

UNIVERSITÉ D'ORLÉANS

École doctorale Energie, Matériaux, Science de la Terre et de l'Univers

Institut des Sciences de la Terre d'Orléans

THÈSE présentée par :

Alex VELLA

soutenue le : **14 Décembre 2022**

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Discipline/ Spécialité : **Sciences de la Terre / Métallogénie**

Highlighting mineralized geological environments through a new Data-driven predictive mapping approach

Mise en évidence des environnements géologiques minéralisés par une
nouvelle approche de cartographie prédictive Data-driven

THÈSE dirigée par :

M. STANISLAS Sizaret

Professeur, Université d'Orléans, ISTO

M. GUMIAUX Charles

Maître de Conférences, Université d'Orléans, ISTO

M. BERTRAND Guillaume

Chercheur senior, Bureau de Recherches Géologiques et
Minière - BRGM

RAPPORTEURS :

M. CARRANZA Emmanuel John

Professeur, Université de Free State

M. SADEGHI Martiya

Chercheur senior, Service de recherches géologiques de
Suède, SGU

JURY :

Mme VRAIN Christel

Professeur, Université d'Orléans, LIFO, **Présidente du jury**

M. CARRANZA Emmanuel John

Professeur, Université de Free State, Afrique du Sud

M. SADEGHI Martiya

Chercheur senior, Service de recherches géologiques de Suède, SGU

M. LIMA Alexandre

Professeur associé, Université de Porto, Portugal

M. STANISLAS Sizaret

Professeur, Université d'Orléans, ISTO

M. GUMIAUX Charles

Maître de Conférences, Université d'Orléans, ISTO

M. BERTRAND Guillaume

Chercheur senior, Bureau de Recherches Géologiques et Minières, BRGM

Remerciements

Ces trois années de thèse auront autant été un apprentissage qu’une épreuve à surmonter. Comme pour toute épreuve, on ne peut y arriver que si l’on est bien épaulé et conseillé. C’est pourquoi je tiens ici à exprimer ma gratitude envers ceux qui m’ont aidé à finaliser ce doctorat et à rédiger ce manuscrit.

Je tiens d’abord à remercier mon directeur de thèse, Stanislas Sizaret, pour m’avoir donné l’opportunité de réaliser cette thèse. Je tiens également à remercier mes co-directeurs de thèse, Charles Gumiaux et Guillaume Bertrand, pour leur soutien et leurs conseils tout au long de cette thèse, ainsi que pour m’avoir accueilli au sein de l’ISTO et du BRGM. Ils ont su rester présents tout du long, malgré ce qui a pu se passer dans le monde pendant cette thèse, et je leur en suis reconnaissant.

Je veux aussi remercier mes co-encadrants pour leur aide précieuse tout au long de cette thèse. Merci d’abord à Bruno Tourlière, pour m’avoir transmis ses connaissances sur les arcanes de la CBA. Merci à Vincent Labbé pour son aide dans l’apprentissage de l’énigmatique langage des machines. Merci à Eric Gloaguen pour son aide et ses retours sur l’antimoine et le Massif Armoricaïn, et pour m’avoir permis de faire partie du projet AUREOLE.

Je remercie chaleureusement l’équipe DGR/GEM ainsi que les membres de l’ISTO pour leur accueil. Je remercie Jérémy Rohmer, Yannick Branquet et Andreas Barth pour leurs précieux retours lors de la soutenance à mi-thèse. Leurs apports ont certainement contribué au résultat final de cette thèse. Merci à Aurélie Peyrefitte pour son aide lors de la campagne de terrain à Valongo, et pour les séances de dégustation de la gastronomie nord-portugaise. Merci aussi à Hugo Breuillard pour les discussions enrichissantes sur la CBA, le Random Forest et le Boosting.

Je tiens à remercier les membres du jury, Emmanuel John Carranza, Martiya Sadeghi, Christel Vrain et Alexandre Lima d’avoir accepté de consacrer de leur temps pour évaluer et porter un regard critique sur ce manuscrit et sur le travail que j’y présente.

Je remercie le gang des doctorants et je leur souhaite bonne chance pour la suite (ce sera bientôt votre tour !). Bon courage Julien, maintenant que tu as reconstruit tes machines jusqu’aux derniers capillaires, tu pourras voir tes belles tourmalines. Attention à ne pas trop dégazer Fabien, tu pourrais louper l’azote ! Fais gaffe à tes émanations d’arsenic Valentin et ne t’inquiète pas, on aura le temps de faire un ROOT. Je remercie aussi mes voisins de bureau : Hector et ses chocolateries, ainsi qu’Austin et son concours silencieux. Je remercie aussi mes amis d’enfance pour leur support psychologique. Merci Benoît pour tes pep-talks et ces longues discussions sur tout ce qui roule et canonne. Merci Julien, Rémi et Loïs pour ces soirées (et matinées) d’anthologies.

Enfin, je remercie ma famille pour leur support indéfectible et leurs nombreuses questions. Merci Simon de m’avoir supporté pendant toutes ces années. Merci Papa, Maman, de m’avoir permis d’en arriver là.

Je dédie cette thèse à ceux qui ne sont plus là ; Jean-Paul, Fabrice et Catherine.

Contents

I	General Introduction	17
1	Resources and Prospective mapping	19
1.1	Mineral Resources	19
1.1.1	Mineralization and metallogeny	19
1.1.2	Societal demand on mineral resources	20
1.2	Prospective mapping and mineral exploration	23
1.2.1	Mineral Prospectivity Mapping	23
1.2.2	Predictive models	25
1.2.3	Knowledge-driven approach	27
1.2.4	Data-driven approach	30
1.3	Developing a new approach to constrain relationships between mineralization and geological features	34
1.3.1	Context of this study	34
1.3.2	Aims of this study	35
2	Ibero-Armorican geological and metallogenic context	39
2.1	Orogen and mineralization	39
2.1.1	Orogenesis	39
2.1.2	Orogenic processes and mineralization	41
2.2	The Variscan Range	45
2.3	Armorican Massif	46
2.3.1	General setting	46
2.3.2	Geological units and related events	50
2.3.3	Main metallogenic resources and events	54
2.3.4	Antimony in the Armorican massif	57

2.4	Iberian Massif : Central Iberian Zone	59
2.4.1	Geological setting and characteristics	59
2.4.2	Main metallogenic resources	63
2.4.3	Antimony in the Iberian massif	66
II	Developments	69
3	Spatial analyses	71
3.1	Cell-based association method	71
3.1.1	Principles	71
3.1.2	Advantages	72
3.1.3	Critical parameters	75
3.2	Development of the DBA	78
3.2.1	Overcoming the limitation of the CBA	78
3.2.2	Disc-Based Association principles	81
3.2.3	Overlapping	84
3.2.4	Critical parameters	87
4	Machine learning for prospectivity mapping	91
4.1	Prospectivity mapping with CBA	91
4.1.1	The "exact cell signature" of the CBA method	91
4.1.2	CBA / AHC classification	92
4.1.3	CBA / Ranking	92
4.2	Random Forest-based prospectivity mapping applied on CBA and DBA matrices	96
4.2.1	Workflow of CBA and DBA matrices-based RF prospective mapping	96
4.2.2	Targetting of the RF	96
4.2.3	Random Forest predictive modeling	100
4.2.4	Cross validation and hyper parameters tuning	103
4.2.5	Evaluation of the predictive model efficiency	106
4.2.6	Evaluating the predictors of the model	109
5	Exploratory Data Analyses on CBA and DBA matrices	111
5.1	Clustering analyses of DBA and CBA matrices	111
5.1.1	Agglomerative Hierarchical Clustering	112

5.1.2	Hierarchical Clustering on Principal Components	116
5.2	Favorable geological environments identification	119
III	Applications	125
6	Proof of concept	127
6.1	Presentation of the study	127
6.1.1	Context	127
6.1.2	Previous studies	128
6.1.3	Dataset	131
6.2	Application of the Random Forest to CBA	134
6.2.1	CBA matrix	134
6.2.2	Results	135
6.2.3	Selection of high favorability cells	135
6.2.4	Interpretation	139
6.2.5	Comparison to other MPM methods	140
7	Prospectivity mapping and environment analyses within the Ibero-Armorican Arc	143
7.1	Antimony MPM in the Armorican Massif	143
7.1.1	Context	143
7.1.2	Automation of the methodology	144
7.1.3	Input dataset	145
7.1.4	DBA matrix	147
7.1.5	DBA / RF predictive model	148
7.1.6	Identification of favorable geological environments	151
7.1.7	Favorable areas and best predictors	153
7.1.8	Analysis of the clusters	156
7.1.9	Comparison with previous methods and CBA / RF	162
7.2	Antimony MPM in the Central Iberian Zone	164
7.2.1	Context	164
7.2.2	Input Dataset	166
7.2.3	DBA matrix	166
7.2.4	DBA / RF results	169

7.2.5	Identification of favorable geological environments	169
7.2.6	Favorable areas and best predictors	169
7.2.7	Analysis of the clusters	174
IV	Conclusion and Annexes	183
8	Discussion and synthesis	185
8.1	Methodological developments	185
8.1.1	Spatially based integration of the geological environments: the Disc-Based Association	185
8.1.2	Evaluating the likelihood of favorable geological environments: the DBA / RF favorability mapping	187
8.1.3	From favorability mapping toward automatic metallotect detection	189
8.1.4	Identification of the various favorable environments with HCPC clustering analysis	190
8.1.5	General Workflow of the methodology	191
8.2	Insights from the geological applications	193
8.3	Automation of the methodology	194
8.4	Ongoing and future developments of the methodology	195
8.4.1	Integration of continuous datasets	195
8.4.2	Structural data and other geological features	196
8.4.3	Harmonization of lithological datasets	198
8.5	Concluding points	198
9	Résumé étendu	201
9.1	Contexte et objectifs de l'étude	201
9.2	Développements	204
9.2.1	DBA	205
9.2.2	RF	208
9.2.3	HCPC	210
9.3	Applications	212
9.3.1	Zone d'étude	212
9.3.2	Application au massif Armoricaïn	214
9.3.3	Application à la Zone Centre Ibérique	217
9.4	Synthèse et conclusion	221

Acronyms	225
Bibliographie	228
A Overlapping simulations	257
B Lithological datasets details for the Armorican Massif (50K) and Central Iberian Zone (1M)	263
C Applications: HCPC clusters	269

List of Figures

1.1	Increase in the complexity of metallic assemblages in common products through time (from Reuter and van Schaik [2013]).	20
1.2	Mining potential of the French metropolitan area (after Bertrand et al. [2015]).	21
1.3	Critical and non-critical Raw Materials from 2020 EU commission Evaluation (from Blengini et al. [2020]).	22
1.4	Flood susceptibility index in the Poyang County (China) of different predictive mapping methods (from Wang et al. [2021]).	24
1.5	Mineral Prospectivity Mapping (MPM) flowchart (modified from Carranza [2017]).	26
1.6	Example of Mineral Prospectivity Mapping flowchart from Billa et al. [2004].	28
1.7	A typical Mamdani-type Fuzzy Inference System (FIS) from Porwal et al. [2015].	29
1.8	Example of Weight of Evidence (WoE) application on a 3D binary factor maps based on faults from the study of the Baixiangshan mining area by Yuan et al. [2014].	31
1.9	Schematic presentation of the Support Vector Machines method (SVM; from Sun et al. [2019]).	32
1.10	An example of Multilayer Perceptron (MLP) neural network from Wang et al. [2021].	33
2.1	Schematic cross-sections through (a) collisional, (b) accretionary and (c) intracratonic orogens from Cawood et al. [2009].	40
2.2	Orogenesis during the Phanerozoic, in 5 phases.	42
2.3	Source-Transport-Deposit sketch, fluids play a critical role in the formation of ore deposit in multiple environments (modified from Marcoux and Jébrak [2008]).	44
2.4	Simplified structural map of the Western Variscan belt (from Cochelin [2016]). IM: Iberian Massif, AM: Armorican Massif, FMC: French Massif Central, BM: Bohemian Massif.	46
2.5	Simplified sketch of the different orogenesis between Laurentia and Gondwana plates (from Ballèvre et al. [2013]).	47
2.6	Synthetic map of the Armorican Massif with simplified cross-sections fo the collision (beginning of the Late Carboniferous), modified from Ballèvre et al. [2014].	48

2.7	Simplified Geological map of the Armorican Massif (modified from the 1M geological map of France, BRGM).	51
2.8	Summary of the main events recorded in the Armorican Massif during the Devonian and Carboniferous plate convergence.	53
2.9	Main metallogenic districts and mineralization in the Armorican Massif.	55
2.10	Main metallogenic events related to the Variscan orogeny in France (from Lescuyer et al. [1997]).	56
2.11	Synthetic geological map of the Armorican Massif with location of the Sb mineralization and the main Sb districts (modified from Pochon et al. [2019]).	58
2.12	Main Variscan domains of the Iberian Peninsula (from Catalán [2011]).	60
2.13	Simplified geology and main structural domains of the CIZ (from Dallmeyer and Garcia [1990]).	61
2.14	Geodynamic and tectono-thermal evolution of the CIZ (ZCI) and GTMZ (ZGTM) during the main deformations events, modified from the study of the Palaçoulo region (east of Morais Massif, NE Portugal) of Silva [2010]).	62
2.15	Main deposits in the CIZ (from Dallmeyer and Garcia [1990]).	65
2.16	Geologic sketch of the Iberian peninsula giving structural zoning and location of the main deposits of antimony in the Iberian Peninsula, as well as the type of occurrences (from Gumiel and Arribas [1987]).	67
3.1	Theoretical sketch example of the creation of the CBA matrix from a lithological dataset.	72
3.2	Addition of new data to the CBA matrix.	73
3.3	Influence of the size of the cells used to produce the CBA matrix on the resulting associations.	76
3.4	Impact of the cell size on different parameters of the CBA matrix.	77
3.5	Identification of cells in the neighbourhood of a formation using an anisotropic search (CBA) and disc-buffers.	79
3.6	Limitation of the CBA method regarding the cell size.	80
3.7	Schematics of the DBA method for lithological association search.	82
3.8	Integration of new data in the DBA matrix.	83
3.9	Effect of the parameters R and R on the resulting DBA grid.	85
3.10	Mathematical approach to evaluate the overlapping of multiple discs for a R/d ranging from 0 to 1.	86
3.11	Numerical simulation of the overlapping surfaces between discs in a regular dot plot.	88
3.12	Impact of the search radius R and the R/d ratio on different parameters of the DBA matrix.	90
4.1	Sketch example of the application of CBA and AHC classification method.	93

4.2	Sketch example of the application of CBA and Ranking predictive mapping method.	95
4.3	Workflow for the CBA / DBA RF predictive mapping method, CBA or DBA can both be used before the RF computation.	97
4.4	Selection of prospect and non-prospect locations in a study from Daviran et al. [2021].	99
4.5	Selection of prospect and non-prospect locations for CBA and DBA matrices.	100
4.6	Creation of Random Forest prospectivity maps from CBA or DBA matrix (modified from Harris et al. [2015]).	102
4.7	Decision tree produced from a stratified bootstrap sampling of the sketch example, with a CBA grid of 5 000 x 5 000 m.	104
4.8	Example of optimization of the Predictive model using hyper-parameters tuning on 2 hyperparameters: the number of trees n and the depth of the trees d	105
4.9	Stratified cross-validation method to evaluate the Random Forest predictive model results.	107
4.10	Example of Receiver Operating Characteristic (ROC) curve produced with a 3-folds Stratified Cross Validation for the Random Forest predictive model.	108
5.1	Euclidean distance matrix applied on the CBA matrix with 5 000 m cells.	113
5.2	2D Sketch example of the calculation of Ward's distance for 2 clusters (μ_1 and μ_2) with different population (m_1 and m_2).	114
5.3	Hierarchical clustering dendrogram and resulting cluster maps from the 5 000 m CBA matrix.	115
5.4	Creation of the complete disjunctive table from a CBA / DBA matrix.	118
5.5	Projection of rows (cells) and column (modalities of our lithological associations) on the first 2 factorial axis.	120
5.6	3D HCPC clustering tree.	121
5.7	Identification of the geological environments surrounding the favorable areas on a sketch DBA / RF prospective maps (1 000 m resolution).	123
6.1	Results from the study by Gloaguen and Tourlière [2016a]: CBA / AHC method with the 12 main environments identified from the 16 deposits selected.	129
6.2	Results from the study by Gloaguen and Tourlière [2016a]: WoE favorability map computed with all the antimony deposits located in the study area.	130
6.3	Results from the study by Gloaguen and Tourlière [2016a]: CBA / Ranking favorability map computed from the 13 main antimony deposits.	132
6.4	Maps of the datasets used in this CBA / RF study.	133
6.5	Application of the multi-scale iterating process of CBA matrices on the dataset.	135
6.6	Favorability maps of antimony in the Armorican Massif produced from the integrated CBA / RF method.	136

6.7	Receiver Operating Characteristic (ROC) curve produced from the 5-folds Stratified Cross Validation of the Random Forest predictive model.	137
6.8	Analysis of the cells displaying a score higher than 0.5 in the Armorican Massif favorability map from the integrated CBA / RF method.	138
6.9	Application and comparison of various methods for prospectivity mapping of antimony in the Armorican Massif.	141
7.1	Maps of the datasets used for the DBA / RF antimony prospectivity mapping across the Armorican Massif.	146
7.2	Application of the multi-scale iterating process of DBA matrices on the lithological dataset of the Armorican Massif for various values of R and R/d	148
7.3	Favorability map of antimony in the Armorican Massif produced from the integrated DBA / RF method.	149
7.4	Receiver Operating Characteristic (ROC) curve from the 5-folds Stratified Cross Validation of the DBA / RF predictive model of the Armorican Massif.	150
7.5	Receiver Operating Characteristic (ROC) curve from the final DBA / RF predictive model of the Armorican Massif.	151
7.6	Identification of the favorable areas of the favorability map of antimony in the Armorican Massif based on the G-Means and J-Score best threshold.	152
7.7	Map of the 4 clusters produced by the Hierarchical Clustering on Principal Component (HCPC) classification of the favorable cells (score ≥ 0.506) of the Armorican Massif DBA / RF favorability map.	154
7.8	Predictors importance of the predictor of the DBA / RF predictive model of the Armorican Massif.	155
7.9	Favorability map of antimony in the cluster 1 of the AM.	157
7.10	Favorability map of antimony in the cluster 3 of the AM.	159
7.11	Favorability map of antimony in the cluster 2 of the AM.	161
7.12	Favorability map of antimony in the cluster 4 of the AM.	163
7.13	Comparison of various methods for antimony prospectivity mapping in the Armorican Massif.	165
7.14	Maps of the datasets used for the DBA / RF antimony prospectivity mapping across the Central Iberian Zone.	167
7.15	Application of the multi-scale iterating process of DBA matrices on the lithological dataset of the Central Iberian Zone for various values of R and R/d	168
7.16	Favorability maps of antimony in the Central Iberian Zone (CIZ) produced from the integrated DBA / RF method.	170
7.17	Receiver Operating Characteristic (ROC) curve from the 5-folds Stratified Cross Validation of the DBA / RF predictive model of the Central Iberian Zone.	171

7.18	Receiver Operating Characteristic (ROC) curve from the final DBA / RF predictive model of the Central Iberian Zone.	171
7.19	Identification of the favorable areas of the favorability map of antimony in the Central Iberian Zone based on the G-Means and J-Score best threshold.	172
7.20	Map of the 4 clusters produced by the Hierarchical Clustering on Principal Component classification (HCPC) of the favorable cells (score ≥ 0.51) of the Central Iberian Zone DBA / RF favorability map.	173
7.21	Predictors importance for the DBA / RF predictive model of the Central Iberian Zone. . .	175
7.22	Favorability map of antimony in the cluster 1 of the CIZ.	176
7.23	Favorability map of antimony in the cluster 2 of the CIZ.	178
7.24	Favorability map of antimony in the cluster 3 of the CIZ.	179
7.25	Favorability map of antimony in the cluster 4 of the CIZ.	181
8.1	General Workflow of the developed DBA / RF / HCPC methodology.	192
8.2	Example of integration of continuous data or predictive factors into the DBA matrix. . . .	197
9.1	Organigramme général de la Cartographie Prédictive Minérale (Mineral Prospectivity Mapping; MPM) flowchart (modifié d'après Carranza [2017]).	202
9.2	Workflow général de l'application de la méthodologie DBA / RF / HCPC.	206
9.3	Schéma de la méthode Disc-Based Association (DBA) pour l'identification d'environnements géologiques.	207
9.4	Création de cartes de prospectivité à partir de la méthode DBA / RF (modifié à partir de Harris et al. [2015]).	209
9.5	Identification des environnements géologiques entourant les zones favorables sur un schéma de la carte de favorabilité DBA / RF (résolution de 1 000 m).	211
9.6	Carte de favorabilité de l'antimoine dans le Massif Armoricain, produite à partir de la méthode DBA / RF.	215
9.7	Carte des 4 clusters produits par la Classification Hiérarchique sur Composantes Principales (Hierarchical Clustering on Principal Component, HCPC) des cellules favorables (score $\geq 0,506$) de la carte de favorabilité DBA / RF du Massif armoricain.	216
9.8	Carte de favorabilité de l'antimoine dans la Zone Centre Ibérique (CIZ), produite à partir de la méthode DBA / RF.	218
9.9	Carte des 4 clusters produits par la Classification Hiérarchique sur Composantes Principales (Hierarchical Clustering on Principal Component, HCPC) des cellules favorables (score ≥ 0.51) de la carte de favorabilité DBA / RF de la Zone Centre Ibérique (CIZ). . .	220

Part I

General Introduction

Chapter 1

Resources and Prospective mapping

1.1 Mineral Resources

1.1.1 Mineralization and metallogeny

Mineral deposits are high concentration of minerals or group of minerals containing chemical element(s) of economic interest. Their parameters (e.g. size, shape, ore grade, host rocks ...) can greatly vary from one deposit to the other, together with the processes leading to their formation. Mineral deposits can be considered rare events from a statistical point of view because they are formed and preserved by a conjunction of many processes/causes: a source for metal, a pathway to transport and/or concentrate the elements of interest and favorable environments to trap the elements and preserve the mineralization from remobilization, erosion or alteration.

Mineralization can form in various geological contexts. In particular, development of orogenies are characterised by tectonic, thermal and magmatic evolutions and processes showing favorable conditions to develop mineral concentrations. The formation of an orogen span a large amount of time and is composed of multiple phases, each accompanied with multiple orogenic processes. Deformations of various natures take place, creating localized structures favoring the circulation of fluids through the crust and local vacuum space to deposit new minerals. Changes in temperature and pressure occurs across the orogen from one tectonic unit to the other, leading to metamorphic processes reaching partial melting conditions, changing the composition of the initial lithology and liberating fluids of various composition [Cox et al., 1986, Li et al., 2019b]. Though average concentration of the crust in substances of economical interest is quite low (Clarke concentration), magmatic or hydrothermal fluids produced by the orogenic processes are often concentrated in such elements and are able to form mineralization of economic interest, called ore deposit [Cathelineau et al., 2011].

The study of mineralisation and their formation is called metallogeny. Metallogeny is the study of the genesis and regional-to-global distribution of mineral deposits, with emphasis on their relationship in space and time to regional petrologic and tectonic features of the Earth's crust. Based on the definition of metallogenic models, this science describes relationships between physical and geochemical processes and the known mineralization, allowing the identification of the recurring patterns between mineralization of the same nature and geological contexts or environments [Marcoux and Chauris, 1994]. As such,

metallogenic models are conceived from observations and insights from known mineral deposits in order to comprehend their genesis and facilitate the discovery of analogs (i.e. deposits of similar type). From selected natural examples regarded as ‘typical’, these models give clues about the relationships that usually exist between the mineralization and the geological features commonly associated with it; those patterns, called metallogenes, are geological features favorably linked to the presence of the mineralization [Marcoux and Chauris, 1994]. They can result from systematic expert study of the known mineralization, or be produced through spatial and statistical analysis of gathered data from those mineralization.

1.1.2 Societal demand on mineral resources

Ore deposits and their associated commodities have always been sought by mankind for the production of ever-evolving tools. This demand has been increasing ever since and includes more and more materials while new developments are made, and some substances, previously ignored, are now actively prospected and used [Charles et al., 2017] (Figure 1.1). With such increasing demand, some substances are becoming critical for the economy of countries: as an example, the European Union’s last Critical Raw Materials (CRM) list contains now 30 materials of critical value for its economy [COM(2020) 474 final, 2020, Blengini et al., 2020, Bobba et al., 2020], most of them found in ore deposits (Figure 1.3). Those substances are considered critical depending on their economic importance, such as their use in high-tech technologies, and their supply risk. This leads to the re-evaluation of the mining potential within some countries, such as the studies done by the BRGM within France [Gloaguen and Tourlière, 2016a,b, Vic and Tourlière, 2016, Gloaguen et al., 2018, Bertrand et al., 2013] (Figure 1.2).

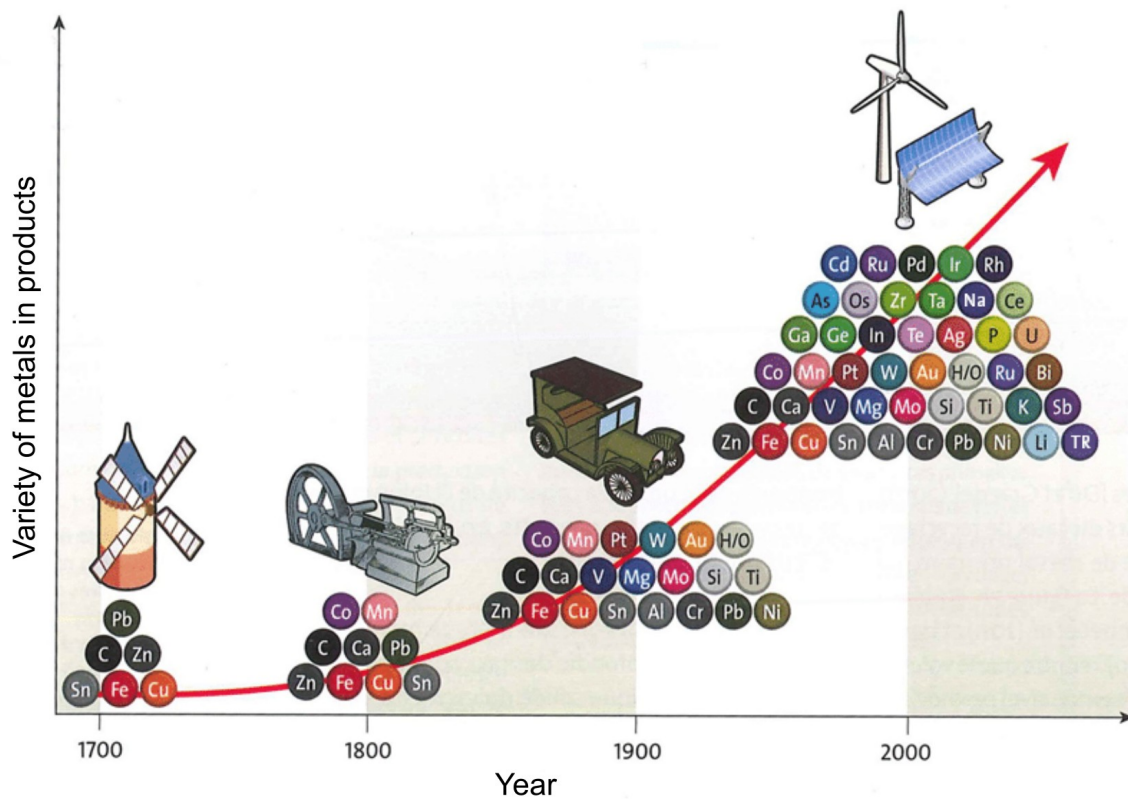


Figure 1.1: Increase in the complexity of metallic assemblages in common products through time (from Reuter and van Schaik [2013]).

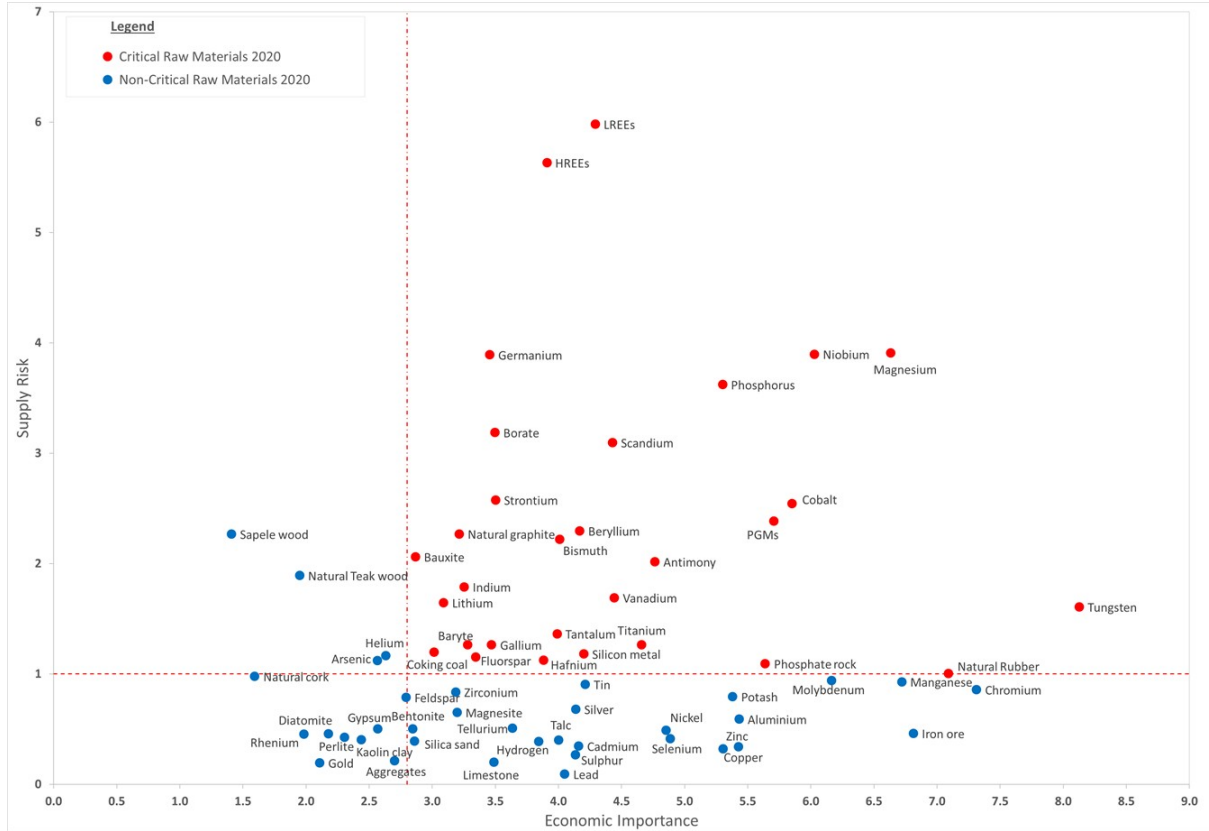


Figure 1.3: Critical and non-critical Raw Materials from 2020 EU commission Evaluation (from [Blengini et al. \[2020\]](#)).

covered mineral resources,

- The Smart Exploration project (<https://smartexploration.eu/>) aimed at developing cost-effective, environmentally friendly tools and methods for geophysical exploration,
- The New EXploration Technologies (NEXT, <https://new-exploration.tech/>) aims at the development of new exploration techniques and of new minerals system modelling, based for example on Self-Organizing Maps (SOM),
- The FRAME project (Forecasting and Assessing Europe's Strategic Raw Materials Needs, <https://www.frame.lneg.pt/>) produced a compilation of mineral potential and metallogenic areas of CRM resources in Europe,
- More recently, the Exploration Information System (EIS) has been launched in May 2022 to study the potential for different types of new mineral resources for the most important metallogenic belts of the EU.

The number of past and ongoing projects shows the interest of the EU and the governments to seek new resources. In addition, we also have more and more cartographic information available as the number of studies increases, and therefore more data that can be used by to work on with MPM.

1.2 Prospective mapping and mineral exploration

1.2.1 Mineral Prospectivity Mapping

In order to prevent potential risk or to manage natural resources, humans seek to predict and locate phenomena of diverse types, such as: fires, floods, winds, rain-showers or mineral concentrations. Predictive mapping aims at determining the spatial distribution and the degree of favorability of the likelihood of a non-random event to occur, to highlight areas of interest. It is used in many application areas, such as cartographic risk assessment, including flood [Das and Gupta, 2021, Jiang et al., 2021, Wang et al., 2021], soil erosion [Arabameri et al., 2020, Karbalaee Ramezanali et al., 2020], wildfires [Montorio et al., 2020, Tuyen et al., 2021] or spatial epidemiology [Rajabi et al., 2014] as well as for Mineral Prospectivity Mapping [Carranza, 2017, Liu et al., 2022b]. An example of predictive map for flooding [Wang et al., 2021] is displayed in **Figure 1.4**, showing the susceptibility of specific areas to suffer floods across the Poyang County in China. Each map is produced with the same original dataset but with a different predictive model, a different way of processing the data. Differences between the displayed maps show the variability of the results depending on the model used.

Mineral Prospectivity Mapping (MPM) is a branch of predictive mapping seeking to determine the likelihood of an area to host specific mineral deposits [Carranza, 2017]. According to McKay and Harris [2016], MPM is defined as the likelihood that mineral ore deposits of the type sought can be found in a specific area. The objective is to determine geographically the probability of occurrence of mineralization, the areas where further exploration could provide access to mineral resources of economic interest. The search for the most favorable areas for mining began as early as 1556, with the treatise "De re Metallica ex libere" by Georgius Agricola describing the methods and activities of mining exploration. These methods have continued to evolve since then, with the first geological (19th century) and mining maps (20th century), initially considered as uncorrelated, until the definition of metallogeny in 1905 by Launay [Orcel, 1954], establishing relationships between mineralization and the environment and geological processes [Barth et al., 2017].

Mineral Prospectivity Mapping (MPM) is part of the mineral exploration processes [Carranza and Sadeghi, 2010, McKay and Harris, 2016, Xiong and Zuo, 2017]. Mineral exploration seeks to find new ore deposits of economical interest for mining purposes. In this process, MPM is used to search areas of economic interest for the mining industry [Carranza, 2004, Yuan et al., 2014, Gloaguen and Tourlière, 2016a, Vic and Tourlière, 2016, Cassard et al., 2012, Feltrin, 2008, Carranza and Sadeghi, 2010, Roy et al., 2006, Billa et al., 2004, Ford et al., 2015, Knox-Robinson and Groves, 1997, Cassard et al., 2008, Wang et al., 2017, Sun et al., 2019, Xiao et al., 2015, Porwal et al., 2015, Harris et al., 2015, McKay and Harris, 2016, Hronsky and Kreuzer, 2019, Karbalaee Ramezanali et al., 2020, Singer and Kouda, 1996]. The mineral exploration process is composed of 4 steps, each one narrowing the study area [Carranza, 2008, McKay and Harris, 2016]:

1. The selection of the area to be explored. This area must be presenting geological characteristics suitable for the presence of targeted substances and mineral deposits. Those characteristics are based on the knowledge of the metallogenic processes and the metallogenetic structures associated with the type of mineral deposits sought. When the selected area is a well-known mineralized district or prospect, which has already been explored and where known mineralization occurs, this is a "Brownfield" context. In this case, the goal of the mineral exploration is to discover additional deposits, or

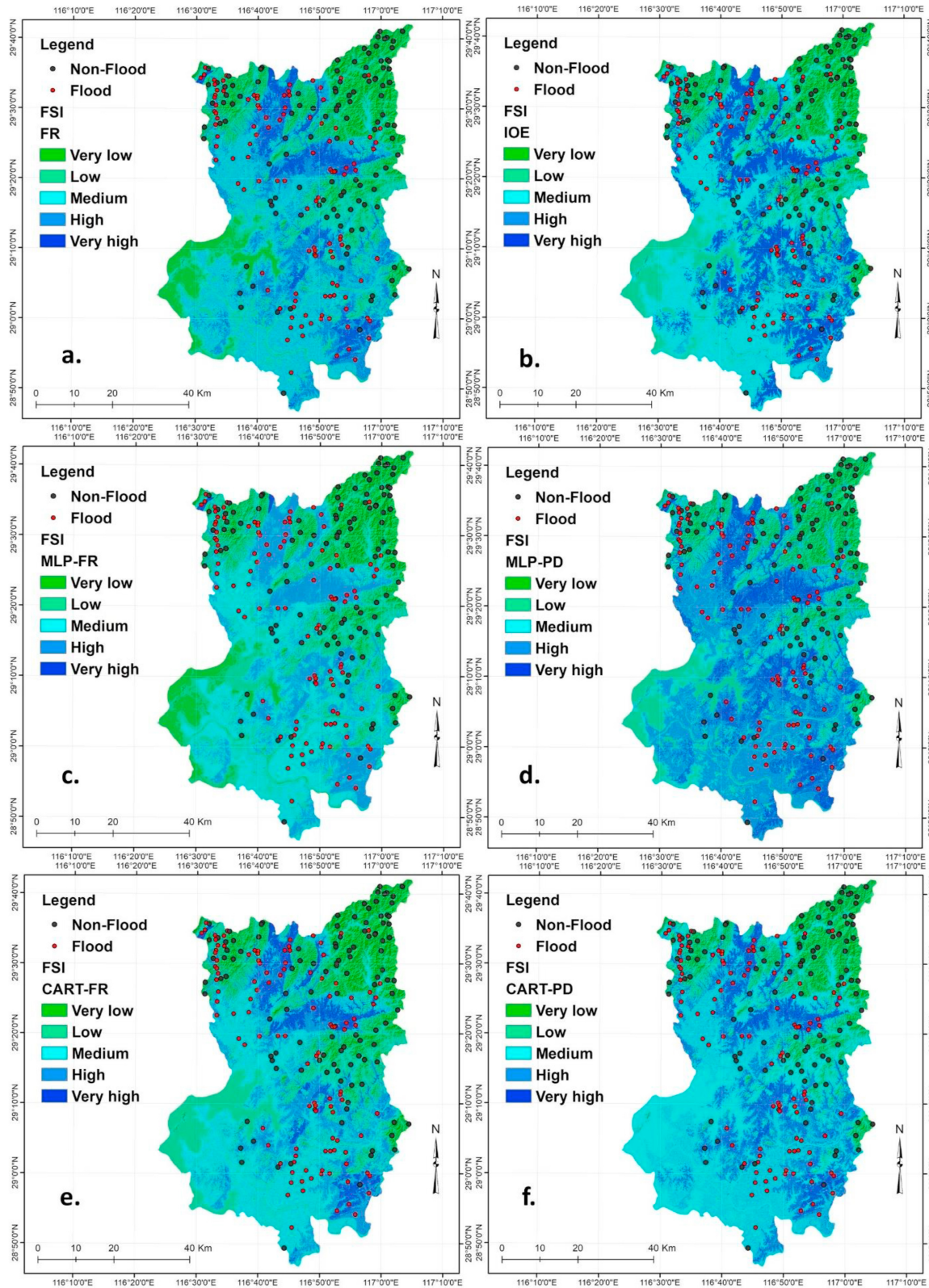


Figure 1.4: Flood susceptibility index in the Poyang County (China) of different predictive mapping methods (from [Wang et al., 2021]) a. Frequency Ratio (FR) b. Index of entropy (IOE) method c. Multilayer Perceptron and Frequency Ratio hybrid method (MLP-FR) d. Multilayer Perceptron and Probability Density hybrid method (MLP-PD) e. Classification and Regression Trees and Frequency Ratio hybrid methods (CART-FR) f. Classification and Regression Trees and Probability Density hybrid method (CART-PD)

extension of known deposit, using existing knowledge regarding the ones located in the area. If an unexplored area is targeted, where no mineralization has been identified, this is a "Greenfield" context. In this context, mineral exploration uses existing knowledge and ore genesis models to discover new potential deposits, their types depending on the geological context of the area.

2. The establishment of prospecting targets needs to identify within the area of interests clues of the presence of mineralization to further delineate the area of interest. This is usually done through grassroots exploration, with airborne survey, geological, geophysical and geochemical ground based prospecting.
3. The resource assessment phase aims at determining the high-grade mineral deposits within the prospecting targets and estimate the grade and size of the deposits. This allows the identification of deposits of economical interest and is done through drilling and geophysical studies of the ore deposits.
4. The reserve estimation is the final step of this process and is done based on the deposits of economical interest. More precise drilling and sampling are done on those deposits, in order to precisely delineate them in various ore classes depending on the grade of the ore and its location.

During the 1980s and 1990s, the evolution of computer hardware allowed the development of more advanced geo-computing techniques, with the use of Geographic Information Systems (GIS) to process spatialized information, particularly in the context of mining exploration [Barth et al., 2017, Carranza, 2011]. Those new tools allowed the development of Mineral Prospectivity Mapping, making the first 2 phases more easy to carry on and less time-consuming, allowing to quickly select area of interests and to identify the prospecting targets within those areas. Prospective mapping is also not restricted to 2D cartographic data and can be developed and applied on a tridimensional environments [Yuan et al., 2014, Xiao et al., 2015, Fu et al., 2021].

MPM uses the data and knowledge gathered about the known mineralization and ore deposits to create models to predict the areas showing the highest likelihood to contain unknown mineralization, allowing to quickly delineate and study areas with possible ore deposits by using diverse cartographic data (e.g. geological, geophysical, geochemical ones) [Bonham-Carter, 1994, Bonham-Carter et al., 1989]. Favorable areas can be highlighted, where the likelihood of hosting mineralization is considered high enough to allow further investigations. The quantification of this likelihood is estimated from a model (either deterministic or statistical) where relationships are evidenced between mineralization occurrences and given surrounding cartographic features.

1.2.2 Predictive models

Predictive models are simplified representations of complex natural events, as the process they aim to predict are made of too many complexes phenomena (i.e. physicochemical reactions, fluids circulation, crystallization kinetics, etc.) which cannot be all integrated in deterministic modelling as of now. Predictive models use another approach, by trying to establish and characterize the relationships between spatial variables (both quantitative and qualitative) and an event of interest. They use those variables to assess the favorability for the presence of this event across the studied area. This obviously requires direct observation data of the event, such as geographically known occurrences, to be able to design and validate these models [Carranza, 2008]. Those known evidences are used as a mean to train a predictive

model. Predictive mapping uses predictive models to quantify and characterize the likelihood of occurrence of an event, seeking to determine the relationships between measured variables and an event, and more specifically the variables that can determine the presence or absence of this event.

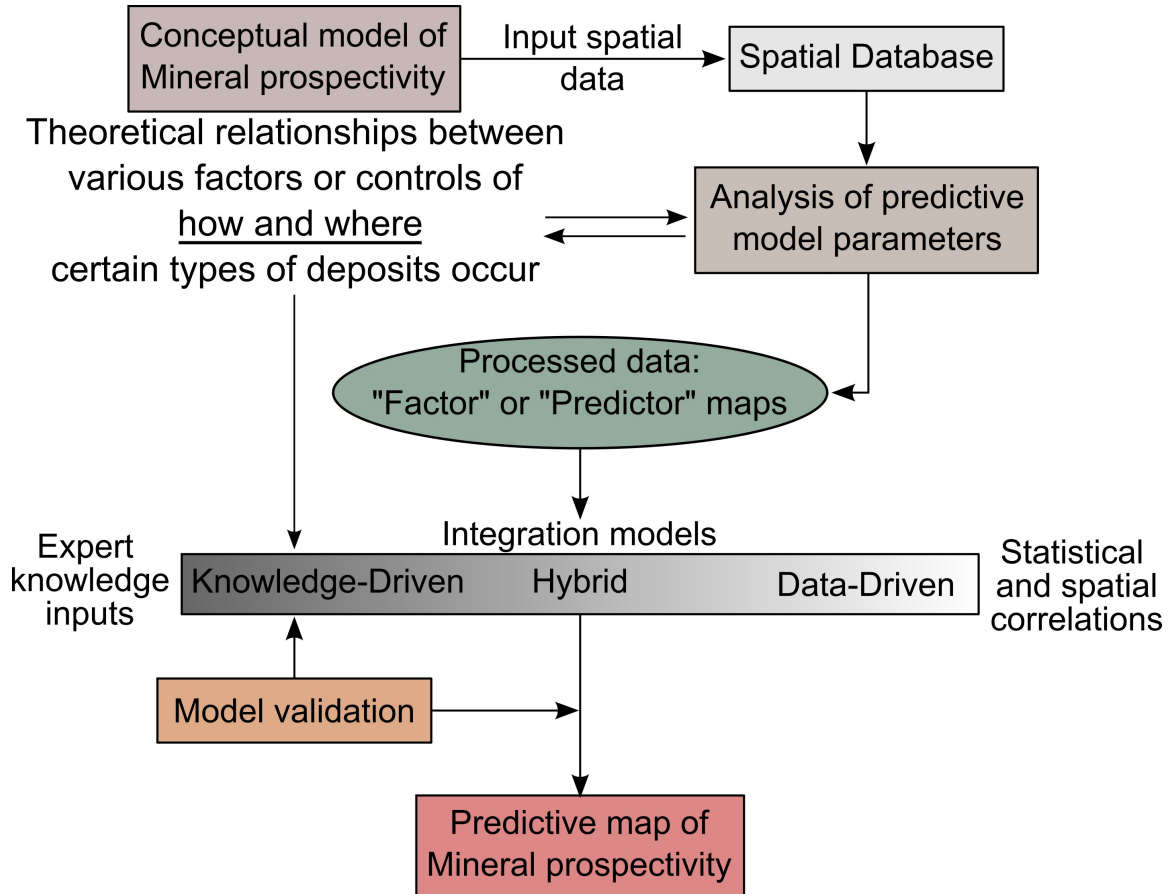


Figure 1.5: Mineral Prospectivity Mapping (MPM) flowchart (modified from Carranza [2017]).

A general workflow to produce a Mineral Prospectivity map can be described as follows (see **Figure 1.5**): it first requires a conceptual model establishing the relationships between various geological and cartographic factors and the targeted mineralization. This model, used for the creation of a favorability map, is based on some assumptions: an area is considered favorable if it shares evidential features (geochemical or geophysical signatures, geological features ...) with known locations of mineral deposits, those evidential features being independent to each others and varying across the study area. The more of those features is present in an area the more favorable it is deemed [Carranza, 2011]. As such, metallogenic models and metalotects are commonly used as factors of controls in MPM to estimate the zones where mineralization are most likely to be hosted, by giving insight about the favorable features. Most of the existing prospective models are dependant to the known/established metallogenic models, defined from the analysis and study of some occurrences considered as 'typical'.

The conceptual model guides the selection of the spatial data that will be used to produce the favorability map (**Figure 1.5**). This process requires to gather and select cartographic data (i.e. geological formation or structures, geophysical and geochemical surveys...) of the explored area from a spatial database. Those data needs to be analyzed and correctly processed to be integrated to the predictive model, resulting in "Factors" or "Predictors" maps (**Figure 1.5**). This means analysing and extracting,

manually or based on statistical and spatial analysis, the characteristics of those data that are meaningful for the predictive model.

Those processed data are then integrated into a predictive model which is designed accordingly to the established conceptual model (**Figure 1.5**). This predictive model is able to highlight the areas that are the more likely to contain targeted mineralization from the processed cartographic data layers available. It defines the characteristics that are favorable to the presence of mineralization. If the main principle of MPM remains the same in many studies, the way the prospective maps are computed can be quite different as the predictive models used for the production of prospective maps are quite diversified.

Two main types of mineral prospective modelling approaches exist (**Figure 1.5**): Knowledge-driven models and Data-driven models [McKay and Harris, 2016, Harris et al., 2015, Carranza, 2011]. According to Dubois et al. [2000], the first modelling approach "emphasizes logic as a tool for representing beliefs held by an agent", while the second modelling approach "claims that the main source of knowledge is made of observed data, and generally does not use [human] logic as a modelling tool." As such, the Knowledge-driven approach heavily relies on human input to produce prospective maps, while the Data-driven one heavily relies on spatial and statistical analyses, which can be automated, to produce prospective maps [McKay and Harris, 2016, Harris et al., 2015, Carranza, 2011]. Hybrid models also exist, combining Knowledge-driven and Data-driven elements to try to improve the predicted results. Knowledge-driven and Data-driven models can both produce good results in some given field of application, the results depending on the model used and its application.

Some comparison between Knowledge-driven and Data-driven methods have been made by Harris et al. [2015], McKay and Harris [2016]. In their work, the performance of Fuzzy Logic (FL) and Weighted Index Overlay (WOI) Knowledge-driven methods are compared to the Random Forest (RF) Data-driven methods. The results of these studies show that, when applicable (i.e. study area with known deposits and multiple datasets), Data-driven methods can provide better favorability maps than Knowledge-driven ones. Rajabi et al. [2014], in his paper made the same conclusion when comparing a Neural Network (NN) method with two Fuzzy Knowledge-driven methods.

The robustness of the predictive model is evaluated by using its effectiveness on known deposits, presenting similar characteristics with the targeted deposits, allowing to validate the results of the model (**Figure 1.5**). The results of the application of a predictive model on the processed data are mainly the production of prospective maps of various scales, used to identify more precisely the prospecting targets and set up the next stage of the mineral exploration campaign [Carranza, 2011, 2008, Cassard and Bertrand, 2017].

1.2.3 Knowledge-driven approach

The Knowledge-driven approach has been the first to be used for MPM. It aims at determining the relationships between mineralization and cartographic features based on a pre-existing conceptual metallogenic model defined for the mineralization studied. It relies on expert inputs and is based on well-known areas that have been thoroughly studied and considered as reference examples in the literature [Carranza, 2017] to determine the evidential features associated to the targeted deposits and their weights. With such approach, the variables used by the model (criteria) are identified and weighted using expert knowledge, based on empirical analyses of the reference case studies.

As they rely on such pre-existing knowledge, Knowledge-driven approaches do not require to collect

data concerning the known mineral occurrences or deposits within the study area, at the cost of adding human bias to the produced models. They rely on the use of mathematical and probabilistic rules associated with GIS tools to produce prospective maps. Predictive maps can thus be computed from a dataset including one layer per criterion with no need of compiling places where the studied event is known to have occurred. This makes those methods particularly fit for greenfield exploration, when seeking to discover mineral deposits in new areas, away from any known deposit.

Knowledge driven approaches include multiple methods as shown in Carranza [2017]. The boolean logic or overlay [Thiart and de Wit, 2000, Harris et al., 2001, Skirrow et al., 2019, Carranza et al., 1999] methods rely on the addition of binary predictor maps selected or produced from an initial dataset based on expert knowledge. Binary or multi-class index overlays methods [Harris et al., 2015, Billa et al., 2004, Roy et al., 2006, Cassard et al., 2008, Madani, 2011, Carranza et al., 1999] function in a similar fashion, with the favorability score being the summing of binary or multi-classes predictor maps (Figure 1.6). In this case, each predictor map can also be assigned a weight by the expert, according to its supposed importance ("Weighted index overlay" or WIO, [Harris et al., 2015]).

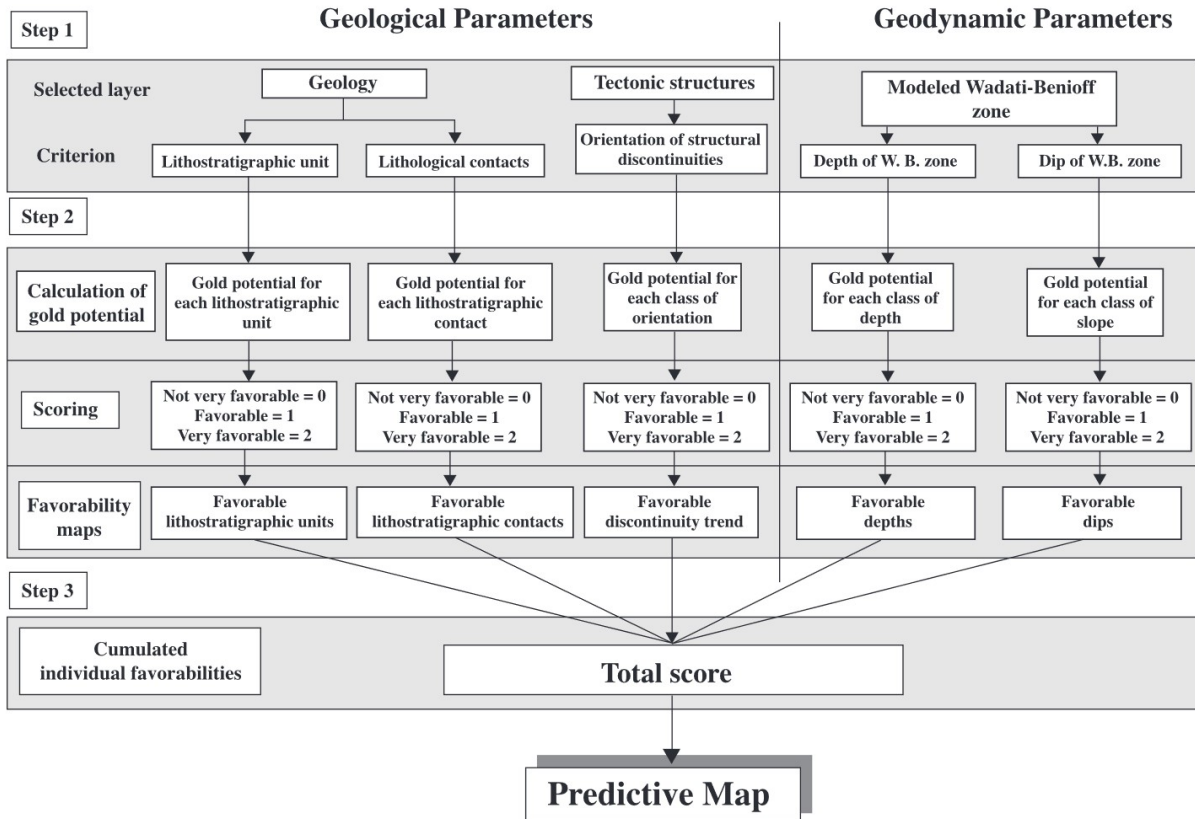


Figure 1.6: Example of Mineral Prospectivity Mapping flowchart from Billa et al. [2004]. Favorable parameters are selected and analyzed by an expert, which attribute a favorability score to the features contained in each layer (i.e. the lithologies inside the lithostratigraphic units). The total favorability score is the addition of each score from each layer selected across the studied area.

More advanced methods are, for example, the Fuzzy Logic method [An et al., 1991, McKay and Harris, 2016, Moon and An, 1991, Porwal et al., 2015, Knox-Robinson and Groves, 1997, Ford et al., 2015, Liu et al., 2014, Ma et al., 2020, Torppa et al., 2019, Tabaei et al., 2017, Yousefi and Carranza, 2015, Nykänen et al., 2015, Feltrin, 2008, Porwal et al., 2003a], relying on the fuzzy sets theory developed

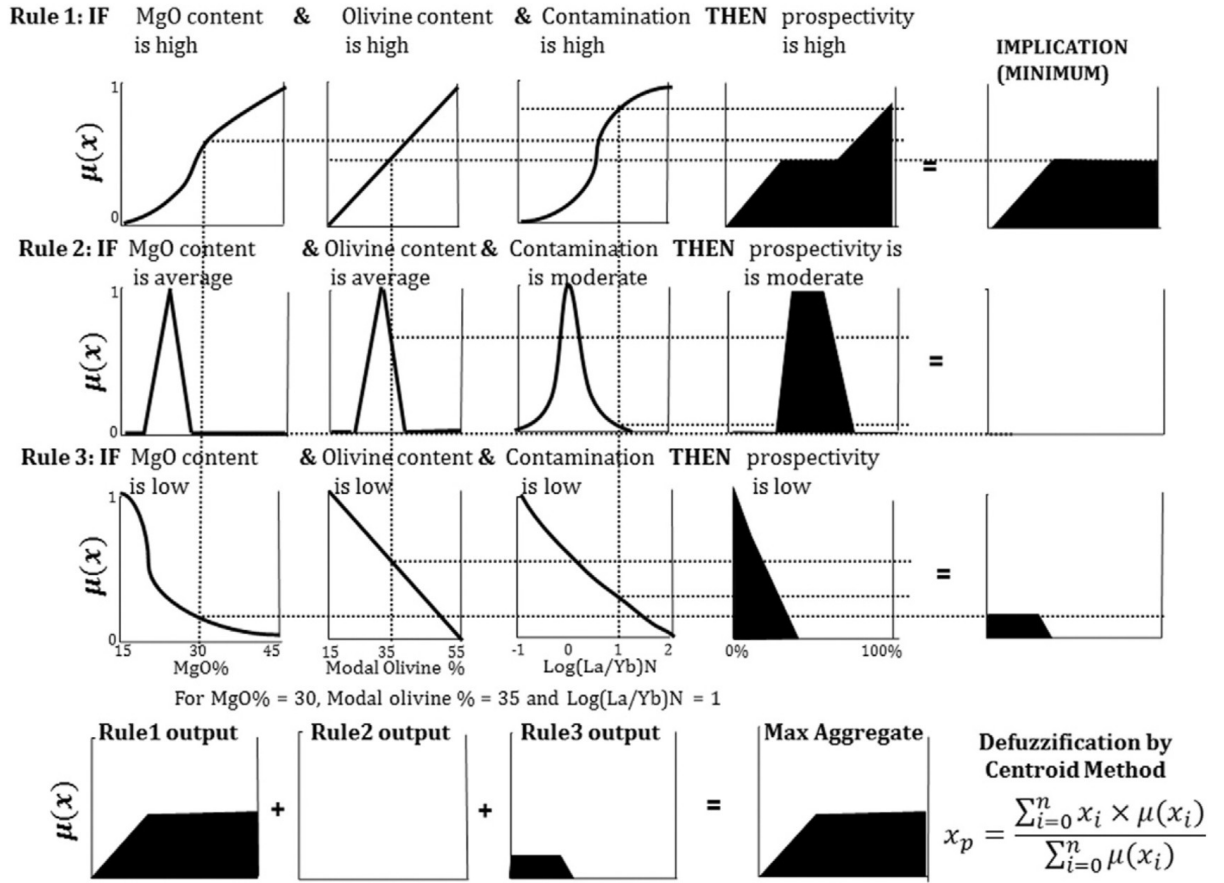


Figure 1.7: A typical Mamdani-type Fuzzy Inference System (FIS) from Porwal et al. [2015]. Three layers are fuzzified: the MgO and Olivine content and the Contamination content. Each of those layers has different membership functions defining if their values are considered as "high", "medium" or "low". The prospectivity potential of the studied area is then computed by applying rules to the combination of the memberships of the layers. The final favorability score is obtained by applying a defuzzification procedure combining the outputs of each rule to compute a point value.

by Zadeh [1965] to handle vague or imprecise concepts. With ordinary sets, the variables or classes "True" and "False" can be handle by two unique values: 1 and 0. One element would belong either to the "True" or the "False" class. In fuzzy logic, the fuzzy sets can take an infinite number of values in the interval [0: "False", 1: "True"], and use the concept of "membership" to accurately characterize the relationship between elements and classes of the fuzzy sets. For example, the closer to 1 an element is, the more similar or the closer it is to the "membership" True. The conversion from continuous or discrete values to fuzzy membership is done through a fuzzy membership function. In MPM, the expert defines the fuzzy memberships rules of multiple predictor maps and combines them to produce the prospectivity map using a Fuzzy Inference System (FIS; **Figure 1.7**).

Evidential belief method [Tangestani and Moore, 2002, Chen, 2004, Rogge et al., 2006, Carranza and Sadeghi, 2010] also allow dealing with vague concepts and is used in Knowledge-driven MPM. This method relies on the belief function theory also called Dempster-Shafer theory, formalized by Dempster [1968] and by Shafer [1976]. With this method, the uncertainty of a value is represented by various mass functions: the belief or support function estimate the probability that a value supports a proposition (conservative estimate for a proposition: i.e. this value means "this cell is mineralized"), the plausability function evaluate the probability that at a value is coherent with a proposition (optimistic assessment

that the evidence supports a proposition: i.e. this value could mean that "this cell is mineralized") [Tangestani and Moore, 2002]. Those functions are defined by the expert for each predictor used, and each one can be mapped separately.

It is noteworthy that new innovative Knowledge-driven methods are also developed, such as the Best-Worst Method and Additive Ratio Assessment (BWM-ARAS) method [Karbalaei Ramezanali et al., 2020].

1.2.4 Data-driven approach

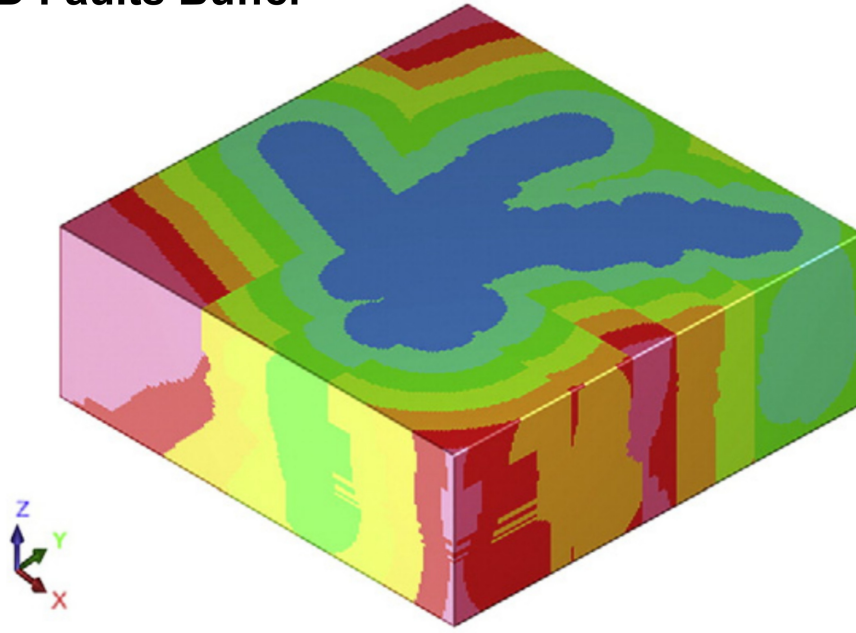
This approach is more recent and its development has been largely favored in the past decades by the growth of databases and computing capacities. They have seen an increased use for favorability mapping in geosciences [Dramsch, 2020]. The Data-driven approach mainly relies on Machine Learning (ML) methods, as it requires processing large amounts of data through statistical and spatial analyses to identify and quantify the relationships between the event and the variables within a reference-training zone. The reference-training zone contains known occurrences of the targeted event used as a training set for statistical computations; this requires a sufficient amount of data to provide relevant results. By principle, a maximum number of input dataset layers should be considered, with a minimum apriori pre-selection, in order to automatically highlight from the statistical results the best criteria to use for the model.

The conceptual models deduced with this kind of approach rely on this analysis to identify factors or predictors that are the most probably linked to the mineralization process. The main hypothesis is that if an area displays common features and/or map configuration with the reference-training zone, it is most likely to host similar mineralization in comparison to other areas. So, the more similar an area is to the reference one (i.e. the more common features or configurations are shared), the higher are the chance of finding new mineralization evidences there, while, conversely, areas with limited similarity with the reference one have low probabilities to host the same kind of mineralization [Carranza, 2017]. This approach is best suited for brownfield exploration, as pre-existing mineralization and deposits can be compiled into a training dataset to train data-driven predictive models.

Data-driven methods need to be based on data about the known mineral occurrences or deposits in the study area while also relying on only few inputs from the geologist. It allows us to avoid biases or subjectivity that could be linked to the extrapolation of the knowledge acquired on few orebodies (e.g. metallogenic models). However, if human-related biases are less evident in those models, most Data-driven methods are far from being completely devoid of human input, as the initial set of data is still selected based on the operator's subjectivity and is pre-processed manually. The results of those methods are also validated by experts and can lead to adjustments and modifications in the final predictive model.

The Data-driven approach includes multiple methods as shown in Carranza [2017] and Liu et al. [2022b]. The Weight of Evidence method (WoE) [Carranza, 2004, Wang et al., 2017, Ford et al., 2015, Liu et al., 2014, Feltrin, 2008, Gloaguen and Tourlière, 2016a, Agterberg et al., 1990, Porwal et al., 2010, Maepa et al., 2021, Yuan et al., 2014, Mao et al., 2019, Li et al., 2019a, Wang et al., 2017, Zhang et al., 2016, Gao et al., 2016, Zuo et al., 2015, Xiao et al., 2015, Nielsen et al., 2015, Ziaei et al., 2011, Partington, 2010, Feltrin, 2008, Porwal et al., 2006, Harris et al., 2003, Asadi and Hale, 2001, Bonham-Carter et al., 1989], which is presented in details in Bonham-Carter et al. [1990] and Agterberg et al. [1990], is a linear predictive model based on the Bayesian theory of conditional probability. This method relies on the

3D Faults Buffer



Weight and contrast of the 3D binary intervals

Buffer distance from (m)	Buffer distance to (m)	Positive weight	Negative weight	Contrast
0	20	1.310	−0.026	1.336
20	40	1.366	−0.062	1.428
40	60	1.574	−0.067	1.641
60	80	1.603	−0.074	1.677
80	100	1.632	−0.076	1.708
100	120	1.667	−0.103	1.770
120	140	1.518	−0.070	1.589
140	160	1.439	−0.079	1.518
160	180	1.478	−0.081	1.559
180	200	0.905	−0.038	0.943
200	220	−0.167	0.004	−0.170
220	240	−0.561	0.010	−0.571
240	260	−0.587	0.011	−0.598
260	280	−0.496	0.010	−0.506
280	300	−0.543	0.010	−0.553
300	320	−0.665	0.012	−0.677
320	340	−0.533	0.009	−0.542
340	360	−0.535	0.010	−0.544

Weight and contrast of final fault 3D binary map

Prospective range	Positive weighting	Negative weighting	Contrast
0–180 (m)	1.524	−1.29	2.814

Figure 1.8: Example of Weight of Evidence (WoE) application on a 3D binary factor maps based on faults from the study of the Baixiangshan mining area by [Yuan et al. \[2014\]](#). The dataset is binarized using buffers intervals and positive weights, negative weights and contrast are computed for each interval. The final buffer map used for the posterior probability map is determined using the buffer displaying optimal contrast and positive weight, with both values above 0.

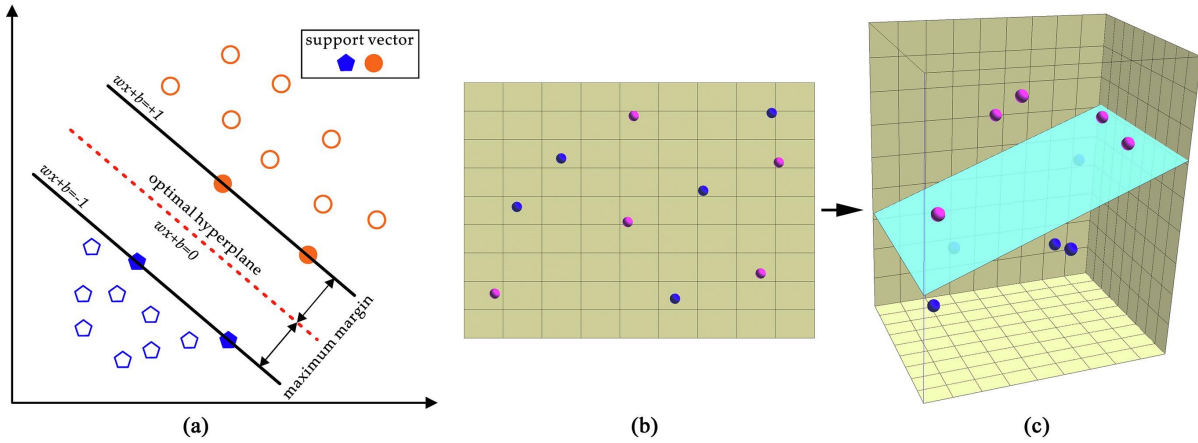


Figure 1.9: Schematic presentation of the Support Vector Machines (SVM) method (from [Sun et al., 2019]). (a) Schematic description of SVM in a linearly separable feature space. The optimal hyperplane is defined by maximizing the margin between the classes. (b) Schematic description showing a transformation of original input vectors from a 2D linearly inseparable feature space into (b) a higher 3D dimensional space where a 2D hyperplane is constructed to separate different classes.

computation, for each predictor map, of a prior probability to encounter mineralization in a random unit area. A positive and negative weight is computed for each binary predictor map using a log function and the conditional probability ratio. The positive weight indicates a positive spatial association between the evidence in a unit area and the mineralization while a negative weight indicates a negative one. The strength of the relationship between the predictor map and the mineralization, or contrast value, can be computed as a difference between the positive and negative weight and allows determining a cut-off value when searching for an optimal way to binarize the dataset. Such example can be given when one wants to define favorable zones around faults (**Figure 1.8**). The final predictive map, or posterior probability map, is the result of the combination of the various final weighted factor maps.

Logistic Regression (LR) [Porwal et al., 2010, Harris et al., 2003, Carranza and Hale, 2001, Li et al., 2019a, 2015, Harris and Pan, 1999] is a non-linear predictive model, developed by Berkson [1944] and presented in details in Hosmer et al. [2013]. It is a supervised ML classification method based on a mathematical model, a sigmoid function transformation applied to a linear regression model. This sigmoid function displays on the map the probability of an event depending on the value of a variable, allowing to perform a classification by setting a threshold value for the probability. In MPM, the LR approach allows modelling the probability of the occurrence of mineralization as a function of binary, numeric and/or categorical predictor maps [Porwal et al., 2010].

Support Vector Machines (SVM) [Sun et al., 2019, Maepa et al., 2021, Sun et al., 2020, Prado et al., 2020, Rodriguez-Galiano et al., 2015, Granek and Haber, 2015, Abedi et al., 2012, Zuo and Carranza, 2011] is another Data-driven supervised ML method detailed in Boser et al. [1992]. This method seeks to create a linear or a hyperplane classifier to separate different classes using the widest decision boundaries. A hyperplane is produced when the problem is non-linear by transforming the input data into a higher dimensional plane. In this new dimensional space, it is possible to find an optimal hyperplane to classify the data, maximizing the margins between the classes. A schematic example is shown in **Figure 1.9**.

Artificial Neural Network (ANN) methods [Sun et al., 2019, Singer and Kouda, 1996, Chen and Wu, 2017, Köhler et al., 2021, Xiong and Zuo, 2017, Rodriguez-Galiano et al., 2015, Oh and Lee, 2010] were first introduced by McCulloch and Pitts [1943] as a way to create mathematical or numerical analogues to

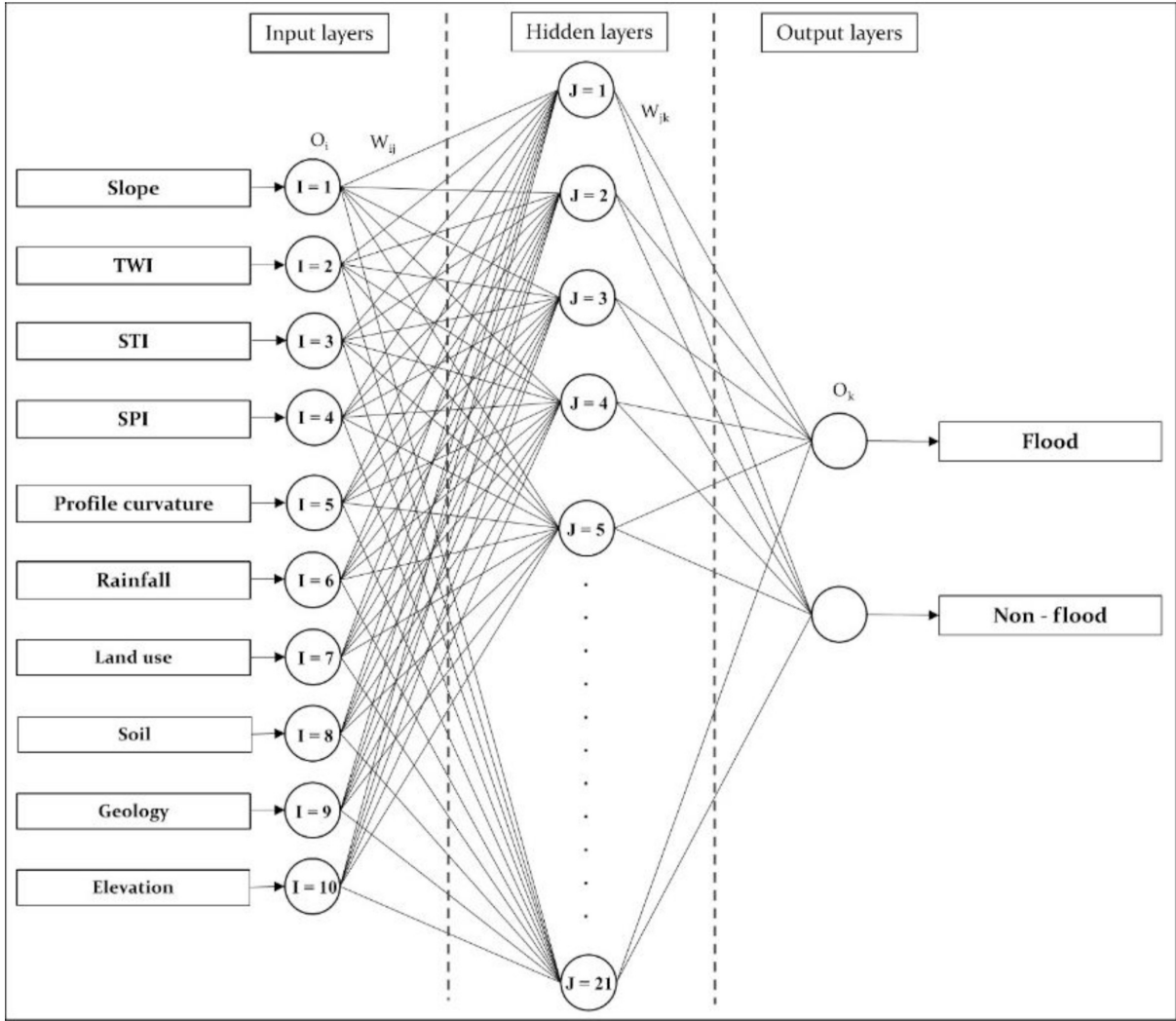


Figure 1.10: An example of Multilayer Perceptron (MLP) neural network from Wang et al. [2021]. Each input layer is processed by perceptrons in the hidden layers. W_{ij} weights are attributed between each connection. The MLP compute multiple iterations of the forward propagation of the input dataset through the layers. Output values from the training dataset are then evaluated and the values of the weights W_{ij} are adjusted to minimize the root mean square error between observed and computed predicted values.

the biological neurons [Aggarwal, 2018]. ANN simulates the biological neural network by using computation units, called neurons or perceptrons, applying a function to imputed values. Multiple neurons are connected in ANN, and the strength of their connections is simulated by weights applied to their inputs (Figure 1.10). The ANN is capable of changing its output and to learn by changing the weight applied to the neuron inputs. ANN can be supervised or unsupervised ML methods, applied for classification or regression. Depending on their architecture, their complexity, the type of neurons and their objectives, multiple models of ANN exist, such as Multilayer Perceptron (MLP) [Skabar, 2007, Brown and Groves, 2003, Brown et al., 2000], Probabilistic Neural Network (PNN) [Wang et al., 2011, Harris et al., 2003, Harris and Pan, 1999], Adaptive Resonance Theory Network (ARTN), Radial Basis Function Neural Network (RBFNN) [Porwal et al., 2003b], Restricted Boltzmann Machine (RBM) [Keykhay-Hosseinpour et al., 2020], Recurrent Neural Network (RCN) or Convolutional Neural Networks (CNN) [Aggarwal, 2018, Maepa et al., 2021, Sun et al., 2020, McMillan et al., 2019], Deep Autoencoder Network (DAN) [Xiong et al., 2018], Back-Propagation Neural Network (BPNN) [Fu et al., 2021], Group Method of Data

Handling (GMDH) [Parsa et al., 2022] or Self-Organizing Maps (SOM) [Klus et al., 2017, Leite and de Souza Filho, 2010, Kohonen, 2013, Torppa et al., 2019].

Among others, the Random Forest (RF) method has also been used for MPM, showing good effectiveness to produce favorability maps [McKay and Harris, 2016, Sun et al., 2019, Harris et al., 2015, Carranza and Laborte, 2015a,b, 2016, Rodriguez-Galiano et al., 2015, Zhang et al., 2019, 2016, Sun et al., 2020, Liu et al., 2022a, Daviran et al., 2021, Keykhay-Hosseinpour et al., 2020, Gao et al., 2016]. Further details about this method are given in **Chapter 4**.

1.3 Developing a new approach to constrain relationships between mineralization and geological features

1.3.1 Context of this study

This work is part of the AUREOLE project (tArgeting eU cRitical mEtals (Sb, W) and predictibility of Sb-As-Hg enviroNmentaL issuEs, ANR-19-MIN2-0002, <https://aureole.brgm.fr/>). This research project is coordinated by the Bureau des Recherches Géologiques et Minières (BRGM), as part of an international consortium made of the BRGM, the Institut des Sciences de la Terre d'Orléans (ISTO), the Escuela de Ingeniería Minera e Industrial de Almadén of the University of Castilla-La Mancha (IGeA-UCLM), the Faculdade de Ciências da Universidade do Porto (FCUP) and the Antea Group. This project target Sb $\pm$ W ore deposits in high-potential areas illustrated by the Variscan massifs of Iberia and France. In this project, this thesis is part of the Work Package WP3 "Mineral resources prospectivity and associated environmental risk assessment", aiming at characterizing the spatial distribution of known Sb-(W) mineralisation and computing prospective maps based on statistical approaches. In this WP, this thesis is part of the task T3.1, focusing on the analysis of (i) spatial distribution of known ore/mineral deposits and (ii) spatial associations between those deposits and the host lithologies and surrounding structures s.l. It also aims at harmonizing, at a given scale, the map representation of heterogeneous datasets to produce the statistical and spatial analysis of the Sb deposits in the studied area to constrain and quantify their spatial relationships with known map features.

As such, this work focuses on the study of antimony (Sb) occurrences as a proof of concept for the new prospectivity approach developed. Sb is used in a variety of industrial operations such as production of flame-retardants, plastics, paint pigments, glassware and ceramics, ammunition alloys and battery manufacturing plants. The global consumption of Sb has increased to more than 1.4×10^5 tons each year [Guo et al., 2014, Henckens et al., 2016] as the demand for use in flame-retardants has increased in recent years. The Sb market is likely to increase in the incoming years as an up-and-coming use for Sb may be for rechargeable high-performance Li or Na-ion batteries, since Sb-based materials seems to be a new promising candidate for ultra-fast high-capacity anode materials [Sun et al., 2010, Luo et al., 2017, He et al., 2018] thanks to their high theoretical capacity and suitable operating voltage [Zoller et al., 2018].

Sb ore deposits are present along the Variscan range within France, Portugal and Spain and have been thoroughly studied [Gumiel and Arribas, 1987, Gumiel, 1982, Gumiel et al., 1976, Arribas and Gumiel, 1984, Couto et al., 1990, Ortega and Vindel, 1995, Ortega et al., 1996, Neiva et al., 2008, Pochon et al., 2016, 2019, 2018, Cheval-Garabedian, 2019, Gloaguen and Tourlière, 2016a,b, Pochon, 2017, Marcoux and Chauris, 1994, Munoz et al., 1992, Marcoux et al., 1984, Bouchot et al., 2005]. The Variscan Range

is an orogen resulting from the collision between the Laurussia to the North and the Gondwana continent to the South during the Paleozoic [Ballèvre et al., 2009, 2013, von Raumer et al., 2003], with the Armorica and the Avalonia microcontinents squeezed in-between. In Europe, the Variscan Range extends from the Iberian peninsula to Central Europe. This study focuses on the Western part of the Variscan Range, more precisely the Central Iberian Zone (CIZ) and the Central Armorican Domain (CAD), covering rather large parts of the external domains of the Variscan Range. The CIZ and the CAD are both part of the Ibero-Armorican Arc (IAA), an arcuate structure extending from the Iberian Massif (IM) to the Armorican Massif (AM) across the Bay of Biscay [Dias et al., 2016]. The association of major faults and ductile shear zone with strong magmatic activities favors the fluid circulation through the crust and thus the development of mineralization [Pereira et al., 1993, Cox et al., 1986, Mateus and Noronha, 2010, Almeida et al., 2002, Pochon et al., 2016, Hedenquist and Lowenstern, 1994, Yardley and Cleverley, 2015], adding more possible constraints to the predictive model. Many recent studies have been done about the mineralization and metallogenic districts in the IAA for multiple elements of interest [Launay, 2018, Silva et al., 2018, Palero-Fernández et al., 2015, Blundell et al., 2005, Cassard et al., 2012, COM(2020) 474 final, 2020, Hartai et al., 2016, Tornos et al., 2004]. There is thus a high quantity of data gathered in this area, necessary to feed a Data-driven model for prospective mapping. Despite all those studies, the Variscan Range is still an interesting target for mineral exploration [Hartai et al., 2016], as multiple "local" studies, focusing on part of this Range, have been performed [Vic and Tourlière, 2016, Gloaguen and Tourlière, 2016a,b, Cassard et al., 2012, Köhler et al., 2021, Cassard et al., 2012] but only few "regional" studies have been produced [Sadeghi et al., 2020, Bertram, 2020]. There is an established continuity between the main geological units and structures across this Arc [Ballèvre et al., 2014, Catalán et al., 2007, Quesada and Oliveira, 2019, Dias et al., 2016, Aerden, 2004, Martínez Catalán, 2012, Dias and Ribeiro, 1995, Pereira et al., 1993], allowing to produce coherent predictive maps from a 1:50,000 scale to a 1:1,000,000 scale.

1.3.2 Aims of this study

Most of the uses of MPM aim to produce favorability maps in order to delineate areas of interest for mining exploration [Carranza, 2017]. MPM allows highlighting areas with a high chance of containing mineralization based on geological and metallogenic knowledge of the relationships between mineralization and geological features (i.e. Knowledge-driven approach), and/or based on data analysis to extract and highlight such relationships (i.e. Data-driven approach). Both approaches use data gathered throughout decades of studies, the first one relies on human analysis of the data with the production of metallogenic models via an expert approach, based on local example of mineralization and their results are extrapolated to other mineralization. The second one nowadays relies on the development of Artificial Intelligence (AI) and Machine Learning (ML) tools to perform systematic spatial and statistical analyses of the data on a large scale to identify and constrain relationships between mineralization and features within the dataset.

Comparison between Knowledge-driven and Data-driven predictive models have shown the effectiveness of the latter in context with high density of data [McKay and Harris, 2016, Harris et al., 2015, Porwal et al., 2003a]. As the Data-driven approach is able to provide tools to evaluate the relevance of the relationships between map data and mineralization, it could be used to identify potential "metallotects" in an automatic manner to enrich existing metallogenic models, by providing new data constraining the geological environments that may contain mineralization, through a more integrative view of the metallogenic processes. Besides, the use of a Data-driven approach allows minimizing the contribution of

pre-existing metallogenic knowledge to the predictive model, avoiding to include an expert bias in the model. This enables comparison of the automatic metallotects to pre-existing models and knowledge to validate the Data-driven model produced and therefore to include them.

As such, this thesis work aims to develop a Data-driven method for automated identification of relationships between mineralization and geological features through associations of characterizing features. This automated Data-Driven method process would allow to identify and differentiate specific geological environments surrounding known mineralization, allowing an inverse approach to MPM where a conceptual model could be drawn from the spatial data available, rather than based on pre-existing knowledge. Thus, this method aims at being the most Data-driven possible to avoid as far as possible an expert bias on the outcome of prospectivity mapping and on the identified relationships. This would allow observing the effectiveness of AI and ML approaches to highlight complex geological links between mineralization and geological features and quantify the strength of these relationships.

The aim of this thesis can be summarized with the following key questions :

- **Using an inverse Data-driven MPM approach, can we highlight lithologies and geological structures s.l. related to mineralization as well as potential metallotects to produce the equivalent of a conceptual model, inferred from the data used and the favorability maps produced ?**
- **Can we identify local geological environments with the available data ?**
- **How to constrain the likelihood of the local geologic environments and their geological features to be linked to mineralization ?**
- **How to synthetize all favorable environments into “key” signatures, explaining the unique features associated with the mineralization ?**
- **Can this whole process be carried out in a Data-driven, automated way, avoiding as much as much as possible the use of pre-existing knowledge ?**

Results obtained from this approach will be compared to the metallogenic models proposed by more "naturalists" published studies, to estimate if the metallotects and criterion highlighted by this new automated approach stand out from the known published ones. By highlighting new links between geological features and mineralization, it should allow a better understanding of metallogenic models and mineralization through the precision of those various relationships. Applying this method at multiple scales allows the observation of its effectiveness and consistency at different scales and for different purposes, with different resolution of data. As the geological, structural, geochemical or geophysical data covering the study area (i) rarely present the same spatial resolution, (ii) are of various types (i.e. continuous, discrete, categorical ..) and (iii) can originate from various scales, the developed method must be able to harmonize and process the data. As this method relies on prospective mapping, its development should be accompanied by the development of new tools for mineral prospectivity mapping.

The approach developed and presented in this manuscript is based on the combined use of several methods and is referred to as the DBA / RF / HCPC method. At first, a spatial analysis and data harmonization method, the Disc-Based Association (DBA), is developed, allowing to gather the cartographic data available in a study area into a single multivariate grid, presenting local geological environments. A ML method, the Random Forest (RF) supervised classification, is then used to analyze these environments in relation to known mineralization to produce a favorability map. This favorability map and these

environments are finally processed using an unsupervised clustering method, the Hierarchical Clustering on Principal Component (HCPC), in order to identify and group together the geological environments in the most favorable areas that share common features. The combination of these methods allows (i) the aggregation and harmonization of data of different natures and various scales, (ii) the identification and quantification of the relationship between these data and known mineralization and (iii) the discretization of the environments displaying the strongest relationships with the mineralization (i.e. the highest favorability scores) to identify associations of potentially mineralized data. Each method is further detailed in the following **Chapters 3, 4 and 5**. Beyond the theoretical developments and corresponding tests, this thesis work aims at producing an effective tool to be used for prospective mining. It results in a code made in Python language to automate the processing as much as possible and allowing to integrate large amount of map data layers including geological maps of different scales.

In order to test the robustness and effectiveness of this new predictivity method, application has been performed on a case study as a proof of concept. This approach is here applied for the study of antimony mineralization along the Western European Variscan Range, spanning from the French Armorican Massif (AM) to the Central-Iberian Zone (CIZ) of the Iberian Massif (IM), in Portugal and Spain. A detailed presentation of the geological setting for each Massif is made in **Chapter 2**. A first case study on the AM is presented in **Chapter 6** to present the effectiveness of the Random Forest method to produce favorability maps from the Cell-Based Association (CBA) method, the precursor of the DBA method (see further details below). Then, applications of the DBA / RF / HCPC methods along the Ibero-Armorican area, at different scales, is presented in **Chapter 7**. A synthesis and a discussion on the developments and results obtained from our new approach is made in **Chapter 8** as well as insights about future developments, improvements and applications.

The chapters of this thesis are organised as follows :

- **Chapter 1** presents a brief introduction to mineral resources and MPM as well as giving the frame of this thesis work and its main objectives.
- **Chapter 2** shows the geological and metallogenic context within the studied areas, describing in more details the actual knowledge about the targeted substance (antimony) and its relationship with geological environments within the AM and CIZ. This set a frame of reference for those relationships that can be used to evaluate the result of the inverse methodology developed during this thesis. This inverse methodology is described in the following **Chapter 3, Chapter 4 and Chapter 5**.
- **Chapter 3** shows the development of a method allowing to integrate the geological and structural spatial data into local geological environments. The development of this method called DBA is described, starting with a presentation of its predecessor, the CBA, and its limitation that the DBA aimed at overcoming.
- **Chapter 4** describes the following stage of this methodology, with the creation of prospective maps to constrain the likelihood of the local geologic environments and their geological features, produced with the method described in **Chapter 3**, to be linked to mineralization. Predictive methods used in previous works are first described, then the application of a ML RF supervised classification on the CBA and DBA geological environments is shown in details as well as the optimization and validation tools used in the frame of this thesis.
- **Chapter 5** displays the post-processing method, used in the final stage of the methodology, to identify the most favorable environments using the predictive maps produced by the RF method

shown in **Chapter 4**. This method synthesizes them into fewer “key” mineralized environments. The unsupervised HCPC classification method is described in details in this chapter, as well as the processing of the predictive maps to identify the most favorable environments, using statistical analysis.

- **Chapter 6** presents a first proof of concept of this methodology, focusing on the production of prospectivity maps. Using the CBA associated with the RF methods to produce a favorability map of antimony within the AM, this first application aims at displaying the advantages of using a ML based method with the CBA or the DBA methods to perform the prospectivity mapping rather the methods previously used in combination with the CBA.
- **Chapter 7** presents a more detailed application of the methodology developed during this thesis, with the combined use of DBA, RF and HCPC to study antimony mineralization in the AM and the CIZ. For each study area, the application of the DBA, RF and HCPC is described, as well as the dataset used. Each unique favorable environments produced at the end of this process is described in details and compared to pre-existing knowledge outlined in **Chapter 2**.
- **Chapter 8** synthesises and discusses the results of the previous applications and developments. It also presents new development path to enhance this methodology and discuss other potential applications of this methodology.

Chapter 2

Ibero-Armorican geological and metallogenic context

*This **Chapter 2** presents the geological and metallogenic framework of the study area, giving more insight about the targeted substance (i.e. antimony) and its relationship with the geological context within the study areas first presented in **Chapter 1**. This **Chapter 2** set a general framework of reference to evaluate the results shown in **Chapter 6** and **Chapter 7**, where the applications of the inverse methodology, described in the following **Chapter 3**, **Chapter 4** and **Chapter 5** is displayed. In this **Chapter 2**, the metallogenic and geological context of the study area is described, first from a global point of view, starting with the definition of an orogenesis and its relationship with the formation of mineralization. A short presentation of the Variscan Range follows this section, before the presentation of both the Armorican Massif (AM) and the Central Iberian Zone (CIZ) of the Iberian Massif (IM). For each, the geological setting is described, as well as the main resources and associated metallogenic events. Further details are given about the main antimony deposits and their genesis in each area.*

2.1 Orogen and mineralization

2.1.1 Orogenesis

The orogeny is a period of relief formation [Gilbert, 1890] linked to the formation of mountain systems by convergence and accretion between two continental margins [DiPietro, 2018]. It results in a penetrative deformation of the earth's crust, associated with phases of metamorphic and magmatic activities of limited duration and on restricted areas, generally spatially linear [Dennis, 1967]. Many hypotheses have been developed to conceptualize the formation of orogens, but it is now commonly accepted that the formation of orogens is related to the movement of tectonic plates [Dewey and Spall, 1975], and that they are essentially located at the plate margins, in convergence zones between two plates [Wilson, 1966, Dewey, 1969, Brown, 2009]. We can classify the orogens according to the nature of the lithospheres implied in these convergence zones [Cawood et al., 2009] (**Figure 2.1**).

