48 to point out favourable zones in terms of mining potential.

From Carranza (2009) : "The term mineral prospectivity (...) is similar to the terms mineral potential
and mineral favourability, both of which imply the chance or likelihood that mineral deposits of the
type sought are contained in a region or district under investigation. (...)

52 Modeling of mineral prospectivity is a regional- or district scale mapping activity, whether field 53 based or GIS based, which aims to delineate prospective areas for further exploration at the next 54 higher scales of mapping. Notwithstanding the scale of mapping, mineral prospectivity is related to 55 the degrees of presence and degrees of importance of individual pieces of spatial evidence of 56 occurrence of mineral deposits of the type sought. That is, in a region or district under 57 investigation, if there are more important pieces of spatial evidence in an area than in another 58 area, then the former is considered to have higher prospectivity than the latter".

59 The most basic case relates to measuring the statistical link between a mineral occurrence (MO) 60 data set (points) and a geological map (polygons). This can roughly be translated as finding what 61 makes places where mineral occurrences were observed distinguishable and trying to target 62 similar locations.

63 End result is often a favorability map that displays the odds of finding something of interest.

64 Cell Based Association (CBA) is meant to be used during the strategic stages of exploration when 65 available data often does not meet the requirements (independence, representativeness of MO) of 66 usual favorability methods such as Weight of Evidence, Fuzzy Logic, Artificial Neural Networks to 67 be used effectively (Bonham-Carter, 1984, Carranza, 2011, Schaeben, 2011)).

68 Usually the requirements are not met due to poor quality of the mineral occurrence data set.

Indeed, the mineral occurrence data set is normally considered as a unique representative sample of unique elements of a single population (or at least a known number of different populations). This is not always the case as mineral occurrences data sets are most of the time classified according to the (potential) resource observed (e.g.: zinc mineral occurrences) which can lead to mixing different types of mineral occurrences that relate to different genesis processes (hence different lithological environments) into a single mineral occurrence data set. Ultimately this means the user will obtain a mixed favorability map that, at best, roughly ranks the different types of mineral occurrences on a single scale that may make no sense because each type should be analysed separately from the others or, at worst, a global fuzzy favorability map that makes even less sense.

Also, when the mineral occurrences data set is too scarce it is impossible to assess if it is representative which is one major assumption/requirement of most favorability methods.

Lastly, considering each mineral occurrence as individual leads to an overestimate of the importance of clustered areas that in fact describe a single object geologically wise which is even more critical in the case of mixed mineral occurrence data sets.

To address these problems, we propose a procedure called 'Cell-Based Associations' (CBA). Using a square grid (of appropriate cell size fixed by user) in which each cell records the presence or absence of lithological units from the geological map and a hybrid driven (combining data and user driven approaches) classification of all cells, this procedure allows to generate a synthetic map that can be coupled to the mineral occurrence data set. CBA thus allows better characterization and sorting of a number of different complex environments associated with the looked for phenomena.

91

A technique consisting in applying a grid with included attributes over a study area was previously described by McCammon et al. (1983) for generating predictive maps, but rather than using their Characteristic Analysis method of data processing of a logical type, CBA uses statistical ranking of coded cells for identifying the different associations/environments contained in the map.

96

97

#### 2. The principles of CBA

The case study presented here in for illustrating the principles of the CBA procedure is a mapping of mineral favorability, where the points are mineralized occurrences (MO) and the polygons are geological units (lithologies). The artificial map in Figure 1 shows the type of document used by geologists when looking for prospective areas that are *a-priori* favourable to the presence of mineralization. On this map, mineral occurrences are present in the B, D and E units. Occurrence 103 "OCC-4" is located in unit C, but its position very close to the limit between units E and D might be104 imprecise.

The CBA procedure was applied to that dataset. First, the optimal cell size of the grid had to be defined. This size relies on two main constraints: 1) it should be coherent with the size of the studied phenomena (Carranza, 2009), and 2) it must allow a sufficient variability of the geological units. Too small cells will contain only one unit, but too large ones will contain too many (if not all) units.