If the collision takes place between two continental plates initially separated by an ocean, we can speak of a collision orogen [Wilson, 1966, Dewey, 1969, Brown, 2009, Burg and Ford, 1997]. This corresponds

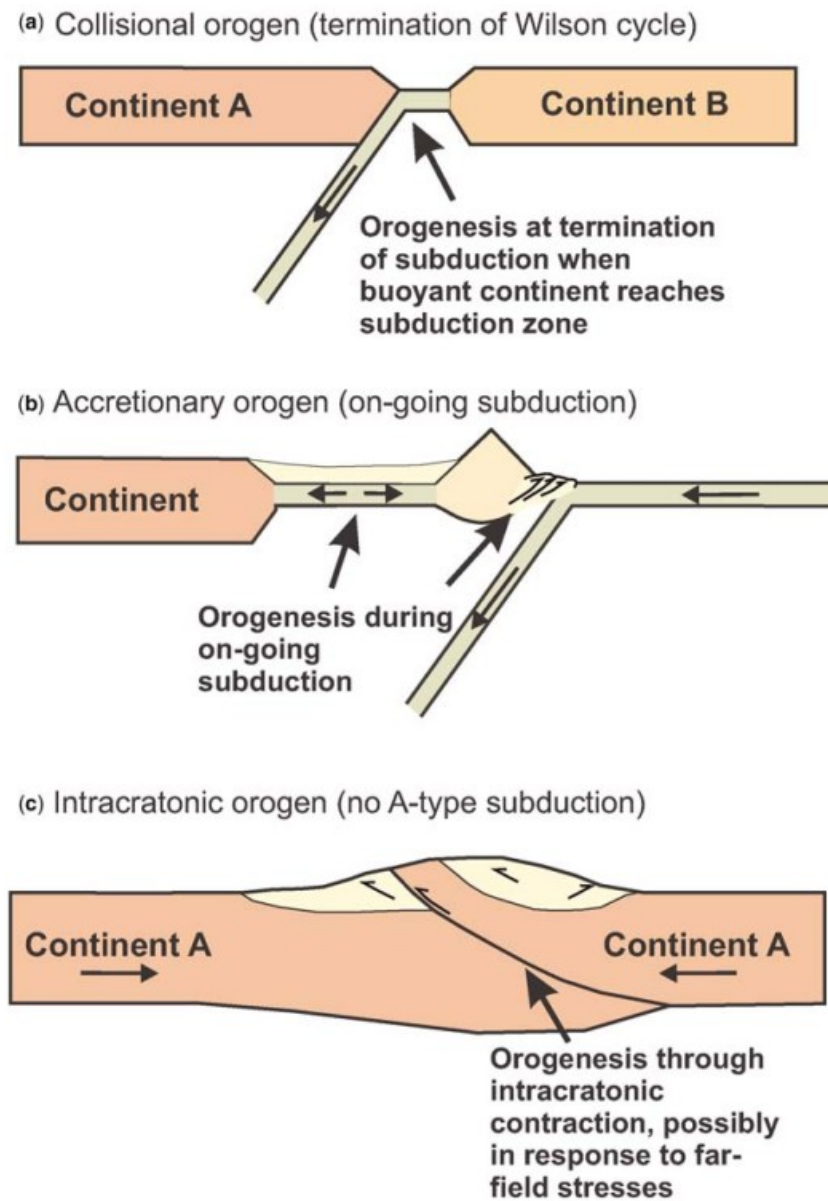


Figure 2.1: Schematic cross-sections through (a) collisional, (b) accretionary and (c) intracratonic orogens from [Cawood et al. \[2009\]](#).

to, for instance, the Variscan, Alpine, Himalayan, Appalachian and Cadomian orogenies [Cawood et al., 2009]. The development of this type of orogen is described in 5 main steps (**Figure 2.2**). First, the oceanic plate separating the two continents enters in a first stage in subduction under one of the continental plates (**Figure 2.2.1**). Relics of this oceanic plate, strongly metamorphosed, can be found in the orogen with the presence of ophiolitic complexes, formed from basic rocks of the oceanic crust metamorphosed in High Pressure (HP) and Low Temperature (LT) condition [Berger et al., 2010, Paquette et al., 2017]. The two continental plates then enter in subduction (**Figure 2.2.2**), with part the subducted continental slab undergoing a HP/LT metamorphism. This subduction will last some times before the continental slab rises up, forming nappe stacks and exhumation of HP rocks (**Figure 2.2.3**). Later intracontinental evolution (**Figure 2.2.4**), results in crustal thickening and the deformation of the lithosphere associated with Barrovian metamorphism (Low Pressure LP / High Temperature HT) [Rubio Pascual et al., 2016, Dias da Silva et al., 2017] and magmatism from crustal melting [López-Moro et al., 2018, Sawyer, 1994, Brown, 1994, Vanderhaeghe and Teyssier, 2001]. The final stage is the post-orogenic extension (**Figure 2.2.5**) with the gravitational collapse of the orogen.

Orogenesis can also occur between a continental lithosphere and an oceanic lithosphere or between two oceanic lithospheres, forming an accretionary orogen, such as the Andes Mountains or the Pacific Orogenic Belt. The continuous subduction of an oceanic lithospheric plate under another lithospheric plate leads to a thickening of the lithosphere with the formation of an accretionary prism, as well as magmatism and subduction metamorphism (HP/LT) [Cawood et al., 2009].

Finally, some orogens can develop within continental plates, far from the active margins of the main lithospheric plates, forming intracratonic orogens, as in central Australia with the Petermann orogen [Raimondo et al., 2010]. Because of their distance from the active margins, these orogens are more complex to interpret. It would seem that this type of orogen may be related to deformations caused by the two previous types of orogens or to their structural heritage [Cawood et al., 2009] (i.e. Moroccan High Atlas linked to an extensive basin, Tian Shan Belt from the Tertiary reactivation of Paleozoic structures [Chen et al., 2011], Pyrenees Mountains with overly stretched continental plate). In the case of the Petermann orogen, the structural, metamorphic and magmatic (crustal melting) characteristics seem similar to those of a collision orogen [Raimondo et al., 2010].

The topography created by the orogeny is bound to disappear with time, when the stresses causing the compression and thickening of the lithosphere dissipate [Dewey, 1988, Fernández and Pereira, 2016, Rey et al., 2001]. Two main forces will then lead to the disappearance of the relief: erosion, which removes material at crust's surface and allows the crustal root of the orogen to rise up and the gravitational collapse of the chain, leading to its disintegration through ductile and fragile deformations of the lithosphere [Jadamec et al., 2007, Rey et al., 2001].

2.1.2 Orogenic processes and mineralization

The formation of an orogen is associated with numerous geological processes. Thus, many tectonic stresses are present throughout the formation of the orogen, leading to the formation of faults or shear zones networks of various kinds [Arthaud and Matte, 1975, Balda et al., 1995, Cochelin, 2016, Ponce de Leon and Choukroune, 1980, Marques et al., 2002, Mateus and Noronha, 2010]. Those stresses are various and depend on the temporal and spatial evolution of the orogen. Compressive deformation occurs as a result of the collision between the continental plates, but extensive deformation can also occurs in

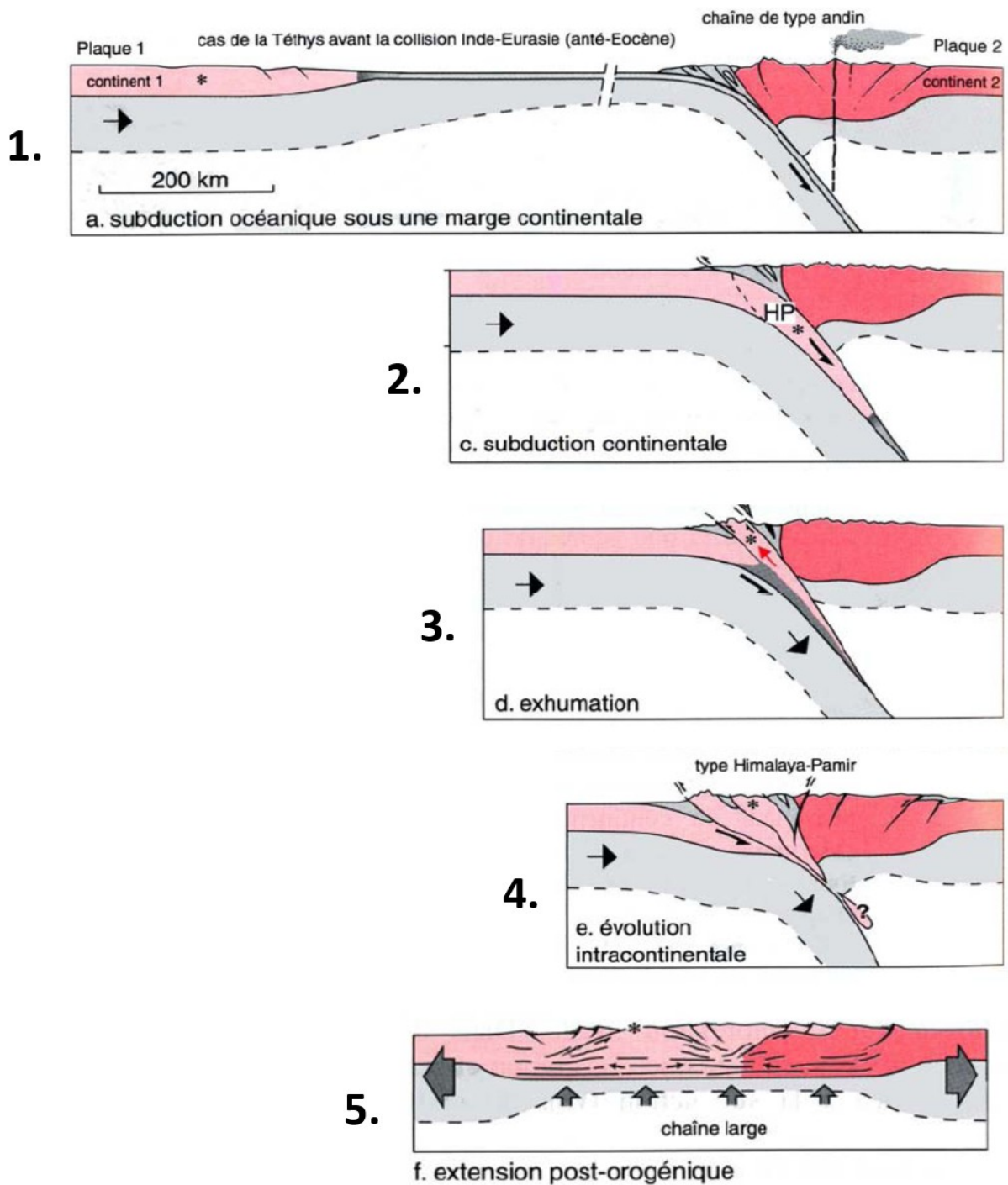


Figure 2.2: Orogenesis during the Phanerozoic, in 5 phases. 1. Oceanic subduction under continental margin. 2. Continental subduction. 3. Continental Collision s.s. (Exhumation of HP rocks). 4. Intracontinental evolution (thickening $\pm$ syn-orogenic extension). 5. Post-orogenic extension.

parts of the orogen that were already thick or that reaches a mechanical equilibrium with over-thickened zones. Transcurrent deformation is also possible, as inherited structures are also reactivated and are rarely strictly perpendicular to the compression main direction [Holdsworth et al., 2001]. These deformations can be ductile or brittle, the rheological properties of the affected rocks varying according to their nature and their structural level in the lithosphere (T°/P variations). Changes in temperature and pressure are also at the origin of metamorphic processes [Ballèvre et al., 2014] .

In a collisional context, the fault system is created at first in a compressional regime, with reverse faults, thrust sheets and shears. Thrust sheets are formed when allochthonous nappes are thrust over autochthonous formations (e.g. Galicia Trás-os-Montes or GTMZ in Spain and Portugal) [Rubio Pascual et al., 2013, Ballèvre et al., 2014, Matte, 2007]. When transitioning from a thickening phase, in a compressional regime, to a collapse stage in an extensional regime, normal or shear faults are formed with sometimes reactivation of previous shallowly dipping thrusts [Dewey, 1988, Jadamec et al., 2007, Fernández and Pereira, 2016, Bulois et al., 2018, Rey et al., 2001]. The thickening phase is also accompanied by metamorphism and magmatic activity, with the increase in temperature and pressure.

Before the collision, during the convergence, the subduction of the oceanic slab leads to the metamorphism of the oceanic crust and the passive margin part of the continental one in context of HP and LT. This can result, in particular, in the formation of ophiolitic complexes, which are relics of the oceanic slab during this subduction stage. Some occurrences can be found across the orogen if they are preserved [Arenas et al., 2016, Cawood and Suhr, 1992]. The continental crust undergoes different metamorphism along the evolution of the orogen. The first stage is a HP/LT metamorphism with the formation of the orogenic prism [Cochelin, 2016] and nappe-stacking resulting in the thickening of the continental crust. The second stage is a prograde metamorphism with the increase of the temperature in the deepest and most thickened part of the continental crust, named crustal root [Rubio Pascual et al., 2013]. The increase in temperature can reach partial melting of the crustal root, leading to the formation of migmatites and granites from the extracted melts [López-Moro et al., 2018].

All these processes are associated to fluids circulation, which can be of magmatic origin [Hedenquist and Lowenstern, 1994], resulting from the differentiation of magmas (granitic plutons) [Audétat and Edmonds, 2020], or of metamorphic origin, from the metasomatism of minerals [Yardley and Cleverley, 2015, Cox et al., 1986]. Hydrothermal fluids also play a role when magmatic and metamorphic fluids are dried, as intrusive granites can work as thermal engines for the circulation of fluids (e.g. meteoric water reaching a shallow granite intrusion) [Lamy-Chappuis et al., 2020, Huber et al., 2012]. Fluids play a major role in the formation of mineralization, as a medium of transport for some elements (**Figure 2.3**). They can be initially rich in elements of economical interest or become enriched in these elements as they interact with fertile rocks during their transport through the continental crust [Heinrich, 2007, Lecumberri-Sanchez et al., 2017, Ding et al., 2018]. These fluid circulations are facilitated by the networks of faults, fractures and shear zones produced during the different deformation phases of the collision [Kerrick, 1986]. This network can play as a trap for enriched fluids if the physical-chemical properties are favorable (e.g. changes in temperature, fluid/rocks reaction, fluid/fluid reaction ...). Mineralization can thus form through diverse physico-chemical processes inside these structures or in their surroundings [Cathelineau et al., 2011, Hartai et al., 2016].

In a geodynamical context of collision, many magmatic and hydrothermal mineral deposits can occur [Groves and Bierlein, 2007]. Before the collision, during the convergence, Arc-related magmatic and hydrothermal events linked to the subduction of the oceanic crust can lead to the formation of porphyry

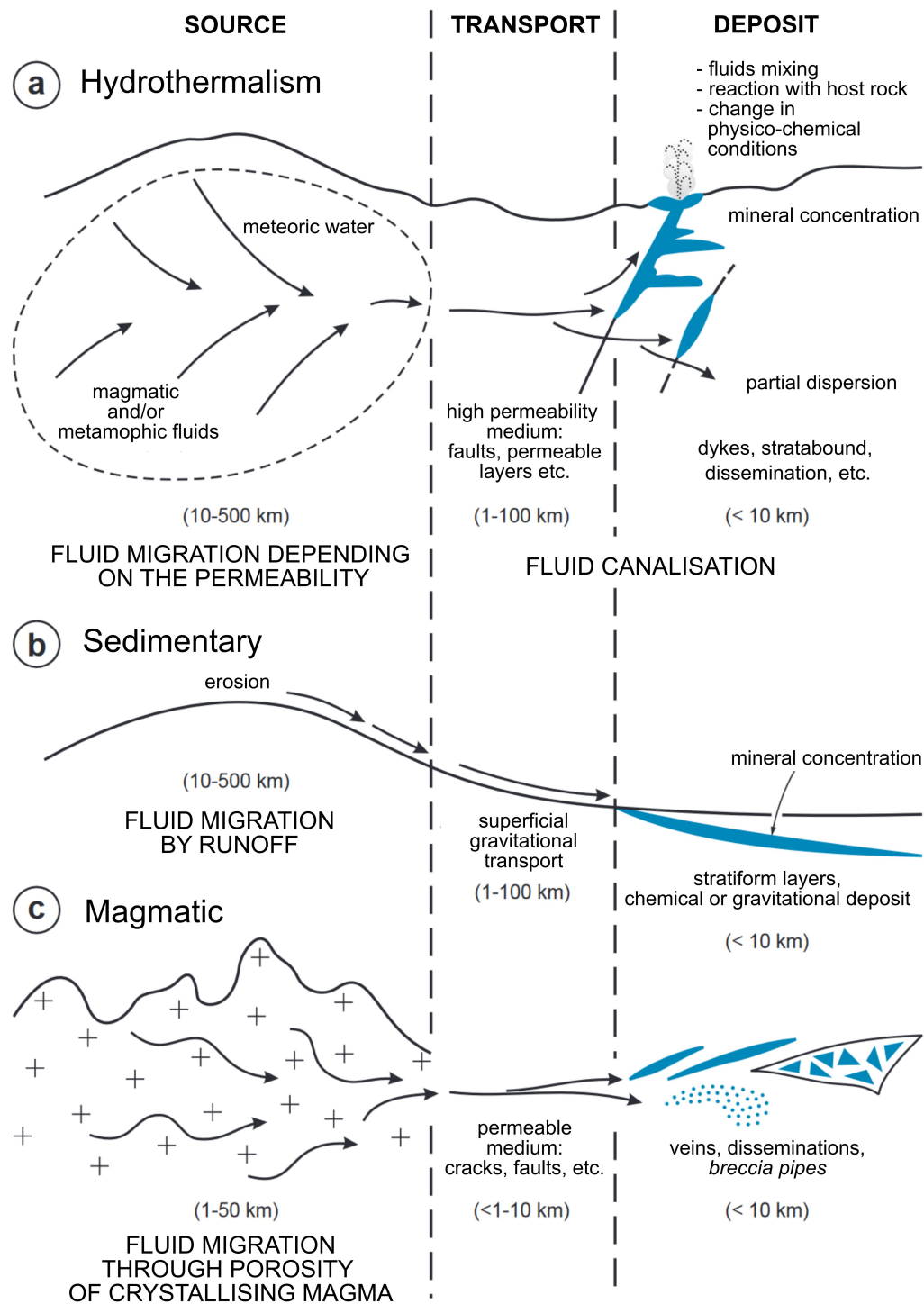


Figure 2.3: Source-Transport-Deposit sketch, fluids play a critical role in the formation of ore deposit in multiple environments (modified from Marcoux and Jébrak [2008]).

or epithermal deposits [Ford et al., 2015, Billa et al., 2004, Seedorff et al., 2005]. Backarc environments, also related with the subduction of the oceanic crust, can result in Volcanogenic Massive Sulfide (VMS) type deposits above the subduction [Arias et al., 2011] or in some case the formation of intrusion-related deposits within the continental crust [Couto and Roger, 2017]. The main deformation events occurring during the collision also favor the setting of gold and base metals in the most deformed part of the orogen [McMillan et al., 2019, Ashley et al., 2006, Groves et al., 1998]. Finally, the erosion of the relief created by the collision and the circulation of meteoric fluids also produce deposits within the foreland basins with erosion-related deposits (i.e. placers and paleoplacers [Roy et al., 2018]) and late hydrothermal deposits [Ballouard et al., 2018].

2.2 The Variscan Range

The Variscan Range developed during the Paleozoic (480 - 250 Ma) and extends in Western Europe from the Iberian Peninsula to Central Europe, through France, Germany and the Czech Republic (**Figure 2.4**). This range results from the collision between two major continental lithospheric plates: Gondwana to the South and Laurussia to the North, itself resulting from the collision between two continental domains, Laurentia and Baltica, separated before its closure by the Iapetus Ocean [Martínez Catalán et al., 2021] (**Figure 2.5**). This collision between Gondwana and Laurussia takes place during several phases, the two domains being separated by microplates detached from the Gondwanian margin: the Avalonia microplate, separated from Laurentia by the Iapetus, and the Armorica microplate, separated from Avalonia by the Rheic Ocean and from Gondwana by the Galico-Moldanubian Ocean. The Ibero-Armorican ensemble formed a micro-plate detached from the North Gondwana margin during Cambrian and mostly Ordovician times. The orogen developed from late Devonian (360 Ma) to the end of Carboniferous (300 Ma) [Ballèvre et al., 2009], during the closing of the Rhaetian Ocean and the Galico-Moldanubian Ocean [Ballèvre et al., 2013, Franke et al., 2017, Matte, 2001].

Currently, the Variscan Range outcrops in several massifs in Western Europe: in the Iberian Massif (IM), the Pyrenees, Cornwall, the Armorican Massif (AM), the French Massif Central (FMC), the Corso-Sardinian block, the Rhenish Massif and the Bohemian Massif (BM). These different massifs can be divided into several domains according to their affinities with the different continental plates or microplates, including their former passive margins. This first order structure is still discussed today and can vary from one author to the other [Ballèvre et al., 2013, Catalán, 2011, Franke et al., 2017].

The Western Variscan Range shows an arcuate structure from the Iberian Peninsula to the French Armorican Massif, known as the Ibero-Armorican Arc (IAA) [Matte, 1991]. In the Iberian Peninsula, the IAA shows NW-SE trends to N-S trends from the SE to the NW of the Peninsula. In the AM, the main structures (e.g. SASZ and NASZ) trend E-W [Quesada and Oliveira, 2019, Dias et al., 2016, Choubert, 1935, Carey, 1955, 1958, Cogné, 1967, 1971, Bard et al., 1971, Lefort, 1979, Perroud and Bonhommet, 1981, Burg et al., 1987].

The inner part of the IAA is the Cantabrian Arc (CA), composed of the Cantabrian Zone (CZ) [Pérez-Estaún et al., 1988, Pastor-Galán et al., 2015]. This CZ is considered to be the foreland thrust and belt of the IM. The West Asturian Leones Zone (WALZ) is considered as the transition between this foreland part of the belt and the internal part [Pérez-Estaún et al., 1991]. The CIZ and its counterpart, the AM, are considered as the Axial Domain of the Western Variscan Range, with abundant granitic plutonism and low to high-grade metamorphism [Quesada and Oliveira, 2019].

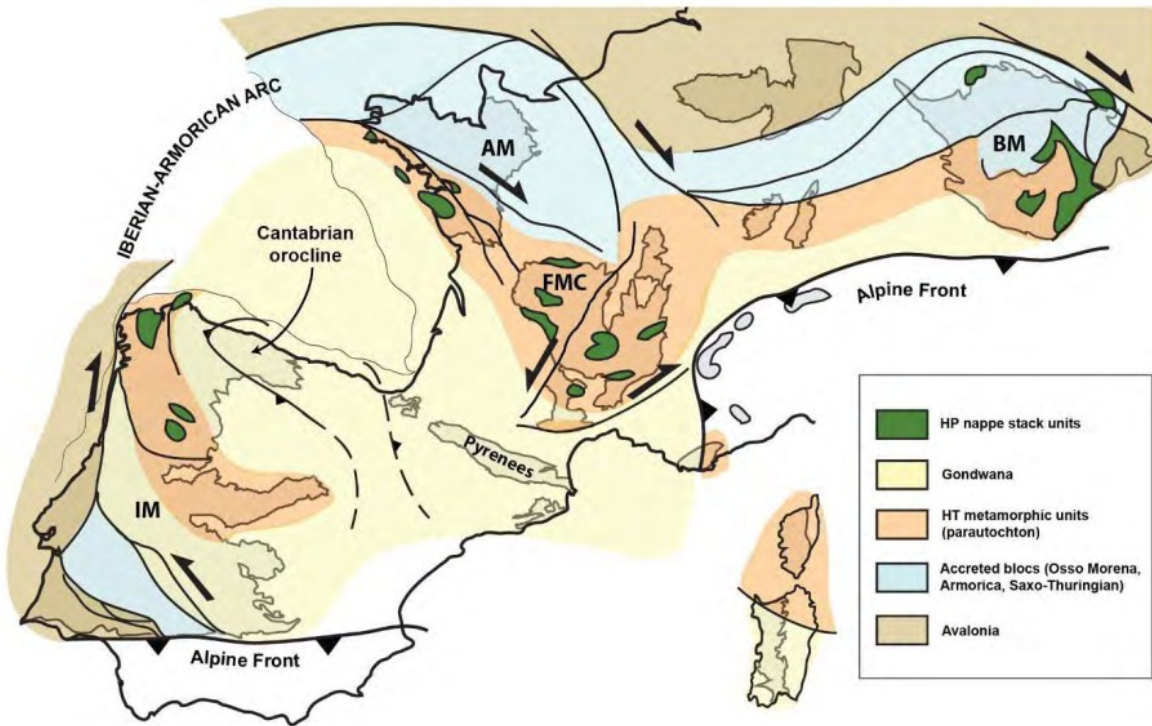


Figure 2.4: Simplified structural map of the Western Variscan belt (from [Cochelin \[2016\]](#)). IM: Iberian Massif, AM: Armorican Massif, FMC: French Massif Central, BM: Bohemian Massif.

The existence of another Arc, the Central-Iberian Arc (CIA), in the South of the IAA, is also discussed [[Dias et al., 2016](#), [Dias and Ribeiro, 1995](#), [Quesada and Oliveira, 2019](#)]. The CIA would be located within the IAA and would mainly cover the CIZ, while the CA would form a S-shaped arc from the WALZ to the Southern part of the CIZ [[Ribeiro et al., 2007](#)].

In addition to the complexity of its orogenic cycle, the Variscan Range is also distinguished by the numerous mineralization present within the outcropping massifs. The substances present along the Variscan Range are diverse and generally related to different phases of formation of the orogen.

2.3 Armorican Massif

2.3.1 General setting

The Armorican Massif (AM), in northwestern France, is part of the Variscan Range, which results from a collision cycle taking place from late Devonian until the end of Carboniferous, between the Laurussia Plate, the Gondwana Plate and the Iberian-Armorican Microplate. The AM groups together all the Proterozoic and Paleozoic formations outcropping in western France. It is bounded to the East by the Paris Basin, to the South by the Aquitaine Basin, to the North by the English Channel and to the West by the Bay of Biscay. As shown by geophysical data, the Paleozoic basement continues under these late coverings but is not outcropping [[Ballèvre et al., 2013](#), [Baptiste, 2016](#)].

The AM is subdivided into four major domains (**Figure 2.6**), according to the location of large Carboniferous crust-scale shear zones [[Gumiaux et al., 2004b](#)], the intensity of deformation and the grade of

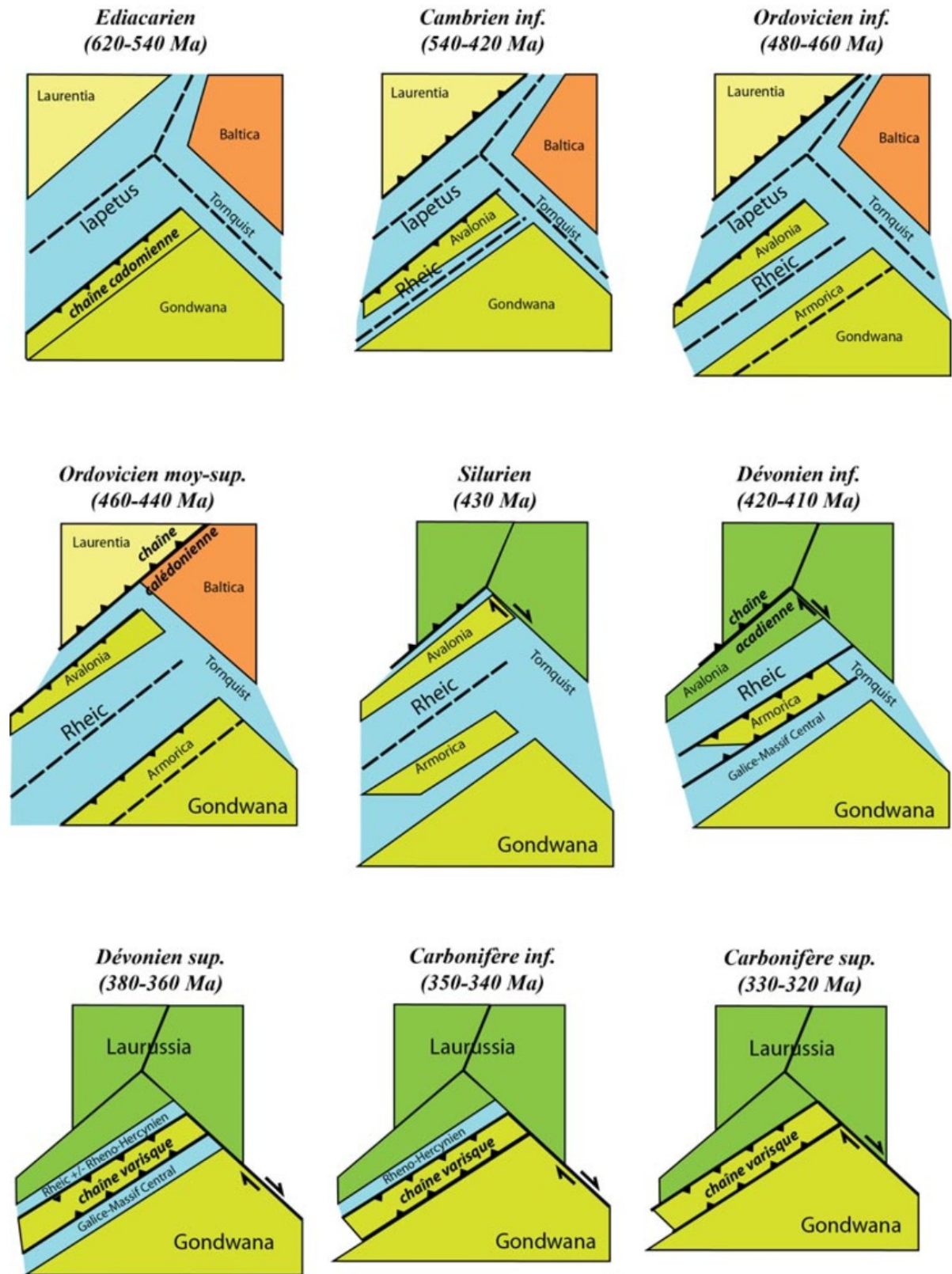


Figure 2.5: Simplified sketch of the different orogenesis between Laurentia and Gondwana plates (from Ballèvre et al. [2013]).

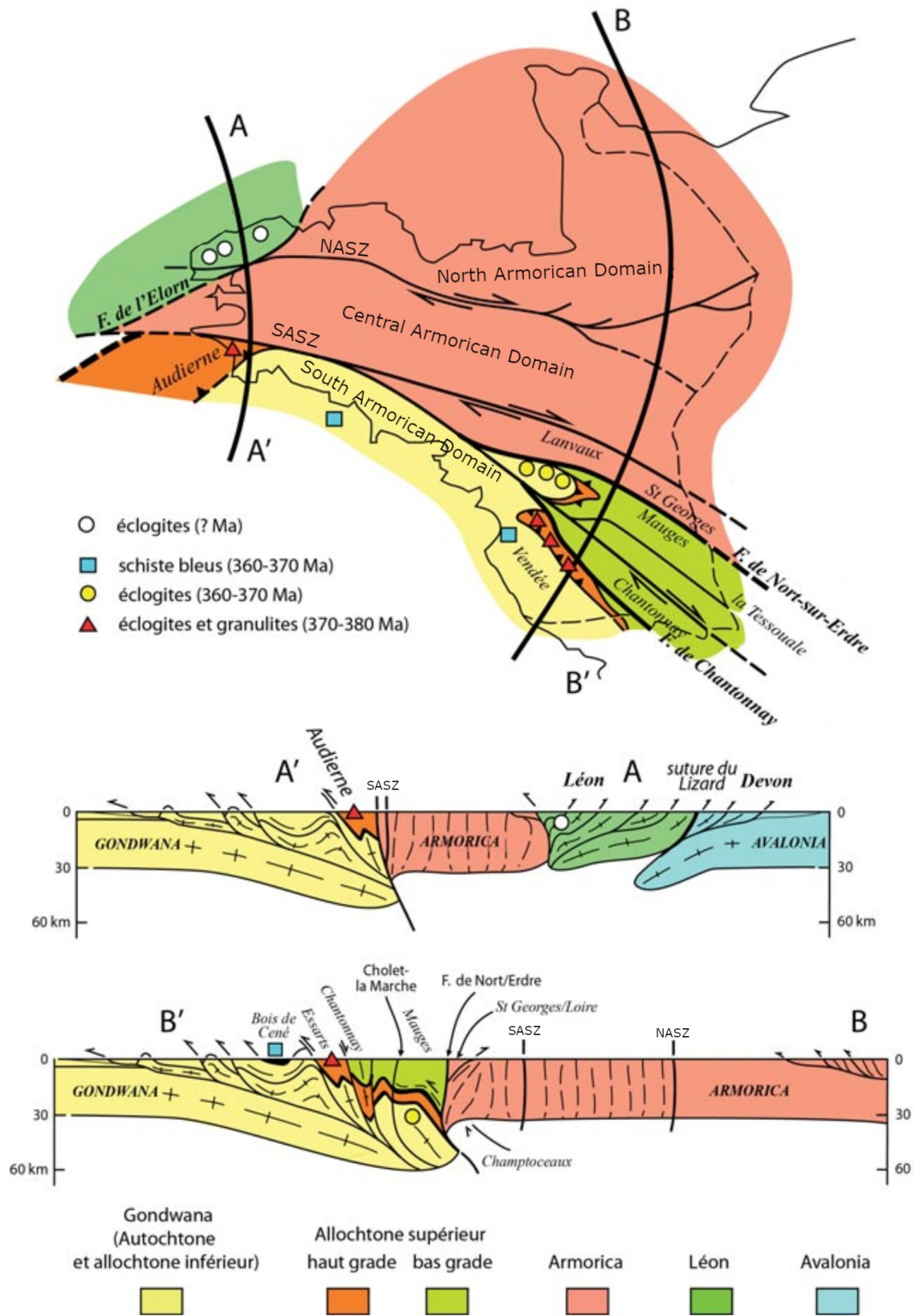


Figure 2.6: Synthetic map of the Armorican Massif with simplified cross-sections for the collision (beginning of the Late Carboniferous), modified from [Ballèvre et al. \[2014\]](#).

metamorphism [Ballèvre et al., 2013, 2009]:

1. The South Armorican Domain (SAD) is located south of a crust to lithosphere-scale structure: the South-Armorican Shear Zone (SASZ), a dextral simple shearing zone. It is composed of a stack of Armorican and Gondwanian units, with multiple phases of synschistous folding with a main deformation phase during the Late Carboniferous. From a structural point of view, the early compressive deformation of the Variscan orogeny is mainly preserved within the SAD [Ballouard et al., 2017b], this domain showing evidences of late extension, syn- to post-compressive (e.g. formation of migmatites domes [Gapais et al., 1993, 2015]). Main structures results from syn- to late-orogenic extension with both flat-lying pervasive foliation developed and large extensional shears zones showing EW trending stretching of the crust during late Carboniferous. The SAD is marked by metamorphic units and corresponds to some internal segments of the orogen, ranging in intensity from HP to Low to Moderate Temperature rock units (eclogite, blue schist and migmatite). These units are preserved from the continental subduction phase and collision of the North Gondawana margin with the Armorican micro-continent. Some of those MP-HT units form low amplitude domes below the preserved stacked nappe units. These metamorphisms and structures result from the pervasive partial melting of the crust occurring during mid-Carboniferous and their up-rising during late orogenic extension, around 300 Ma [Ballèvre et al., 2009]. As such, the Variscan magmatism is quite abundant with many "two-micas" granites along the shear zones.
2. The North Armorican Domain (NAD), north of the North-Armorican Shear Zone (NASZ). This domain is mostly made of geological units developed during the Cadomian orogeny (Cadomian basement of the Variscan Massif). If considering only the Variscan orogeny, there is very few traces of metamorphism and only of lower grade. There is also no trace of syn-orogenic magmatism, but there is trace of post-orogenic magmatism known as "pink granites". This domain also contains most of the outcropped Cadomian basement and units of the AM. It behaved mostly like a rigid block during the Variscan collision.
3. The Central Armorican Domain (CAD), is bounded by the two dextral crustal-scale wrench zones crossing the area: the North Armorican Shear Zone (NASZ) and the South Armorican Shear Zone (SASZ). It is mostly composed of a Proterozoic to Palaeozoic sedimentary series showing limited metamorphism beside localized contact metamorphism around late granitic intrusions. During the Carboniferous, the whole CAD underwent a N123°E trending regional-scale dextral simple shearing producing E-W trending upright folds and sub-vertical concomitant pervasive cleavage [Gapais and Le Corre, 1980, Gumiaux et al., 2004b]. Local and map-scale structural analysis of the granite plutons located along the shear zones show that the intrusions are syn-tectonic and mark a regional-scale shearing during Carboniferous within this external domain.
4. The Leon Domain, on the Northwestern part of the AM. It is composed of strongly metamorphosed units of disputed origin, with a main deformation event of Early Carboniferous to Late Devonian age and many phases of synschistous folds. This domain is covered by abundant syn-orogenic magmatic units, with also the post-orogenic "pink granites" similar to those found in the NAD.

2.3.2 Geological units and related events

The lithological setting of the AM is described in [Ballèvre et al. \[2009, 2013, 2014\]](#), [Franke et al. \[2017\]](#) and the main events associated with it are also described in [\[Ballèvre et al., 2009\]](#). A simplified lithological map is presented in **Figure 2.7**.

The oldest rocks of the AM are Icartian in age (Lower Proterozoic 2200-1800 Ma) and are located from Lannion to Cherbourg areas [\[Auvray et al., 1980a, Samson and D'Lemos, 1998\]](#); but they only display as enclaves within earlier intrusive plutons. The AM comprises units from the Cadomian Range (750-540 Ma), a subduction belt. It includes 4 large units located in the NAD [\[Egal et al., 1995, Chantraine et al., 2001b\]](#): The Trégor-la Hague unit, The Saint-Brieuc unit, The Saint-Malo unit and the Fougères unit (**Figure 2.7**). In the CAD, the Cadomian Range is represented by terrigenous detrital Brioverian formations (580-540 Ma) located from the Douarnenez Bay to the Rouez area. The deformation associated with the Cadomian orogeny is very weak or absent there [\[Gumiaux, 2003\]](#). In the SAD, some Proterozoic rocks that are part of the continental crust that underwent the Cadomian orogeny can be encountered, such as the Mauges and Roc-Cervelle units [\[Bouton and Camuzard, 2012\]](#). Other units, deformed during the Variscan orogeny, preserve pre-Variscan relics. The Proterozoic basement, deformed by the Cadomian orogeny, forms the base of the Paleozoic sedimentary formations.

From the Cambrian to the Ordovician a continental extension phase occurred, with the fragmentation of the Northern margin of the Gondwana plate and the formation of the Avalonia microplate with the opening of the Rheic ocean. This phase is associated to detrital sediments deposit. In the NAD, Cambrian detrital formations unconformably overlie the Cadomian basement, followed by Lower Ordovician reddish-purple sandstone and clay series derived from variable sedimentary environments [\[Suire et al., 1991\]](#) and locally overlain by volcanic rocks [\[Auvray et al., 1980b, Galerne et al., 2006\]](#). In the SAD, the Cambrian series are composed of fossiliferous marine series and thick detrital series overlying the Proterozoic basement. In the CAD, these Cambrian formations are absent or not yet identified, as Brioverian sedimentation may have continued [\[Guerrot et al., 1992\]](#). This difference suggests that the NASZ may have been active since the Cambrian. Ordovician formations in the CAD are composed of "red series" of sandstone or silty levels with a basal conglomerate. The thickness of this formation is quickly varying within hemi-grabens, delimited by NE-SW normal faults [\[Ballard et al., 1986, Brun et al., 1991, Dabard and Simon, 2011\]](#).

Some magmatic activity is also linked to the Cambro-Ordovician rifting and formation of the Rheic Ocean [\[Ballèvre et al., 2013, Paquette et al., 2017\]](#). In the NAD, it is mostly Cambrian magmatism [\[Le Gall, 1993\]](#) with calc-alkaline and sometimes peraluminous series. During the Ordovician, magmatism is almost absent in the NAD and is found interbedded in the "red series" [\[Auvray et al., 1980b, Galerne et al., 2006\]](#). In the CAD, this Ordovician magmatism is also interbedded in the Ordovician sedimentary formations. In the SAD however, it is at the origin of orthogneiss protoliths (480 Ma) and volcanic porphyroids [\[Ballèvre et al., 2013\]](#). The geochemical diversity of this Ordovician magmatism could be due to the generation of these magmas by melting of different materials, such as crustal materials or more basic asthenospheric materials. The composition of the magma will thus vary according to the source material [\[Ballouard et al., 2018\]](#).

These formations are then covered by the "Armorican sandstone" of Lower Ordovician age, white sandstones of variable thickness, sometimes with levels of iron ore or heavy minerals (zircon and monazite). The thickness of this formation largely varies from tens to hundreds of meters [\[Noblet, 1984, Durand,](#)

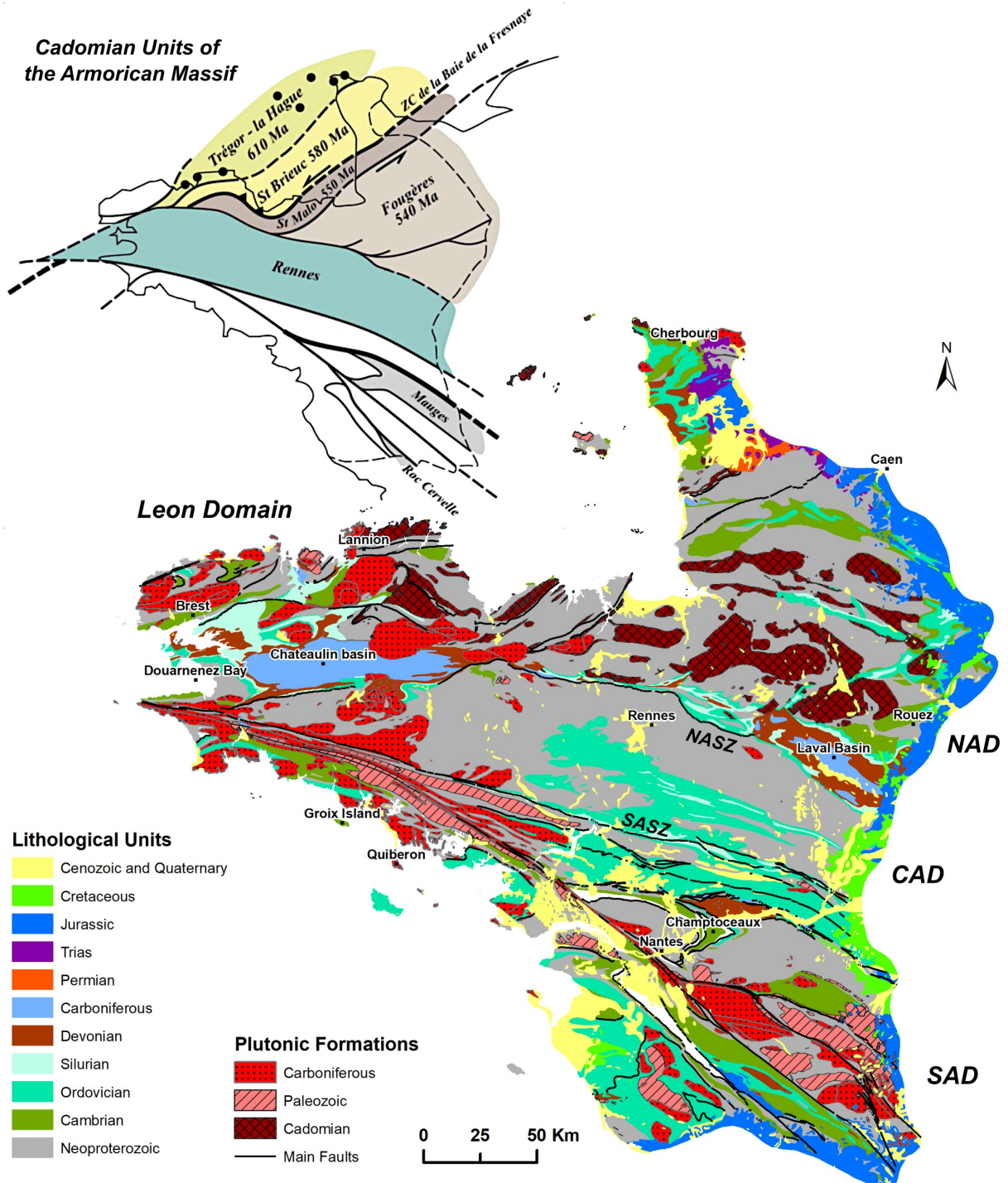


Figure 2.7: Simplified Geological map of the Armorican Massif (modified from the 1M geological map of France, BRGM). Only the age of the protolith is displayed here. Post-Variscan formations on the East and South of the Massif are part of the Paris Basin and the Aquitaine Basin respectively. The top-left insert shows the various Cadomian units of the Armorican Massif (modified after Ballèvre et al. [2013]). NAD: North Armorican Domain; CAD: Central Armorican Domain, SAD: South Armorican Domain; NASZ: North Armorican Shear Zone; SASZ: South Armorican Shear Zone.

1985]. In the Middle Ordovician, marine sedimentation followed, covering most of the AM [Dabard et al., 2007, Vidal et al., 2011] with more clayey and silty facies and sandy levels from North to South. The sedimentary formations deposited during the Upper Ordovician are characterized by successions of sandy facies in the north of the AM, their levels decreasing towards the north and becoming more pelitic. Carbonate levels overlie these formations, sometimes associated with a volcanic episode with breccias and pillow flows [Juteau et al., 2007, Caroff et al., 2009]. At the end of the Upper Ordovician, glaciomarine facies were deposited as well as silty or sandstone facies.

Silurian is characterized by fine terrigenous and black shale facies, rarely carbonate. In the CAD and NAD, these are black clayey/silty sediments ("ampelites"), with numerous graptolites. In the SAD, we also find these ampelites associated to graphitic quartzites ("phtanites"). Farther South, in the Vendée region, this series has been intensely deformed. Some levels of the Silurian display abundant nodules often with pyrite ("schistes à sphéroïdes"). The Silurian ampelites served as a detachment level during the Variscan deformation due to their low mechanical strength, between the Lower Paleozoic formations and the Devonian series [Ballèvre et al., 2013].

Devonian corresponds to the onset of the convergence (**Figure 2.8**). Unlike previous sedimentary formations, the Devonian series highly vary in lithology across the AM [Ballèvre et al., 2013]. In the NAD and CAD, the Devonian is conformably lying over Silurian. Devonian formations are often incomplete with only the early Devonian formations being preserved. Those formations are detrital with a succession of araneous to argillaceous-calcareous formation and black-shales for the latest ones [Robardet et al., 1994]. In the SAD, Devonian formations are found in two locations : the Mauge Unit and at Villedée d'Ardin (Roc-Cervelle unit). Those formations show that most of the AM was covered by the sea water during the Devonian, with some emerged elevation linked to the beginning of the collision. The Devonian is also accompanied by the deformations associated with the subduction, which lasted until Early Carboniferous. Relics of this phase of HP-LT are visible along the SAD and the Leon Domain. Eclogites, such as the one included into the Champtoceaux nappes system, and blue schist like the one present inside the island of Groix are evidences of this HP-LT metamorphism [Ballèvre et al., 2013].

The continental collision s.s. occurs during the Carboniferous, causing the formation of lithological stacks and the thickening of the crust. In the NAD, the deformation is weak, the collision causing a folding of the Paleozoic formations, with a weak or unobserved schistosity. The deformation is more important in the North, following a N-S gradient. In the SAD, the collision leads to the formation of complex nappe stacks, with the superposition of allochthonous formations on autochthonous ones [Ballèvre et al., 2003] and the formation of the carboniferous dextral SASZ, with it's Northern and Southern branches and the Nord-sur-Erdre fault. Granitoids intrusions are mainly localised near the major shear zones, with most of them emplaced along the SASZ [Ballèvre et al., 2013].

The CAD shows a synschistous folding, with upright folds showing an E-W trending. Those folds reach regional scales as shown by the Rennes area. Those upright folds are associated with sub-vertical axial plane cleavage, with a sub-horizontal mineral lineation showing E-W stretching [Gapais and Le Corre, 1980]. This plane strain deformation is associated with a regional scale and pervasive dextral shearing deformation affecting a ≈ 90 km wide band [Gapais and Le Corre, 1980, Le Corre, 1978] also shown in the leucogranites north of the SASZ [Berthé et al., 1979]. The restoration of the Variscan strain within the CAD shows an overall N123 dextral shearing deformation on the CAD during the Carboniferous [Gumiaux, 2003, Gumiaux et al., 2004a].

The Devonian and early Carboniferous period is characterized by the emplacement of gabbros and

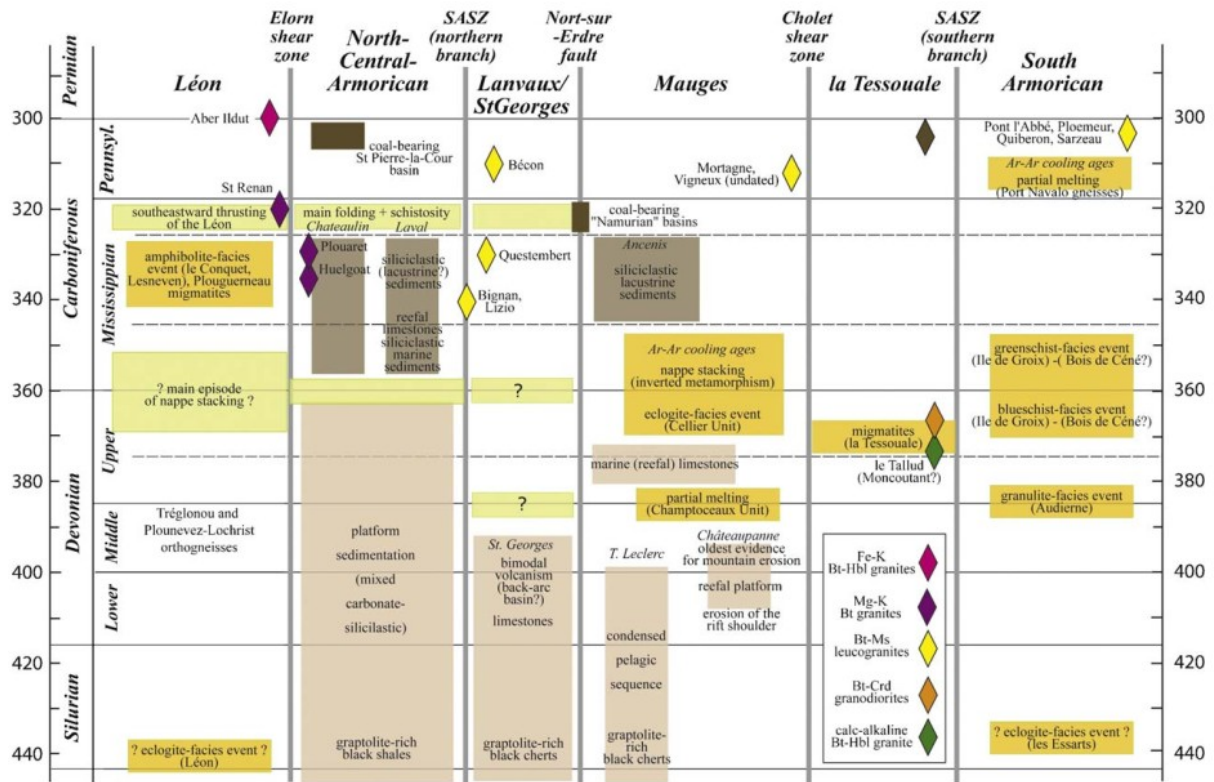


Figure 2.8: Summary of the main events recorded in the Armorican Massif during the Devonian and Carboniferous plate convergence. In pale brown the pre-orogenic sedimentary sequences events, in dark brown the synorogenic sedimentary formations, in pale yellow the metamorphic rocks and the lozenges are events of intrusive granitic rocks (from [Ballèvre et al., 2009]).

various granitoids [Capdevila, 2010], and by the intrusion of dykes around 360 Ma [Pochon et al., 2016]. The Late Carboniferous period (320-300) is marked by the development of granitoids and the exhumation of migmatites core complex [Gapais et al., 2015] linked to the crustal melting and the dextral shearing of the AM. Three main granitoids are differentiated depending on their composition [Ballouard et al., 2018]. The muscovite bearing granites are peraluminous granites located mostly along the SASZ and are of crustal origins [Ballouard et al., 2017b, Euzen, 1993, Tartèse and Boulvais, 2010, Ballouard et al., 2015, 2017a] with ages ranging from 317 Ma to 305 Ma [Ballouard et al., 2017b]. Cordierite-bearing granites are located on the Western part of the CAD and are derived from mantle sources [Ballouard et al., 2018], with ages around 315 Ma [Ballouard, 2016, Ballouard et al., 2017b]. Finally, the calc-alkaline belongs to the magnesio-potassic and ferro-potassic suites and were formed between 325 Ma to 300 Ma [Ballouard et al., 2018] from mantle sources [Peucat et al., 1984, Capdevila, 2010, Ballouard, 2016].

The sedimentation during the Carboniferous is characterized by the formation of multiple and independent basins in unconformity with the previous formations. The basins of Châteaulin, Ménez-Belair and Laval are located along the CAD. The basin of Ancenis is located along the northern border of the SAD. Those basins were formed through local subsidence areas. Those formations are composed of detritic rocks and carbonate ones [Ballèvre et al., 2013].

During Late Carboniferous, (320-300), i.e. during the end of the collision, the SAD shows an extensive regime of E-W trending [Gapais et al., 1993, Le Hebel et al., 2000, 2002] with the formation of shear zones and the intrusion of leucogranites. Brittle deformations also occur with tilted blocks showing an

E-W extension [Ballèvre et al., 2013]. In the CAD, the Late carboniferous deformation is linked to the intrusion of leucogranites along the SASZ and very localised there [Gumiaux, 2003].

2.3.3 Main metallogenic resources and events

The Armorican Massif (AM) contains a wide variety of resources. Metallic resources include Au, Sb, U, Fe, Pb, Zn, Ag, Sn, W, Cu, Ta, Nb, Mo (**Figure 2.9**). Industrial resources are also located in this Massif, with the andalusite and kaolinite deposits [Béziat and Bornuat, 1994]. The AM presents an important metallogenic interest, due to the presence of several metallogenic episodes [Chauris and Guigues, 1969, Marcoux and Chauris, 1994, Chauris, 1978, Hartai et al., 2016, Lescuyer et al., 1997]. Those metallogenic episodes and some associated mineralization are as follows (**Figure 2.10**):

1. Before the formation of the Variscan Range, mineralization related to the Cadomian orogeny were already present. They have been preserved in the northern part of the AM. In the Brioverian, volcanosedimentary and hydrothermal polymetallic sulphide deposits were formed in the Rouez deposit (Zn-Pb-Cu-Au) and also granitic intrusions with cassiterite-wolframite associations in the Montbelleux deposit.
2. Sedimentary deposits of Iron and Zircon-Titanium formed during the extensional sedimentation phase of the Ordovician, following the Cadomian orogeny. The main Iron deposits, along with the sedimentary Zr-Ti deposits, are located in the eastern part of the Massif, with the Segré district and the Soumont District.
3. Polymetallic sulfide deposits were formed during the Silurian, linked to the volcanic activities of a back-arc basin with the main deposit of La Haie Claire (Zn-Pb-Cu-Ag).
4. The Variscan orogeny induced the formation of various mineralization. These mineralization are related to the numerous granites' intrusions within the AM, with the circulation of magmatic and hydrothermal fluids, eased by the faults and shear zones crossing the Massif. Mineralization include deposits of Cassiterite and Wolframite veins (Abbaretz, Montbelleux, Saint Rénan), Gold veins (La Bellière, Lopérec), Antimony veins (La Lucette, Semnon, Quimper, Rosnoën), Lead-Zinc veins (Pontpean, Huelgoat, Plelauff), Tin-tantalum-niobium veins (Tréguennec) and Uranium deposits (La Commanderie, L'Ecarpière, Le Chardon). At the end of the Devonian [Lescuyer et al., 1997], massive polymetallic sulfide bodies also formed (Bodennec, La-Porte-aux-Moines and Menez-Albot). Finally, during the late-Variscan period, Molybdenum deposits linked to potassic metasomatism surrounding peraluminous intrusions were formed with the Beauvin deposit.
5. The recent works from Pochon [2017] and Pochon et al. [2019] have highlighted another metallogenic event around 360 Ma. This event is linked to mafic magmatism in the NAD and SAD, with a network of dykes and sills (dolerites). It resulted in the formation of Sb-Au deposits related to mafic dykes, such as Le Semnon, during the paroxysm of the HP-LT metamorphism in the SAD. This metallogenic event is thus probably linked to the subduction of the North Gondwana continental margin below Armorica with a clear temporal concordance between mafic magmatism, deposits and HP/LT pic metamorphism.

Later erosion and sedimentation of enriched rocks led to the formation of some deposits derived from the Variscan orebodies, like the Tin deposit of Saint Renan. Today, the substances and associations

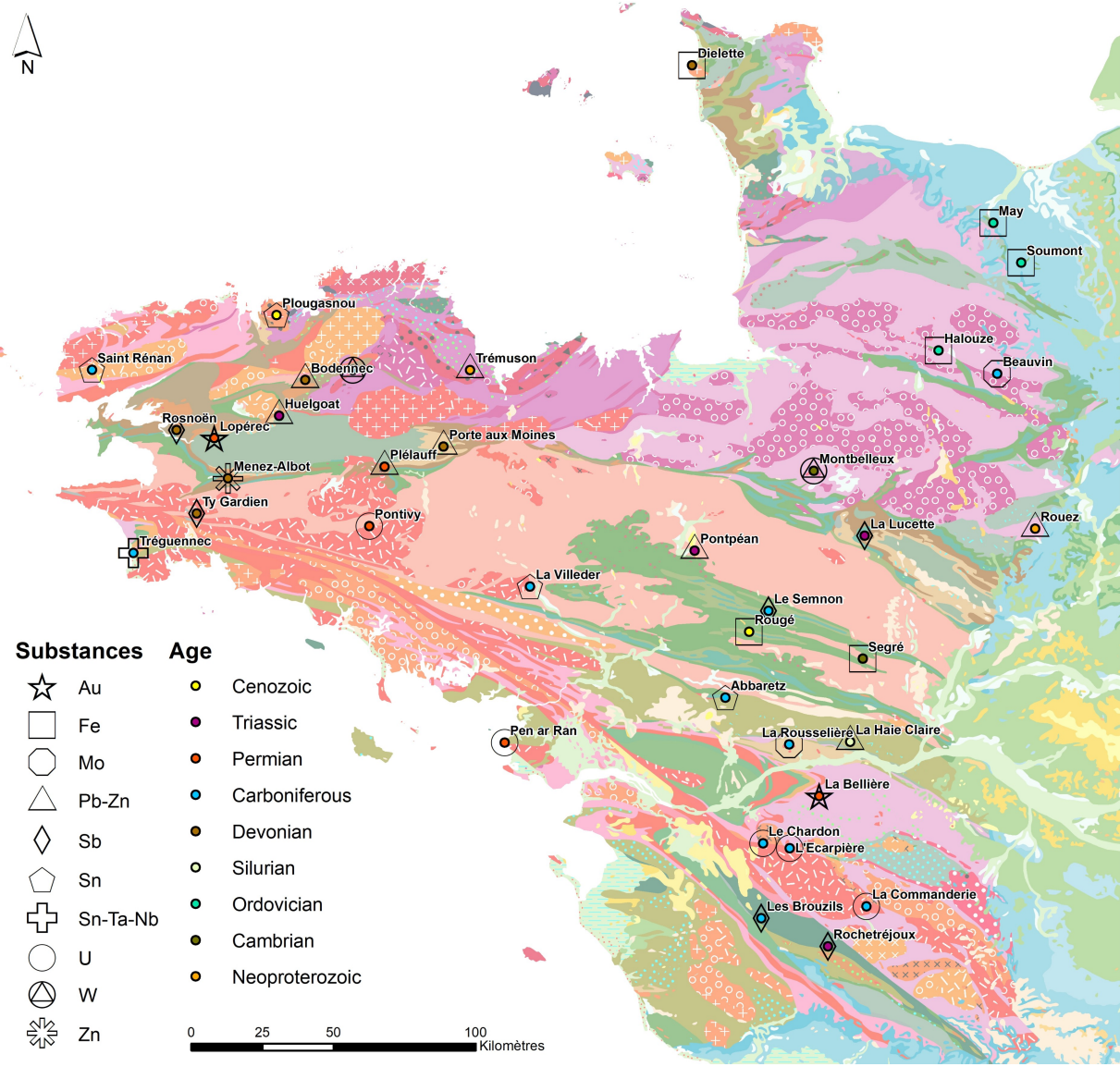


Figure 2.9: Main metallogenetic districts and mineralization in the Armorican Massif. Edited from [Marcoux and Chauris, 1994, Cassard et al., 2012, Chauris, 1989] and BRGM Data, ages from [Marcoux and Chauris, 1994, Cheval-Garabedian, 2019, Cheval-Garabédian et al., 2020, Pochon, 2017, Ballouard et al., 2018].

Background image is the simplified lithology from the 1M map of France.

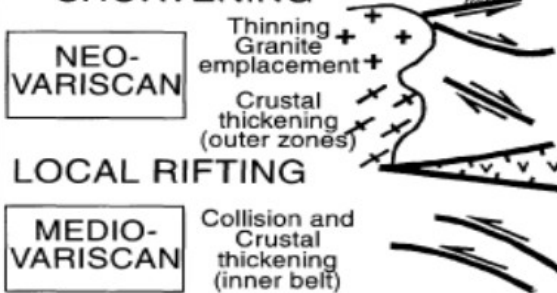
Age (Ma)		Tectonic processes	Ore deposits
Jurassic	310 -200	 POST-HERCYNIAN EXTENSION LATE-HERCYNIAN SHORTENING	Ba, F
Trias			Zn, Pb
Permian			Pb
Carboniferous			U
Steph.	350 -360	 NEO-VARISCAN LOCAL RIFTING MEDIO-VARISCAN	Zn, Ge
Westp.			Au, As, Sb
Namu.			mesothermal gold
Dinan.			W, U, Sn, Li
Devonian	400	 EO-VARISCAN	Cu, Zn, Pb, Ba massive sulphides
Silurian			(Zn, Ba, Ag, Au) La Haie Claire
Ordovician			
Cambrian	450	 PRE-HERCYNIAN EXTENSION	Zn, Pb

Figure 2.10: Main metallogenic events related to the Variscan orogeny in France (from [Lescuyer et al. \[1997\]](#)).

of substances considered of main economic interest in the AM are Lead-Zinc-Silver, Gold, Antimony, Tungsten, Molybdenum and Tin-Tantalum-Niobium (from Mineral-info.fr and [Cassard et al., 2012]).

2.3.4 Antimony in the Armorican massif

In this study, data covering the Armorican Massif (AM) are used to test the predictive approach developed with a regional-scale natural case study. The studied area covers most of the AM, apart from the Northern part which is mainly composed of Cadomian basement rocks. The studied area is mostly composed of Variscan geological units of various natures. The AM contains 5 main antimony districts (**Figure 2.11**) : the districts of La Lucette (LU), Vendée (V), Semnon (S), Quimper (Q) and Rosnoën (R). Most of these districts are vein-types, related to major faults along the North and South Armorican Shear Zones (NASZ, SASZ). The Semnon ore deposits are located exclusively in and around dolerite dykes. The Rosnoën deposits, located within the Carboniferous syn-compression Chateaulin Basin, are the only stratabound deposits known in the AM.

Antimony is mostly found as a sulfide, in the shape of Stibnite. It is quite common in geothermal systems and often accompanied by other metals (As, Au, Ag ...). Antimony deposits can be formed in a large variety of lithological (magmatic intrusion, veins ...) and structural contexts (rifting extension, collision compression ...). The main ore deposits morphologies are limited. The most common ones are vein-type deposits, often linked to structural events (i.e. faults, shear-zone ...). The other type is strata-bound deposits, replacing sedimentary rocks, most of the time carbonates [Gloaguen and Tourlière, 2016a].

In the AM, the origin and formation of antimony mineralization are explained by many metallogenic models. Those models draw relationships between antimony and multiple geological features, such as granite intrusions [Fouquet, 1980, Marcoux et al., 1984], deformational events [Marcoux and Chauris, 1994] and mafic intrusions [Pochon et al., 2018]. Some antimony mineralization are explained as perigranitic veins (Vendée, Quimper), where magmatic fluids circulation in the vicinity of a granite intrusion results in the formation of multiple mineralized zonations around it [Marcoux and Jébrak, 2008]. In this case a spatial relationship is observed between the granite and the mineralization, as the antimony veins should be formed within a specific surrounding from the granite. Granite intrusions inside the AM are often associated to major structural features (i.e. the NASZ and SASZ) which also play a role in the formation of antimony mineralization, creating structures able to host the mineralization while the granitic intrusion provides fluids and a thermal engine.

Usually, metallogenic models associated the antimony mineralization to post-orogenic extensional processes in the AM and in the neighbouring FMC (300-310 Ma) [Munoz et al., 1992]. In this case, the formation of the antimony mineralization is linked to the intracontinental and post-orogenic evolution, during the Late-Carboniferous with the emplacement of granitic intrusion and the development of shear zones. The antimony in the mineralized fluid could be of different origins: the fluids could already be enriched in antimony from a deep crustal source [Fouquet, 1980] or the antimony could be provided by enriched lithological formation(s). The latter model has been proposed by [Marcoux et al., 1984] for the Vendée antimony district with the remobilisation of antimony inside Cambro-Ordovician volcano-sedimentary units, enriched in antimony. A quite detailed study of antimony mineralization through the AM has been realized recently by Pochon et al. [2016, 2019], with the acquisition of new petrographic, geochemical and radiochronological data. This study evidenced a link between the presence of antimony

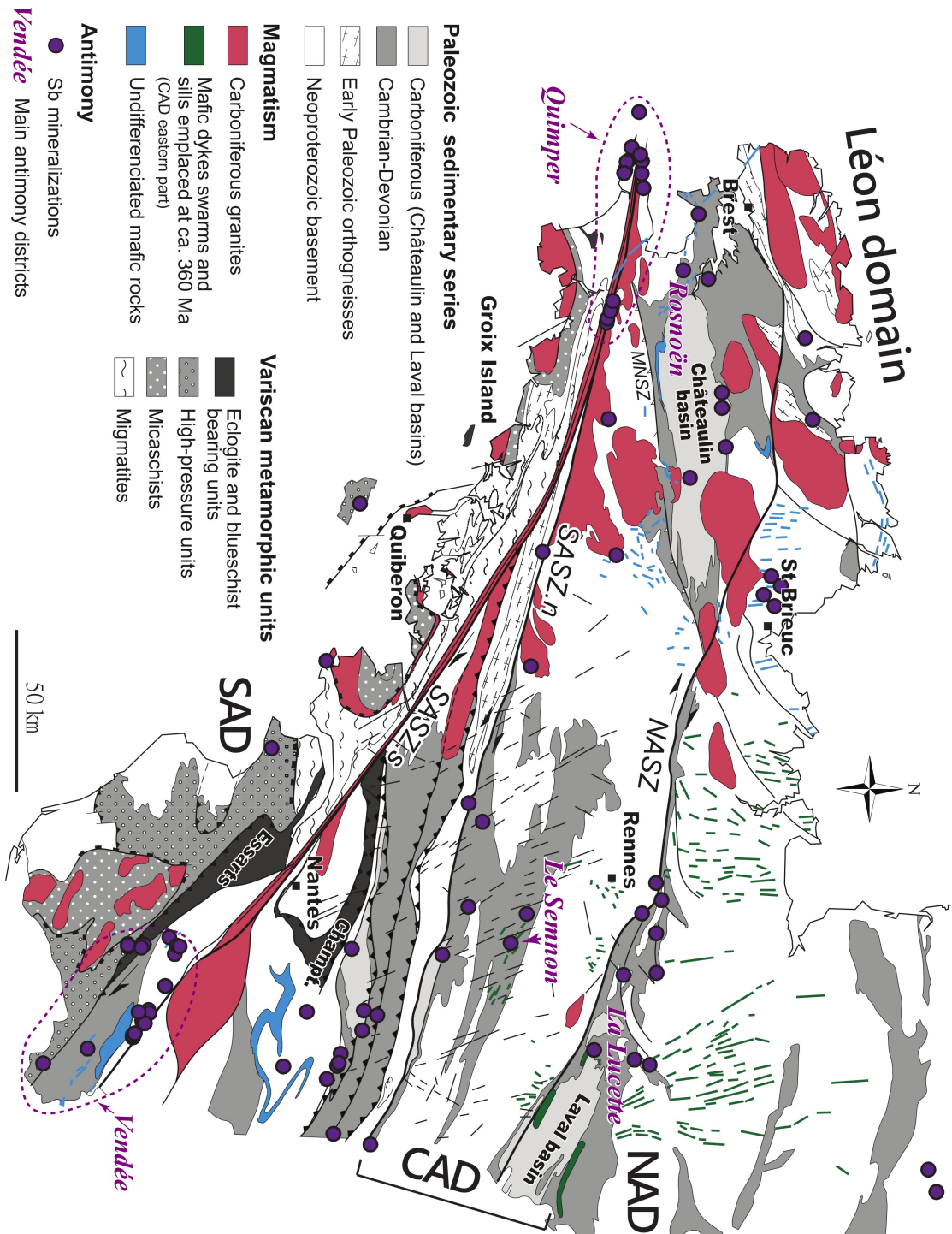


Figure 2.11: Synthetic geological map of the Armorican Massif with location of the Sb mineralization and the main Sb districts (modified from [Poehon et al. \[2019\]](#)). NAD: North Armoricain Domain; CAD: Central Armoricain Domain; SAD: South Armoricain Domain; NASZ: North Armoricain Shear Zone; SASZ: South Armoricain Shear Zone (n.: northern branch; s.: southern branch); MNSZ: Montagnes Noire Shear Zone; Champt: Champtocéaux metamorphic complex.

mineralisation in the Semnon and La Lucette districts to trace of mafic magmatism, such as dolerite dykes and geophysical anomalies interpreted as underlying deep mafic magmatic intrusions. This event would have taken place during the continental northward subduction of the North Gondwana margin during Lowermost Carboniferous, with datation of around 360 Ma for the Semnon district. This mafic magmatism happened when the Northern Gondwana margin subducted below the Ibero-Armorican micro-plate, to the north, leading to the intrusion of numerous dolerite dykes throughout the area. The antimony would have thus been provided by the mafic intrusions and the dolerite dykes.

In regard to a potential conceptual model to perform predictivity mapping on antimony in the AM, some main metallotects can be identified from previous metallogenic models. Most antimony mineralization within the AM are veins-types, which are often hosted in quartz veins and dolerites dykes. Those two lithological formations could be used in a predictive model as metallotects of antimony mineralization and potential predictors to investigate the presence of unknown mineralization. Carboniferous granite intrusions, as well as mafic intrusions and Cambrian volcano-sedimentary units are presented as potential source for the genesis of antimony mineralization, which could also be integrated as potential predictors to map antimony favorability. Faults and structural frameworks also present an interest, as they favor the circulation of fluids and the genesis of ore deposits. As such, they should also be considered in a conceptual model, especially the Carboniferous shear zones, like the SASZ and NASZ, that are the main deformation events in the AM. If available, geophysical data should be considered, as geophysical anomalies could display the presence of deep mafic magmatic intrusions as well as the presence of shallow granitic intrusion that could be related to the formation of antimony deposit. Geochemical data could also be integrated, as Sb, Hg or As geochemical anomalies could be footprints of antimony mineralization.

2.4 Iberian Massif : Central Iberian Zone

2.4.1 Geological setting and characteristics

In the Iberian Peninsula, the Variscan Range is divided into six domains (**Figure 2.12**); from north to south: the Cantabrian (CZ), The West Asturian-Leonese (WALZ), the Central Iberian (CIZ), The Galicia-Tras Os Montes (GTMZ), The Ossa Morena (OMZ) and the South Portuguese Zone (SPZ). The CIZ is located in the Western part of the Iberian Peninsula, extending from the North-Western part of Spain to its center, through North Portugal. It is bordered to the North by the Ollo de Sapo anticline, forming its contact with the WALZ, and to the South by the Ferreira do Zêzere-Portalegre thrust [[Quesada and Oliveira, 2019](#)]. The GTMZ is sometimes considered as being part of the CIZ. The GTMZ is composed of allochthonous units overlying autochthonous units, while the CIZ is made of autochthonous units. The GTMZ presents strong similarities with the SAD [[Ballèvre et al., 2014](#)]. For this work, we will focus on the CIZ sensu-stricto, corresponding to the autochthonous units.

The autochthonous CIZ is composed of four sequences of sedimentary rocks separated by three main discontinuities. The Lower Alcuadian and the Upper Alcuadian-Pusian are the 2 first sequences of Precambrian ages and form the basement of the Variscan belt in the CIZ. The Lower Alcuadian is mainly composed of turbidites formed in an extensional basin. The Upper Alcuadian-Pusan is made of a more varied range of lithologies with turbidites from deep basins and shore deposits. On this pre-Cambrian basement lies the Paleozoic sequences of the Lower Ordovician-Lower Carboniferous, made of sedimentary units evolving from fluvial sediments deposits to coastal ones and the Middle to Late Carboniferous

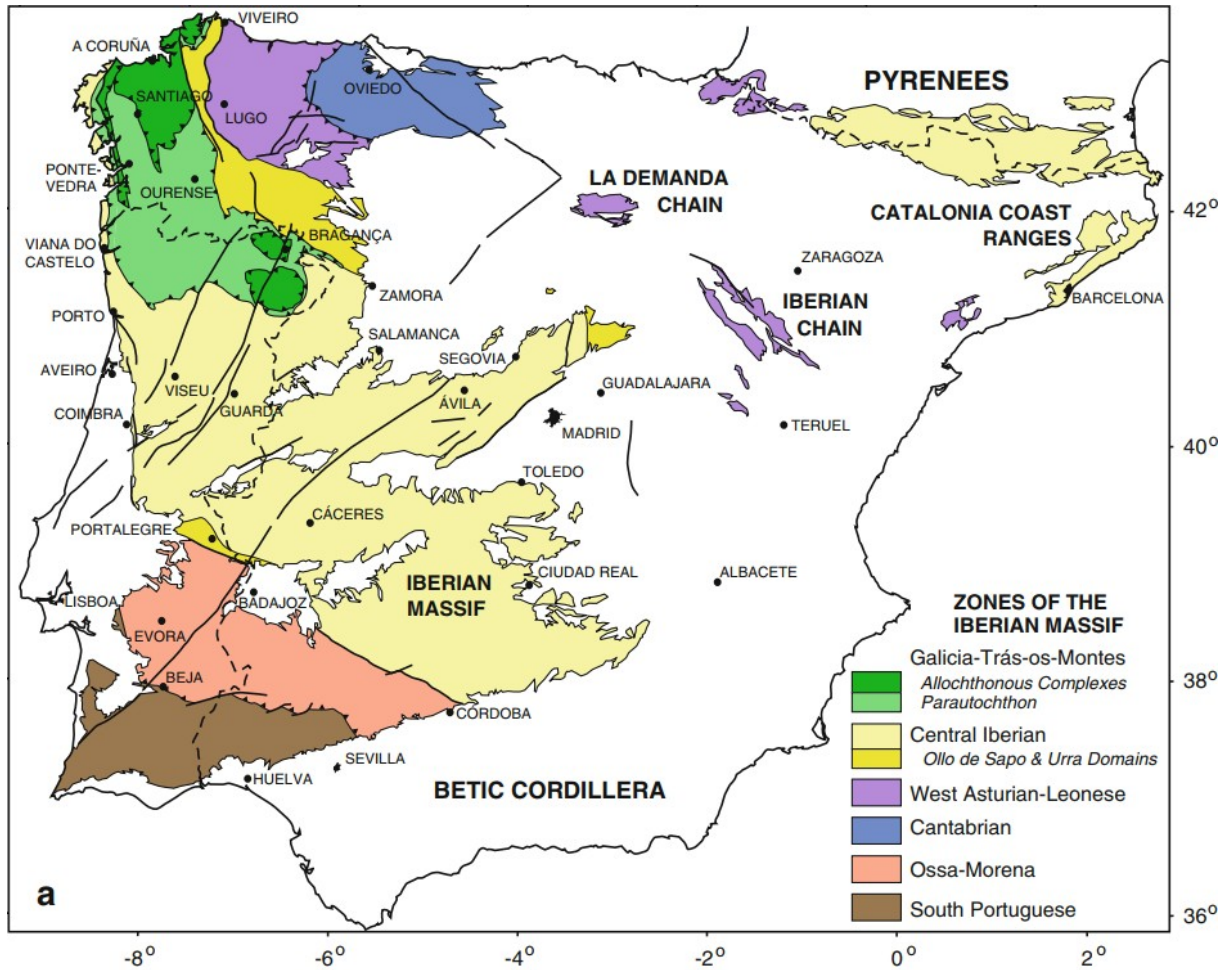


Figure 2.12: Main Variscan domains of the Iberian Peninsula (from [Catalán \[2011\]](#)).

with continental deposits [[Dallmeyer and Garcia, 1990](#), [Quesada and Oliveira, 2019](#)].

The CIZ is first marked by the opening of the Rheic Ocean around 500 Ma, and then by the subduction of the south-eastern part of this ocean, generating the back-arc basin of the Paleotethys and the dextral shearing of Porto-Tomar-Ferreira. From 390 to 370 Ma, the collision between the Armorican plate and the Iberian plate follows, with the obduction of the Paleotethys and the setup of the allochthonous unit of the GTMZ. After 370 Ma, the collision leads to an intracontinental deformation and a NW-SE horizontal shortening of the Ibero-Armorican arc (IAA) [[Gutiérrez-Alonso et al., 2012](#), [Ribeiro et al., 2007](#)].

Some structures present in the CIZ are marked by the curved geometry of this arc. In the north, near the Cantabrian Arc (CA), structures trend E-W to NNE-SSW and gradually change to WNW-ESE trending structures in the center of Iberia. The same can be observed for shear zones, with dextral shears generally parallel to these trends and associated sinistral shears oblique to them. The precise geodynamic origin of these shears zones is still debated [[Gutiérrez-Alonso et al., 2015](#)]. The CIZ can be divided in two structural domains: the North is the recumbent fold domain and the South the domain of vertical folds [[Dallmeyer and Garcia, 1990](#)] (**Figure 2.13**).

The deformation undergone by the CIZ during the Variscan orogeny can be separated in three main phases [[Dallmeyer and Garcia, 1990](#), [Ribeiro, 2013](#), [Dias da Silva et al., 2017](#)] (**Figure 2.14**). The first phase, called D1 (**Figure 2.14.1**), occurs in a compressive regime with the formation of vertical axial plane

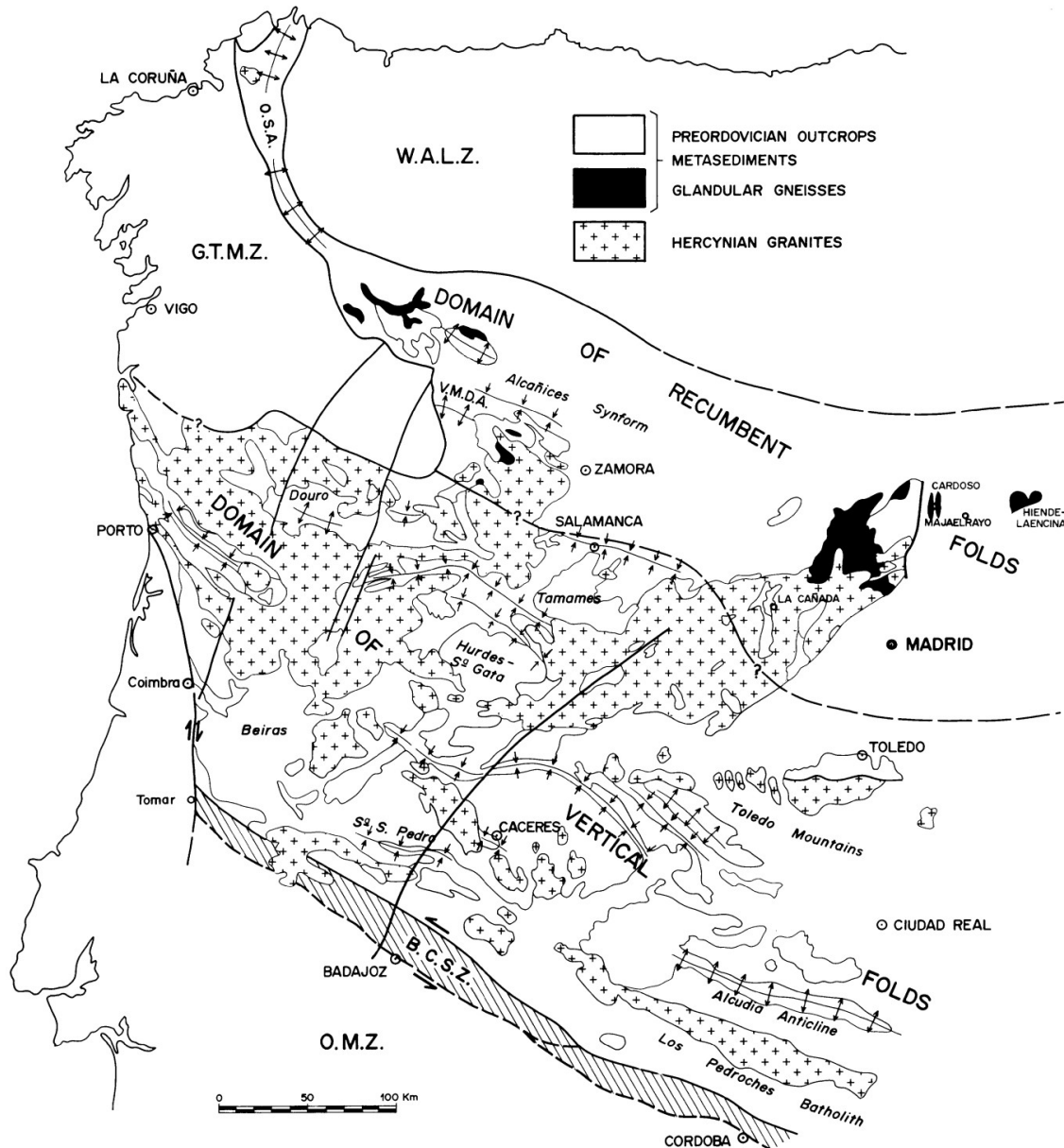


Figure 2.13: Simplified geology and main structural domains of the CIZ (from [Dallmeyer and Garcia, 1990]).

G.T.M.Z. Galician-Tras-Os-Montes Zone; W.A.L.Z. West Asturian-Leonese Zone; O.S.A. Ollo de Sapo Anticlinorium; V.M.D.A. Villadepera-Miranda Do Douro Antiform; B.C.S.Z. Badajoz-Cordoba Shear Zone; O.M.Z. Ossa Morena Zone.

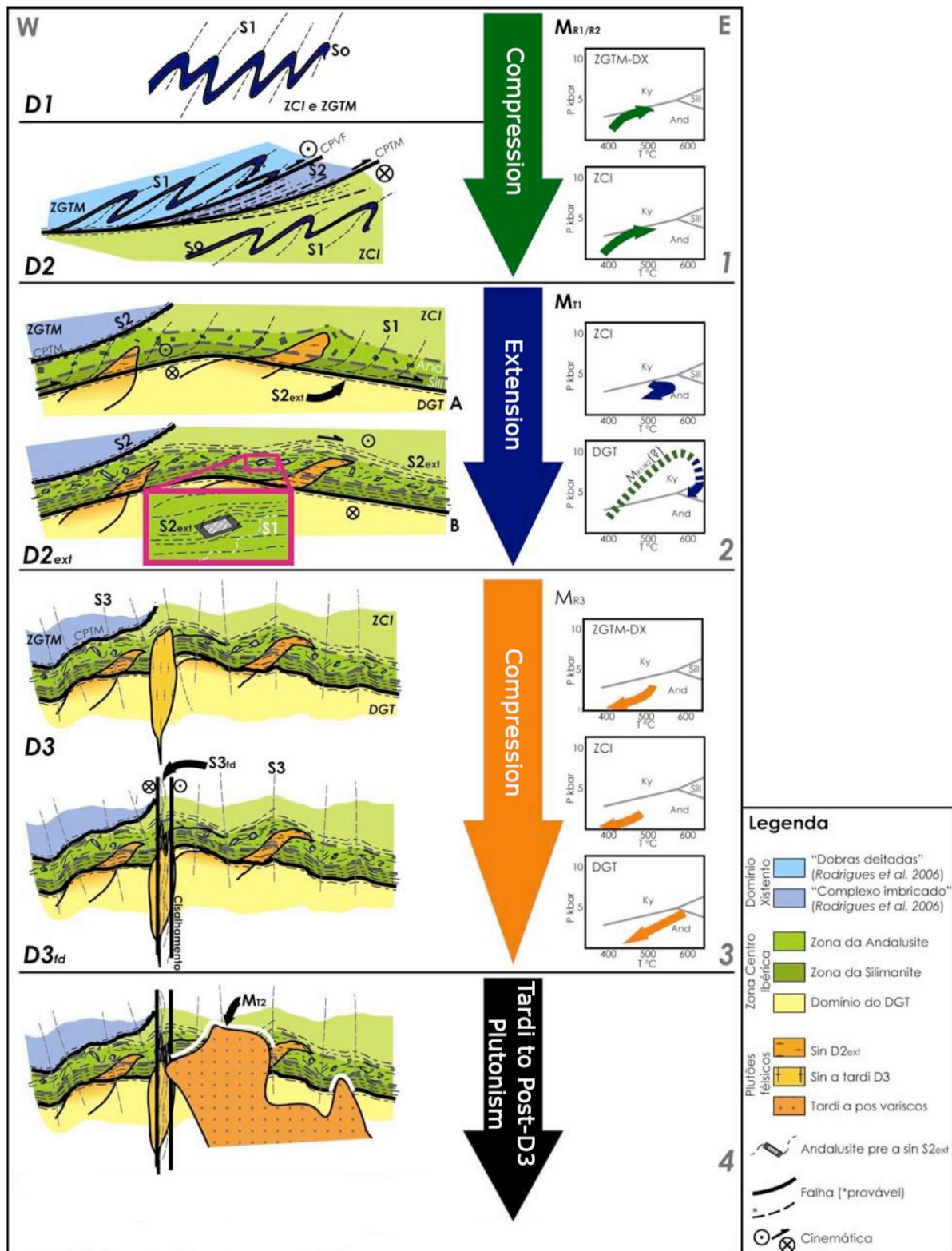


Figure 2.14: Geodynamic and tectono-thermal evolution of the CIZ (ZCI) and GTMZ (ZGTM) during the main deformations events, modified from the study of the Palaçoulo region (east of Morais Massif, NE Portugal) of Silva [2010]). D1, D2 and D3 are the deformation events and S1, S2 and S3 the associated schistosity. DX: Schistous domain; DGT: Gneiss Dome of Tormes; CPVF: Palheiros-Vila Flor thrusting; CPTM: Main thrusting of Tras-Os-Montes.

folds of trending NW-SE and the formation of an axial plane cleavage S1 of N120 trending. It is associated to a low-grade metamorphism (greenschists facies) [Rubio Pascual et al., 2013]. The second phase, called D2 (**Figure 2.14.2**), is tangential to the first one, with a subhorizontal foliation S2 with a N120 stretching lineation, plunging more or less steeply from 25 to 75° to the SE [Rubio Pascual et al., 2013]. This phase is also associated to a subhorizontal shear-zone that outcrops mostly on the northern part of the CIZ and is mainly linked to an extensional regime. The third phase is called D3 (**Figure 2.14.3**), with the formation of subvertical axial plane folds of ESE-WNW trend and a crenulation cleavage S3, during a compressive regime [Dallmeyer and Garcia, 1990]. The end of this phase is linked to the formation of granitic intrusive bodies (**Figure 2.14.4**).

Late-Variscan structures can also be observed within the CIZ, with a complex framework of faults and shear zones developed by a N-S to NNW-SSE compression [Arthaud and Matte, 1975, Marques et al., 2002]. Multiple interpretations exist, with Arthaud and Matte [1975] recognizing three main Late-Variscan strike – slip fault systems in Iberia, a dominant sinistral NE–NNE with a subordinate dextral NW–NNW and a sinistral E–ENE. Ribeiro [1974], Ribeiro et al. [1979] and Pereira et al. [1984, 1993] defined two main episodes of variscan faulting, with a Late-Variscan sinistral N80 and a dextral N25 conjugate systems formed during the D3 deformations [Marques et al., 2002], and a Late-Variscan Permian strike-slip faulting with a dextral N135 and sinistral N25 conjugated fault-systems. N155 dextral faults and N110 strike- to dip-slip left-lateral faults are also present, most likely Alpine reactivation of early Variscan ductile deformations or normal faults [Ribeiro, 1974, Arthaud and Matte, 1975].

The CIZ comprises numerous granitic intrusions linked to the Variscan orogeny. These formations are divided in three categories, depending on their relationships with the D3 tectonic event. The Syn-D3 granites (321-312 Ma) are mostly peraluminous granites (two-micas granites, biotite granites and leucogranites). Late-D3 granites (312-300 Ma) are more diverse with biotite and dominant-biotite granites (monzogranites and granodiorites with biotite), sometimes associated with basic rocks, and some leucogranites. Finally, Post-D3 granites (290-299 Ma) are sub-alkaline granites (biotite to biotite-muscovite granites) [Mateus and Noronha, 2010, Fernández and Pereira, 2016, Quesada and Oliveira, 2019].

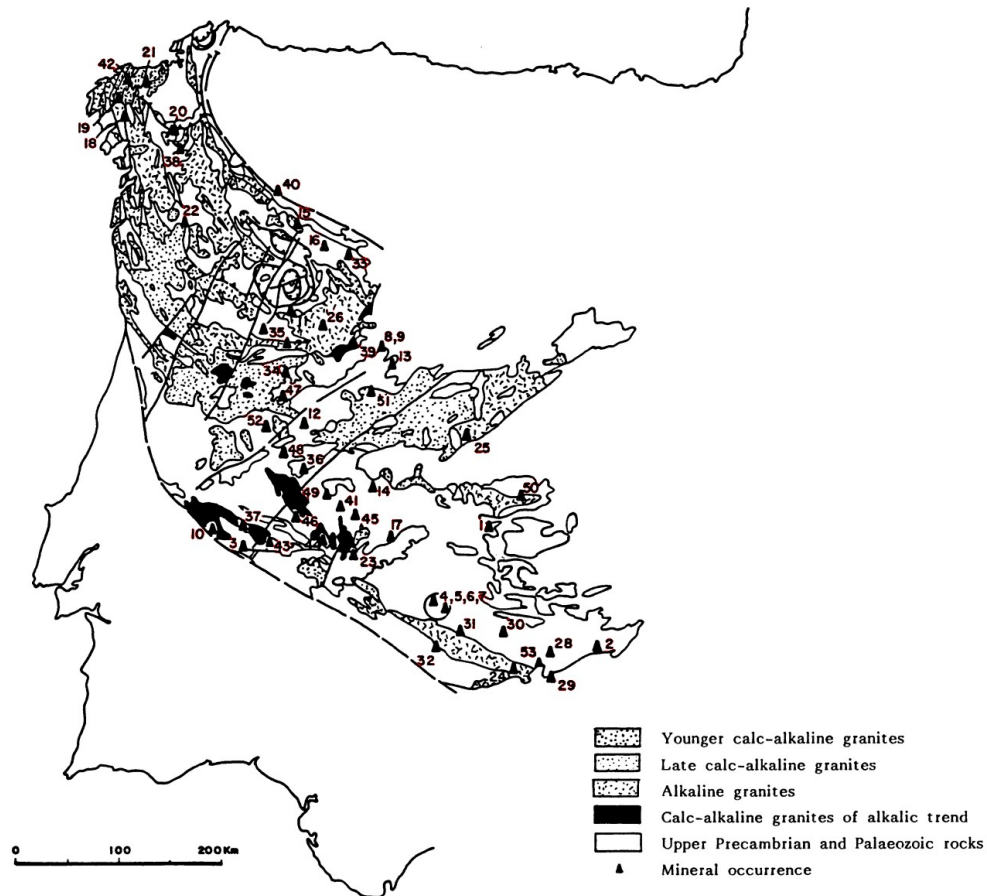
Metamorphism of the CIZ shows a first prograde event, named $M_{R1/R2}$ (**Figure 2.14.1**). It is linked to a barrovian metamorphism due to crustal thickening, associated to, in particular, thrusting of the allochthonous units (GTMZ) in the Northern part of the CIZ [Viruete et al., 1994, Rubio Pascual et al., 2013]. This barrovian metamorphism reaches its peak conditions during the M_{T1} event (**Figure 2.14.2**) with HT and rather LP conditions recorded in the deeper levels and the onset of granites intrusion within the area. Those Syn-D2 granites show similar compositions than the Syn-D3 ones [Fernández and Pereira, 2016]. Finally, the retrograde M_{R3} event shows the diminution of pressure and temperature condition, the formation of Syn-D3 granitic intrusion and shear zones. The final event M_{T2} is a local event of contact metamorphism caused by Late- to Post-Variscan intrusions. As a result of these successive events, the CIZ displays as elongated metamorphic belts, following the arcuate structure of the IAA, with metamorphic grades ranging from granulite facies to greenschists one. Those belts are crossed by metamorphic isograds [Dallmeyer and Garcia, 1990, Quesada and Oliveira, 2019].

2.4.2 Main metallogenic resources

According to [Ribeiro et al., 2014], 4 main metallogenic periods can be recognized in the Iberian Peninsula before and during the Variscan orogeny, each related to specific mineralisation type (**Figure 2.15**), within

the CIZ [Dallmeyer and Garcia, 1990]:

1. The Cadomian (587-532 Ma) and Proterozoic-Cambrian periods (560-540 Ma), with several metamorphic, magmatic and magmatic-hydrothermal ore-systems. In the CIZ, this corresponds to the Fontanarejo phosphate deposit of upper Precambrian age and related to the Cadomian synorogenic sequences.
2. The Eo-Variscan period, with the Cambrian-Ordovician (530-470 Ma) and Lower-Middle Devonian (390-370 Ma) intervals. This period is characterised by diverse sedimentary, exhalative hydrothermal, volcanogenic and magmatic processes. It is during this period that some Antimony (La Nazarena, San Antonio) and Mercury (Almadén, El Entredicho, Vieja Concepcion, Las Cuevas) stratabounds deposits related to volcanic activity occurs.
3. The Meso-Variscan period (340-315 Ma), and especially the 320-315 Ma interval (D3 event). It corresponds to all the synorogenic processes, with large-scale fluid flow, regional metamorphism and nappe stacking, crustal anatexis of the deeper structural levels and shear zone development. The Tungsten skarn deposits of Alegria, Explotada and Morille, formed during the regional metamorphism are likely to be linked to this period, as the Gold deposit of La Codosera.
4. The Late-Variscan period (312-260 Ma) with magmatic and hydrothermal processes linked to the emplacement of late and postorogenic granites and the widespread hydrothermal activity that followed. As granite intrusives are widespread within the CIZ, there are a lot of deposits linked to Syn-, Post- or Late-granites. Different kind of deposits can be linked to different type of granites. Alkalines granites (Syn- to Late-D3) are related to Tungsten (Barruecopardo, La Lapa), Tin (Teba, Calabor), Uranium (La Virgen, Esacalona) Lead-Zinc-Silver (Linares, La Carolina), Bismuth (Conquista, Pozoblanco) and Antimony deposits (Losacio). Some mineralization are linked to Muscovite leucogranites cupolas with Tin mineralization (Golpejas, Penouta). Calc-Alkaline granites are linked to Phosphate and Uranium deposits (Albuquerque, Trujillo). Late-Hercynian calc-alkalines granites are related to minor Lead-Silver deposits (Toledo, Béjar). Finally, pegmatite, enriched in Tin and lithium (Barroso-Alvao, Castillejo, Tres Arroyos) are also emplaced inside the CIZ, linked to late-D3 events.



Location of the main deposits and occurrences in the Spanish Central-Iberian Zone. 1 P, Fontanarejo (Ciudad Real); 2 Sb, La Nazarena, Almuradiel (Ciudad Real); 3 Sb-W-(Hg), San Antonio (Badajoz); 4 Hg, Almaden (Ciudad Real); 5 Hg, El Entredicho, Almaden (Ciudad Real); 6 Hg, Vieja Concepción (Ciudad Real); 7 Hg, Las Cuevas (Ciudad Real); 8 W, Alegria Mine, Morille (Salamanca); 9 W, Explorada Mine (Salamanca); 10 Au, La Codosera (Badajoz); 11 W, Barruecopardo (Salamanca); 12 W, La Lapa (Cáceres); 13 W, Santa Genoveva (Salamanca); 14 Sn-W, Teba (Cáceres); 15 Sn-W, Calabor (Zamora); 16 Sn-W, Arcillera (Zamora); 17 Sn-W-P, Logrosán (Cáceres); 18 Sn-W, San Finx (La Coruña); 19 Sn-W, Santa Comba (La Coruña); 20 Sn-W, Fontao; 21 Sn-W, Monteneme (La Coruña); 22 Sn-W, Lovios (Orense); 23 Sn-W, La Parrilla (Badajoz); 24 U, La Virgen (Jaen); 25 U, Escalona (Toledo); 26 U, Peralonso (Salaman-

ca); 27 U, Sobradillo (Salamanca); 28 Pb-Zn, Linares (Jaen); 29 Pb-Zn, La Carolina (Jaen); 30 Pb-Zn, Diógenes (Ciudad Real); 31 Bi, Conquista (Cordoba); 32 Bi, Pozoblanco (Cordoba); 33 Sb, Losacio (Zamora); 34 Sn-Li, Castillejo (Salamanca); 35 Sn-Li, La Fregeneda (Salamanca); 36 Sn-Li, Las Navas (Cáceres); 37 Sn-Li, Tres Arroyos (Badajoz); 38 Sn-Li, Lalin (Pontevedra); 39 Sn, Golpejas (Salamanca); 40 Sn, Penouta (Orense); 41 Sn, El Trasquilón (Cáceres); 42 Au, Corcoesto (La Coruña); 43 U, Albuquerque (Badajoz); 44 U, Albalá (Cáceres); 45 U, Trujillo (Cáceres); 46 U, Cabeza de Araya (Cáceres); 47 U, Fe Mine, Ciudad Rodrigo (Salamanca); 48 P, Zarza La Mayor (Cáceres); 49 P, Aldea Moret (Cáceres); 50 Pb, Ag, Guajaraz (Toledo); 51 Pb, Ag, Bejar (Salamanca); 52 Pb, Ag, Cilleros (Cáceres); 53 Sb, San Agustín (Jaen)

Figure 2.15: Main deposits in the CIZ (from Dallmeyer and Garcia [1990]).

2.4.3 Antimony in the Iberian massif

Numerous antimony mineralization can be found across the IM, as shown by [Gumiel and Arribas \[1987\]](#) and [Gumiel \[1982\]](#) (**Figure 2.16**). Most of the occurrences are located along the CIZ, with mineralization being located in the Córdoba, Jaén, Ciudad Real, Badajoz, and Cáceres provinces in Spain, and in the Trás-os-Montes region, in Northern Portugal [[Martins, 2012](#), [Neiva et al., 2008](#)] and near Porto, in the Dúrico-Beirã area [[Couto and Borges, 2005](#), [Couto and Roger, 2017](#)]. In the CIZ, Sb mineralization can be distinguished into three categories depending on the morphology of the orebody according to [Gumiel and Arribas \[1987\]](#): veins-type, which are mainly located in Precambrian and Paleozoic rocks and are the most common occurrences, stratabound deposits, located in Paleozoic sedimentary layers, with fewer occurrences located mainly along the Southern border of the CIZ and finally the volcanic-hosted mineralization, located in post-orogenic felsic dykes and located on the South-Eastern part of the CIZ. Sb is mainly hosted by stibnite, associated with other minerals, the most common association among the mineralization being stibnite and quartz.

According to [Gumiel and Arribas \[1987\]](#), antimony mineralization within the CIZ is either related to pre-orogenic volcanism, to Carboniferous intrusive rocks or to post-orogenic volcanism. Multiple pre-orogenic volcanic events can be identified, with a Lower Ordovician event (Sierra Morena and Alcudian Valley), an Ordovician-Silurian event (Villarbacú, Alcudian Valley and Almadén district), a Devonian event, related to the major Sb-W-(Hg) antimony deposit of San Antonio [[Gumiel et al., 1976](#), [Arribas and Gumiel, 1984](#), [Couto et al., 1990](#), [Combes et al., 1992](#)]). A lower Carboniferous event, which lead only to minor Sb mineralization in the Ossa Morena zone, is of major importance for the Iberian Pyrite Belt, south of the CIZ. Those volcanic events are responsible for the formation of the major Sb-W-(Hg) deposits within the CIZ and its enrichment in Sb. Later remobilization occurred during the formation of the Variscan orogen, as the Variscan plutonism is responsible for the formation of most of the veins-types deposits from the Carboniferous to Permian across the CIZ. Finally, some mineralization hosted in felsic volcanic rocks are related to a post-orogenic volcanic event (San Agustin).

A conceptual model for antimony prospectivity mapping in the CIZ could rely on similar metallogenes as the one presented for the AM: geophysical data, displaying the potential presence of magmatic intrusions could be used, as well as geochemical anomalies. Faults and structural frameworks are also to be taken into account, favoring, as much as possible, Carboniferous active faults. The host of antimony mineralization are similar, with antimony mineralization located mainly in quartz dykes and also mafic and felsic dykes for veins-types orebody. Like in the AM, carboniferous granitic intrusions and mafic ones are potential metallogenes and predictors for antimony. In the CIZ, the various lithological formations related to the pre-orogenic volcanic events are also possible metallogenes for antimony. In addition, as most stratabound deposits are located Paleozoic sedimentary layers, those could be considered as potential metallogenes for antimony mineralization as they could host them. However, these criteria span a large number of formations and a large surface of the study area, and integrating such predictors in a conceptual model could make it less accurate.

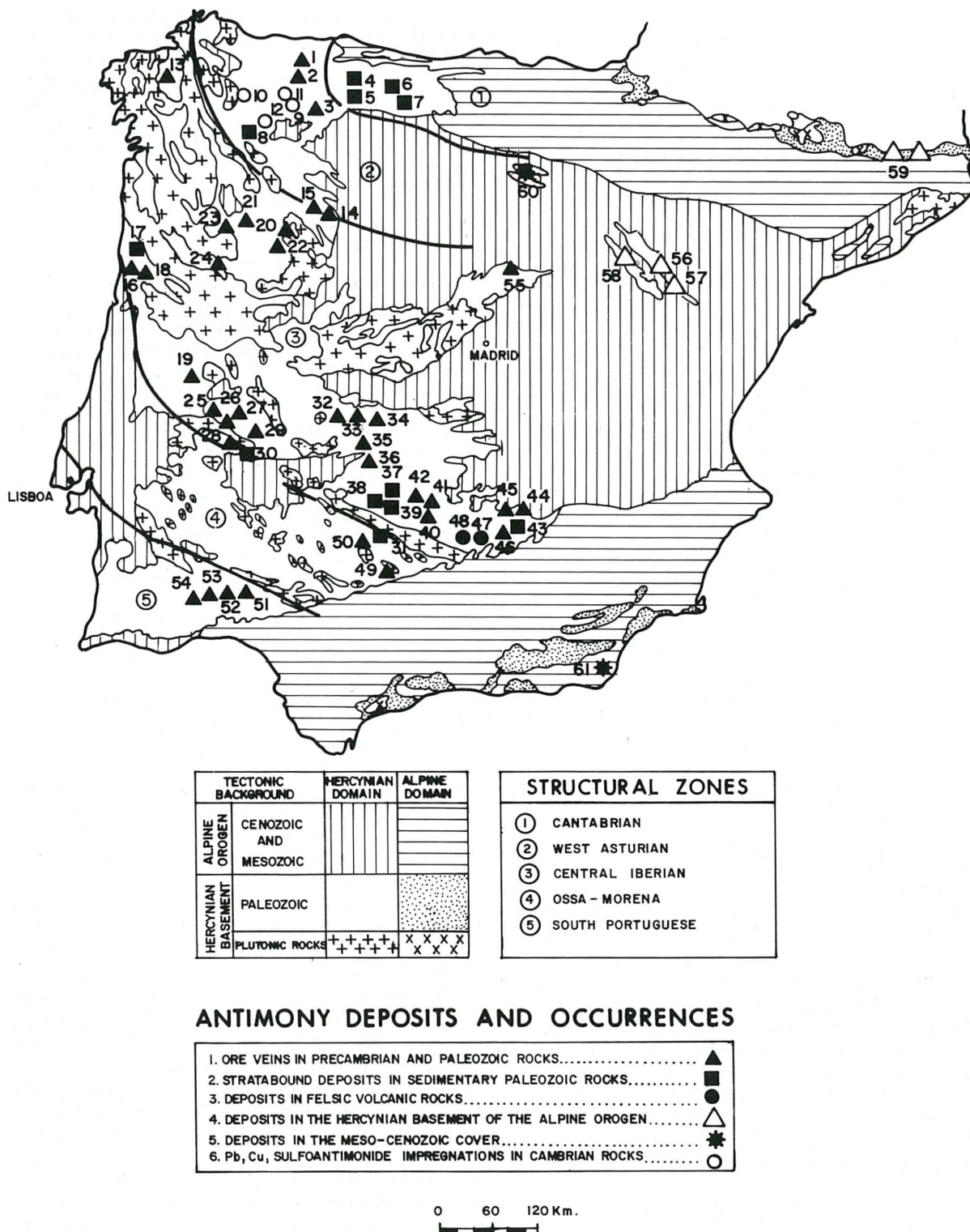


Figure 2.16: Geologic sketch of the Iberian peninsula giving structural zoning and location of the main deposits of antimony in the Iberian Peninsula, as well as the type of occurrences (from Gumiel and Arribas [1987]).

Part II

Developments

Chapter 3

Spatial analyses

While previous **Chapter 2** displays the geological and metallogenic context in which this study takes place, **Chapter 3** and the following **Chapter 4** and **Chapter 5** aim at describing the development and operating of the methodology developed to produce the Data-driven automated identification of relationships between mineralization and geological features. This **Chapter 3** shows the development of a method allowing to summarize the geological and structural spatial data into local geological environments, before the application of a method quantifying the relationships between those geological environments and the mineralization, as described in **Chapter 4**. A first method, named Cell-Based Association (CBA), is presented, allowing to discretize lithological and structural spatial data into a regular grid with an associated multivariate matrix. While this method allows to produce the geological environments needed to perform our methodology, it also presents some limitations. To overcome those limitations, we developed a new method based on the CBA, the Disc-Based Association (DBA), allowing to improve the resulting geological environments found. This method is presented in details in this **Chapter 3**, after the introduction to the CBA and its limitations.

3.1 Cell-based association method

3.1.1 Principles

The Cell-Based Associations method (CBA) is a prospective mapping method developed by [Tourlière et al. \[2015\]](#), based on the previous work of [McCammon et al. \[1983, 1984\]](#), in order to produce favorability maps for mineral resources. It was developed with the particular aim of better integrating geological maps data, a qualitative data by essence, in the calculation of mineral resources prospectivity maps. This approach is based on a discretization of the study area using a square grid cell to synthesize the geological data, such as lithology, into geological environments. From a technical point of view, considering a list of n lithological formations contained within the study area, a n digits binary number is computed for each cell of the matrix resulting from the square cell grid. This binary code can be defined as a lithological spectrum of a cell, where each lithological formation is converted into binaries: if the lithology is present inside the cell, it will be coded as "1", otherwise it will be coded as "0" (**Figure 3.1**).

Each cell is then defined by this binary code representing the "local" lithological association within

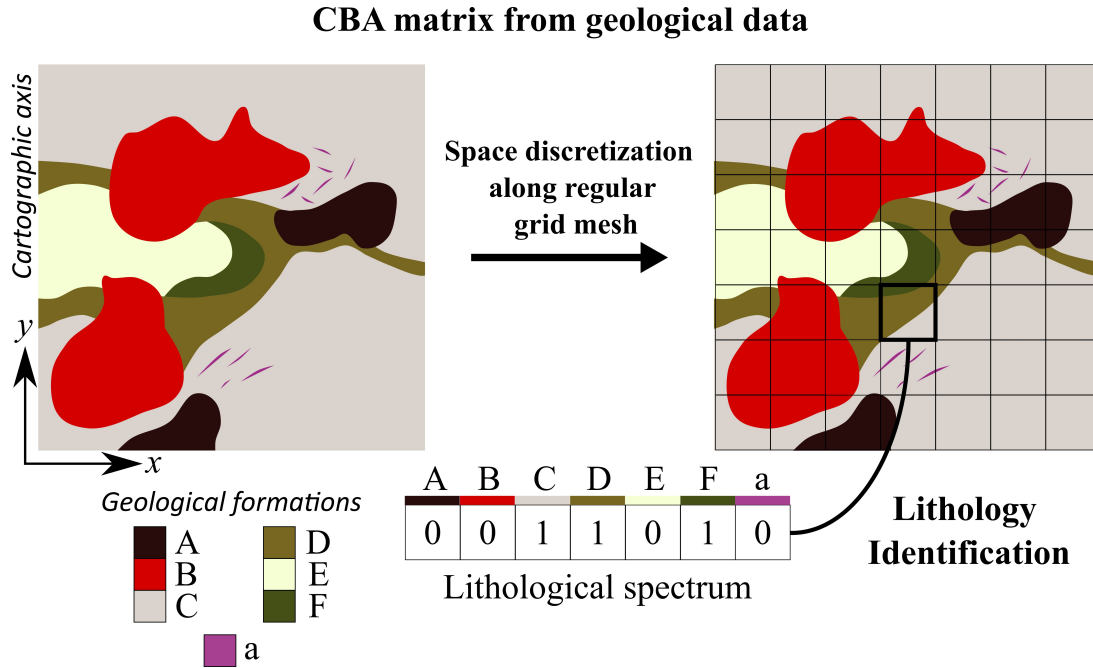


Figure 3.1: Theoretical sketch example of the creation of the CBA matrix from a lithological dataset. The map space is discretized using a regular square grid based on the cartographic coordinates.

the cell area. Its length corresponds to the number of lithologies that are comprised within the study area covered by the grid. The maximum number of associations is 2^n . If the number of lithologies is high, it becomes harder to process the associations, as their number becomes too high. It is then possible to use a classification or clustering method, such as the Agglomerative Hierarchical Clustering method (AHC) to diminish the number of associations, to group and classify them while keeping their consistency.

The matrix produced by the CBA can integrate other data than lithological one. Structural data, for example, can be added to the matrix using the same process as the lithological dataset: a new column is added to the matrix and when a fault intersects a cell, it is noted as “1” in this new column, otherwise as “0” (**Figure 3.2.a**). For Mineral Prospectivity Mapping (MPM) known targeted mineralized occurrences are integrated in the matrix. One supplementary dimension is added to the spectrum, determining if the cells are mineralized or not. A cell containing at least one mineralization occurrence is considered mineralized with an attributed binary “1”, while a cell containing no mineralization at all is considered non-mineralized with a binary “0” (**Figure 3.2.b**). In order to consider the possible inaccuracies in the location of the occurrences compiled in the metallogenic databases used, computation of a buffer zone around the corresponding point features can be considered, as it allows avoiding the case where, due to the space discretization process, a deposit which would unluckily be located along a cell border would not be taken into account in the neighbouring cell. Moreover, buffers simulate the fact that some kind of mineralization can occupy, by essence, a significant volume and thus cover a non-null surface on the map.

3.1.2 Advantages

The CBA is a Data-driven approach designed to be used during the exploration stages, when the limited data available do not allow the use of other more traditional methods such as Weight of Evidence (WoE),

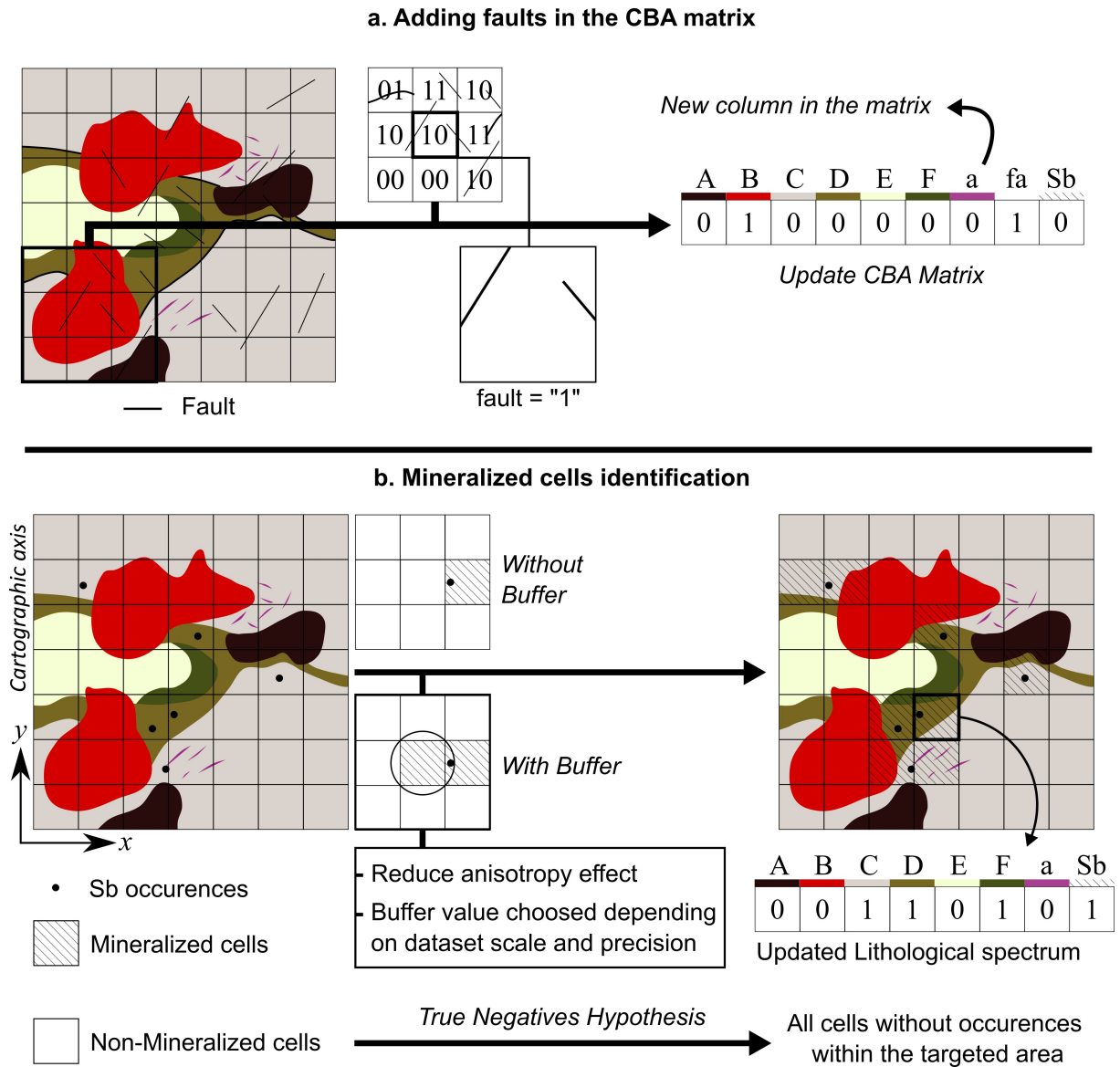


Figure 3.2: Addition of new data to the CBA matrix. a. Integration of fault to the CBA matrix. The corresponding binary variable stands for the presence (1) or the absence (0) of faults within the cell. b. Identification of mineralized cells in the CBA matrix. Mineralized cells can be identified with or without buffer; the latter is defined from the center of the cell and can be defined by the operator. After the mineralized cells are identified, all other cells are considered as non-mineralized and the state of each cell is shown in the CBA matrix with either a "1" (mineralized) or a "0" (non-mineralized).

Fuzzy Logic (FL) or Artificial Neural Network (ANN) [Carranza, 2011, Bonham-Carter, 1994, Schaeben, 2012]. Implementing this approach overcomes some of the problems inherent in using geological maps as input datasets. Such problems comprise (i) the inaccuracies in polygons contouring (e.g. lithological formations) and points location (e.g. mineralization), (ii) the outline curves tortuosity depending on the scale of the geological map used, (iii) the presence of superficial cover unrelated to the studied events, (iv) the inappropriate generalization of a phenomenon related to mixed environments into a whole unique formation or polygon (e.g. mineralization occurring in a skarn should not be related to the carbonate formation, but to its contact with the intrusive), and (v) clustering effects, when occurrences related to the same event are clustered together, leading to an overestimation of the importance of spatially associated variables.

This method does not take into account the surface of geological units and only consider their absence or presence in a given area. Indeed, one can assume that the surface covered by a geological feature on a geological map has no importance for the predictive models used, as lithological formations are 3D objects and the surface area displayed on the map, and in the case of the CBA, in the cells, cannot be strictly correlated to their importance. As such, it cannot be reliably used to weight them as long as the volume of the formation is unknown. For instance, a lithological unit covering a small area on the surface may represent a much more consequent volume in depth. Small occurrences of certain lithologies at the surface, such as granite apex, could correspond to a much larger volume at depth, inducing fluids circulation through a wider zone. In addition, other types of formations, such as veins, aplites or pegmatites, can also show a strong relationship with certain types of mineralization while only covering a small surface of the map. Their volumes can also be underestimated, as it is the case with the granite apex, giving that only the veins intersecting the surface are shown [Launay et al., 2021]. Their surface on the map can also be often exaggerated, this type of geological object being reported on the geological map without strict respect of the scale, their size being increased to make them visible on a map. With the CBA, the relative proportion of each lithology is not used to weight them, as it only takes into account their presence or absence. It allows avoiding any bias related to the 3D arrangement of lithologies that can't be taken into account on the 2D space displayed by a geological map. As such, if one cell from a CBA grid contained four different lithologies, covering respectively 95%, 2%, 2% and 1% of the cell's surface, while another cell contained the same lithologies in an equal proportion (25% of the cell's surface each), those two cells would be for the CBA, as they present the same environment, even if the lithological proportion are not the same.

Rather than seeking to obtain the finest possible resolution to establish the most favorable areas for mineralization, the CBA seeks to define over a larger area the signatures of environments that may be related to these mineralization. This approach is therefore interesting for identifying and defining the geological environments associated with mineralization and allows, through a statistical approach, to mimic the thinking that a geologist interested in the genesis of these mineralization might have by looking at combinations of geological factors [Billa et al., 2004, Roy et al., 2006, Cassard et al., 2008]. While other predictive methods focus mainly on computing favorable criteria and favorability score maps related to the target mineralization, the CBA method is able to highlight the geological association signatures associated to the mineralization, in addition to the estimation of the favorability score map. However, the resulting resolution of the CBA is coarser than other predictive methods because of the aggregation of data in cells. As such, this method is more suitable to identify the lithological environments of the mineralization, when applied on larger study areas, highlighting the spatial relationships between the mineralization and their surroundings and giving more insight about their genesis.

If different events of mineralization have occurred for the same substance and in the same area, the CBA approach allows separating different classes of geological associations signatures attached to the known deposits, and thus to distinguish different potential distinct metallogenic events. Moreover, this method allows the differentiation of diverse types of mineralization within data containing mineralization of various natures. Indeed, databases are generally organized according to the substances or associations of substances searched. However, mineralization of the same nature can be generated by different processes, which implies different spatial relationships with their environment. The application of predictivity methods in this context does not allow to differentiate these relationships and tends to provide an integrated favorability map (i.e. grouping mineralization from various events, processes), not very representative of the diversity of the mineralization genesis modes. The CBA addresses this problem by proposing to characterize the different environments to which the mineralization belongs before performing favorability calculations. This allows identifying the mineralization that can be the result of similar processes according to the similarities highlighted between their environments. The CBA approach is one of the only approaches able to calculate a level of favorability to find mineralization in addition to the association model of the favorable areas for several distinct association models.

It is noteworthy that the CBA also allows compensating for other defects in this type of database, notably in the case of mineralization forming "clusters". Those can be representative of a single object geologically wise. It can be, for example, multiple orebodies in a mineralized field linked to the same intrusion [London, 2018, Pochon et al., 2016, Launay, 2018]. If we consider each individual belonging to this cluster as a single mineralization, this can lead to an overestimation of the importance of the single associated object, which can strongly affect the quality and outcomes of the predictive model if we are in the case of a database containing mineralization of various natures. The CBA, by discretizing the map surface, can group these clusters within the same cell or neighboring ones, allowing to group these occurrences presenting a unique signature.

3.1.3 Critical parameters

From its principles and the first applications made of the CBA (see Gloaguen and Tourlière [2016a,b], Tourlière et al. [2015]), it clearly comes out that the size of the grid used to discretize the cartographic data is a critical parameter. The absolute location of the grid is also an important parameter. The square grid used is anisotropic by definition, so this parameter allows compensating, in a certain limit, the creation of environments where the mineralization are present only in the edge vicinity of the cells and thus are less representative of the environment around this mineralization. The computation of the grid is an important step in the CBA in order to outline the favorable areas with enough precision. To achieve precision, the optimal CBA resolution should result in the smallest possible cell size as it corresponds to the resolution of the final output for both the lithological environment found and the final favorability scores grid. Meanwhile, the cell size should be large enough to allow covering significant neighbouring areas and, thus, as much diversity as possible in the resulting local lithological associations to analyse. The size of the grid cells is a crucial parameter since it directly influences the possible number of geological environments present in each cell during the classification. It is necessary to find a good balance: each cell must contain on average more than one lithological formation (otherwise, one cannot speak of geological environments), but the number of geological environments must be sufficiently varied as a whole so that the classification has a meaning; cells should thus not be too large not to be too much integrative.

Generally speaking, the number of cells will inversely increase with the size of the grid cells, increasing simultaneously the computation time. Increasing the resolution (decreasing cell size) results in an increasing number of unique lithologies associations (i.e. unique spectra), at the expense of the diversity of lithologies each cell contains. Conversely, increasing the grid cell size leads to a decrease in the number of cells, largely reducing the calculation times. Lowering the resolution (increasing cell size) results in more contrasted associations (with more variety of lithologies) but much less unique associations (**Figure 3.3**). At very low resolutions, those associations become irrelevant as they regroup a wide variety of lithologies, adding “noise” in the classification. This leads to a degradation of the quality of the classification: with so few individuals presenting increasingly strong similarities between them with the integration of common lithologies, the obtained classes do not have any real significance, or do not even have any interest when the number of individuals is extremely low. To obtain suitable results, it is necessary to adjust the spatial resolution to be coarse enough for the associations to contain at least more than one lithology and to have enough distinct associations to make a proper classification process. The resolution also needs to be high enough to keep some meaning in regard to the targeted area and the usefulness of the MPM maps computed at the end.

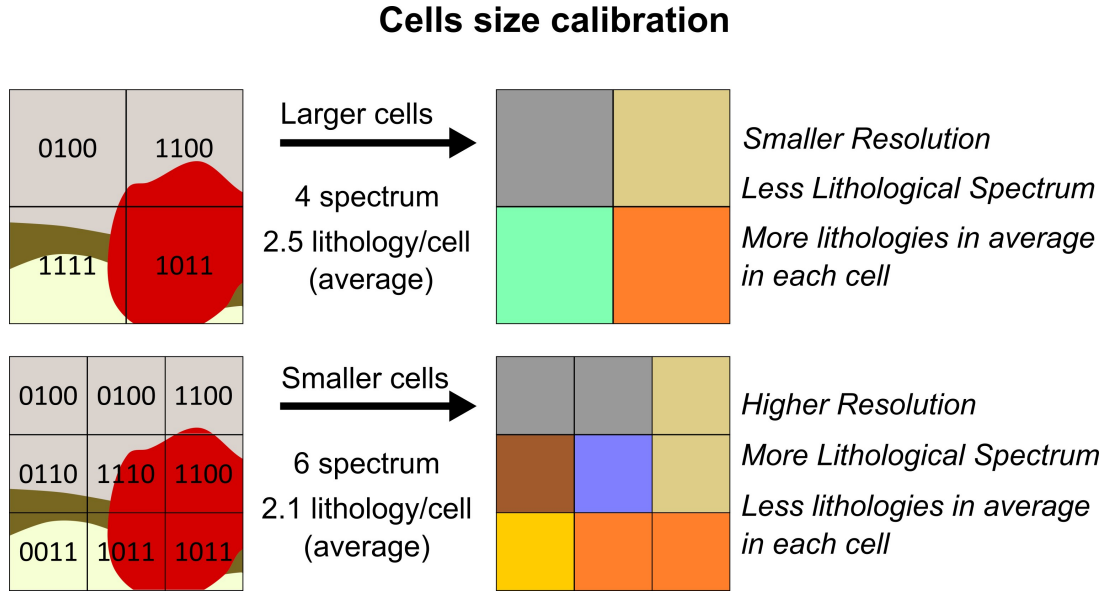


Figure 3.3: Influence of the size of the cells used to produce the CBA matrix on the resulting associations.

These optimal parameters are initially determined through a multi-scale iterating process of creation of CBA matrices, from a minimum to a maximum cell-size, changing at each iteration the size of the grid cells, and if necessary, the absolute location of its origin. For each step, four parameters are computed including (i) the total number of output distinct geological associations (**Figure 3.4.a**), (ii) the total number of cells in the grid (**Figure 3.4.b**), (iii) the average number of lithologies in a cell (**Figure 3.4.c**) and (iv) the frequency of cells including only one lithology (**Figure 3.4.d**). The number of cells containing occurrences can also be computed, its variation depending mainly on the clustering of the mineralization: widespread mineralization tends to keep a constant number of mineralized cells, while with highly clustered mineralization the number of mineralized cells will sharply decrease with the cells size increase. The realization of various graphs comparing the evolution of these parameters as a function of the cell size forming the grid, specific to each study area, makes it possible to determine an interval of optimal values for the size of the grid. This interval can be seen when studying the evolution of the number of unique

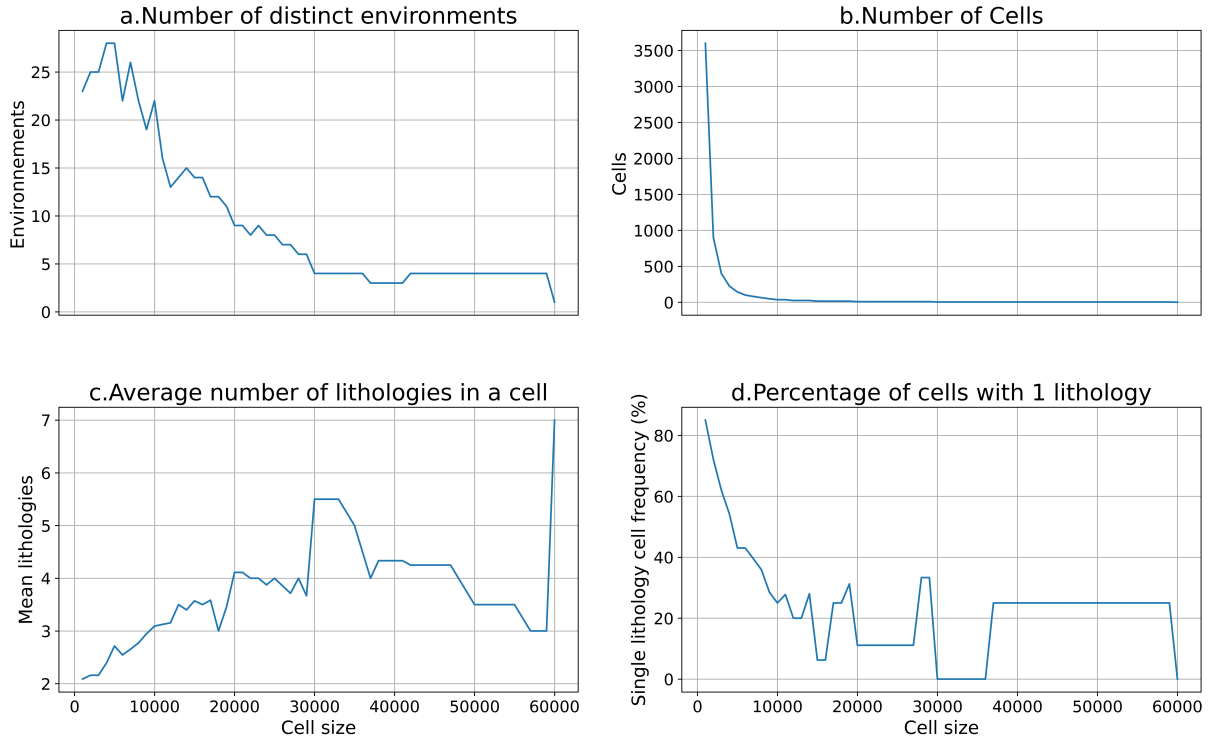


Figure 3.4: a. to d. Impact of the cell size on different parameters of the CBA matrix, applied on the previous sketch map with cell size ranging from 1 000 to 60 000 m with steps of 1 000 m. The size of the area covered by the sketch map is a square of 60 000 x 60 000 m.

associations as a function of cell size, generally in the form of a slope break in the values curve. This interval of optimal cell size should allow maximizing the total cells number while minimizing the ratio of cells containing only one geological unit, which are useless for the subsequent classification process. It should also ensure to preserve a number of distinct environments high enough to get significant results from the classification.

A careful selection of the geological data can also help improving the results of the CBA matrix. The processed lithologies used to create the CBA matrix should be restrained to the relevant ones in regard to targeted mineralization. For example, ignoring Quaternary covers and formations more recent than targeted mineralization gives more coherent results, as only the lithologies that could be linked to their formation will be processed. The main drawback is the possible gaps in the resulting matrix, in area mostly covered by such formations. However, this pre-selection will avoid considering those recent formations as potential favorable factors for mineralization during the favorability scoring stage. If in this case, we know the age of the mineralization events, in other situations where such information is unknown, it would be better to rely on the model to draw its conclusion than to include possible bias into it.

3.2 Development of the DBA

3.2.1 Overcoming the limitation of the CBA

The CBA is an efficient method to analyze the discrete geological data. However, it comes out with some limitation mainly due to the anisotropic nature of the square cell grid used to identify the lithological environments. The use of square cells induces a bias related to the relative position between the cell and the lithological formation or the mineralization occurrences. The shape of the cell means that the outline of a lithological formation or a point representing a mineralization occurrence has more chance to be overlapped by the cell if it is located near its corners, while it has less chance to overlap the cell if it is located near the center of its vertices, even if it is at the same distance from the center of the cell as an overlapped formation near the edge. **Figure 3.5** displays an example for a square grid matrix paving the space along the cartographic X and Y axis, with two similar geological formations of various orientations. As seen, depending on their main cartographic orientations and the position of the cells, they are not overlapped by the same number of cells. This anisotropic search can be problematic for some specific formation such as dykes, veins or all other anisotropic formations with strained shapes or structural control on their development/emplacement. If an isotropic search is performed, using discs to identify the cells close to the formation (**Figure 3.5**), the number of selected cells remain the same, despite the change of orientation, and is only slightly affected by the shift of the grid.

Another problem induced by the use of square cells to determine the lithological associations is the restriction it imposes to the resolution and the number of unique associations. As explained before, the cell size is a critical parameter that determines multiple variables of the CBA, mainly affecting the number of cells (i.e. the resolution of the final grid), the number of unique lithological spectrum found and the diversity of the lithologies inside those spectrum. As the cell size determines both the final resolution of the resulting grid and the association that will be found, there is a threshold to the precision allowed by this method. By improving the resolution, the number of associations will increase, but their diversity will drop, while decreasing the resolution will have the opposite effect (**Figure 3.6**). When a fine resolution is needed, as it could be the case in some MPM applications, the optimal threshold found for the cell size (i.e. the balance between the spatial resolution and the number and diversity of association) may not be fine enough.

Improvements on the CBA method must focus on (i) limiting or suppressing the anisotropic effect induced by the square cells, and (ii) improving the final resolution of the grid without degrading the quality of the lithological spectra found by the matrix. Solutions to those problems must improve the resulting matrix while keeping effectiveness in computation time. One solution to solve those limitation would be to change the shape of the cells used in the grid to a more isotropic geometric shape. This could be achieved using hexagonal shape cells instead of square cells. Indeed, hexagonal cells allow covering all the study area with a more isotropic shape than a square, without overlapping. This solution should improve the results, as it partially solves the anisotropic problem. However, it does not change the limitation in resolution due to the lithological spectrum quality as the cell size optimisation problem would be exactly the same than for square cells; the output grid map being the same as the computation grid to treat and classify associations. In addition, if plotting a hexagonal grid is possible, it also comes with its own disadvantages compared to a usual square grid. First, the offset between even and odd columns could lead to more complex border effects and makes this grid more sensitive to positioning effects. A hexagonal grid would also be more complex to work on existing geophysical or geochemical

Anisotropy and grid shifting effects on identification of lithologies

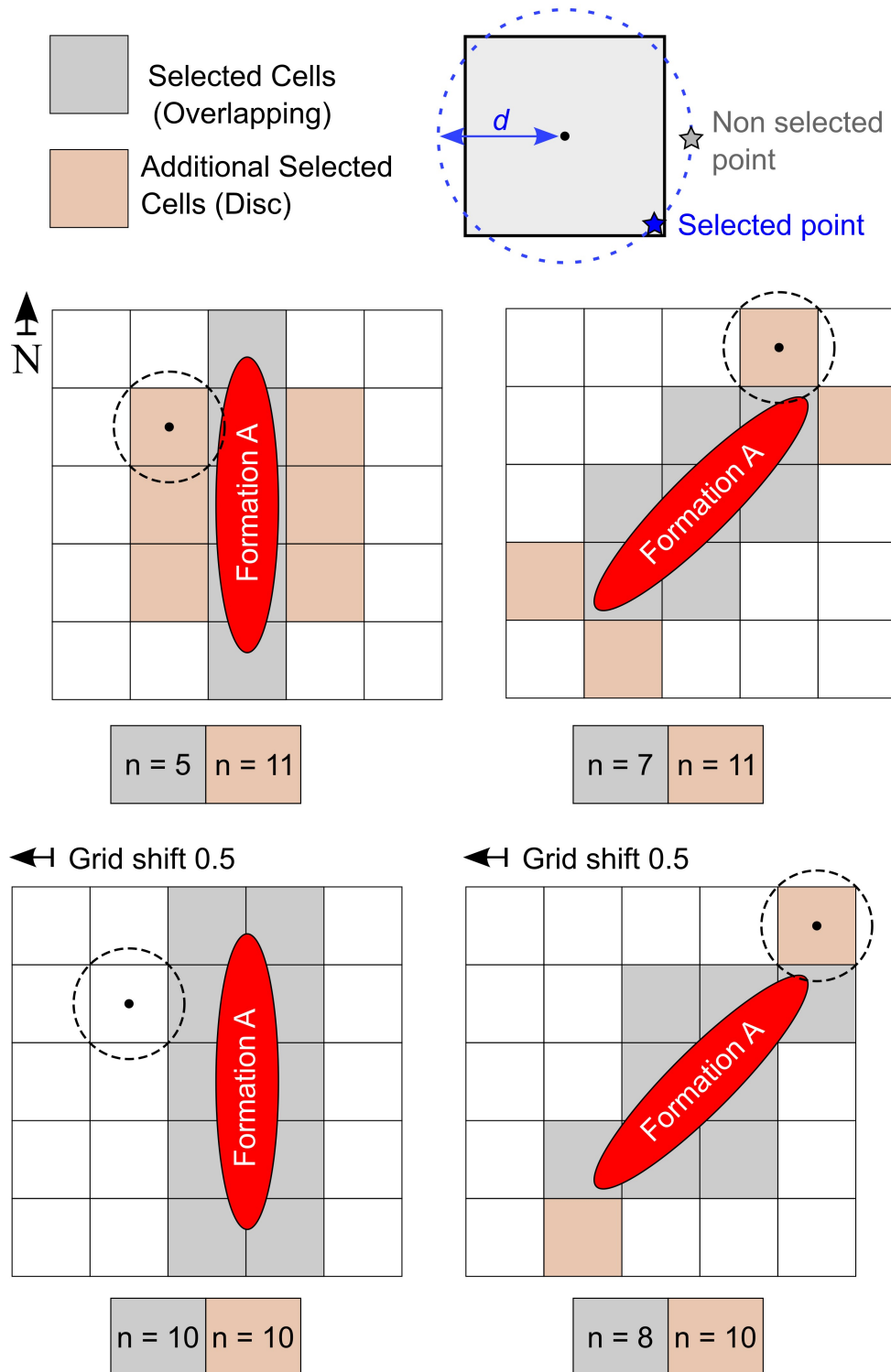


Figure 3.5: Identification of cells in the neighbourhood of a formation using an anisotropic search (CBA) and disc-buffers. The grey cells are overlapping the formation in red, while the pale orange cells are cells within buffer range of the formation. The number of cells of each type are displayed below each situation, the grey cells are taken into account for the number of cells within buffer range.

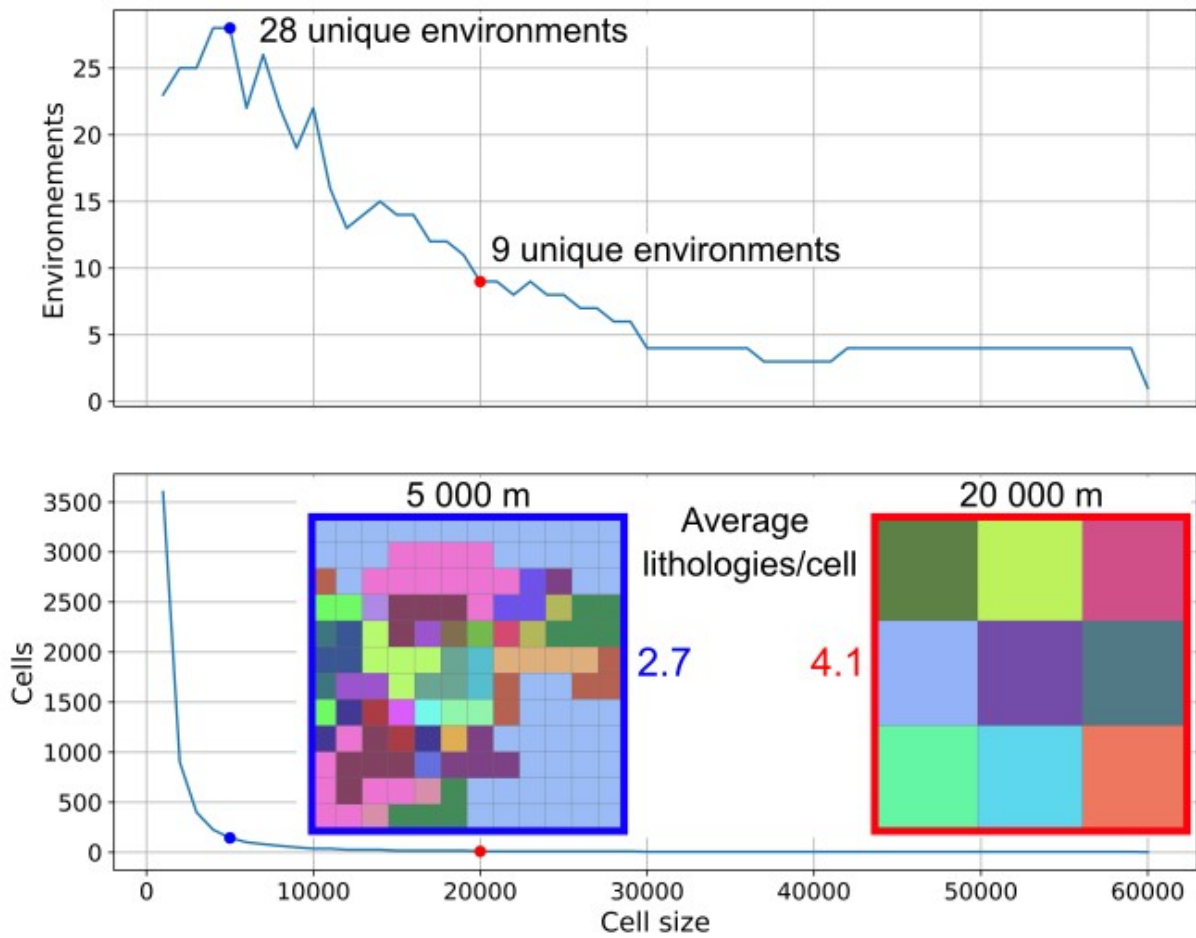


Figure 3.6: Limitation of the CBA method regarding the cell size. Cell size parameter defines the resolution of the grid and the diversity in the lithological association in the cells composing this grid. The 5 000 m cell size (blue) allows good resolution and the highest number of unique associations, at the cost of the diversity of lithology in those cells. The 20 000 m cell size (red) shows a very low resolution with few lithological association, but each association contains more lithologies. This limitation will scale with bigger datasets, covering more surfaces and containing more lithologies.

map which are always coded and furnished as squared raster grids. When exporting the data into raster formats, the produced raster grid, composed of square pixels, would require higher resolution than a square grid to properly render the hexagonal grid without deteriorating its quality. All these features lead to not retaining the hexagonal grids as an appropriate solution to deal with the problems related to the traditional CBA method.

Another solution to solve anisotropy problem is to use discs surrounding areas to explore local geological environments and create the lithological spectrum. Disc radius can thus be easily tuned in order to adapt the best possible area to characterise associations. Yet, disc shape does not allow a regular paving of surfaces over a study area. Here, the grid of discs used for analysing local associations is distinct from the output representation squared cells grid, as follows. The disc buffers are used to identify the associations of geological features (e.g. the lithological formations, mineralization occurrences and structural features) within the studied area. Those discs are plotted across the study area based on points forming their centers over a regular grid. Each point will thus be associated to a local lithological and/or structural environment, and the final lithological association classes can be displayed using a regular grid

based on this point set. This approach has the advantage that the size of the disc buffer can be set independently to the size of the grid; the quality of the lithological spectrum is thus not restricted to the resolution of the grid. It is rather defined by the combination of the size of the buffer and the spacing between each disc (i.e. the resolution of the grid). We define this new method, allowing to overcome the main problems of the CBA, as the so called Disc-Based Association (DBA) which is developed and tested below.

3.2.2 Disc-Based Association principles

The Disc-Based Association (DBA) developed here is a prospective mapping method based on the Cell-Based Association (CBA). Like the CBA, it aims at better integrating geological maps or other qualitative data in the calculation of mineral resources prospectivity maps. Like the CBA, it is based on the discretization of the geological features and the production of lithological spectrum to synthesize and identify various geological environments across the study area. In this case, the discretization is done by plotting a regular dot plot grid across the area. This allows the identification of local surrounding geological environments around each dot and the creation of an associated matrix containing the lithological spectrum. A lithological spectrum is composed of n binary values corresponding to the n total number of lithologies present inside the overall study area. A lithological spectrum is created for each point produced by this dot plot using a neighbouring search based on disc-shaped buffer. The resulting lithological spectrum displays the lithologies found by this neighbouring search: if the lithology is present in the neighbourhood of the point, it will be coded as "1", otherwise it will be coded as "0" (**Figure 3.7**). A squared cells grid is created and centered on the dot plot grid to display results from the oncoming statistical analysis. It is noteworthy that this final grid resolution depends on the interdistance between disc centers and not their radius value. A high number of lithological spectrum can be produced, and an unsupervised classification (AHC, HCPC) can be applied to highlight the main classes.

Other data can be added to the DBA matrix, to complement the geological environment and give more "explanatory" variables, or predictors, to the following predictive model. Structural data can be added using the neighbouring search process used for the lithologies (**Figure 3.8**). In this case, a new column is added in the DBA matrix to indicate the presence or absence of structural features in the neighbourhood of each point of the DBA matrix. Even if the final square cell matrix is produced, the initial plot grid can be recovered using the center of the square cells. Using the plot grid, a new neighbouring search is performed for the structural features: if they are found in the vicinity of a point, it is translated by a "1" in the lithological spectrum, otherwise by a "0". As this process is done independently to the main lithological associations identifications, it is possible to set specific disc buffer values when searching structural features. This allows taking into account the imprecision of the input data that could be different from one to the other map data set (lithological and structural maps, for instance) or to simulate the fact that some mineralization or structural features can occupy a significant volume and thus cover a surface that is not represented on the map.

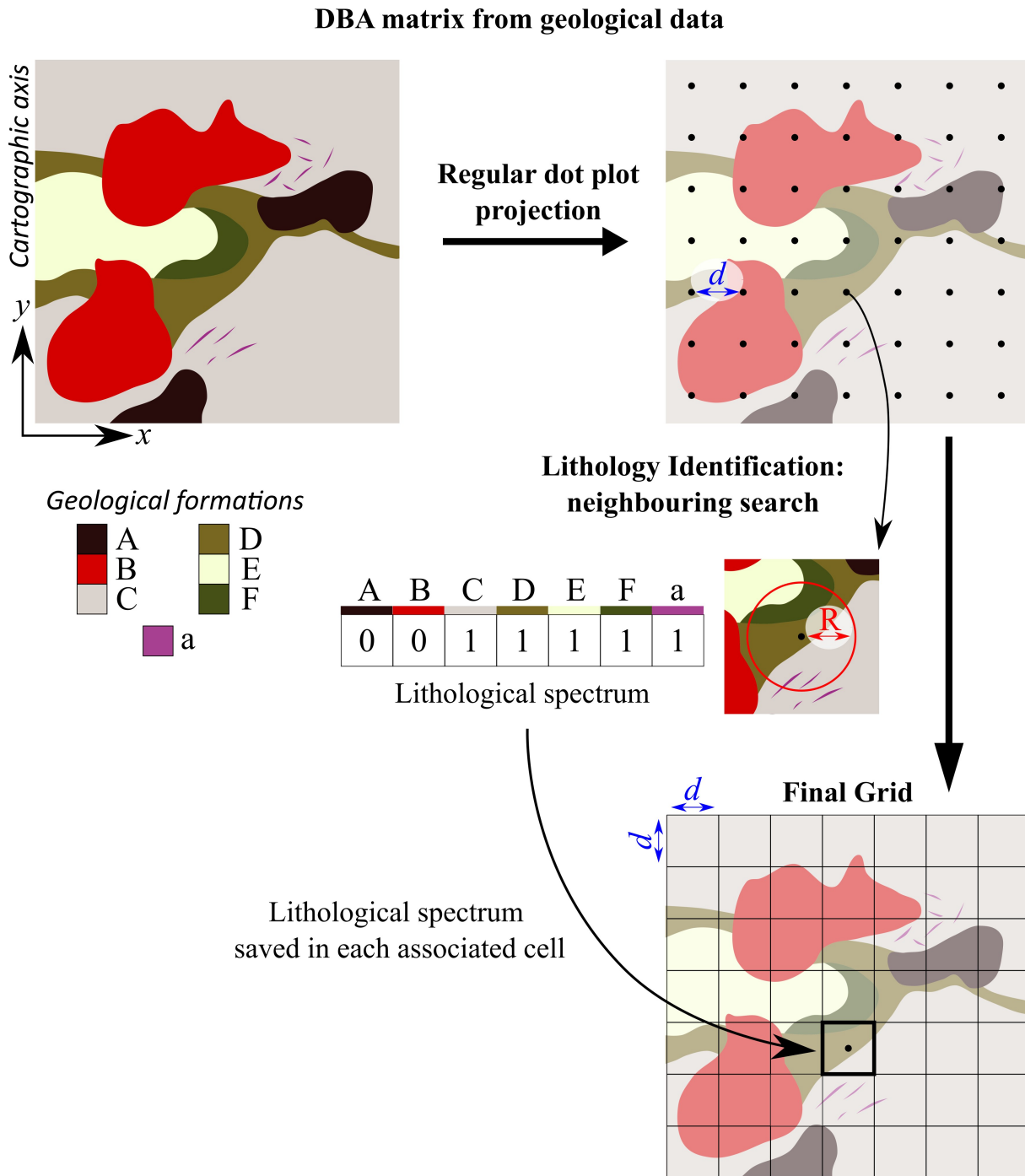


Figure 3.7: Schematics of the DBA method for lithological association search. The main parameters of this method are the spacing between the points d , defining the dot plot, and the radius of the neighbouring search R . The spacing d also defines the resolution of the resulting grid.

Adding new data in the DBA matrix

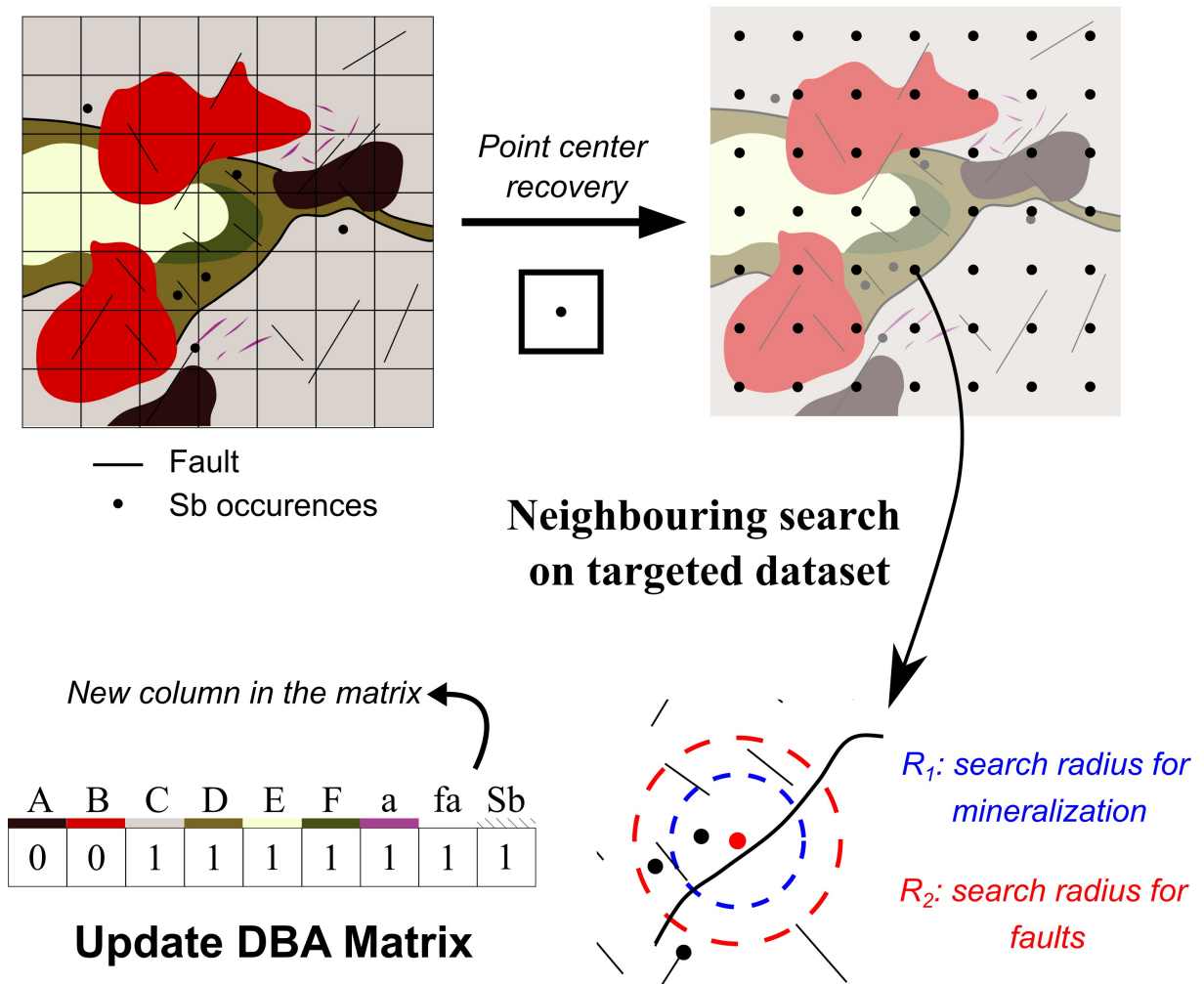


Figure 3.8: Integration of new data in the DBA matrix. The resulting grid can be reverted into the dot plot using the center of each cell. From this dot plot, a neighbouring search is performed using the R value provided by the operator and the DBA matrix is updated by adding new columns. A different value can be set for each dataset added in the matrix.

Mineralization are also added to the DBA matrix, this feature being the target of the MPM study and used to train the following predictive model. This addition follows the same process: a new column is added in the DBA matrix to indicate the presence or absence of mineralization occurrences in the neighbourhood (i.e. buffer) of the point of the DBA matrix using the center of the square cells from the final matrix. A specific disc buffer value can be set to take into account the imprecision of the input data or the fact that some types of mineralization can occupy a significant volume and thus cover a certain surface on the map.

The DBA is controlled by two parameters: the size of the neighbouring search R (i.e. the radius of the disc buffer used) and the value of the spacing d between the points of the regular plot grid. The first is critical regarding the diversity of lithological spectrum found using the DBA: increasing its value raises the diversity of the lithological spectrum (i.e. the average number of lithologies present in the lithological spectrum). The spacing value determines the resolution of the final grid, as its value is equivalent to the

cell size of this output representation grid. As so, it plays a role in the number of lithological associations found using the DBA: decreasing the spacing results in more points in the dot plot and a higher resolution for the final grid. More points means that more neighbouring searches are computed and more lithological associations are produced, while the location of those points will also change depending on the spacing value.

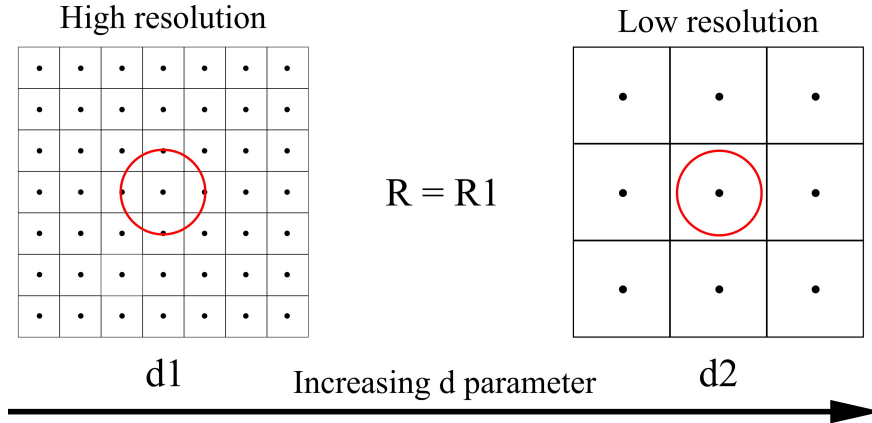
The value of the spacing and the size of the neighbouring search should not be fully set independently to one another: for a fixed buffer value, changing the spacing (**Figure 3.9.a**) will not have the same effect as changing the buffer value for a fixed spacing (**Figure 3.9.b**). In the first case, the resolution and the number of lithological spectrum will both change. In the other case, it will mainly affect the lithological spectrum found. Furthermore, increasing the spacing without increasing the size of the disc buffer can lead to some parts of the studied area which are not treated: if the spacing is higher than the size of the buffer disc diameter, some area will not be covered by the neighbouring search and the resulting lithological spectrum will be degraded. In order to simplify managing the parameter of the DBA grid, rather than using the radius of the neighbouring search R and the value of the spacing d independently we will use one of these parameters and the R/d ratio (**Figure 3.9.c**). This allows keeping a more stable relationship between those parameters and to avoid the case of untreated areas in the process.

3.2.3 Overlapping

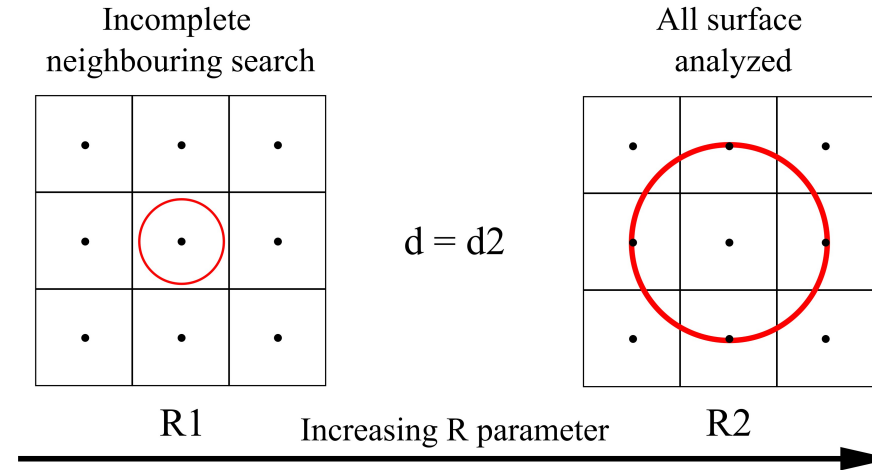
In the DBA method, the surface covered by the standard disc will vary as a function of the R/d ratio value. If this value is low enough, some part of the map could be ignored as the discs will not cover all of the surface area. This will results in degraded lithological spectrum, as some formations, depending on their location, will be under-represented or even ignored. To avoid this problem, the final grid must cover all of the studied area and the radius of the disc buffer must be large enough to, at least, cover the grid cell area. This is done by determining a threshold value for the R/d ratio where the disc buffer produced from a point will cover the whole surface of its associated cell. This threshold value can be determined mathematically as it corresponds to the half length of the diagonal of a 1×1 square: $\sqrt{2}/2$, a value of approximately 0.71. As long as the R/d ratio is higher than this value, all of the studied area is covered by the neighbouring search.

However, another issue appears as for R/d value higher than 0.5, corresponding to an inscribed circle within the square cell, the buffer discs used will be overlapping; the higher the R/d ratio value is, the more discs overlap. This overlapping means that some areas will be covered by more buffer disc than others and will be investigated several times. If this characteristic of the method has less impact on the quality of the resulting lithological associations than dismissing some areas when the R/d ratio is too low, it is worth evaluating its evolution depending on the R/d value. As such, the overlapping of buffer discs have been studied using mathematical and empirical methods. Using the mathematical approach, equations have been produced to quantify the overlapped surface depending on the number of discs overlapping the surface (**Figure 3.10**). The use of mathematical formulae to evaluate the overlapped surface is efficient but quite tricky to handle as a specific equations must be defined for each number of overlapping, and those equations will change each time the number of discs overlapping increases. The complex shapes formed by the overlapping discs makes this method difficult to apply for high R/d ratio. Mathematical development is presented here for a configuration with a maximum of 4 overlapping discs (**Figure 3.10**), with a maximum R/d ratio of 1, resulting in nine different formulae to compute the overlapping surfaces.

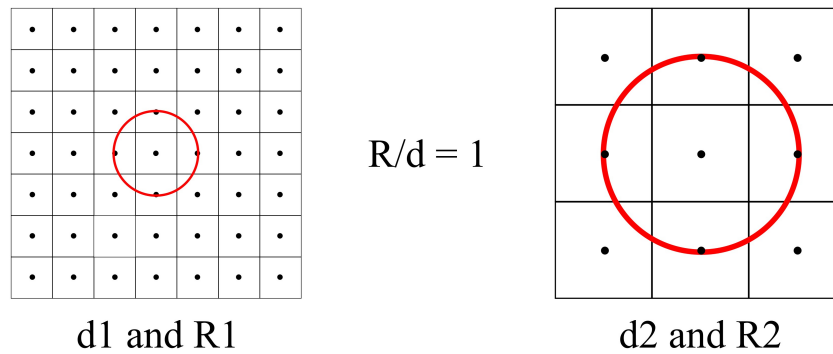
a. Changing the d parameter for a fixed R value



b. Changing the R parameter for a fixed d value



c. R/d Ratio



Constant R/d parameters for various R and d values

Figure 3.9: Effect of the parameters R and R/d on the resulting DBA grid. a. Effects of an increasing d parameter for a fixed R parameter on the DBA grid. b. Effects of an increasing R parameter for a fixed d parameter on the DBA grid. c. Introduction of the R/d ratio to keep a constant relationship between the size of the grid and the size of the cells.

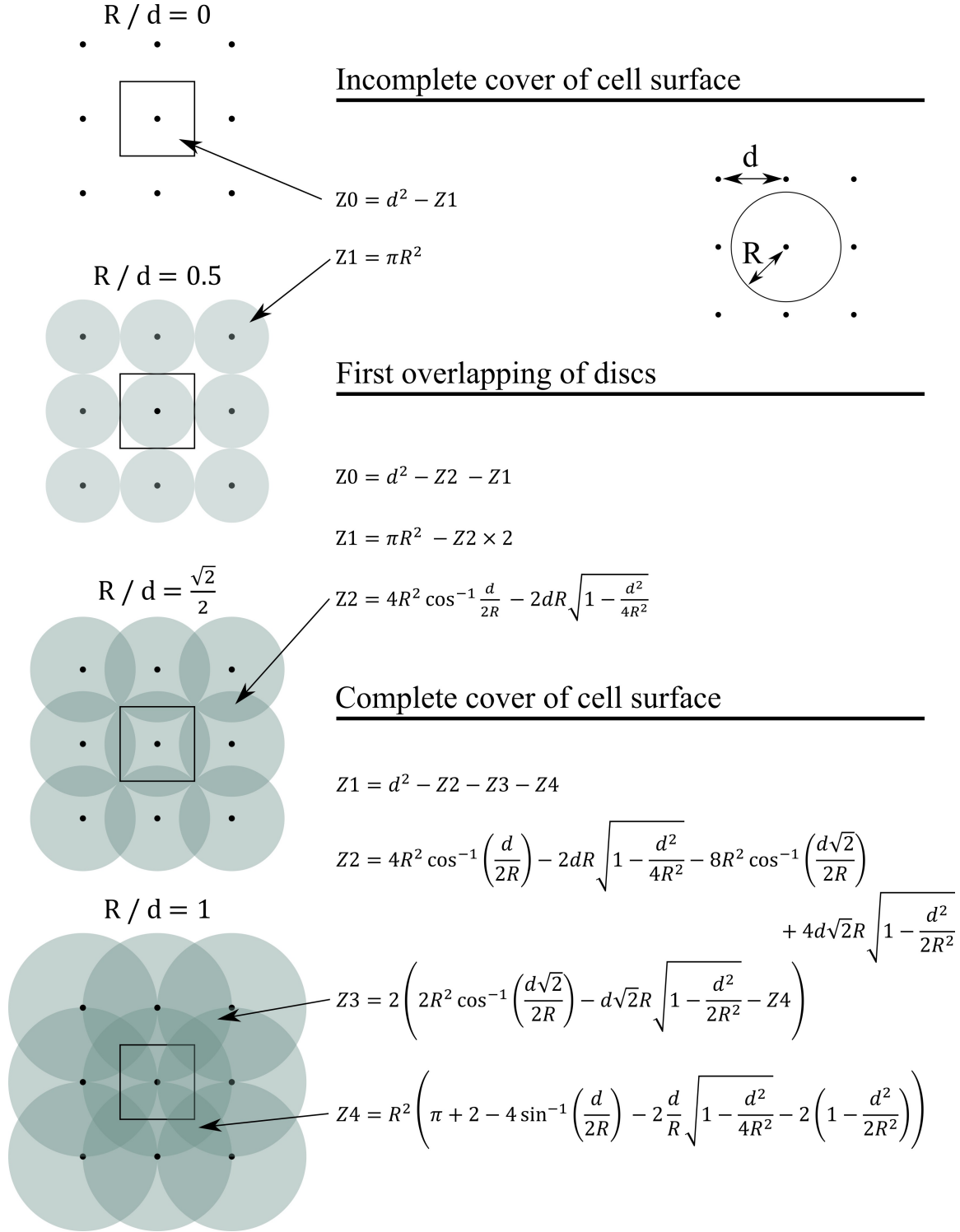


Figure 3.10: Mathematical approach to evaluate the overlapping of multiple discs for a R/d ranging from 0 to 1. 4 thresholds are identified ($R/d = 0, 0.5, \sqrt{2}/2$ and 1). Between each threshold, the mathematical formulae displayed allow the calculation of the overlapped surface: Z_0 for the surface covered by no disc, Z_1 for the surface covered by one disc, Z_2 for a surface covered by 2 discs etc... The Z_n surface are calculated within the black square cell.

The empirical method, on the other hand, relies on numerical simulations to evaluate the overlapping surfaces of the discs (**Figure 3.11**). The overlapping is simulated using a DBA matrix of 22×22 cells, centered on one point with a $2\,000 \times 2\,000$ pixel matrix to compute the surfaces, for a R/d ranging from 0 to a maximum of 21 with an increasing step of 0.005 between each iteration. In this simulation, the inter-distance d between the points is fixed and only the radius R of the discs changes. The area covered by this simulation is centered on one point and extends to its 8 adjacent points (see **Figure 3.11**). The area obtained through this simulation can be different to the one obtained in real application of the DBA method, as this simulation doesn't take into account borders effect, especially when considering high R/d ratios. The values of the pixel matrix are recovered, the value of each pixel corresponding to the number of discs overlapping in it. We consider the number of pixels associated to a specific number of overlapping to be the surface covered by this number of overlapping at a given R/d ratio. This allows estimating the coverage of each overlapping discs, based on their surface (i.e. the number of pixels covered by each overlapping discs). For each R/d ratio computed, the average, maximum, minimum and median value of overlapping is also recovered. Comparison of values between the range of 0 to 1 R/d ratio shows similar results with the mathematical solution. However, compared to the mathematical solutions, this method is easier to apply and allows the computation of overlapping for higher value of R/d without adding manual inputs at each overlapping steps.

These numerical and mathematical simulations allow to produce abacus linking the R/d ratio to the percentage of surfaces of the cells being overlapped by other discs and the number of discs sharing the corresponding overlapped area (**Figure 3.11**). Due to the complex nature of the shapes with multiples overlapping, their proportional surfaces values are highly variable depending on the R/d ratio. As mentioned above, a R/d ratio under 0.71 results in areas of the cells not being covered by any disc buffers (see **Z0, Figure 3.10 and 3.11**).

3.2.4 Critical parameters

The two parameters R and d are key parameters to implement the DBA. As shown before, we recommend to focus on one of those parameters associated with the R/d ratio value to determine the property of the DBA grid. In this study, we choose to rely on the R parameter (i.e. the radius of the discs buffers used in the neighbouring search). Using the R/d ratio rather than the d spacing will allow to keep the same overlapping values between the discs when changing the size of the buffer discs R during the different tests of the implementation. As such, changing R parameter for a fixed value of R/d will not only affect the size of the neighbouring search for the lithological association, but also the resolution of the grid: increasing its value will decrease the resolution of the grid, while decreasing the number of unique associations and increasing the diversity in those associations. On the other hand, changing the R/d value for a fixed R value will have a major effect on the overlapping between the discs, as increasing its values will increase the number of overlapping in the studied area. Decreasing this ratio value also has similar effect than increasing the R value, as it will decrease the resolution and therefore the number of unique associations, while increasing the diversity of lithologies inside the lithological associations. A balance between high output spatial resolution and the quality of the lithological association is needed to compute an effective DBA matrix.

Both R and R/d are critical parameters determining the outcome of the DBA method. It is necessary to determine an optimal range of value for both parameters. In the case of the R/d ratio, the first criteria is to determine a value equal or higher than 0.71 in order to cover all the study area during the

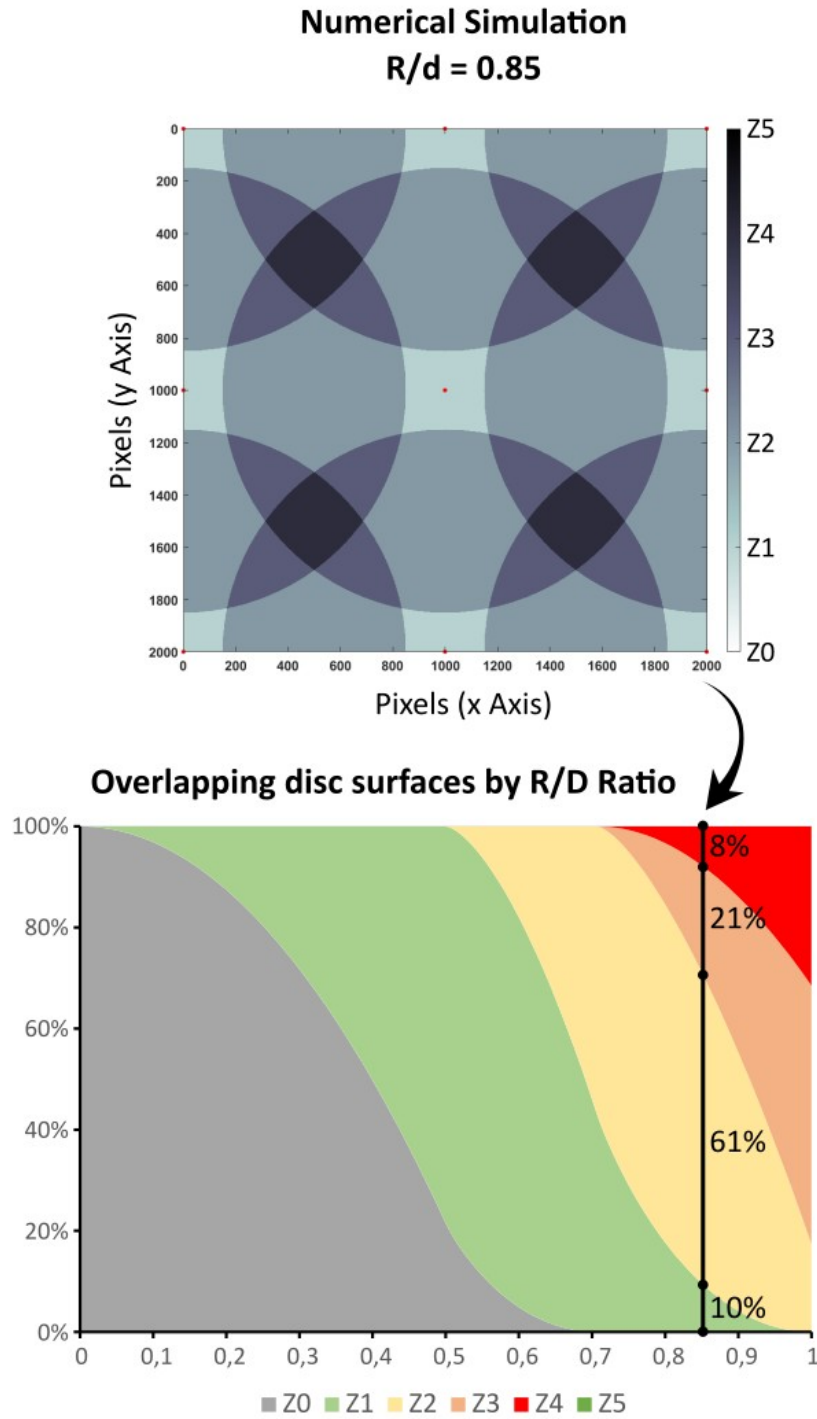


Figure 3.11: Numerical simulation of the overlapping surfaces between discs in a regular dot plot. The evaluation of the different overlapping surfaces is done using a 2 000 x 2 000 pixel matrix. The position of the dot plot is fixed and only the radius R of the discs are changed for each iteration, depending on the R/d value. The step between each iteration is 0.005, ranging from 0 to 21. Results of each iteration are saved, resulting in the abacus presented below.

neighboring search (see **Annex A.1 and A.2** for corresponding figure). Determining an optimal range of values for both parameters is complex, as these two parameters are not fully independent. A first approach to determine an optimal range of values for both parameters is to use a multi-scale iterating process. This allows the production of similar graphs showing (i) the total number of distinct geological associations extracted from the DBA (**Figure 3.12.a**), (ii) the total number of cells in the grid (**Figure 3.12.b**), (iii) the average number of distinct lithologies in a cell (**Figure 3.12.c**) and (iv) the frequency of cells including only one lithology (**Figure 3.12.d**). Based on this abacus, it is possible to choose the range of values for both R/d and R resulting in the best balance between a high enough grid resolution and quality of the resulting lithological associations (i.e. quality of the information they contain).

The total number of distinct geological associations in the DBA matrix allows to assess the evolution in the number of unique geological environments found by the DBA. The optimal range of values for the matrix should allow a high number of unique environments, generally found at lower values of R . On certain case study, under a certain R threshold, this value can decrease depending on the case study. This behaviour is displayed here in **Figure 3.12.a**, with a drop in the number of distinct environments below 8 000 m for a R/d of 3, and below 3 000 m for a R/d of 0.5. As such, the optimal range of values is located after this drop and before the plateau at higher R values, around 30 000 m in **Figure 3.12.a**. Increasing the R/d ratio will increase the number of distinct environments. The total number of cells in the grid display the evolution of the resolution depending on R and R/d , sharply increasing at low values of R and greatly increasing with the value of the R/d ratio, as shown in (**Figure 3.12.b**). For the lowest value of R , 1 000 m, a R/d of 0.5 results in 900 cells, while a R/d of 3 results in 32 041 cells, 35 times more. However, the number of cells quickly drops, with a total number of cells of 3 600 for R of 3 000 m a R/d of 3. Setting the R and R/d parameter to have the highest resolution would result in a lower number of distinct environments (**Figure 3.12.a**), less diversity within those environments (**Figure 3.12.c**) and a high percentage of 1 lithology cells (**Figure 3.12.d**). As such, the optimal range has to be set aiming at a reduced number of cells, in the downward trend of the curve or even the plateau. The average number of lithologies in a cell shows a different trend (**Figure 3.12.c**) than previous curves, as increasing R will increase this value, showing the increase in diversity within the unique environments. This value usually shows a linear increasing, and is more indicative than determinant when searching for an optimal range of R and R/d values. Finally, the percentage of 1 lithology cells shows a steady decrease with increasing value of R , while decreasing R/d tends to result in sharper variations in this trend, with values jumping between 40 and 0% with a R/d of 0.5. As we want to minimize the number of 1 lithology cells, the optimal range of values should be located near the plateau, with high R values. However, this optimal range of value has to be low enough to allow having enough unique environments and a resolution high enough to perform the predictive mapping.

The overlapping values depending on the R/d ratio can be evaluated using the results of the numerical simulations. The produced abacus allows the evaluation for each value of R/d of the number of overlapping with its associated surface along with the mean, median, maximum and minimum number of overlapping obtained with the simulation. Using this abacus, another way to define the optimal R/d value is to use a parameter called "*Dispersion*" in this study. This parameter is computed for each value of R/d , using the following formula:

$$Dispersion = \frac{Maximum - Minimum}{Median} = \frac{Range}{Median}$$

The *Dispersion* variable aims at estimating the variability of the overlapping coverage of the discs.

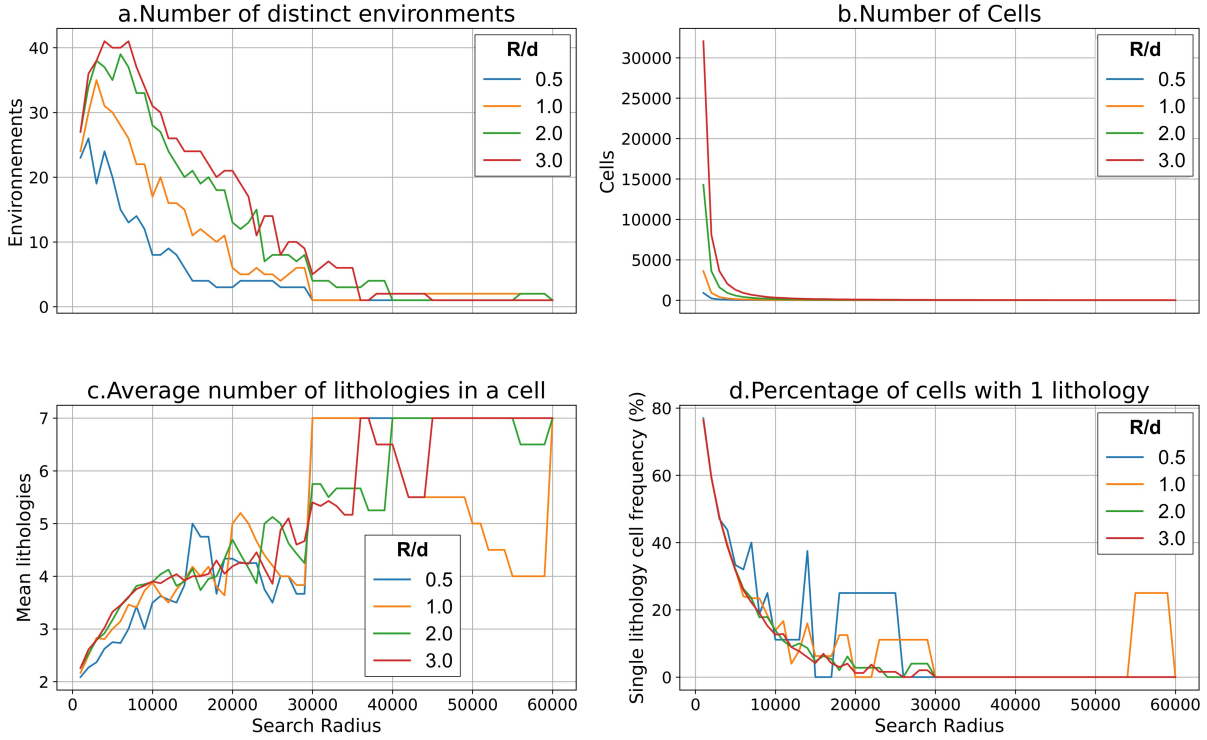


Figure 3.12: a. to d. Impact of the search radius R and the R/d ratio on different parameters of the DBA matrix, applied on the previous sketch map with search radius ranging from 1 000 to 60 000 m with steps of 1 000 m and various R/d values. The size of the area covered by the sketch map is a square of 60 000 x 60 000 m.

The use of this value to determine the optimal R/d value is based on the hypothesis that minimizing the discrepancy in the association searches increases the quality of the resulting grid as it leads to a "homogeneous" search in the study area. This discrepancy is here evaluated using the *Range* and median values of the pixels. The *Range* value evaluates the discrepancy of the analysis by comparing the maximum value of overlapping (e.g. the most investigated areas) to the minimum value of overlapping (e.g. the least investigated areas). This discrepancy is weighted by the median value, evaluating the number of overlapping associated with this *Range* value. The mean value could be used to replace the median value, allowing to take into account the relative surface of the overlapping. Increasing the range will increase the discrepancy and the resulting *Dispersion*, while increasing the median value will decrease the importance of this discrepancy. The abacus and resulting *Dispersion* graph are shown in (Annexes A.5, A.4, A.3) with the graphs of mean, maximum, minimum, median, *Dispersion* and *Range* produced for a R/d ranging from 0 to 21. The *Range* values increase at a slow rate compared to the median values with increasing R/d ratio. This shows that the *Dispersion* value decreases as the R/d ratio grows, approaching 0. As such, a high *Dispersion* will occur at low R/d values corresponding to a configuration with limited overlapping, but also with some area being comparatively much more overlapped than others. As the *Range* varies highly even if following a stable trend, the *Dispersion* values vary also locally, even if following a trend toward lowering. Using the *Dispersion* to find an optimal R/d value will mainly consist in determining a threshold for the maximum desired *Dispersion*, and find the lowest R/d ratio resulting in a value under this threshold. This method can be used to select one or few R/d ratio to perform the multi-scale iterating process, saving computing times, or to determine the properties of each R/d value used in the process.

Chapter 4

Machine learning for prospectivity mapping

*This **Chapter 4** displays the second step of the developed inverse methodology. After the development of a method to produce local geological environments, described in **Chapter 3**, **Chapter 4** shows the development of a method allowing the quantification of the relationships between local geological environments and the mineralization, using prospectivity mapping. First, methods that were commonly used with the CBA in previous studies are presented to give a baseline of the applications of CBA in prospectivity mapping. The development of a method pairing either a CBA or a DBA grid with a machine learning supervised classification method, namely the Random Forest (RF), to produce prospective maps is then detailed. This coupling aims at producing better prospective maps than the methods already in use with the CBA. The evaluation and validation processes of this paired method are also described, as well as the characteristics of this supervised classification. Resulting prospective maps and favorability scores are then used during the post-processing and final step of the developed methodology, presented in the next **Chapter 5**.*

4.1 Prospectivity mapping with CBA

4.1.1 The "exact cell signature" of the CBA method

The *exact signature method* [Gloaguen and Tourlière, 2016a] is a prospectivity map method relying on the direct output from the CBA matrix. In this matrix, each cell of the square grid contains an association of geological features defined by their intersection with the cells. By selecting the cells associated with the mineralization, it is possible to determine the various lithologies spatially linked to mineralization. The resulting list of 'favorable' lithological associations is then used to select all cells presenting the same lithological associations. If considering that cells showing the same geological environments — i.e. exactly the same lithological associations here — than 'mineralized' ones, this selection allows outlining the favorable zones for further exploration in the study area. Moreover, this method allows to differentiate potential distinct favorable contexts from distinct lithological associations and to differentiate the various types of mineralization.

Such favorability map is only qualitative: no score are attributed to the recognized favorable areas, even if the exact similitude with the mineralized environments ensures that these zones are favorable. However, this method seems rather restrictive in the sense that the resulting CBA matrix can produce a very high number of associations depending on the input datasets. As mineralization are most likely to be scarce across the study area, the lithological association(s) linked to them will cover a limited surface, resulting in limited surfaces of favorable areas with similar lithological associations.

4.1.2 CBA / AHC classification

Depending on the size of the grid made during space discretization, CBA can lead to a rather large number of possible lithological spectra. Indeed, the theoretical maximum number of environments increases drastically with the number n of lithological formations within the study area, the theoretical maximum number of associations being 2^n . In order to estimate the similitude of a given spectrum in a cell with the reference 'favorable' ones (i.e. the ones containing mineralized occurrence(s) in our case), another approach consists in using classification or clustering techniques to segment the distinct associations to be considered.

The Agglomerative Hierarchical Clustering (AHC, [Lance and Williams \[1967\]](#)) has already been tested and applied by [Tourlière et al. \[2015\]](#) in association with the CBA matrix. The whole matrix produced by the CBA is classified using the AHC to obtain a more synthetic map of the lithological associations (**Figure 4.1.a**). The AHC consists in grouping individual observations (here the lithological associations) by successive aggregations, until obtaining a single class [[Rham, 1980](#)]. These successive aggregations are based on a distance matrix defining the likelihood between each individual. Various metrics can be used to establish this distance matrix, in the case of the CBA matrix, an euclidean distance is computed between each lithological association, based on the binary digits composing the spectrum (**Figure 4.1.b**).

These successive aggregations are presented in the form of a tree-shaped graph called dendrogram. The number of distinct classes to retain is defined using this dendrogram and an expert evaluation of the quality of the classes produced. This quality can be evaluated by assessing the homogeneity of the associations obtained within each class or by using the inertia parameter produced by the AHC. Each class with at least one occurrence within one of its corresponding cells will be selected. Then, all the cells belonging to one of those classes are extracted (**Figure 4.1.c**). These classes and their related cells are considered as favourable ones. It could be possible to discriminate these classes depending on the number of occurrences they contain, but such approach could be biased if considering possible clustered occurrences, as discussed before. It is alternatively also possible to consider the total area they cover by cumulating the total number of intersected cells.

This segmentation procedure is rather simple to handle, and it allows a quick identification of the geological framework surrounding mineralization. Yet, it is mainly a qualitative way of evaluating the area of interest in the study area as it does not give weight to the results, aside from using the number of mineralization occurrences within the classes.

4.1.3 CBA / Ranking

Favorability maps can be computed from CBA grids without using a classification or clustering method to simplify the associations. Both [Gloaguen and Tourlière \[2016a\]](#) and [Vic and Tourlière \[2016\]](#) used CBA to

CBA favorability maps with AHC classification

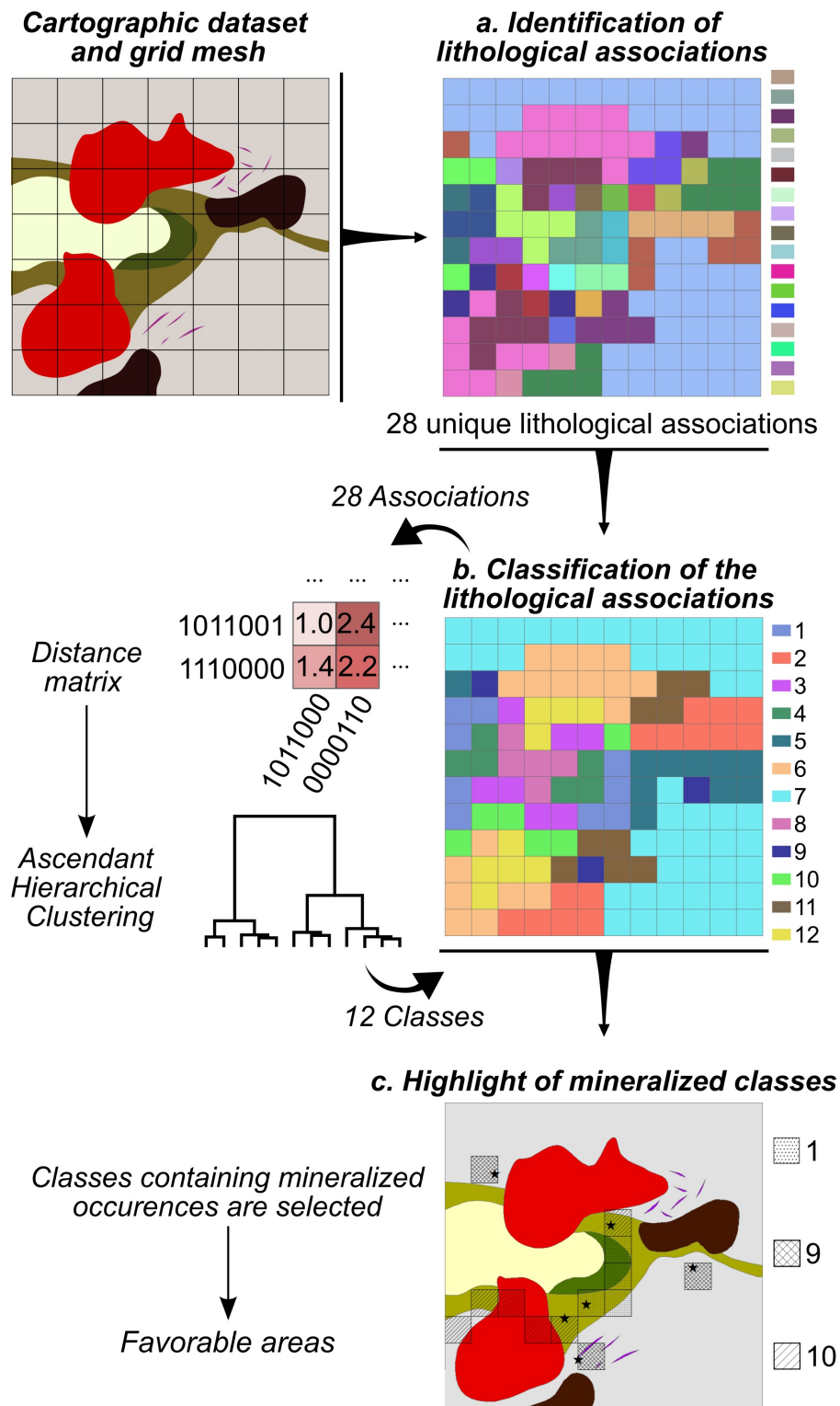


Figure 4.1: Sketch example of the application of CBA and AHC classification method. a. The CBA is applied on the lithological data to identify the lithological associations. b. Resulting lithological spectrum are classified using the AHC method. A distance matrix is produced from each unique association; the aggregation is computed and the optimal numbers of classes are selected. c. Selection of mineralized classes to highlight favorable areas.

produce favorability maps using a "Ranking" method in addition to the favorability maps produced with the AHC classes. This method, which is broadly inspired by the Weight of Evidence (WoE) approach [Agterberg et al., 1993], is based on a frequency estimation of the presence of lithological formations within the cells produced by the CBA. After the production of lithological associations using CBA (**Figure 4.2.a**), a *Ratio* is computed for each lithology i , based on the Frequency in Standard Cells (FSC), which is the frequency on mineralized cells containing i , and the Frequency in Study Area (FSA), the frequency of cells containing i . From this *Ratio* score, a Ranking score is calculated for each cell based on the sum of the *Ratio* of the formations present inside this cell (**Figure 4.2.b**).

$$Ratio_i = \frac{FSC_i}{FSA_i}$$

$$FSC_i = \frac{\text{number of mineralized cells with the lithology } i}{\text{total number of mineralized cells}}$$

$$FSA_i = \frac{\text{number of cells with the lithology } i}{\text{total number of cells in the CBA grid}}$$

FSC: Frequency in Standard Cells (mineralized cells).