110 The relationship between the variability inside the grid cell and its size has been studied in the dataset in order to define the optimized size for a cell. Variability has been estimated using 2 111 112 parameters: the average number of geological units per cell and the number of lithological spectra 113 (or associations, see definition below) generated by the 'cut-out' of the map by grids with different 114 cell sizes (Table 1). The graphic representation (Figure 2) of this result shows that the optimal size 115 of a cell ranges from 7 to 12 km (side length of square cells), which corresponds to a number of 116 spectra comprised between 15 and 17 (for a theoretical maximum of 36 [i.e. 2<sup>5</sup>] possible 117 combinations, 17 being the maximum number of combinations present in the map) and an average 118 number of lithologies per cell ranging from 1.8 to 2.2 (Table 1). In this example, the cell size was 119 arbitrarily set at 10 km that is between 7 and 12km.

120 Once the cell size has been defined by user, the grid is generated and superimposed on the 121 geological map (Figure 3). All cells have for attributes all the units represented in the map. The 122 geological information contained in each cell is coded as presence = 1 or absence = 0 of each 123 geological contained in the map (Figure 4 and Table 2). Each cell is then characterized by a 124 binary-type attribute spectrum showing the presence/absence of each unit within it. There is thus a 125 transition from a discrete categorical variable (rock type) to a set of binary variables describing a 126 lithological spectrum or - more generally - an attribute spectrum. This operation does not consider 127 the relative or absolute surface of the geological units, but only their presence or absence. This 128 reasoning is based on the idea that the projected surface of a unit on a geological map is irrelevant 129 for determining its real importance as a small area unit may be the critical factor that determines a 130 particular lithological environment (e.g.: dyke fields) or may be the end of a much larger body (e.g.: 131 plutonic intrusion) it also may be partially masked surficial deposits. It seems preferable to 5

emphasize the presence of poorly represented units than to lose that information because mining favorability is ultimately a 3D problem. The processing of each unit as an independent binary variable assumes that a geological map is a combination of exclusive events, in other words: at every location of the map a single lithology may be observed.

This technique avoids the loss of attribute information (Bai et al., 2011) and renders it directly usable for multivariate analysis. The maximum number of possible combinations is 2<sup>n</sup>, where n is the number of attributes, here geological units. If the number of attributes is large, the individual processing of all present combinations becomes impossible, implying the need to use a method that allows grouping the attribute spectra into representative classes of existing combinations.

Same processing should have been possible in raster mode but that implies a multiplication of layers (one per unit) and a possible loss of information for small surface covering units. The creation of a single table containing all the attributes is less time consuming, makes it easier to treat and is coherent with the notion of spectrum created by concatenation of the attributes.

145 In the present work, the hybrid driven Ascendant Hierarchical Clustering (AHC) method has been146 used to classify the associations within the cells.

147

157

#### 148 3. The principles of AHC

AHC uses a system of aggregation by pairs of elements according to: 1) their proximity based on a
dissimilarity function that measures the multivariate statistical distance between the elements; and
2) an aggregation function that groups those that are iteratively the closest (Rolet and Seguin,
1986)

In the present case, the calculation relies on a dissimilarity index based on "percent
disagreement" (Equation 1) and an aggregation function based on the reciprocal neighbours
method (Rham, 1980) after centring and reduction of the data.

156 The dissimilarity index is as follows:

$$dP(I_i, I_j) = \frac{number \ of \ (x_{ik} \neq x_{jk})}{K}$$
(1)

158 Where i and j are the identifiers of the compared elements I, k is the variable (geological unit) 159 identifier, x is the variable value, and K the number of variables (number of units). This type of dissimilarity function is used for variables of the categorical/discrete type, and is the most suitablefor evaluating the distance between elements.

However, in our specific case it is also possible to use the usual Euclidean distance(Equation 2) as it is easier to apply instead of the dissimilarity function.