FSA: Frequency in Study Area.

$$Ranking_{cell} = \sum_{i=1}^n Ratio_i$$

The resulting predictive map shows the value of Ranking computed for each cell (**Figure 4.2.c**). A *Ratio* value greater than 1 corresponds to a favorable lithological formation. Indeed, this means that this formation is statistically encountered more frequently within the mineralized cells than in non-mineralized ones. Formations that have no links with the mineralized cells will have lower *Ratio* values: a widespread formation will have a high FSC, but also a high FSA, and a rare or even absent formation in the mineralized cells will have a low FSC.

The higher the score of a cell is, the higher the number of favorable formations inside this cell is. The cells with the higher scores will be the one containing lithological associations similar to the cells containing occurrences of mineralization. The most favorable areas will be composed of cells with higher scores, with more similarity with the geological environments of the mineralization. However, this method favours complex geological environments, as cells with various lithologies are more likely to contain formations with good *Ratio* and could end up with a higher score than mineralized cells. There is also a cumulative effect, as the sum of the *Ratios* is considered: if several lithologies with an average or even low *Ratios* are summed, they can reach a significantly high value. It is also unable to differentiate distinct environments surrounding the mineralization, depending on the nature of the mineralization. However, this method gives interesting results and requires low computation time when applied to the CBA.

CBA favorability maps with Ranking

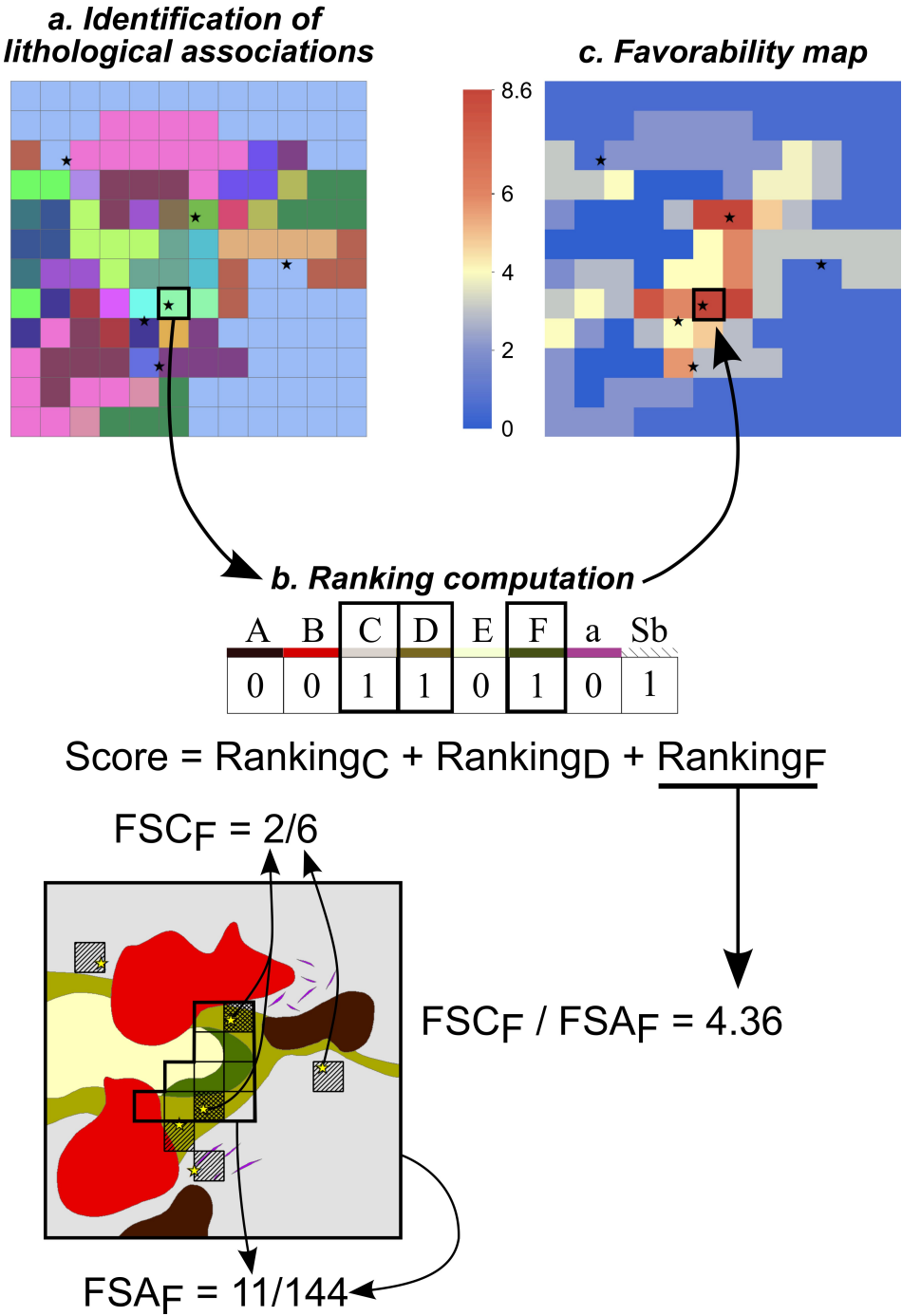


Figure 4.2: Sketch example of the application of CBA and Ranking predictive mapping method. a. The CBA is applied on the dataset to identify the lithological associations and the mineralized cells. b. The Ranking of each cell is computed based on the *Ratio* of each lithology present in its association. This *Ratio* is based on the frequency of mineralized cells containing the lithology and the frequency of the lithology in the grid. c. The favorability map can be displayed: in red the cells with the highest score, in blue the lowest score.

4.2 Random Forest-based prospectivity mapping applied on CBA and DBA matrices

4.2.1 Workflow of CBA and DBA matrices-based RF prospective mapping

Random Forest (RF) is a supervised machine learning method used for classification and regression. This method was formalized by Breiman [2001] and has been used in multiple works of MPM [McKay and Harris, 2016, Sun et al., 2019, Harris et al., 2015, Carranza and Laborte, 2015b,a, Rodriguez-Galiano et al., 2015]. In this section, the development of a method coupling CBA and DBA matrices analyses of the geological dataset with RF is presented. This method is employed to build the predictive model through a data driven approach and with the less possible expert guided input. In details, RF is used to (i) classify cells (i.e. mineralized/non-mineralized classes) using the multi-variable information or predictors – including lithologies, structures and occurrences – contained in the matrix produced by the CBA / DBA and (ii) quantify the similitude level between different zones of the study area, and more specifically with the ones with known deposits.

The main objective of the integration of the RF to the CBA method is to improve its results compared to other predictive methods, such as the Ranking scoring or the AHC classification presented above. As detailed below, the application of RF on the matrices allows the integration of both continuous and discrete variables (**Figure 4.3.a**) to perform classification; a big advantage as the methods previously shown require the discretization of continuous variables before the classification process. Here, variables are included in the associations matrix produced either through CBA or DBA process (see above and **Figure 4.3.b**). RF predictive modelling is then performed using the information contained in the association matrix, computing a binary classification between the mineralized and non-mineralized cells of the matrix (**Figure 4.3.c**). RF uses a more advanced classification process instead of statistical computations, allowing to identify environments showing the highest likelihood to host mineralization. Favorability scores are then attributed to each map cell based on the results of the RF (**Figure 4.3.d**). Results are then validated using expert knowledge and statistical analysis (**Figure 4.3.e**).

The CBA and DBA methods allow an analysis of metallotects and geological features by using the geological associations found by the matrix in both mineralized cells and favorable areas. The CBA / DBA RF method also evaluates the performance of the integrated data and allows identifying which variables are the most efficient to identify potential mineralized areas, owing to the capacity of RF to evaluate the discriminating potential of each variable. This feature is useful to provide a more precise guide for exploration and to improve the metallogenic models associated with studied mineralization and in turn their genetic models. Such application might allow defining ‘automatic’ metallotects. The application of this method for antimony prospective mapping through the Armorican Massif is displayed in the **Section IV - Applications (p.185)**, as a proof of concept, using discrete datasets comprising lithological units and structural data.

4.2.2 Targetting of the RF

RF classification of multiple variables is based on targets. Basically, this predictive modeling aims at determining whether a cell is more likely to host mineralization than the others using the lithological association in the mineralized cells, as found by the CBA or the DBA. The RF classification is based on

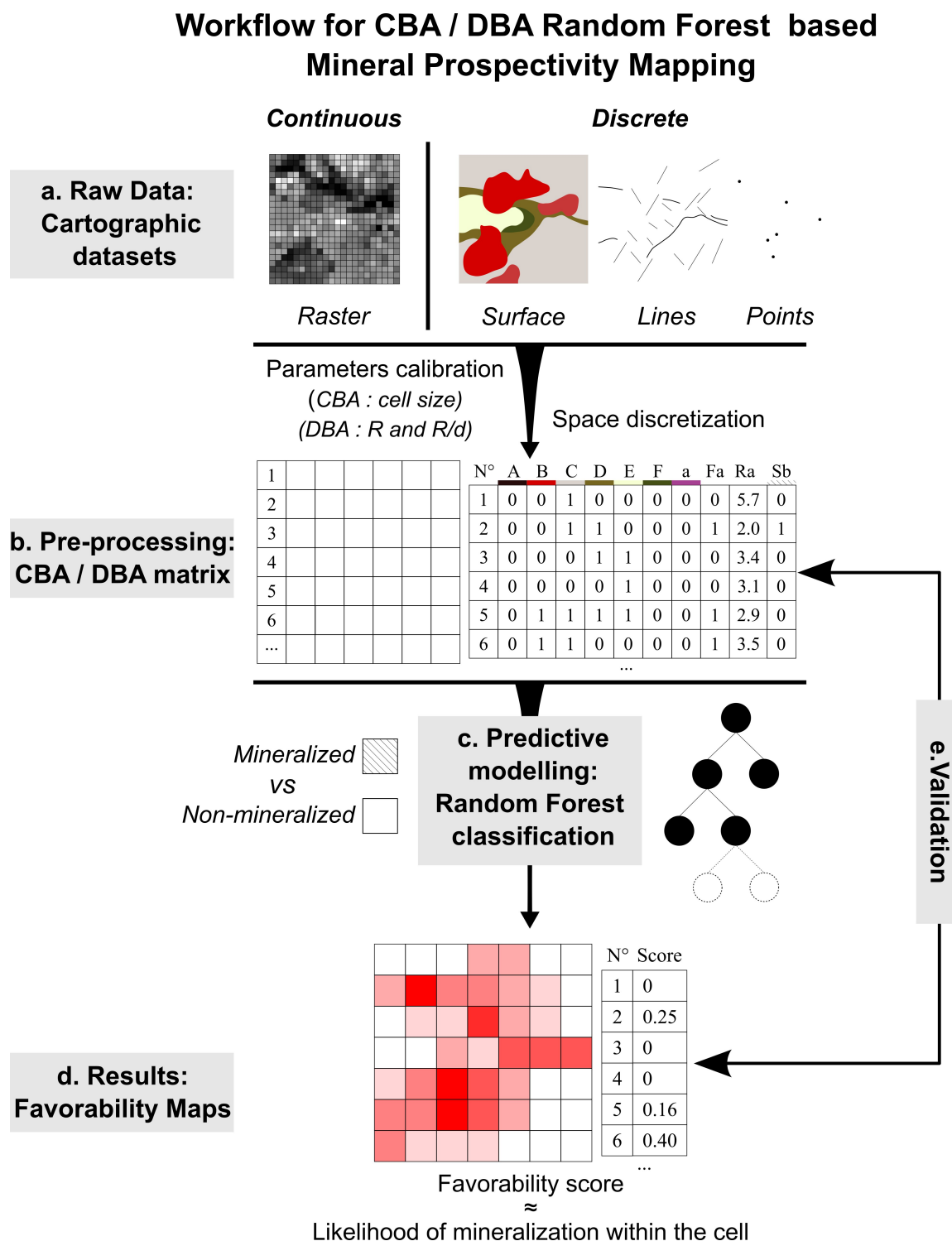


Figure 4.3: Workflow for the CBA / DBA RF predictive mapping method, CBA or DBA can both be used before the RF computation. a. The raw data are composed of cartographic datasets of various natures. b. The matrix produced by the CBA or the DBA methods show the “local geological associations” (i.e. a collection of variables extracted from the cartographic dataset) in the shape of a square grid. c. The predictive model is based on Random Forest (RF) binary classification of mineralized and non-mineralized cells using the matrix. d. The binary classification produces a favorability map based on the selected mineralized and non-mineralized environments, estimating the likelihood of each cell to be mineralized according to the model. e. The validation of the results is performed through statistical and expert means, linking the resulting favorability maps to the parameters and variables of the matrix.

a training set comprising both mineralized and non-mineralized cells to (i) learn and then (ii) produce a predictive model. However, if mineralized cells are easy to identify, non-mineralized cells are trickier to define, as it would require establishing with certainty that a cell contains no mineralization at all. Even if the probability of existing mineralization is unknown in most of those cells, for the MPM approach to have any sense it is assumed to be very low across the map as mineralization are by essence rare. This assumption induces a strong imbalance between the two classes with very few mineralized cells compared to many non-mineralized cells. This major imbalance means that we are facing an “anomaly detection” problem, with a classification containing a minority class (mineralized cells) which we need to outline to produce the favorability map. From the data science point of view, this kind of problem can be processed using either supervised, unsupervised or semi-supervised approaches [Chandola et al., 2009].

This classification problem can initially be qualified as semi-supervised, as the labels used for the classification are partially unknown (i.e. the non-mineralized cells). In fact, we know the existence and location of the mineralization that were discovered but we do not know if more could exist nor their exact location, so only the known mineralized cells are really labelled. The problem we are facing is also transductive a priori: we aim at classifying the unlabeled cells of our dataset to produce a favorability map rather than producing a general classification model for both mineralized and non-mineralized cells (inductive problem) to apply it on other datasets. Many supervised techniques require a large-sized matrix of labelled samples to develop a decent predictive model capable of handling the bias-variance tradeoff [Kuhn and Johnson, 2013] of AI-based algorithms, such as RF. Research into AI-based MPM suggests using equal positive and negative samples to balance supervised algorithms’ input [Porwal et al., 2003a]. Given the intrinsic scarcity of positive sites, it is challenging to develop a decent-sized matrix from available labelled samples, impeding models’ generalization and leading to the over-fitting problem. On the other hand, determining negative samples can also prove to be challenging, as the selection of negatives sites would require either on-site study to ensure the absence of the targeted mineralization. Anyways, negative sites will be more commonly found than positive sites, due to the scarce nature of mineralization, meaning the sample data is bound to be imbalanced in favor of the negative sites. Dealing with this imbalance is necessary to perform the classification, meaning either balancing the input samples before applying the model (over-sampling the positives sites or under-sampling the negative sites) or weighting the data within the model itself. With the CBA or DBA, under-sampling could be applied by selecting only few non-mineralized cells, at random or using expert knowledge. However, in the frame of this study, we choose to rely on an inbuilt weighting tool within the RF model used to deal with the imbalance, meaning that all known positive sites (mineralized cells) and negative sites (non-mineralized cells) are used during the classification process.

Other studies on MPM approaches [Carranza and Laborte, 2015b, 2016, Daviran et al., 2021, Rodriguez-Galiano et al., 2015] relies on a supervised and inductive machine learning method. They try to label both mineralized and non-mineralized targets for the training process to produce a general classification model and then fit the whole study area using this model. In those studies, to solve this labelling problem, only parts of the studied area, outside the known mineralization, are considered as non-mineralized for the predictive model. Non-mineralized samples are selected using specific methods to increase the likelihood of these samples to be non-mineralized, mainly by selecting areas that have the least resemblance to mineralized areas. One of the simplest way to define these non-mineralized occurrences is by using random selection, producing a random number of points within the studied area which are considered as non-mineralized. More advanced methods use spatial and statistical criteria to determine the non-mineralized occurrences, as described in Carranza and Laborte [2015b], where the non-mineralized occurrences are

identified at distances where the multivariate environments will not match the one surrounding the mineralization occurrences (**Figure 4.4**). Ultimately, those locations can also be selected manually based on an expert knowledge by determining which areas are the most unlikely to host mineralization. However, these non-mineralized points are often selected in contexts so unlikely for mineral occurrences that they do not allow a proper discrimination, from a statistical point of view.

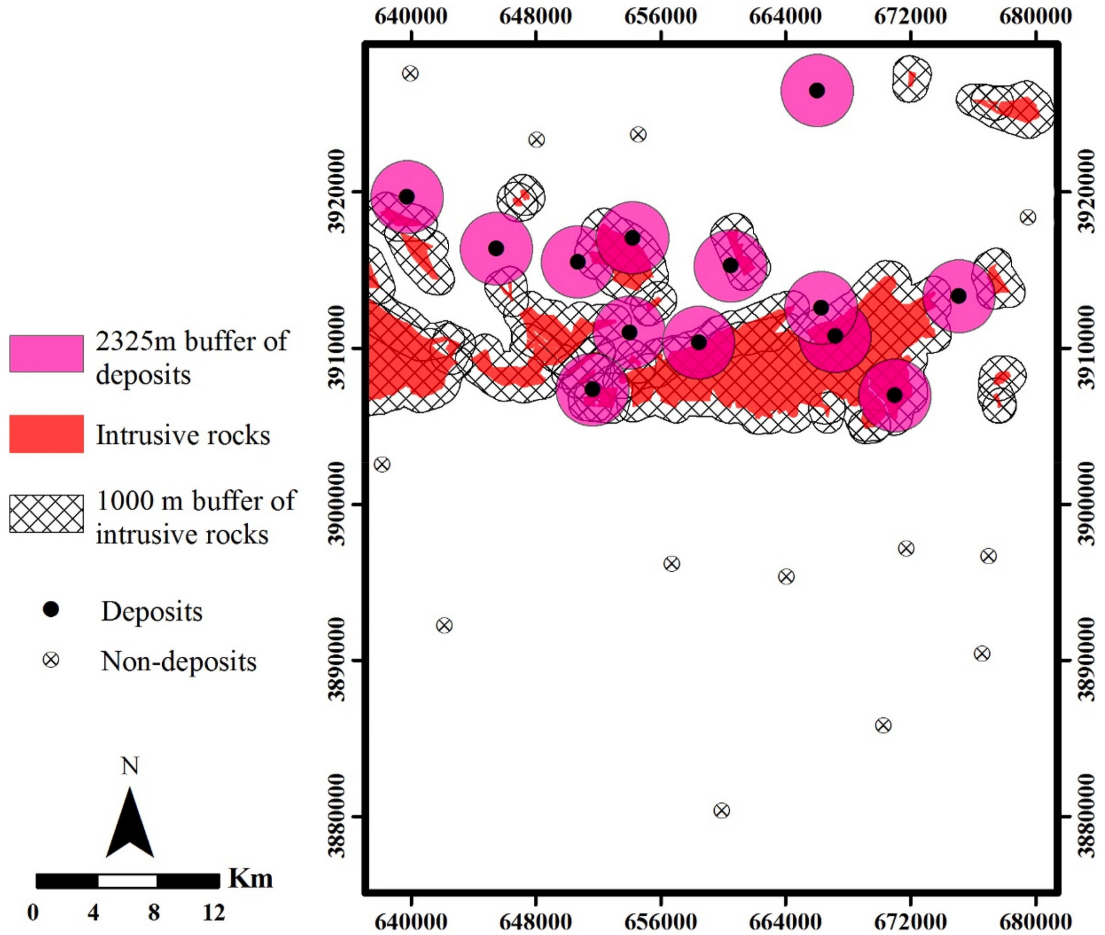


Figure 4.4: Selection of prospect and non-prospect locations in a study from [Daviran et al. \[2021\]](#). Following the rules established by [Carranza and Laborte \[2015a\]](#) the number of non-prospect sites equals the number of prospect sites. Non-prospect sites are distal to prospect sites and follow a random distribution.

With the combination of the CBA / DBA method with the RF classification, the problem is viewed as a fully supervised inductive problem. One can consider in that case that all cells that do not belong to the mineralized class are to be considered as non-mineralized, resulting in most of the studied area being considered as non-mineralized (**Figure 4.5**). This outcome seems logical as MPM is fundamentally based on the hypothesis that mineralization is by nature rare, meaning that most of the study area would end up being “barren” (i.e. devoid of mineralization). It is possible to consider that, outside the known mineralized cells, unknown mineralization do exist but most of the studied area will end up being “non-mineralized”. This hypothesis allows a simple and robust selection of non-mineralized cells for the classification process. This approach is expected to produce False Positives (FP) cells (i.e. non-mineralized cells with high favorability score, possibly undiscovered mineralized cells) as the result of the RF classification, which are the more likely locations of mineralization, while the True Negatives (TN) cells (i.e. non-mineralized cells with low favorability score) are the probable non-mineralized areas.

Selection of non-mineralized occurrences

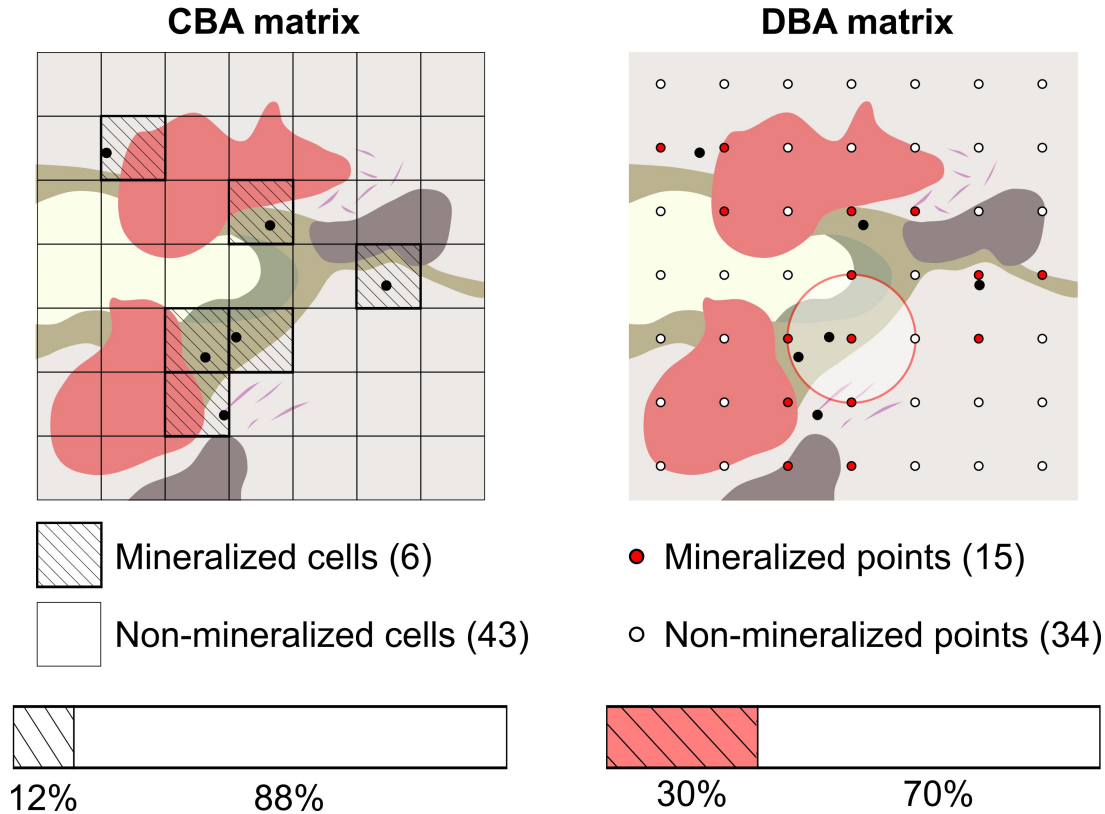


Figure 4.5: Selection of prospect and non-prospect locations for CBA and DBA matrices. This method of selection induces an imbalance between the number of mineralized sites and number of non-mineralized sites. This imbalance will mainly depend on the number of mineralized cells found by the matrix and thus depend on the parameters used as well as the number of known mineralization and their distribution across the studied area.

By definition, there is no a priori fully reliable, unbiased, and automated labelling process, which is precisely the goal of the MPM methodology. The only known labels are the rare mineralization, which is why the CBA / DBA RF approach is conceived to focus on the mineralized cells and consider all other cells as non-mineralized. As such, both the unreliability (ratio of zones classified as "non-mineralized" a priori and wrongly) and the bias (non-representativeness on the whole map of areas classified a priori as "non-mineralised") are upper bounded by the probability of mineralization on the whole considered area. Completely and very easily automatable, this labeling method also provides more stable results than methods based on random selection.

4.2.3 Random Forest predictive modeling

Random Forest (RF) is a supervised machine learning method used for classification and regression as formalized by Breiman [2001]. This method is based on the aggregation of multiple decision trees combined with a randomized selection of samples and variables, known as a predictors, to predict either qualitative or quantitative values associated with known classes. RF approach is here employed in combination with the CBA or DBA matrices to build the predictive model through a data driven approach and with the

less possible expert guided input. RF is used to (i) classify cells (i.e. mineralized/non-mineralized classes) using the multi-variable information provided (e.g. lithologies, structures, occurrences) contained in the matrix produced by the CBA or the DBA and (ii) quantify the similitude between different zones of the study area, more specifically the ones with known deposits.

With the CBA / DBA RF method, the samples used in those trees are the cells produced by the matrices. When combined with the CBA and DBA methods, RF is used to perform a binary classification of the cells of the matrix, between a mineralized and a non-mineralized class. This classification also gives an estimation of the probability of each cell to belong to the mineralized class: the RF method gives for each cell a score between 0 and 1, estimating the likelihood of the cell to belong to the mineralized class. The closest to 1 this score is for a given cell, the higher this probability gets. However, it is noteworthy that this estimated probability cannot strictly be considered as the probability of a cell to be mineralized; it is thus used here as a favorability score. A high favorability score associated with known mineralization is expected for an efficient model, while high favorability areas without known mineralization in their vicinity are likely to hold undiscovered ones and will be preferred targets for exploration.

The matrix is the training dataset of this RF model (**Figure 4.6.a**). The predictors of the RF model are the variables associated with the sample and used by the decision trees to perform the classification. While the predictors can be any pertinent data from the cartographic dataset, in the frame of this work, we use lithological formations and structural features as predictors to be aggregated in the matrix. This RF binary classification is based on the matrix column determining if a cell is mineralized or not, the target of this binary classification being the mineralized and non-mineralized cells determined earlier. The number n of trees produced by the model is defined manually. For each tree, a randomized selection of samples (**Figure 4.6.b**) is done using stratified bootstrap aggregation [Breiman, 1996]: two thirds of the cells are randomly selected to be classified (i.e. training set), while the remaining third is used to validate the results (i.e. testing set). Stratified bootstrapping allows keeping the same proportion of each class (i.e. mineralized vs. non-mineralized) in test and training samples regarding the original dataset. This method helps overcoming the imbalance classes and results in more coherent trees. Each random sampling is done with replacement. In addition, the test and training bootstrap samples are automatically balanced during the RF classification, their weights being inversely proportional to class frequencies in the input data, to compensate the initial imbalance and focus more on the minority "mineralized-cells" class.

Each of the n trees also uses a random selection of i predictors to perform its classification (**Figure 4.6.c**), which can be defined manually even though it is generally defined as the square root of the total number of predictors [Breiman, 2001, Hastie et al., 2009]. The samples used by each tree are also balanced automatically to adjust the weights: each class is weighted inversely proportional to its frequencies in the input data for each bootstrap sample used by each tree. This sampling method is commonly used when dealing with imbalanced classes, such as the one resulting from the previous selection of target cells with only a few mineralized cells.

In a second stage, classification is performed on the training sampled cells and the selected predictors. Each iteration results in one decision tree trained and validated with one bootstrap sample. The decision trees classify the training sample by partitioning the sample multiple times using the predictors. As shown in **Figure 4.7**, this process can be visualized using a tree-shape graphic showing the partitioning of the sample. Those decision trees are composed of nodes and leaves [Genuer and Poggi, 2017]. Each node represents a subset; the final nodes being called leaves while the first one (e.g. the training dataset before

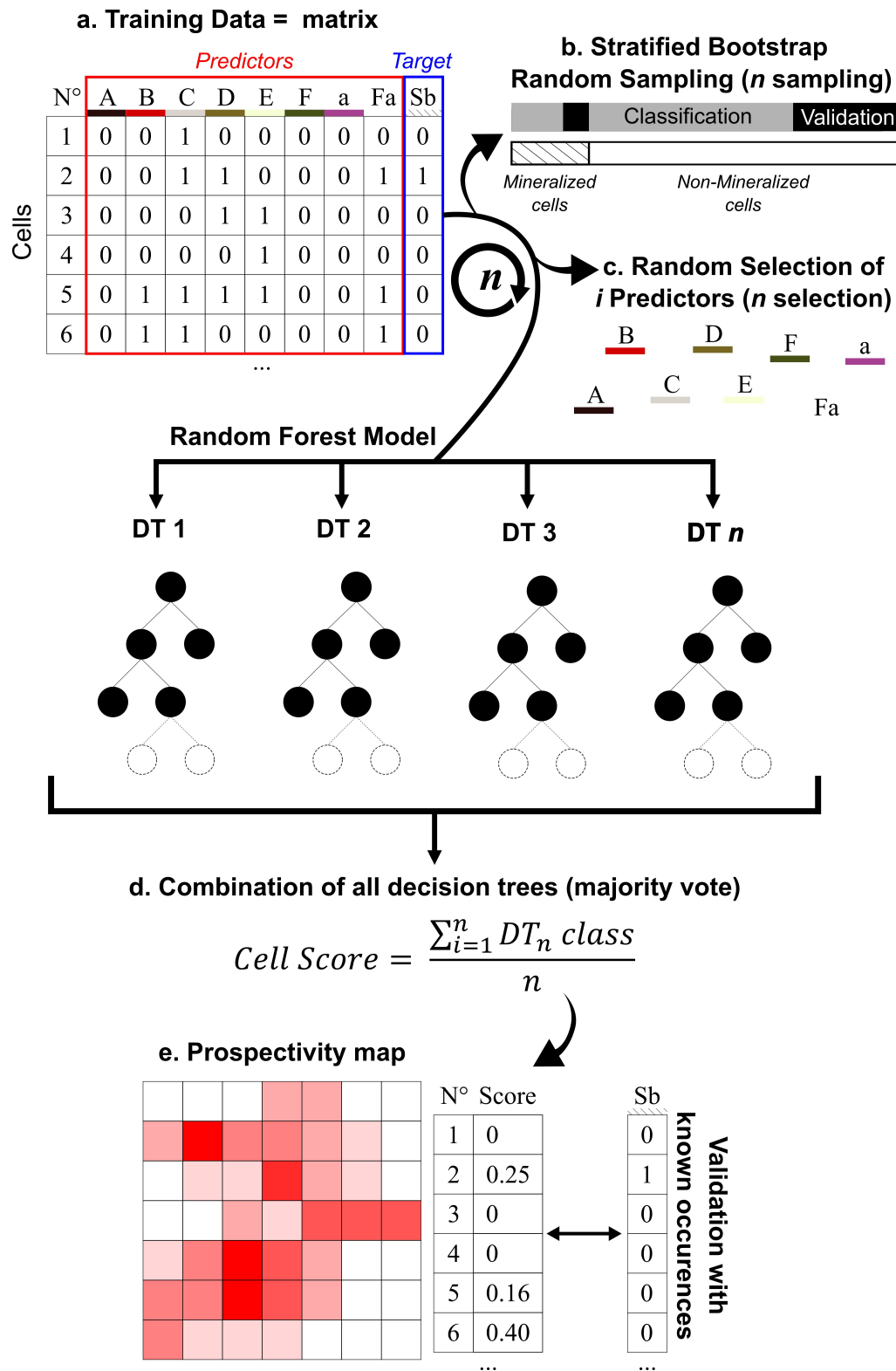


Figure 4.6: Creation of Random Forest prospectivity maps from CBA or DBA matrix (modified from Harris et al. [2015]). a. The training dataset used by the RF predictive model is composed of the cell of the matrix and can be divided between the predictors used by the model and the label of the cells. b. Stratified bootstrap random sampling is applied on the matrix for each tree. c. A random selection of predictors is done for each tree. d. The random forest model is applied on each cell of the matrix, and its results are combined using a majority vote. e. The favorability scores are saved in the matrix and can be evaluated using known occurrences or using check occurrences not used in the model.

the first partition) is called root. The complexity of a decision tree can be measured using its depth: the distance (i.e. number of partition) between the final leaf and the root.

For each partition, the most efficient predictor is selected using impurities measurement, such as the Gini impurity from the CART algorithm [Breiman et al., 1984]. It evaluates the ‘purity’ of a node: a value of 0 corresponds to a pure node (e.g. contains only cells labelled as mineralized), while a value of 0.5 is an impure one (e.g. contains the same number of cells labelled as mineralized and non-mineralized). The Gini impurity measures the disorder of a set element, evaluating the probability of mislabeling the cells if they were randomly labeled, a pure node being the less likely to mislabel the cell while an impure one has the highest chance of mislabeling. At each partition, the Gini impurity is computed for the parent node and for each possible children node to determine which predictor maximizes the Gini gain (i.e. the diminution of impurity value). This predictor is then used to partition the data into subsets. For categorical variables, the partition is done using its classes, while continuous variables are converted into categorical variables using an optimal threshold automatically determined through the algorithm. The produced subsets, or nodes, are then partitioned again using the same process if enough individuals of different classes are still present. This process is iterated until reaching a breaking rule (i.e. reaching a predefined maximum depth, a minimum number of individuals in the subsets, etc.).

The validation process is performed after the training of the decision tree is done. The validation and the evaluation of each tree are based on the remaining individuals, the validation sample, composed of individuals not used during the training process. This sample is classified using the trees produced by the training. The error in the prediction of this testing sample is called the Out Of Bag error ("OOB") [Genuer and Poggi, 2017]. This OOB error is similar to an evaluation of the model using cross-validation and is used to estimate the prediction error of the model. This metric evaluates the effectiveness of each decision tree in the RF model, and its mean value evaluates the effectiveness of the whole RF model.

Once the n trees have been trained on their respective bootstrap samples and evaluated by their test samples, the whole forest composed of the n decision trees is used to predict the classes of the cells ("0" or "1") of the CBA or DBA matrices. Each cell of the grid is classified by each tree in the forest. The probability of a cell to belong to the mineralized class, also defined as its favorability score, is determined through a majority vote of all the decision trees inside the RF model (**Figure 4.6.d**), the final score of the cell being the mean value of this vote:

$$\text{Cell Score} = \frac{\sum_{i=1}^n \text{Tree class } i}{\text{Number of tree } n}$$

This favorability score is then added to a new column in the CBA / DBA matrix, allowing to display the favorability map using the square grid (**Figure 4.6.e**). Cells with the highest score show more likelihood with mineralized environments than the ones with lower scores. Mineralized areas can display lower values depending on the effectiveness of the model: an efficient predictive model is expected to assign higher values to the mineralized cells of the matrix.

4.2.4 Cross validation and hyper parameters tuning

The results obtained from the RF model will vary depending on the tunable algorithm parameters used to train it, such as (i) the number of trees n , (ii) their depths d or (iii) the size of the randomly selected predictors set. Those tunable algorithm parameters are called hyper-parameters and define the training

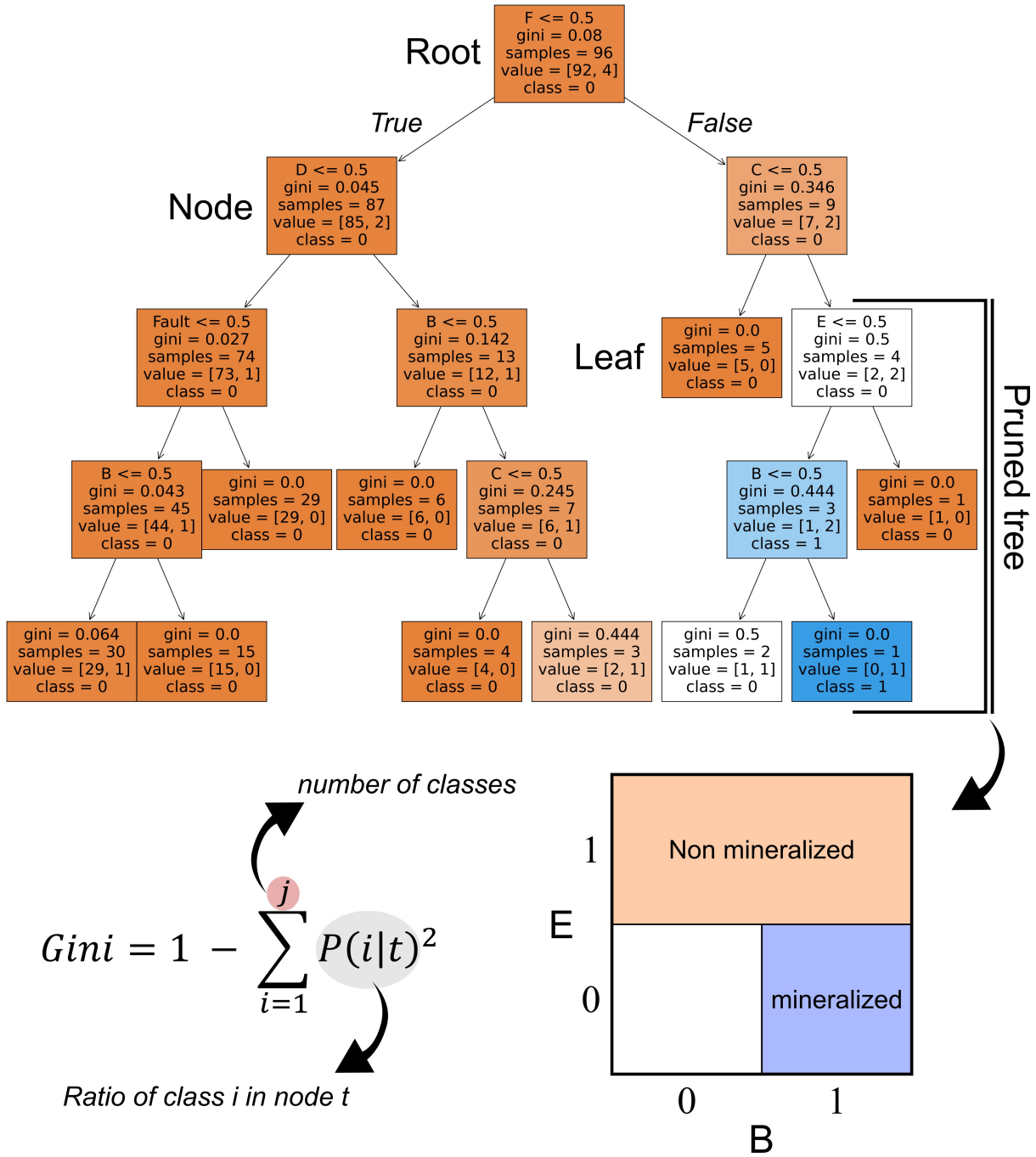


Figure 4.7: Decision tree produced from a stratified bootstrap sampling of the sketch example, with a CBA grid of 5 000 x 5 000 m. Of the 144 cells, 96 are selected, with 4 of the 6 mineralized cells. For each node, the following is specified: the rule used to partition the node, the Gini score of the node, the number of cells, the classes of the cells (['non-mineralized', 'mineralized']) and the class attributed to the sample. Nodes in orange are classified as non-mineralized, in blue as mineralized and in white the classification cannot discriminate. Results can be shown using decision boundary, showing the classification results regarding pair of variables, as shown here using a pruned portion of the tree.

process of the RF model, changing the results and overall effectiveness of the model. In order for the model to perform the most efficiently, it is important to “tune” those hyper-parameters and find an optimal configuration. In order to find the optimal hyper-parameters, machine learning method such as the RF can rely on various hyper-parameters tuning methods. These methods will try different sets of hyper-parameters in order to find one able to improve the results.

Two main approaches of hyperparameters tuning exist: the Grid Search and the Random Search (Figure 4.8). With the former, a list of values is given for each hyperparameter. The predictive model is then tested using each possible association of hyperparameters values from this list, pairing each value with the other ones. Once all possible associations have been tested, the one showing the best results is kept, the best model being the one with the best OOB score (i.e. the lowest). If this method allows a strong improvement of the effectiveness of the model, it can also be quite computing-intensive for a marginal enhancement of the results. The Random Search approach is less intensive in computing power: the operator gives a number of tries i for the algorithm to find the optimal hyperparameters. For each try, a random value is set for each parameter using, for example, a given range and a distribution rule (e.g. normal distribution) and the predictive model is tested using those randomly selected values. At the end, the set of hyperparameters giving the best results is kept. Both methods use Cross Validation to evaluate the hyperparameters.

Hyperparameters tuning with cross-validation

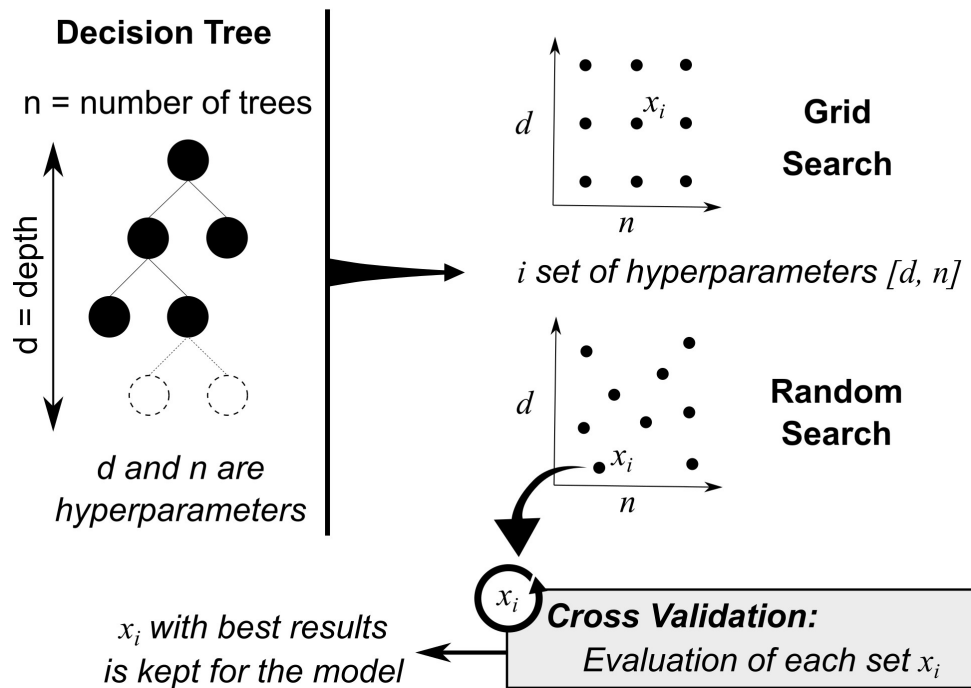


Figure 4.8: Example of optimization of the Predictive model using hyper-parameters tuning on 2 hyperparameters: the number of trees n and the depth of the trees d . Grid Search Cross Validation and Random Search Cross Validation method evaluate multiple sets of random hyper-parameters values x_i using cross validation and select the optimal one. For grid search, a set of value is given for each hyperparameter and all i possible associations of these parameters are tested. For Random Search, the number of set i is defined by the operator and the values of the hyperparamters are computed randomly for each set using a random distribution.

Cross-validation is a resampling method allowing the creation of multiple training and testing samples from a single dataset. The number of samples produced is determined by a number k of folds: a 3 folds

cross-validation will produce three sets of training and testing data, each one different from the others. The Stratified Cross Validation (SCV) is a variant of this method that is used in this study. It allows each sample to be balanced and to contain the same proportion of mineralized and non-mineralized cells (**Figure 4.9**). Folds are made of random balanced samples, meaning that the relative proportion of mineralized and non-mineralized cells randomly selected for the samples is equivalent to the one in the original data. This SCV method is commonly used for machine learning processes relying on imbalanced data. With cross-validation, for each fold the model is trained using the training sample and then evaluated using the associated testing set. The model can be evaluated using different metrics, in this case the OOB is used to define the most effective model. Increasing the number of folds will increase the number of sets of training and testing dataset on which the model is trained and evaluated. While increasing the computation time, increasing the number of folds also means that the model is trained on a larger training set and tested on a smaller test fold, which should lead to a lower prediction error as the model see more of the data. On the other hand, lowering the number of folds k result on the model being trained on a smaller training set and tested on a larger test fold, which should result in a higher prediction error on average. In the previous example, with a 3-folds cross-validation, the model is trained on 2 third of the data and evaluated on the remaining 1 third. In the application described in **Chapter 6** and **Chapter 7**, we rely on a 5-folds cross-validation to evaluate the model with the Random search method. This 5-folds cross-validation allows a compromise between performance (computation time) and the prediction error.

Cross-validation is commonly used to avoid overfitting of the model when dealing with small dataset [Lever et al., 2016]. A predictive model can be evaluated using its training dataset to compute its estimation error: the estimation of the predictive model is compared to the known label of the individuals of the training set. It can also be evaluated using test samples made of individuals that have not been used by the model to train itself, computing the prediction error, the performance of the model when it comes to predict the classes of new data. Both prediction error and estimation error are influenced by bias, the error relative to the incapacity of the model to grasp the natural model we try to simulate, and variance, the error due to sensitivity to noises in the dataset. The more complex a model is, the lower the bias is and the higher the variance is.

Overfitting occurs when the complexity of the model is too high for the given training dataset. The predictive model will have a low bias but a very high variance: the estimation error is low, meaning that the model will give excellent results with the training set, but the prediction error will be higher, as the model is not able to correctly classify new data. Correctly fitting the model means finding a balance between bias and variance, between the estimation error and the prediction error: the predictive model should give good results on both training and testing sets to avoid over- or underfitting. Cross-validation helps to avoid overfitting by evaluating the dataset over multiple training and testing samples. For example, with the tuning parameters method, the best model will be the one with the lowest averaged OOB across the k -folds.

4.2.5 Evaluation of the predictive model efficiency

The performance of the predictive model can be analysed using the resulting predictive classes. As each cell is attributed a favorability score, they can be classified during a post-processing stage between mineralized and non-mineralized using a threshold value based on their associated favorability score. As such, classified cells can be considered as True Positive (TP), True Negative (TN), False Negative (FN)

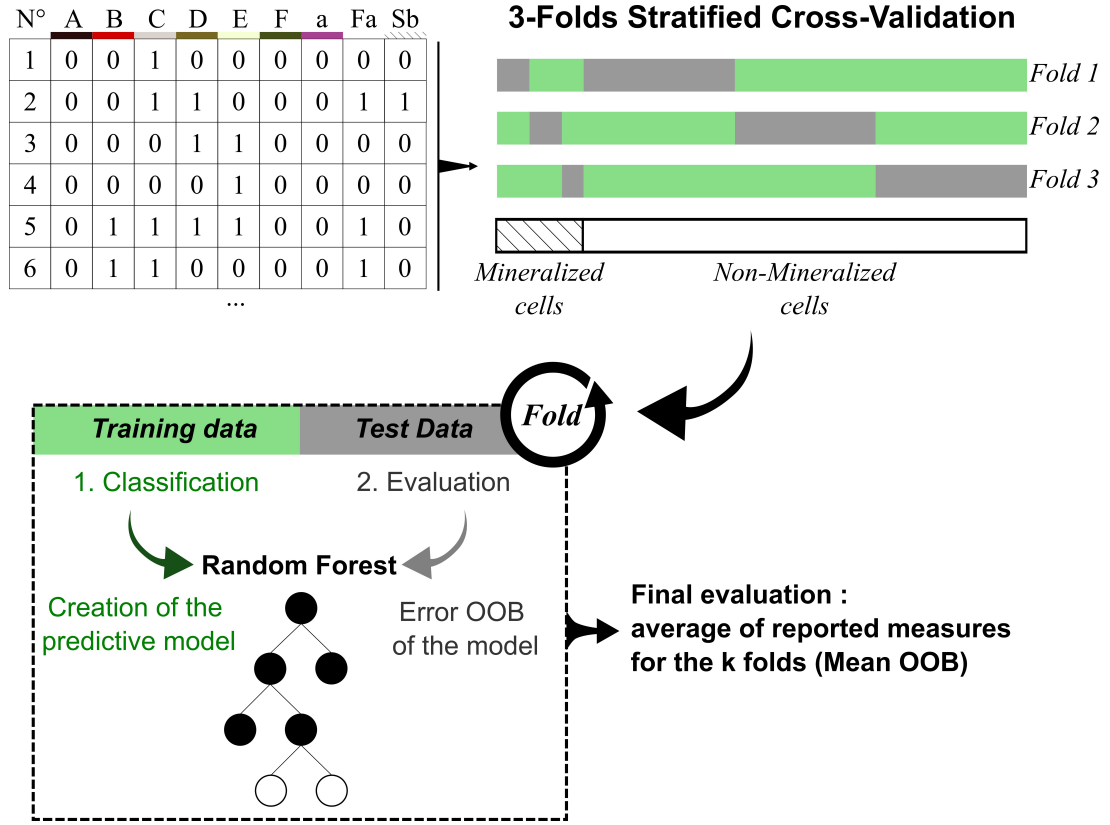


Figure 4.9: Stratified cross-validation method to evaluate the Random Forest predictive model results. This method evaluates the model using balanced samples of training and testing datasets extracted from the CBA / DBA matrix. In this example, a 3-folds Stratified Cross Validation is applied: both the non-mineralized and mineralized rows are split in three sub-samples. For each fold, the training sample is used to classify the cells, while the testing sample evaluates its results using the specified metric or measure (in this case, the Out-Of Bag error or OOB). The final evaluation is the average of the reported measure from all folds.

and False Positive (FP) depending on their initial value. The first two types correspond to cells which predictive classes match their initial classes: TP are mineralized cells classified as mineralized (i.e. high favorability score) by the RF model and TN are non-mineralized cells classified as non-mineralized (i.e. low favorability score). FP and FN are “misclassification”, the former being non-mineralized cells classified as mineralized and the latter mineralized cells classified as non-mineralized. All those predictions depend on the score threshold retained to classify the cells: above this threshold, a cell is considered as mineralized (class “1”). As such, a high threshold results in less predicted mineralized cells, potentially increasing TN and FN, while decreasing it will results in more predicted mineralized cells and increase TP and FP.

The Receiver Operating Characteristic (ROC) curve allows the evaluation of the predictive model using those predicted classes, through the True Positive Ratio (TPR) and the False Positive Ratio (FPR). Those ratios are computed for a continuous range of threshold values and projected in a graph (**Figure 4.10**).

$$TPR = \frac{TruePositive}{TruePositive + FalseNegative}$$

$$FPR = \frac{FalsePositive}{FalsePositive + TrueNegative}$$

4.2. RANDOM FOREST-BASED PROSPECTIVITY MAPPING APPLIED ON CBA AND DBA MATRICES

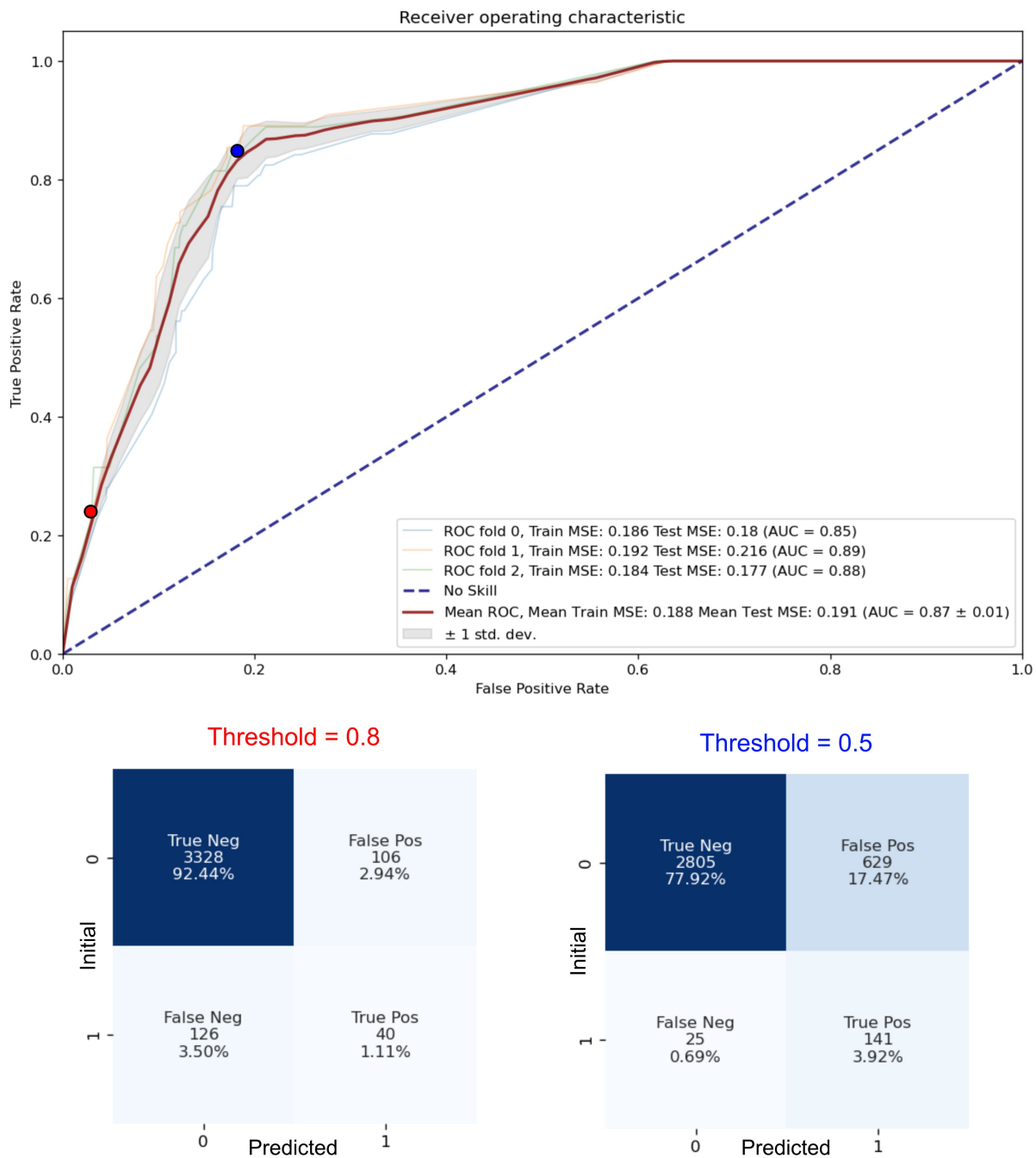


Figure 4.10: Example of Receiver Operating Characteristic (ROC) curve produced with a 3-folds Stratified Cross Validation for the Random Forest predictive model. Each fold is displayed with its corresponding Area Under the Curve (AUC) score. The no-skill model displayed in dotted blue corresponds to a theoretical worst-case predictive model. The mean squared prediction error for the training and testing datasets are also displayed. Confusion matrix for a threshold value of 0.5 and 0.8 shows the number of TP, FP, TN and FN cells given by the model.

Comparing the TPR and FPR values allows a quick evaluation of the efficiency of the model to classify the mineralized cells: an efficient model should display high values of TPR for low values of FPR. This relation shows that the model is able to correctly classify the mineralized cells, with a high frequency of TP compared to FN, while minimizing the misclassification on non-mineralized cells with a low frequency of FP compared to TN. An easy way to evaluate the effectiveness of the model using this ROC curve is by computing the Area Under the Curve (AUC). The closer to 1 the value is, the better the model is at correctly predicting the known mineralized cells. This ROC curve can be produced when performing Cross Validation to display the various models created from the different folds to assess its variability depending on the training dataset used (**Figure 4.10**). The Mean Squared Error (MSE) of the prediction errors can be computed with the cross-validation, for each training and testing dataset. A high discrepancy between the MSE of training and testing datasets belonging to the same fold can indicate possible overfitting or underfitting (the model tends to perform better with one of the two datasets).

4.2.6 Evaluating the predictors of the model

RF classification relies on the use of predictors to produce the favorability map. The predictors used to produce the favorability map can be evaluated to determine which ones are the most important for the classification (i.e. the predictors that are the most efficient to classify the cells, to discriminate the two classes). This can be evaluated using the metric used to create the decision tree: the Gini impurity. The Mean Decrease in Impurity (MDI) is computed from the Decrease in Impurity of each predictor across the whole RF [Breiman, 2003]. The MDI is based on the total decrease in node impurity (Gini impurity) from splitting the samples using a given variable in a tree, averaged over all the trees using this predictor. The Decrease in Impurity I is computed for each node:

$$I = G_{Parent} - \sum_{i=1}^n P_i G_i - P_n G_n$$

G_{Parent} : Parent node; n : number of split / children nodes created; P_n : Proportion of cells in children node n ; G_n : Gini index of node n .

This Decrease in Impurity of a predictor in a tree is computed by using each node performing a split based on this predictor. The MDI score is the averaged Decrease in Impurity I computed for each tree in the RF using this predictor. The higher the MDI score is, the better the predictor is at decreasing the impurity of the nodes, at classifying the samples. Predictors with the highest MDI score are the most efficient in the predictive model.

However, MDI is known to be biased toward continuous and high cardinality features [Strobl et al., 2007, 2008]. These types of variables will have more splits than other variables, resulting in a higher probability of a tree to contain multiple splits using them and therefore increasing their MDI score. As the CBA and DBA matrix shown here are based on binary dataset, such problem does not occur, but the addition of continuous data could introduce such bias in the evaluation of the MDI in the approach we develop here. Other metrics allow overcoming this issue, such as the Permutation Importance, based on the Decrease in Accuracy when permuting the predictors in the decision trees [Breiman, 2001].

Chapter 5

Exploratory Data Analyses on CBA and DBA matrices

*The final step of the developed inverse methodology is detailed in this **Chapter 5**. The identification of the local geological environments, described in **Chapter 3**, and their evaluation regarding their relationships with targeted mineralization, displayed in **Chapter 4**, are necessary steps before the identification and discretization of the most favorable ones into few "key" mineralized geological environments. This final step can be considered as an Exploratory Data Analysis (EDA) process applied on the results of the previous steps, as we want to summarize their main characteristics to interpret their results in a geological and metallogenic point of view. With the use of an unsupervised classification method, called Hierarchical Clustering on Principal Components (HCPC), we analyze the results produced by the DBA and the RF. Beforehand, the geological environments are evaluated using their estimated relationships with the targeted mineralization, quantified by their favorability score. The selection process of the most favorable environments, relying on a statistical analysis of the RF predictive model, is described in this **Chapter 5**, as well as the HCPC classification process, which includes a factorial analysis (Multiple Correspondence Analysis, MCA) and an unsupervised classification method (Agglomerative Hierarchical Clustering, AHC). Following **Chapter 6** and **Chapter 7** display proof of concept and applications of the methodology developed and presented over the **Chapter 3**, **Chapter 4** and **Chapter 5**.*

5.1 Clustering analyses of DBA and CBA matrices

Supervised machine learning methods, such as RF, rely on labeled dataset to compute a predictive model to classify the data. These methods can be applied while the outcomes of the dataset, the targeted classes of the predictive model, are known. In the case of the CBA / DBA RF, each cell is labeled as mineralized or non-mineralized before the classification and the model is designed to have similar outcomes. However, if the target is not the identification of mineralized areas but rather the identification of "geological environments" or "geological signatures" across the study area, then the data contained inside the matrix can be considered as unlabeled. In this case, the supervised methods can not be applied and we rely on unsupervised clustering methods, which are machine learning methods allowing the attribution of labels or classes to unlabeled datasets. These methods draw inferences from the unlabeled dataset, trying to

learn the inherent structure of the dataset, the relationships between the individuals and the variables, to produce clusters of individuals presenting similar responses, behaviors, or sharing the most similarities between them. As the CBA and the DBA matrices result in a large quantity of associations, depending on the initial number of lithologies across the study area (2^n with n number of lithologies), their clustering allows to identify a reduced number of geological environments sharing similarities between them. Those more "global" geological environment are easier to analyze and interpret than the numerous associations initially found.

5.1.1 Agglomerative Hierarchical Clustering

Exploratory Data Analysis (EDA) is a term referring to all descriptive methods that allows to describe and visualize multivariate data set [Husson et al., 2010]. EDA methods allow to observe relationships, patterns, resemblances or differences within individuals or variables in a multidimensional space. An example of EDA is Principal Component Analysis (PCA) methods. Those methods allow to reduce the dimensionality of multivariate datasets to facilitate their analysis and visualization. Unsupervised clustering, such as hierarchical clustering, are another example of EDA method where multidimensional data are grouped based on their resemblances or differences. Ascendant Hierarchical Clustering (AHC) is one of those EDA method. It is a clustering method used here to group the various lithological associations produced by either the CBA or the DBA matrix. As an unsupervised learning method for clustering, it allows (i) identifying group of individuals or clusters, sharing higher similarities among them than with the other individuals and (ii) grouping those clusters by successive aggregations [Rolet and Seguin, 1986a,b]. Initially, n individuals exist corresponding to the number of cells produced by CBA / DBA. It is worth noting that all the cells sharing the same binary code (i.e. exactly the same lithological associations) are already clustered together as they are strictly identical. These individuals are grouped using a distance matrix, calculated for all the individual pairs based on a dissimilarity index. This index evaluates the likeness between two observations and is similar to a distance. The higher this distance is between two individuals, the more dissimilar they are.

Many kinds of indexes can be used depending on the type of variable studied (euclidean, manhattan, Gower, ϕ^2 ...). With the CBA, using the euclidean distance has shown to be efficient in previous studies [Vic and Tourlière, 2016, Gloaguen and Tourlière, 2016a, Tourlière et al., 2015]. This euclidean distance to compute the dissimilarity is based on the following formula, where x and y are 2 distinct lithological associations and n the number of variables inside the lithological association [Rham, 1980]. This euclidean distance is computed between each lithological association. The dissimilarities or distances between each individual can be displayed as a matrix (**Figure 5.1**).

$$d_{(x,y)} = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

Once the distance matrix is produced using the dissimilarity index, the aggregation function determines on which criteria the lithological associations will be grouped. As for the dissimilarity index, several methods exist based on the combination of the matrix index produced with various statistical or weight factors. The aggregation function used here is the Ward's method [Murtagh et al., 2011, Ward, 1963], grouping the clusters with the nearest gravity centers. It is based on the maximization of inter-cluster inertia and the minimization of intra-cluster inertia (**Figure 5.2**) and follows this equation:

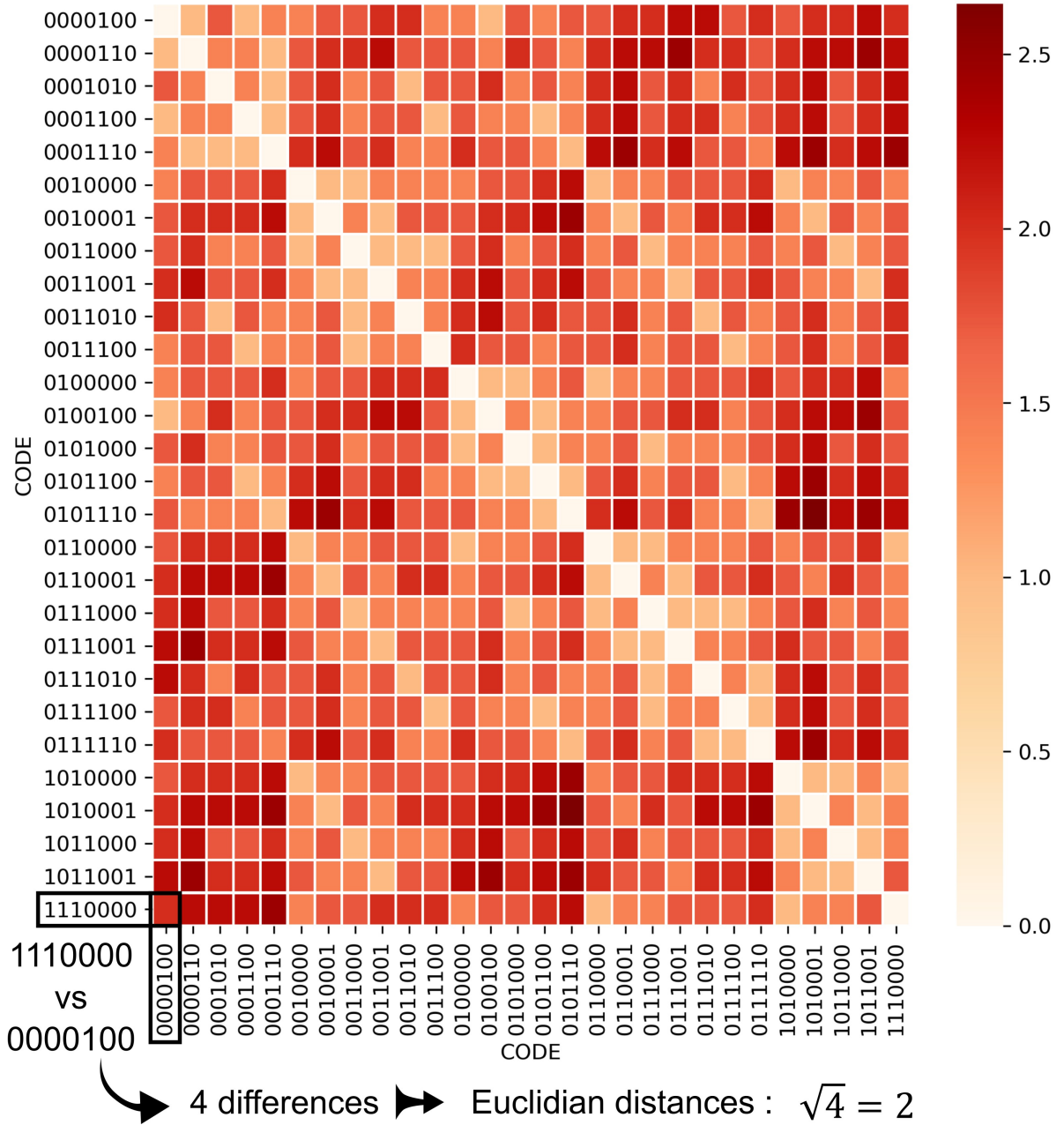


Figure 5.1: Euclidean distance matrix applied on the CBA matrix with 5 000 m cells. The 144 cells results in 28 unique lithological spectra, the euclidean distance between each of them is computed using the binary code, which is equivalent to the square root of the number of differences between each spectrum.

$$\Delta_{ward(x,y)} = \frac{m_x \times m_y}{m_x + m_y} d^2_{(\mu_x, \mu_y)}$$

x and y are 2 clusters; $d^2_{(\mu_x, \mu_y)}$: squared euclidean distance between cluster x and y ; m_x and m_y : x and y cluster's population.

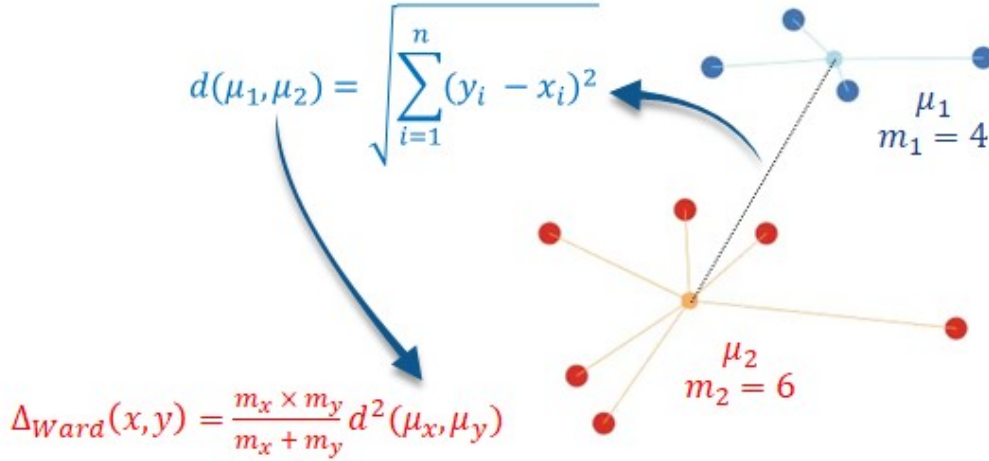


Figure 5.2: 2D Sketch example of the calculation of Ward's distance for 2 clusters (μ_1 and μ_2) with different population (m_1 and m_2).

This aggregation function results in the grouping of two individuals, two clusters or a cluster and an individual, which will form a new cluster. After each grouping, a new distance matrix is produced, using the same dissimilarity index and aggregation function. This process is iterated from the n individuals to only two clusters integrating all the individuals. For the first iteration, this aggregation function does not affect the distance between the lithological associations, so the nearest one will be grouped together. However, after the first iteration, the distance between clusters and individual lithological associations or other clusters will be computed based on the gravity center of the cluster, following Ward's method.

The results of the AHC can be displayed as a dendrogram: individuals are shown along the X-axis and a tree-shaped graph is drawn, showing the grouping of each individual and each cluster as the iteration goes on (**Figure 5.3**). The Y-axis shows the distance between the clusters and its value depends on the dissimilarity index and the aggregation function used. This dendrogram is often used to determine the cut-off level: the number of clusters that seems optimal for the current model. This cut-off level can be determined manually or by using specific function or criteria depending on the objective of the study.

Agglomerative Hierarchical Clustering (AHC) on CBA matrix has been already applied to produce prospective maps [Gloaguen and Tourlière, 2016a,b, Vic and Tourlière, 2016, Gloaguen et al., 2018]. As it is applied on a CBA or DBA matrix, the results of the AHC will vary depending on their properties. In the case of the CBA, the number of cells will inversely increase with the size of the grid cells, increasing at the same time the computation time necessary to perform the AHC. The number of cells with unique lithology increases at the same time, decreasing the quality of the classification: the AHC will be forced more often to gather cells with no common point to realize its grouping, the classes obtained from these cells with unique lithology will see their quality strongly decrease. Conversely, increasing the grid cell size leads to a decrease in the number of cells, largely reducing the calculation times and decreasing the number of unique lithologies in the cells, while the average number of lithologies in the cells will

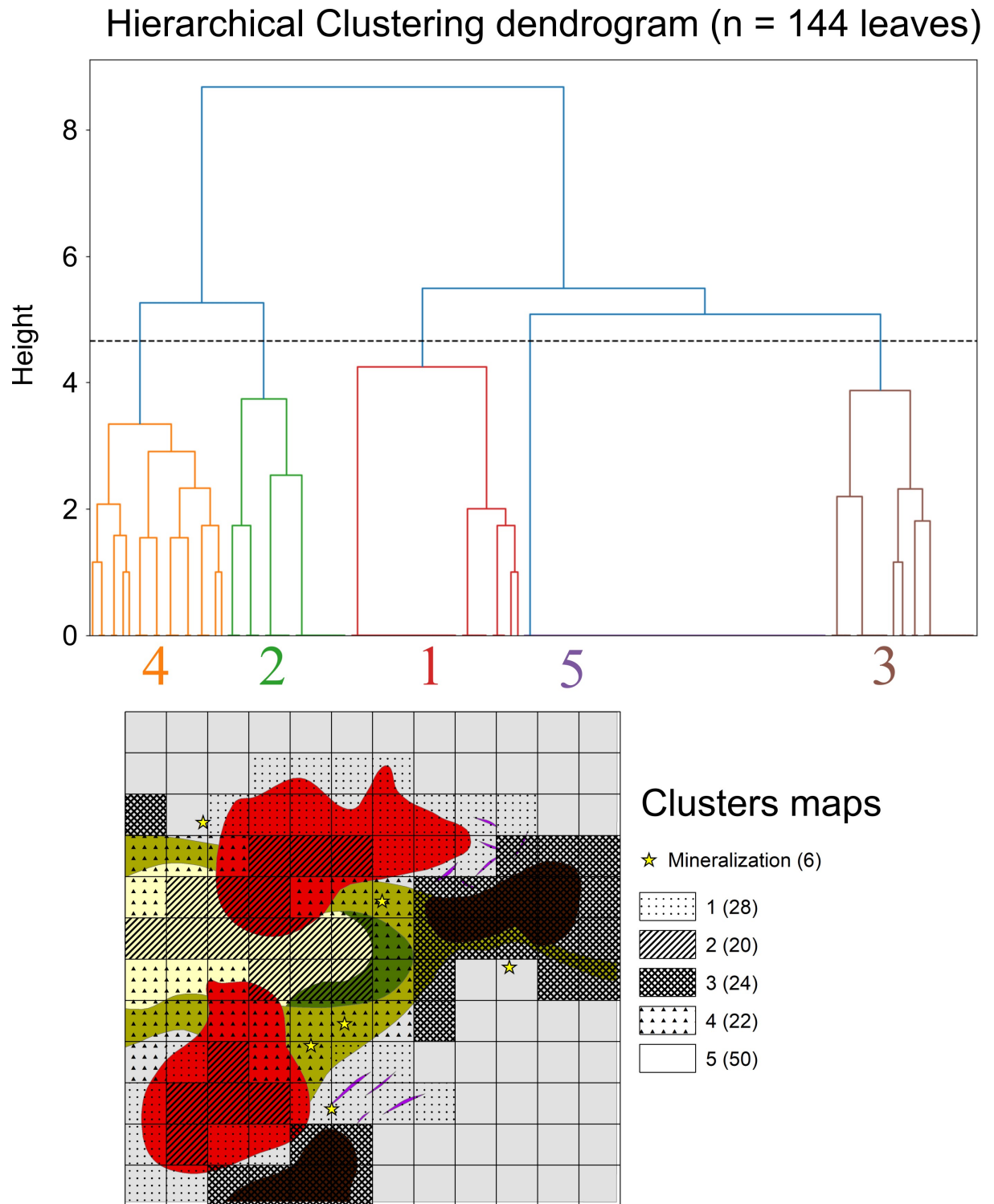


Figure 5.3: Hierarchical clustering dendrogram and resulting cluster maps from the 5 000 m CBA matrix. The AHC is performed on the lithological association (7 lithologies) with 28 unique lithological associations for 144 cells. Height corresponds to the Ward's distances between the nodes of the dendrogram. The dotted line on the dendrogram is the 5 clusters cut.

increase. At very low resolutions, with a low number of cells, those associations become less relevant as they group a wide variety of lithologies, adding “noises” in the classification and leading to a degradation of the quality of the classification. With so few individuals presenting increasingly strong similarities between them with the integration of common lithologies, the obtained classes do not have any more real significance, or even do not have any interest when the number of individuals is extremely low. For the DBA the same balance must be found between the number of associations and the number of cells with one lithology, but with less constraint regarding the resolution.

However, the use of euclidean distances with the AHC is poorly adapted when applied to categorical variable, which is the type of variable we are the much dealing with the lithological associations coming from the geological map information. The data produced by the CBA and DBA matrix are binary codes whose length corresponds to the number of lithologies present in the study area. These data are defined by categorical variables, with only two categories: presence ("1") or absence ("0") of a given lithology. Using AHC with only two categorical variables can be very restraining. With only two different values used to calculate the distance ("0" and "1"), the number of lithology will have a big impact on the resulting distance matrix: the distance between each association will be calculated using n numbers (n being the number of lithologies) of "0" or "1" values. In details, for most of the distance used (e.g. euclidean, manhattan ...), it will result in a limited number of calculated distances: all the associations sharing the same number of differences between each other will also share the same distances. As such, by using euclidean distances to compute the dissimilarity between the lithological associations, we only take into account the number of differences between each association to perform the clustering, the distances found will mainly depend on the number of lithologies in the lithological spectrum. This also restrains the results obtained with most of the aggregation function. Most of them are based on the distance between the individuals (in our case, the lithological associations). With only few different distances shared between multiple associations, this will result in very rough classes.

5.1.2 Hierarchical Clustering on Principal Components

The Hierarchical Clustering on Principal Components (HCPC) [Husson et al., 2010] is an alternative EDA method which performs an analysis of the pattern of relationships between several categorical variables (e.g. the lithologies inside the lithological spectrum) before performing the clustering method itself. The use of the HCPC with the CBA and DBA matrix aims to improve the classification and get rid of the main limitation of the AHC when applied on the matrix: only few distances are computed in the distance matrix, due to the discrete nature of the variables, resulting in rough classes as the distances between the associations are often similar. Rather than producing clusters based on the binary data, as the AHC, the HCPC consists in performing clustering on the factor axes resulting from a factorial analysis, where the euclidean distances have more meaning. The classification can also be reinforced with a partitioning ("k-means" or moving average). Factorial analysis allows the integration of large amounts of data by identifying relationships between individuals (i.e. cells) and variables (i.e. lithologies). The selection of the most relevant relationships allows reducing the dimensions of the studied data. Several types of factor analysis exist depending on the type of data considered (Principal Component Analysis, Multiple Correspondence Analysis, Multiple Factor Analysis ...) [Greenacre and Blasius, 2006, Jolliffe, 2002, Le Dien and Pagès, 2003]. The factorial analysis combined to the AHC method developed here is the Multiple Correspondence Analysis (MCA), a method similar to the PCA that allow the analysis of qualitative multivariate data and can be applied when studying data described by more than

two qualitative variables [Greenacre and Blasius, 2006, Chavent, 2015]. By analysing the pattern of relationships of several categorical variables, it allows transforming qualitative data into quantitative data and can be combined with an automatic classification method to improve the results [Husson et al., 2010].

With the MCA, the variables modalities need to be coded into a matrix, called "disjunctive table" or "Burt table" depending on its shape (Figure 5.4). Each number x_{ij} of the matrix is translated into modalities k_{is} . Each modality represents one of the possible outcomes of each categorical variable, the number of modalities associated with one variable is the number of categories associated with this value. The lithologies inside our matrix have two value: "1" or "0" and thus two modalities, each corresponding to a binary number. For example, if the variable A of the cell i is "1" (i.e. presence of the lithology), it will be coded into two modalities: $A1 = "1"$ and $A0 = "0"$. The first binary is the modality "presence of the lithology", the second one is the modality "absence of the lithology". Decomposing the variables into categories allows comparing modalities between each variable. It is worth reminding that if in this case, each variable can only belong to one of the two classes determined ("absence" or "presence"), but the MCA method allows processing data with various numbers of classes for the different variables.

The row-profile and column-profile tables are computed from the complete disjunctive table to perform the MCA. Point clouds of the individuals and modalities can be plotted from those tables (Figure 5.4). The distance between two individuals or between two modalities inside those cloud points can be measured with the ϕ^2 distance (« Phi-square ») [Briatte and Soetewey, 2021], a variant of the χ^2 distance (« Chi-square ») [Yadolah, 2008]. This ϕ^2 distance is used in the MCA to compare individuals or to compare modalities of a variable.

ϕ^2 distance between two individuals: two individuals are near when they share the same modalities, even more if this modality is rare as this metric gives more weight to modality with low n_s .

$$d_{(i,i')}^2 = \frac{n}{p} \sum_{s=1}^m \frac{1}{n_s} (k_{is} - k_{i's})^2$$

ϕ^2 distance between two modalities: two modalities are near if they are shared by the same individuals.

$$d_{(s,s')}^2 = n \sum_{i=1}^n \left(\frac{x_{ik}}{n_s} - \frac{x_{i'k}}{n_{s'}} \right)^2$$

The MCA creates a new dimensional space to represent the individuals and variables. This new space is defined by new axis, called "factorial axis", aiming at better representing the variability of the individuals. This variability is defined as the inertia I , the square distance between the individuals or the modalities and the origin 0 of the corresponding point cloud, as defined by the row- and column-profile. The total inertia is the sum of the inertia of all individuals or modalities, which is equivalent to the average number of modalities per variable (i.e. lithologies). Rare modalities and variables containing a large number of modalities will affect more the inertia and could introduce a bias.

$$I = \sum_{i=1}^n \frac{1}{n} d^2(i, O) = \frac{m}{p} - 1$$

The factorial axes are produced sequentially and orthogonally, the first one resulting in the maximum

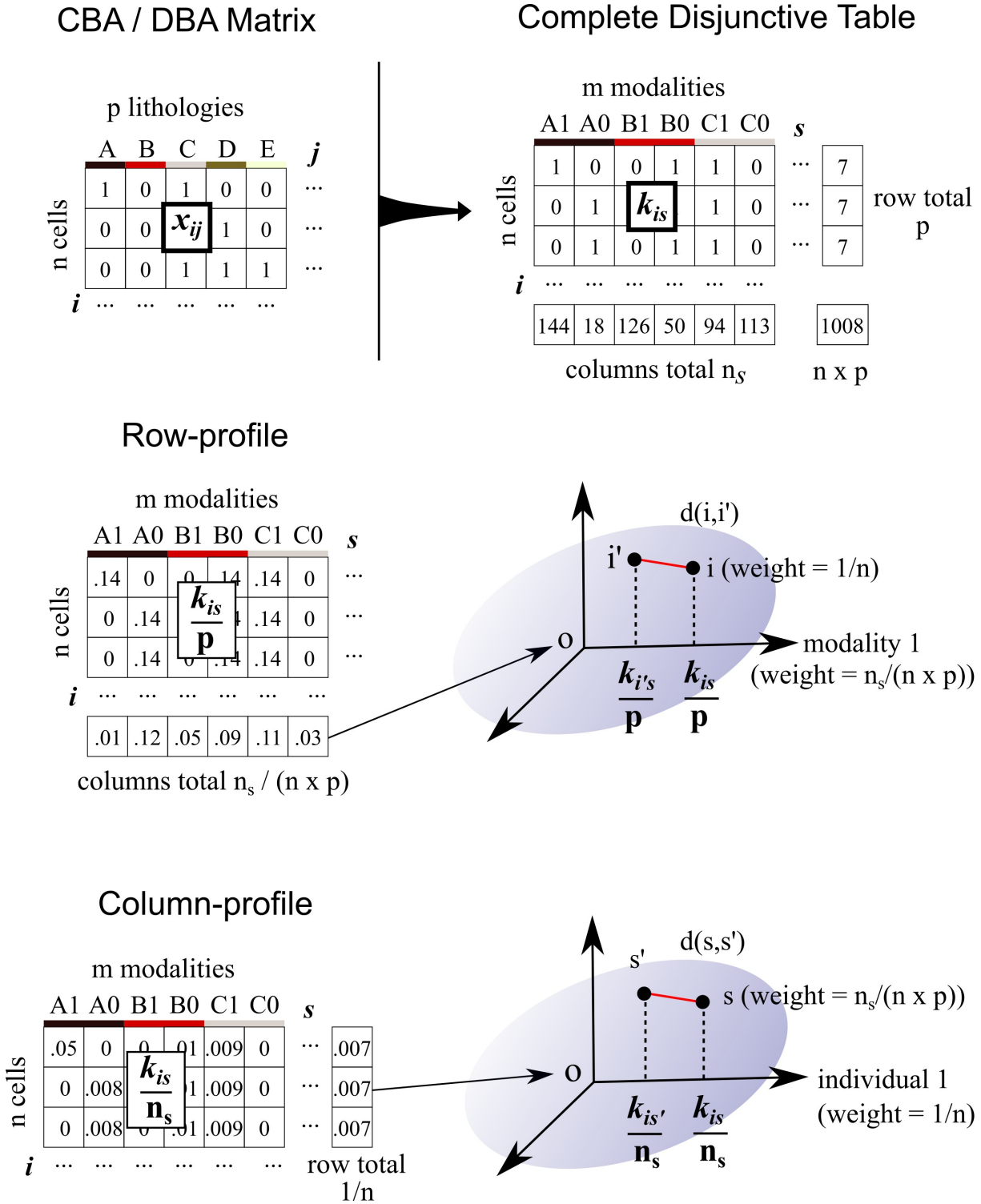


Figure 5.4: Creation of the complete disjunctive table from a CBA / DBA matrix. The disjunctive table is used to produce the row- and columns-profile representing the weighted projection of the individuals or modalities in a centered multidimensional space.

inertia while consecutive one will have decreasing inertia. The individuals and modalities are associated to new factorial coordinates based on their projections on those new factorial axes (**Figure 5.5**). Each factorial axis can be evaluated based on the percentage of total inertia it represents. The first axis will have the highest value while consecutive axis will decrease. The results for the best selected factorial axis can be shown in 2D plots to see the patterns between individuals or modalities. The AHC clustering is then performed on the new factorial coordinates for the individuals (i.e. the cells of the matrix) (**Figure 5.6**) to produce the clusters and group the cells depending on their theological associations. Instead of binary values, the Ward's distance is calculated with continuous values given by the MCA, the new coordinates giving more insight about the relationships between each cell and thus their respective lithological association. The results of the classification have more strength compared to the application of the AHC on the binary values, where only the number of differences between each cell is taken into account.

5.2 Favorable geological environments identification

The application of clustering analyses on the CBA or DBA matrix allows synthetizing all the local lithological associations into fewer geological signatures. This application can also be used in addition to the MPM methods to outline the geological environment in areas deemed favorable by the favorability maps. The favorability maps produced allow the identification of favorable zones across the studied areas, and, based on the MDI scores outcoming from the RF, can also give insight on the metallotects associated with the studied mineralization. However, favorability maps do not identify the typical mineralized geological environment related to the favorable areas. Being able to identify these associations allows more insight about the possible metallotects of the various mineralizations and enable the differentiation of the various type of mineralization used to perform the MPM.

Synthesizing the favorable lithological associations from the CBA / DBA RF matrix requires first the definition and selection of the cells and corresponding lithological associations with the highest favorability score. A threshold must be set on the favorability score in order to select the cells considered as favorable ones. Then the cells presenting favorability score above this threshold are classified using the HCPC clustering method.

This threshold is here determined based on the ROC curve of the CBA / DBA RF model. From this curve, it is possible to determine the optimal score threshold to classify the cells as mineralized and non-mineralized ones: it corresponds to the point of the ROC curve giving the best trade-off between the TPR and the FPR. This point would be located on the curve, near the top-left corner of the graph. Two metrics can be used to locate this point: the Geometric Means (G-Means) and the Youden's J statistic (J-Score) [Youden, 1950].

$$G-Mean = \sqrt{TPR * (1 - FPR)}$$

$$J-Score = TPR - FPR$$

$$TPR = \frac{TruePositive}{TruePositive + FalseNegative}$$

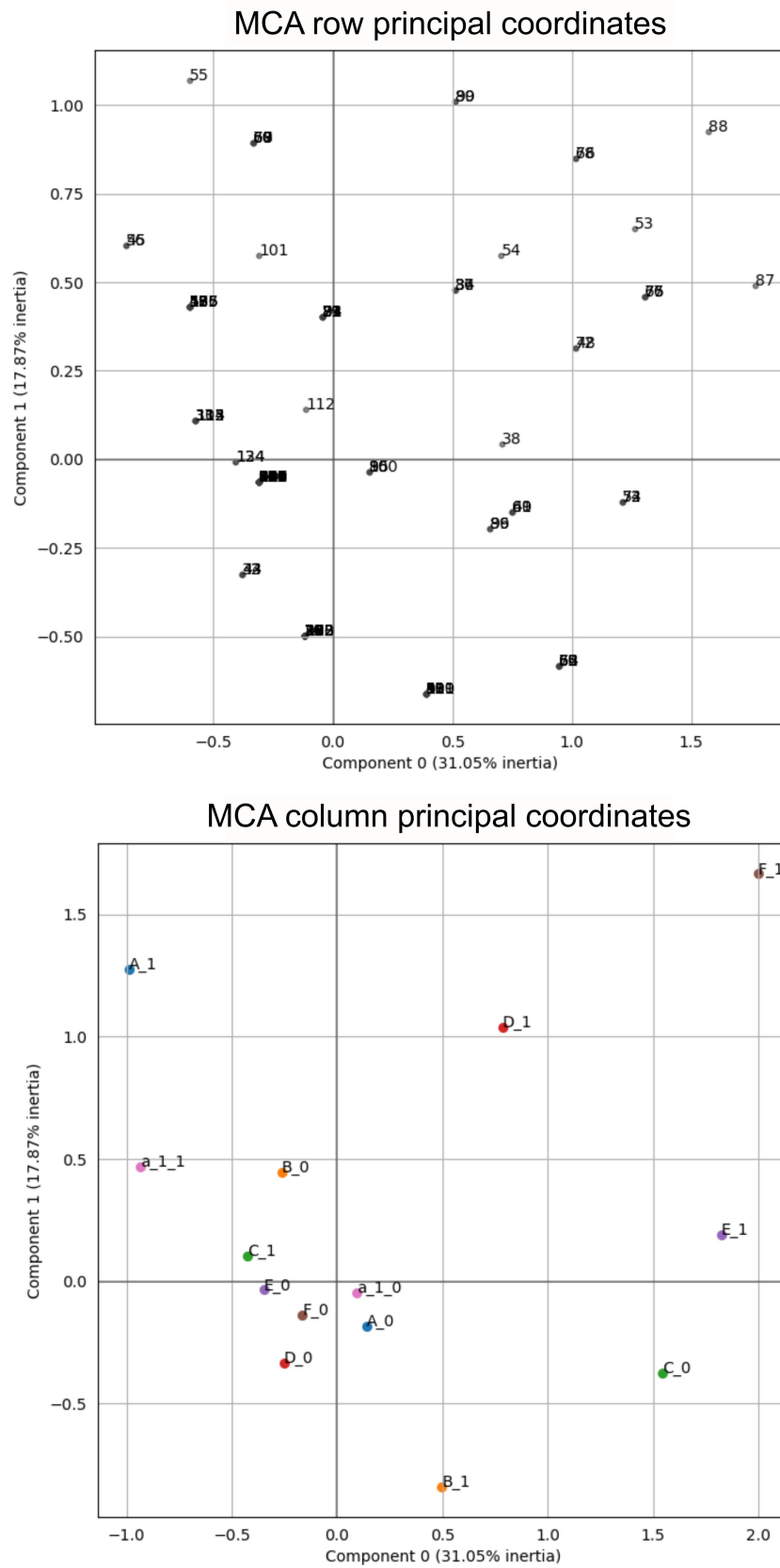


Figure 5.5: Projection of rows (cells) and column (modalities of our lithological associations) on the first 2 factorial axis. Each factorial axis explains a percentage of the total inertia of the data.