164 
$$dE(I_i, I_j) = \sqrt{\sum_k (x_{ik} - x_{jk})^2}$$
(2)

165 For this, equations 3.1 to 3.2 are used. As x can only take the values 0 or 1, we have:

166

167 
$$\sqrt{\sum_{k} (x_{ik} - x_{jk})^2} = number \ of (x_{ik} \neq x_{jk}) \ \forall x = \{0,1\}$$
 (3.1)

168 Where

169  $dE(I_i, I_j) = dP(I_i, I_j) * K \forall x = \{0,1\}$  (3.2) 170 Where i and j are the identifiers of the compared elements I (cells), k the variable (unit) identifier, x

the value of the variable, and K the number of variables (units).

The choice of Euclidian distance is arbitrary as it is the most "intuitive" but other kind of distance could be used as Manhattan distance for example. The aggregation factor for classes is minimum in-class variance. Varying those parameters could lead to significantly different results in classification that have not been tested.

176 AHC produces a binary hierarchical grouping of the attributes characterizing each element in 177 a class. The quality of such grouping progressively degrades with a diminishing number of classes 178 (Hastie et al., 2009). This is generally expressed as a dendrogram (Figure 5) showing intermediate 179 grouping results and their hierarchical relations, which allows to select the number of classes to 180 represent lithological environments. That number of classes is based on the morphology of the 181 dendrogram and has to be balanced by the user between a low level of aggregation (numerous 182 classes) that will give a fine description of lithological associations with the risk to describe very 183 local associations and a high level of aggregation (few classes) that will be too comprehensive and 184 thus not applicable to the searched result. Once the level of cutting chosen by the user, an attribute is added to the table to store the AHC class attribute (Table 3) in order to mark the classes 185 186 attributed to each cell, each class corresponding to a group of associated units.

187 It is also possible to calculate classification quality for each cell and each level. This quality 188 criteria corresponds to the Euclidean distance between the cell and the barycentre of the class to 189 which the cell belongs. The quality value associated with each cell ('Qual' columns in Table 3) thus 190 shows the strength of the link between the cell and its attributed class. This parameter ideally 191 should be 0, but this is conceptually impossible when grouping the lithological associations by 192 using AHC. This quality parameter is useful for comparing different classification levels (number of 193 classes) of the same dataset.

194

#### 195 4. Results

Figure 6 shows a cartographic plot of classes and thematic analysis for a 10 class cut-off ofthe dendrogram.

198 Cells of the same class do not exactly present the same lithological spectrum. This is the 199 case, for instance, for the four cells of class 7 (no filling) from which 2 are of type 00110 and 2 of 200 type 00111. The grouping of class 7 was thus based on a common spectrum of type 0011x (x 201 being 1 or 0). Class 5 (horizontal and vertical cross-hatching) shows a similar case with a spectrum 202 of the 01x11 type and class 10 (left inclined wide-spaced hatching) with a spectrum of the 1x101 203 type. This clearly illustrates that classification quality decreases along with the number of classes. 204 Nevertheless, the fundamental characteristics of the associations forming the classes are 205 conserved.

Producing the favorability map (Figure 7) consists in (i) selecting the classes of the cells that contain occurrences, and (ii) extracting all the cells showing one of the favourable classes. In this example, cells of class 6 (association of units D and E), 7 (units C, D and E) and 8 (units C and D) are associated to mineral occurrences and are, thus corresponding to geological environments favourable for the presence of mineralisation. In this example, other classes do not contain any occurrence and are thus considered as non-favourable

#### 213 5. Discussion

The CBA procedure goes counter to the natural tendency of trying to obtain the finest possible resolution. CBA, on the contrary, aims to define an **association** of elementary signatures present in a larger area, rather than looking for the smallest elementary signature possible. This led to the idea of considering a larger cell that integrates an environment defined as the presence or absence of all lithological units in the study area.

Moreover, this type of approach mimics the thought process of geoscientists as it associates phenomena with a multifactorial context rather than a single criterion. In mineral exploration, for instance, the exploration geologist will favour a complex geological setting over a single unit (e.g. Billa et al., 2004; Roy et al., 2006; Cassard et al., 2008).