HCPC classification

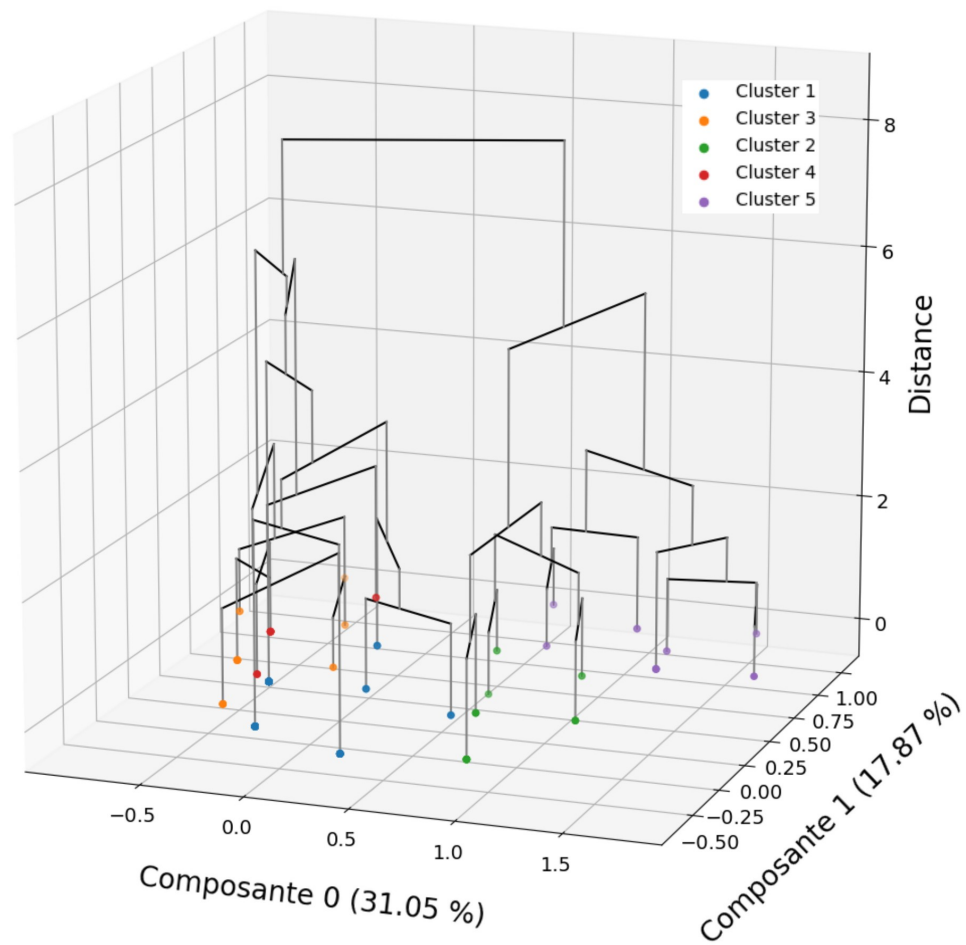


Figure 5.6: 3D HCPC clustering tree. The 2D dimensional space corresponding to the 2 best factorial axes given by the MCA are displayed on the XY axis. The inertia explained by the 2 axes is shown in percentage. The Z axis shows the height (Ward's distance) between the nodes of the AHC classification tree.

$$FPR = \frac{FalsePositive}{FalsePositive + TrueNegative}$$

Those scores are computed for each threshold used to produce the ROC curve. The threshold giving the highest G-Means or J-Score is the one giving the best trade-off between TPR and FPR. This threshold is then used to select the cells to be classified: all cells with a score equal to or greater than this threshold are selected (**Figure 5.7.a**). Then, the HCPC classification can be applied on the selected cells to identify the various geological environments (**Figure 5.7.b**), the number of classes being determined manually. It can come out that all favorable "mineralized" cells could be linked to a unique lithological association signature (which would correspond to a single and rather simple metallogenic process) or to several ones. If needed, the threshold can be set manually to determine the geological signature of more specific areas: increasing the threshold will result in fewer cells and target the most favorable areas, while decreasing the threshold will select more cells and target less favorable areas. However, the use of G-Means or J-Score to determine this threshold allows the automation of this selection from the results of the predictive model.

The resulting clusters can be further analyzed by computing the frequency of the lithologies inside them (**Figure 5.7.c**). This allows identifying the geological environment (i.e. the global association of lithologies) inside this cluster. This geological environment can give clues regarding the association of environments found in the favorable areas. The geological environment of the clusters can be compared to the one of the study area: comparing the frequency of the lithologies inside the clusters to the one across the map displays the specificity of each of the favorable geological environments and the possible anomalous behaviour of some lithologies. As shown for example in **Figure 5.7.c**, cluster number 5 displays very high frequencies for the formations D and F. Those formations can be considered as a marker of this favorable geological environment and their association a possible "statistical" metallogenic. In addition to the MDI given by the RF method, which only gives information on the most efficient lithologies for the classification, those favorable geological environments give more insight about the lithologies and association of lithologies that are linked to the mineralization according to the predictive map.

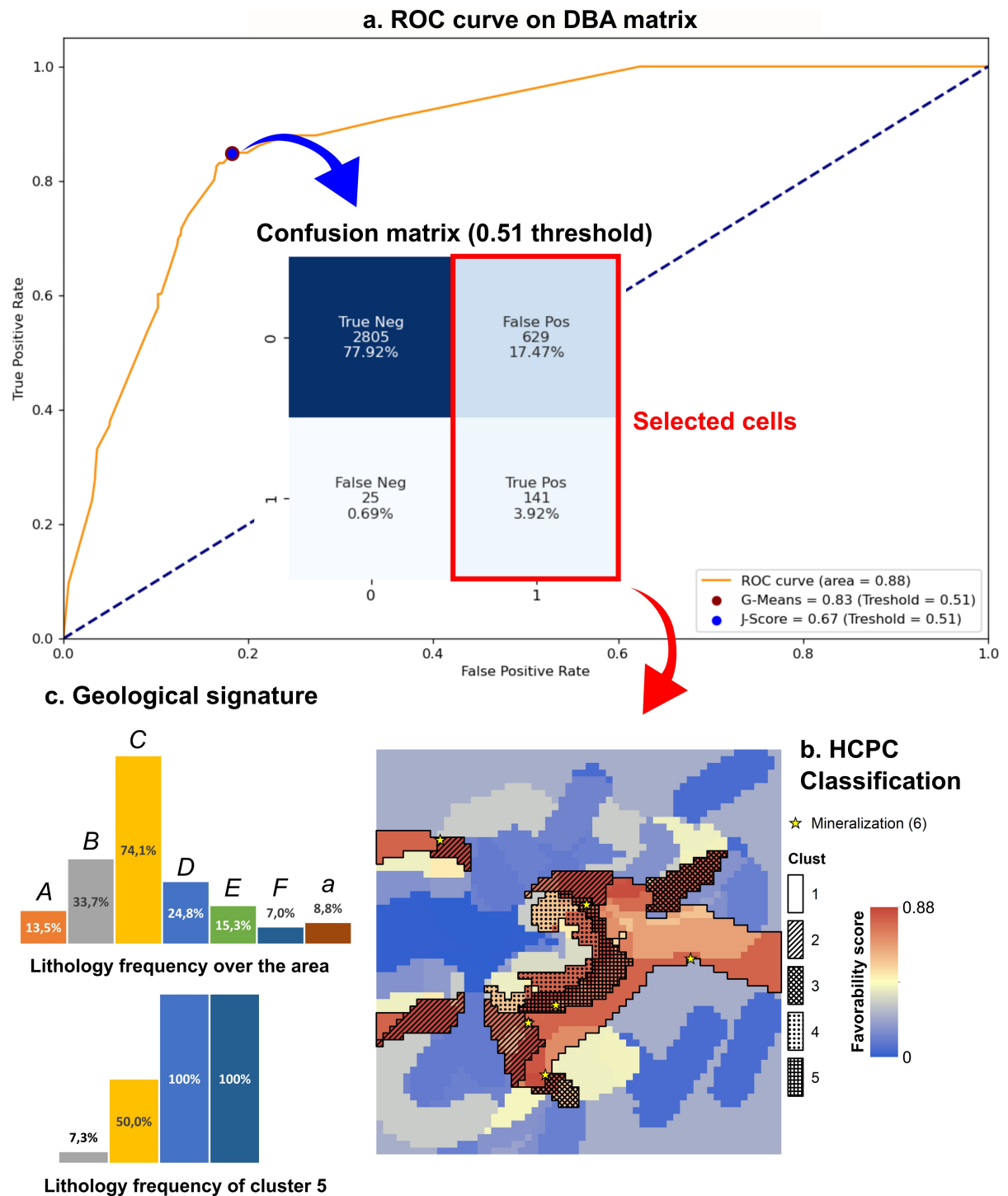


Figure 5.7: Identification of the geological environments surrounding the favorable areas on a sketch DBA / RF prospective maps (1 000 m resolution). a. The ROC curve is used to determine the threshold determining the favorable cells, using the G-Means and J-Score. b. The HCPC classification is performed on the lithological associations of the selected cells, producing clusters resuming the multiple lithological associations into fewer geological environments. c. The frequency of the lithologies inside the cluster is computed and compared to the ones across the study area to identify the geological environment of this cluster.

Part III

Applications

Chapter 6

Proof of concept

*The methodology developed and presented in previous **Chapter 3**, **Chapter 4** and **Chapter 5** is here partially applied as a proof of concept. While this methodology integrates the DBA, RF and HCPC methods, the proof of concept detailed in **Chapter 6** is limited to the combination of the CBA and RF method to produce prospectivity maps of antimony within the Armorican Massif. This proof of concept aims at displaying the advantages of using a machine learning based favorability scoring method in combination with the geological environments identified by the CBA instead of the methods that were previously used, such as the ones presented in the beginning of **Chapter 4**. The data used for this proof of concept are similar to the one used in the application of the DBA / RF / HCPC method on the Armorican Massif shown in **Chapter 7**, where the full application of the new methodology developed in the frame of this thesis is explained. Previously used prospective mapping methods, as detailed in **Chapter 4**, have been applied in previous studies on a very similar case, to map the favorability of antimony in the Armorican Massif. The results of those studies are presented at the beginning of this **Chapter 6**. This allows a comparison between the results given by those methods and the one based on the CBA / RF method presented in this **Chapter 6**, as well as a comparison to the application of the complete methodology shown in the following **Chapter 7**.*

6.1 Presentation of the study

6.1.1 Context

In order to test the application and automation of a new predictivity method developed in this thesis work, the Armorican Massif (AM) has been chosen as a natural study area. Before the applications of each method described above in **Chapter 3**, **Chapter 4** and **Chapter 5**, which are shown in the following **Chapter 7**, the predictive method applied here is comprised of the Cell-Based Association (CBA) and the Random Forest (RF). First, the CBA is applied to the dataset presented below, then the RF classification is computed on the CBA matrix to produce the favorability maps. The application of this CBA / RF method allows us to estimate the effectiveness of this predictive method in real conditions (i.e. with data similar to the ones used in other prospective mapping methods and for rather classical case studies conditions) before comparing it to the developed DBA / RF / HCPC methodology. As part of the AUREOLE project, aiming at a better understanding of antimony mineralization throughout the Variscan

Range, the application of the developments is made here to study antimony mineralization within the AM, where five major antimony districts have been recognized to date. Preliminary spatial analysis [Pochon, 2017] and predictive map computation [Gloaguen and Tourlière, 2016a] have been recently carried out through the AM. Based on the same datasets than the ones used for these previous studies, results outcoming from the developed CBA / RF method can be easily compared to both (i) the metallogenic models proposed, up to now, for antimony deposits through the AM and (ii) the numerical results from the two mentioned previous works. The production of favorability maps using the CBA combined with RF on this area would allow testing and evaluating the results of those obtained by the application of the CBA with different methods of scoring. The use of the CBA instead of the DBA allows first comparing the various favorability mapping methods based on the CBA, without modifying all the parameters used in those methods. The DBA is applied in a second phase, as presented in **Chapter 7, p.143**.

The new CBA / RF method is applied to the AM in order to produce favorability maps for antimony deposits. The lithological associations resulting from the CBA space discretization are used to produce favorability maps based on the RF method described above. As explained in the end of **Chapter 1**, the goal of the developed methodology is to produce a "statistical" conceptual model in an automated way by analyzing the spatial data. As such, no conceptual models are used beforehand to select the metallotects / predictors identified in **Chapter 2** and most data are imputed as is in the CBA method. The only exception is the focus on Paleozoic units: recent sedimentary covers are ignored to reduce the number of lithology entries in the CBA matrix and reduce the noise regarding the studied phenomenon, which is of Paleozoic age. The RF method is combined with Cross-validation to compute Receiver Operating Characteristic (ROC) curves and evaluates the robustness of the predictive model. The ROC curves will then be used to select the cells presenting an optimal favorability score. Results of this CBA / RF approach are compared to previous studies and favorability maps obtained by Gloaguen and Tourlière [2016a].

6.1.2 Previous studies

Gloaguen and Tourlière [2016a] performed a first prospective and favorability study of antimony within the AM using both WoE and CBA. The Ranking and AHC methods were applied to the CBA process as described in details in **Section 3.1, p.71**. Those methods were carried out with the same lithological dataset used in the present study, in addition to geophysical data (e.g. gravimetric and magnetic anomalies) but without structural data.

The CBA / AHC method (**Figure 6.1**), rather than estimating likelihood of hosting mineralization, points out areas containing similar lithological associations, according to the AHC classification. In a second stage, environment classes are linked to the known main antimony deposits in order to highlight potential favorable areas. In Gloaguen and Tourlière [2016a], only 16 deposits, considered as the most significant ones, were used. The resulting maps highlight areas sharing similar geological associations with the main antimony deposits, with high discrepancy in the number of cells and a high proportion of non-selected mineralization that are not located in those output favorable areas. This method gives mainly qualitative results, as the resulting classes do not carry out favorability scores.

The WoE method (**Figure 6.2**) is able to highlight some of the areas known to host mineralization, but suffers mainly to its sensitivity regarding (i) clustering and positioning of mineralization and (ii) the relative surface occupied by the geological formation. As such, the WoE method has the disadvantage of

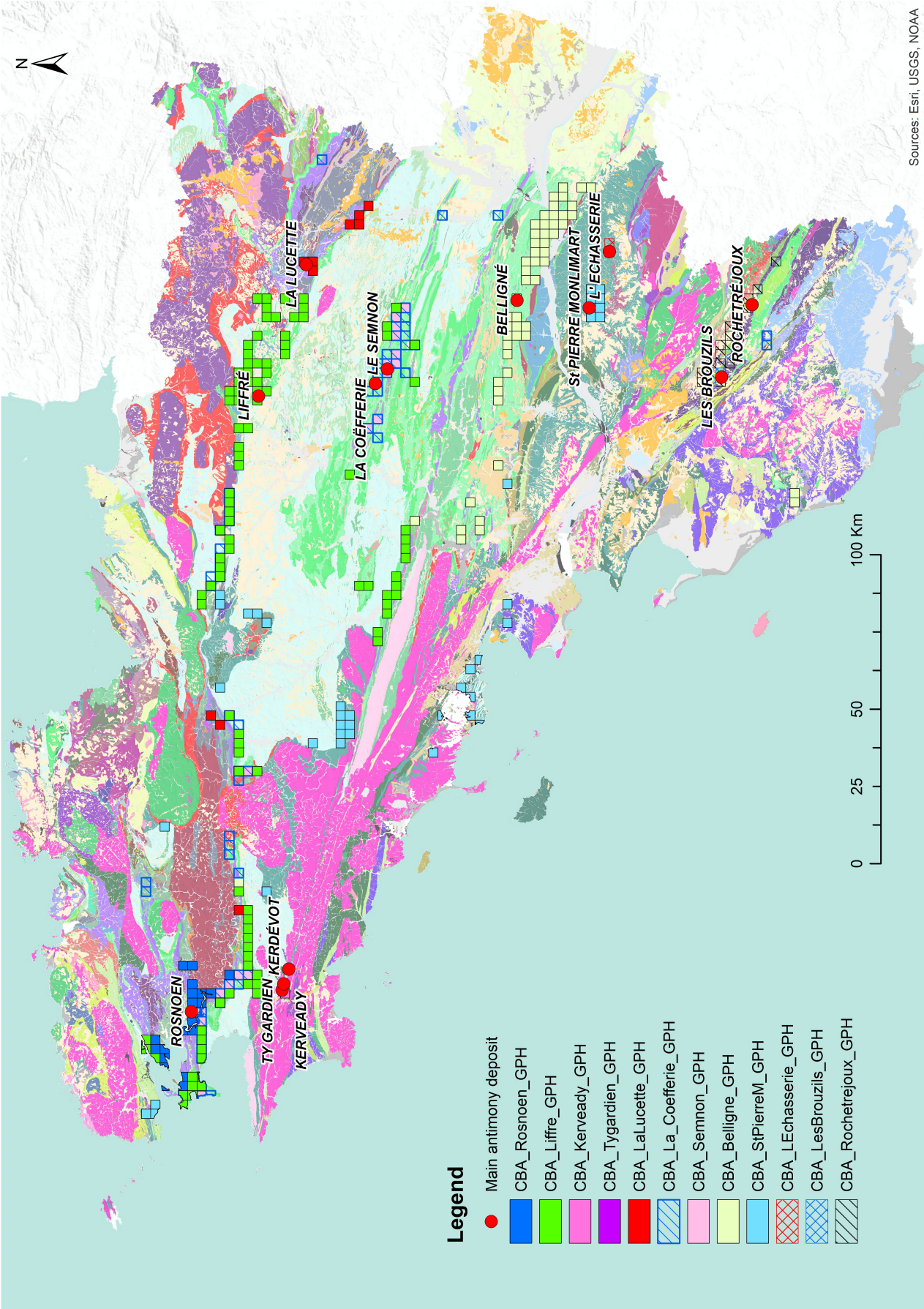


Figure 6.1: Results from the study by Gloaguen and Tourlière [2016a]: CBA / AHC method with the 12 main environments identified from the 16 deposits selected.

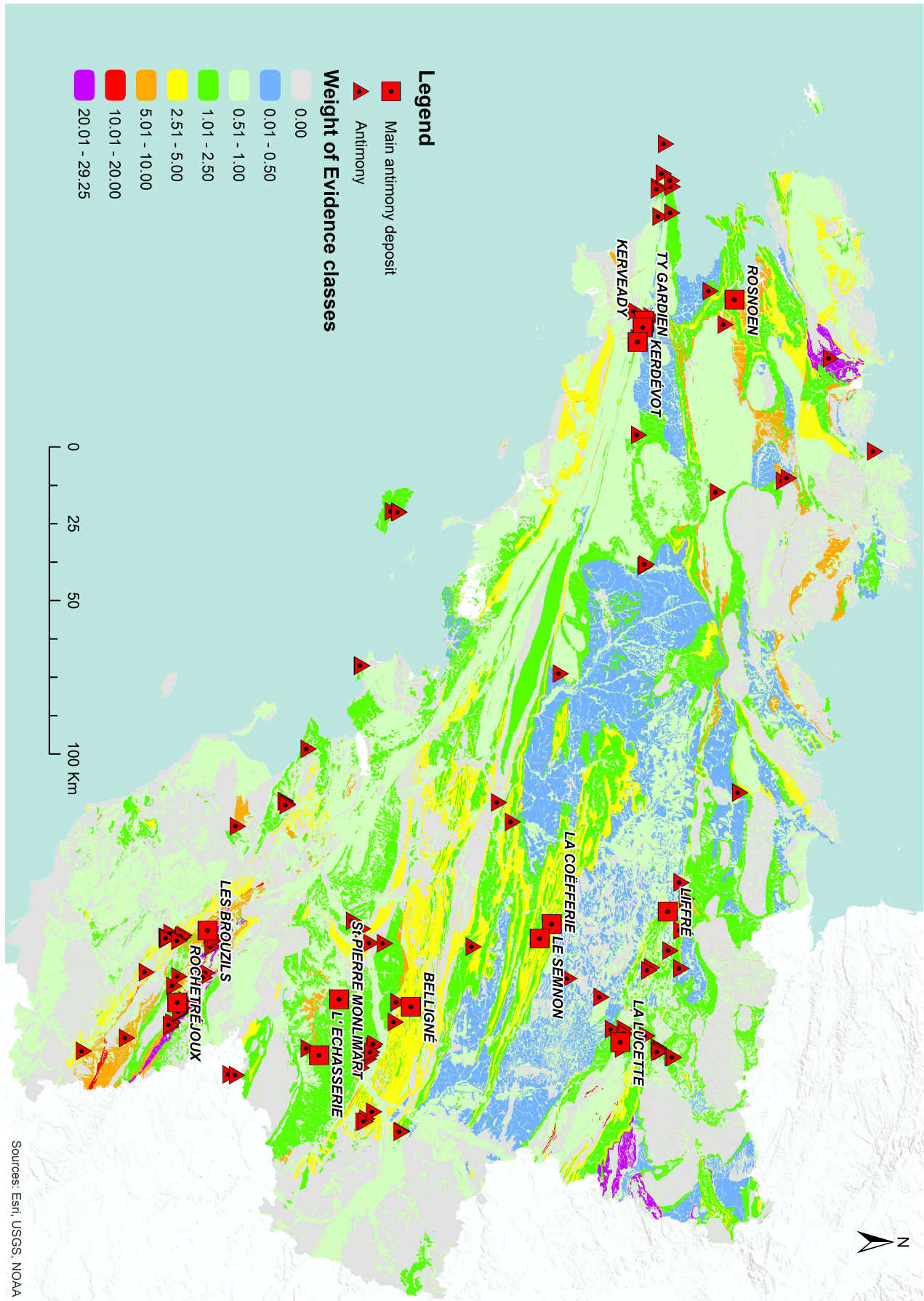


Figure 6.2: Results from the study by Gloaguen and Toulière [2016a]: WoE favorability map computed with all the antimony deposits located in the study area.

over highlighting the small formations containing mineralization, while larger ones have lower scores, even though they contain most of the mineralization. Indeed, WoE requires large datasets of deposits to have a statistical meaning and avoid this kind of bias. This method is also constrained by the geological data used, as the final results are displayed based on the same geological formations outlines as the one used as input. Finally, it is also required to select the mineralization or pre-process the geological formations in order to avoid highlighting covers formations (e.g. quaternary, anthropogenic formations, alluviums ...).

The CBA Ranking method (**Figure 6.3**) was done based on 13 deposits which define the “Standard Cells” for the method. In contrast with the CBA / AHC method, this method is able to estimate quantitative favorability scores for each cell. Clustering bias is ignored here due to the use of the CBA matrix, but the bias induced by the relative surface covered by each formation in the map remains, as the favorability score depends on the number of cells containing a given lithology. The Ranking method thus favours formations covering a small surface, just like the WoE. Using this approach, all the lithologies with a Ratio above 1 are statistically regarded as favorable. However, the favorability score is not normalized and, by definition, it scales up to the maximum resulting score over the cells. For instance, in the application of [Gloaguen and Tourlière \[2016a\]](#), scores reach up to 37.9 in some cells. It comes out that, even if this method gives qualitative results, they only have a relative – instead of an absolute – meaning. Also, post-processing is required, and a threshold has to be manually setup to clearly delineate the area(s) of interest within the final prospective map.

Geophysical datasets, including gravity Bouguer and magnetic anomaly maps, were integrated in the [Gloaguen and Tourlière \[2016a\]](#) study, by discretizing the gravity anomaly into heavy and light anomalies and the magnetic anomalies into positive and negative anomalies. Those categorical variables can then be integrated into the CBA matrix using the same binary process: each type of anomaly is a variable, which is either present ("1") or absent ("0") of each cell. It is noteworthy that the previous used classification methods, compared to the CBA / RF, have no possibility to handle directly continuous datasets grids (i.e. Regionalised Variables) and require these ones to be discretized in order to be integrated in the MPM process. The geophysical data available for the AM clearly supplements the geological dataset, as geophysical data tends to improve predictive models results and allows to better constrain favorable areas [[Gloaguen and Tourlière, 2016a](#), [Moon and An, 1991](#), [Riahi et al., 2021](#), [Rodriguez-Galiano et al., 2015](#)]. However, the discretizing process (see heavy vs. light anomalies and positive vs. negative magnetic anomalies respectively, [Gloaguen and Tourlière \[2016a\]](#)) can be questioned and could induce bias in a pre-processing stage.

6.1.3 Dataset

The regional case-study of our new integrated CBA / RF prospectivity approach is based on an exhaustive and homogeneous digital 1:50,000 set of geological data available for the overall AM (**Figure 6.4**; [Gloaguen and Tourlière \[2016a\]](#)). It comprises one polygons-layer for the geological units. This data (including geographic and attribute data) originally come from the CHARM harmonized dataset (see InfoTerre website: <http://infoterre.brgm.fr/>). These 1:50,000 digital geological maps correspond to some harmonization made for administrative districts (‘départements’ in French) based on the different sheets published by the BRGM independently and at different periods, over the French territory. The AM covers 8 of these administrative districts with one harmonization map for each: ‘départements’ n°85 [[Béchenec, 2009](#)], n°56 [[Cagnard, 2009a](#)], n°53 [[Vernhet, 2010](#)], n°49 (unpublished), n°44 [[Béchenec,](#)

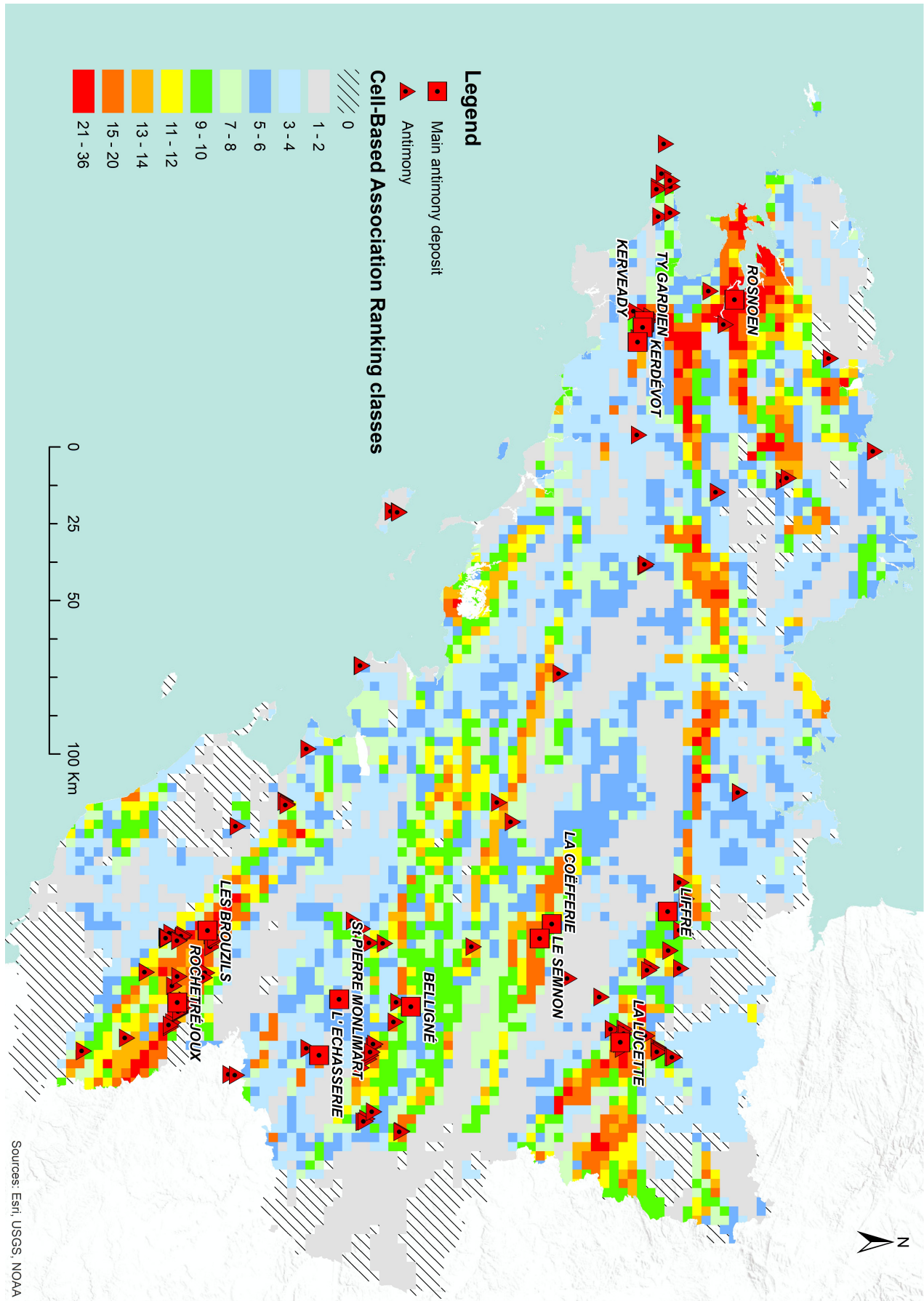


Figure 6.3: Results from the study by Gloaguen and Toulière [2016a]: CBA / Ranking favorability map computed from the 13 main antimony deposits.

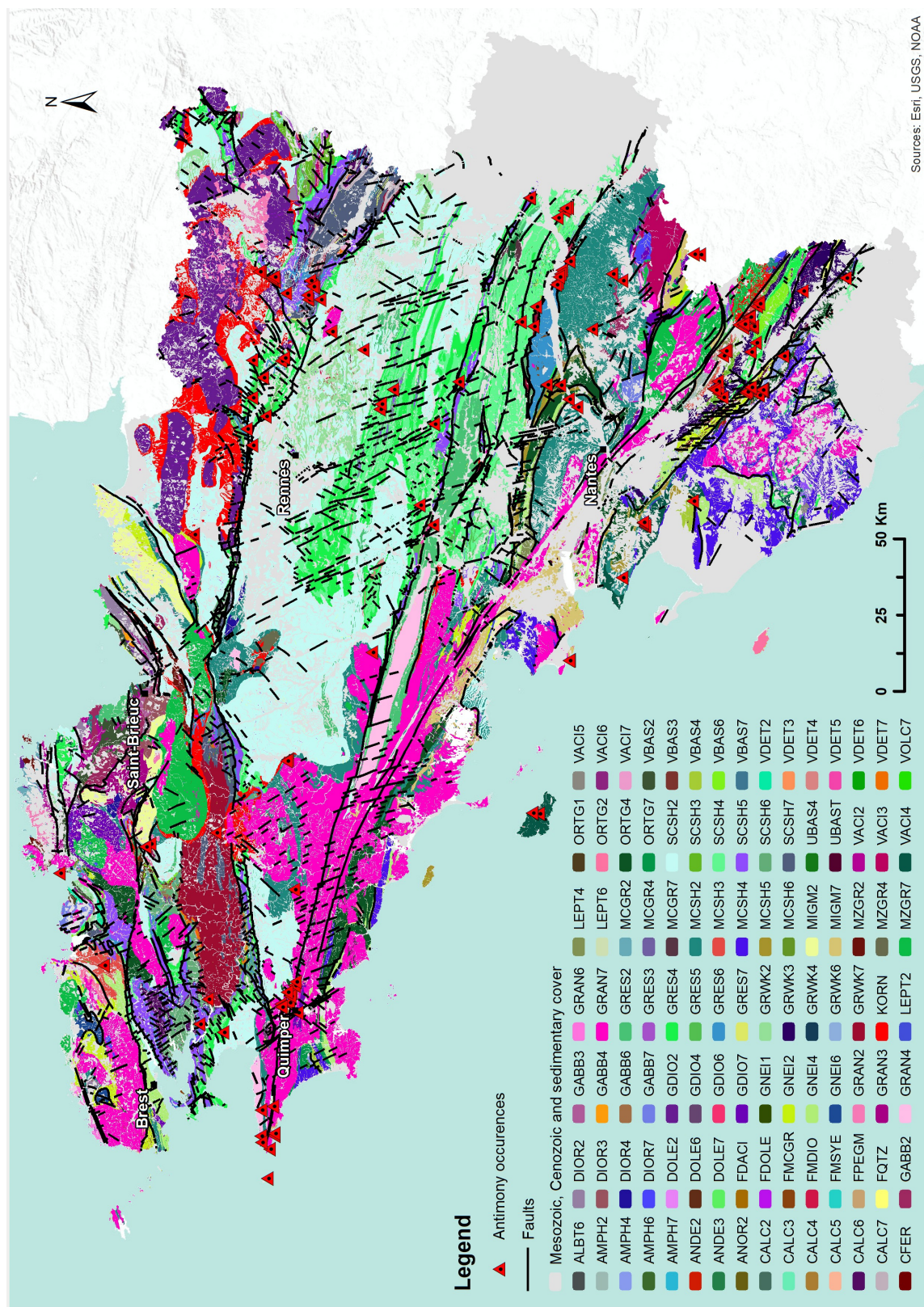


Figure 6.4: Maps of the datasets used in this study. Palaeozoic and Precambrian formations (in colors): cover 65.35% of the studied area; Mesozoic and Cenozoic formations (in grey): cover 34.65% of the studied area. Faults are from the 250K sythetic map of the Armorican Massif, produced by [Chantraine et al. \[2001a\]](#) (unpublished) and mineralization are from the BRGM Database [[Cassard et al., 2015](#)].

2007], n°35 [Cagnard, 2009b], n°29 [Cagnard, 2008] and n°22 [Egal, 2010]. This corresponds to 128 original sheets of the 1:50,000 geological map. A preliminary integration and harmonization work has been made for the entire study area [Gloaguen and Tourlière, 2016a]. Concerning the geological units description in particular, these have been homogenised and simplified based on a code defined from the lithological and stratigraphic exhaustive descriptions made for each individual sheet. It results in a set of letters corresponding to the lithology completed by a sequential number defining the relative age of the formation (deposition, intrusion, metamorphism). Details about this harmonization index can be found in **Appendix B.1, p.264**. This homogenized 1:50,000 scale dataset allows developing a regional-scale prospectivity approach while maintaining the precision and local-scale complexities in (i) the shape of geological outlines, (ii) the size of isolated small map features such as the dykes, for instance and (iii) the distinction between different lithologies.

Antimony mineralization in the AM are mainly linked to Palaeozoic metallogenic events during the Variscan orogeny [Chauris, 1977, Marcoux and Chauris, 1994, Munoz et al., 1992, Pochon et al., 2019]. In order to avoid noisy formations, such as quaternary sedimentary covers or Mesozoic sedimentary formations, only Palaeozoic and older formations of the harmonized map dataset of the AM are retained in this study. Those Palaeozoic formations represent about 65% of the surface of the original dataset shown in **Figure 6.4**.

The digital geological dataset also includes one single polylines layer for the map-scale structures. These have been extracted from the 250K synthetic map of the AM, produced by Chantraine et al. [2001a] (unpublished), outlining the major faults in the studied area. Finally, in addition to the digital geological map information, a set of 141 known antimony occurrences of various sizes has been extracted from the BRGM database [Cassard et al., 2015] updated by Gloaguen and Tourlière [2016a] over the AM. Those antimony deposits are mostly veins-type deposits, with some exceptions as the stratabound-type deposit of Rosnoën. No geophysical or geochemical datasets are used in this application.

6.2 Application of the Random Forest to CBA

6.2.1 CBA matrix

CBA method is applied on the harmonized 50K geological dataset of the AM as compiled in Gloaguen and Tourlière [2016a,b]. The spatial resolution optimization process results in a 3 000 m cell size grid of exactly 5 865 cells. Results of the four parameters used to define the optimal resolution correspond to the graphs displayed in **Figure 6.5.a** to **Figure 6.5.d**. For a 3000 m cell size, only 12.3% of the cells contain a single lithology and 2 574 unique environments are defined. On average, cells comprise 5 distinct lithologies with a maximum of 14 for one single cell. This configuration allows producing a MPM map of a resolution high enough to be exploited and cells large enough to integrate multiple map features for further classifications on the environments. This grid is consistent with the size of explorations permits, which can reach standard areas between 20 to 100 km² [Hronsky and Groves, 2008, McCuaig et al., 2010], around 3 to 12 cells of the CBA.

In order to integrate map scale structures in the predictivity model, the absence or presence of faults intersecting the cells are added to the matrix, whatever their orientations or natures. It is noteworthy that some distinctions between fault classes could be performed before their integration in the matrix. Finally, each cell with its center closer than 3 000 m from a Sb occurrence is considered as mineralized. This CBA

Cell-Based Association iterating process on the Armorican Massif

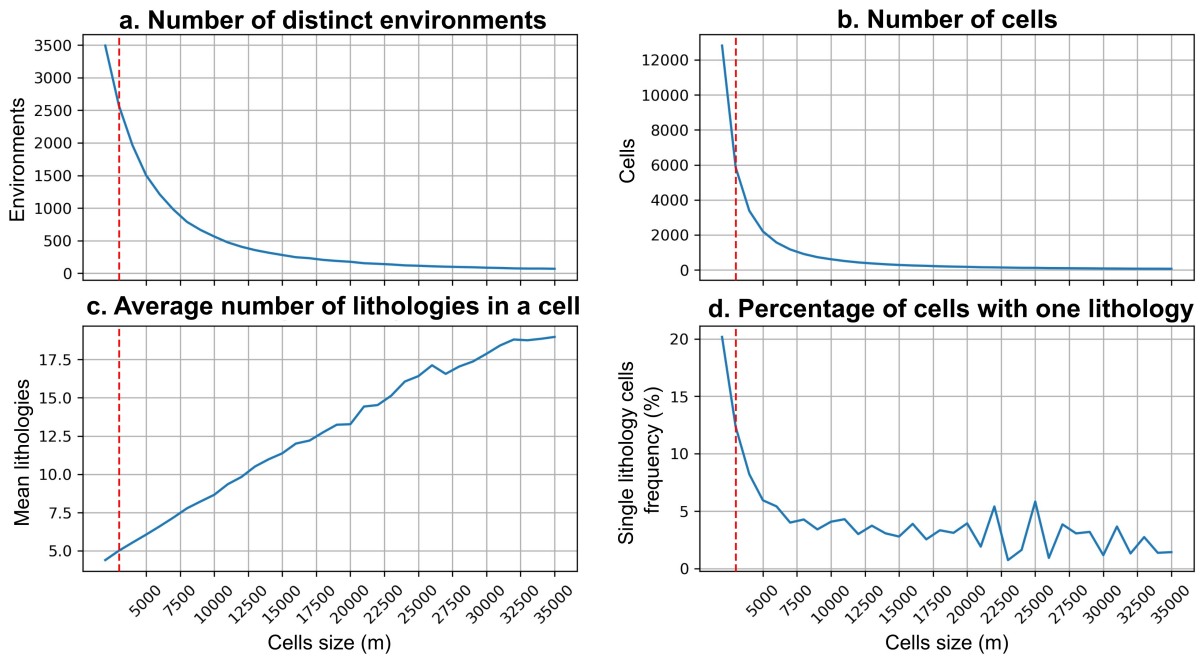


Figure 6.5: a. to d. Application of the multi-scale iterating process of CBA matrices on the dataset. Those graphs show the impact of the cell size on different parameters of the CBA matrix. The red dashed line is the optimal value found for the cells size (3 000 m), where grid resolution, number of environments and single lithology cells frequency are considered optimal.

processing results in a cells grid with a corresponding matrix summarizing the presence or absence of lithologies or faults within each cell, and the initial metallogenic class (“mineralized” or “non-mineralized”) of each cell.

6.2.2 Results

RF predictive modelling is applied on this matrix to compute the favorability score of each cell, based on the known antimony mineralization occurrences within the studied area. Overall, the resulting favorability scores show a large range of values and the corresponding map (**Figure 6.6**) clearly displays areas showing scores consistent with the location of the known mineralization, with distinctive high values (up to 0.86) near them, and with little noise effect (e.g. isolated cells with high score values). The most and least favorable areas regarding antimony mineralization are displayed in red and blue, respectively. At first glance, the known antimony deposits are well located within favorable areas, as displayed in the **Figure 6.6**.

6.2.3 Selection of high favorability cells

The performance of the predictive model can be analysed using the resulting predictive classes: cells can be classified as True Positive (TP), True Negative (TN), False Negative (FN) and False Positive (FP) in a post-processing stage. The first 2 types correspond to cells which predictive classes match their initial classes, while the others are regarded as “misclassification”. Those predictions depend on the score

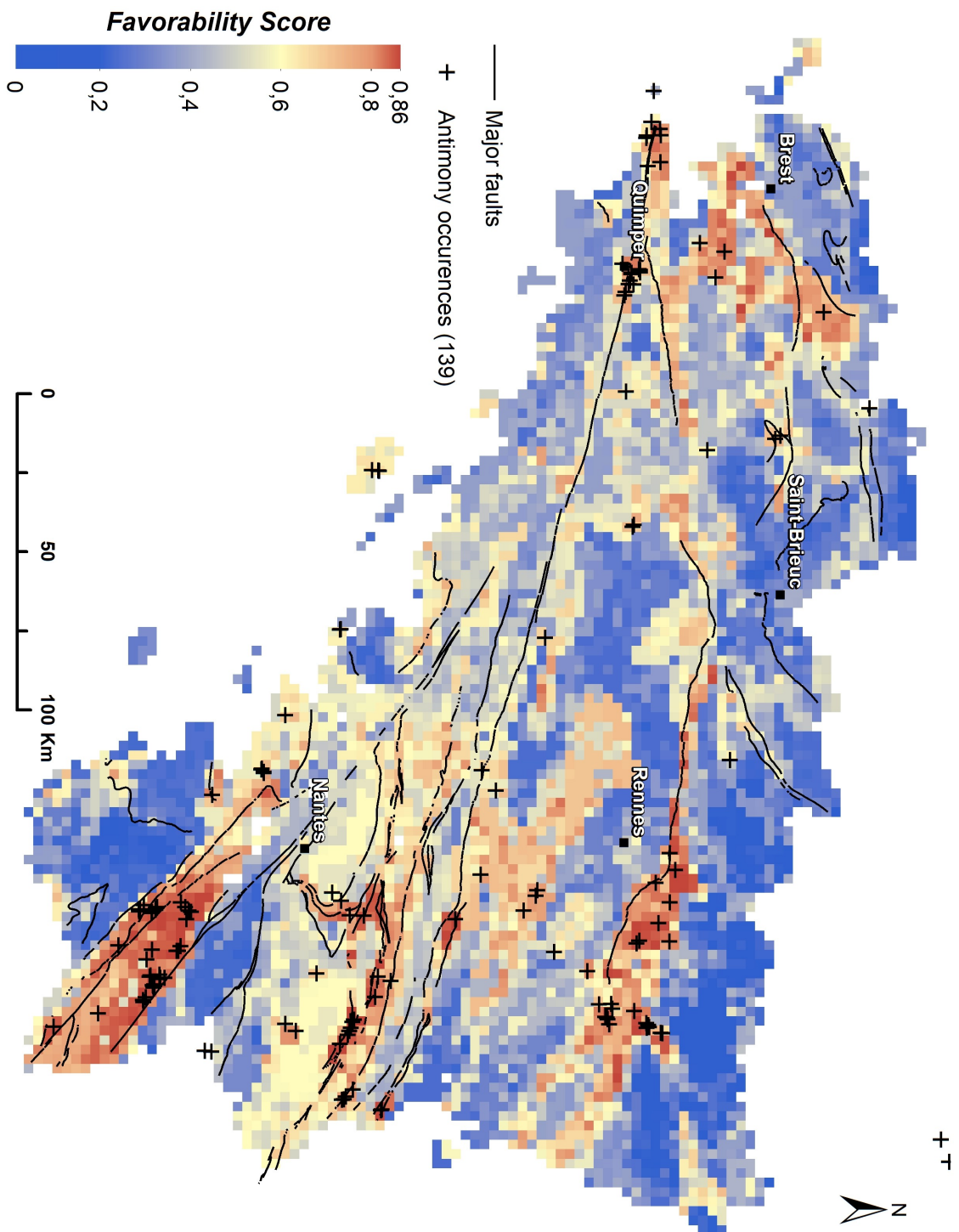


Figure 6.6: Favorability maps of antimony in the Armorican Massif produced from the integrated CBA / RF method. The CBA matrix is based on the absence and presence of faults (from 250K synthetic map) and of each lithological formation (from 50K harmonized dataset). Cells size of 3 000 m. Mineralized cells are produced using buffers of 3 000 m around the center of each cell. High favorability score are displayed in red and correspond to a higher likelihood for the cell to contain antimony mineralization.

threshold retained to classify the cells: above this threshold, a cell is considered as mineralized (class “1”). Changing this threshold will then change the populations of the mineralized and non mineralized classes, and thus the TP, TN, FN and FP populations. The Receiver Operating Characteristic (ROC) curve allows the evaluation of the predictive model using those classes, through the True Positive Ratio (TPR) and the False Positive Ratio (FPR). Those ratios are computed for a continuous range of threshold values and projected in a graph.

In the ROC curve, an efficient model should show high values of TPR for low values of FPR. In our case, it means that an efficient model should be able to correctly classify the mineralized cells while minimizing the misclassification of non-mineralized cells. A ROC curve has been computed for a 5-fold Stratified Cross-Validation (SCV) of the model to evaluate it after the hyperparameters tuning (see resulting average value curve and standard deviation in **Figure 6.7**). In this case study, the ROC curve resulting from the 5-folds SCV of the CBA / RF model (**Figure 6.7**) shows a mean AUC of 0.80 ± 0.02 , meaning that the classification is quite effective, as a model working with random prediction would give an AUC around 0.50.

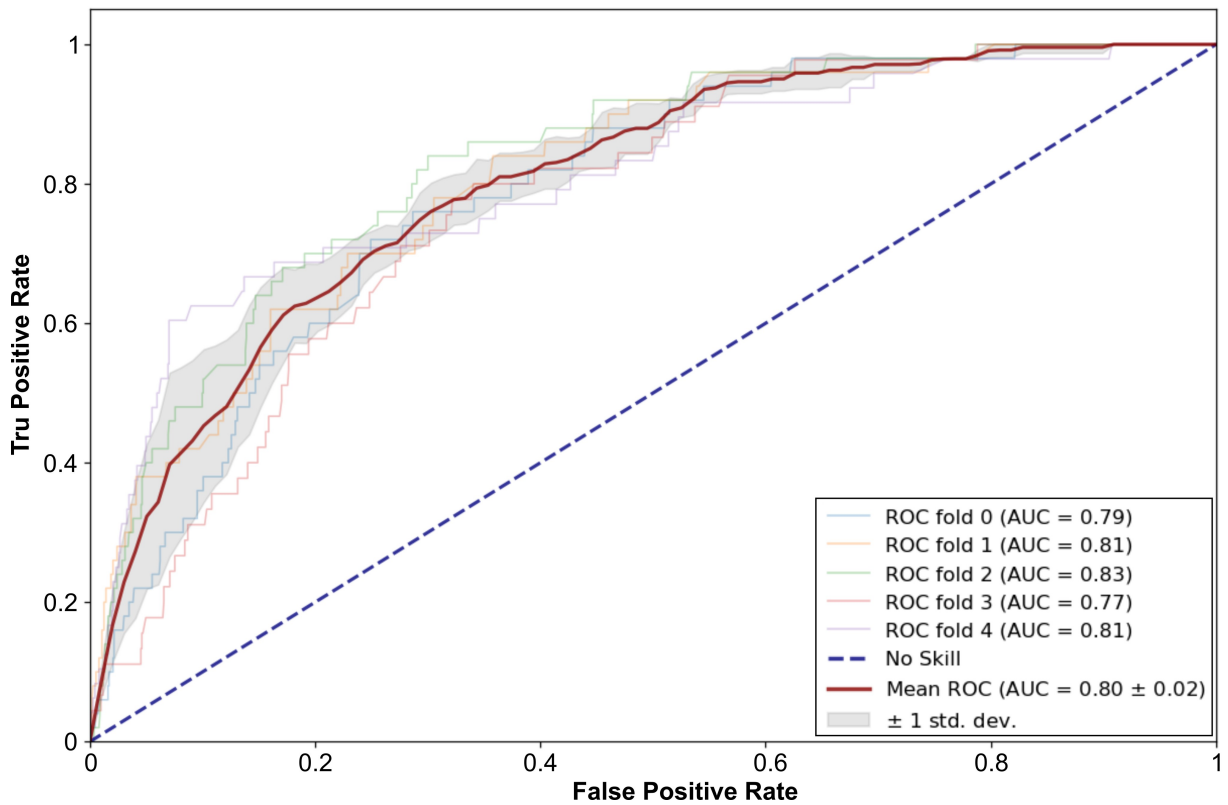


Figure 6.7: Receiver Operating Characteristic (ROC) curve produced from the 5-folds Stratified Cross Validation of the Random Forest predictive model. Each fold is displayed with its corresponding Area Under the Curve (AUC) score. The no-skill model displayed in dotted blue corresponds to a theoretical worst-case predictive model

The favorability map resulting from the new CBA / RF method developed here (**Figure 6.6**) highlights most of the mineralized cells. The most favorable areas, with scores above a threshold of 0.5, are displayed in **Figure 6.8.a**, with 78% of the mineralized cells having a score above this threshold (**Figure 6.8.b**). These cells are mostly located within the main antimony districts. Cells showing the highest scores are located around the *Vendée* and *La Lucette* districts, with a mean value of 0.744 and 0.708 respectively. The *Rosnoën* and *Quimper* districts show slightly lower favorability scores with mean values of 0.673 and

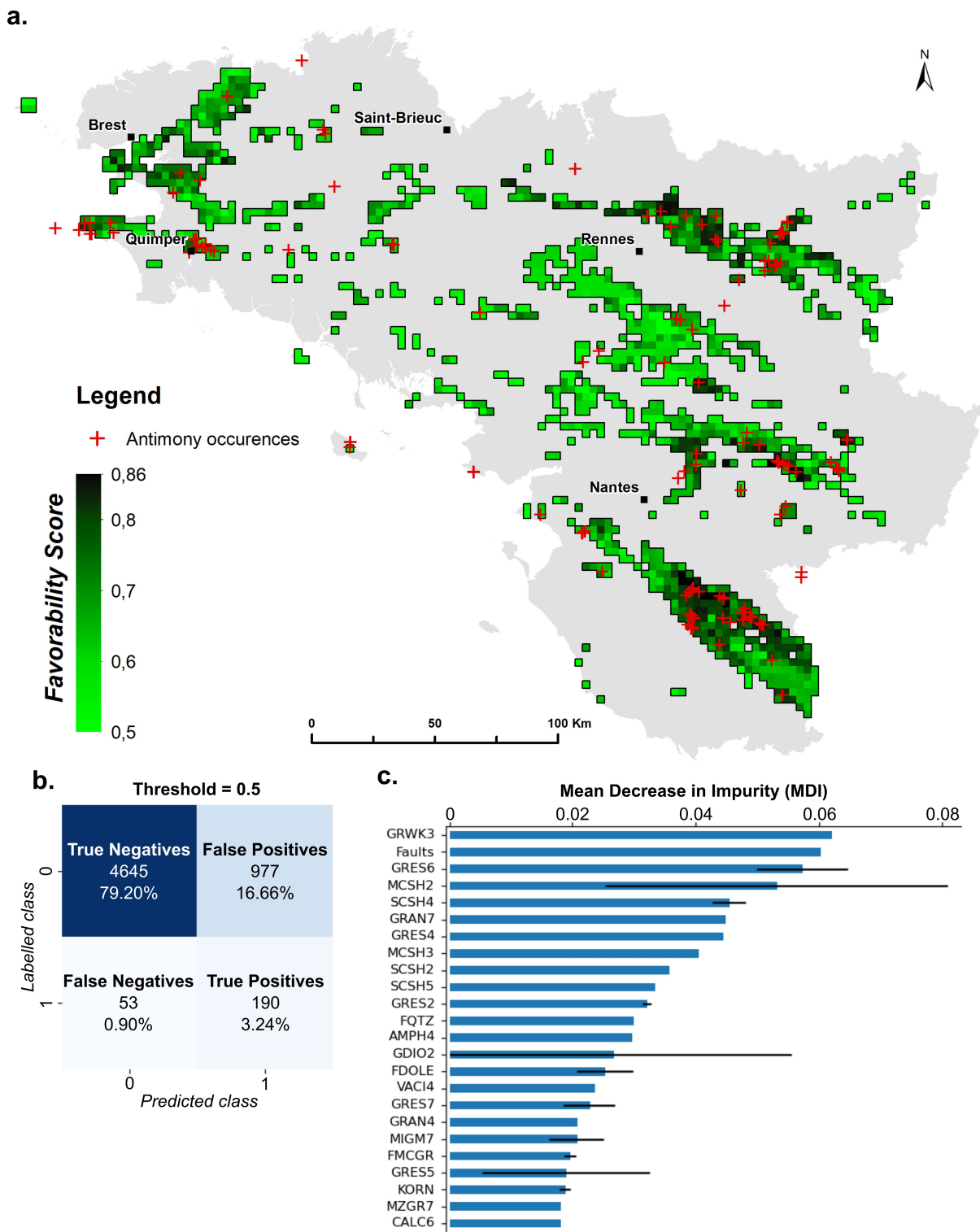


Figure 6.8: Analysis of the cells displaying a score higher than 0.5 in the Armorican Massif favorability map from the integrated CBA / RF method. a. Favorability map of cells displaying a score higher than 0.5, the color ramp is scaled for score higher than 0.5 b. Confusion matrix of the cells based on a 0.5 threshold for the favorability score. The number of cells in each category is displayed, the percentage is relative to the total number of cells. The x-axis corresponds to predicted values ("1": mineralized, "0": non-mineralized) while the y-axis corresponds to the initial classes. c. Mean Decrease in Impurity (MDI) score of the main predictors (see text for further details)

0.680, respectively. The district showing the worst favorability scores from our model is the *Semnon* with a mean value of 0.644 for mineralized cells. Cells with the lowest favorability scores are mainly located outside of those areas, far from the main districts.

6.2.4 Interpretation

Based on their similarity with the mineralized environments identified from the RF classification, new zones (i.e. cells with no occurrences known to date) with high antimony potential mineralization also appear in the map. The predictive model results in 17.4% of the non-mineralized cells showing a favorability score above the 0.5 threshold, meaning that they show similarities with known mineralized cells environments. Most of those cells are located near the main districts, where the environments are most likely to host antimony, but, interestingly, some are also located outside those areas. These cells are mainly located along the SASZ and the NASZ, while others are located in small patches within the SAD or in the folded Palaeozoic sedimentary series of the CAD, south of Rennes.

The Mean Decrease in Impurity (MDI, presented above in [Section 4.2.6, p.109](#)) score is displayed in [Figure 6.8.c](#) for each predictor. This score is based on the Gini gain associated with each predictor, averaged for the overall computed trees. The best predictors produce higher MDI scores, as they have a higher Gini gain and are used more frequently than the others. The error bars displayed in this graph show the Standard Deviation (SD) of this MDI value for each predictor. This SD shows the inter-trees variability of the MDI score for the predictors. Predictors with no error bars shown have a very low (i.e. negligible) SD.

According to the MDI of this RF predictive model, the predictor showing the best results to classify the cells, and are therefore the most likely to be associated to mineralization, are Cambrian Grauwackes (GRWK3) located within the Vendée district ([Figure 6.8.c](#)). This district contains most of the mineralized occurrences referenced through the AM ([Figure 6.8.a](#)) and also shows the highest favorability scores. The second best predictor for this model are the faults ([Figure 6.8.c](#)). According to the metallogenic models described in the literature, this result also seems reasonable as most known antimony deposits in the AM are linked to structural features [[Cheval-Garabédian et al., 2020](#), [Munoz et al., 1992](#)], and in particular regional-scale shear zones such as, among others, the NASZ, the northern and southern branches of the SASZ or the MNSZ. It would be interesting to integrate more details about the faults in the matrix, such as distinguishing the types or orientations of the faults, to further evaluate the relationships with the structural features. Another good predictor is the Carboniferous sandstones unit (GRES6), with a low variability in their MDI score. This lithological formation is mainly outcropping in the carboniferous basins, such as the Laval and the Châteaulin ones. These two basin comprise, in particular, the La Lucette and Rosnoën districts respectively, and are reputed for hosting fairly large antimony mineralization [[Marcoux and Chauris, 1994](#), [Pochon, 2017](#)].

Other predictors show lower MDI score. Although MDI scores evaluate the predictors, their SD ranges also show their reliability: the Neoproterozoic micaschists formation (MCSH2), despite showing the 4th higher score, shows a huge variability in its score, performing better in some trees than in others. This formation is widespread in the study area, and is present in some mineralized cells around the Quesnon district and near the Champtoceaux Nappe. As the cell samples used by each tree are randomized, it is possible that the MDI of this formation is highly variable depending on the training sample. Depending on the sample used by the decision trees produced by the RF model, this formation can either be very

good to classify the cells or underperforming compared to the other ones.

6.2.5 Comparison to other MPM methods

As a proof of concept, the new approach developed here is successfully applied to the example of Sb mineralization in the AM. As illustrated by the resulting favorability map (**Figure 6.6**) and the corresponding ROC curve (**Figure 6.7**), coupling CBA and RF methods allows estimating scores very consistent with all the known Sb mineralization and districts through the AM and with limited arbitrary user's decision making during the computation process. In order to compare the results of this CBA / RF approach with the other methods previously used, two supplementary predictive maps have been computed using WoE and CBA / Ranking approaches. The same dataset as the one presented in **Section 6.1.3, p.131** for the present study have been used to produce those predictive maps, in order to avoid any bias related to the data. The results of each method are displayed for comparison in **Figure 6.9**. The comparison relies on the estimation of their Prediction Rate (PR) and Percentage of Prospective Area (PPA) as shown in the Prediction Rate and Percentage of Prospective Area (PRPPA) graph shown in **Figure 6.9.a**. The PR is equivalent to the TPR shown in **Section 6.2.3, p.135** and evaluates the capacity of the model to correctly predict known mineralization occurrences. The PPA shows the percentage of the surface considered as favorable (i.e. TP and FP). An efficient model should minimize the area deemed as favorable (i.e. the PPA) while being able to correctly predict the known mineralization (i.e. the PR). As with the ROC curve, the model efficiency can here again be evaluated from the computation of the AUC.

For a PR threshold of 80% the WoE model (**Figure 6.9.b**), with an AUC of 0.67, results in more than 50% of the map being favorable (i.e. with score equivalent or superior to 0.5). For the same threshold, the CBA / Ranking method (**Figure 6.9.c**), with an AUC of 0.74, results in around 50% of the map being favorable, even though the corresponding curve clearly shows that the 80% PR threshold is almost reached at 40% PPA. Finally, the CBA / RF approach developed here (**Figure 6.9.d**), with an AUC of 0.88, highlights only $\approx 20\%$ of the total map surface as favorable for the same 80% PR threshold. This shows that the integrated CBA / RF method is clearly more effective than the WoE or the CBA / Ranking ones to delineate precisely favorable areas.

When compared to the WoE, the CBA / RF method, in addition to better performances according to the PRPPA curve analysis, is able to produce favorability scores in the shape of a raster grid covering the overall study area, without retaining the lithological boundaries, as the WoE method does. It thus produces more continuous information and favorability maps that are easier to handle and interpret. It is also less sensitive to the effect of clustering and positioning of mineralization. The favorability scores produced are easier to interpret, with a computed value ranging between 0 and 1, while the WoE requires some post-processing to correctly assess the meaning of each score. A WoE score above 1 means that this formation contains statistically more mineralization than the others, but some formation can score way higher than others (up to 25 in **Figure 6.9.b**) due to the bias induced by the outcropping surface of the formation.

The CBA / RF also performs better than the CBA / Ranking method, with a better AUC. Their respective favorability maps highlight comparable zones, with prospective zones mainly located along the SASZ and the NASZ, as well as the areas showing similarities with the Carboniferous basins of Laval and Châteaulin. However, the CBA / RF is significantly more effective to delineate the Semnon district and isolated mineralization located outside of the main known districts. As for the WoE, application of the

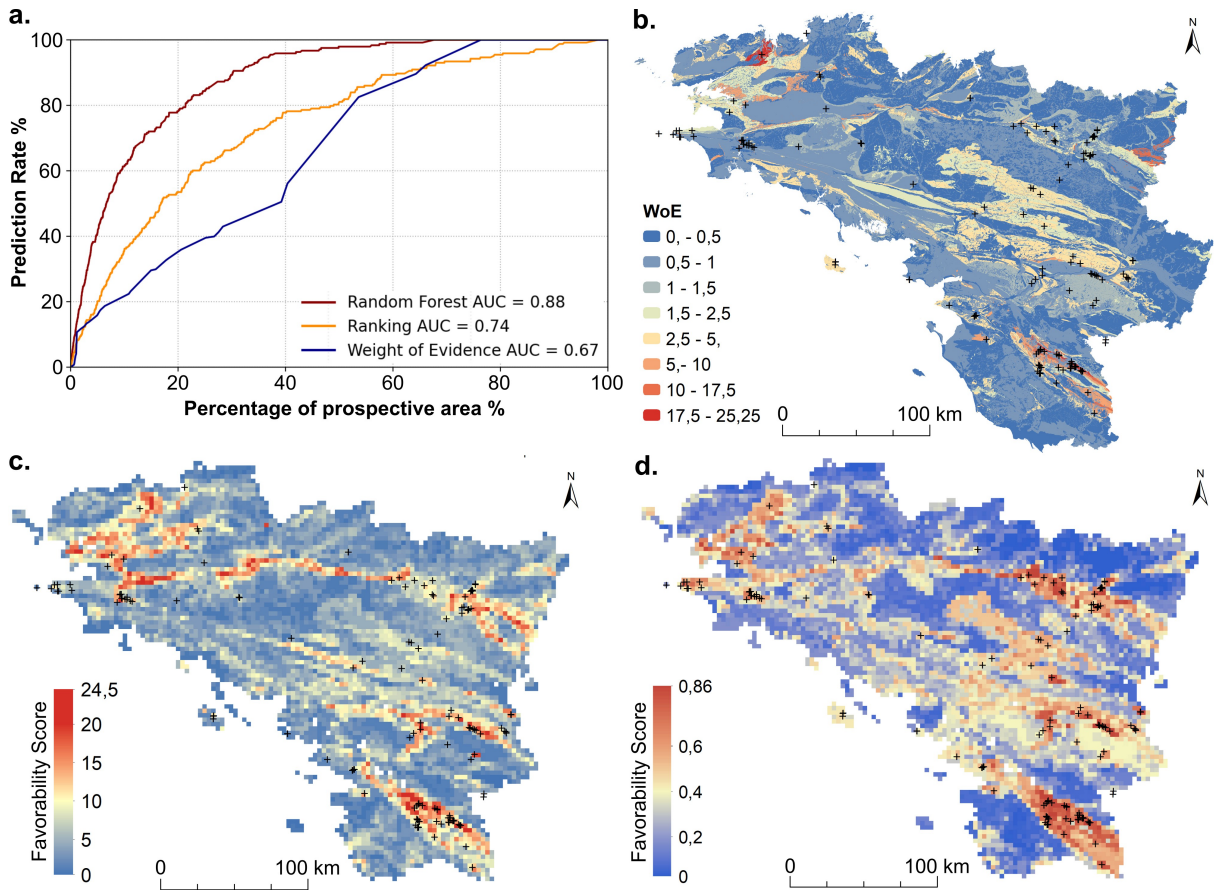


Figure 6.9: Application and comparison of various methods for prospectivity mapping of antimony in the Armorican Massif. a. Predictive Rate and Percentage of Prospective Area (PRPPA) curve of the different predictive models b. Favorability map of the WoE model based on all harmonized lithologies c. Favorability map of the CBA / Ranking model based on Paleozoic and pre-Paleozoic lithological formations d. Favorability map of the CBA / RF model, using the same CBA matrix as the Ranking method. Black crosses are antimony mineralization used to perform the prospectivity mapping method.

Ranking method results in favorability scores that need some post-processing. Indeed, if the clustering effect is ignored, thanks to the CBA matrix, the bias induced by the relative surfaces of the geological formations remains, as lithological formation contained in few cells but associated with mineralization will have a higher score than larger formations. Such bias does not appear when using the RF method, where the relative surfaces of lithologies and other predictors are not considered. Instead, the RF uses decision trees and impurity measurement (e.g. Gini impurity) to evaluate the capacity of the predictors to discriminate non-mineralized and mineralized cells during the classification process. These results show that the processing of CBA matrix with RF machine learning method significantly improves the predictive maps produced when compared to its association with statistical methods such as Ranking.

Chapter 7

Prospectivity mapping and environment analyses within the Ibero-Armorican Arc

*A complete application of the methodology developed during this thesis is displayed in this **Chapter 7**, using DBA, RF and HCPC as described in **Chapter 3**, **Chapter 4** and **Chapter 5**. The previous **Chapter 6** focused on the application of CBA and RF to produce favorability maps of antimony in the Armorican Massif (AM). In this **Chapter 7**, a similar dataset is used to apply the DBA / RF / HCPC to produce favorability maps of antimony in the AM, at a scale of 1:50 000. Another application of this methodology is detailed, with the production of antimony favorability maps of the Central Iberian Zone (CIZ) s.l. at a scale of 1:1 000 000. The AM application is the first to be presented, followed by the case study on the CIZ. Processing of the data is described chronologically, beginning with a description of the data used for each study case, followed by the application of the DBA to identify geological environments. Then, the favorability scores quantifying the relationships between geological environments and mineralization are computed using the RF predictive model. Finally the HCPC is applied to identify the "key" mineralized environments. Resulting "key" environments for both study cases are compared to existing knowledge about the geological environment and the metallogenic processes leading to the formation of antimony mineralization in those areas, as presented in **Chapter 2**. Results of the applications detailed in this **Chapter 7**, and in a lesser extent, previous **Chapter 6**, are further investigated in the following **Chapter 8**.*

7.1 Antimony MPM in the Armorican Massif

7.1.1 Context

The previous chapter highlights the benefits of using the CBA / RF method compared to previously used MPM methods, with the RF being the only new feature added to the CBA method. In this chapter, we test the application of the full new methodology by applying the DBA, RF and HCPC on a similar study of antimony on the AM. First, the DBA matrix is produced from a dataset comprising lithological,

structural and metallogenic data. The RF method is then used to compute the favorability scores map, which will be compared to the results of the previous methods. Finally, the HCPC clustering method will allow pointing out the relationships between the favorable areas produced by the RF predictive model and the known mineralization and metallogenic models of antimony in this area. Use of this comprehensive approach could automatically allow highlighting new metallotects for given commodities in a given geological domain. The developed DBA / RF / HCPC methodology aims at producing a "statistical" conceptual model, based on an automated Data-driven analysis of the spatial data. No conceptual models are used beforehand to select the metallotects / predictors to use in the MPM process, as we want to avoid as much as possible human bias and favor a Data-driven approach. However, the resulting clusters produced at the end of this process will be analyzed in regard to the metallotects mentioned in **Chapter 2**, to check the consistency of the results. Like in **Chapter 6** above, most lithological and structural data are imputed as is in the DBA method, with the exception of the focus on Paleozoic units (i.e. recent sedimentary covers are ignored).

This DBA / RF / HCPC application is expected to produce better results when compared to the CBA / RF method. It will also increase the complexity and computation time required in comparison. It is necessary with this application to display the interest of the HCPC clustering method as a post-processing method, as it is able to analyze the dataset and automatically (or semi-automatically) give returns concerning the geological environments, the associations of geological variables located in the area deemed the most likely to be mineralized by the RF model. Associating the DBA, RF and HCPC methods is the only methodology able, in this case, to (i) identify the local geological environments (ii) estimate the likelihood of each association of predictors to host mineralization and (iii) discretize among the most favorable associations, various favorable geological environments.

7.1.2 Automation of the methodology

The methodology developed during this thesis requires numerous spatial and statistical manipulations and computations. This thesis aims at automating those processes, and as such, we rely on the use of Python to automate and streamline most of the processes, including: the creation of the DBA grid, the RF predictive modelling and the HCPC clustering method. Indeed, this programming language presents an extensive library, which is commonly used in science applications and is also compatible with the most used GIS software. In the near future, the developments and programs made during this thesis work are planned to be integrated within QGIS, an open-source GIS software, for the development of an integrated prospective mapping tool developed by the BRGM.

To automate the process of the DBA / RF / HCPC method, we created multiples Python algorithms. Those algorithms allow the production of the DBA matrix and grid, as presented in **Chapter 3**, using *Pandas* and *Geopandas* for the processing of georeferenced data, *Numpy* for advanced mathematical and scientific computations and *Matplotlib* for the generation of graphical outputs. A regular cell grid is first produced using the boundaries of the lithological dataset and a value d to set the interdistance between the points. From this regular grid, the cells centers are recovered to perform the neighbouring analysis of the geological environments for all input datasets (i.e. lithologies, structures and occurrences) and produce the matrix using a search radius R . Binary values are added in the matrix, cell by cell, based on the intersected lithologies, structures and occurrences, producing the binary spectrum defining the geological environment. The produced matrix is then exported in a shapefile format and can be displayed in any GIS software.

This DBA matrix is then processed using *Scikit-Learn* to perform the RF predictive modelling. This RF classification uses *Scikit-Learn Random Forest Classifier* and *Randomized Search Cross Validation* classes to produce the optimal model. The hyperparameters are randomized, and the scoring method for the RF model is set to the AUC score of the ROC curves. After the autotuning process and the definition of the optimal hyperparameters with the Randomized Search Cross Validation, the model is trained using the whole dataset to produce the favorability map quantifying the relationships between the mineralization and the environments identified by the DBA. Training with the whole dataset means that the model is optimized to predict to the known mineralization rather than generalizing and predicting possible unknown mineralization. The favorable areas are as such limited and outline mainly the known mineralization, which means that the cluster and the "statistical" conceptual model is indeed based on the known mineralization and will extrapolate as little as possible. The favorability scores produced by the RF predictive model for each cell are added to the DBA matrix.

The final step of the method, the HCPC clustering EDA method, is produced using *Prince* and *Scikit-Learn*. First, the G-Means of the ROC curve are computed. The threshold corresponding to the highest G-Mean is then saved and used to select only the cells showing a favorability score higher than this threshold. Those cells are then processed with *Prince* to produce the MCA factorial analysis. This analysis produces a new matrix of continuous coordinates for each cell. Those coordinates are the inputs used for the *Agglomerative Clustering* function of *Scikit-Learn*, which produces x clusters depending on the imputed value x . Multiple value of x are tested to determine the best balance between precision and generalization for the mineralized environments. The cluster attributed to each cell is stored in the matrix.

Initial and final georeferenced dataset are manipulated and processed using *ArcGIS*® (<https://www.esri.com/en-us/home>) for the selection of the dataset, pre-processing and for the edition and creation of the displayed favorability maps and cluster maps.

7.1.3 Input dataset

This case study uses similar datasets than the previous CBA / RF study (**Figure 7.1**). The lithological dataset used is the exhaustive and homogeneous digital 1:50,000 set of geological data of the AM produced by [Gloaguen and Tourlière \[2016b\]](#) from the CHARM harmonized database. Details about this harmonization can be found in **Appendix B.1, p.264**. In contrast to the previous case study, this one relies on the structural map features extracted from the same 1:50,000 CHARM harmonized dataset rather than from the 250K harmonization from [Chantraine et al. \[2001a\]](#) (unpublished), allowing to keep a consistent scale of all the geological features in this study. The sheets (or "départements") used for both the harmonized dataset and the structural features are n°85 [[Béchenec, 2009](#)], n°56 [[Cagnard, 2009a](#)], n°53 [[Vernhet, 2010](#)], n°49 (unpublished), n°44 [[Béchenec, 2007](#)], n°35 [[Cagnard, 2009b](#)], n°29 [[Cagnard, 2008](#)] and n°22 [[Egal, 2010](#)], corresponding to 128 sheets from the original 1:50,000 geological map of France.

The antimony mineralization dataset used for the predictive model is the same 141 occurrences over the AM extracted from the BRGM database [[Cassard et al., 2015](#)] and updated by [Gloaguen and Tourlière \[2016b\]](#), with mostly veins-type deposits. Likewise, the lithological formations have been filtered in order to use only Paleozoic and older formations to perform the analysis, as antimony mineralization are linked to Paleozoic events during the Variscan orogeny [[Chauris, 1977](#), [Marcoux and Chauris, 1994](#), [Munoz et al.,](#)

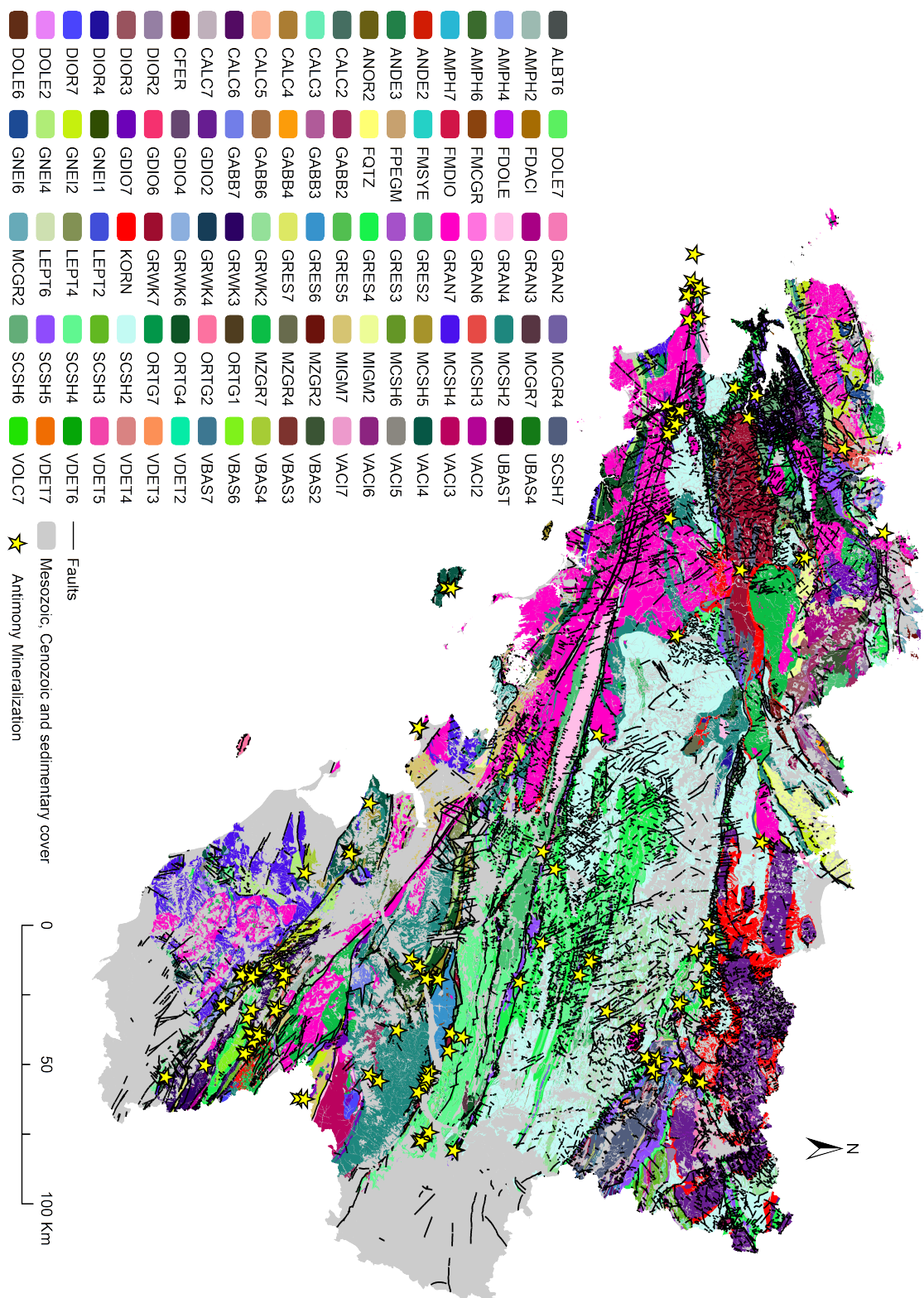


Figure 7.1: Maps of the datasets used for the DBA / RF antimony prospectivity mapping across the Armorican Massif. Palaeozoic and Precambrian formations (in colors, from [Gloaguen and Tournière \[2016b\]](#)): cover 65.35% of the studied area; Mesozoic and Cenozoic formations (in grey): cover 34.65% of the studied area. Faults are from the 50K CHARM dataset, mineralization from BRGM Database [[Cassard et al., 2015](#)].

1992, Pochon et al., 2019]. Selection of Paleozoic faults was not possible due to the lack on information relative to the age of the faults development. As the previous application, no geophysical or geochemical datasets are used in this application, and we rely solely on lithological and structural data to identify the geological environments.

7.1.4 DBA matrix

The DBA method is first applied on the harmonized 50K geological dataset of the AM as compiled in Gloaguen and Tourlière [2016a,b]. It results in a regular grid covering Palaeozoic and older formations of the AM. In order to find the optimal cell size, the DBA is iterated on the studied area with different values for the neighbouring search distances, from 2 000 m to 30 000 m with a 1 000 m step (**Figure 7.2.a** to **.d**) as presented in **Section 3.1.3, p.75**. As we use neighbourhood search discs, it is necessary to determine the value of both the cell size d and the disc radius R . Both values are linked with a unique ratio R/d .

The optimal neighboring search distance and R/d ratio is selected using the results produced by the iterations. The slope breaks of the curves showing cells count (**Figure 7.2.a**) and uni-lithological cells percentage (**Figure 7.2.d**) is considered to correspond to the optimal radius for the neighbouring search. This optimal distance allows to maximize the total cells number (and, hence, the resolution of the grid map) while minimizing the ratio of grid nodes corresponding to only one single geological unit in the local surrounding. The optimal values found for the lithological dataset used are a R/d ratio of 3 and a R search radius of 3 000 m. This R value insures a balance between resolution and diversity of the lithological associations. The R/d ratio of 3 allows a high number of lithological environments for this R value compared to lower ratio, while avoiding the high number of cells to process compared to a higher ratio. Higher R values have been considered, with a test performed with a R of 5 000 m and a R/d of 4, values that are also located along the slope breaks of the cells count and uni-lithological cells percentage curves. Application of the DBA / RF on those values displayed similar results to the ones mentioned above, with similar AUC scores for the ROC curves and the same highlighted areas on the map, at the cost of a coarser resolution. As such, the R value of 3 000 m and R/d of 3 are retained for the rest of this application.

With a R/d of 3 and a neighbouring search distance of 3 000 m, the cell size will be of 1 000 m, the *Dispersion* of 21.4%, with a *Range* of 6, a minimum of 26 overlapping, a maximum of 32 and an average of 28.3 overlapping. Besides, only 3% of the cells (1 477) show a unique lithology inside. In total, this DBA grid produces 48 871 cells and 9 529 unique environments. On average 7.5 lithologies lie in a cell with a maximum of 21 for a single cell.

After processing the lithologies, mineralization and faults are added to the CBA matrix. This process is done in this order as we must first consider the lithologies to define the R/d and R values from which the matrix is produced. The search radius for the neighbouring search of the faults and mineralization is fixed to 1 000 m around each node, resulting in 30 335 (i.e. 62%) cells containing faults. A lower search radius is used than for the lithologies as faults are widespread across the study area: with a search radius of 3 000 m (the same as the one used to search the lithologies) it would result in 90% of the cells (44 293) which would cross faults. Reducing the search radius allows to avoid an over-representation of the fault in the matrix and to avoid inducing noise. With this neighbouring search of 1 000 m, 321 cells 0.67% of the cells (321 cells) contain mineralization.

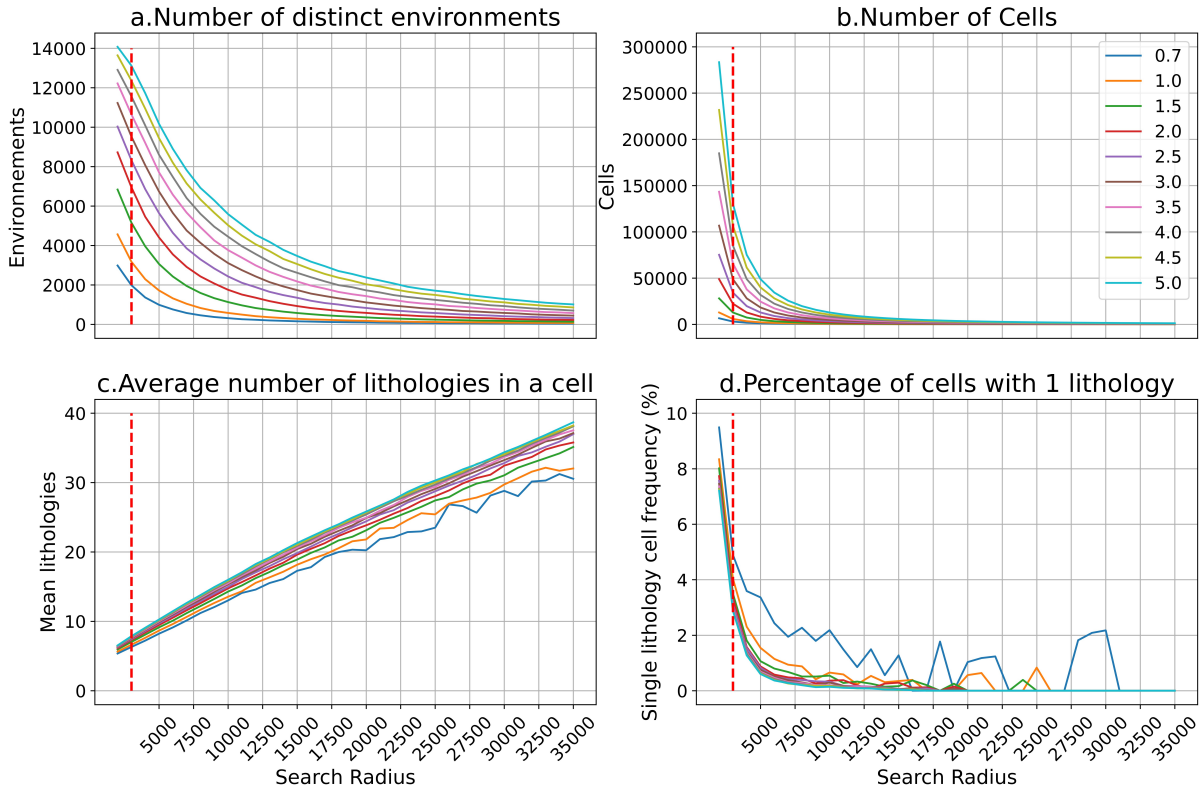


Figure 7.2: Application of the multi-scale iterating process of DBA matrices on the lithological dataset of the Armorican Massif for various values of R and R/d . The optimal value found for the disc size is 3 000 m for a R/d ratio of 3, showed with the red dashed line. a. Number of lithological associations b. Number of cells. c. Average lithologies per cell d. Frequency of cells containing only one lithology.

7.1.5 DBA / RF predictive model

The RF predictive modeling is applied on the DBA matrix. A binary classification is made on all of the cells of this matrix: all the non-mineralized cells are used as non-occurrences of antimony and all the mineralized ones are set as occurrences of antimony. The optimal hyperparameters are defined using *Random Search* with *Scikit-Learn* based on 500 iterations and a 5 fold Stratified Cross-Validation (SCV). As an example, the optimal hyperparameters found determine that the model is the most performing with 659 trees with a maximum depth of 71.

The final predictive model is computed to the whole of the matrix using those optimal hyperparameters to define the favorability score of each cell, i.e. the most likely to host antimony mineralization. Results are shown in **Figure 7.3**: cells showing the higher favorability scores are displayed in green; they correspond to the most likely to be mineralized and sharing the most features with the known mineralized cells. On the other hand, the cells that are the less likely to host antimony mineralization and sharing the least or no features with the known mineralized cells are displayed in deep purple.

The most favorable cells have a score of 0.973, while the less favorable cells have a score of 0. For the mineralized cells, the average favorability score is 0.81 with a Standard Deviation (SD) of 0.14, well above the 0.5 threshold. Some mineralized cells, however, have a favorability score below this threshold, with 9 cells having a score between 0.5 and a minimum value of 0.31. If we look at the non-mineralized cells, the average favorability score is 0.11 with a SD of 0.18 and values ranging between 0 and 0.97. The

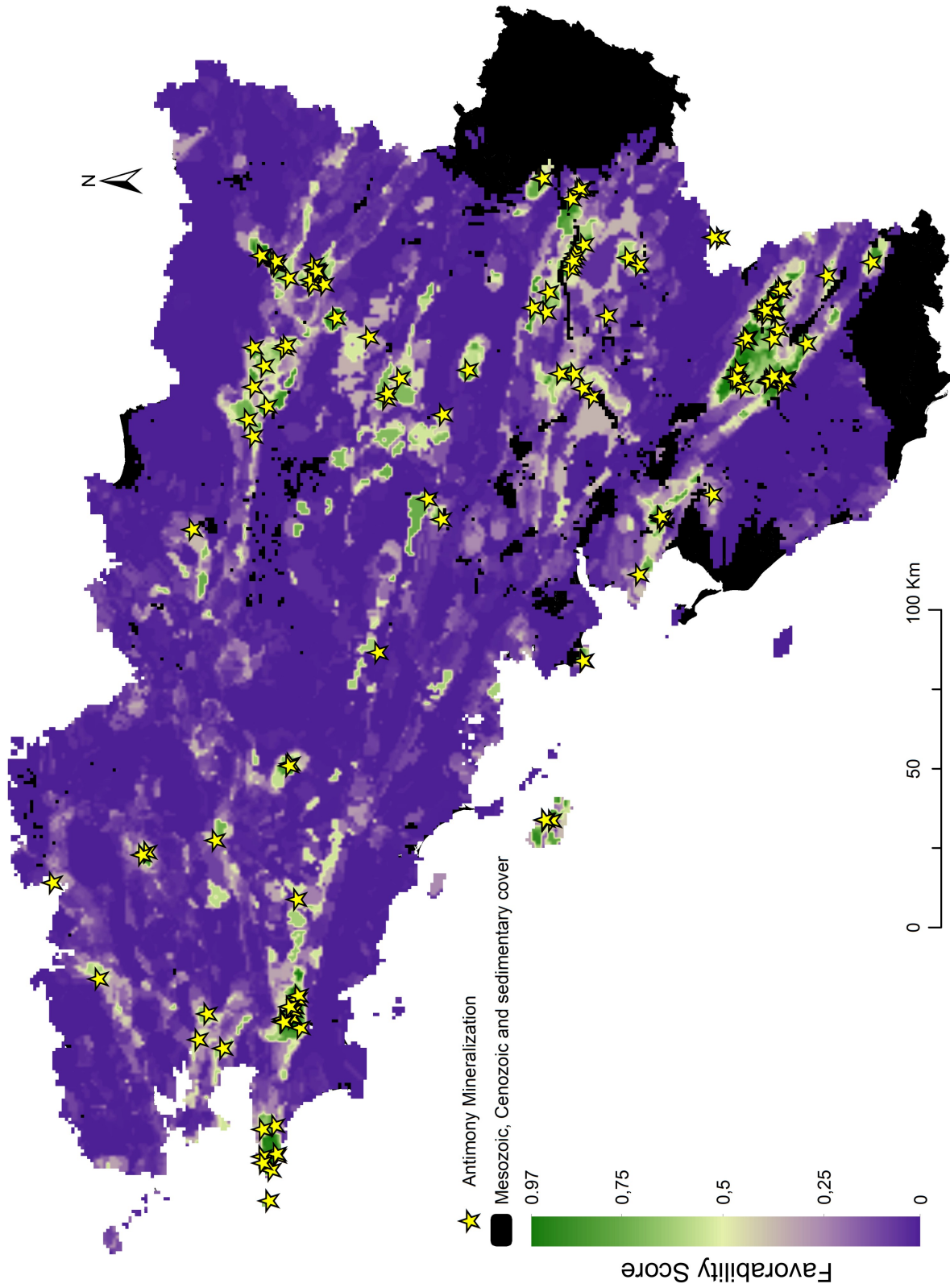


Figure 7.3: Favorability map of antimony in the Armorican Massif produced from the integrated DBA / RF method. With a neighbouring search R of 3 000 m and a R/d of 3, the resolution of the grid is 1 000 m. Mineralization and faults are added to the matrix with a neighbouring search R of 1 000 m. High favorability score are displayed in green, corresponding to a higher likelihood for the cell to contain antimony mineralization.

high average score for the mineralized cells shows that the predictive model is able to correctly predict most of them and give higher scores to known mineralized areas, while the low average score for the non-mineralized cells shows that most of them are indeed predicted as non-mineralized.

A deeper evaluation of the resulting favorability scores is established using the ROC curve. With this curve, one can evaluate the ability of the model to correctly predict the mineralized cells (True Positive Rate; TPR) and the number of non-mineralized cells predicted as mineralized (False Positive Rate; FPR) depending on the threshold value. The first ROC curve (**Figure 7.4**) is established using 5-folds SCV on the optimized predictive model and shows the variability of the prediction of the model when applied on various sets of cells. The shape of the curves varies between the training set, with AUC scoring between 0.96 and 0.93, for an average value of 0.94 ± 1 .

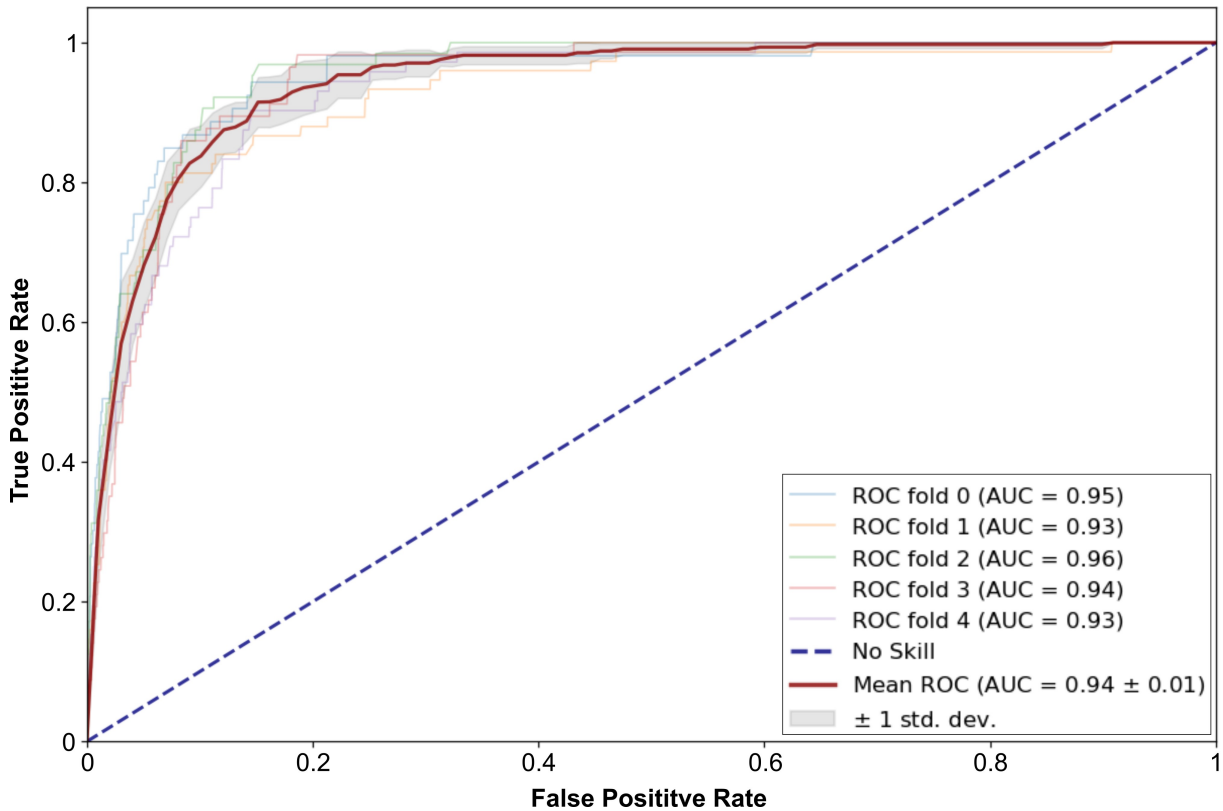


Figure 7.4: Receiver Operating Characteristic (ROC) curve from the 5-folds Stratified Cross Validation of the DBA / RF predictive model of the Armorican Massif. Each fold is displayed with its corresponding Area Under the Curve (AUC) score. The no-skill model displayed in dotted blue corresponds to a no-skill predictive model.

The second ROC curve (**Figure 7.5**) is obtained by computing the predictive model on the whole dataset, without Cross-Validation samplings, resulting in the associated favorability map of **Figure 7.3**. All the mineralized and non-mineralized cells are used as training data for the RF classification process, resulting in a ROC curve showing a higher AUC score of 0.98. Higher AUC is to be expected, as this model only relies on the stratified bootstrapping sampling method of the RF to avoid overfitting and does not use the SCV method in addition. If this model is more likely to overfit the data, it is also able to restrain the area that will be analyzed regarding the geological environments surrounding the known mineralization. Fewer cells will be highlighted using this model, but most of the known mineralization can be favorably classified with less "noise", less False Positives (FP), making the identification and analysis of the geological environment easier. For prospective purposes, the favorable area that could be targeted

will be more restrained, which can be advantageous when trying to delineate specific targets rather than searching for larger prospective areas to further investigate.

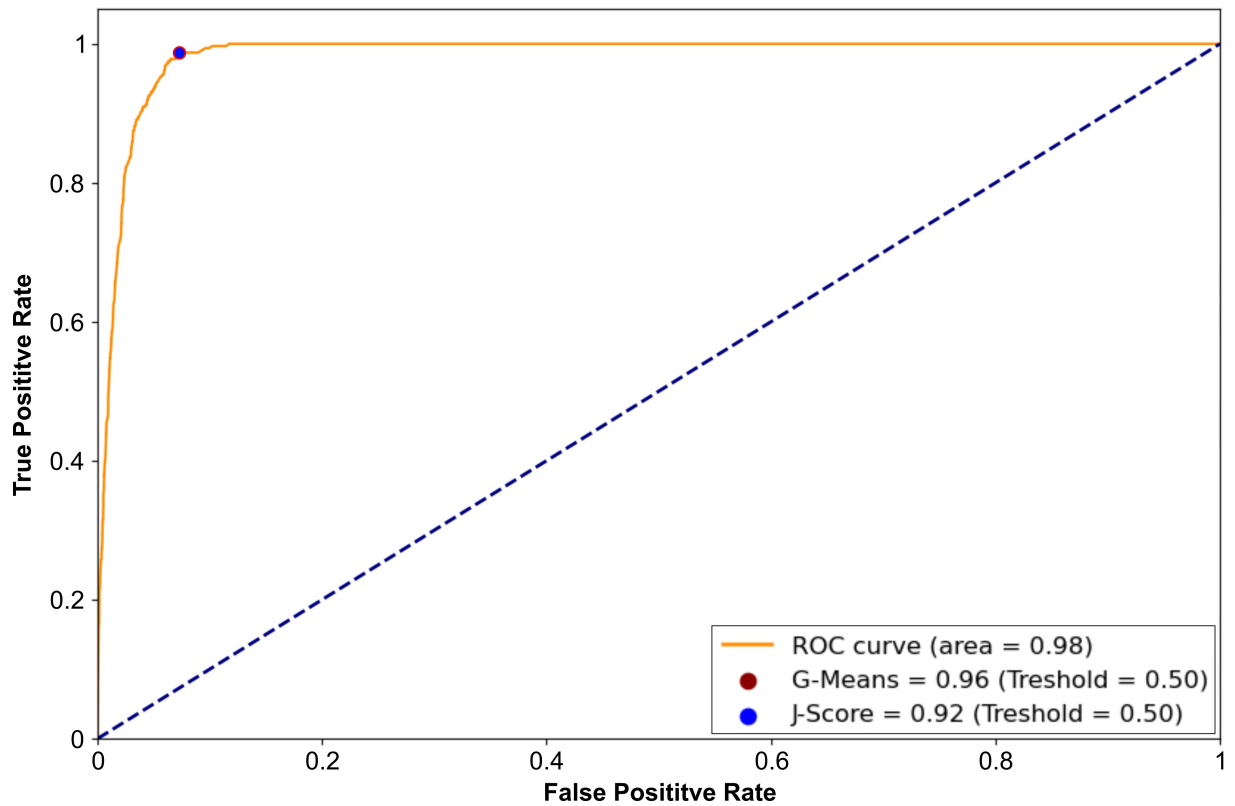


Figure 7.5: Receiver Operating Characteristic (ROC) curve from the final DBA / RF predictive model of the Armorican Massif. All cells from the DBA matrix are used to train the predictive model. The no-skill model is displayed in dotted blue.

7.1.6 Identification of favorable geological environments

With the DBA / RF method, we calculated the local geological environments and assigned them favorability scores. Now, from those results, we intend to identify and highlight the possible geological environments associated with the most favorable areas, that could serve as potential guide for exploration. To identify the most favorable cells, we use the previous ROC curve. From this ROC curve, the Geometric Means (G-means) and the Youden's J statistic (J-Score) are computed to establish the threshold to use. This threshold determines which cells will be considered as favorable regarding the predictive model, with cells having higher or equal scores than the threshold being considered as favorable. The computed G-Means and J-Score are displayed on the ROC curve (**Figure 7.5**), both pointing to a threshold value of exactly 0.506.

Selected favorable cells (e.g. cells with favorability score ≥ 0.506) are displayed in **Figure 7.6.a**. Associated confusion matrix (**Figure 7.6.b**) shows the resulting classes of this predictive model, depending on initial and predicted classes. Selected favorable cells correspond to the False Positive (FP) and the True Positive (TP) showed on the right of this matrix. This selection result in 3 346 cells, 312 TP (97% of mineralized cells) and 3 034 FP (6.2% of non-mineralized cells). This selection contains 661 unique lithological associations.

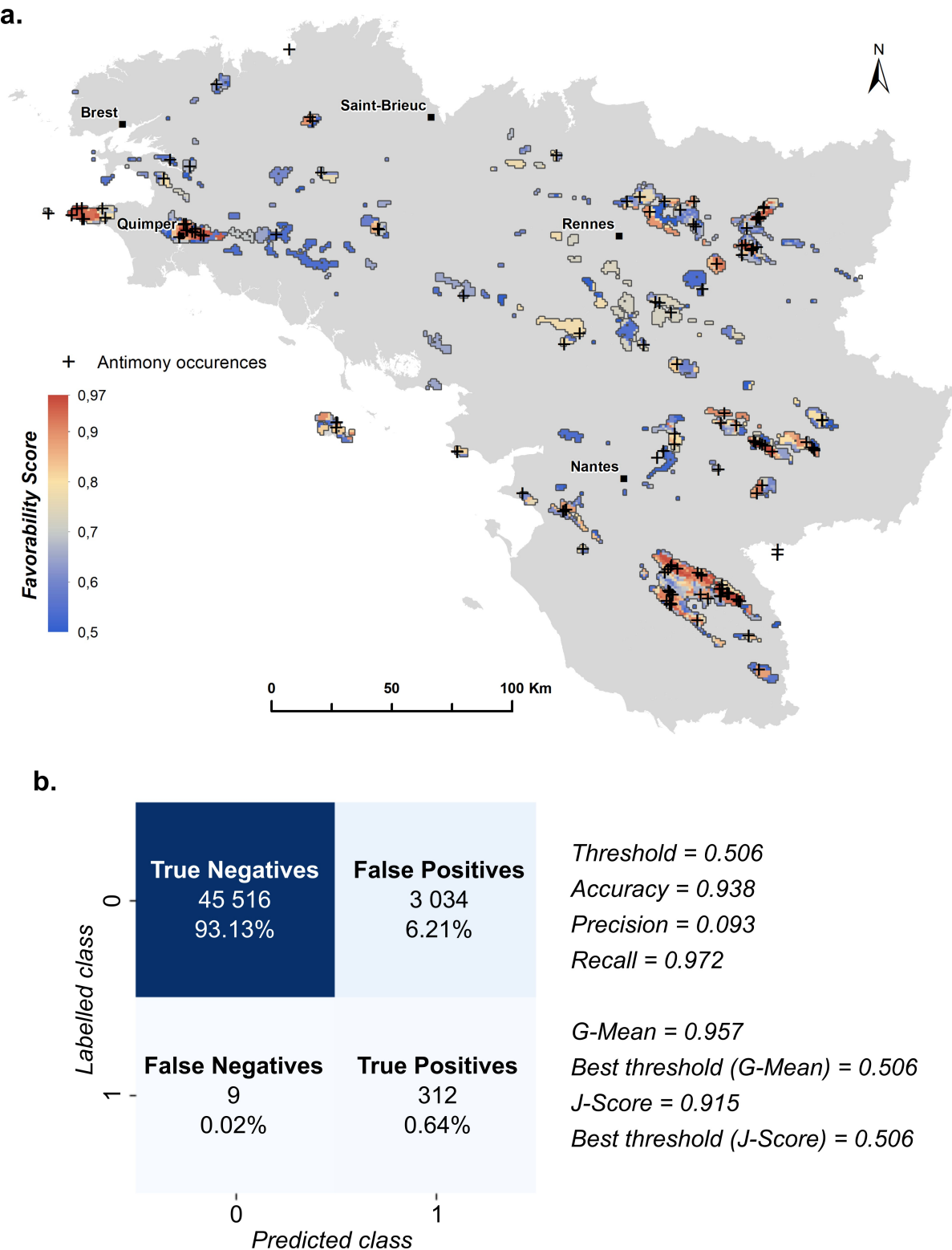


Figure 7.6: Identification of the favorable areas of the favorability map of antimony in the Armorican Massif based on the G-Means and J-Score best threshold. a. Favorability map of cells displaying a score higher than 0.506, the color ramp is scaled for score higher than 0.506 b. Associated confusion matrix of the classification of the cells based on a 0.506 threshold for the favorability score.

The HCPC clustering method is applied on the selected cells to reduce those numerous lithological associations into a more reduced number of clusters, based on the likelihood between the lithological associations. First, a factorial analysis is performed on the selected cells, using the Multiple Correspondence Analysis (MCA), to project the selected cells into new factorial axes. Then, cells are clustered together using the Agglomerative Hierarchical Clustering (AHC) method applied to those new factorial axes coordinates. The agglomerative function used here is Ward's method, while the distance used is euclidean.

Selection of the number of clusters from the HCPC is done here manually using the resulting dendrogram and inter-cluster inertia gain to select a "cut-off" value. Multiple cut-off values are selected and displayed in a map, the cluster's identification associated to each cell being added to the matrix. The "best" cut-off value can be determined by analyzing the resulting clusters: the distribution of the cells between the various clusters should result in a few clusters, each of them presenting a number of individuals in the same order of magnitude, to identify "global" trends in the geological environment across the favorable areas. It is better to avoid a high number of clusters at this stage, as we want the most comprehensive analysis possible. In this natural case study, we determine that the best cut-off value is 4 clusters. If necessary, the identified geological environments can be clustered again using a larger cut-off value to produce more clusters and analyze the geological environments in more restrained areas, or identify sub-classes within the more "global" clusters.

The resulting 4 clusters are displayed in **Figure 7.7**. For each cluster, one can analyze the geological environments they contain: the frequency of lithologies and faults within each cluster can be computed and compared to the other clusters, or alternatively to their respective frequency across the favorable areas or the studied area. This comparison allows identifying specific lithologies belonging to a given cluster, lithologies that are more frequent in this cluster than in other areas, or that are only present in this cluster. Differences between clusters and the overall studied area can highlight features or associations of features that are likely to be hosting mineralization, while differences between each cluster and between a cluster and the favorable cells highlight the specific signature of each cluster.

7.1.7 Favorable areas and best predictors

From the RF predictive model, one can evaluate the effectiveness of each predictor used by the decision trees based on the Mean Decrease in Impurity (MDI) score, evaluating the tendency of each predictor to be used during the classification process. The higher the MDI score is, the more frequently the predictor is used in the Decision Trees, and the more important a predictor is for the classification process. **Figure 7.8.a** displays the 20 most important predictors for this DBA / RF predictive model, with their MDI score and corresponding SD value.

The Cambrian Greywackes (GRWK3; **Figure 7.8.b**) and Carboniferous granites (GRAN7; **Figure 7.8.c**) are the most important predictors for this model, showing the highest MDI score with low SD. The Cambrian Greywackes are located on a limited area on the South of the AM, along the Vendée antimony district, with 1 485 cells containing it, 58 of them being mineralized. Cells containing this lithology have an average favorability score of 0.37 with a SD of 0.32.

On the other hand, the Carboniferous granites cover a widespread area across the SAD, the West of the CAD and the Leon Bloc, with 16 317 cells, 79 of them being mineralized. The most notable antimony mineralization related to this formation are located in the Quimper district, to the West of the AM. Cells

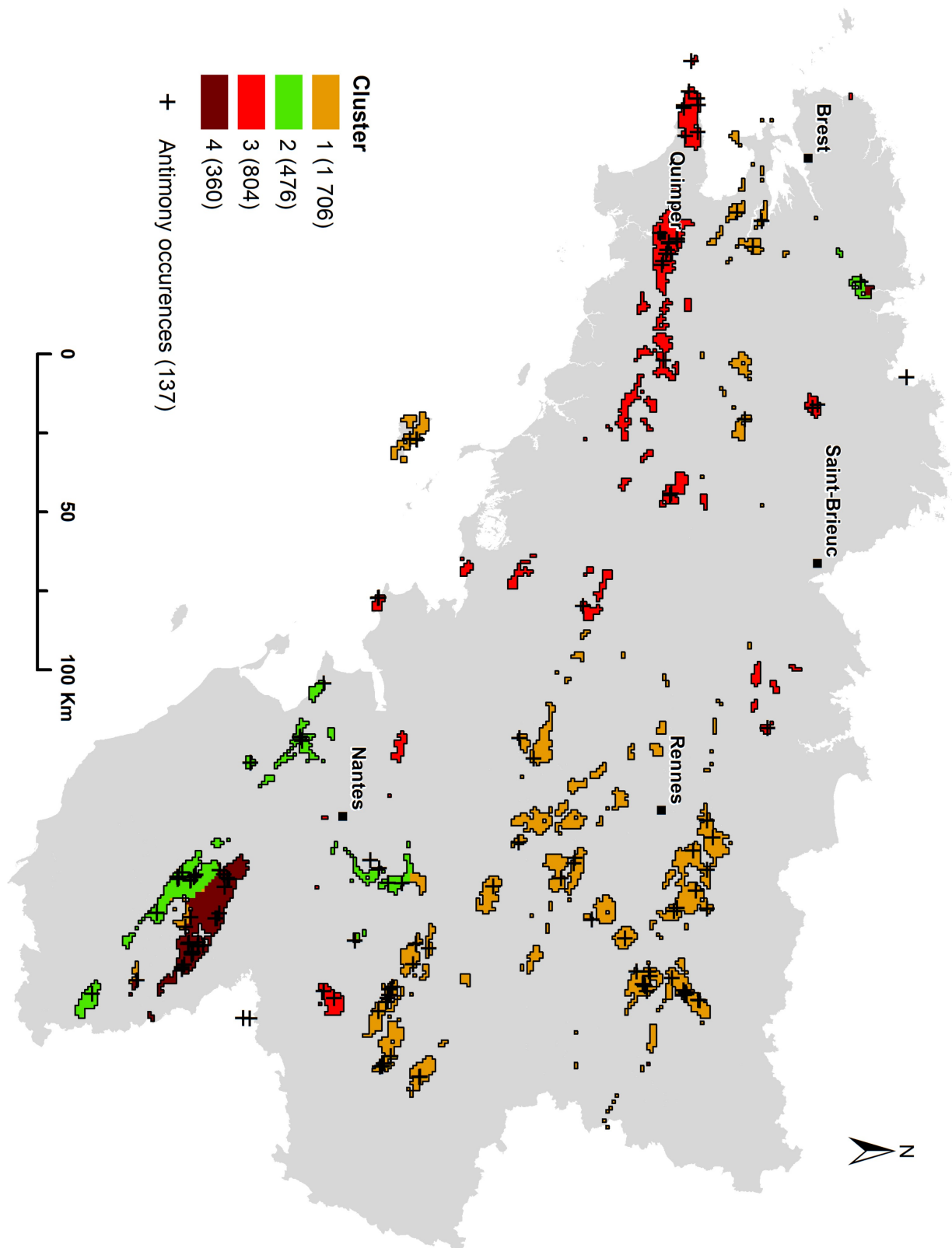
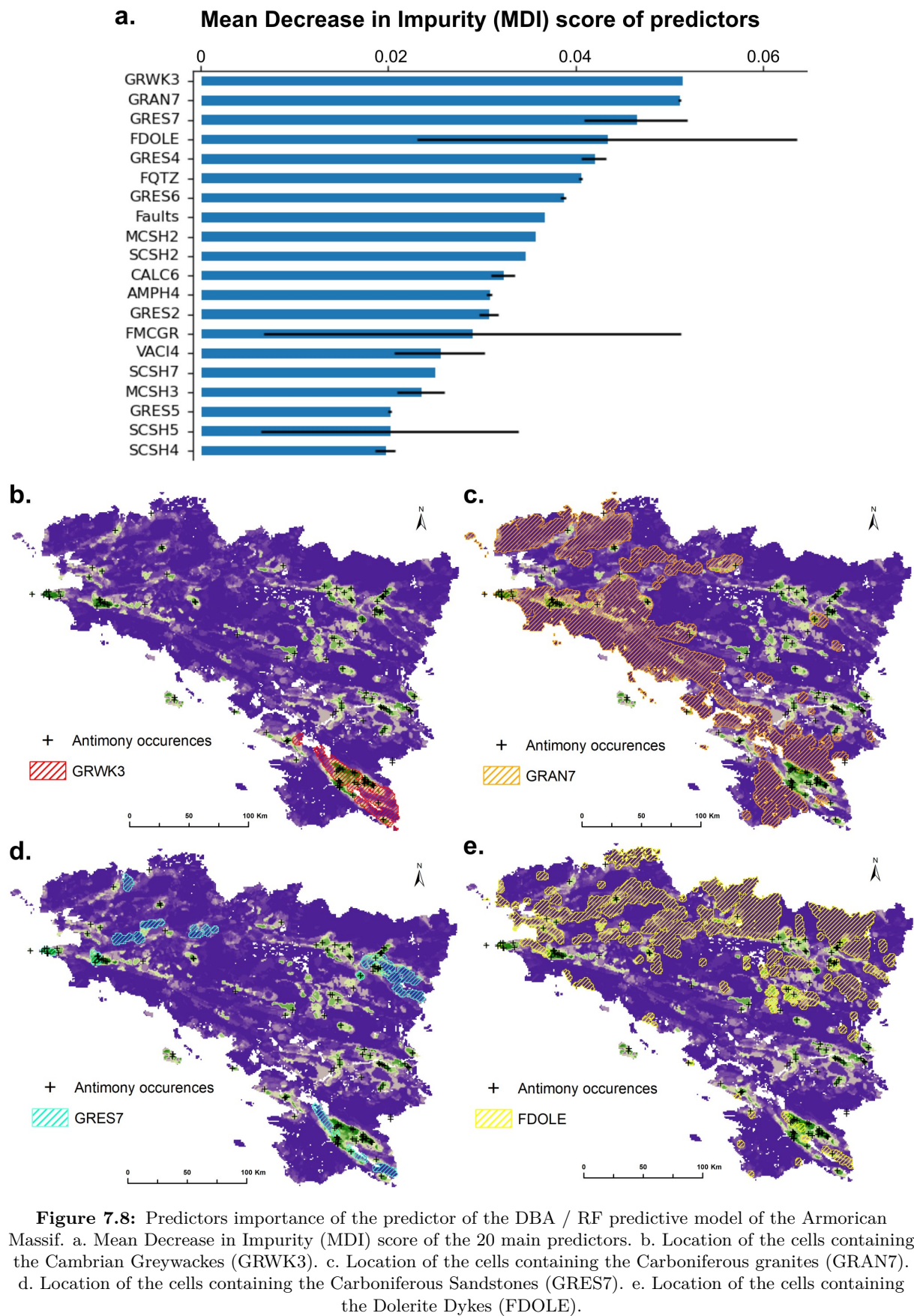


Figure 7.7: Map of the 4 clusters produced by the Hierarchical Clustering on Principal Component classification (HCPC) of the favorable cells (score ≥ 0.506) of the Armorican Massif DBA / RF favorability map. The number of cells is displayed for each cluster, as well as the number of antimony occurrences across the study area.



containing this lithology have an average favorability score of 0.08 with a SD of 0.16, as it covers mostly areas with low favorability scores.

The Carboniferous sandstones (GRES7; **Figure 7.8.d**) are the third most important predictors in the predictive model, with a higher SD compared to previously shown predictors. This lithology is located in various areas across the AM and cover a small area, with 1 989 cells, 44 being mineralized. This lithology unit has an average favorability score of 0.20 with a SD of 0.27.

The Dolerite Dykes (FDOL; **Figure 7.8.e**) are the fourth most important predictors for this predictive model with a very high SD. This high SD shows that its usage in the classification process of the decision trees varies widely, possibly depending on the bootstrap selection process. It could also mean that the importance of this predictor is highly dependant to the mineralized cells selected by the bootstrap sample, or that the number of trees produced was not high enough for this predictor. This lithology is located mostly in the NAD and CAD, covering a large area with 12 020 cells, 71 being mineralized. Cells containing this lithology have an average favorability score of 0.10 with a SD of 0.18.

Despite those four examples being the best predictors for the predictive model developed here, they show different relationships with the mineralization and favorable cells. While the Cambrian Greywackes and the Carboniferous sandstones are "rare" lithologies, covering only restrained parts of the studied area and showing relatively high average favorability score, the Carboniferous Granites and Dolerites Dykes are widespread lithologies with low favorability scores in comparison. It is easier to define the first two (GRWK3 and GRES7) as possible "metallotects" to be extracted from this model, as they allow targeting more specific area. However, this analysis is restrained to one lithology at a time, even if the Carboniferous Granites and Dolerites Dykes covers large areas, their association with other lithologies could be an indicator for the presence of antimony mineralization.

7.1.8 Analysis of the clusters

The presentation of the clusters is done accordingly to their relative size, with first the cluster 1, then the cluster 3 and finally cluster 2 and 4.

Cluster 1

The cluster 1 is the biggest cluster produced by the HCPC, containing 1 706 cells, 50.99% of the favorable area. It contains 134 or 42.9% of the mineralized cells of the favorable area, 7.9% of its cells being mineralized. The average favorability score inside this cluster is 0.68 ± 0.12 . A detailed map of the favorability score inside this cluster and the antimony occurrences and deposits within it is displayed in **Figure 7.9**.

The lithologies inside this cluster are dominated by Ordovician sandstones (GRES4) and schists (SCSH4), as more than 74% of its cells contain those lithologies. They are also specific to this cluster, as more than 78% of the favorable cells containing those formations are localized within this cluster. Faults are widespread in this cluster, as they cross 84% of the cells, this cluster containing 53% of the cells including faults. Other notable formations are Silurian schists (SCHS5), Devonian sandstones (GRES6) and Devonian limestones (CALC6), which are contained in more than 40% of the cells for the 2 first, and more than 30% for the last. They are present almost exclusively in this cluster. This cluster also contains a large amount of dolerites dykes (FDOL) and quartz dykes (FQTZ), with 28.7% and 40.4% of the cells containing such geological map features respectively. This cluster is the one containing most

draft

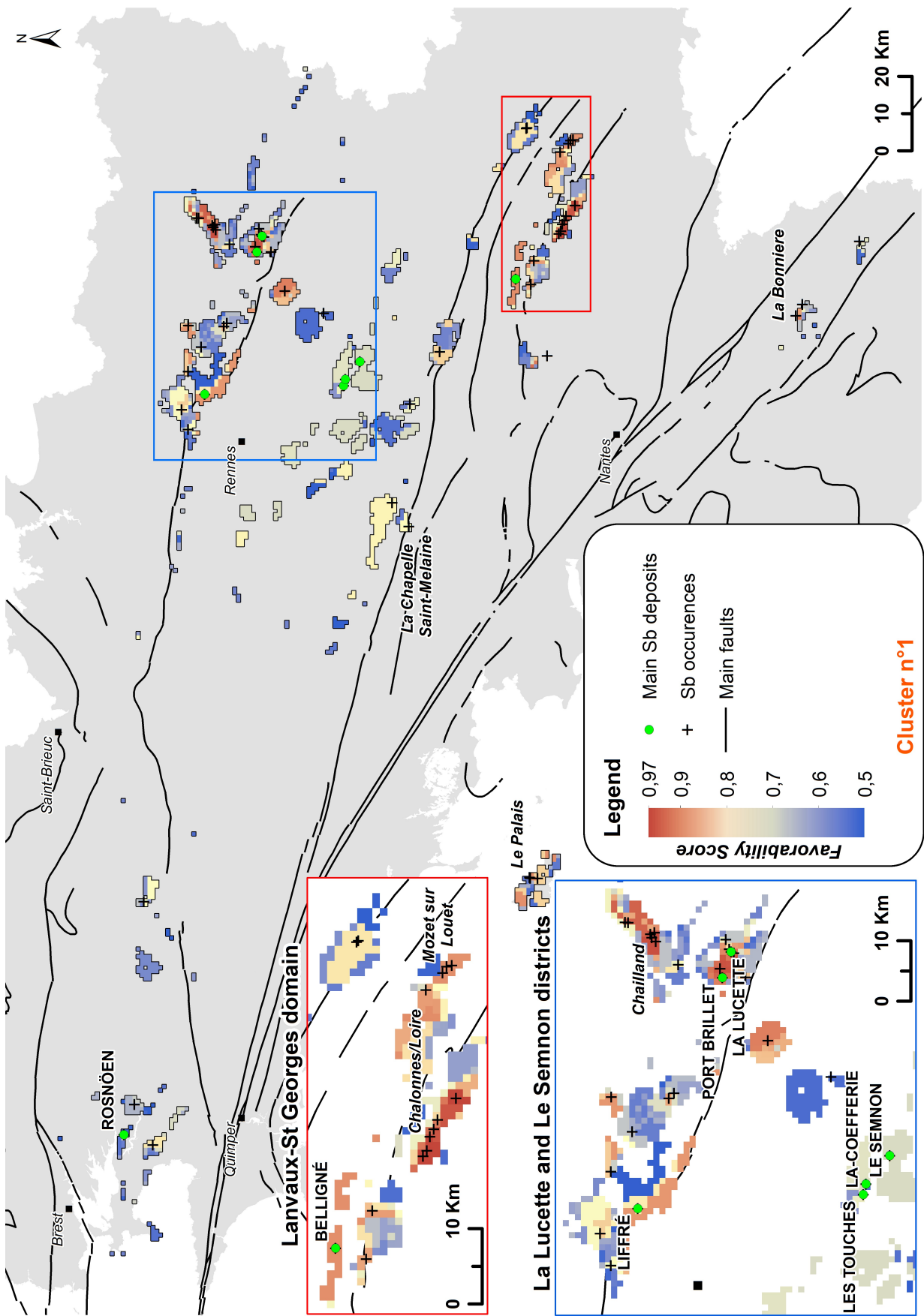


Figure 7.9: Favorability map of antimony in the cluster 1 of the AM.

of the dolerite and quartz dykes, containing 64% and 43% of all cells with those lithologies. More details about the frequency of lithologies in this cluster are available in **Appendix C.2**. From this global analysis, one can see that this cluster is clearly dominated by sedimentary formations of various ages (Ordovician to Devonian with few Carboniferous and Neoproterozoic units), mainly of Ordovician age (Ordovician sandstones and schists), associated with dolerites and quartz veins.

This cluster contains some of the most important antimony districts [Marcoux and Chauris, 1994, Gloaguen and Tourlière, 2016b] in the AM, with the Lucette district [Guiollard, 2009, Moyroud et al., 1979, Serment, 1978], the Semnon-La Coëfferie district [BRGM, 1987, Chauris et al., 1985, Gloaguen and Tourlière, 2016b, Moyroud et al., 1979] and the Rosnoën deposit [BRGM, 1988b, Gentilhomme et al., 2001]. The Lucette district is located inside the Laval Basin, along the NASZ and is mainly made of the Lucette deposit. The antimony mineralization is contained in quartz veins hosted in sandstones and metapelites from Ordovician to Silurian ages [Marcoux and Chauris, 1994, Serment, 1978]. The Semnon-La Coëfferie deposit is located along the Rennes syncline, the antimony being located inside quartz veins hosted in dolerites dykes and Ordovician schists [Chauris et al., 1985]. The Rosnoën deposit is a recently discovered stratiform deposit located in the Devonian limestones of the Châteaulin Basin, West of the CAD.

The highest favorability scores inside this cluster are located around the Lucette district and the Chailland mineralization in the Laval Basin. High favorability scores are also displayed in the Lanvaux-St Georges domain, with Sb-Hg mineralization like Belligné [BRGM, 1983], Pb-Sb mineralization like Chalonnès sur Loire or orogenic gold mineralization with antimony like Mozet sur Louet. Finally, high favorability scores are also displayed near the Palais mineralization, SW of the AM. Besides, the Semnon and Rosnoën districts show relatively low favorability scores in comparison.

This cluster groups various mineralization, the main common point between them being the predominance of sedimentary formations in this environment, with the grouping of various mineralization located within sedimentary stratified units such as the Laval and Châteaulin Basin and the Rennes syncline. Dolerite dykes and quartz veins make also a major part of the geological environment, which comes as no surprise as they are the main hosting formations for the various antimony districts, as described in the literature. Map-scale structures are also predominant, as most of the favorable areas are located near major shear zones or in the vicinity of faults swarms (e.g. Le Semnon district).

Cluster 3

The cluster 3 is the second biggest cluster produced by the HCPC, containing 804 cells, 24.03% of the favorable area. It contains 84 or 26.9% of the mineralized cells of the favorable area, 10.4% of its cells being mineralized. The average favorability score inside this cluster is 0.68 ± 0.15 . A detailed map of the favorability scores and the antimony occurrences and deposits within this cluster is displayed in **Figure 7.10**.

Carboniferous granites (GRAN7) and Neoproterozoic micaschists and schists (MCSH2 and SCSH2) are the most predominant lithological formations in this cluster with frequency of 88%, 82% and 65% respectively. Carboniferous granites and Neoproterozoic micaschists are also overrepresented in this cluster as it contains more than 70% of the cells containing those lithologies. Quartz veins (FQTZ) are present in more than half of the cells of this cluster, this cluster contributing to 27% of this lithology. Ordovician granites (GRAN4) and Carboniferous migmatites (MIGM7) account for more than 20% of the cells and are almost exclusively contained in this cluster. This cluster also contain few Neoproterozoic

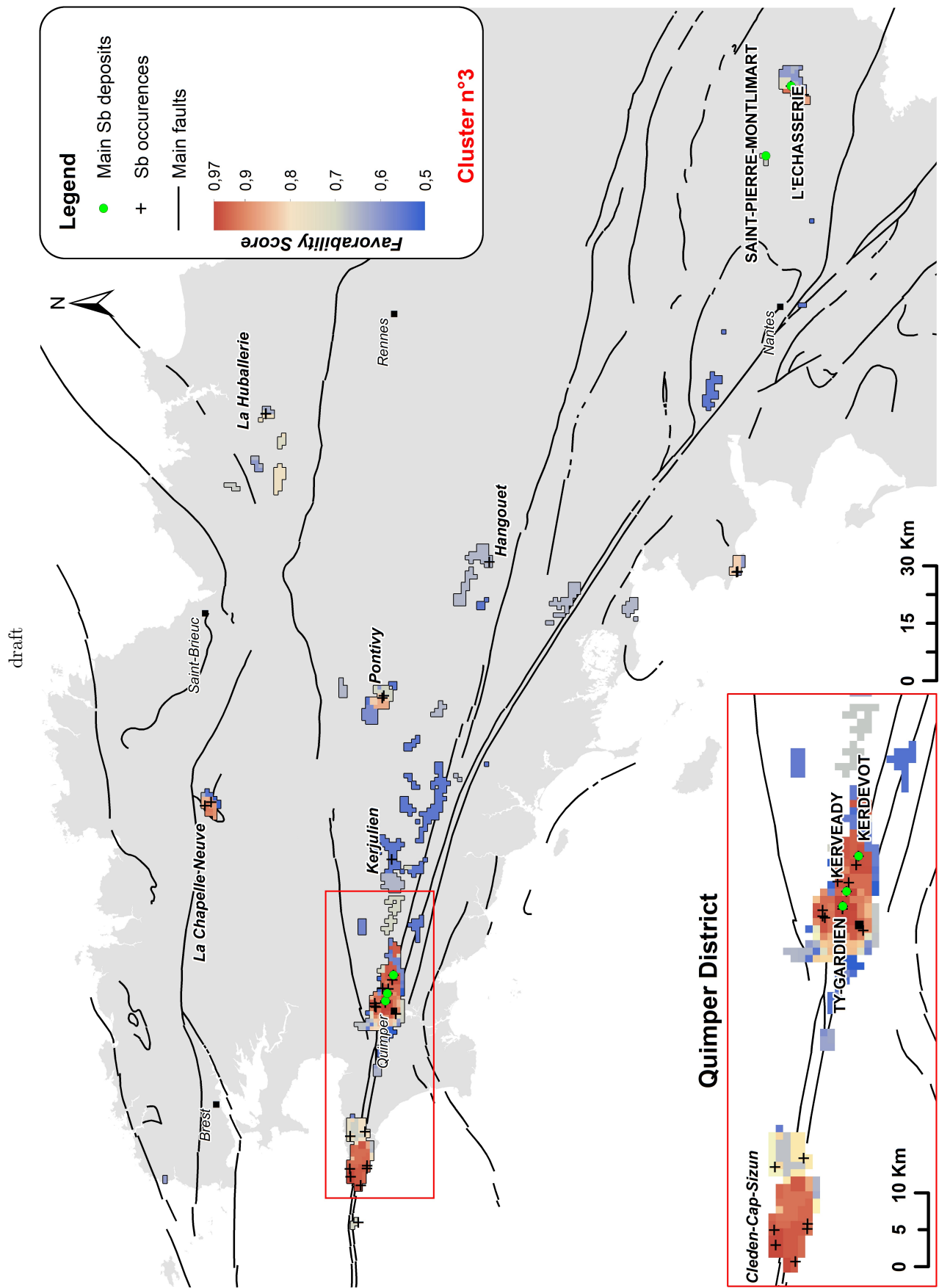


Figure 7.10: Favorability map of antimony in the cluster 3 of the AM.

doleritic formations (DOLE2) and dolerites dykes (FDOLE) with less than 20% of its cells containing those formations. Faults are predominant in this cluster, being contained in 86% of its cells. More details about the frequency of lithology in this cluster are available in **Appendix C.4**. To synthesize, this cluster is composed of Carboniferous granites and migmatites associated with Neoproterozoic basement formations, in a faulted and veins-rich environment.

This cluster contains one of the most important antimony districts of the AM, with the Quimper-Cap Sizun district [BRGM, 1990, Fouquet, 1980, Gorichon and Kerjean, 1979, Moyroud et al., 1979] located in the Western part of the CAD. Antimony mineralization is located in quartz veins hosted in Ordovician and Carboniferous granites, in vicinity of the major shear zone (i.e. South Armorican Shear Zone) [Fouquet, 1980, Chauris, 1977, Marcoux and Chauris, 1994].

The main deposits in this district are the Ty-Gardien, Kervéady and Kerdévot deposits, where the highest favorability scores are located. High scores are also displayed on the other part of the district in the Cap-Sizun veins mineralization. Another deposit with high scores is highlighted inside this cluster, with the Echasserie [Gloaguen et al., 2018] W-Sb polymetallic veins deposits, located above a leucogranite cupola in the Mauges Nappes in the SAD. The Saint-Pierre Montlimart Au-Sb deposit [Braux et al., 1991] with orogenic gold veins hosted in Brioverian formations is highlighted by this cluster with only 4 cells in this domain. High scores are also displayed near the Pontivy (Ag-Pb-Sb-Zn) and the Chapelle Neuve mineralization. Other mineralization show favorability scores only slightly above 0.5.

This cluster groups mineralization occurring in granitic setup, with the Quimper-Cap Sizun district, the Echasserie deposit and the Pontivy mineralization being in granitic environments. Most of its cells are also located near major faults such as the SASZ. The Saint-Pierre Montlimart is an oddity in this cluster as it does not contain granites in its environment. This cluster outlines mainly the mineralization occurring in the Quimper-Cap district and as such seems to highlight more the mineralization sharing similar environment.

Cluster 2

Cluster 2 contains 476 cells, 14.23% of the favorable area. It contains 45 or 14.4% of the mineralized cells of the favorable area, 9.5% of its cells being mineralized. The average favorability score inside this cluster is 0.69 ± 0.13 . A detailed map of the favorability score inside this cluster and the antimony occurrences and deposits within it is displayed in **Figure 7.11**.

Ordovician metamorphic units are frequent in this cluster, with high frequency of orthogneisses (ORTG4; 71%), amphibolites (AMPH4; 64%), micaschists (MCSH4; 55%) which are also mostly present in this cluster. Ordovician sandstones are also frequent (GRES4; 40.8%), but this cluster only accounts for 10% of them in the favorable area. Other metamorphic formations of various ages are present, with Silurian and Neoproterozoic micaschists (MCSH5 and MCSH2), Ordovician schists (MCSH4) and Neoproterozoic gneisses (GNEI2). Ultrabasites (UBAST) are the third most frequent formations, with a frequency of 64% in this cluster, which contains almost 90% of this formation. Faults are a major component of this cluster, being the predictor with the highest frequency (75%). More details about the frequency of lithology in this cluster are available in **Appendix C.3**. This cluster presents a geological environment principally made of metamorphic formations, mainly of Ordovician ages, associated to ultrabasites in a faulted environment.

This cluster contains part of the Vendée district [Blouin, 1991, BRGM, 1988a, Marcoux and Chauris, 1994, Moyroud et al., 1979] located in the Eastern part of the SAD. Antimony mineralization in this

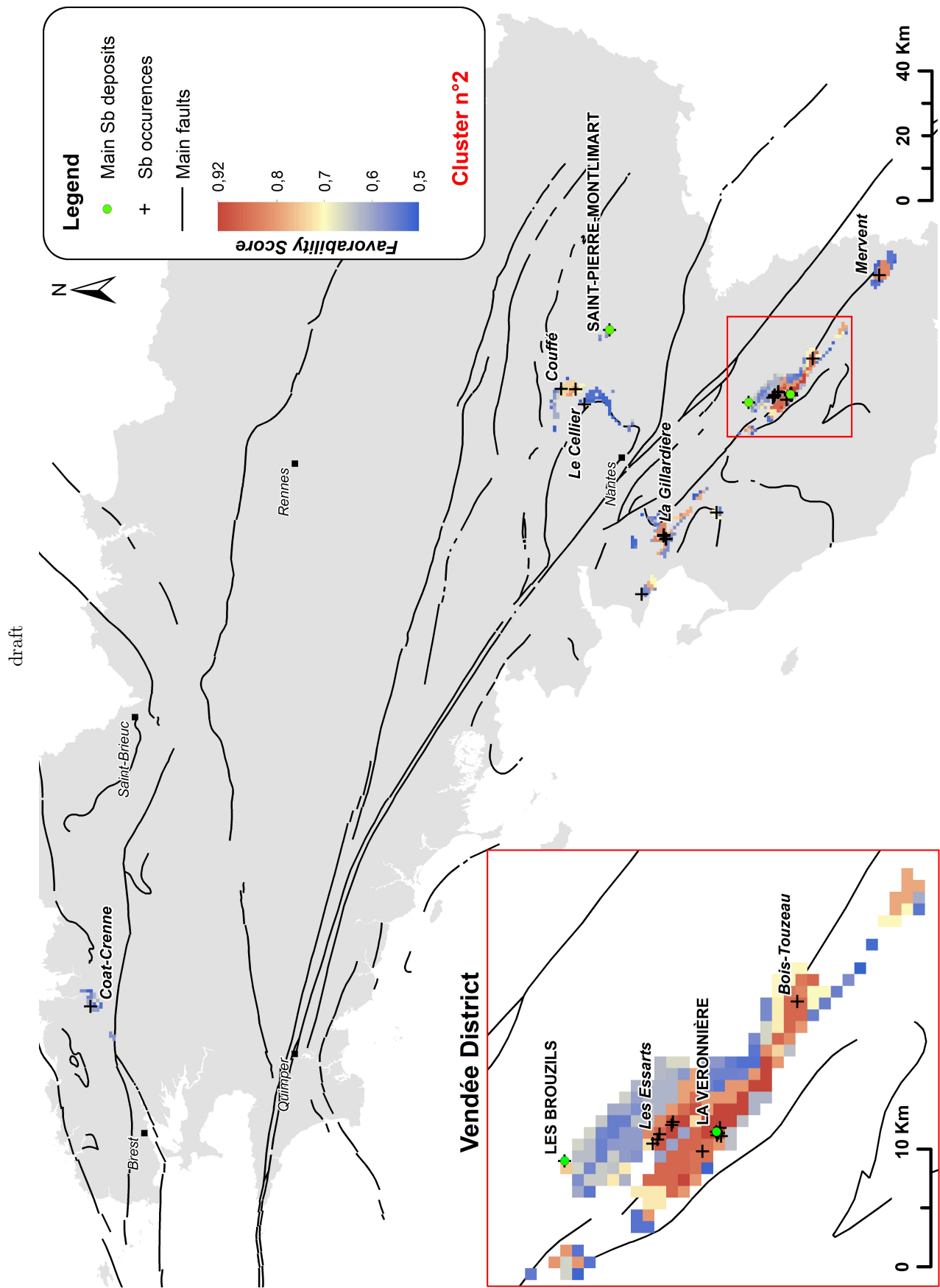


Figure 7.11: Favorability map of antimony in the cluster 2 of the AM.

district are located in quartz veins hosted in Brioverian and Cambrian schists and in the vicinity of Cambrian volcanic formations [Marcoux et al., 1984].

The highest favorability scores are located in one of the main deposits of the Vendée district, the Veronnière and the Essarts [Moyroud et al., 1979, Marcoux et al., 1984] veins-types mineralization. Les Brouzils [Le Fur et al., 1988, Blouin, 1991] despite being another important deposit in the Vendée district, shows lower favorability score. Other favorable areas are displayed along the Champtoceaux Nappe with the Couffé veins mineralization. The Gillardièrre mineralization, a Mo-Sb-Sn placer also displays high score in this cluster. Few cells are located near the Saint-Pierre-Montlimard deposits, previously seen in the cluster 3.

The main mineralization in this cluster comes from the Vendée districts, with a geological environment dominated by metamorphic rocks and basic and ultrabasic formations which are found both in the Essarts Nappes and in the outer parts of the Champtoceaux Nappe, which explains the inclusion of mineralization such as La Gillardièrre, Couffé and Le Cellier in this cluster. The Saint-Pierre-Montlimard deposit is an oddity in this cluster, much like it was in the cluster 3, as the model seems to struggle to compute a favorable score for it or to assign it to 1 cluster.

Cluster 4

The cluster 4 is the smallest one, containing 360 cells, 10.76% of the favorable area. It contains 49 or 15.7% of the mineralized cells of the favorable area, 13.6% of its cells being mineralized. The average favorability score inside this cluster is 0.81 ± 0.14 . A detailed map of the favorability score inside this cluster and the antimony occurrences and deposits within it is displayed in **Figure 7.12**.

This cluster contains a majority of Cambrian formations, with greywackes (GRWK3; 92%), schists (SCSH3; 89%), micaschists (MCSH; 63%) and Volcano-detritic formations (VDET3; 52%). Apart the greywackes, those formations are also almost exclusive to this cluster, with a contribution of around 90% from this cluster. Most of the cells in this cluster contain quartz veins (FQTZ; 95%). This cluster is the one where faults are the rarest, with only 61% of the cells containing faults, which is still a high value. More details about the frequency of lithology in this cluster are available in **Appendix C.5**. Generally, the geological environment within this cluster is characterized by Cambrian basement units and volcano-detritic formations associated with quartz veins.

This cluster is limited to the Vendée district, with the main mineralization being the Brouzils [Le Fur et al., 1988, Blouin, 1991] deposit, like the previous cluster, accompanied with the Rochetretoux [Serment, 1980], La Ramée, Sainte Marie [Moyroud et al., 1979] and la Baussonnière [BRGM, 1988b, Gentilhomme et al., 2001] veins-type mineralization. This cluster is very localized and presents a very high favorability score in comparison to the 3 other clusters and as such, most of the mineralization within it are located in high favorability areas.

This cluster can be associated with the cluster 2, as both are associated with the Vendée district. This cluster contains the Northern part of the Vendée district, here the mineralization are located afar from the Essart Nappes and more into Cambrian lithologies along the SASZ.

7.1.9 Comparison with previous methods and CBA / RF

The result of this DBA / RF approach is here compared to methods previously used and presented in the previous chapter, the results of each method being compared using their Prediction Rate (PR) and

draft

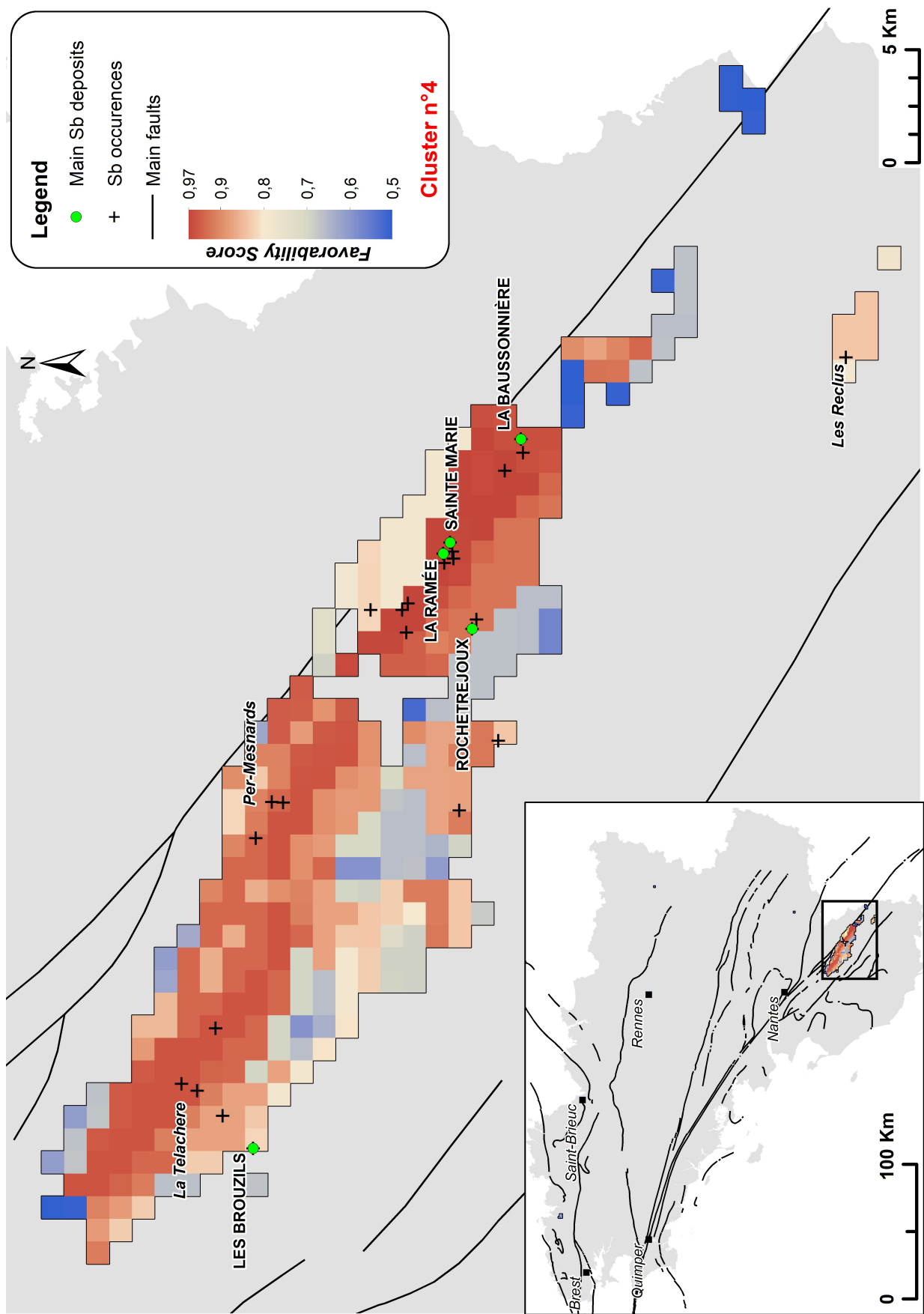


Figure 7.12: Favorability map of antimony in the cluster 4 of the AM.

Percentage of Prospective Area (PPA) in **Figure 7.13.a**. The PR is equivalent to the TPR in the ROC curves and evaluates the capacity of the model to produce True Positives (TP). The PPA shows the percentage of the surface considered as favorable (i.e. TP and FP). As with the ROC curve, the model efficiency is evaluated with the AUC.

The result of the DBA / RF method is shown **Figure 7.13.c**. The three other favorability maps have been computed using CBA / RF (**Figure 7.13.b**), CBA / Ranking (**Figure 7.13.d**) and WoE approaches (**Figure 7.13.e**), based on a slightly different dataset.

From the favorability maps, one can conclude that the DBA / RF model allows a finer resolution compared to the CBA / RF and the CBA / Ranking model. The favorable areas are more localized, covering a more restrained area, while other methods highlight larger surfaces as favorable. This result seems more reasonable considering the fact that mineralization are rare phenomena (as shown by the large imbalance between the "mineralized" and "non-mineralized" classes during the classification) and, as such, are expected to cover only a fraction of the study area. This result is also more interesting from a prospective point of view, as it limits the area to investigate and explore to seek new mineralization. With the PRPPA curve, one can see that this model is able to correctly predict 100% of the mineralized cells by considering only about 12% of the map as being favourable, while the other methods reach the same PR for more than 60% of PPA.

7.2 Antimony MPM in the Central Iberian Zone

7.2.1 Context

The DBA / RF / HCPC methodology is here applied again to the Antimony deposits of another part of the Variscan Range, in the Central Iberian Zone (CIZ) covering both Spain and Portugal. The same methodology as for the AM (see above) is applied to test the effectiveness of this new predictive approach to various datasets and various — while comparable — contexts. Indeed, the scale of the dataset used in this case study is different. While geological maps covering the CIZ are available at a scale of 1:50,000 for both the Portuguese and Spanish sides, harmonization of those geological maps at the scale of CIZ is not available. Harmonization covering this area exists in the Spanish part, with the GEODE maps (Mapas IGME: <http://info.igme.es/cartografiadigital/geologica/geode.aspx>) produced from the 1:50,000 geological maps of Spain. However, this harmonization does not fully cover the study area and would require further processing to allow the application of the DBA / RF / HCPC methodology at the scale of the whole CIZ.

In the Portuguese part, no harmonization is available at a 1:50,000 scale, but 1:200,000 geological map covering Portugal exists, with two sheets covering the CIZ and GTMZ. Integration of both the Spanish and Portuguese geological maps would require either the harmonization of around 70 geological sheets at a scale of 1:50,000 from Portugal in combination with the two set of harmonized maps of Spain (GEODE) or the harmonization of two 1:200,000 Portuguese geological maps with these harmonized maps. Though such dataset would be of prime interest to understand the structural framework of Iberia and to perform statistical map analyses at various scales, this task is out of the scope of this thesis.

As such, we rely for this application on the 1:1,000,000 geological map of the Iberian Peninsula [Rodríguez Fernández and Oliveira, 2015] which covers the whole CIZ. While a smaller spatial resolution

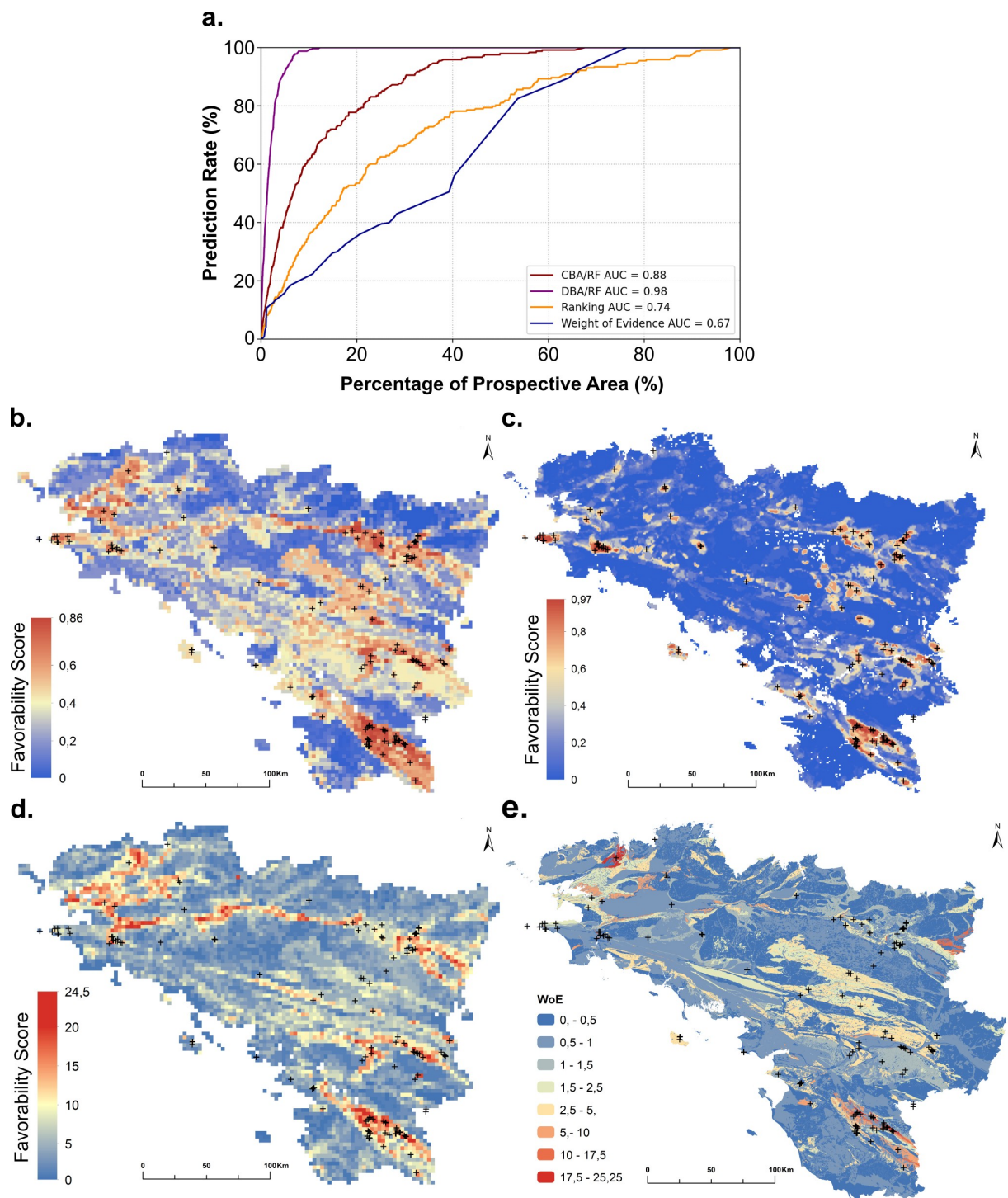


Figure 7.13: Comparison of various methods for antimony prospectivity mapping in the Armorican Massif. a. Prediction Rate and Percentage of Prospective Area (PRPPA) curve of the different predictive models b. Favorability map of the CBA / RF model (resolution: 3 000 m) c. Favorability map of the DBA / RF model (resolution: 1 000 m) d. Favorability map of the CBA / Ranking model, using the same CBA matrix as the Ranking method (resolution: 3 000 m) e. Favorability map of the WoE model (score computed for each lithology).

means that the precision of the results are expected to be lower, as the diversity of lithological formations will decrease compared to a study at a 1:50,000 scale, using this map allows the application of this method on a large area, comprising both the Portuguese and Spanish parts of the CIZ. This study also includes the Galicia-Trás-os-Montes (GTMZ) zone of the Iberian Massif (CIZ s.l.). It will allow us to test relevance of our new predictivity approach with data set layers at country-scale.

7.2.2 Input Dataset

In this case study, the lithological dataset used is thus extracted from the 1:1,000,000 geological map of the Iberian Peninsula (2015 edition) produced by the Spanish Geological Survey (IGME) in collaboration with the Portuguese Geological Survey (LNEG) [Rodríguez Fernández and Oliveira, 2015]. As shown in **Figure 7.14**, the lithological formations are restrained to Paleozoic and older formations, as for the AM, in order to study the Variscan orogen and its relative "basement" rock units. This selection results in 87 unique lithological formations covering 154 000 Km². If the number of formations is similar to the AM (106 unique lithologies), precision of the lithological boundaries is clearly not the same: in the AM, 1:50,000 lithological formations have been harmonized and integrated into a reduced number of lithologies but the precision of their boundaries has been preserved.

In addition to the lithological formations, faults are extracted from the same geological map, at the same 1:1,000,000 scale. Only faults intersecting the selected lithological formations are considered, and all types of faults are used. The antimony occurrences are provided by both the LNEG (SIORMINP database, <https://geoportal.lneg.pt/pt/bds/siorminp/>; Patinha et al. [2020]) and IGME (BDMIN database, <http://doc.igme.es/bdmin/>), resulting in 140 geolocalised point features across the study area. As for the AM, all sizes and types of mineralization are used in this study, with a mix of various veins-type deposits and stratabound-type deposits. No geophysical or geochemical datasets are used in this application.

7.2.3 DBA matrix

The DBA method is applied on the lithological dataset of the CIZ to produce the multivariate grid. To find the optimal neighbouring search distance R and cell size d , the DBA is iterated over multiple values of R and R/d , from 2 000 m to 35 000 m with a 1 000 m step, with R/d ratio ranging from 0.71 to 5. Results from those iterations are shown in **Figure 7.15**, with (i) the number of distinct environments produced (**Figure 7.15.a**), (ii) the number of cells (**Figure 7.15.b**), (iii) the average number of lithologies in the produced cells (**Figure 7.15.c**) and (iv) the percentage of cells containing only one lithology (**Figure 7.15.d**).

Compared to the previous study, the number of distinct environments (**Figure 7.15.a**) is lower, reaching a maximum value of 4 000, as the number of lithologies in the study area is lower: 87 unique lithologies are located in the CIZ while the harmonized dataset of the AM contains 106 unique lithologies. Each lithology also covers a larger area, as the extension of the study area is much larger, with 35 000 km² for the AM and 120 000 km² for the CIZ. As such, the number of cells (**Figure 7.15.b**) is much higher, with values reaching a maximum of 1 million cells, while the matrix for the AM only reaches a maximum size of 280 000 cells. In comparison with the AM study, the average number of lithologies is decreased (**Figure 7.15.c**), while the percentage of unilithological cells is increased (**Figure 7.15.d**).

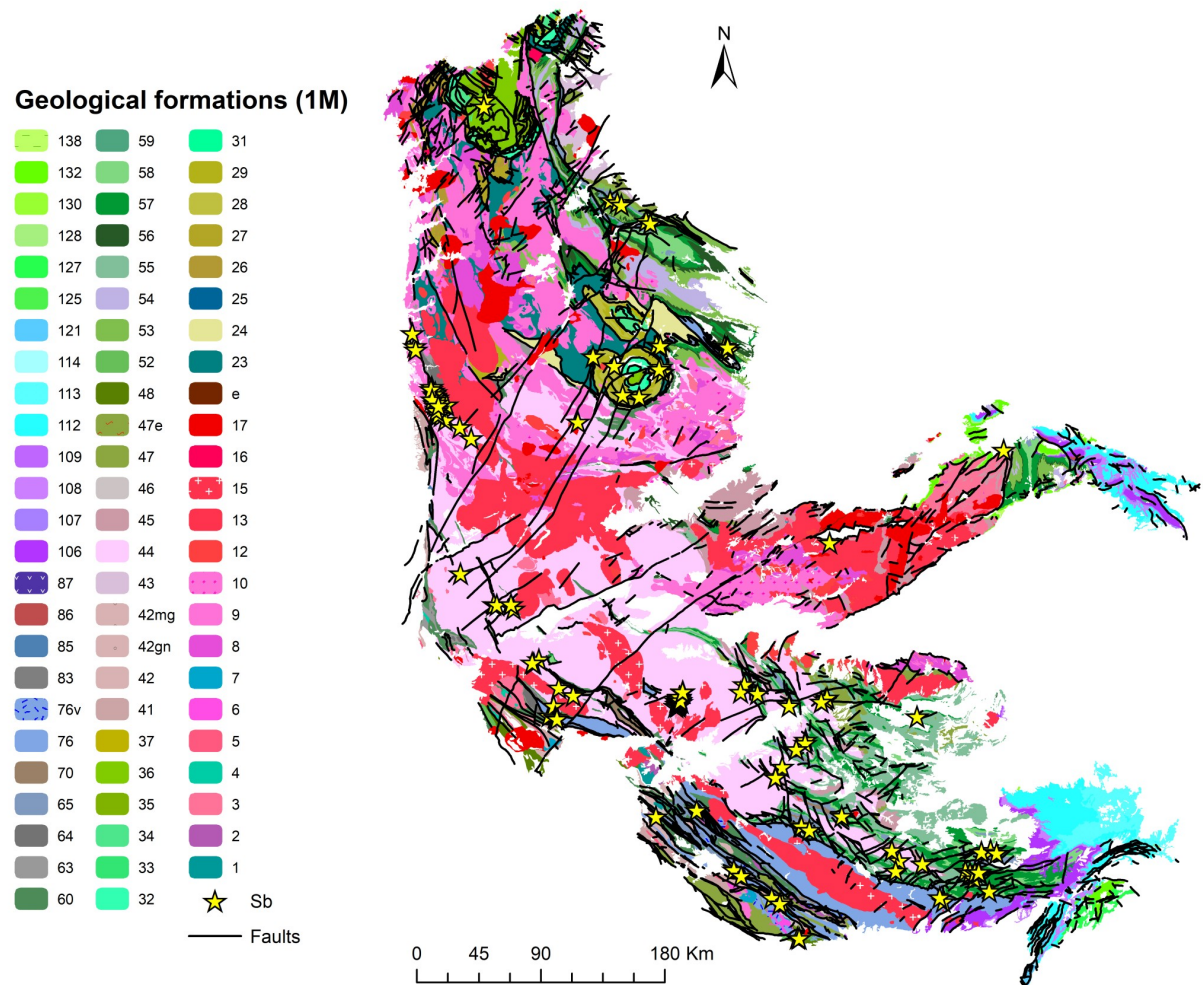


Figure 7.14: Maps of the datasets used for the DBA / RF antimony prospectivity mapping across the Central Iberian Zone. Selected formations are of Palaeozoic and Precambrian ages, from the 1M geological map of the Iberian Peninsula (2015; Rodríguez Fernández and Oliveira [2015]). Faults are from the same geological map. Antimony mineralization is from the LNEG (SIORMINP database, <https://geoportal.lneg.pt/pt/bds/siorminp/>; Patinha et al. [2020]) and IGME (BDMIN database, <http://doc.igme.es/bdmin/>) database.

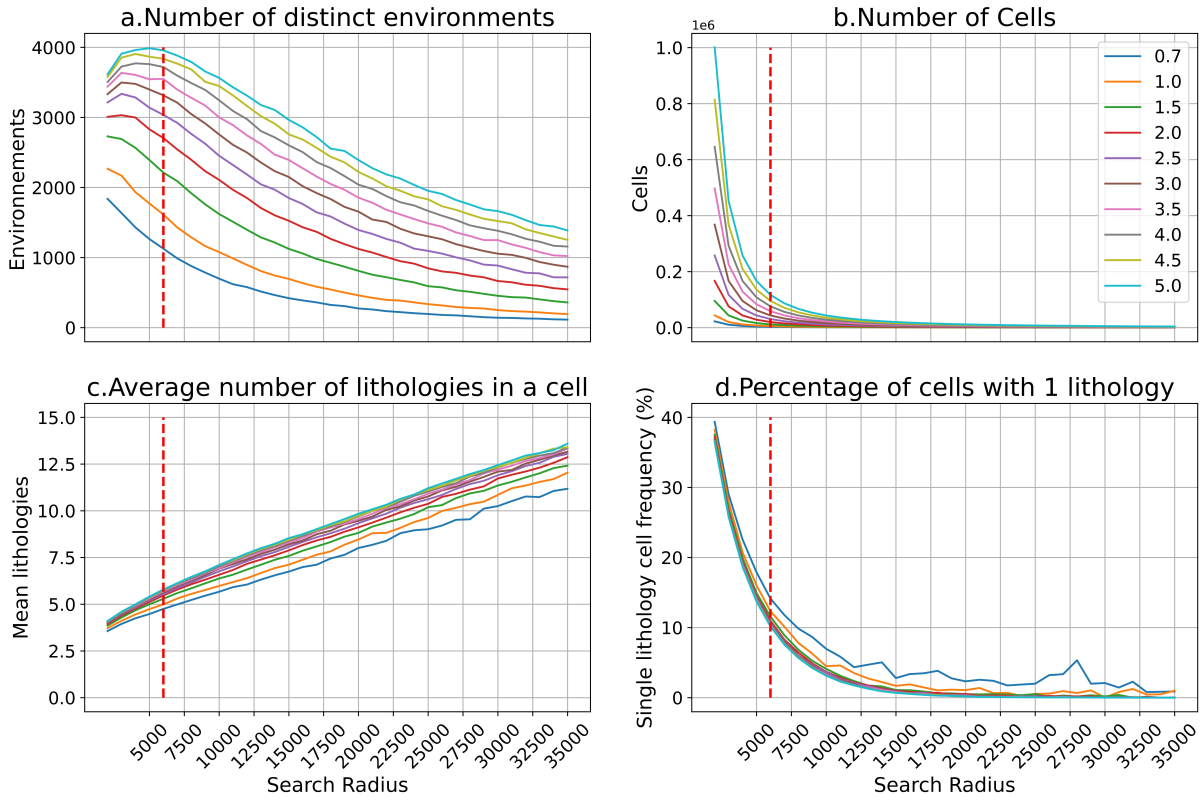


Figure 7.15: Application of the multi-scale iterating process of DBA matrices on the lithological dataset of the Central Iberian Zone for various values of R and R/d . The optimal value found for the disc size is 6 000 m for a R/d ratio of 3, showed with the red dashed line. a. Number of lithological associations b. Number of cells. c. Average lithologies per cell d. Frequency of cells containing 1 lithology only.

It results in a higher optimal value for the neighbouring search with $R = 6\,000$ m for a R/d of 3, similar to that of the AM. Keeping the same R/d ratio in-between the two case studies allows keeping the same proportion of overlapping and also allows having a high number of unique environments without drastically increasing the number of cells. In the CIZ case, DBA space discretizing results in 3 315 unique lithological associations, while keeping the percentage of unilithological around 10% with 10.5% of the cells containing only 1 lithology. This number is higher than for the Armorican Massif DBA matrix (3%), but reaching a similar value would require decreasing the resolution even further. The final resolution of the matrix is 2 000 m, with 43 505 cells, each cell containing 5.6 lithologies on average.

Once the DBA matrix has been produced from the lithologies, the faults and mineralization are added to the matrix. A 2 000 m neighbouring search distance is used for both datasets, resulting in 16 303 cells containing faults or 37.5% of the cells and 305 cells containing mineralization or 0.70% of the cells. This search distance has the same proportion of overlapping as the one used for the AM, as it has the same size as the DBA cells, or a R/d ratio of 1 (R_{min} and $R_{fault} = 2\,000$ m, $d = 2\,000$ m, $R_{fault/min}/d = 1$). The proportion of mineralized cells is relatively similar to the Armorican Massif (0.67%), but the proportion of faults is almost twice as low (62%) due to the nature of the fault dataset, restricted to the main faults of the domain at a scale of 1:1,000,000, while the AM case study contained all faults at a scale of 1:50,000.

7.2.4 DBA / RF results

The antimony favorability map is produced by applying the RF predictive model on the DBA matrix, including the lithologies, the faults and mineralization synthetized in the multivariate matrix. The binary classification is applied with all cells, mineralized and non-mineralized. Random Search is first used to define the optimal hyperparameters with a 5-fold Stratified Cross-Validation (SCV) and then the optimal hyperparameters are used to produce the final model with the whole matrix.

The resulting favorability map is shown in **Figure 7.16**, high favorability scores being displayed in green, with a maximum value of 0.973, and low favorability scores in deep purple reaching a minimum of 0. The mineralized cells have an average favorability score of 0.79 with a SD of 0.14. The lowest value for a mineralized cell is 0.30, with 18 cells having favorability score below 0.5, most of them (15) having a score between 0.4 and 0.5. For non-mineralized cells, average favorability is 0.15 with a SD of 0.23 and values ranging from 0 to 0.973. As for the case study of the AM, we have high average favorability scores across mineralized cells and low favorability scores for non-mineralized cells, with some of them considered as favorable to host antimony.

A ROC curve is produced with a 5-folds SCV using the optimal hyperparameters (**Figure 7.17**). The average AUC of the cross-validation ROC curves is 0.94 ± 0.01 , with AUC varying between 0.96 and 0.93. The shape of the curve is stable, the model performing efficiently despite the various training set subsamples used by the Cross-Validation. After the production of the final model, using the whole matrix to train the predictive model, a final ROC curve is produced (**Figure 7.18**), with an AUC of 0.97.

7.2.5 Identification of favorable geological environments

The final ROC curve (**Figure 7.18**) is used to determine the most favorable cells, i.e. to establish the threshold above which the cells are selected to perform the clustering analysis. The G-Means and J-Score are computed from the ROC curve, showing both an optimal threshold value of 0.51 to perform the binary classification. Cells with a score above this 0.51 threshold are displayed in **Figure 7.19.a**. The resulting classes are shown in **Figure 7.19.b**, the selected cells corresponding to the 4 653 FP (10.7% of the non-mineralized cells) and 286 True Positives (93.8% of the mineralized cells) for a total of 4 929 cells. Those selected cells contain 261 unique lithological associations.

HCPC clustering is applied on the selected cells. First, the MCA factorial analysis produces new factorial coordinates based on predictors and cells relationships. The AHC clustering is then applied to group the cells together from those new coordinates. The same euclidean distance and Ward's method from the previous application are used to group the clusters. The "cut-off" value is determined manually using statistical analysis and spatial visualisation of the results. 4 clusters are then recognised and displayed in **Figure 7.20**. Each cluster is then analyzed regarding the frequency of the lithologies comprised in its cells.

7.2.6 Favorable areas and best predictors

The effectiveness of each predictor used by the decision trees is evaluated using the Mean Decrease in Impurity (MDI) score. The higher the MDI is, the more frequently the given predictor is used in the classification process of the Decision Trees of the RF model. In **Figure 7.21.a** are presented the MDI

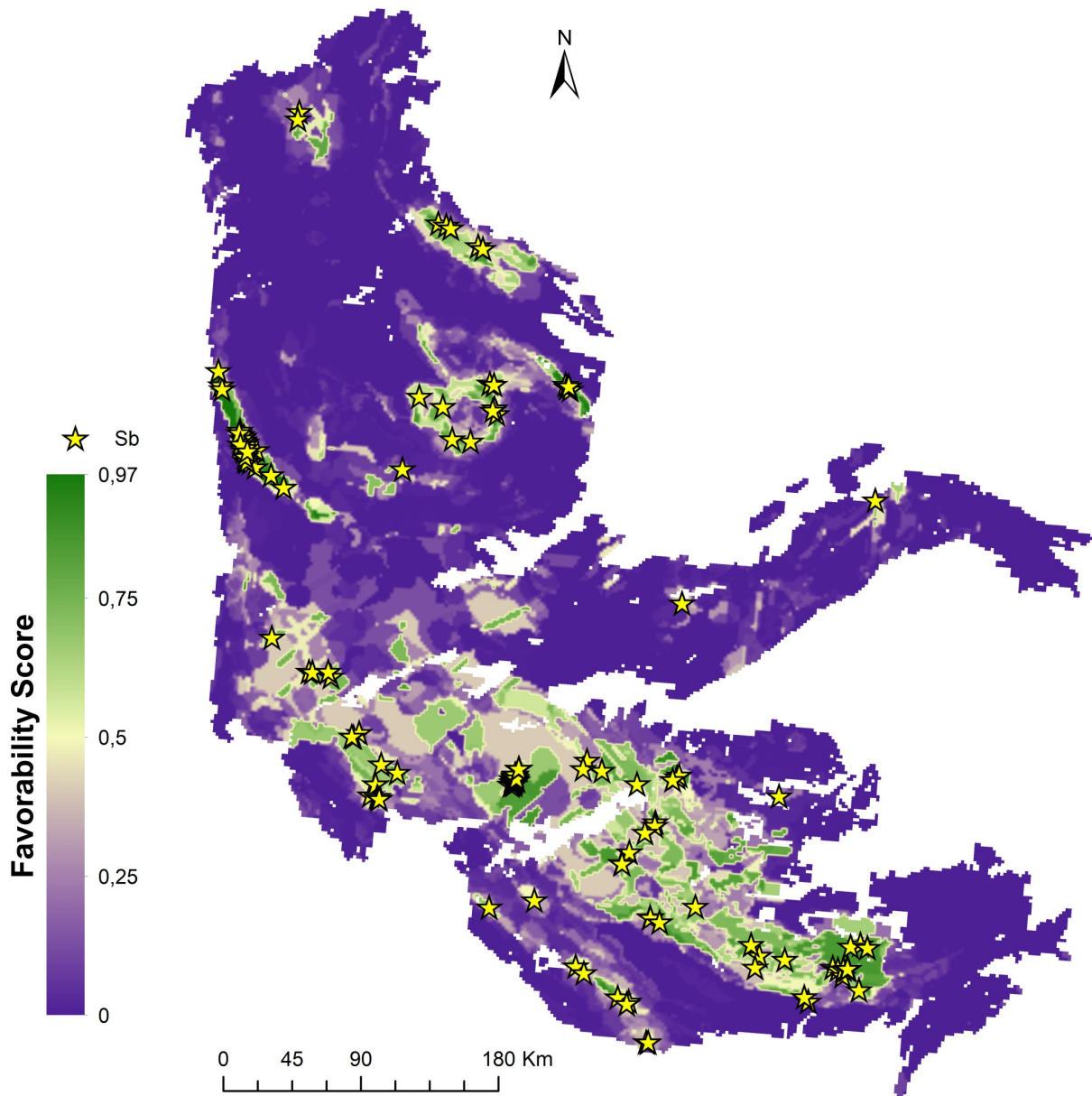


Figure 7.16: Favorability maps of antimony in the Central Iberian Zone (CIZ) produced from the integrated DBA / RF method. With a neighbouring search R of 6 000 m and a R/d of 3, the resolution of the grid is 2 000 m. Mineralization and faults are added to the matrix with a neighbouring search R of 2 000 m. High favorability score are displayed in green, corresponding to a higher likelihood for the cell to contain antimony mineralization.

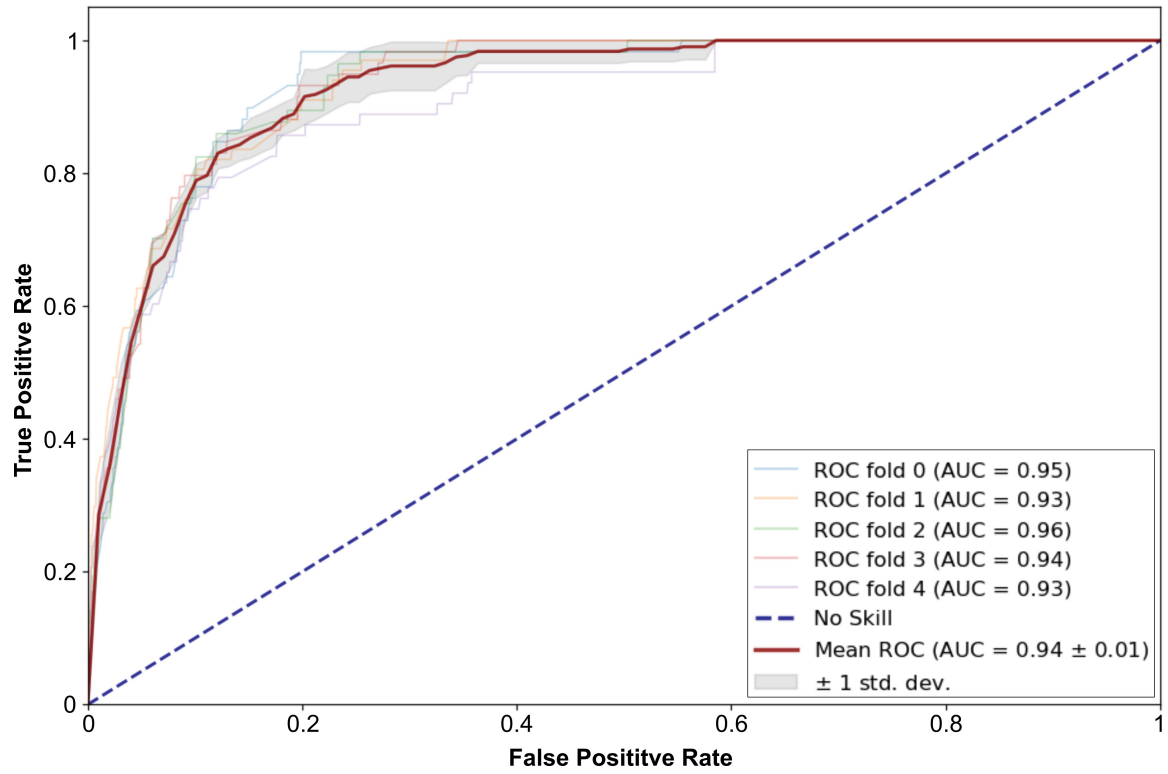


Figure 7.17: Receiver Operating Characteristic (ROC) curve from the 5-folds Stratified Cross Validation of the DBA / RF predictive model of the Central Iberian Zone. Each fold is displayed with its corresponding Area Under the Curve (AUC) score. The no-skill model displayed in dotted blue corresponds to a no-skill predictive model.

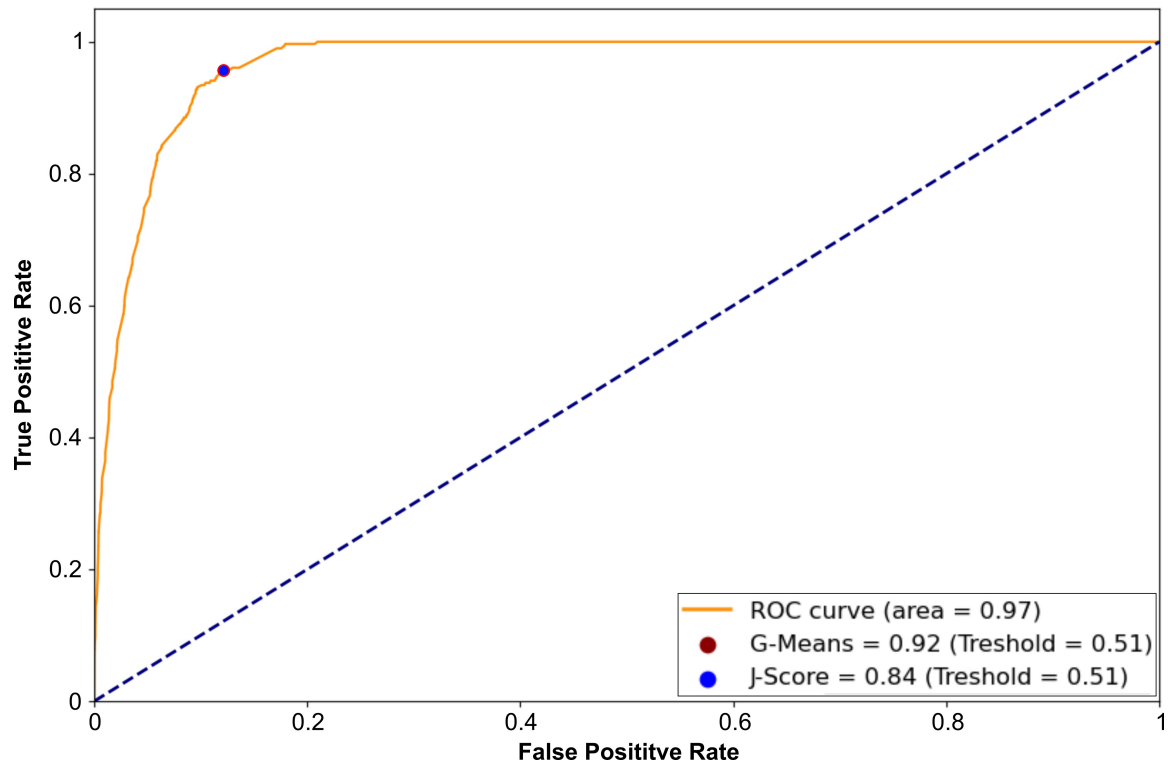


Figure 7.18: Receiver Operating Characteristic (ROC) curve from the final DBA / RF predictive model of the Central Iberian Zone. All cells from the DBA matrix are used to train the predictive model. The no-skill model is displayed in dotted blue.

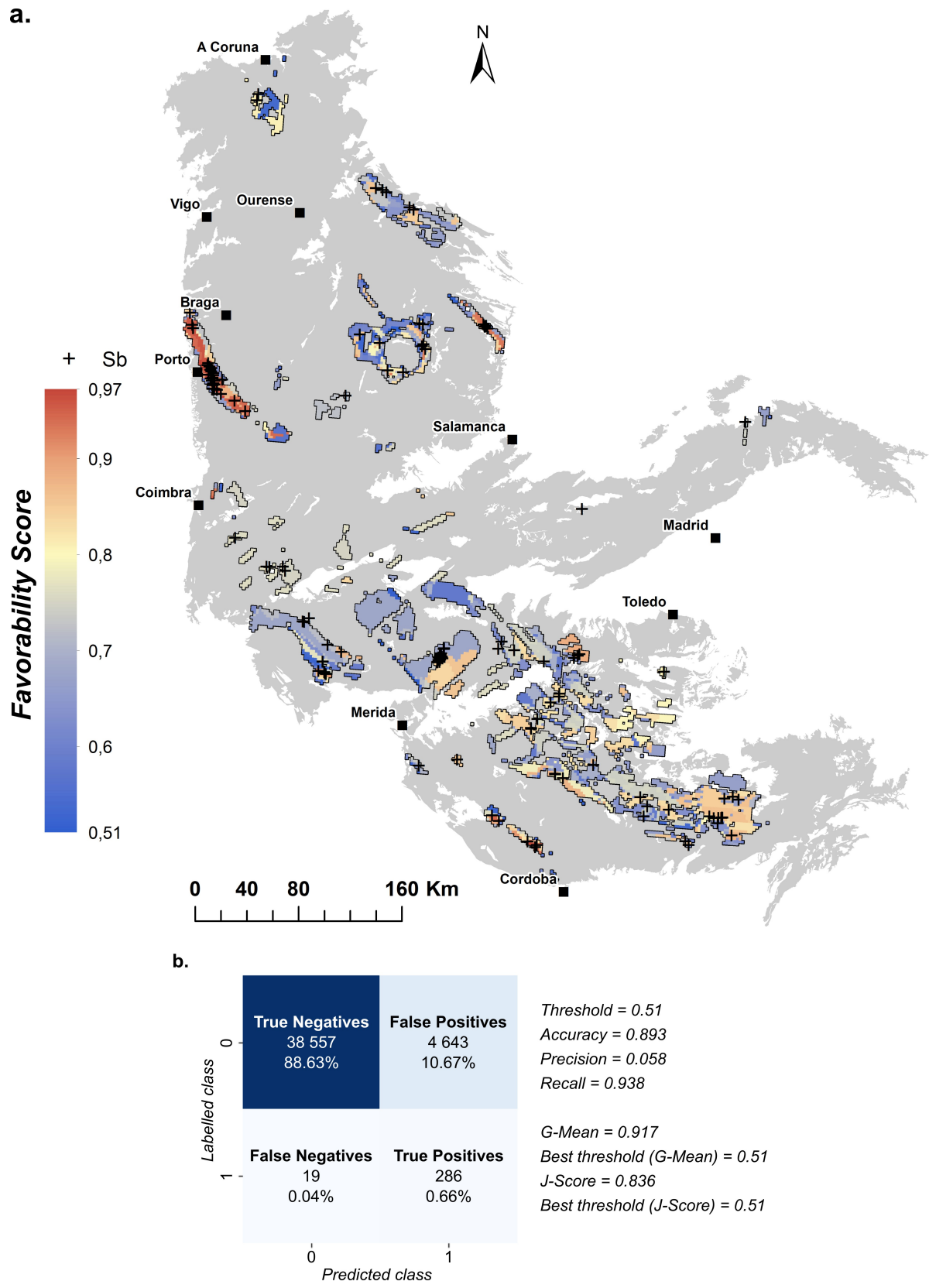


Figure 7.19: Identification of the favorable areas of the favorability map of antimony in the Central Iberian Zone based on the G-Means and J-Score best threshold. a. Favorability map of cells displaying a score higher than 0.51, the color ramp is scaled for score higher than 0.51 b. Associated confusion matrix of the classification of the cells based on a 0.506 threshold for the favorability score.

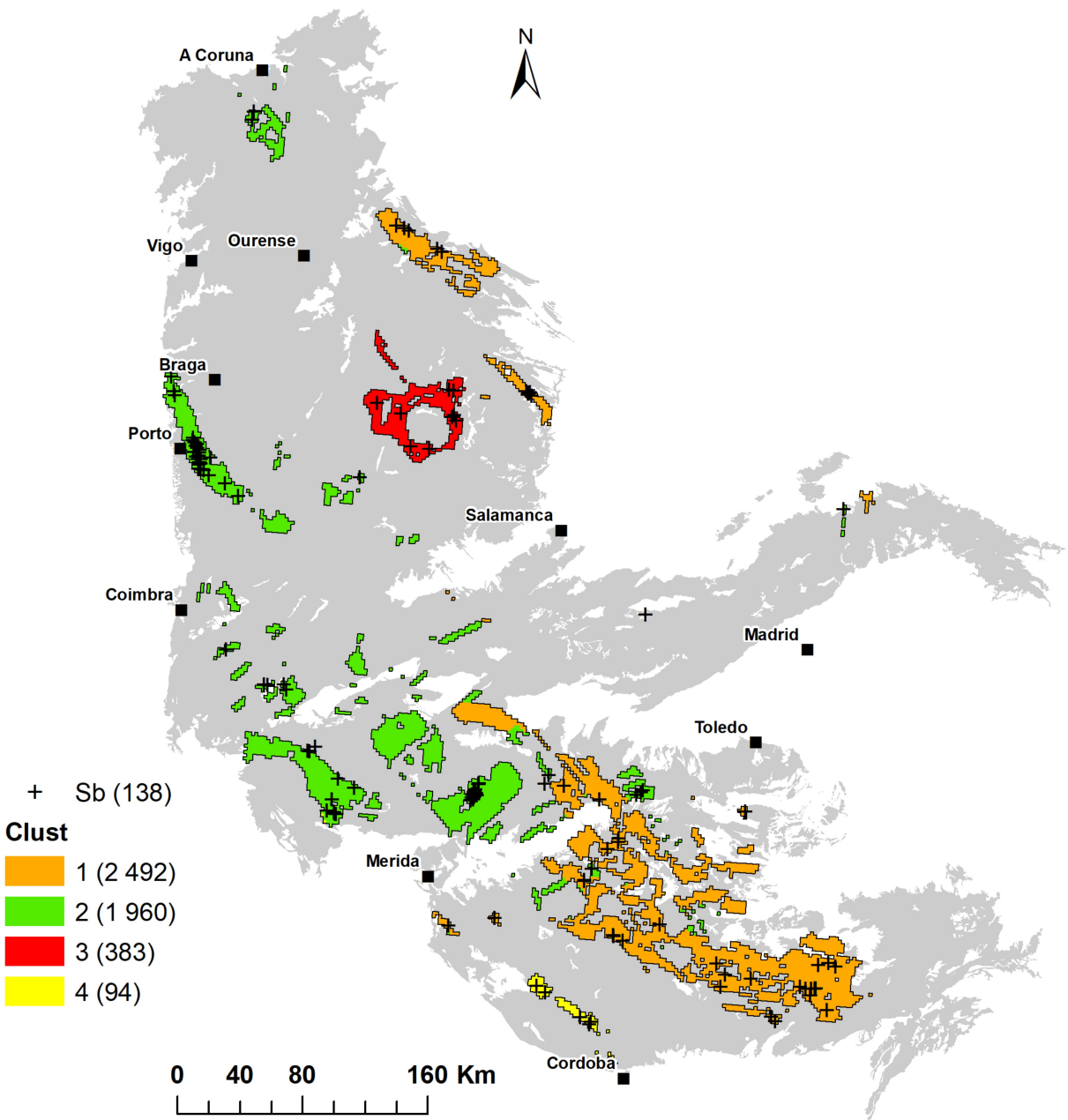


Figure 7.20: Map of the 4 clusters produced by the Hierarchical Clustering on Principal Component (HCPC) classification of the favorable cells (score ≥ 0.51) of the Central Iberian Zone DBA / RF favorability map. The number of cells is displayed for each cluster, as well as the number of antimony occurrences across the study area.

scores and SD of the 20 most important predictors for this predictive model. Map extension of the 4 most important predictors are shown in **Figure 7.21.b to e.**

The most important predictor is the Cambrian schists and graywackes complex ("44"; **Figure 7.21.b**) with a MDI of 0.8. It is located over a large area across the CIZ, covering most of the Central Iberian Domain s.s with 15 243 cells containing it (35% of all cells) and 178 mineralized ones. The favorability scores of cells containing this formation is on average 0.29 ± 0.27 .

The Carboniferous basin of Bierzo and Ciñera-Matallena ("85"; **Figure 7.21.c**) comes in second place with a MDI of around 0.064. This formation is very localized and only present in 509 cells (1.2% of all cells) and 10.4% of its cells are mineralized (53 mineralized cells). Its average favorability score is quite high with an average of 0.62 ± 0.31 .

Carboniferous two-mica granites and leucogranites ("9"; **Figure 7.21.d**) are the third most important formation with a MDI of 0.062, slightly lower than the previous predictor. It covers a large area over the Northern part of the CIZ, mainly within the GTMZ domain, with 11 843 cells (27.2% of all cells), 305 of them being mineralized. The favorability score within cells containing this formation is lower than for the previous ones with an average score of 0.16 ± 0.24 .

Faults (Faults; **Figure 7.21.e**) are also a major predictor in this predictive model with a MDI of 0.06 with a lower SD than previous predictors explained. This predictor is widespread across the study area. 16 303 cells are crossed by faults (37.5% of all cells), those cells scoring an average favorability score of 0.20 ± 0.27 .

Like in the previous case of the AM study, it is necessary to investigate the relationships between the predictors and the mineralization across the study area to better understand their relationships. Some predictors are effective due to their small extension and localized nature (see predictor "85", **Figure 7.21.c**), in specific areas showing evidence of mineralization, while other are localized but covering larger areas ("44", "9") and some are widespread (Faults). The interpretation of their relationship with the mineralization will differ depending on their spatial distribution. In any case, with a high MDI, those formation are worth investigating in regard to antimony mineralization according to this model and this dataset.

7.2.7 Analysis of the clusters

Cluster 1

Cluster number 1 comprises 2 492 cells, 50.6% of the favorable area. It contains 104 or 36.4% of the mineralized cells from the favorable area, 4.2% of its cells being mineralized. The average favorability score inside this cluster is 0.72 ± 0.09 . A detailed map of the favorability scores inside this cluster and the antimony occurrences and deposits within it is displayed in **Figure 7.22**.

This cluster is composed of Ordovician schists and quartzites associated with Silurian black shales, in a faulted environment. The 3 main lithologies in frequencies in this cluster are Ordovician metamorphic shales and schists formations with the formation "57" (Formation of Luarca, Róo and others) in 94.2% of the cells, "58" (Formation of Agüeira and others) in 80.6% and "55" (Serie of Púrpura and others) in 76%. A Silurian formation also represents a major part of this cluster, with formation "65" being represented in 73.3% of the cells. Those 4 formations are mostly exclusive to this cluster, as it contains more than 90% of them. Faults are present in 63.9% of the cells and this cluster contain 62.5% of all

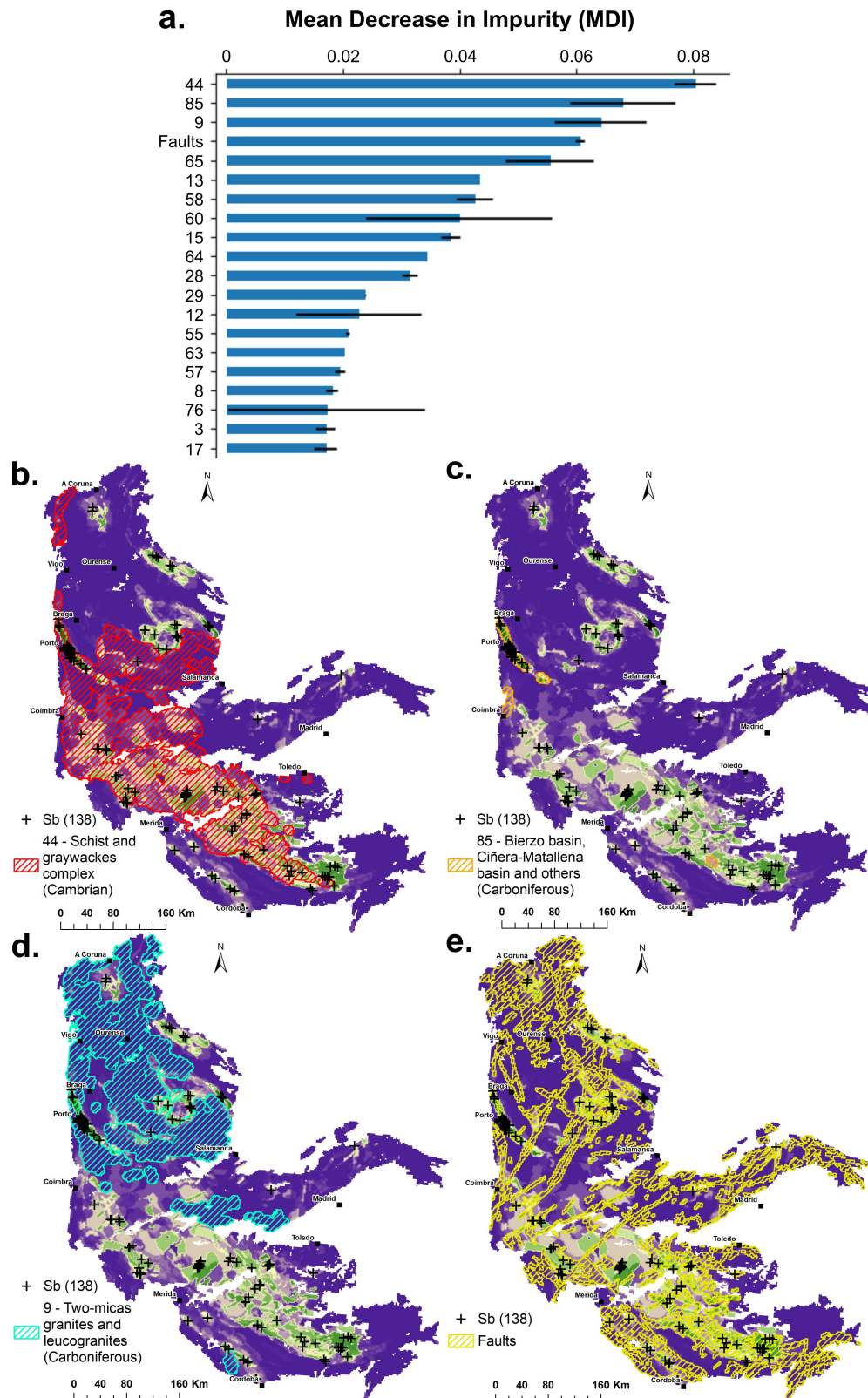


Figure 7.21: Predictors importance for the DBA / RF predictive model of the Central Iberian Zone. a. Mean Decrease in Impurity (MDI) of the 20 main predictors. b. Location of the cells containing the Cambrian Schist and Graywackes complex (44) (GRWK3) c. Location of the cells containing the Carboniferous basin (85) d. Location of the cells containing the 2 micas granites and leucogranites (9) e. Location of the cells containing Faults.