Cell size is a critical parameter to the procedure and must comply with two major constrains: coherence with the size of the phenomena of interest (Carranza, 2009), and sufficient variability of the attributes per cell. For correct operation of the procedure, the cells should neither be too small (mono-attribute cells that pixelate the map), nor too large (cells containing most or all attributes). It also is the cell size parameter that allows CBA to deal with over and under sampling issues of the mineral occurrence data set.

229 As discussed in section 2, given that the MO display minute to no clustering effect, a grid of 230 suitably-sized meshes can be calculated, based on the variability of the lithology parameter 231 estimated either on the average number of unit contained in grid cells or on the number of unique 232 lithological associations 'generated' by this grid. Conversely, statistics based on MO spatial 233 behaviour (i.e. "kernel density method for univariate data" (Silverman, 1986) or "pair correlation 234 function g(r)" (Van Leeuwen et al, 1959)) may be used as indicators to optimize cell size 235 provided that they allow for sufficient variability of the units in the cells. As a last resort, the size of 236 the cells is a decision of the user who must reconcile a size compatible with the map(s) scale and 237 the searched for events. CBA, as described in this paper, is based on a regular grid, but it will 238 work just as well with cells of variable size adapted to the influence of the units, so as to limit the 239 number of mono-attribute cells.

The idea of 'specific surface' of a unit, whether by cell or within the study area, was considered as irrelevant as this is a variable induced by the intersection between topography and units that may be partly covered by overlapping units. It is supposed that the area of a mapped unit is not a critical criterion and that mineral deposits are likely to be linked to "rare" units as much as more frequent ones. This is why only existence is taken into account (0 or 1) which in return favours presence over absence: it highlights details over global trends because we consider that only the proximity to observed MO is important in this case.

247 CBA uses positive reasoning assuming that "non-links" are impossible to demonstrate. In reality, it 248 is impossible to conclude that the absence of a certain unit reduces the odds of finding a MO. 249 Geological maps and geologists observation are never perfect and, having no MO in a certain 250 lithological setting does not make it unfavourable, it makes it non-favourable. As such, positive 251 reasoning (observations lead to consider the environment in which they are found as favourable) 252 should be preferred over negative reasoning (environments in which no observation are found are 253 unfavourable). Ultimately, this relates to the way field geologists work, describing the environment where the presence of MO proves that it is a favourable one, whatever are the respective surfaces 254 255 of the units on the geological map.

For other subjects, however, such as agricultural or town-planning studies, this surface factor may be introduced as the respective percentage covered by each attribute in the cell, with 0 indicating its absence. This type of approach has not yet been tested but could be an interesting development of this work.

260 For cell ranking, the Ascendant Hierarchical Clustering (AHC) has been selected because, in 261 addition to being a hybrid driven classification function, it produces a hierarchical tree 262 (dendrogram) that shows the relations between the different classes. The number of classes is a 263 user decision helped by the shape of the branches of the dendrogram and the level of details he 264 requires. Based on the cut-off level (i.e., the number of classes) of the tree, the classes will be 265 more or less informative, thus more or less focused on relevant lithological spectrum. For rather 266 low cut-off levels (i.e., few classes), AHC commonly produces a "garbage" class that covers all 267 cells poorly classified for this level. In the case of several occurrences falling within this class, it is 268 however always possible to extract the related cells for an independent re-classification.

Other methods of multivariate classifications – as, for example, IsoData, K-means clustering - may
yield different results. They are still to be tried.

The result obtained from such data processing is a new synthetic map of the assemblage of classified cells showing the different groups of associations in the study area.

Family of cells containing MO will be regarded as favourable and the capability of the procedure to describe the contained association of lithologies is helpful for exploration teams to point out the cells showing the same environment than the ones with MO. Even in that case, results must be interpreted to check their relevance.

277 CBA can induce a bias for points located near a cell boundary that might correspond to a 278 different lithological spectrum from that effectively favourable. This drawback is limited if the 279 number of occurrences is sufficiently large. However, if too many showings lie close to cell 280 boundaries, it may be necessary to shift the grid origin or change the cell size (Pakyuz-Charrier, 281 2013).