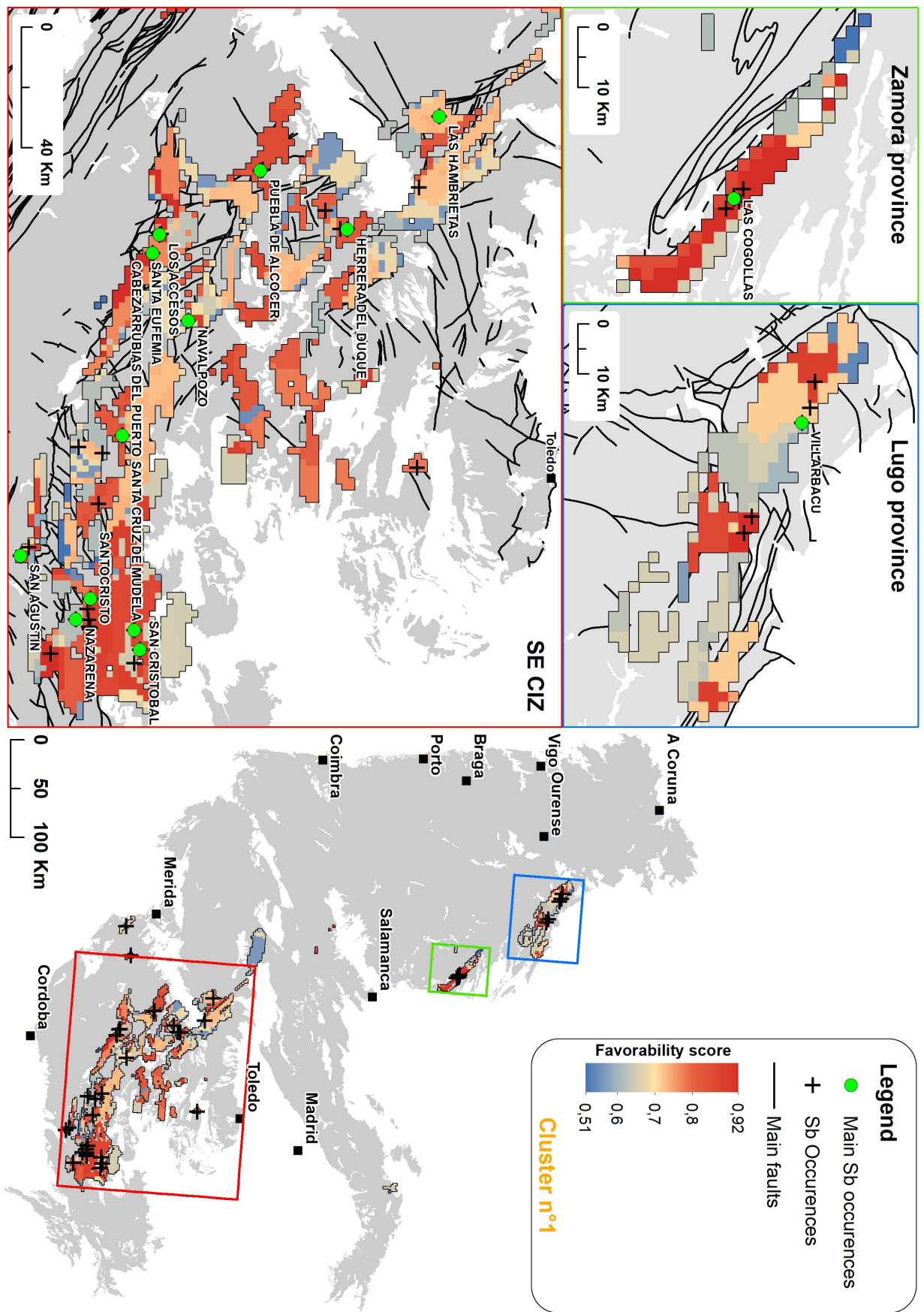


Figure 7.22: Favorability map of antimony in the cluster 1 of the CIZ.

the faulted cells in the favorable area. More details about the frequency of lithology in this cluster are available in **Appendix C.7**.

Most of the cells of this cluster are located on the SE part of the CIZ, with mineralization belonging to the Cáceres (las Hambrientas), Badajoz (Herrera del Duque, Puebla de Alcocer), Cordoba (Los Accessos, Santa Eufemia), Jaén (San Agustín), and Ciudad Real (Cabezarrubias del Puerto, San Cristobal, Santo Cristo, Nazarena, Navalpozo) provinces of Spain. Most of those mineralization are veins-types antimony, with the exception of the Los Accessos, Santa Eufemia and Nazarena mineralization, which are strata-bound type deposits [Gumiel and Arribas, 1987]. The highest favorability scores in this cluster are located in the Eastern province of Ciudad Real and in the Southern part, in the Cordoba province, reaching values of 0.86 around the Los Accessos, Santa Eufemia and Nazarena mineralized zones.

Some parts of this cluster are also located to the Northern part of the CIZ in Spain, along the Olla de Sapo domain, with the Lugo Province, with the strata-bound deposits of Villarbacu, and the Zamora province with the Las Cogollas vein-type mineralization. The latter presents the highest score inside this cluster, with the maximum value of 0.92, while other areas show lower scores on average.

Cluster 2

The cluster 2 comprises 1 960 cells, i.e. 39.8% of the favorable area. It contains 134 or 46.9% of the mineralized cells of the favorable area, 6.8% of its cells being mineralized. The average favorability score inside this cluster is 0.73 ± 0.11 . A detailed map of the favorability scores inside this cluster and the antimony occurrences and deposits within it is displayed in **Figure 7.23**.

The main geological environment of this cluster is made of Cambrian schists and graywackes complexes with tardi-orogenic biotites and cordierites granites. The formation "44", (schists and graywackes complex of Neoproterozoic to Cambrian age) is present in 93% of the cells, constituting the main lithological formation of this cluster. The tardi-orogenic granites of the formation "15" account for 48% of the cells and are only present in this cluster. The formation "60", an Ordovician formation composed of schists with basic volcanic units, is the third lithological formation with 37.1% of the cell containing it and is mostly comprised in this cluster (85.6% cluster contribution). Faults are less present, as only 29% of the cells are crossed by faults. More details about the frequency of lithology in this cluster are shown in **Appendix C.8**.

Most of the cells of this cluster are located on the South Western part of the CIZ, with mineralization from the Cáceres (Portoviejo, Serafina, Plansenzuela), Badajoz (Mari Pepa, Santa Aurelia, San Antonio) and Ciudad Real (Anchuras de los Montes, Puebla de Alcocer) provinces, and the Castelo Branco mineralization (Portugal). All those mineralization are veins-type, except for the San Antonio mineralization, which is a stratabound deposit [Gumiel and Arribas, 1987]. This area shows a low score on average, with maximum values located to the east, around the Plasezuela, Serafina and Anchura de los Montes mineralization.

The cells with the highest favorability scores in this cluster are located in its western part, near Porto, in the Dúrico-Beirō Au-Sb district (Ribeiro de Igreja, Pinheirinhos, Alto de Sobrido) [Gumiel and Arribas, 1987, Couto et al., 1990] with a maximum score of 0.97. This cluster also contains cells with lower scores located near La Coruña in the North of Spain (Pandeiro). In those areas too, vein-type is predominant, with the exception of the stratabound deposit of Ribeiro de Igreja.

Cluster 3

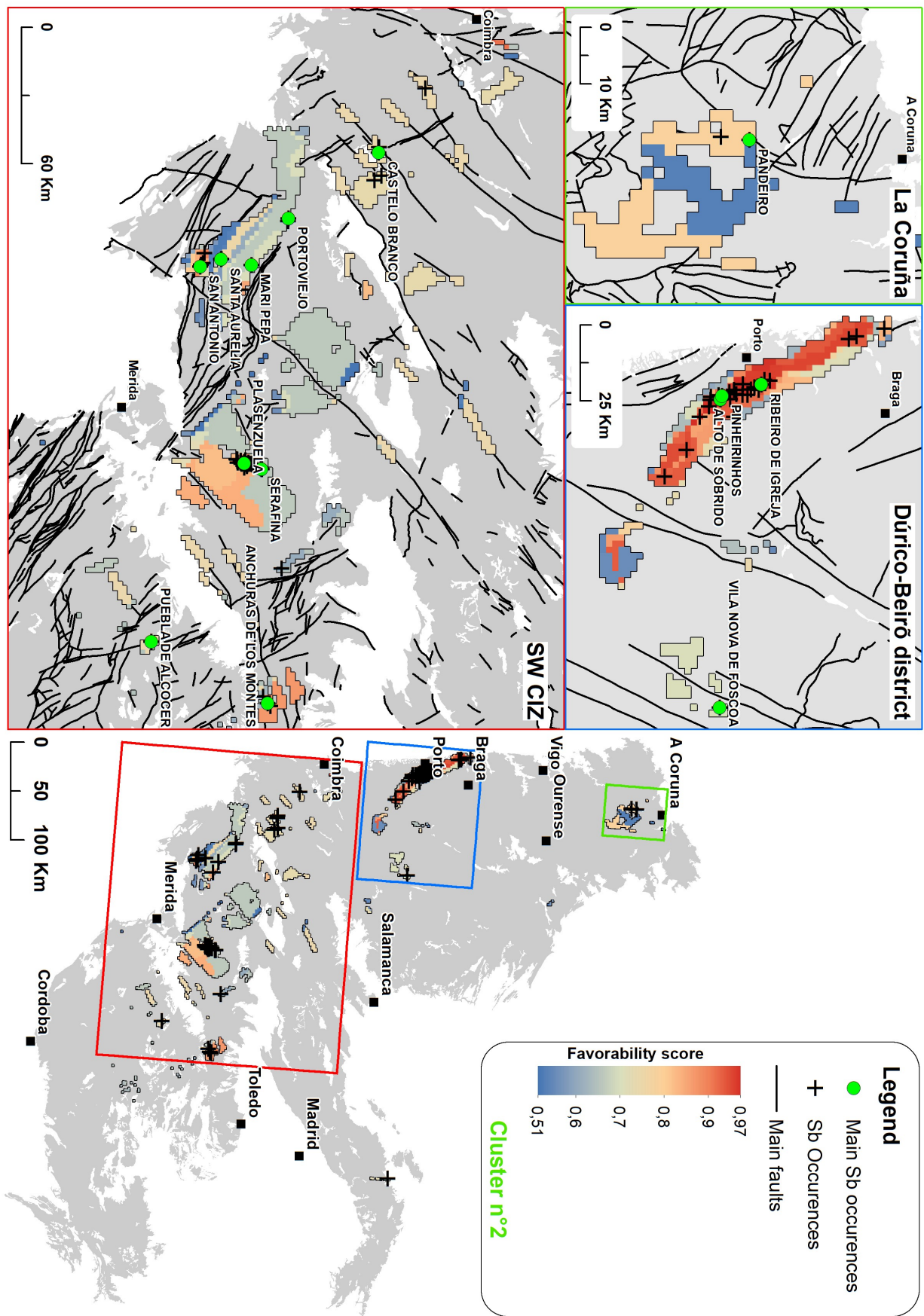


Figure 7.23: Favorability map of antimony in the cluster 2 of the CIZ.

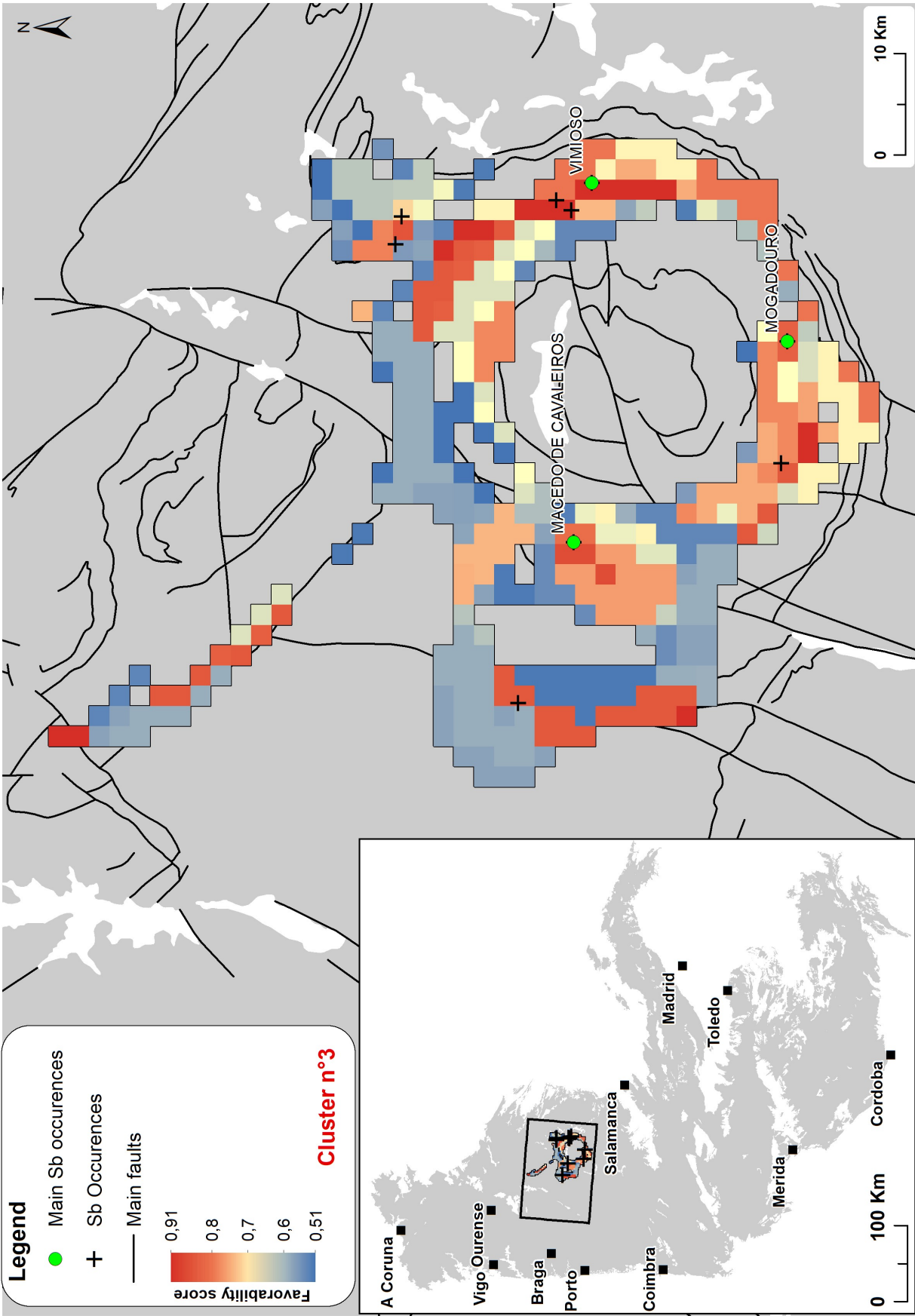


Figure 7.24: Favorability map of antimony in the cluster 3 of the CIZ.

Cluster 3 comprises 383 cells, 7.8% of the favorable area. It contains 29 or 10.1% of the mineralized cells of the favorable area, 7.6% of its cells being mineralized. The average favorability score inside this cluster is 0.68 ± 0.11 . A detailed map of the favorability scores inside this cluster and the antimony occurrences and deposits within it is displayed in **Figure 7.24**.

The lithology inside this cluster is dominated by the Ordovician to Devonian schists, quartzites and volcanic rocks of the GTMZ. The 3 main lithologies inside this cluster are from the basal units of the allochthonous complex of the GTMZ, with the formation "53", with schists and vulcanites, corresponding to 94% of the cells, the formation "28", made of phyllites, quartzite and vulcanites account for 92.7% and the formation "23" with schists, graywackes and vulcanites, accounts for 82%. All 3 of them are only comprised in this cluster. Faults are more frequent, as they are contained in 78.3% of the cells. More details about the frequency of lithologies in this cluster are available in **Appendix C.9**.

This cluster is located to the NE of Portugal, in the GTMZ domain and contains 3 main mineralization with the Vimioso, Mogadouro, and Macedo de Cavaleiros vein-type Sb occurrences, forming the Bragança district [Neiva et al., 2008]. This cluster is very localized, with the mineralization being located in the basal and parautochthonous units, forming a ring around the ophiolitic units of the GTMZ. Due to its localization inside the GTMZ and the specificity of the lithological association inside this area, which are separated from the main CIZ lithology in the geological map of the Iberian Peninsula, it presents a very singular signature compared to the other clusters.

Cluster 4

Cluster 4 comprises 94 cells, 1.9% of the favorable area. It contains 19 or 6.6% of the mineralized cells of the favorable area, 20.2% of its cells being mineralized. The average favorability score inside this cluster is 0.76 ± 0.14 . A detailed map of the favorability scores inside this cluster and the antimony occurrences and deposits within it is displayed in **Figure 7.25**.

This cluster is mainly made of Carboniferous volcanic rocks, schists and sandstones with faults. Most cells contain faults (93.6%), and 3 of the most important lithologies are Carboniferous formations, with the formation "76" with schists, sandstones and limestones (89.4%), the formation "83" with sandstones and conglomerates (86.2%) and the formation "76v" with vulcanites (83%). This cluster also contains Neoproterozoic basement units with the formation Negra "42", with schists and graywackes (84%). More details about the frequency of lithologies in this cluster are given in **Appendix C.10**.

This cluster is located in the Southern part of the CIZ, with mineralization belonging to the Cordoba province (El Membrillejo, Espiel, Sierra Morena). Like the previous one, this cluster is very localized and presents a different signature than the other, as it is localized near the Ossa-Morena area. Despite its limited map extension, it is not grouped with the other clusters as its distance is greater in the MCA coordinates system (see **Appendix C.6**).

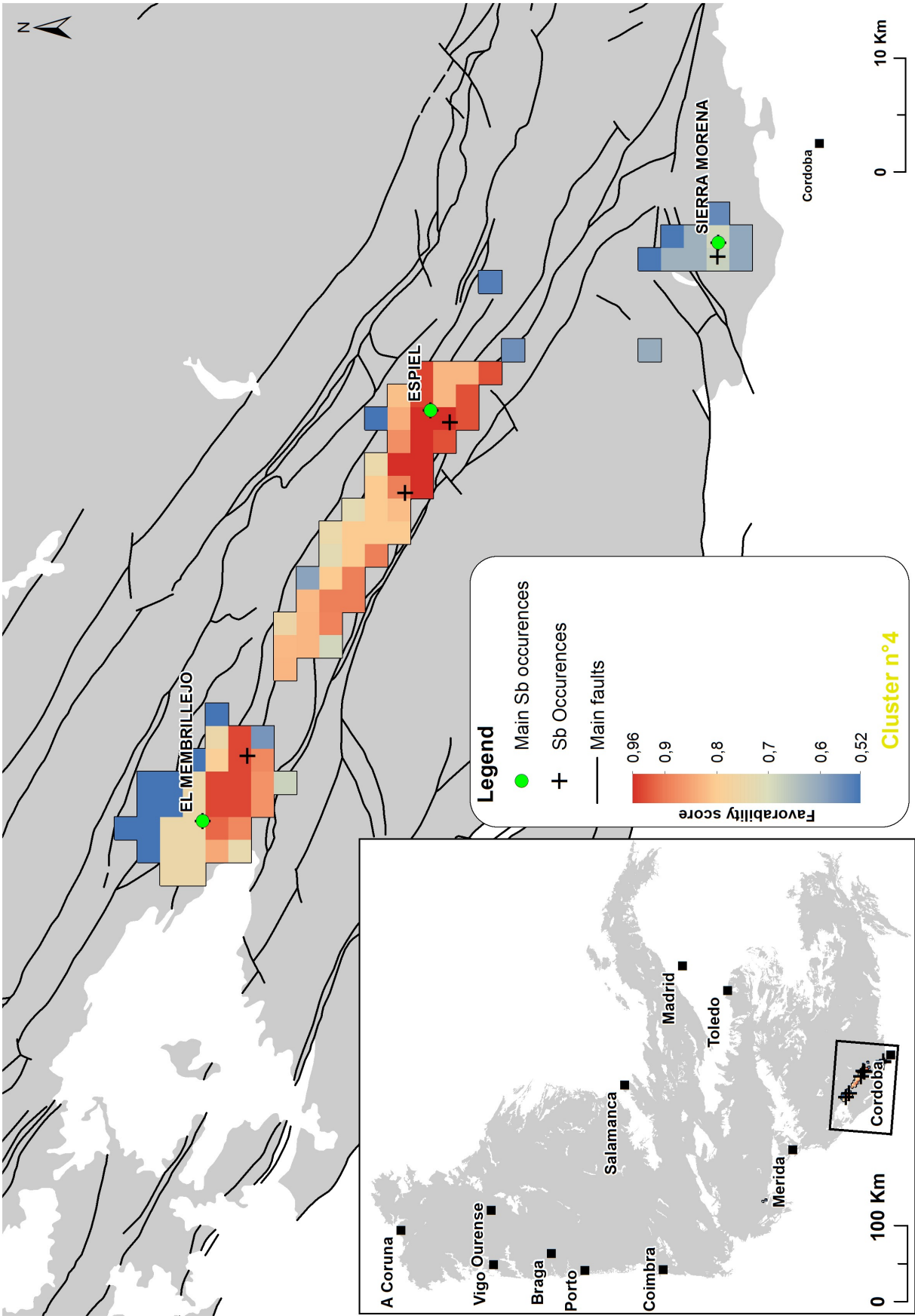


Figure 7.25: Favorability map of antimony in the cluster 4 of the CIZ.

Part IV

Conclusion and Annexes

Chapter 8

Discussion and synthesis

*This final **Chapter 8** synthesises and discusses the developments of the DBA / RF / HCPC methodology displayed in **Chapter 3**, **Chapter 4** and **Chapter 5**. The methodological developments are first analysed in regard to the main questions drawn in the **Introduction**, with the identification of geological environments, the evaluation of the relationships between geological environments and mineralization and the identification and synthesis of mineralized environments. The general workflow of the DBA / RF / HCPC methodology is presented in detail. Results of the applications of this new methodology, described in **Chapter 6** and in **Chapter 7** are summarized in regard to the general frame and context given in **Chapter 1** and **Chapter 2**. Automation of this methodology is also presented, regarding the tools used and possible ways of optimizing the various required processes. New development paths are also presented, to either enhance this methodology or apply it to other fields of applications. A short conclusion, summarizing the main points raised during this discussion is presented at the end of **Chapter 8**.*

8.1 Methodological developments

8.1.1 Spatially based integration of the geological environments: the Disc-Based Association

The Cell-Based Association (CBA) and Disc-Based Association (DBA) are more spatial than statistic methods. While statistic methods focus more on a global analysis of the trends between the predictors and mineralization, those spatial approaches focus on the local spatial associations of geological variables and features of various natures to describe the relationships between the predictors and the mineralization. Such associations are referred to as "geological environments" which allows describing various map features located in the same surrounding. In this study, those geological environments are mainly composed of lithological and structural features, but this concept also encompasses other data such as geophysical or geochemical anomalies, for instance. The CBA and DBA methods are focusing more on a local approach when compared to other favorability mapping methods, which are more integrative. For example, the well known Weight-of-Evidence (WoE) method, which principles are detailed in this study directly integrates the whole lithological dataset, all over the study area, to compute scores (and, thus, favorability maps). Usually, predictive mapping methods process the lithological dataset in an integrative way, using identified favorable lithological and structural features to produce new datasets.

Those targeted features are selected and processed into discrete or continuous datasets, using for example buffers or distance analysis (e.g. distance to permissive lithologies, distance to structures, contact between lithologies, etc.) [Carranza and Laborte, 2015b, Parsa et al., 2022, Maepa et al., 2021, Xiong and Zuo, 2021, Torppa et al., 2019, McKay and Harris, 2016]. The produced datasets are then integrated in the predictive model. Here, the lithological and structural data are analyzed locally and not considered over the entire study area. The multivariate information (i.e. various data layers) is integrated into a cell grid, defining local geological environments evenly spaced across the study area. This approach based on the identification, in a given study area, of the geological environments (i.e. associations) can be performed at multiple scales. In this work, we aim at defining the highest possible resolution to produce precise favorability maps, with the smallest cell sizes possible. However, this approach based on the automated identification of geological environments (i.e. map features associations) can be performed at various scales from the same study area. If this aspect is not exploited here, it could be interesting to explore the possibility of this new approach to study the evolution of the geological environments and favorability maps depending on the scale of the local geological environment observed.

The new DBA method has been developed during this study from the CBA method from Tourlière et al. [2015]. The CBA was developed with the aim of better integrating geological maps data in the calculation of mineral resources prospectivity maps, by discretizing them into a multivariate matrix representing local geological environments. It uses a regular square grid cell to integrate the intersecting map features into the multivariate matrix. It allows overcoming some of the problems inherent this type of dataset, such as (i) inaccuracies in dataset contouring and locations, (ii) variations of the outline of map features at different scale, (iii) presence of superficial cover unrelated to the studied event, (iv) inappropriate generalization of a phenomenon and (v) clustering effects. However, it comes with some limitations. First, its final grid resolution is limited by the size parameter d of the grid and its relationship with the resulting local geological associations. Second, the anisotropy of the square cells used by the grid introduces a bias in the identification of the geological associations, which depends on the initial position of the grid and the preferential orientation of the lithologies inside the dataset.

The DBA method has been developed to overcome the limitations of the CBA, while keeping its ability to overcome the problems inherent to the studied dataset. Instead of using a regular grid cell to identify the lithological associations, it relies on (i) the projection of a raster grid of square cells, with a given resolution d over the area and (ii) a neighbouring search performed with a radius R , constituting a search disk around the cell centers and identify the intersecting map features. Those are then integrated into the raster grid and its multivariate matrix, resulting in the local geological environments. This development allows to overcome the anisotropy of the CBA, as the neighbouring search is done using discs, an isotropic shape. It also allows an improvement of the resolution of the final grid of the associations: the final resolution is determined by the value of d , but the relationship between the geological association and the final grid depends on both R and d values. Lowering d allows to have a higher resolution, and it is then possible to modify R to change the properties of the geological associations found by the DBA.

Overlapping between the discs used to "explore" the neighboring around cells centers plays a significant role in the count and the diversity of the associations found by this method. A low overlapping (small radius) can lead to some parts of the map being untreated, while a high overlapping (large radius) would result in associations with very little meaning as they would integrate map data from too large surroundings and, then, wouldn't be specific enough for the classification to be meaningful. As such, the production of an abacus allows quantifying this overlapping, determining a minimum threshold to ensure covering the whole study area ($\sqrt{2}/2 = 0.71$) and linking the R/d ratio to the cumulated number of the

overlaps and corresponding surfaces. It is hard to evaluate the impact of the overlapping on the quality of the resulting DBA grid, as those means that some areas, mostly the one located at the border of the cell, will be taken into account in the geological associations more frequently than the one located near the center of the discs. Reducing this overlapping would be possible by distributing the point more evenly on the studied area, using a hexagonal grid or a square grid based on a hexagonal distribution [Birch et al., 2007] to determine the center of the points instead of a regular square grid. However, despite the advantages of a hexagonal grid [Brimkov and Barneva, 2001] this would induce more complexity when processing the matrix or using it in combination with other datasets, more frequently based on grid square, up to now.

The *Dispersion*, based on the *Median*, *Maximum* and *Minimum* number of overlapping, has been tested to estimate the variability of the overlapped surfaces of the discs. The differences between the maximum and minimum number of overlapping (*Range*) estimate the differences between the most investigated area and the least investigated one. This value is divided by the *Median* to compare this difference to the number of overlapping over the whole area. This *Dispersion* metric mainly shows that at low R/d value, the overlapping differences are important when compared to the *Median*, while increasing the R/d will decrease this value, meaning that the differences between the overlapped areas have less and less importance regarding the total number of overlapping. While low *Dispersion* values can be reached with high R/d ratios, we recommend to avoid reaching those values as much as we avoid low value of R/d ratio. The *Dispersion* curve is mainly used to determine specific values of R/d where the *Dispersion* reaches low values, allowing to diminish the overlapping differences.

With the DBA method, it is also more complex to rationalize the optimal parameters to produce the grid, as two parameters (d and R) have to be defined. It also brings a slight increase in computation time, the processing of the dataset requiring more steps. As a matter of fact, the DBA method still requires many iterations to determine the optimal inputs. Automation of this process allows to facilitate the determination of those optimal parameters, but in the end, those parameters are defined manually by the operator depending on the results of this analysis and the target of the study.

8.1.2 Evaluating the likelihood of favorable geological environments: the DBA / RF favorability mapping

The Random Forest (RF) method has been used in numerous Mineral Predictive Mapping (MPM) studies and has demonstrated its efficiency [Sun et al., 2020, Carranza and Laborte, 2015a,b, Rodriguez-Galiano et al., 2015, Carranza and Laborte, 2016, McKay and Harris, 2016, Zhang et al., 2016, 2019, Daviran et al., 2021]. It allows, like other Machine Learning (ML) methods, to evaluate the importance of the predictors used by the predictive model [Sun et al., 2020, Maepa et al., 2021, Carranza and Laborte, 2015b, 2016, McKay and Harris, 2016] enabling better interpretability of the results. Neural Networks (NN) methods have also been often used recently to perform MPM, giving results as promising as the RF methods [Maepa et al., 2021, Parsa et al., 2022, Köhler et al., 2021, Rodriguez-Galiano et al., 2015]. However, NN studies result in an estimation of the favorability but with very poor outcomes on the associations scheme leading to mineral deposits; these are clearly much less easily interpreted in terms of ore deposits mechanisms than the results obtained from RF [Zhang et al., 2021]. RF, outside of MPM, gives results comparable and sometimes better than the NN methods [Ahmad et al., 2017, Nawar and Mouazen, 2017, Fernandez-Delgado et al., 2014] with an easier interpretability compared to the NN, even if it is still discussed [Aria et al., 2021, Bénard, 2022]. In our case, the interpretability of the model is established

on its capacity to give insight about the relationships found between the geological environments and the mineralization. We rely on the evaluation given by the RF model of the predictors used and on the analysis of the final favorability maps to highlight such links.

This work shows that the association of the DBA pre-processing method with the RF predictive model is able to produce relevant favorability maps for antimony, based on lithological data. Using the DBA to homogenize the lithological data into a grid showing their local associations, it is possible to use the RF to evaluate which associations show the most likelihood to the ones present around known mineralization. The use of the DBA / RF to produce a predictive model allows a more ‘purely’ Data-driven approach, requiring fewer expert inputs to perform the predictive mapping. Most of the human interventions needed during the DBA / RF process is linked to the definition of the parameters of the DBA, searching an optimal R/d ratio as a compromise between higher resolution and more diverse geological associations (see details in the preceding section). By minimizing the amount of metallogenic data to be injected in the predictive model, this Data-driven approach allows to stray from the metallogenic models previously established and published on the basis of more traditional and sometimes local metallogenic studies. Relying less on the guidance of a conceptual model to produce the MPM, we can highlight "statistical" metallotects which could help produce a "statistical" conceptual model, defined from the analysis of the data rather than existing knowledge. This approach aims at enriching these models through spatial and statistical analyses rather than having to rely on their knowledge to produce favorability maps. Of course, the DBA / RF predictive mapping still requires the selection of reliable and meaningful data before the computation of the model to perform efficiently, but, compared to the ‘purely’ Knowledge-driven approaches, the operator does not need here assigning weights to those data, as the predictive model will automatically define them.

In the case of the DBA / RF, all non-mineralized cells are used in the classification. This feature clearly differs from the other methods based on ML, relying on handpicked non-mineralized area to overcome the imbalance between the non-mineralized areas and the mineralized ones [Carranza and Laborte, 2015a, 2016, Daviran et al., 2021, Rodriguez-Galiano et al., 2015]. This selection process is less subjective. Recently, studies on some ML methods tried to overcome this imbalance using methods such as Synthetic Minority Over-sampling Technique (SMOTE) to over-sample the mineralized area [Prado et al., 2020]. In our study, all cells that do not contain a known mineralization are considered non-mineralized (or True Negative) during the learning stage and are used in the classification process. Despite this large imbalance, the DBA / RF method is able to correctly predict the mineralized areas and to highlight favorable ones located outside of known mineralized districts too. This imbalance is accounted for in the RF model used, as it relies on in-built tools such as balanced sample and stratified bootstrapping. The former automatically adjusts weights inversely proportional to class frequencies in the input data for each bootstrap sample used by each tree. The latter allows each tree to perform the classification with the same proportion of each class (i.e. mineralized vs. non-mineralized) in its test and training samples, in the same way the Stratified Cross-Validation (SCV) does. This avoids determining non-mineralized areas, which depending on the method used can induce some bias on the model (e.g. in the case of insufficient data to determine areas where no mineralization should occur).

In the applications of our new approach to the Armorican Massif (AM) and Central Iberian Zone (CIZ), most of the known antimony occurrences have been highlighted as favorable areas. Better results could be achieved with the addition of more data correlated with the Sb mineralization, such as geophysical surveys, as used by Gloaguen and Tourlière [2016a] for their prospectivity maps. Geochemical or more varied structural data could also be used, as antimony deposits can show strong correlations with hydrothermal

activities. An advantage of combining RF with the DBA is its ability to process multiple types of datasets. RF is able to perform classification on either continuous or discrete data or a joint combination of the two kinds of datasets. In this study, we mainly rely on discrete/qualitative data, with lithological formations and structural features extracted from the geological map. However, as shown in [Moon and An \[1991\]](#), [Rodriguez-Galiano et al. \[2015\]](#), [Gloaguen and Tourlière \[2016a\]](#), [Riahi et al. \[2021\]](#); geophysical data can be very useful to evaluate prediction scores for some types of mineralization. The addition of continuous data will be a direct extension of this thesis work.

When compared to previous studies across the AM [[Gloaguen and Tourlière, 2016a](#)] and to the application of other prospective mapping methods, such as the WOE, the Ranking method and the AHC clustering method, the DBA / RF method clearly performs better. The favorability maps produced show its ability to better predict the mineralized areas and to restrict the number of False Positives (FP), of potentially mineralized areas. Having a more restricted favorable area allows determining more precisely the potentially mineralized areas, the areas to prospect when searching commodities. In the case of this study, being able to more accurately delineates the areas showing the highest likelihood to host mineralization also allows to more precisely determine their relationships with their geological environment s.l. If larger areas were to be considered, the resulting geological environments found by the HCPC method would include more cells, more lithological associations and could integrate geological environments that would have no link with the mineralization. Besides, this spatial precision is also a great advantage for minerals targeting, as it allows better targeting of permit applications and exploration campaigns, and therefore a strong optimization of the cost/benefit ratio in the early stage of exploration.

In order to obtain proper results, the application of DBA for prospective mapping purposes requires cartographic datasets that are homogeneous throughout the studied area. This means that the lithological data must have been harmonized at the overall scale. A harmonized dataset implies that all lithological formations of the same nature share the same name or description, which is not always the case if the dataset used is a collection of multiple cartographic datasets produced with different standards. This is an important part of the process as DBA / RF applied to a heterogeneous cartographic dataset will give incorrect results. If not available, harmonized datasets have to be produced, which can be costly as it requires a lot of processing time and expert knowledge.

8.1.3 From favorability mapping toward automatic metallotect detection

The DBA / RF predictive modelling approach developed here is creative in that it allows defining automatically geological associations of the mineralization, giving further insights about the relationships between the results of the model and the input data. Compared to the alternative predictive models shown for the AM (i.e. WoE, CBA / Ranking and CAH), the RF classification is able to estimate the importance of each predictor derived from the input datasets, using the MDI score associated. The highest the MDI value of a predictor is, the more frequently it is used in the RF classification and the better it is to differentiate mineralized cells and non-mineralized ones. This means that high MDI predictors are most likely linked to the presence or absence of mineralization, as MDI scores evaluate the ability of a predictor to discriminate the two classes during the RF classification process. The interpretation of the RF model gives some feedback about the link between the input cartographic data and the favorability scores, as the highest MDI predictors can be considered as proxies for the occurrence of mineralization and as possible metallotects. Those predictors could give more insights about the relationships between the mineralization and their geological associations, highlighting specific geological features that tend to

be linked to the mineralization. The highest MDI features can display different kind of relationships with the mineralization as seen in the applications: some features are localized and related to specific districts and mineralization, while others look more widespread and are linked to all or the majority of the districts in the considered area. Those predictors could also be used as potential metallogenetic for further expert-guided studies in areas where datasets are inappropriate for Data-driven methods (e.g., insufficient number of known deposits). The integrated CBA / RF method could thus be used in areas where data are available but few "traditional" metallogenetic studies have been made to automatically identify associations between mineralization and the gathered data. However, "metallogenetic" identified through this automated process are only "suggestions" of potential metallogenetic, which must be critically evaluated and reviewed by metallogeneticists before being validated and implemented into other methods, other approaches or for others case studies.

8.1.4 Identification of the various favorable environments with HCPC clustering analysis

While identifying unique geological features as targets or metallogenetic linked to the mineralization is possible with the evaluation of the predictors importance, the analysis of the association of map features surrounding the favorable areas is also of prime interest. However, analyzing the whole favorable area would result in mixing the geological associations of various types of antimony mineralization over all of the study area. Indeed, all antimony mineralization were targeted to produce the favorability map, regardless of the type of deposits or the related metallogenetic event. This is why we rely on an EDA method, the Hierarchical Clustering on Principal Components (HCPC), to differentiate the various geological environments in the favorable areas. Instead of relying on unique geological features, we can investigate the possible relationships between multiple variables and the mineralization. Those relationships can complete the "statistical" conceptual model mentioned above in addition to the "statistical" metallogenetic identified through the DBA / RF, with different types of mineralization associated with different geological environments (i.e. specific association of spatial factors) identified by the HCPC analysis. This requires the selection of a specific area to analyze these relationships. This selection is performed from the favorability map produced by the predictive model.

The use of the HCPC unsupervised classification method [Argüelles et al., 2014] allows differentiating various geological environments amongst the favorable areas identified by the RF predictive model. Compared to the AHC [Rolet and Seguin, 1986a,b] method previously associated with the CBA and DBA, the HCPC is a strong improvement as the application of an additional factor analysis to the clustering process allows a more coherent delineation of the clusters, with more weight attributed to unique predictors or specific geological associations, whereas the AHC was only based on the differences between each geological association. Each cluster produced by the HCPC displays different geological associations that can be attributed to different types of mineralization or different known districts across the study area. This process avoids mixing geological associations that could be attributed to different types of mineralization. Each cluster can be analyzed in details to outline specific associations of predictors and potential metallogenetic. However, an analysis of the relationships between specific typologies of mineralization and geological environments could be planned. Separating the various typologies of mineralization when performing the DBA / RF / HCPC could better delineate lithological associations that are linked to specific type of mineralization. The various types of mineralization could be identified during the production of the DBA matrix, and the RF / HCPC process could be done using mineralized cells where specific

typologies occur. Results from this approach could be compared to the one presented in this study where all types of mineralization are integrated during the RF / HCPC process, as the RF favorability map is produced on all known mineralization, without selecting specific typologies.

Results from the HCPC classification, which highlight specific environments surrounding mineralization, can also benefit from a deep analysis and comparison with known metallogenic models for antimony. This HCPC classification has been able to highlight the geological frameworks associated with the most favorable areas. These geological frameworks present similarities with the known districts, which looks rational as each district is characterized by specific relationships between the geological features and the mineralization. The application of the HCPC classification and the analyses of its results requires a lot of human interventions to process the data, such as the determination of the optimal number of clusters or the identification of a specific geological signature within each cluster. If the interpretation of its results should inherently be performed by an expert, the developments of statistical analysis to produce unique lithological signatures linked to each class would make this interpretation easier. With it, the operator could concentrate on the geological analysis, such as the geological interpretation of the environments within each cluster or the evaluation of the relationships between clusters and known mineralization and districts.

8.1.5 General Workflow of the methodology

The general workflow of the new methodology developed in this thesis is presented in **Figure 8.1**. This methodology relies on a combination of three methods: the Disc-Based Association (DBA), the Random Forest classification (RF) and the Hierarchical Clustering on Principal Components (HCPC). The association of those methods allows the analysis of the relationships between mineralization and geological features, while delivering favorability maps to identify the areas with the most probable exploration interest. It is very important to follow the sequence of application of these three methods: first the local geological environments are identified by the DBA, then their relationships with the mineralization are evaluated with the RF predictive model and finally the favorable geological environments are discretized and characterised through HCPC.

The available cartographic datasets across the study area are selected depending on their relevance. They can be of various natures, with discrete ones containing surfaces, lines and points together with continuous datasets. DBA is applied on all the selected datasets to produce the local geological environments. This produces a grid and an associated matrix of attributes presenting for each cell the geological features that are present within it, their association forming the local geological environment. The parameters defining this grid are optimized depending on the characteristics identified by the DBA.

Once the matrix and the grid have been produced, the RF predictive model is applied to the matrix and grid to produce the favorability map. As for all predictivity mapping approach, a favorability score is computed for each cell of the map grid but, in addition, geological environments having the highest likelihood to host mineralization can be recognized while scoring the importance of each geological feature (i.e. predictors) according to the model. The best parameters to produce the predictive model are then automatically defined and optimised. This method is more effective than the Ranking method previously used in combination with the CBA or the DBA. The most favorable areas are selected by evaluating the predictive model.

The HCPC is applied on this selection to classify the favorable geological environments. The HCPC

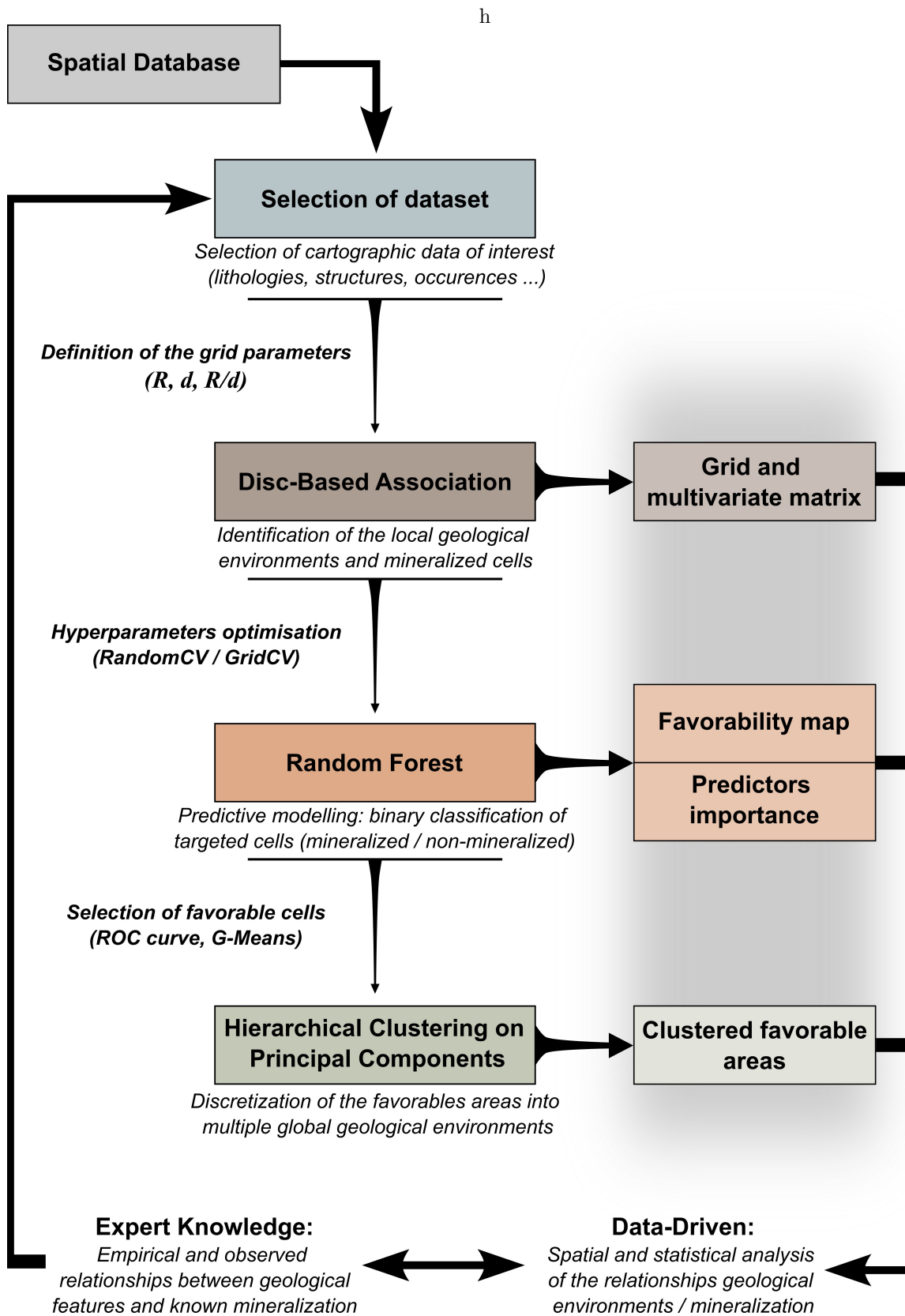


Figure 8.1: General Workflow of the developed DBA / RF / HCPC methodology.

gives results allowing to segment the favorable areas into clusters; their number is defined by the operator depending on the knowledge of the study area and the objectives of the study. The HCPC is an evolution of the AHC toward a more consistent analysis of discrete variables. The results obtained at the different steps of the process give important insights on the spatial and statistical relationships between the geological features s.l. and the mineralization. The relationships found through this method can complement existing knowledge about mineralization processes and metallogenic models, determining various specific associations of map features spatially associated with the mineralization, and help defining a "statistical" conceptual model to guide future applications or complement existing ones. Results of this method have to be verified with expert knowledge to assess its validity and identify those relationships, by analyzing the environment within the clusters and comparing it to existing knowledge about the mineralization and districts associated with it.

8.2 Insights from the geological applications

Predictors importance given by ML models have been used to evaluate the predictors processed by the methods and gives returns about the most effective ones. Studies have also shown the importance of using this parameter to better optimize the results given by ML methods [Wang et al., 2022, Zhang et al., 2022]. In our study, we use the predictors importance to evaluate the possible metallotects, the geological features favorably linked with the presence of mineralization. However, while this value gives some insight on the geological feature / mineralization relationships, the identification of favorable geological environments is the real novelty with the DBA / RF / HCPC methodology.

The application of the DBA / RF / HCPC approach on selected case studies is able to highlight the favorability of antimony across the study area. The resulting computed favorability scores are linked to the likelihood between the various local geological environments identified by the DBA and the ones that are mineralized. As such, these favorability scores give a first insight about the relationships between the geological associations and the mineralization: the associations showing the highest scores are the most similar to the geological environments surrounding the known mineralization. The favorability maps highlight the areas showing strong similarities with the environments of the known mineralization.

The produced favorability maps display those high scores in specific areas, mostly around the known mineralization. Almost all the known mineralization are highlighted by the favorability maps produced, meaning that the model is able to correctly predict the mineralized cells most of the times. The selection and discretization of the favorable environments based on the ROC curves allow to target only the most favorable areas and study their properties. The discretization of the various geological environments into fewer clusters, representing global geological environments, allows the identification of various trends, various types of geological environments in these favorable areas. In the application shown previously, these clusters tend to correspond to specific districts of antimony and to areas displaying unique geological associations. Instead of relying on the favorability scores to identify a unique favorable environment, mixing various geological units and domains together, the use of an unsupervised clustering method allows the identification of several favorable environments and the identification of specific lithological associations within them.

With this method we were able to highlight different favorable geological environments across the study area. In the case of the AM, the Vendée and Quimper districts could be associated with one or two clusters, showing the ability of the method to highlight the specific geological environments of

the mineralization in these districts. In the case of the La Lucette, Le Semnon and Rosnoën districts, the best level of sectioning was not able to distinguish them, but their geological environments are quite similar. Overall we find lithological signatures typical of the different antimony districts, with the importance of volcano-sedimentary formations for the Vendée district, the association of faulted structures and Carboniferous granites for the Quimper district and the presence of dolerites and quartz veins in Ordovician to Carboniferous sedimentary environments for the La Lucette, Le Semnon and Rosnoën districts. Further analyses of the environments, with a greater number of clusters, could allow the identification of more specific favorable geological environments in this area, but still requires a lot of manual work at this stage. Separating the typologies of mineralization, between strata-bound and veins for example, during the HCPC clustering or at an earlier stage, when performing the DBA / RF, could also yield interesting results.

In the CIZ, the identified clusters have variable sizes, and present environments that are above all marked by the domains in which they are located. This is especially the case for clusters 3 (**Figure 7.24**) and 4 (**Figure 7.25**) which are located in lithologies specific to certain domains of the Iberian Massif (IM), in this case the GTMZ and the northern boundary zone of the Ossa Morena. In this study based on the Spanish 1:1,000,000 scale geological map, the mineralized zones and the different districts are pretty well highlighted at the scale of the domains of the study area. Still, it should be noted that it seems more difficult to target more specific districts within the CIZ s.s. Those districts are grouped in two massive clusters, making it difficult to perform a fine differentiation of specific mineralized districts, like the district of Dúrico-Beirõ in Portugal. It is possible that the results could be improved by limiting the application of the method to the geological units of the CIZ domain s.s. to specifically identify the mineralized districts. One could also alternatively use more precise data, at 1:200,000 or 1:50,000 scale to have more numerous and varied lithologies over the study area. As for the AM, the possibility of automating the determination of the cut-off value of the HCPC to allow a larger number of clusters would be an interesting option for a more detailed analysis.

8.3 Automation of the methodology

One of the objectives of this thesis work was the automation of the favorability map creation process, from the creation of the DBA matrix, the creation of the predictive model to the analysis of the favorability map and the production of the geological environments. This process must be Data-driven, minimizing expert inputs and allowing for statistical evaluation of the data. The CBA / AHC and CBA / Ranking methods that served as the basis for the development of this DBA / RF / HCPC methodology were originally very poorly automated, with most operations (except for the AHC) requiring manual processing on GIS software and spreadsheets. Several *Python* algorithms were thus produced as part of the development of our methodology to automate as much as possible the data processing and the production of favorability maps and clusters. One of the interests of this automation is to allow "non-experts" to make predictive maps, leaving the interpretation to experts and allowing to save time. This automation is also a way to optimize exploration costs for a private actor.

The use of *Python* and its associated libraries has made it possible to automate most of the data processing. Nevertheless, the selection and pre-processing of the data is done manually, and could hardly be automated. The creation of the DBA grid is largely automated, requiring only to determine a range of values for the parameters to be tested (R and R/d). The algorithm then produces the different graphs

allowing to evaluate the grid according to the indicated parameters. The operator must select the optimal value(s) of the grid to be used for the predictive model, according to the graphs produced and the data from the overlap abacus (*Dispersion*, number of overlaps *Maximum*, *Minimum*, *Average* depending on the value of R/d). Adding new data to the DBA grid, such as lithologies or mineralization, requires that the search distances R to be used be determined beforehand, with the d value of the grid remaining unchanged. Automation of the creation of the DBA grid would require establishing metrics that allow to evaluate the pertinence of the grid regarding the study area, which is not an easy task as even with the graphs produced multiple choices of combination of R and R/d can be of interest.

The calculation of the predictive model from the DBA grid is also automated in large part thanks to the *Scikit-learn Python* library. The definition of the hyperparameters is automated with the *Random Search* method. The main contribution of the operator is to determine the number of random searches to be performed to determine the optimal hyperparameters and the number of folds of the cross-validation. Several models can be produced, the results of the RF varying according to the iterations. If too much variability is observed, it may be necessary to increase the number of iterations of the random search.

The production of clusters by HCPC is a step requiring a lot of expert analysis and manual processes. Indeed, if this method allows to automatically discretize our data by grouping them according to their "similarity", the final number of clusters must generally be determined manually based on geological knowledge as well as on the statistical properties of the clusters. It would be interesting to optimize and automate this process at least partially. The validation of the geological environments has to be done by an expert, but an automated process could be developed to identify the most important lithologies inside the clusters. It could be evaluated in term of frequency of this lithology or in its relative proportions across the favorable or the whole study area. On the other hand, the selection of the favorable cells that serve as a basis for the production of the cluster can easily be automated with the use of the G-Means or the J-Score to determine the optimal threshold of favorability score to perform the selection.

Some optimizations of the method could allow for more automation and less expert inputs. Performing DBA and RF together to produce predictive models could allow direct determination of the combination of DBA grid and predictive model providing the best estimate of the study area. This approach would be much more time and energy consuming but would avoid having to identify optimal grid parameters manually and would likely results in more efficient models. However, it would still be necessary to verify the consistency of the grid used with the study area.

8.4 Ongoing and future developments of the methodology

8.4.1 Integration of continuous datasets

The approach developed in this thesis relies on the integration of discrete variables into the DBA matrix. However, continuous datasets are commonly used to produce favorability maps in MPM including geophysical, geochemical, or even hyperspectral remote sensing ones [Rodriguez-Galiano et al., 2015, Daviran et al., 2021, Riahi et al., 2021]. Continuous data can also be produced from initial discrete features dataset, with the production of factor map, such as, for instance, computing the distance from specific lithological features or structures into continuous raster grid. Integrating such continuous data into the multivariate grid of the DBA matrix would allow having more features integrated into the local geological environments produced, giving a more detailed insight about them and giving more relevant

predictors to process for the predictive model. This feature would allow a great flexibility in the application of the DBA / RF method and the use of more data to produce favorability maps, while remaining as Data-driven as possible and minimizing expert inputs in the pre-processing stage of the predictive model.

The integration of existing continuous dataset into the DBA grid would require either a discretization, as the one done in the previous studies presented in **Subsection 6.1.2 p.71**, or a re-sampling of the data in order to match the resolution of the DBA grid (**Figure 8.2.a**). This resampling can be done using the final grid produced by the DBA. If continuous datasets are to be produced from the initial ones, their processing can be done based on the resolution and location of the DBA grid too (**Figure 8.2.b**). Integration of continuous data into the DBA matrix would however change its nature. If the RF predictive model is able to handle continuous data, the evaluation of the predictors importance could not be done using the MDI, as this metric is susceptible to high cardinality categorical variables and to continuous ones. Those variables tend to be over-represented in the decision trees produced by the RF, resulting in high MDI scores for this variables kind. Permutation importance [Breiman, 2001] is another method used to evaluate the importance of the predictors used by the RF, which is not susceptible to high cardinality or continuous dataset and could thus replace the MDI method applied here. The HCPC method would also be less effective, as the MCA is a factorial analysis targeting discrete data. However, it is possible to perform a factorial analysis with mixed data (i.e. containing both continuous and discrete variables) with the Factorial Analysis of Mixed Data method (FAMD; Pagès [2004]).

8.4.2 Structural data and other geological features

Discrete data can be added to complete the geological environments characterised through the DBA method. For example, other geological features of interest could be used in addition to the lithological dataset, such as the metamorphic isograds or the alteration zoning when available, as those features can display strong relationships with some mineralization like antimony. The structural dataset used can be completed too, with the addition of other structures than faults, like folds or shear zone [Maepa et al., 2021]. In addition, while faults are integrated in the DBA with this current methodology, it only concerns the absence or presence of the faults within the local geological environments. However, structural features have more characteristics that can be of interest for MPM, such as their nature, orientation, their density within the local area or even their age [Parsa et al., 2022, Parsa, 2021, Carranza and Laborte, 2015b, Sun et al., 2020]. The latter can be one of the most important, as possible reactivation(s) can lead to a complex history for specific faults and families of faults. The complexity of faults networks can also be evaluated and is sometimes related to mineralization orebodies [Zhao et al., 2011]. Complementary information concerning the faults can be added into the matrix, allowing to study in more details the relationships between faults and mineralization. This could require either a pre-processing of the dataset containing the faults, in order to classify them according to their age and orientation, for example, or the integration of new tools inside the DBA method to identify these parameters in a specific local environment, such as their main orientation or their density. Both approaches are not equivalent: the first one means that the faults are analyzed on a global scale before being processed, while the latter computes their local characteristics to integrate them in the matrix. It is noteworthy that lineament study can also be performed to complete the known structural features at a regional scale [Yeomans et al., 2019]. Metallogenic data could also be added to enrich the geological environments. Multiple substances can be correlated, sharing similar processes, metallogenic events and geological environments, so it is possible to use deposits and mineralization of other substances as potential predictors for the

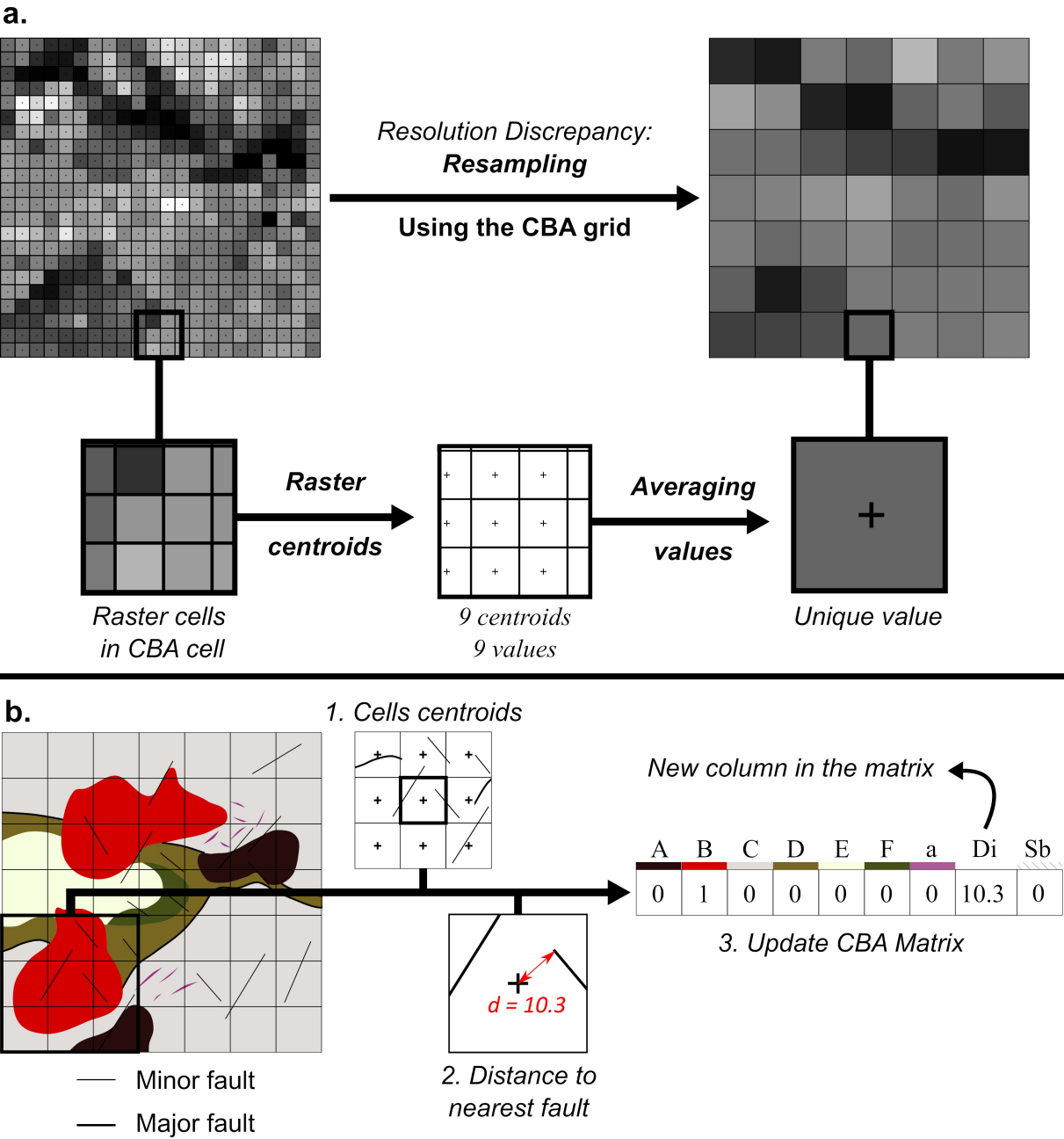


Figure 8.2: Example of integration of continuous data or predictive factors into the DBA matrix. a. Re-sampling (local averaging) of continuous data using the DBA grid b. Distance to faults computed from the DBA grid using shortest distance from faults to the center of the cell.

presence of the targeted one.

8.4.3 Harmonization of lithological datasets

The main limitation when using lithological datasets to produce favorability maps is the necessity of having a unified, harmonized dataset covering all the study area. Indeed, lithologies sharing the same properties must have the same denomination to be correctly integrated in the DBA matrix and treated as similar predictors in the predictive model. Such difficulty more particularly arises when applying this methodology to a vast regional study, like this one across the Ibero-Armorican Arc (IAA) where the study area covers three countries. Indeed, the lithological numerical datasets description of the geological units are most often (if not always) consistent for each country, with their own lithological descriptions and indexing applicable within their respective borders. It results that the raw dataset cannot be directly used and must be harmonized: lithologies sharing common properties must have a common denomination. This requires in the approach (i) the translation of the lithological descriptions, (ii) the harmonization of the lithological indexes and (iii) the processing of the shapefiles in order to produce a cross-border consistent map. Up to now, such harmonization tasks must be performed manually through experts knowledge and are a rather time-consuming work. Yet, some statistical automation processes could be considered for development to help comparing and/or grouping geological map features from two distinct databases, based on their characteristics. Indeed, we argue that the use of the DBA and HCPC developed here could be used to simplify lithological datasets; the DBA identifies the local association of lithologies, which is then simplified with the HCPC clustering analysis. The association of those two steps from our methodology could be used for the creation of simplified geological maps and their harmonization. The use of DBA and HCPC would allow to produce maps showing simplified geological environments. DBA and HCPC classification would be performed on various datasets to simplify the lithological data, and similar classes would be unified to form unique features across the datasets. This method is, however, actually far from being an automated harmonization process. In its current state of development, it would require the automation of the attribute tables processing and of the spatial analyses of the shapefiles. Moreover, other discrete digital data could be harmonized too, such as structural data. Information about faults types or kinematics may sometimes need to be harmonized, as well as their spatial distribution, with the problem of their continuity across the study area.

8.5 Concluding points

This thesis work aims at highlighting mineralized geological environments through a Data-driven predictive mapping approach, allowing an "inverse" approach to the creation of conceptual models for Mineral Prospectivity Mapping (MPM). A new DBA / RF / HCPC approach is developed combining space segmentation and local parameters associations (Disc-Based Association, DBA) with predictive modelling (Random Forest, RF) and hierarchical clustering (Hierarchical Clustering on Principal Component, HCPC) performed on these local associations signatures. As a proof of concept, this new method is applied to the study of the favorability of Sb within the Armorican Massif (AM) and Central Iberian Zone (CIZ) of the Variscan Massif.

- The development of the DBA / RF / HCPC methods allows analyzing the relationships between geological features and mineralization through the use of a Data-driven automated analysis. It allows

a more Data-driven approach compared to the previously used methodologies as the predictive model is able to identify relationships between mineralization and geological features rather than relying on pre-established statistical rules.

- From this work, we infer that the DBA / RF / HCPC approach would allow defining new undiscovered metallotects from its application on various natural cases.
- The DBA is an effective improvement of the Cell-Based Association (CBA) method with a better resolution and less bias regarding the anisotropy of the square grid.
- The RF classification, used as a predictive model, can be coupled to the DBA to produce favorability maps and give insight on the relationships existing between the geological environments identified by the DBA and the known mineralization used in the classification process. This DBA / RF method gives better results than previously used methodology coupling CBA and RF or CBA with Ranking and is less affected by clustering effects and surface-related biases.
- The HCPC is an improvement over the Agglomerative Hierarchical Clustering (AHC). It allows the definition and discretization of various favorable geological environments across the study area based on the results of a statistical analysis of the geological features and cells produced.
- The application of the DBA / RF / HCPC methodology on the AM and CIZ is able to highlight consistent relationships between the antimony deposits and the geological framework of those areas. Known districts and provinces of antimony are highlighted and discretized thanks to the predictive model and clustering method.
- The selection of non-mineralized targets is simpler with the DBA / RF method when compared to similar methods: all cells without mineralization are considered as non-mineralized and are used for the classification, allowing an automated selection and avoiding potential bias related to the selection process.
- Improvements can be made to this methodology to increase the degree of automation and to complete the geological environments processed by the DBA. With new developments, continuous datasets could be added to the DBA without necessitating previous discretization and further discrete geological information features could be added and processed.
- This new methodology developed in this work could be tested and used in applications other than MPM, such as land cover mapping or risk predictivity, as the DBA allows the combination of various datasets and its application is not restrained to geological datasets.
- With further developments, the DBA / HCPC method could be used to help in the harmonization of contiguous map datasets from various origins (such as cross-borders data sets).

Chapter 9

Résumé étendu

9.1 Contexte et objectifs de l'étude

Les gisements sont constitués de concentration élevée de minéraux ou de groupes de minéraux contenant des éléments d'intérêt économique. Leurs paramètres (par exemple, la taille, la forme, la teneur en minerai, les roches hôtes...) peuvent varier considérablement d'un gisement à l'autre, tout comme les processus qui ont conduit à leur formation. Ces gisements sont formés et préservés par la conjonction de nombreux processus/causes : une ou des sources métallogéniques, une voie de transport et de concentration des éléments d'intérêt, et des environnements favorables pour piéger les éléments et préserver la minéralisation de la remobilisation, de l'érosion ou de l'altération [Marcoux and Jébrak, 2008]. A ce titre, les gisements peuvent être considérés comme des événements rares d'un point de vue statistique / probabiliste. Ces gisements fournissent des substances essentielles pour l'industrie actuelle, dont les besoins ne font que croître, augmentant la quantité et le nombre d'éléments utilisés en industries, notamment avec le développement des hautes-technologies [Legaz et al., 2021, Charles et al., 2017]. Certaines substances deviennent critiques pour l'économie des pays en raison de leurs importances industrielles et technologiques, l'Union Européenne a ainsi identifié une liste de 30 substances critiques (Critical Raw Materials, CRM) pour ses membres [Bobba et al., 2020, Blengini et al., 2020].

Pour ces raisons, la recherche de nouveaux gisements est un problème critique, les découvertes devenant de plus en plus complexes et coûteuses [SP2, 2021]. Cette difficulté a mené aux développements de nouvelles disciplines cherchant à identifier les zones les plus favorables à l'exploration pour l'industrie minière et ce dès le XVI^{ème} siècle (1556 : "De re Metallica ex libre" de Georgius Agricola; [Barth et al., 2017]). L'exploration minérale actuelle peut compter sur de nombreux outils pour appuyer cette recherche avec l'utilisation d'imagerie ou d'analyse géochimique, géophysique, terrestre, aérienne ou même satellitaire. Cette recherche de nouveaux gisements est aussi appuyée par une autre discipline : la métallogénie, définie en 1905 par Launay comme étant "la recherche des lois qui ont présidé à la répartition, à l'association ou à la séparation des éléments chimiques dans les parties abordables de l'écorce terrestre" [Orcel, 1954]. La métallogénie cherche ainsi à étudier les relations entre les minéralisations connues et les processus géochimiques ou géologiques afin d'identifier des motifs récurrents entre des minéralisations de même nature [Marcoux and Chauris, 1994]. Ces observations mènent aux développements de modèles métallogéniques décrivant la genèse d'un type ou d'une famille de gisements, à partir de gisements considérés comme "type". Les motifs récurrents, reliant favorablement certains objets géologiques à la

présence de minéralisations, sont appelés "métallotectes". L'exploration minérale repose ainsi beaucoup sur l'utilisation de ces métallotectes et des modèles métallogéniques les utilisant pour définir des zones favorables à l'activité minière.

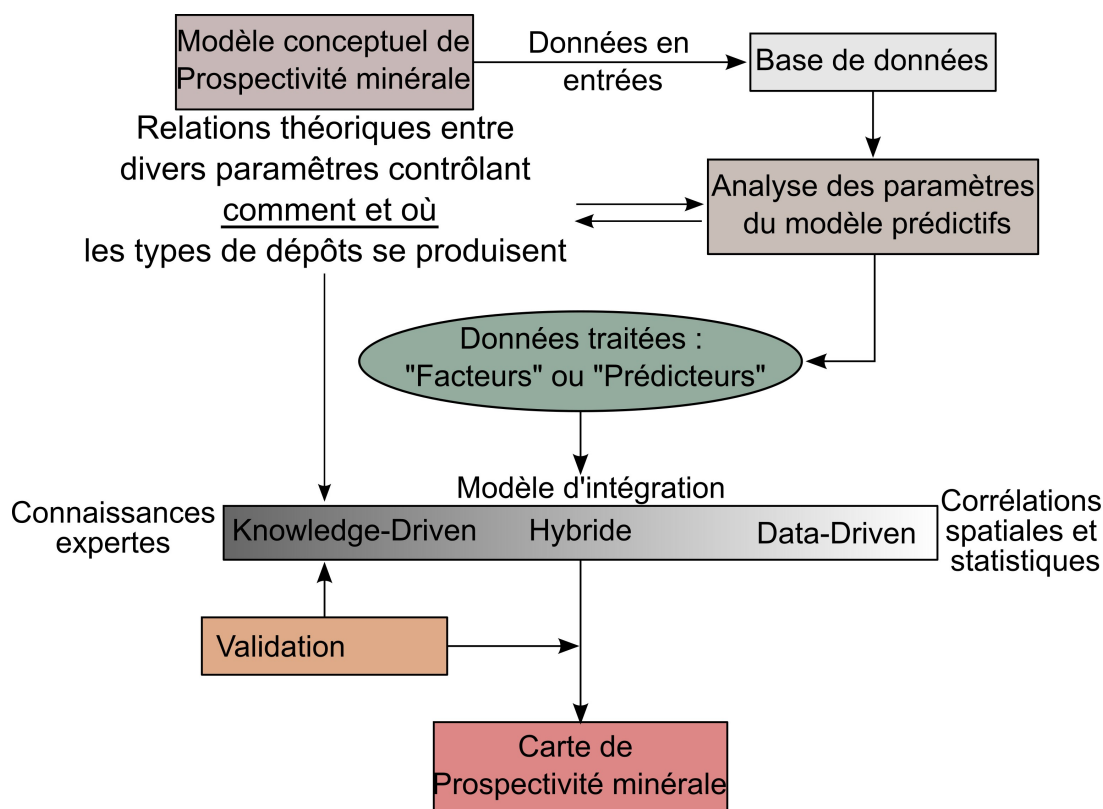


Figure 9.1: Organigramme général de la Cartographie Prédictive Minérale (Mineral Prospectivity Mapping; MPM) flowchart (modifié d'après Carranza [2017]).

Un autre outil couramment utilisé dans le cadre de la prospection minérale est la cartographie prédictive. Les méthodes de cartographies prédictives consistent à déterminer la distribution spatiale et le degré de probabilité d'occurrence d'un événement non-aléatoire. Elles sont couramment utilisées dans de nombreuses disciplines, notamment pour cartographier et évaluer les aléas et risques liés à des événements naturels, comme des inondations [Das and Gupta, 2021, Jiang et al., 2021, Wang et al., 2021], des feux de forêts [Montorio et al., 2020, Tuyen et al., 2021], l'érosion des sols [Arabameri et al., 2020, Karbalaeei Ramezanali et al., 2020] ou l'étude de la propagation des épidémies [Rajabi et al., 2014]. La cartographie prédictive minérale (souvent abrégée en MPM pour "Mineral Prospectivity Mapping", **Figure 9.1**) est une branche de la cartographie prédictive appliquée à l'exploration minérale : elle cherche à déterminer la distribution spatiale potentielle des minéralisations dans une zone (explorée ou non: "Brownfield" vs. "Greenfield"), afin de trouver des zones cibles pour l'exploration, pouvant contenir de nouveaux gisements [Carranza, 2017]. La MPM repose sur l'utilisation de modèles prédictifs pour déterminer la distribution spatiale potentielle de minéralisations non-connues. Bien que ces modèles cherchent à déterminer une "probabilité" qu'une zone soit minéralisée ou non, on préfère utiliser le terme de "favorabilité" en raison de la nature complexe du phénomène que l'on cherche à modéliser et pour lequel une probabilité s.s. peut difficilement être calculée.

Ces modèles prédictifs se basent sur un principe a priori simple : une zone est considérée comme favorable lorsqu'elle partage des caractères communs (i.e. lithologies, structures, signatures géochimiques

ou géophysiques ...) avec les zones minéralisées connues, ces points communs étant indépendants les uns des autres à travers la zone d'étude. Plus ces points communs sont nombreux, plus la zone est favorable [Carranza, 2011]. Cette relation de contrôle entre les facteurs "explicatifs" (géologiques s.l. dans notre cas) et les minéralisations définit un modèle conceptuel (**Figure 9.1**). Dans le détail, celui-ci caractérise les données à sélectionner pour créer le modèle prédictif, la manière dont ces données sont analysées pour produire des facteurs ou prédicteurs et le modèle qui servira à calculer les scores et créer la carte de favorabilité. Pour cette dernière étape de modélisation, où l'on intègre les prédicteurs, on peut se reposer sur 2 approches différentes : l'approche dite "Knowledge-driven" et l'approche dite "Data-driven" [McKay and Harris, 2016, Harris et al., 2015, Carranza, 2011].

Les modèles Knowledge-driven sont les premiers à avoir été utilisés en MPM et se basent sur des connaissances expertes (i.e. métallotectes, modèles métallogéniques) pour déterminer quels sont les prédicteurs les plus importants pour modéliser les zones potentiellement minéralisées, ainsi que leur poids ou importances relatives [Carranza, 2017]. Cette méthode est notamment avantageuse en exploration "Greenfield", lorsque peu ou aucune minéralisation connue ne sont situées dans la zone cible. On dispose alors de peu de données sur des minéralisations connues et on doit s'appuyer sur des règles, des métallotectes pré-établies à partir d'autres exemples d'études. Néanmoins, elle introduit un biais humain important : les résultats vont très fortement dépendre du modèle choisi par l'expert et de ses critères, alors même que le cas d'étude ne correspond peut-être pas à ce dit modèle. L'approche Data-driven, elle, est plus récente et repose sur l'analyse de larges quantités de données pour identifier et quantifier l'importance des relations entre les prédicteurs et les minéralisations. De fait, cette approche a beaucoup profité des progrès technologiques et de l'enrichissement des bases de données géologiques spatialisées [Dramsch, 2020]. Outre des traitements statistiques et spatiaux, elle repose aussi sur l'utilisation de méthodes d'apprentissages machine pour identifier les relations (et en quantifier la force) entre les données cartographiques et les minéralisations, à partir des facteurs ou prédicteurs d'une zone d'entraînement de référence. L'approche Data-driven se base sur le concept de base de la MPM pour définir ces relations : plus une zone partage des facteurs communs avec les zones minéralisées, plus elle est favorable. On peut de même évaluer l'importance des facteurs en fonction de leurs relations avec ces zones minéralisées. Cette approche nécessite de disposer d'une quantité de données importantes sur la zone à étudier afin que le modèle puisse "apprendre". Il en résulte que cette approche est plus applicable en contexte de "Brownfield", lorsque l'on cherche à déterminer des zones d'exploration au sein d'un prospect ou d'un district connu. Elle permet un biais humain moins important : même si la sélection des données et la création du modèle sont basées sur des connaissances expertes, sa mise en application nécessite peu d'apports expert et est généralement largement automatisée en raison de la quantité de données à traiter.

Ce travail de thèse cherche à explorer la capacité de l'approche Data-driven à contraindre et quantifier les relations complexes existant entre minéralisation et objets géologiques. Cette approche a démontré son efficacité comparée à l'approche Knowledge-driven dans des environnements à haute densité de données [McKay and Harris, 2016, Harris et al., 2015, Porwal et al., 2003b], ce qui pourrait montrer sa capacité à mettre en évidence des relations qui ne sont jusqu'alors pas contenues dans les modèles Knowledge-driven et donc dans les modèles métallogéniques et les connaissances existantes. Cette approche pourrait donc contribuer à fournir des informations, voir de potentiels nouveaux métallotectes pour enrichir les modèles métallogéniques existants et mieux contraindre les relations liant les minéralisations à l'ensemble des objets géologiques les entourant ; ces objets définissent ce que nous appelons "l'environnement géologique" des minéralisations.

On propose donc ici le développement d'une nouvelle méthode Data-driven de cartographie prédictive

pour identifier de manière automatique les relations entre des minéralisations et des environnements géologiques. Cette méthode Data-Driven automatisée permettrait une approche inverse de la MPM "conventionnel", avec un modèle conceptuel établi à partir des données spatiales disponibles en identifiant et différenciant les environnements géologiques spécifiques autour des minéralisations connues, plutôt que sur la base d'une connaissance préexistante des relations entre minéralisations et géologie. Cette méthode doit tendre à être la plus automatisée et la plus Data-driven possible afin de minimiser l'apport de connaissances expertes et d'éviter d'inclure un biais inhérent. Cette méthode vise à obtenir des métalotectes résultant strictement de l'analyse du modèle d'apprentissage machine. L'application de cette approche sur une substance et dans une zone déjà étudiée et riche en données permet de tester cette nouvelle méthode dans le cadre de cette thèse. Les résultats de cette approche sont comparés aux connaissances métallogéniques actuelles, produites via une approche naturaliste, afin de vérifier la pertinence des métalotectes produits de manière automatique par notre approche inverse.

L'objectif de ce travail de thèse peut être résumé par les questions suivantes :

- **En utilisant une approche MPM Data-Driven inverse, est-il possible de mettre en évidence les lithologies et les structures géologiques s.l. liées aux minéralisations, ainsi que les potentiels métalotectes associées, pour produire l'équivalent d'un modèle conceptuel qui serait déduit des données utilisées et des cartes de favorabilité produites ?**
- **Peut-on identifier des environnements géologiques locaux en se basant sur différentes couches de données cartographiques ?**
- **Comment déterminer ou quantifier la probabilité que les environnements géologiques locaux identifiés et leurs caractéristiques géologiques soient liés à la minéralisation ?**
- **Est-il possible de synthétiser les environnements géologiques favorables en quelques environnements "clés" permettant d'expliquer les caractéristiques uniques associées aux minéralisations ?**
- **L'ensemble de ce processus peut-il être effectué de manière Data-driven, automatisé, en se détachant le plus possible des connaissances pré-existante ?**

9.2 Développements

L'approche méthodologique présentée dans ce manuscrit repose sur l'utilisation combinée de 3 méthodes pour (i) agréger et homogénéiser des données cartographiques hétérogènes et de diverses natures et échelles, (ii) identifier et quantifier des relations entre ces données cartographiques et les minéralisations connues et (iii) discrétiser les environnements géologiques (i.e. associations de facteurs géologiques) ayant les relations les plus fortes avec les minéralisations. L'application de cette méthodologie est présentée dans le Workflow de la **Figure 9.2**. Une fois les données à traiter et les prédicteurs produits, la première étape à réaliser est l'application d'une méthode de pré-traitement des données cartographiques développée dans le cadre de cette thèse, la Disc-Based Association (DBA). La seconde étape est effectuée à l'aide d'une méthode d'apprentissage automatique permettant de produire un modèle prédictif de classification supervisée : les Forêts Aléatoires ("Random Forest", RF). Enfin, la dernière étape résulte de l'application d'une méthode de classification non-supervisée : la Classification Hiérarchique sur Composantes Principales (Hierarchical Clustering on Principal Components; HCPC). Le développement de cette méthode

dite DBA / RF / HCPC a été réalisé sous *Python* pour permettre l'automatisation des processus de traitements.

9.2.1 DBA

La DBA est une méthode de cartographie prédictive basée sur la Cell-Based Association (CBA; [Tourlière et al. \[2015\]](#)) pour la production de cartes de favorabilité. Ces méthodes ont été développées afin de mieux intégrer les données de cartographie géologique, de nature discrète, dans les calculs de cartographie prédictive minérale. La DBA a été développée dans le cadre de ce travail de thèse, comme évolution de la CBA. Elle permet de résoudre les limitations de cette méthode, principalement la limite en termes de résolution pratique de la CBA, ainsi que l'effet d'anisotropie observé avec l'utilisation d'une grille carrée. La DBA permet, comme la CBA, de surmonter des problèmes inhérents à ces données, comme (i) l'imprécision du contour des polygones et de la localisation des données ponctuelles (ii) le changement dans la représentation de ces contours en fonction de l'échelle des données, (iii) la présence d'une couverture superficielle sédimentaire ou anthropique et (iv) les effets de regroupements, lorsque des occurrences minéralisées liées au même événement sont regroupées, menant à une surestimation de l'importance des variables qui leur sont spatialement associées.

La DBA discrétise les données cartographiques pour identifier des environnements géologiques locaux, représentés par l'association des différentes variables cartographiées située dans une unité spatiale constante le long de la zone d'étude (**Figure 9.3**). La DBA se base sur une grille de points pour effectuer cette identification : la grille de points est projetée selon une maille carrée régulière, d'interdistance d . Depuis ces points, on effectue une recherche par voisinage pour identifier les variables présentes dans un rayon R autour des points. Une fois ce calcul effectué pour l'ensemble des points de la grille, l'environnement identifié est enregistré dans une grille carrée de résolution d . Les environnements géologiques de chaque cellule sont représentés dans une matrice par un code binaire dont la longueur équivaut au nombre de variables pouvant être identifiée dans la zone d'étude. La présence d'une variable dans une cellule est indiquée par un "1", son absence par un "0". La relation entre ces environnements et les paramètres d et R est complexe, d permettant de modifier la résolution de la grille et modifiant aussi la répartition des points, changeant donc les environnements identifiés, tandis que R modifie la distance à laquelle les variables sont agrégées et donc la diversité des environnements. Ces environnements sont dans notre cas constitués de données lithologiques et structurales, servant de prédicteurs pour le modèle, et de données métallogéniques : les occurrences minéralisées, qui sont la cible de cette analyse. Les environnements géologiques contenant des minéralisations sont qualifiés de "minéralisés", les autres de "non-minéralisés".

L'utilisation de cette méthode implique aussi des recouvrements entre les disques de recherches, certaines zones étant analysées de multiples fois par la recherche par voisinage, dont l'importance varie selon les valeurs de R et de d . Ces recouvrements ont été calculés de manière empirique par des simulations numériques pour déterminer un seuil minimum assurant une couverture complète de la zone d'étude. Cela a aussi permis l'établissement d'abaques permettant de déterminer le nombre et la surface des recouvrements en fonction du rapport R/d . Un seuil minimal de la valeur de R/d a été déterminé pour garantir la couverture de l'ensemble de la zone d'étude ($\sqrt{2}/2 = 0.71$). Il est aussi nécessaire de définir une grille et une distance de recherche optimale pour identifier les environnements géologiques. La détermination de ces valeurs se fait de manière empirique, en analysant les résultats obtenus en termes de nombre d'environnements, de diversité au sein de ces environnements (i.e. nombre moyen de lithologie par environnements) et de résolution. Une valeur optimale est déterminée à partir des résultats en

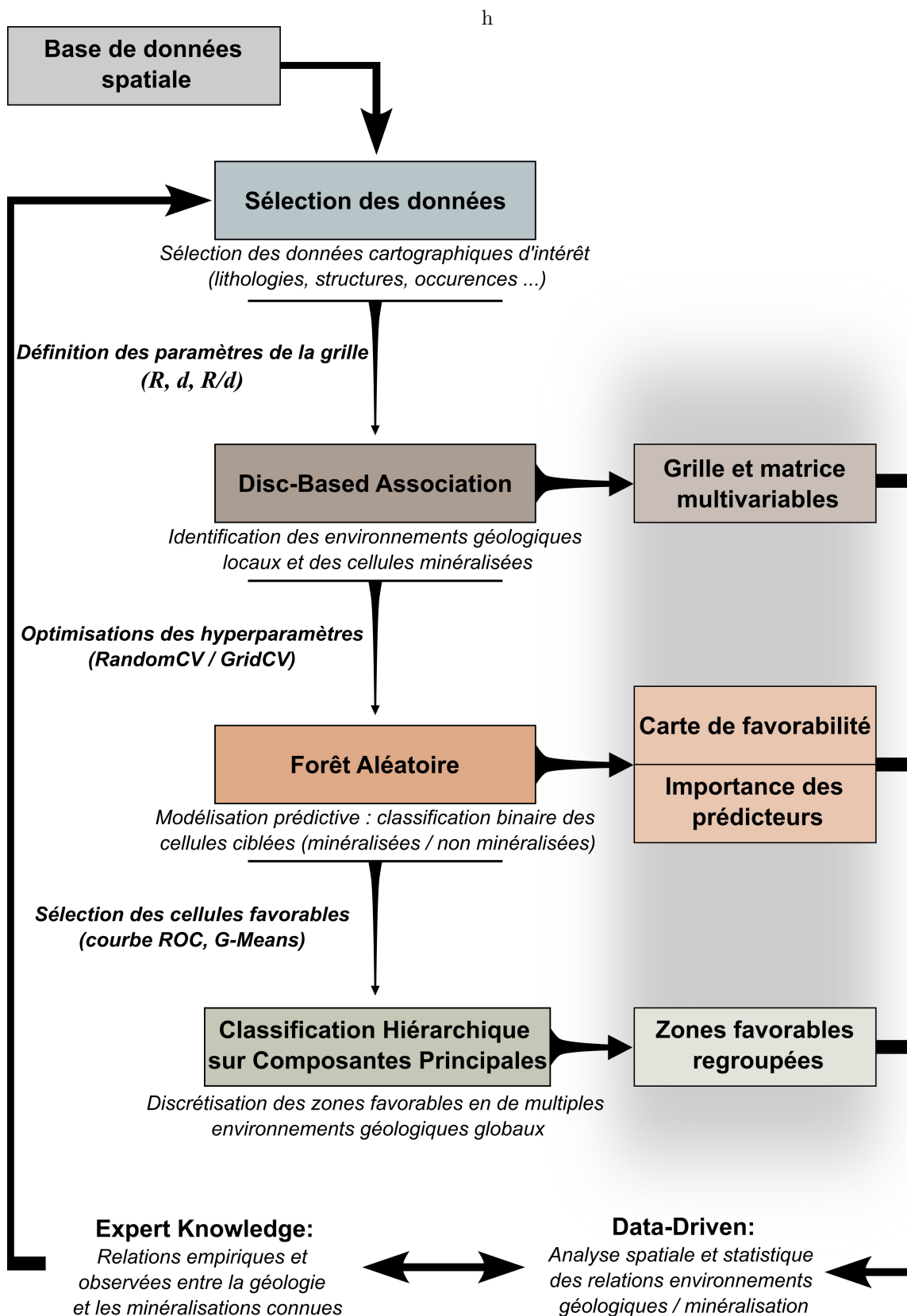


Figure 9.2: Workflow général de l'application de la méthodologie DBA / RF / HCPC.

Création d'une matrice DBA à partir de données lithologiques

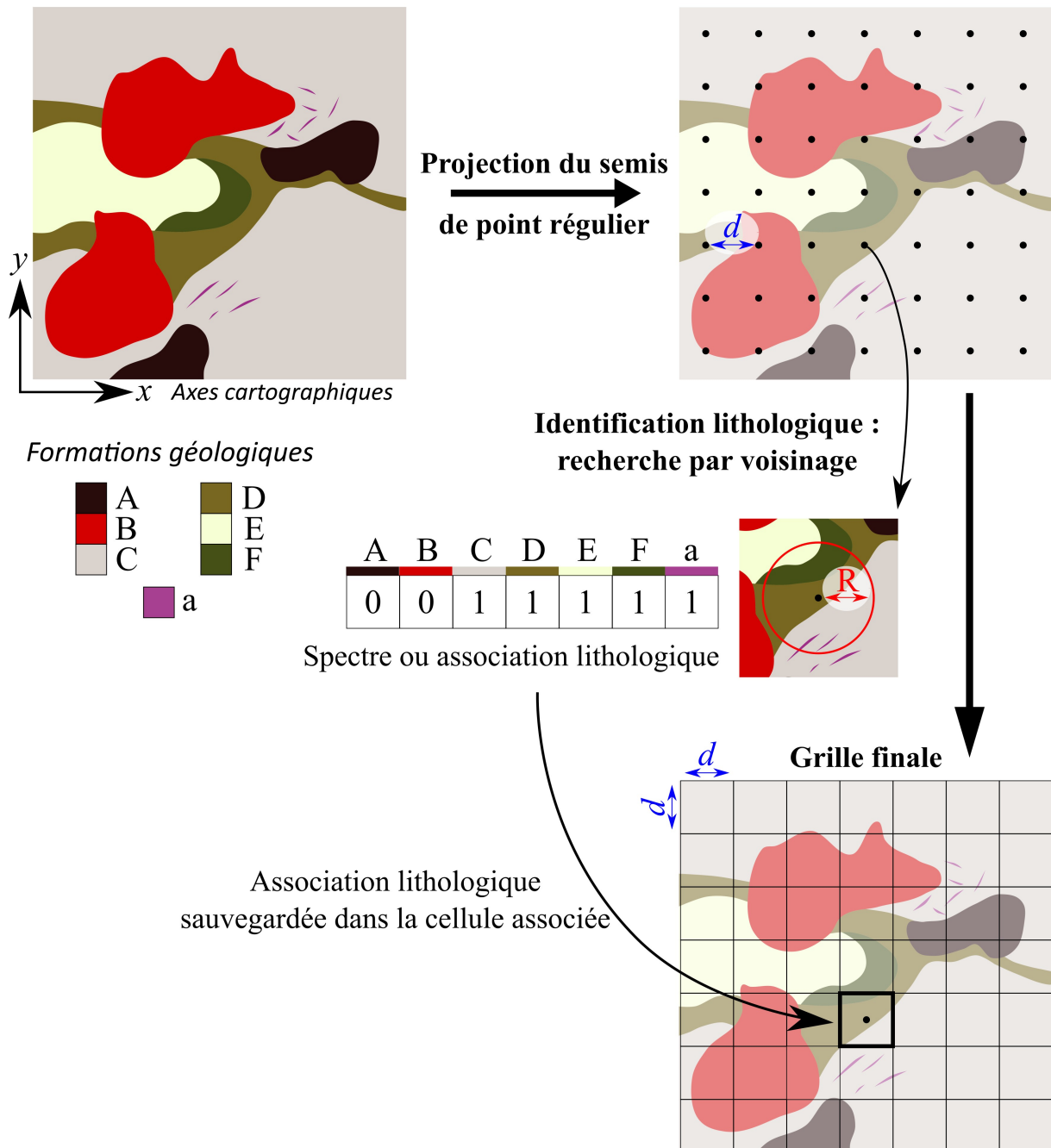


Figure 9.3: Schéma de la méthode Disc-Based Association (DBA) pour l'identification d'environnements géologiques. Les principaux paramètres de cette méthode sont l'espacement entre les points d , définissant également la résolution de la grille finale, et le rayon de la recherche par voisinage R .

fonction des objectifs de l'étude en matière de précision ou d'échelle. Dans le cadre de cette étude, nous avons cherché à maximiser la résolution tout en conservant des environnements présentant une certaine diversité.