According to points presented and discussed here above, predictive results of the CBA procedure could possibly be further improved by:

- Ignoring specific units (such as alluvium) that clearly are irrelevant to the phenomenon
   looked for. Classification will then be based on a partial lithological spectrum, which might
   improve the quality of the result.
- Introducing relevant parameters into the cells that come from other thematic layers, such
  as the presence/absence of faults or geophysical anomalies. Such additional attributes
  enrich the information contained in the spectrum and can be used for classification, which
  then becomes multi-thematic. The approach would then shift from a "bivariate-data driven"
  to a "multivariate-data driven" method (Carranza, 2011) that resembles McCammon's
  "characteristic analysis" (1983).

Producing more relevant classes by using other classification methods.

In supervised mode, CBA can be used by processing only cells that contain occurrences, provided they are sufficiently numerous for statistical processing. In this case, only the lithological spectra associated to mineralization would be classified. A predictive map would then be produced by selecting all cells with comparable spectra to those deriving from the classification. This approach (Cassard et al., 2012; Tourlière et al., 2012) is not illustrated here. In simple cases, it
may allow dispensing with classification, becoming comparable to the "characteristic analysis"
approach (McCammon, 1983).

301

In addition the synthetic map produced by spectra classification with the CBA procedure can help
in (i) interpreting a thematic map and (ii) highlighting the presence of distant but similar association
areas.

305

#### 306 6. Conclusions

The Cell-Based Associations (CBA) procedure was developed for overcoming the limitationscaused by low quality input data sets often used in early stages of mining exploration.

Application of a cell grid (*'cell based'*) of suitable size (fixed by the user) over a study area and the coding of the units found in each cell as 1 (presence) or 0 (absence), allows obtaining a binary attribute spectrum (*'association'*) that can be statistically classified.

312 Decrease in resolution of the original data induced by the procedure is compensated by the 313 concept of association that, after classification, allows the extraction and regrouping of more 314 complex environments and their cartography as a synthetic map.

This approach better considers the complexity of geological phenomena favouring the presence of mineralization and is able to distinguish different types of favourable lithological environments. It is also possible to integrate data from other thematic layers (e.g., tectonic structures, geophysical properties...) into the cells which enriches the attribute spectrum, thus opens the possibility of creating multi-thematic synthetic maps.

As presented in that paper, CBA can be a guide to mining exploration essentially at strategic scale when the mining occurrences are scarce or clustered and units areas are irrelevant or small areas are suspected to be important.

The hybrid driven statistical method (Ascendant Hierarchical Clustering) has been used in this study for calculating the grouping of geological environments by cell but other methods of classification can be relevant too. CBA can also be used as a dataset for predictive methods of the supervised type by using cells containing showings as learning sets.

328 Apart from any search for predictions the CBA procedure produces a synthetic thematic 329 representation that may help in the interpretation of complex and/or very large scale maps.

# 330331 Acknowledgements

This work was initiated as part of the EU-FP7 funded "ProMine" project (grant agreement n° 228559) and finalized in the frame of the "Favorability" project of BRGM's "RSM 16" Scientific Programme.

- 335 Many thanks to the two anonymous reviewers who guided us into redirecting this paper to make it
- 336 more relevant.
- 337

#### 338 References

#### 339

- Bai, Y., Liao, S., Jiulin, S., 2011. Scale effect and methods for accuracy evaluation of attribute information loss in rasterization. Geographical Science, pp. 1089-1100.
- Billa M., Cassard D., Lips A.L.W., Bouchot V., Tourlière B., Stein G. and Guillou-Frottier L., 2004.
  Predicting gold-rich epithermal and porphyry systems in the central Andes with a continental-scale
  metallogenic GIS. Ore Geology Reviews, 25, pp. 39-67.
- Bonham-Carter, G.F., 1994. Geographic Information Systems for Geoscientists: Modelling with GIS: Pergamon, Oxford, 398 pp.
- Carranza, E.J.M., 2009. Objective selection of suitable unit cellsize in data-driven modeling of mineral prospectivity: Computers & Geosciences 35, pp. 2032–2046.
- Carranza, E.J.M., 2011. Geocomputation of mineral exploration targets. Computers & Geosciences
   37, pp. 1907-1916.
- Cassard D., Bertrand G., Monthel J., Gloaguen E., Melleton J., Husson Y., Billa M. et Tourlière B.,
  2012. Réévaluation du potentiel français en ressources minérales : élaboration d'une méthodologie
  et application à une zone d'étude test (sud-ouest du Massif central). Final report. BRGM/RP61745-FR, 126 p. Public document in French downloadable at: HYPERLINK
- 355 "http://infoterre.brgm.fr/rapports/RP-61745-FR.pdf" <u>http://infoterre.brgm.fr/rapports/RP-61745-</u> 356 <u>FR.pdf</u>
- Cassard D., Billa M., Lambert A., Picot J.-C., Husson Y., Lasserre J.-L., Delor C., 2008. Gold
   predictivity mapping in French Guiana using an expert-guided data-driven approach based on a
   regional-scale GIS. Ore Geology Reviews, 34, pp. 471-500.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining,
   Inference, and Prediction. Stanford: Springer.
- McCammon R.B., Botbol J.M., Sinding-Larsen R., Bowen R.W., 1983 HYPERLINK
   "https://www.researchgate.net/publication/226193164\_Characteristic\_analysis1981\_Final\_program
   and\_a\_possible\_discovery?ev=prf\_pub"
   Characteristic Analysis-1981: Final Program and a
   Possible Discovery. Mathematical Geology 15(1), pp.59-83.
- Pakyuz-Charrier E., 2013. Cell Based Associations: An alternative to Weight of Evidence using cell
   based lithological associations. PhD Thesis Univ. Orleans, 63 pp.
- 368 Rham, C. D., 1980. La classification ascendante hiérarchique ascendante selon la méthode des 369 voisins réciproques. In *Les cahiers de l'analyse des données*, pp. 135-144. Dunod.
- Rolet, P., Seguin, J.J., 1986a. Traitement de donnees multivariables (Application aux sciences de la Terre) Approche theorique, Tome 1 - Version 1.0. BRGM Report BRGM/86-DT-005-ISA, 120 pp.
   Public document in French downloadable at: HYPERLINK "http://infoterre.brgm.fr/rapports/86-DT-005-ISA.pdf" <u>http://infoterre.brgm.fr/rapports/86-DT-005-ISA.pdf</u>
- Rolet, P., Seguin, J.J., 1986b. Traitement de donnees multivariables (Application aux sciences de la Terre) Aspect pratique, Tome2 - Version 1.0. BRGM Report BRGM/86-DT-006-ISA, 62pp Public document in French downloadable at: HYPERLINK "http://infoterre.brgm.fr/rapports/86-DT-006-ISA.pdf" <u>http://infoterre.brgm.fr/rapports/86-DT-006-ISA.pdf</u>
- Roy R., Cassard D., Cobbold P.R., Rossello E.A., Billa M., Bailly L. and Lips A.L.W., 2006.
  Predictive mapping for copper-gold magmatic-hydrothermal systems in NW Argentina: use of a
  regional-scale GIS, application of an expert-guided data-driven approach, and comparison with
  results from a continental-scale GIS. Ore Geology Reviews, 29, pp. 260-286.
- Schaeben, H., 2011. Comparison of Mathematical Methods of Potential Modeling. Mathematical
   Geoscience, pp. 101-129.
- 384 Silverman, B. W. (1986). Density estimation for statistics and data analysis, Monographs on
- 385 Statistics and Applied Probability, London: Chapman and Hall/CRC. pp. 7-11
- Wan Leeuwen, J. M. J., Groeneveld, J., De Boer, J. (1959). New Method for the Calculation of the
   Pair Correlation Function. <u>Physica</u>, 25: pp. 792-208

Mis en forme : Anglais (États Unis)

Mis en forme : Espagnol (International)

- 391 Figures :
- 392 Figure 1: Geological map and mineral occurrences (theoretical example)
- 393 Figure 2: Optimal cell size range estimated by the number of different spectra generated by the
- 394 gridding
- 395 Figure 3: Grid mesh of 10km applied to the geological map of Figure 1 and numbering of cells
- Figure 4: Grid applied to the geological map of Figure 1 and binary coding of the lithologicalspectra of cels
- 398 Figure 5: Dendrogram of lithological associations; data are those of Figure 4; dotted lines indicate
- 399 cut-offs according to the desired number of classes (15 10, 7 and 5 classes)
- 400 Figure 6: Result of cell classification with AHC in the case of a cut-off at 10 classes. The binary
- 401 series correspond to the lithological spectra of the different cells (Figure 4)
- 402 Figure 7: Favorability map: cells of the same class as those containing occurrences, bold numbers
- 403 indicating the class of association.
- 404
- 405 Tables :
- Table 1: Relationships between the cell size and lithological variability. Grey area = optimal cell size range.
- Table 2: Attributes of the grid after extracting the lithological information. For each cell (line) a lithological spectrum is defined by the presence/absence of all lithologies on the map (columns)
- Table 3: Attributes of the grid after AHC. The different AHC columns correspond to different cut-off levels in the dendrogram. Whole numbers correspond to the lithological association for a given cell and for a set cut-off level (see Figure 5). 'Qual' measures the quality of each cell in terms of the dissimilarity between it and the multivariate centre of gravity of the class comprising the cell
- 414
- 415
- 416







423 Figure 1: Geological map and mineral occurrences (theoretical example)





Figure 2: Optimal cell size range estimated by the number of different spectra generated by the gridding



431 Figure 3: Grid mesh of 10 km applied to the geological map of Figure 1 and numbering of cells





433 Figure 4: Grid applied to the geological map of Figure 1 and binary coding of the lithological spectra of cells



436 Figure 5: Dendrogram of lithological associations; data are those of Figure 4; dotted lines indicate cut-offs

437 according to the desired number of classes (15, 10, 7 and 5 classes)



Figure 6: Result of cell classification with AHC in the case of a cut-off at 10 classes. The binary series correspond to the
 lithological spectra of the different cells (Figure 4)



Figure 7: Favorability map: cells of the same class as those containing occurrences, bold numbers indicating the class of association.

Number of cells	Cell size (km)	Average number of lithologies by cell	Number of Speciific lithologic spectra	Number of cells with occurrences
4	30	4.25	3	4
9	20	3.11	6	4
16	15	2.56	9	5
25	12	2.16	15	4
36	10	2.06	15	5
64	7.5	1.78	17	5
100	6	1.57	17	5
144	5	1.51	17	5
225	4	1.4	17	6
400	3	1.3	17	6
900	2	1.2	17	6
3600	1	1.1	17	6

456	able 1: Relationships between the cell size and the lithological variability. Grey area	=
457	ptimal cell size range.	

	Formations				Formations								
Cell#	А	в	с	D	Е	Lithological Spectrum	Cell#	А	в	с	D	Е	Lithological Spectrum
00	1	1	0	0	0	11000	30	0	0	1	0	1	00101
01	1	1	0	0	0	11000	31	0	0	1	0	1	00101
02	1	1	0	0	0	11000	32	0	0	1	1	1	00111
03	1	1	0	0	0	11000	33	0	0	1	1	0	00110
04	1	1	0	0	0	11000	34	0	0	1	1	1	00111
05	1	0	0	0	0	10000	35	0	0	1	0	1	00101
10	0	1	0	0	0	01000	40	1	0	1	0	1	10101
11	0	1	0	0	1	01001	41	1	1	1	0	1	11101
12	0	1	0	0	1	01001	42	0	1	1	1	1	01111
13	0	1	0	0	1	01001	43	0	0	1	1	0	00110
14	1	1	0	0	1	11001	44	0	0	1	0	0	00100
15	1	1	0	0	0	11000	45	0	0	1	0	0	00100
20	0	1	0	0	1	01001	50	1	0	0	0	0	10000
21	0	1	0	0	1	01001	51	1	1	0	0	0	11000
22	0	0	0	1	1	00011	52	0	1	1	0	0	01100
23	0	0	0	1	1	00011	53	0	1	1	0	0	01100
24	0	1	0	1	1	01011	54	0	0	1	0	0	00100
25	0	1	0	0	1	01001	55	0	0	1	0	0	00100

461 Table 2: Attributes of the grid after extracting the lithological information. For each cell (line)

462 a lithological spectrum is defined by the presence/absence of each lithology on the map

463 (columns)

464

Lithological spectrum AHC10 Qual10 Cell# ABCDE AHC15 Qual15 AHC7 Qual7 AHC5 Qual5 0--0 11000 11 0.00 8 0.00 6 0.44 4 0.44 0.44 0--1 11000 11 0.00 8 0.00 6 4 0.44 0.44 11000 11 8 0.00 0.44 4 0--2 0.00 6 11000 11 8 0.44 4 0.44 0--3 0.00 0.00 6 0--4 11000 11 0.00 8 0.00 6 0.44 4 0.44 0--5 9 10000 12 0.00 0.00 1.54 4 1.54 6 1--0 01000 3 0.00 3 1.32 2 1.32 1 2.34 1--1 01001 5 0.00 4 3 0.00 0.00 0.00 2 1--2 01001 5 0.00 4 0.00 3 0.00 2 0.00 5 4 2 1--3 01001 0.00 0.00 3 0.00 0.00 1--4 11001 13 0.00 10 1.48 7 1.48 5 1.48 1--5 11000 11 0.00 8 0.00 6 0.44 4 0.44 2--0 01001 5 0.00 4 0.00 3 0.00 2 0.00 0.00 2--1 5 4 0.00 2 0.00 01001 0.00 3 2--2 00011 8 0.00 6 0.00 4 1.11 3 1.42 2--3 00011 8 0.00 6 0.00 4 1.11 3 1.42 2--4 01011 7 0.00 5 0.99 4 1.11 3 2.00 5 4 0.00 0.00 2--5 3 0.00 2 01001 0.00 3--0 00101 1 0.00 1 0.00 1.13 1 1.52 1 3--1 00101 1 0.00 0.00 1 1.13 1.52 1 1 3--2 10 7 00111 0.00 0.99 5 0.99 3 1.02 3--3 00110 9 0.00 7 0.99 5 0.99 3 1.73 3--4 00111 10 0.00 7 0.99 5 0.99 3 1.02 3--5 00101 1 0.00 1 0.00 1 1.13 1 1.52 14 10 1.48 4--0 10101 0.00 1.48 7 5 1.48 15 0.00 0.94 4--1 11101 10 0.94 7 5 0.94 4--2 01111 6 0.00 5 0.99 4 1.79 3 1.74 4--3 00110 9 0.00 7 0.99 5 0.99 3 1.73 4--4 2 00100 2 0.00 0.00 0.85 0.86 1 1 0.00 4--5 00100 2 2 0.00 1 0.85 0.86 1 5--0 12 0.00 9 1.54 10000 0.00 6 4 1.54 5--1 11000 11 0.00 8 0.00 6 0.44 4 0.44 0.00 3 0.66 5--2 01100 4 2 0.66 1 1.52 5--3 01100 4 0.00 3 0.66 2 0.66 1.52 1 5--4 00100 2 0.00 2 0.00 0.85 0.86 1 1 00100 5--5 2 0.00 2 0.00 0.85 0.86 1 1

465

466 Table 3: Attributes of the grid after AHC. The different AHC columns correspond to different cut-off levels in the

467 dendrogram. Whole numbers correspond to the lithological association for a given cell and for a set cut-off level (see

468 Figure 5). 'Qual' measures the quality of each cell in terms of the dissimilarity between it and the multivariate centre of

469 gravity of the class comprising the cell