9.2.2 RF

La Forêt Aléatoire ("Random Forest", RF) est une méthode de classification et de régression supervisée, formalisée par Breiman [2001]. Cette méthode est ici couplée à l'analyse des environnements géologiques effectuée par la DBA pour produire des cartes de favorabilité "Data-driven". Le RF est appliqué pour fournir une classification binaire des cellules, entre une classe "cellule minéralisée" et une classe "cellule non-minéralisée". Cette classification est effectuée sur l'ensemble des cellules produites par la DBA, entraînant un large déséquilibre entre la classe cellule minéralisée, minoritaire, et la classe cellules non-minéralisée, largement majoritaire (**Figure 9.4.a**). Ce déséquilibre s'apparente à un problème de détection d'anomalie : l'événement que l'on cherche à prédire, la présence de minéralisations, est rare par nature, il est donc logique qu'il soit sous-représenté. D'un autre côté, si les cellules non-minéralisées sont forcément majoritaires, il existe une probabilité non-négligeable que certaines d'entre elles puissent contenir une minéralisation encore inconnue. Il existe donc une 3^{ème} classe sous-jacente dans cette analyse : les cellules minéralisées non-connues, dont on cherche à déterminer l'existence et l'emplacement à travers ce modèle prédictif. Pour cela, il est nécessaire que le modèle produise des Faux Positifs (FP) : des cellules non-minéralisées présentant un score de favorabilité élevée et ayant donc une forte chance de contenir une minéralisation selon le modèle. Étant donné que les classes des cellules sont partiellement incertaines ("non-minéralisée" et "minéralisée inconnue") nous sommes dans le cas d'une classification semi-supervisée.

Pour obtenir les scores de favorabilité, le modèle RF se base sur l'utilisation d'arbres de décision (**Figure 9.4**). L'opérateur détermine le nombre n d'arbres de décision produit par le modèle. Cette classification se base sur une sélection aléatoire d'un échantillon d'individus ("Bootstrap Random Sampling") servant à entraîner le modèle, généralement les 2/3 du nombre total d'individus (**Figure 9.4.b**). Le tiers restant sert à valider l'arbre de décision, calculant l'erreur "Out-Of Bag" permettant de déterminer l'efficacité de chaque arbre, sa capacité à correctement prédire les classes. Cette sélection est effectuée de manière stratifiée, afin de conserver la proportion relative entre les différentes classes aussi bien entre l'échantillon d'entraînement et l'échantillon de validation. Ces échantillons sont aussi pondérés de manière inversement proportionnelle à la fréquence des classes pour diminuer l'impact du déséquilibre entre la classe majoritaire et la classe minoritaire. Chaque arbre de décision va aussi utiliser un nombre i de prédicteurs sélectionnés aléatoirement pour effectuer la classification (**Figure 9.4.c**). Celle-ci est effectuée par chaque arbre, et se base sur l'indice d'impureté de "Gini", déterminant à chaque partition de l'échantillon d'entraînement les prédicteurs permettant une meilleure séparation entre les 2 classes. L'objectif de l'arbre de décision est d'appliquer des séparations itérativement sur l'échantillon d'entraînement jusqu'à obtenir des sous-échantillons les plus purs possibles, c'est-à-dire ne contenant qu'une seule des 2 classes dans le meilleur des cas. Le score de favorabilité de chaque cellule est calculée une fois la forêt aléatoire produite : il s'agit de la moyenne des votes de l'ensemble des arbres de décision (**Figure 9.4.d**). Ce score de favorabilité, attribué à chaque cellule, permet d'obtenir la carte de favorabilité finale. Celle-ci est validée par un expert, via une comparaison avec les minéralisations et districts connus ou à des données de validations retenues avant la classification (**Figure 9.4.e**).

Les hyperparamètres du modèle prédictif, c'est-à-dire les paramètres déterminés en amont de l'analyse

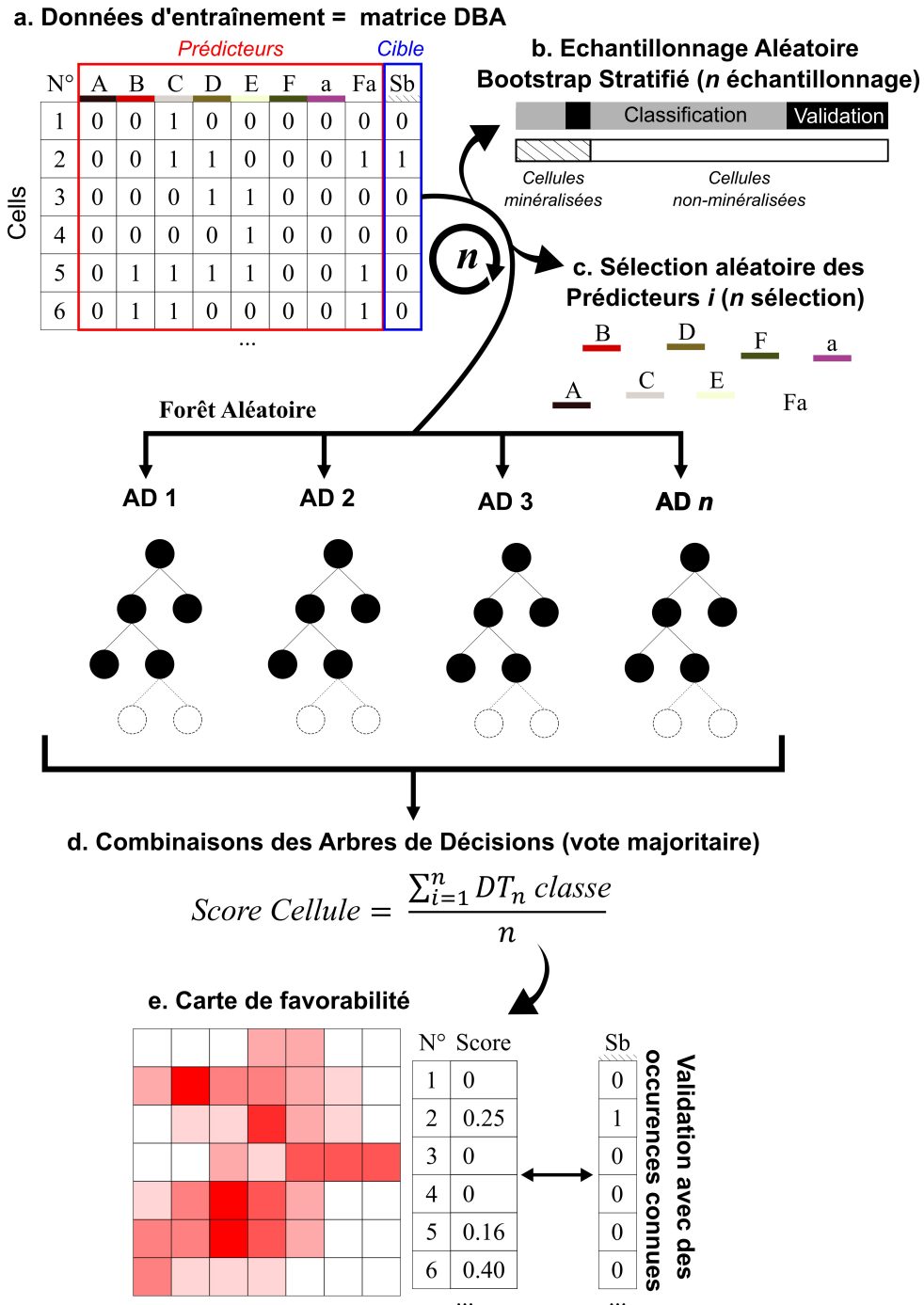


Figure 9.4: Création de cartes de prospectivité à partir de la méthode DBA / RF (modifié à partir de Harris et al. [2015]). a. Les données d'entraînement utilisées par le modèle prédictif RF sont composées des cellules. On peut distinguer dans la matrice les prédicteurs utilisés par le modèle (formations lithologiques "A" à "a" et faille "Fa") et l'étiquette des cellules (absence / présence de Sb). b. Un échantillonnage stratifié aléatoire dit "Bootstrap" est appliqué sur la matrice pour chaque arbre. c. Une sélection aléatoire de prédicteurs est effectuée pour chaque arbre. d. La Forêt Aléatoire résultante est appliquée sur chaque cellule de la matrice, le score de favorabilité de la cellule étant le résultat du vote majoritaire. e. Le score de favorabilité est sauvegardé dans la matrice et peut être évalué en utilisant des occurrences connues ou en utilisant des occurrences de contrôle non utilisées dans le modèle.

des données (nombre d'arbres n , nombre de prédicteurs i , profondeur des arbres ...) sont défini par une méthode d'optimisation automatique appelée "Random Search". Les résultats du modèle sont évalués à partir de la courbe ROC (Receiver Operating Characteristic), permettant, pour différentes valeurs seuils de classification, de comparer la capacité du modèle à prédire correctement les cellules minéralisées, ou Taux de Vrais Positifs ("True Positive Rate", TPR), à la proportion des cellules non-minéralisée prédites comme étant minéralisée, ou Taux de Faux Positifs ("False Positive Rate", FPR). Les prédicteurs utilisés par le modèle sont aussi évalués à l'aide du "Mean Decrease in Impurity" (MDI), définissant l'importance de chaque prédicteur dans la classification en se basant sur sa capacité à diminuer l'impureté des nœuds des arbres l'utilisant. Une comparaison entre la méthode DBA / RF et d'autres méthodes de cartographie prédictive utilisées au sein du BRGM, comme la CBA / Ranking, la CBA par *signature exacte*, la CBA / CAH et le WoE, a démontrée l'efficacité du couplage DBA / RF pour produire des cartes de favorabilité. Celle-ci permet en effet de minimiser l'emprise de la carte considérée comme favorable tout en permettant une bonne classification des cellules minéralisées.

9.2.3 HCPC

La Classification Hiérarchique sur Composantes Principales ("Hierarchical Clustering on Principal Components", HCPC) est une méthode de classification non-supervisée [Husson et al., 2010] permettant de regrouper les environnements géologiques identifiés par la DBA en fonctions de leurs similitudes. Cette méthode se base sur l'application d'une analyse factorielle, l'Analyse des Correspondances Multiples ("Multiple Correspondance Analysis", MCA), combinée à une méthode de regroupement, la Classification Ascendante Hiérarchique ("Agglomerative Hierarchical Clustering", AHC). La combinaison de cette méthode avec la DBA représente une évolution du couplage entre la CBA et l'AHC. L'AHC est limitée par la nature discrète et binaire des données produites par la CBA ou la DBA : l'agrégation effectuée par l'AHC se base sur un indice de dissimilarité euclidien, plus adapté à des données continues. La HCPC, elle, est optimisée pour l'utilisation de variables qualitatives et discrètes. La HCPC, avec l'application d'une analyse factorielle avant l'agrégation, permet de transformer les données discrètes en données continues, la distance entre chaque couple de cellules correspondant à une estimation de leurs similarités selon la MCA. Les résultats de l'AHC sont donc plus cohérents lors d'une HCPC que lorsqu'elle est appliquée seule.

La HCPC est ici appliquée après les calculs de la DBA / RF. À partir des résultats de celle-ci, il est possible de déterminer les cellules les plus favorables selon le modèle prédictif (**Figure 9.5.a**). Cette sélection est effectuée à partir de la courbe ROC en utilisant un indice appelé G-means, permettant de déterminer un seuil de coupure optimale, minimisant le taux de Faux Positif (FP) et maximisant le taux de Vrais Positifs (TP). Une fois le seuil de coupure déterminé, les cellules présentant un score $\geq$ à ce score sont sélectionnées. Ces cellules sont alors traitées avec la HCPC pour identifier les regroupements d'environnements géologiques favorables (**Figure 9.5.b**). L'analyse factorielle de la MCA est d'abord appliquée, la distance entre chaque individu et chaque modalité des prédicteurs de la matrice est calculée en se basant sur la distance au ϕ^2 . Avec cette distance, 2 individus sont proches s'ils partagent les mêmes modalités et 2 modalités sont proches si elles sont partagées par les mêmes individus. Plus une modalité est rare, plus son poids est important. Des axes factoriaux sont produit à partir de ces distances afin de reprojeter les individus dans un nouvel espace dimensionnel, cherchant à mieux représenter la variabilité des individus en maximisant l'écart à l'origine.

Les individus sont alors regroupés avec l'AHC. Ce regroupement se base sur les distances euclidi-

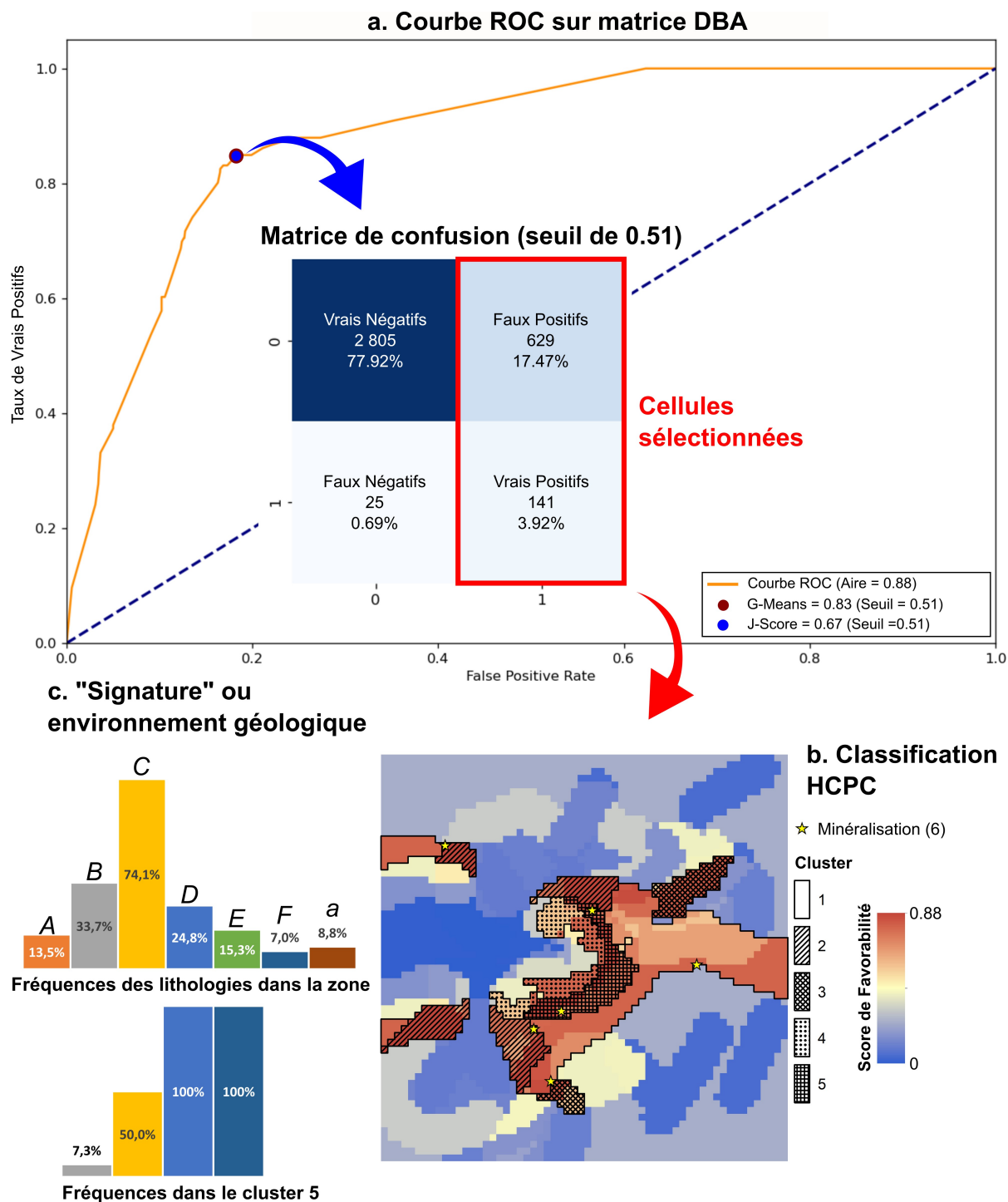


Figure 9.5: Identification des environnements géologiques entourant les zones favorables sur un schéma de la carte de favorabilité DBA / RF (résolution de 1 000 m). a. La courbe de ROC est utilisée pour déterminer le seuil déterminant les cellules favorables, en utilisant le G-Means et le J-Score. b. La classification HCPC est effectuée sur les associations lithologiques des cellules sélectionnées, produisant des clusters représentant différents environnements géologiques favorables. c. La fréquence des lithologies à l'intérieur du cluster est calculée et comparée à leur fréquence à travers la zone d'étude afin d'identifier les particularités de son environnement géologique, les lithologies et associations lithologiques remarquables.

ennes entre les individus dans cet espace factoriel : une matrice de distance est produite, calculant la distance entre chaque individu lors de la première itération. Les deux individus les plus proches sont alors regroupés, formant un regroupement ou cluster. Pour les agrégations suivantes, on utilise la méthode de Ward. Avec cette méthode, la distance entre deux clusters correspond au carré de la distance entre leurs centres de gravité respectifs, pondéré par leurs populations. Cette méthode permet de maximiser l'inertie inter-clusters (écarts entre le centre de gravité de chaque cluster et le centre de gravité de l'ensemble des individus) et de minimiser l'inertie intra-clusters (écart entre les individus de chaque cluster et leur centre de gravité respectif). Cela permet, théoriquement, que les individus au sein de chaque cluster soient les plus semblables possibles entre eux, et les plus différents possibles des individus des autres clusters. L'agrégation des individus et des clusters est faite de manière itérative, une nouvelle matrice de distance étant calculée à chaque itération, jusqu'à ce qu'il n'y ait qu'une classe. L'application de l'AHC sur la MCA peut être représentée sous forme d'un dendrogramme 3D représentant l'agrégation des cellules (**Figure 5.6, p.121**). Le nombre de clusters optimal est défini manuellement par l'opérateur, en fonction des distances (ou "inertie") entre les clusters, et selon ses connaissances géologiques, permettant une généralisation des environnements géologiques locaux les plus favorables en un nombre restreint d'environnements géologiques globaux. Ceux-ci sont ensuite analysés pour faire ressortir les associations de prédicteurs typiques des différents environnements minéralisés (**Figure 9.5.c**).

9.3 Applications

9.3.1 Zone d'étude

Ce travail fait partie du projet AUREOLE (tArgeting eU cRitical mEtals (Sb, W) and predictability of Sb-As-Hg enviroNmentaL issuEs, ANR-19-MIN2-0002, <https://aureole.brgm.fr/>). L'antimoine est considéré comme l'une des ressources critiques pour l'Union européenne [COM(2020) 474 final, 2020, Blengini et al., 2020, Bobba et al., 2020] en raison notamment de son utilisation comme retardateur de flamme [Guo et al., 2014, Henckens et al., 2016] et son potentiel en tant que matériaux pour les batteries rechargeables à hautes capacités [Sun et al., 2010, Luo et al., 2017, He et al., 2018, Zoller et al., 2018]. Ce projet vise (i) à développer un nouveaux modèle métallogénique 3D pour les minéralisations à Sb et W, (ii) à réaliser une étude des risques environnementaux pour les métaux associées (Sb, Hg, As) et (iii) à produire une carte de prospectivité pour ces minéralisations, couvrant les terrains paléozoïques Français et Ibériques. Au sein de ce projet, cette thèse s'inscrit dans l'étude des minéralisations d'antimoine et de leurs relations avec la géologie, dans la production de cartes de favorabilité pour l'antimoine ainsi que dans le développement d'un outil pour homogénéiser les données cartographiques. La zone d'étude sélectionnée pour ce travail et sur laquelle la méthodologie développée est appliquée couvre la partie Ouest du Massif Varisque Européen, s'étendant du Massif Armoricaïn ("Armorican Massif", AM) en France à la Zone Centre Ibérique ("Central Iberian Zone", CIZ), couvrant le Nord du Portugal et le Nord-Ouest de l'Espagne.

La méthodologie est donc appliquée le long de l'Arc Ibéro-Armoricaïn ("Ibero-Armorican Arc", IAA), une structure arquée s'étendant du Massif Armoricaïn à la Zone Centre Ibérique de part et d'autre du Golfe de Gascogne, présentant une continuité établie entre ses principales unités et structures géologiques [Ballèvre et al., 2014, Catalán et al., 2007, Quesada and Oliveira, 2019, Dias et al., 2016, Aerden, 2004, Martínez Catalán, 2012, Dias and Ribeiro, 1995, Pereira et al., 1993]. Cet Arc fait partie de l'orogène

Varisque, développé durant le Paléozoïque, s'étendant de l'Europe Central à la Péninsule Ibérique. L'orogène résulte de la collision entre deux grandes plaques lithosphériques continentales : le Gondwana au Sud et Laurussia au Nord, elle-même issue de la collision entre deux domaines continentaux, Laurentia et Baltica [Martínez Catalán et al., 2021]. La collision entre le Gondwana et la Laurentia se déroule en plusieurs phases, les deux domaines étant séparés par des microplaques détachées de la marge gondwanienne : la microplaque Avalonia, séparée de la Laurentia par le Iapetus, et la microplaque Armorica, séparée d'Avalonia par l'océan Rhéique et du Gondwana par l'océan Galico-Moldanubien. L'ensemble Ibéro-Armoricain formait une microplaque détachée de la marge du Gondwana Nord au Cambrien et surtout à l'Ordovicien. L'orogène s'est développé essentiellement de la fin du Dévonien (360 Ma) [Ballèvre et al., 2009] à la fin du Carbonifère (300 Ma), durant la fermeture de l'océan Rhéique et de l'océan Galico-Moldanubien [Ballèvre et al., 2013, Franke et al., 2017, Matte, 2001].

Les minéralisations d'antimoine le long de cette portion du Massif Varisque ont été étudiées en détail à travers de nombreux travaux (e.g. [Gumiel and Arribas, 1987, Gumiel, 1982, Gumiel et al., 1976, Arribas and Gumiel, 1984, Couto et al., 1990, Ortega and Vindel, 1995, Ortega et al., 1996, Neiva et al., 2008, Pochon et al., 2016, 2019, 2018, Cheval-Garabedian, 2019, Gloaguen and Tourlière, 2016a,b, Pochon, 2017, Marcoux and Chauris, 1994, Munoz et al., 1992, Marcoux et al., 1984, Bouchot et al., 2005]). L'association de failles majeures, de zones de cisaillement ductiles d'échelle crustale et d'une forte activité magmatique favorise les circulations de fluides dans la croûte le long de l'orogène et donc la formation de minéralisations [Pereira et al., 1993, Cox et al., 1986, Mateus and Noronha, 2010, Almeida et al., 2002, Pochon et al., 2016, Hedenquist and Lowenstern, 1994, Yardley and Cleverley, 2015]. De nombreuses études récentes ont été réalisées sur les minéralisations et les districts métallogéniques dans cette zone pour de multiples substances d'intérêt [Launay, 2018, Silva et al., 2018, Palero-Fernández et al., 2015, Blundell et al., 2005, Cassard et al., 2012, COM(2020) 474 final, 2020, Hartai et al., 2016, Tornos et al., 2004]. Malgré toutes ces études, la chaîne Varisque reste une cible intéressante pour l'exploration minière [Hartai et al., 2016, Vic and Tourlière, 2016, Gloaguen and Tourlière, 2016a, Cassard et al., 2012]. Sur la base des cartes géologiques maintenant disponibles de manière numérique à travers les différents domaines de l'Arc, une approche Data-driven peut être réalisée. La continuité des structures le long de cet arc permet la production de cartes prédictives cohérente d'une échelle 1:50 000 pour le Massif Armoricaïn et d'une échelle de 1:1 000 000 pour le Massif ibérique, comme réalisé au cours de cette thèse.

Le Massif Armoricaïn est situé au Nord-Ouest de la France et regroupe l'ensemble des formations protérozoïques et paléozoïques affleurant dans l'ouest de la France. Il est limité à l'Est par le Bassin parisien, au Sud par le Bassin aquitain, au Nord par la Manche et à l'Ouest par le Golfe de Gascogne. Selon les données géophysiques, le socle paléozoïque se poursuit sous ces couvertures tardives, mais n'affleure pas [Ballèvre et al., 2013, Baptiste, 2016]. La CIZ est située dans la partie occidentale de la Péninsule Ibérique, s'étendant du nord-ouest de l'Espagne à son centre, en passant par le nord du Portugal. Elle est bordée au nord par l'anticlinal d'Ollo de Sapo, formant son contact avec l'Ouest Asturien-Léonais, et au sud par le chevauchement Ferreira do Zézere-Portalegre [Quesada and Oliveira, 2019]. La Zone Galice-Tras Os Montes, composée d'unités allochtones sur l'autochtone de la CIZ, présente de fortes similitudes avec le domaine sud-armoricaïn et est parfois considéré comme faisant partie de la CIZ s.l. [Ballèvre et al., 2014].

9.3.2 Application au massif Armoricaïn

Avant l'application de la nouvelle méthode DBA / RF / HCPC, une démonstration de la faisabilité du couplage entre la CBA et le RF est réalisé sur le Massif Armoricaïn. Cette démonstration de faisabilité montre l'efficacité de la méthode CBA / RF par rapport à d'autres méthodes de cartographie prédictives : la CBA / Ranking, la CBA par *signature exacte*, la CBA / CAH et le WoE, en comparant leurs résultats lors d'une cartographie prédictive de l'antimoine. Une fois cette démonstration réalisée, la DBA / RF / HCPC est appliquée sur le Massif Armoricaïn. Les données lithologiques utilisées sont des données harmonisées à l'échelle de tout le Massif au 1:50 000 et produites par [Gloaguen and Tourlière \[2016b\]](#) à partir de la base de données CHARM. Les failles utilisées pour le modèle prédictif sont elles aussi issues de la base de données CHARM, au 1:50 000. 141 occurrences d'antimoine sont utilisées, extraites de la base de données du BRGM [[Cassard et al., 2015](#)] et mises à jour par [Gloaguen and Tourlière \[2016b\]](#). Les formations lithologiques ont été filtrées afin de n'utiliser que les formations paléozoïques et précambriennes pour réaliser l'analyse, les minéralisations d'antimoine étant liées à des événements paléozoïques au cours de l'orogénèse Varisque [[Chauris, 1977](#), [Marcoux and Chauris, 1994](#), [Munoz et al., 1992](#), [Pochon et al., 2019](#)].

La méthode DBA est d'abord appliquée sur l'ensemble des données géologiques harmonisées. Afin de trouver la valeur de la taille de la cellule d et du rayon du disque R optimales, la DBA est itérée sur la zone étudiée avec différentes valeurs pour la recherche par voisinage, allant de 2 000 m à 30 000 m avec un pas de 1 000 m. Les valeurs optimales trouvées avec ce jeu de données lithologiques sont un rapport R/d de 3 et un rayon de recherche R de 3 000 m, permettant le meilleur équilibre entre la résolution et la diversité des associations lithologiques extraites. La taille des cellules d est donc de 1 000 m. Au total, la grille DBA produite contient 48 871 cellules et 9 529 environnements uniques, avec en moyenne 7,5 lithologies par cellule et un maximum de 21 lithologies dans une cellule. Les failles sont ajoutées avec une recherche par voisinage de 1 000 m afin d'éviter leur sur-représentation dans les cellules, un rayon de recherche de 3 000 m aurait conduit à ce que 90% des cellules contiennent une faille, contre 62% avec 1 000 m. Pour les minéralisations, on utilise la distance de voisinage la plus faible, celle utilisée pour les failles. Ainsi 0,67% des cellules contiennent une minéralisation connue.

Le RF est appliqué sur la matrice DBA. Sur l'ensemble de la matrice, toutes les cellules minéralisées et non minéralisées sont utilisées comme données d'entraînement pour le processus de classification RF. Le modèle utilise les hyperparamètres optimaux défini par le "Random Search". Dans la carte prédictive en résultant (**Figure 9.6**), les cellules les plus favorables ont un score de 0,973, tandis que les cellules les moins favorables ont un score de 0. Les prédicteurs possédant les MDI les plus élevées (**Figure 7.8, p.155**) sont les grauwackes du Cambrien (GRWK3), les granites Carbonifères (GRAN7), les grès Carbonifères (GRES7) et les filons de dolérites (FDOLE). Ils montrent des relations différentes avec les cellules minéralisées : les grauwackes Cambriens et les grès Carbonifères sont des lithologies "rares", couvrant seulement une partie de la zone étudiée et montrant un score de favorabilité relativement élevé, tandis que les granites Carbonifères et les filons de dolérites sont des lithologies répandues avec des scores de favorabilité faibles en comparaison. Les deux premiers sont plus facilement définissables comme de possibles "métallotectes" à extraire de ce modèle prédictif, car ils permettent de cibler une zone plus spécifique. Cependant, même si les autres formations couvrent de grandes surfaces, leurs associations avec d'autres lithologies pourraient être aussi favorablement liées à la présence d'antimoine.

Une évaluation plus approfondie du score de favorabilité est établie à l'aide de la courbe ROC. Celle-ci possède une aire sous la courbe ("Area Under the Curve", AUC) de 0,98, un score conséquent et largement

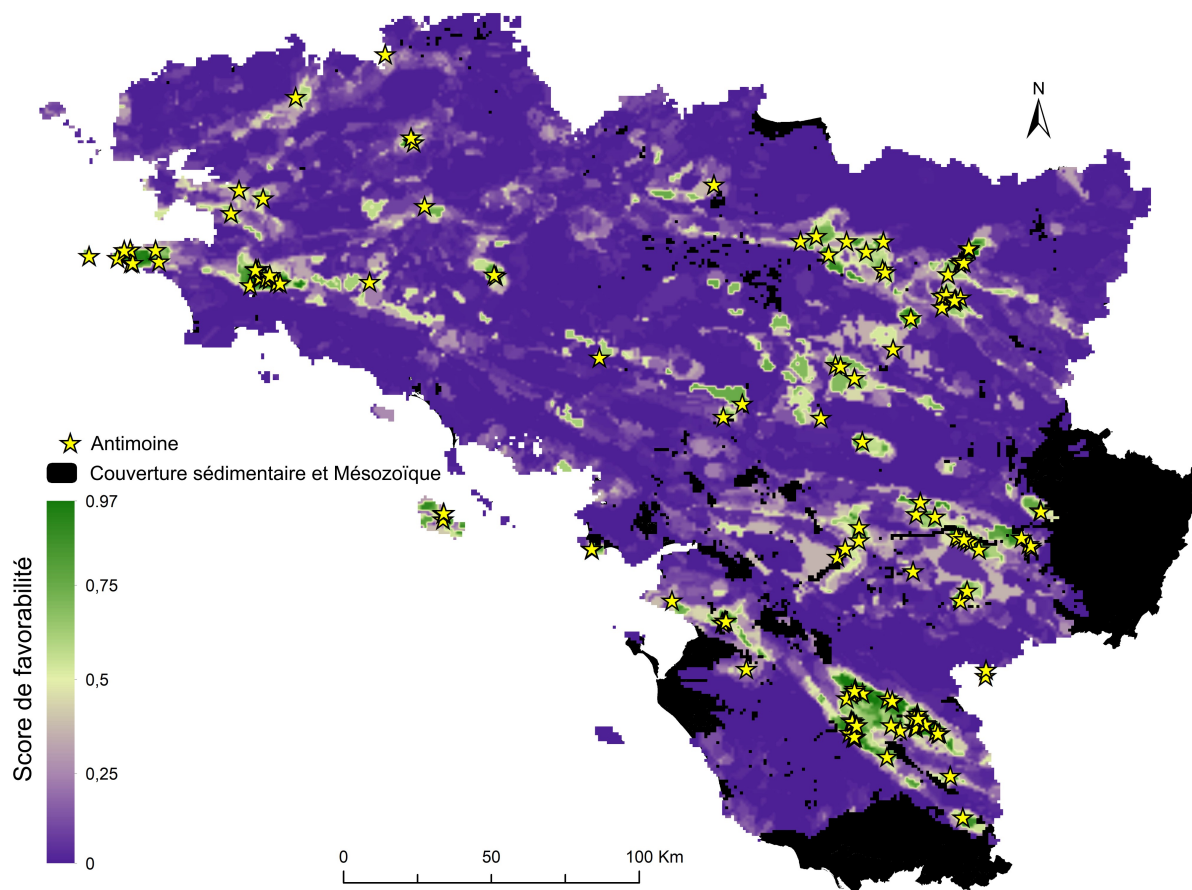


Figure 9.6: Carte de favorabilité de l'antimoine dans le Massif Armoricain, produite à partir de la méthode DBA / RF. Avec une recherche par voisinage R de 3 000 m et un R/d de 3, la résolution de la grille est de 1 000 m. Les minéralisations et les failles utilisent une recherche par voisinage R de 1 000 m. Les scores de favorabilité élevés sont affichés en vert, correspondant à une plus grande probabilité que la cellule contienne une minéralisation d'antimoine.

au-dessus de 0,5 et démontrant l'efficacité de la classification produite par le modèle. Cette courbe ROC est utilisée pour déterminer les cellules favorables à l'aide de la valeur de coupe optimale fournie par le G-Means. Cette valeur est ici de 0,506 et toute cellule ayant un score de favorabilité supérieur ou égal est donc sélectionnée et traitée avec la HCPC. Celle-ci nous permet d'identifier 4 clusters (**Figure 9.7**), correspondant à des environnements géologiques globaux favorables au sein du Massif Armoricain.

Le cluster 1 est le plus grand cluster, représentant 50,99% de la zone favorable. Son environnement géologique global est dominé par des formations sédimentaires d'âges variés, principalement ordoviciennes, associées à des dolérites et des veines de quartz. Il contient certains des plus importants districts à antimoine du Massif Armoricain, avec le district de La Lucette [Guillard, 2009, Moyroud et al., 1979, Serment, 1978], de Le Semnon-La Coëfferie [BRGM, 1987, Chauris et al., 1985, Gloaguen and Tourlière, 2016b, Moyroud et al., 1979] et le gisement de Rosnoën [BRGM, 1988b, Gentilhomme et al., 2001]. Les scores de favorabilité les plus élevés à l'intérieur de ce cluster sont situés autour du district de La Lucette, dans le bassin de Laval. Il regroupe diverses minéralisations dont le principal point commun est la prédominance de formations sédimentaires dans leurs encaissants. Les filons de dolérites et les veines de quartz sont également très présents, ce qui n'est pas surprenant puisque ce sont les principales formations hôtes des minéralisations d'antimoine [Guillard, 2009, Moyroud et al., 1979, Serment, 1978,

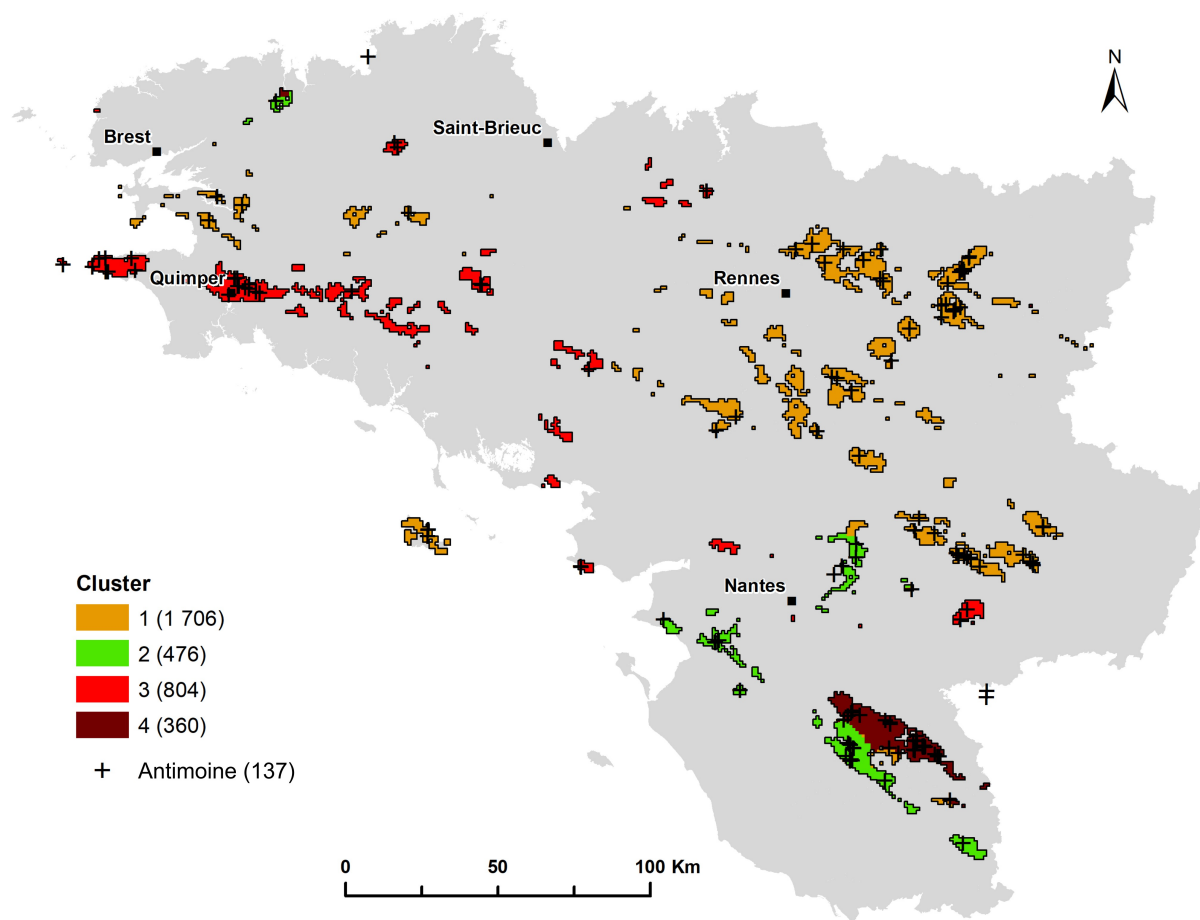


Figure 9.7: Carte des 4 clusters produits par la Classification Hiérarchique sur Composantes Principales (Hierarchical Clustering on Principal Component, HCPC) des cellules favorables (cellules avec un score $\geq 0,506$) de la carte de favorabilité DBA / RF du Massif armoricain. Le nombre de cellules est affiché pour chaque cluster, ainsi que le nombre d'occurrences d'antimoine dans la zone d'étude.

BRGM, 1987, Chauris et al., 1985, Gloaguen and Tourlière, 2016b, Moyroud et al., 1979]. Les failles sont également prédominantes, la plupart des zones favorables étant situées près de zones de cisaillement majeures ou à proximité d'essaims de failles.

Le cluster 3 est le deuxième plus grand cluster, contenant 24,03% de la surface de zone favorable. Il est composé de granites et de migmatites Carbonifère associés à des formations du socle Néoprotérozoïque, dans un environnement à forte densité de failles et de veines. Il contient l'un des plus importants district d'antimoine du Massif Armoricain, le district de Quimper-Cap Sizun [BRGM, 1990, Fouquet, 1980, Gorichon and Kerjean, 1979, Moyroud et al., 1979], situé dans la partie occidentale du Domaine Centre-Armoricain. Les minéralisations d'antimoine y sont situées dans des veines de quartz encaissées dans des granites Ordoviciens et Carbonifères, à proximité de la Zone de Cisaillement Sud-Armoricaine ("South Armorican Shear Zone", SASZ) [Fouquet, 1980, Chauris, 1977, Marcoux and Chauris, 1994]. Les scores de favorabilité les plus élevés sont dans ce district ainsi que dans d'autres gisements: le gisement polymétalliques (W-Sb) de l'Échasserie [Gloaguen et al., 2018], le gisement d'or orogénique (Au-Sb) de Saint-Pierre Montlimart [Braux et al., 1991] et les minéralisations de Pontivy et de La Chapelle Neuve. Ce cluster regroupe les minéralisations se produisant dans des formations granitiques, près de failles majeures. Il décrit principalement les minéralisations présentes dans le district de Quimper-Cap Sizun et semble mettre davantage en évidence les environnements similaires.

Le cluster 2 contient 14,23% en surface de la zone favorable. Il présente un environnement géologique principalement constitué de formations métamorphiques, principalement d'âge Ordovicien, associées à des ultrabasites dans un environnement faillé. Ce cluster contient une partie du district vendéen [Blouin, 1991, BRGM, 1988a, Marcoux and Chauris, 1994, Moyroud et al., 1979] situé dans la partie orientale du Domaine Sud Armoricaire ("South Armorican Domain", SAD). Les minéralisations d'antimoine de ce district sont localisées dans des veines de quartz encaissées dans des schistes Briovériens et Cambriens, à proximité de formations volcaniques Cambriennes [Marcoux et al., 1984]. Les scores de favorabilité les plus élevés sont situés dans 2 des principaux gisements du district vendéen, la Veronnière et les Essarts [Moyroud et al., 1979, Marcoux et al., 1984]. Le site des Brouzils [Le Fur et al., 1988, Blouin, 1991], bien qu'il s'agisse d'un gisement important, présente un score de favorabilité plus faible. Les minéralisations de ce cluster sont principalement situées dans le district de Vendée, avec un environnement géologique dominé par les roches métamorphiques et les formations basiques et ultrabasiques, que l'on retrouve à la fois dans les Nappes des Essarts et dans les parties externes de la Nappe de Champtoceaux.

Le cluster 4 est le plus petit cluster, contenant 10,76% de la zone favorable. Son environnement géologique est caractérisé par des unités du socle Cambrien et des formations volcano-détritiques associées à des veines de quartz. Il se limite au district de Vendée, avec comme principale minéralisation le gisement des Brouzils [Le Fur et al., 1988, Blouin, 1991], accompagné des minéralisations filoniennes de Rochetrejoux [Serment, 1980], La Ramée, Sainte Marie [Moyroud et al., 1979] et la Baussonnière [BRGM, 1988b, Gentilhomme et al., 2001]. Ce cluster est très localisé et présente un score de favorabilité très élevé par rapport aux autres, et de ce fait la plupart des minéralisations qui le composent sont situées dans des zones de haute favorabilité. Ce cluster peut être associé au cluster 2, les deux contenant des minéralisations situées dans le district de Vendée. Le cluster 4 contient la partie nord du district de Vendée, où les minéralisations sont situées loin des nappes de Les Essart et plus au niveau des lithologies Cambriennes situées le long de la Zone de Cisaillement Sud-Armoricaine.

9.3.3 Application à la Zone Centre Ibérique

La méthodologie DBA / RF / HCPC est appliquée sur une autre partie du Massif Varisque, la Zone Centre Ibérique (CIZ, Espagne et Portugal). Le contexte global lié au massif varisque est similaire mais l'échelle du jeu de données utilisé dans cette étude est différente. Alors que des cartes géologiques de la zone sont disponibles à une échelle de 1:50 000 sur les portions portugaises et espagnoles, l'harmonisation de ces cartes géologiques à l'échelle de la CIZ n'est pas disponible. Ainsi, nous utilisons pour cette application le jeu de données lithologiques et les failles de la carte géologique de la péninsule ibérique au 1:1 000 000 (édition 2015; Rodríguez Fernández and Oliveira [2015]). Les formations lithologiques sont limitées aux formations paléozoïques et précambriennes de la CIZ s.l. Les occurrences d'antimoine sont extraites des bases de données de la LNEG (SIORMINP, <https://geoportal.lneg.pt/pt/bds/siorminp/>; Patinha et al. [2020]) et de l'IGME (BDMIN, <http://doc.igme.es/bdmin/>), ce qui résulte en 140 occurrences dans l'ensemble de la zone d'étude. Avec une résolution plus faible, on s'attend à ce que la précision des résultats soit moindre, la diversité des formations lithologiques diminuant par rapport à une étude à l'échelle 1:50 000.

La méthode DBA est appliquée sur le jeu de données lithologiques de la CIZ pour produire la grille multivariée. Pour trouver la distance optimale de recherche de voisinage R et la taille de cellule d , la DBA est itérée sur de multiples valeurs de R et R/d , de 2 000 m à 35 000 m, avec un pas de 1 000 m et un rapport R/d allant de 0,71 à 5. La valeur optimale pour la recherche de voisinage est plus élevée, avec

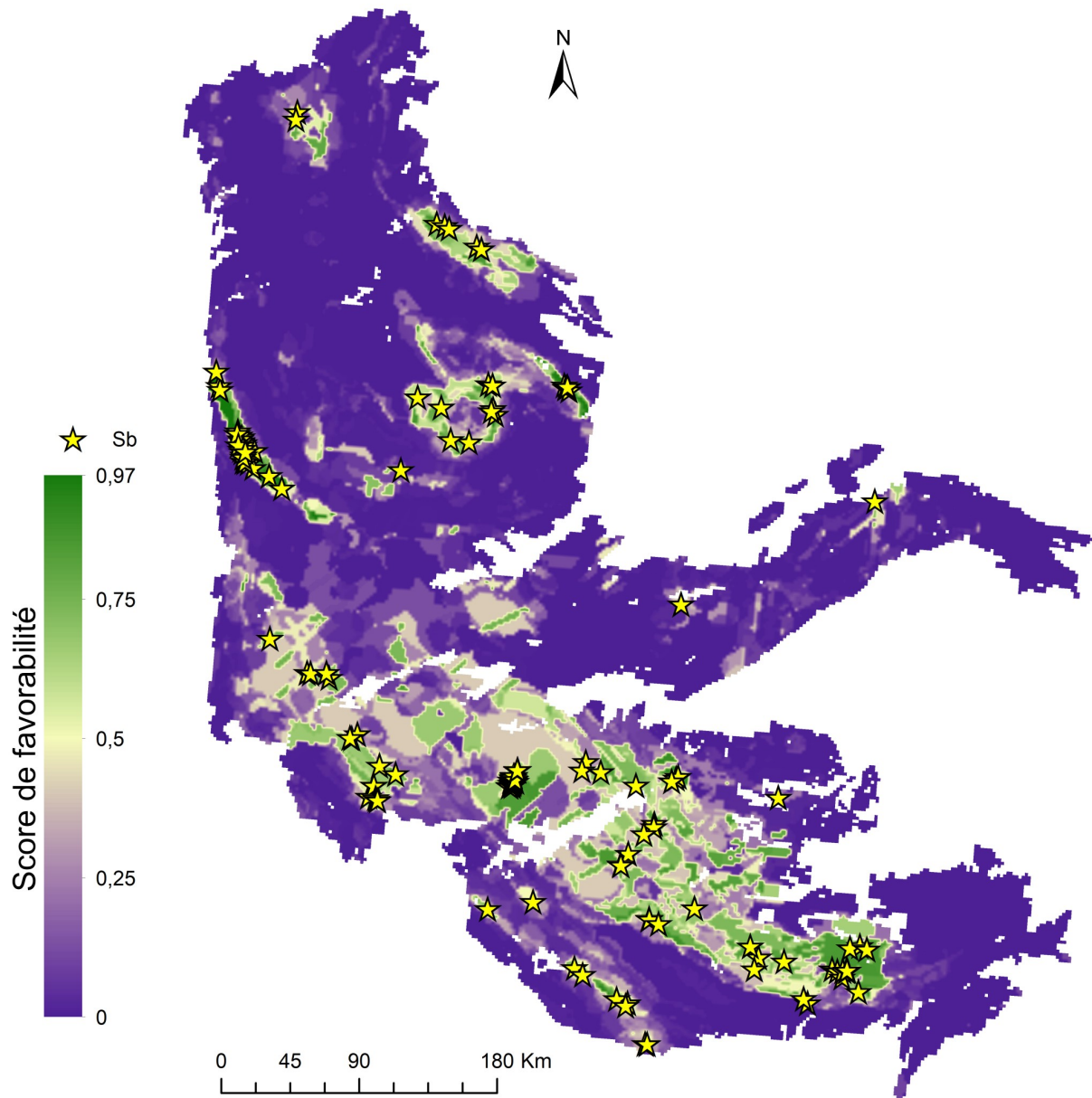


Figure 9.8: Carte de favorabilité de l'antimoine dans la Zone Centre Ibérique (CIZ), produite à partir de la méthode DBA / RF. Avec une recherche par voisinage R de 6 000 m et un R/d de 3, la résolution de la grille est de 2 000 m. Les minéralisation et les failles utilisent une recherche par voisinage R de 2 000 m. Les scores de favorabilité élevés sont affichés en vert, correspondant à une plus grande probabilité que la cellule contienne une minéralisation d'antimoine.

une valeur fixée à $R = 6\,000$ m pour un R/d de 3. Garder le même rapport R/d permet de conserver la même proportion de recouvrement que dans l'application précédente et permet également d'avoir un nombre élevé d'environnements uniques sans augmenter significativement le nombre de cellules. Nous obtenons 3 315 associations lithologiques uniques, avec une résolution finale de 2 000 m pour 43 505 cellules, chaque cellule contenant en moyenne 5,6 lithologies. Une distance de recherche de voisinage de 2 000 m est utilisée pour ajouter les failles et les minéralisations. Ainsi, 37,5% des cellules contiennent des failles et 0,70% contiennent des minéralisations. La proportion de cellules minéralisées est assez similaire à celle du Massif Armoricaïn (0,67%), mais la proportion de failles est presque deux fois plus faible (62%) en raison de l'échelle du jeu de données, contenant les principales failles à l'échelle régionale.

Le modèle prédictif RF est appliqué sur la matrice DBA, contenant les environnements multivariés, sur toutes les cellules, minéralisées et non minéralisées. La méthode "Random Search" est d'abord utilisée pour définir les hyperparamètres optimaux du modèle final. La carte de favorabilité résultante est présentée dans la **Figure 9.8**, les scores de favorabilité élevés étant affichés en vert, avec une valeur maximale de 0,973, et les scores de favorabilité faibles en violet, atteignant un minimum de 0. Les 4 prédicteurs les plus importants (**Figure 7.21**, p.175) sont le complexe schistes-grauwackes Cambriens ("44"), les bassins Carbonifère de Bierzo et Cñera-Matallena ("85"), les granites et leucogranites à deux micas du Carbonifère ("9") et les failles ("Faults"). Comme dans le cas précédent sur le Massif Armoricaïn, les relations entre les prédicteurs et les minéralisations dans la zone d'étude sont différentes. Certains prédicteurs sont efficaces en guise de "métallotectes" en raison de leurs natures localisés (i.e. "85"), couvrant des zones réduites fortement minéralisée, tandis que d'autres sont aussi situés dans des zones spécifiques, mais couvrent de plus grandes surfaces ("44", "9"). Certains sont aussi plus répandus dans la zone d'étude et se retrouvent tout le long de la zone d'étude ("Faults"). L'interprétation de leurs relations avec les minéralisations peut différer en fonction de leurs distributions spatiales. Dans tous les cas, les formations présentant un MDI élevé méritent d'être étudiées.

Comme précédemment, une courbe ROC est produite à partir du modèle final, présentant une AUC de 0,97. Elle est utilisée pour établir le seuil de sélection des cellules les plus favorables. Le G-Mean calculé montre une valeur seuil optimal de 0,51, au-dessus de laquelle les cellules favorables sont sélectionnées. La HCPC est appliquée sur les cellules sélectionnées et 4 clusters sont produits, comme le montre la **Figure 9.9**. Chaque cluster est ensuite analysé en fonction de la fréquence des lithologies comprises dans ses cellules afin d'identifier les différents environnements géologiques favorables.

Le cluster 1 contient 50,6% en surface de la zone favorable. Il est composé de schistes et de quartzites Ordoviciens associés à des schistes noirs Siluriens, dans un environnement faillé. La plupart des cellules sont situées dans la partie SE de la CIZ, avec des minéralisations appartenant aux provinces espagnoles de Cáceres (las Hambrientas), Badajoz (Herrera del Duque, Puebla de Alcocer), Cordoba (Los Accessos, Santa Eufemia), Jaén (San Agustín), et Ciudad Real (Cabezarrubias del Puerto, San Cristobal, Santo Cristo, Nazarena, Navalpozo). La majorité de ces minéralisations sont filoniennes, à l'exception des minéralisations de Los Accessos, Santa Eufemia et Nazarena, qui sont de type stratiforme. Les scores de favorabilité les plus élevés sont situés dans la province orientale de Ciudad Real et dans la province de Cordoba, au Sud. Certaines parties de ce cluster sont situées dans la partie Nord de la CIZ en Espagne, le long du domaine d'Olló de Sapo, avec la province de Lugo (Villabacu) et la province de Zamora (Las Cogollas), cette dernière présentant le score le plus élevé de ce cluster avec une valeur de 0,92.

Le cluster 2 contient 39,8% de la zone favorable. Son environnement est constitué du complexes schistes-grauwackes Cambriens et de granites tardi-orogéniques à biotites et cordiérites. La plupart des

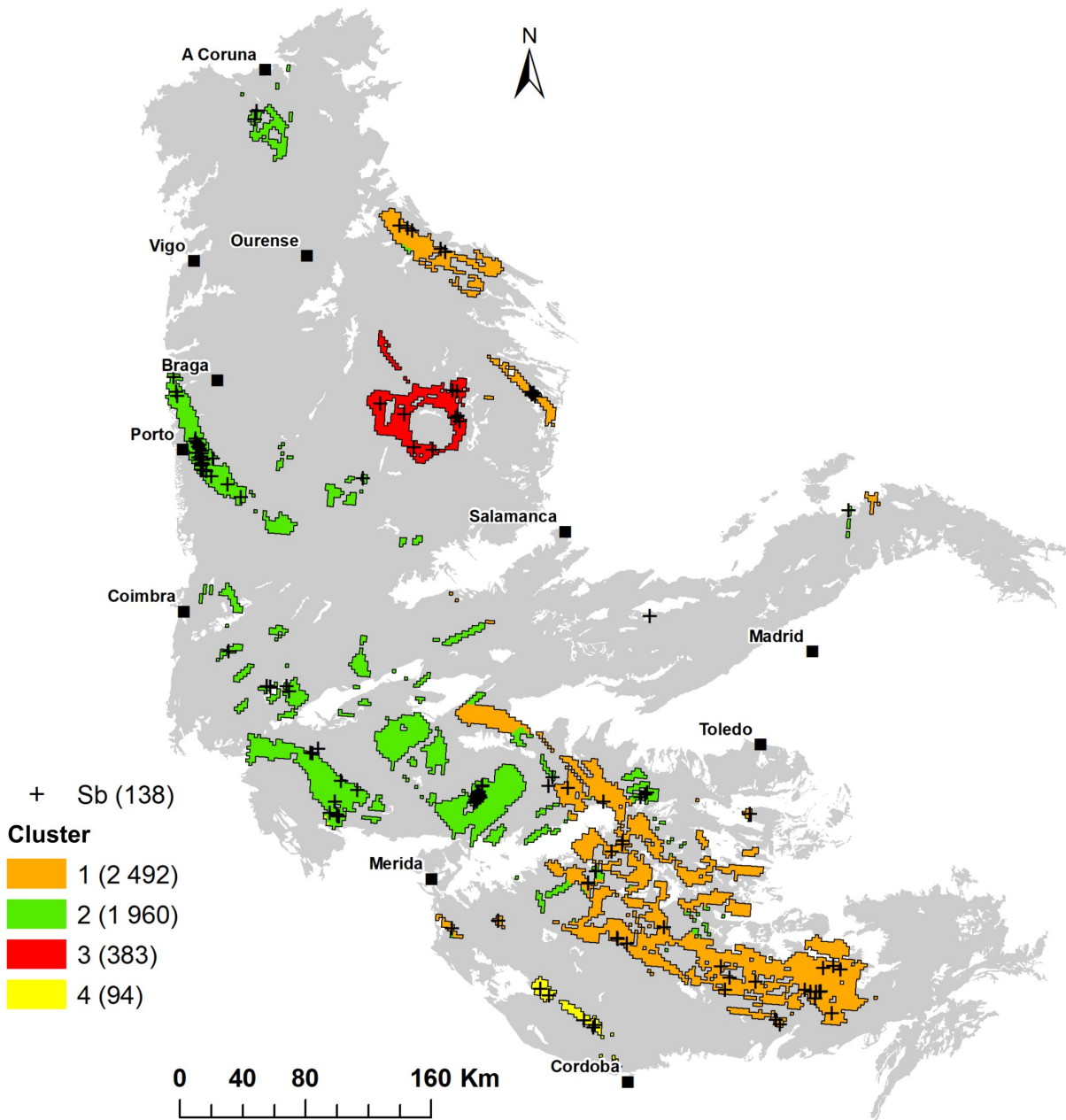


Figure 9.9: Carte des 4 clusters produits par la Classification Hiérarchique sur Composantes Principales (Hierarchical Clustering on Principal Component, HCPC) des cellules favorables (cellules avec un score ≥ 0.51) de la carte de favorabilité DBA / RF de la Zone Centre Ibérique (CIZ). Le nombre de cellules est affiché pour chaque cluster, ainsi que le nombre d'occurrences d'antimoine dans la zone d'étude.

cellules sont situées dans la partie Sud-ouest de la CIZ, avec des minéralisations provenant des provinces de Caceres (Portoviejo, Serafina, Plansenzuela), Badajoz (Mari Pepa, Santa Aurelia, San Antonio) et Ciudad Real (Anchuras de los Montes, Puebla de Alcocer) en Espagne, ainsi que de la minéralisation de Castelo Branco au Nord du Portugal. Toutes ces minéralisations sont filoniennes, à l'exception de la minéralisation de San Antonio, stratiforme [Gumiel and Arribas, 1987]. Les cellules ayant les scores de favorabilité les plus élevés sont situées dans la partie Ouest, près de Porto, dans le district Au-Sb de Dúrico-Beirô (Ribeiro de Igreja, Pinheirinhos, Alto de Sobrido) [Gumiel and Arribas, 1987, Couto et al., 1990], avec un score de 0,97. Ce cluster contient également des cellules situées près de La Coruña dans le nord de l'Espagne (Pandeiro).

Le cluster 3 contient 7,8% de la zone favorable. La lithologie y est dominée par les schistes, quartzites et roches volcaniques de la GTMZ, d'âge Ordovicien à Dévonien. Il est situé dans la partie Nord-est du Portugal, dans le domaine de la GTMZ, et contient 3 minéralisations principales connues : Vimioso, Mogadouro, et Macedo de Cavaleiros. Ces minéralisations filoniennes forment le district de Bragança [Neiva et al., 2008]. Ce cluster est très localisé, la minéralisation étant située dans les unités basales et parautochtones de la GTMZ, formant un anneau autour des unités ophiolitiques. En raison de sa localisation, il présente une signature unique par rapport aux autres clusters.

Le cluster 4 contient 1,9% de la zone favorable. Il est principalement constitué de roches volcaniques Carbonifères, de schistes et de grès avec des failles. Il est situé dans la partie sud de la CIZ, avec des minéralisations appartenant à la province de Cordoba (El Membrillejo, Espiel, Sierra Morena). Ce cluster est très localisé et présente une signature différente des autres en raison de sa proximité avec la zone d'Ossa-Morena. Malgré sa petite taille, ce cluster n'est pas regroupé avec un autre en raison de son environnement géologique très différent, lui donnant des coordonnées extrêmes sur les axes factoriels de la MCA. Son agrégation avec d'autres clusters entraînerait une forte augmentation de l'inertie intra-cluster avec la méthode de Ward, il n'est donc agrégé que tardivement lors de l'AHC.

9.4 Synthèse et conclusion

La méthode DBA est une méthode plus spatiale que statistique, se concentrant sur les associations spatiales locales de variables géologiques, appelées "environnement géologique". Dans cette étude, ces environnements géologiques sont principalement composés de caractéristiques lithologiques et structurales, mais cette notion englobe également d'autres données cartographiques possibles, comme des données géophysiques ou géochimiques. La DBA permet d'intégrer un jeu de données lithologiques dans un processus de cartographie prospective en discrétisant ce jeu de données dans une matrice multivariée. Elle permet de surmonter certains des problèmes inhérents aux jeux de données traités (e.g. imprécision du contour des polygones, effet de regroupement des données...). La méthode DBA a une approche locale par rapport aux autres méthodes de cartographie prédictive minérale, habituellement plus intégratives. Les données lithologiques et structurales sont ici analysées localement et assemblées en une grille de cellules, définissant les environnements géologiques de manière uniforme dans la zone d'étude. Cette identification peut être réalisée à plusieurs échelles avec cette méthode : si lors de cette étude, nous avons cherché à obtenir la plus haute résolution possible, il est aussi possible de produire différents environnements géologiques à plusieurs échelles à partir d'un même jeu de données. Il pourrait ainsi être intéressant d'explorer la possibilité d'étudier l'évolution des environnements géologique en fonction de l'échelle considérée.

Ce travail montre que l'association DBA / RF est capable de produire des cartes de favorabilité basées

sur des environnements géologiques locaux. La RF évalue les environnements géologiques montrant le plus de points communs avec les environnements se trouvant autour de minéralisations connues. Les cartes et scores de favorabilité donnent ainsi un aperçu des relations environnements géologiques / minéralisations : les environnements géologiques avec les scores les plus élevés sont les plus similaires aux environnements minéralisés connus. L'utilisation de la DBA / RF pour produire un modèle prédictif permet une approche plus Data-driven, nécessitant moins de connaissances d'experts. Celles-ci restent cependant nécessaires au cours du processus de la DBA / RF pour définir une sélection de données fiables et significatives à utiliser et lors de la définition des paramètres de la DBA. Néanmoins, en minimisant la quantité de connaissances expertes injectées dans le modèle prédictif, cette approche permet de s'éloigner des modèles métallogéniques connus. Un autre aspect de cette automatisation est l'utilisation de l'ensemble des cellules lors de la classification RF, différenciant ainsi des autres méthodes d'apprentissage automatique utilisant des non-occurrences définies par un expert [Carranza and Laborte, 2015a, 2016, Daviran et al., 2021, Rodriguez-Galiano et al., 2015]. Toutes les cellules ne contenant pas de minéralisation connue sont ainsi considérées comme des non-occurrences lors de l'apprentissage. Malgré le déséquilibre induit entre les classes, la méthode DBA / RF est capable de prédire correctement les zones minéralisées. Dans notre application au Massif Armoricain et à la CIZ, la plupart des occurrences d'antimoine connues ont été mises en évidence. Comparée à des études antérieures menées dans le Massif armoricain avec d'autres méthodes de cartographie prospective, la méthode DBA / RF est aussi plus performante. Elle détermine des zones favorables plus restreintes, permettant de déterminer plus précisément les zones potentiellement minéralisées et donc de déterminer plus précisément leurs relations avec leurs environnements. Si de plus grandes zones avaient été considérées, les environnements géologiques favorables résultants auraient inclus plus de cellules, plus d'associations lithologiques et auraient pu intégrer des environnements géologiques qui n'ont aucun lien avec la minéralisation. Cette précision spatiale est également un grand avantage pour l'exploration minière. Elle permet de mieux cibler les demandes de permis et les campagnes d'exploration, et donc d'optimiser fortement le rapport coût/bénéfice au stade précoce de l'exploration.

L'utilisation du RF permet d'évaluer l'importance des prédicteurs utilisés, permettant une meilleure interprétabilité des résultats obtenus. Cette interprétabilité permet d'étudier le modèle pour avoir un aperçu des relations identifiées entre les environnements géologiques et les minéralisations ciblées. L'évaluation de l'importance des prédicteurs fournie par le RF et l'identification par la HCPC des environnements associés permettent d'analyser ces relations. Le RF quantifie l'importance de chaque prédicteur avec le MDI et l'importance de chaque association avec le score de favorabilité. La HCPC permet une analyse plus qualitative de ces environnements favorables, à un ordre de grandeur plus important. Les prédicteurs aux MDI les plus élevés pourraient être considérés comme des "métallotectes", comme des objets géologiques spécifiques ayant tendance à être liés à la minéralisation. Les métalotectes issues de cette approche Data-driven pourraient être reprises dans des études avec une approche plus Knowledge-driven pour étudier des zones où une application Data-driven serait difficile (i.e. application "Greenfield" dans des zones où les données métallogéniques sont plutôt parcellaires). On pourrait aussi appliquer la CBA / RF dans des zones où les données sont disponibles mais où peu d'études métallogéniques "traditionnelles" ont été réalisées, afin d'identifier automatiquement des métalotectes. L'importance des prédicteurs fournis par des modèles d'apprentissage machine ("Machine Learning", ML) a déjà été mise en évidence pour évaluer les prédicteurs les plus importants et donner un retour sur leur efficacité dans une optique d'optimisation [Wang et al., 2022, Zhang et al., 2022]. Même dans des zones déjà bien étudiées, cette approche peut permettre de préciser les modèles métallogéniques en faisant ressortir des métalotectes alternatifs ou en identifiant des sous-classes dans les métalotectes connus.

L'analyse des environnements géologiques spécifiques entourant les zones favorables est un avantage de la méthode DBA / RF / HCPC. Au lieu de se baser sur des caractéristiques géologiques uniques, comme avec le MDI, nous pouvons étudier les relations entre l'association des prédicteurs spatiaux et les minéralisations. La sélection des environnements favorables sur la base de la courbe ROC permet de cibler uniquement les zones les plus favorables pour en étudier les propriétés. Analyser directement ces zones pour identifier un unique environnement géologique favorable aurait pour conséquence de mélanger les associations géologiques des différents types de minéralisation d'antimoine dans la zone d'étude, indépendamment du type de gisement ou de l'événement métallogénique associé. L'utilisation de la HCPC sur cette sélection permet de différencier dans ces associations géologique favorable les divers environnements pouvant être attribuées à différents types de minéralisation ou à différents districts minéralisés. Chaque environnement (i.e. cluster) peut être analysé en détail afin de souligner les associations spécifiques de prédicteurs qui y sont présents, une "signature" spécifique à cet environnement minéralisé. Cette classification HCPC, appliquée au Massif Armoricaïn et à la CIZ, a été capable de mettre en évidence des environnements géologiques favorables cohérents avec les districts d'antimoine connus. Le traitement des résultats de la classification HCPC nécessite une intervention humaine importante avec, en particulier, le besoin de déterminer le nombre de clusters et d'analyser leurs "signatures".

La DBA / RF / HCPC, appliquée sur les cas d'études, a été capable de mettre en évidence différents environnements géologiques favorables. Elle pu mettre en évidence les environnements géologiques spécifiques à certains districts, comme pour les districts de Vendée, de Quimper dans le Massif Armoricaïn, et les districts de Bragança (cluster n°3) et de Cordoba (cluster n°4). Pour le Massif Armoricaïn, le meilleur niveau de coupe de la HCPC n'a pas permis de distinguer les districts de La Lucette, du Semnon et de Rosnoën, mais leurs environnements géologiques sont assez similaires. Dans la CIZ, les clusters identifiés ont des tailles variables et présentent des environnements marqués par les domaines dans lesquels ils se trouvent. Les clusters 3 et 4 sont localisés dans des lithologies spécifiques à certains domaines du Massif Ibérique (GTMZ et Ossa Morena), tandis que les clusters 1 et 2 regroupent de nombreux districts différents et contiennent des lithologies couvrant de large portion du domaine CIZ. Il semble plus difficile de cibler des districts spécifiques au sein de la CIZ s.s. Une analyse plus poussée des environnements, avec un plus grand nombre de clusters, pourrait permettre d'identifier des environnements géologiques favorables plus spécifiques. Il pourrait aussi être possible d'améliorer les résultats en limitant l'application à la CIZ s.s., pour identifier plus spécifiquement les districts minéralisés, ou en utilisant des données plus précises à une échelle plus grande (1:200 000 ou 1:50 000).

L'utilisation de *Python* et de bibliothèques spécifiques a permis d'automatiser la plupart des traitements de données. Certaines optimisations de la méthode pourraient permettre une plus grande automatisation et une diminution des apports experts, que ce soit lors de la sélection des paramètres de la DBA, de la sélection de la zone favorables ou de l'identification des environnements géologiques favorables. En outre, les résultats de cette méthodologie seraient améliorées par l'ajout d'autres données géologiques et par le traitement de données continues. Ce type de donnée est couramment utilisé pour produire des cartes de favorabilité en MPM avec des données géophysiques, géochimiques ou même satellitaires hyperspectrales. Ces données continues permettraient d'avoir plus de caractéristiques intégrées dans les environnements géologiques locaux, donnant un aperçu plus détaillé des différents prédicteurs au sein de ces environnements. Ces données peuvent être traitées à l'aide du même modèle prédictif RF, bien que certaines modifications des outils intégrés et de la méthode d'analyse factorielle de la HCPC soit nécessaires. D'autres objets géologiques, de nature discrète, pourraient être utilisés en plus de l'information lithologique. Des exemples seraient les isogrades métamorphiques ou les zonations d'altération. Les

données structurales utilisées peuvent également être complétées par l'ajout de structures autres que les failles, comme les plis ou les zones de déformation [Maepa et al., 2021], et par l'ajout de détails supplémentaires sur les événements structuraux (orientation, complexité, âge, etc.) [Parsa et al., 2022, Parsa, 2021, Carranza and Laborte, 2015b, Sun et al., 2020].

Ce travail de thèse a permis le développement d'une nouvelle méthodologie permettant de produire des cartes de favorabilités et de faire ressortir des associations de variables cartographiques géologiques, ou "environnements géologiques", favorablement liées à des minéralisations. L'analyse effectuée par cette méthode permet ainsi de faire ressortir des métallotectes potentiels parmi les variables utilisées, en quantifiant leurs importances durant le processus de modélisation prédictive (RF), et en identifiant et discrétisant les associations de variables favorables lors d'une phase de post-traitement (HCPC). Le développement et l'intégration de la DBA dans cette méthodologie a permis l'identification d'environnements géologiques locaux à partir des différentes couches de données cartographiques disponibles. Ces environnements sont utilisés pour alimenter le modèle prédictif de RF sur lequel repose cette nouvelle méthodologie. Les résultats obtenus avec ce modèle prédictif permettent de mieux contraindre les relations entre les environnements identifiés et les minéralisations cibles connues, le score de favorabilité quantifiant cette relation. La HCPC permet la discrétisation de ces environnements pour identifier des associations types entre les variables et les minéralisations. L'automatisation de ce processus et l'utilisation de méthodes d'apprentissage machine permettent une approche Data-driven et inverse de cartographie prédictive, se basant sur l'analyse de données pour établir des relations entre variables spatiales et minéralisations, plutôt que sur un modèle conceptuel utilisant des connaissances expertes pré-existantes

Acronyms

AHC Agglomerative Hierarchical Clustering or Ascendant Hierarchical Clustering

AI Artificial Intelligence

AM Armorican Massif

ANN Artificial Neural Network

ARTN Adaptive Resonance Theory Network

AUC Area Under the Curve

AUREOLE tArgeting eU cRitical mEtals and predictibility of Sb-As-Hg enviroNmentaL issuE

BM Bohemian Massif

BPNN Back-Propagation Neural Network

BRGM Bureau des Recherches Géologiques et Minières

BWM-ARAS Best-Worst Method and Additive Ratio Assessment

CA Cantabrian Arc

CAD Central Armorican Domain

CBA Cell-Based Association

CIA Central-Iberian Arc

CIZ Central Iberian Zone

CNN Convolutional Neural Networks

CRM Critical Raw Materials

CZ Cantabrian Zone

DAN Deep Autoencoder Network

DBA Disc-Based Association

EDA Exploratory Data Analysis

EU European Union

FAMD Factorial Analysis of Mixed Data

FCUP Faculdade de Ciências da Universidade do Porto

FIS Fuzzy Inference System

FL Fuzzy Logic

FMC French Massic Central

FN False Negative

FPR False Positive Rate

FP False Positive

FSA Frequency in Study Area

FSC Frequency in Standard Cells

GIS Geographic Information Systems

GMDH Group Method of Data Handling

GTMZ Galicia Trás-os-Montes

HCPC Hierarchical Clustering on Principal Components

HP High Pressure

HT High Temperature

IAA Ibero-Armorican Arc

IGeA-UCLM Escuela de Ingeniería Minera e Industrial de Almadén of the University of Castilla-La Mancha

IM Iberian Massif

ISTO Institut des Sciences de la Terre d'Orléans

LP Low Pressure

LR Logistic Regression

LT Low Temperature

MCA Multiple Correspondence Analysis

MDI Mean Decrease in Impurity

MLP Multilayer Perceptron

ML Machine Learning

MNSZ Montagnes Noire Shear Zone

MPM Mineral Prospectivity Mapping

MSE Mean Squared Error

NAD North Armorican Domain

NASZ North-Armorican Shear Zone

OMZ Ossa Morena Zone

OOB Out Of Bag error

PCA Principal Component Analysis

PNN Probabilistic Neural Network

PPA Percentage of Prospective Area

PRPPA Prediction Rate and Percentage of Prospective Area

PR Prediction Rate

RBFNN Radial Basis Function Neural Network

RBM Restricted Boltzmann Machine

RCN Recurrent Neural Network

RF Random Forest

ROC Receiver Operating Characteristic

SAD South Armorican Domain

SASZ South-Armorican Shear Zone

SCV Stratified Cross Validation

SD Standard Deviation

SMOTE Synthetic Minority Over-sampling Technique

SOM Self-Organizing Maps

SPZ South Portuguese Zone

SVM Support Vector Machines

TN True Negative

TPR True Positive Rate

TP True Positive

VMS Volcanogenic Massive Sulfide

WALZ West Asturian Leones Zone

WIO Weighted Index Overlay

WoE Weight of Evidence method

Bibliography

World Exploration Trends March 2021. *S&P Global Market Intelligence*, page 13, Mar. 2021.

M. Abedi, G.-H. Norouzi, and A. Bahroudi. Support vector machine for multi-classification of mineral prospectivity areas. *Computers & Geosciences*, 46:272–283, Sept. 2012. ISSN 00983004. doi: 10.1016/j.cageo.2011.12.014.

D. G. Aerden. Correlating deformation in Variscan NW-Iberia using porphyroblasts; implications for the Ibero-Armorican Arc. *Journal of Structural Geology*, 26(1):177–196, Jan. 2004. ISSN 01918141. doi: 10.1016/S0191-8141(03)00070-1.

C. C. Aggarwal. *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing, Cham, 2018. ISBN 978-3-319-94462-3 978-3-319-94463-0. doi: 10.1007/978-3-319-94463-0.

F. Agterberg, G. Bonham-Carter, and D. Wright. Statistical Pattern Integration for Mineral Exploration. In *Computer Applications in Resource Estimation*, pages 1–21. Pergamon Press, Oxford-New York, Oxford-New York, gaál, g., merriam, d.f. (eds.) edition, 1990. ISBN 978-0-08-037245-7. doi: 10.1016/B978-0-08-037245-7.50006-8.

F. Agterberg, G. Bonham-Carter, Q. Cheng, and D. Wright. Weights of evidence modeling and weighted logistic regression for mineral potential mapping. In *Computers in Geology—25 Years of Progress*, pages 13–32. Oxford Univ. Press, New York, davis, j. c., and herzfeld, u. c edition, 1993.

M. W. Ahmad, M. Mourshed, and Y. Rezgui. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, 147:77–89, July 2017. ISSN 03787788. doi: 10.1016/j.enbuild.2017.04.038.

A. Almeida, H. Martins, and F. Noronha. Hercynian Acid Magmatism and Related Mineralizations in Northern Portugal. *Gondwana Research*, 5(2):423–434, Apr. 2002. ISSN 1342937X. doi: 10.1016/S1342-937X(05)70733-4.

P. An, W. M. Moon, and A. Rencz. Application of Fuzzy Set Theory to Integrated Mineral Exploration. *Canadian Journal of Exploration Geophysics*, 27(1):1–11, 1991.

A. Arabameri, W. Chen, M. Loche, X. Zhao, Y. Li, L. Lombardo, A. Cerda, B. Pradhan, and D. T. Bui. Comparison of machine learning models for gully erosion susceptibility mapping. *Geoscience Frontiers*, 11(5):1609–1620, Sept. 2020. ISSN 16749871. doi: 10.1016/j.gsf.2019.11.009.

R. Arenas, S. Sánchez Martínez, R. Díez Fernández, A. Gerdes, J. Abati, J. Fernández-Suárez, P. Andon-aegui, P. González Cuadra, A. López Carmona, R. Albert, J. M. Fuenlabrada, and F. J. Rubio Pascual. Allochthonous terranes involved in the Variscan suture of NW Iberia: A review of their origin

- and tectonothermal evolution. *Earth-Science Reviews*, 161:140–178, Oct. 2016. ISSN 00128252. doi: 10.1016/j.earscirev.2016.08.010.
- M. Argüelles, C. Benavides, and I. Fernández. A new approach to the identification of regional clusters: Hierarchical clustering on principal components. *Applied Economics*, 46(21):2511–2519, July 2014. ISSN 0003-6846. doi: 10.1080/00036846.2014.904491.
- M. Aria, C. Cuccurullo, and A. Gnasso. A comparison among interpretative proposals for Random Forests. *Machine Learning with Applications*, 6:100094, Dec. 2021. ISSN 26668270. doi: 10.1016/j.mlwa.2021.100094.
- M. Arias, P. Gumiel, D. J. Sanderson, and A. Martin-Izard. A multifractal simulation model for the distribution of VMS deposits in the Spanish segment of the Iberian Pyrite Belt. *Computers & Geosciences*, 37(12):1917–1927, Dec. 2011. ISSN 0098-3004. doi: 10.1016/j.cageo.2011.07.012.
- A. Arribas and P. Gumiel. First Occurrence of a Strata-Bound Sb—W—Hg Deposit in the Spanish Hercynian Massif. In *Syngeneses and Epigenesis in the Formation of Mineral Deposits*, pages 468–481. Springer-Verlag Berlin Heidelberg, a. wauschkuhn et al. edition, 1984.
- F. Arthaud and P. Matte. Les décrochements tardi-hercyniens du sud-ouest de l’europe. Géometrie et essai de reconstitution des conditions de la déformation. *Tectonophysics*, 25(1-2):139–171, Jan. 1975. ISSN 00401951. doi: 10.1016/0040-1951(75)90014-1.
- H. H. Asadi and M. Hale. A predictive GIS model for mapping potential gold and base metal mineralization in Takab area, Iran. *Computers & Geosciences*, 27(8):901–912, Oct. 2001. ISSN 0098-3004. doi: 10.1016/S0098-3004(00)00130-8.
- P. M. Ashley, D. Craw, M. K. Tighe, and N. J. Wilson. Magnitudes, spatial scales and processes of environmental antimony mobility from orogenic gold–antimony mineral deposits, Australasia. *Environmental Geology*, 51(4):499–507, Nov. 2006. ISSN 0943-0105, 1432-0495. doi: 10.1007/s00254-006-0346-6.
- A. Audétat and M. Edmonds. Magmatic-Hydrothermal Fluids. *Elements*, 16(6):401–406, Dec. 2020. ISSN 1811-5217, 1811-5209. doi: 10.2138/gselements.16.6.401.
- B. Auvray, R. Charlot, and P. Vidal. Données nouvelles sur le Protérozoïque inférieur du domaine nordarmoricaïn (France): âge et signification. *Canadian Journal of Earth Sciences*, (17):532–538, 1980a.
- B. Auvray, J. Macé, P. Vidal, and R. Van der Voo. Rb-Sr dating of the Plouézec volcanics, N. Brittany: Implications for the age of the red beds (“Séries rouges”) in the northern Armorican Massif. *Journal of the Geological Society of London*, 137:207–210, 1980b.
- M. A. D. Balda, J. R. M. Catalan, and P. A. Arribas. Syn-collisional extensional collapse parallel to the orogenic trend in a domain of steep tectonics: The Salamanca Detachment Zone (Central Iberian Zone, Spain). *Journal of Structural Geology*, 17(2):163–182, Feb. 1995. ISSN 0191-8141. doi: 10.1016/0191-8141(94)E0042-W.
- J.-F. Ballard, J.-P. Brun, and J. Durand. La discordance Briovérien-Paléozoïque en Bretagne centrale : Signature d’un épisode de distension ordovicienne. *Comptes-Rendus de l’Académie des Sciences de Paris II*, 303:1327–1332, 1986.

- M. Ballèvre, P. Pitra, and M. Bohn. Lawsonite growth in the epidote blueschists from the Île de Groix (Armorican Massif, France) : A potential barometer. *Journal of metamorphic Geology*, 21:723–735, 2003.
- M. Ballèvre, V. Bosse, C. Ducassou, and P. Pitra. Palaeozoic history of the Armorican Massif: Models for the tectonic evolution of the suture zones. *Comptes Rendus Geoscience*, 341(2-3):174–201, Feb. 2009. ISSN 16310713. doi: 10.1016/j.crte.2008.11.009.
- M. Ballèvre, V. Bosse, M.-P. Dabard, C. Ducassou, S. Fourcade, J.-L. Paquette, J.-J. Peucat, and P. Pitra. Histoire Géologique du massif Armoricain: Actualité de la recherche. *Bulletin de la Société Géologique et Minéralogique de Bretagne*, 10–11:93, 2013. doi: <https://hal-insu.archives-ouvertes.fr/insu-00873116>.
- M. Ballèvre, J. R. Martínez Catalán, A. López-Carmona, P. Pitra, J. Abati, R. D. Fernández, C. Ducassou, R. Arenas, V. Bosse, P. Castiñeiras, J. Fernández-Suárez, J. Gómez Barreiro, J.-L. Paquette, J.-J. Peucat, M. Poujol, G. Ruffet, and S. Sánchez Martínez. Correlation of the nappe stack in the Ibero-Armorican arc across the Bay of Biscay: A joint French–Spanish project. *Geological Society, London, Special Publications*, 405(1):77–113, 2014. ISSN 0305-8719, 2041-4927. doi: 10.1144/SP405.13.
- C. Ballouard. *Origine, Évolution et Exhumation Des Leucogranites Peralumineux de La Chaîne Hercynienne Armoricaine: Implication Sur La Métallogénie de l’uranium*. PhD thesis, Université de Rennes 1, Rennes, France, 2016.
- C. Ballouard, P. Boulvais, M. Poujol, D. Gapais, P. Yamato, R. Tartèse, and M. Cuney. Tectonic record, magmatic history and hydrothermal alteration in the Hercynian Guérande leucogranite, Armorican Massif, France. *Lithos*, 220–223:1–22, Apr. 2015. ISSN 00244937. doi: 10.1016/j.lithos.2015.01.027.
- C. Ballouard, M. Poujol, P. Boulvais, J. Mercadier, R. Tartèse, T. Venneman, E. Deloule, M. Jolivet, I. Kéré, M. Cathelineau, and M. Cuney. Magmatic and hydrothermal behavior of uranium in syntectonic leucogranites: The uranium mineralization associated with the Hercynian Guérande granite (Armorican Massif, France). *Ore Geology Reviews*, 80:309–331, Jan. 2017a. ISSN 01691368. doi: 10.1016/j.oregeorev.2016.06.034.
- C. Ballouard, M. Poujol, P. Boulvais, and A. Zeh. Crustal recycling and juvenile addition during lithospheric wrenching: The Pontivy-Rostrenen magmatic complex, Armorican Massif (France), Variscan belt. *Gondwana Research*, 49:222–247, 2017b. doi: 10.1016/j.gr.2017.06.002.
- C. Ballouard, M. Poujol, J. Mercadier, E. Deloule, P. Boulvais, J. M. Baele, M. Cuney, and M. Cathelineau. Uranium metallogenesis of the peraluminous leucogranite from the Pontivy-Rostrenen magmatic complex (French Armorican Variscan belt): The result of long-term oxidized hydrothermal alteration during strike-slip deformation. *Mineralium Deposita*, 53(5):601–628, June 2018. ISSN 0026-4598, 1432-1866. doi: 10.1007/s00126-017-0761-5.
- J. Baptiste. *Cartographie structurale et lithologique du substratum du Bassin parisien et sa place dans la chaîne varisque de l’Europe de l’Ouest : approches combinées géophysiques, pétrophysiques, géochronologiques et modélisations 2D*. PhD thesis, Sciences de la Terre, Université d’Orléans, Orléans, France, 2016.
- J. Bard, P. Matte, and R. Capdevila. La structure de la Chaîne Hercynienne de la Meseta Ibérique: Comparaison avec les segments voisins. *Publ. Inst. Fr. Pétrole*, 1(22):68, 1971.

- A. Barth, A. Knobloch, S. Noack, P. Hielscher, S. Etzold, and A. Brosig. Mineral predictive mapping - from intuition to quantitative hybrid 3D modelling. In *Mineral Prospectivity, Current Approaches and Future Innovations*, page 3, Orléans, France, Oct. 2017.
- F. Béchenec. Carte géologique harmonisée du département de Loire-Atlantique. Technical Report BRGM/RP- 55703-FR, BRGM, Orléans, France, Sept. 2007.
- F. Béchenec. Carte géologique harmonisée du département de Vendée. Technical Report BRGM/RP - 57473 - FR, BRGM, Orléans, France, July 2009.
- C. Bénard. *Forêts aléatoires et interprétabilité des algorithmes d'apprentissage*. PhD thesis, Statistiques [math.ST], Sorbonne Université, Paris, France, 2022.
- J. Berger, O. Féménias, D. Ohnenstetter, O. Bruguier, G. Plissart, J.-C. C. Mercier, and D. Demaiffe. New occurrence of UHP eclogites in Limousin (French Massif Central): Age, tectonic setting and fluid-rock interactions. *Lithos*, 118(3-4):365–382, Aug. 2010. ISSN 00244937. doi: 10.1016/j.lithos.2010.05.013.
- J. Berkson. Application of the Logistic Function to Bio-Assay. *Journal of the American Statistical Association*, 39(227):357–365, 1944. ISSN 0162-1459. doi: 10.2307/2280041.
- D. Berthé, P. Choukroune, and P. Jégouzo. Orthogneiss, mylonite and non coaxial deformation of granites: The example of the South Armorican Shear Zone. *Journal of Structural Geology*, 1(1):31–42, 1979.
- A. Bertram. *PowerShell for Sysadmins: Workflow Automation Made Easy*. No Starch Press, San Francisco, 2020. ISBN 978-1-59327-918-9.
- G. Bertrand, J.-J. Dupuy, J. Melleton, B. Tourlière, D. Cassard, A.-S. Audion, J.-M. Angel, E. Gloaguen, Y. Husson, and H. Berthier. Réévaluation du potentiel français en ressources minérales: Retraitement des données géochimiques de l'Inventaire et établissement de fiches de cibles minières. Technical Report BRGM/RP-62960-FR, Bureau des Recherches Géologiques et Minières, Orléans, 2013.
- G. Bertrand, N. Charles, J. Melleton, B. Tourlière, D. Cassard, J.-J. Dupuy, E. Gloaguen, C. Loiselet, J.-M. Angel, and H. Berthier. Réévaluation du potentiel français en ressources minérales : Retraitement des données géochimiques de l'Inventaire et établissement de fiches de cibles minières. Synthèse des travaux réalisés de 2013 à 2015. Rapport final. Technical Report BRGM/RP-65165-FR, BRGM, Orléans, 2015.
- P. Béziat and M. Bornuat. Carte Minière de la France Métropolitaine, 1994.
- M. Billa, D. Cassard, A. Lips, V. Bouchot, B. Tourlière, G. Stein, and L. Guillou-Frottier. Predicting gold-rich epithermal and porphyry systems in the central Andes with a continental-scale metallogenic GIS. *Ore Geology Reviews*, 25(1-2):39–67, Aug. 2004. ISSN 01691368. doi: 10.1016/j.oregeorev.2004.01.002.
- C. P. Birch, S. P. Oom, and J. A. Beecham. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*, 206(3-4):347–359, Aug. 2007. ISSN 03043800. doi: 10.1016/j.ecolmodel.2007.03.041.
- G. A. Blengini, C. E. Latunussa, U. Eynard, C. Torres de Matos, D. Wittmer, K. Georgitzikis, C. Pavel, S. Carrara, L. Mancini, M. Unguru, D. Blagoeva, F. Mathieux, and D. Pennington. Study on the EU's list of Critical Raw Materials – Final Report. Technical report, European Union, Luxembourg, 2020.

- J. Blouin. District antimonifère de Vendée .Rappel sur les réserves démontrées à l'intérieur des divers P.E.R. Technical Report DMM/DL/NANTES RVL/FD92-04, BRGM, 1991.
- D. Blundell, N. Arndt, P. R. Cobbold, and C. Heinrich. Geodynamics and ore deposit evolution in Europe: Introduction. *Ore Geology Reviews*, 27(1-4):5–11, Nov. 2005. ISSN 01691368. doi: 10.1016/j.oregeorev.2005.07.031.
- S. Bobba, S. Carrara, J. Huisman, F. Mathieux, and C. Pavel. Critical Raw Materials for Strategic Technologies and Sectors in the EU : A Foresight Study. Technical report, European Union, Luxembourg, 2020.
- G. Bonham-Carter, F. Agterberg, and D. Wright. Weights of evidence modelling: A new approach to mapping mineral potential. In *Statistical Applications in the Earth Sciences*, Geological Survey of Canada, pages 171–183. Agterberg, f. p.; bonham-carter, g. edition, Jan. 1990.
- G. F. Bonham-Carter. *Geographic Information Systems for Geoscientists: Modelling with GIS*, volume 13 of *Computer Methods in the Geosciences*. Elsevier, 1994. ISBN 978-1-4831-4494-8.
- G. F. Bonham-Carter, F. P. Agterberg, and D. F. Wright. Integration of geological datasets for gold exploration in Nova Scotia. In J. N. Van Driel and J. C. Davis, editors, *Digital Geologic and Geographic Information Systems*, pages 15–23. American Geophysical Union, Washington, D. C., 1989. ISBN 978-0-87590-709-3. doi: 10.1029/SC010p0015.
- B. E. Boser, I. M. Guyon, and V. N. Vapnik. A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152. ACM Press, 1992.
- V. Bouchot, P. Ledru, C. Lerouge, J.-L. Lescuyer, and J.-P. Milesi. 5: Late Variscan mineralizing systems related to orogenic processes: The French Massif Central. *Ore Geology Reviews*, 27(1-4):169–197, Nov. 2005. ISSN 01691368. doi: 10.1016/j.oregeorev.2005.07.017.
- P. Bouton and J. Camuzard. Le Givétien de la Villedé d’Ardin (Sud du Massif armoricain, France) : Une série discordante sur un socle cadomien ? *Annales de la Société Géologique du Nord*, 19:25–34, 2012.
- C. Braux, F. Apolinarski, and J. Blouin. District Aurifère de Saint-Pierre-Montlimart (Vendée, France). Synthèse des travaux d’exploration. Technical Report R32554, BRGM, Orléans, France, 1991.
- L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, Aug. 1996. ISSN 1573-0565. doi: 10.1007/BF00058655.
- L. Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001. doi: 10.1023/A:1010933404324.
- L. Breiman. Setting up, using, and understanding random forests v3.1. page 29, 2003.
- L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification And Regression Trees*. Routledge, New York, 1st edition edition, 1984. ISBN 978-1-315-13947-0. doi: 10.1201/9781315139470.
- BRGM. Ressources minières francaises. Tome 13. Les indices de Mercure (Situation en 1983). Technical Report RR-41430-FR-13, BRGM, Orléans, France, 1983.
- BRGM. Plan minier breton, exercice 1985, rapport de fin de travaux, les recherches minières. Technical Report 87-DAM-018-OP4, BRGM, Orléans, France, 1987.

- BRGM. Plan minier breton, exercice 1986, rapport de fin de travaux, les recherches minières. Technical Report 88-DAM-019-OP4, BRGM, Orléans, France, 1988a.
- BRGM. Inventaire du Territoire Métropolitain. Les minéralisations Sb de La Baussonnière (Vendée). Technical Report DAM/DL/NANTES. RVL/FD 91-01, BRGM, Orléans, France, 1988b.
- BRGM. Plan minier breton, convention 1988, rapport de fin de travaux, les recherches minières. Technical Report RR-30413-FR, BRGM, Orléans, France, 1990.
- F. Briatte and A. Soetewey. Classification ascendante hiérarchique (CAH), May 2021.
- V. E. Brimkov and R. P. Barneva. “Honeycomb” vs Square and Cubic Models. *Electronic Notes in Theoretical Computer Science*, 46:321–338, Aug. 2001. ISSN 15710661. doi: 10.1016/S1571-0661(04)80995-2.
- M. Brown. The generation, segregation, ascent and emplacement of granite magma: The migmatite-to-crustally-derived granite connection in thickened orogens. *Earth Science Revue*, 36:83–130, 1994.
- M. Brown. Metamorphic patterns in orogenic systems and the geological record. In *Earth Accretionary Systems in Space and Time.*, number 318 in Geological Society, pages 37–74. Special Publications, London, cawood, p. a. and kroner, a. edition, 2009.
- W. Brown and D. Groves. Use of Fuzzy Membership Input Layers to Combine Subjective Geological Knowledge and Empirical Data in a Neural Network Method for Mineral-Potential Mapping. *Natural Resources Research*, page 18, 2003.
- W. M. Brown, T. D. Gedeon, D. I. Groves, and R. G. Barnes. Artificial neural networks: A new method for mineral prospectivity mapping. *Australian Journal of Earth Sciences*, 47(4):757–770, Aug. 2000. ISSN 0812-0099, 1440-0952. doi: 10.1046/j.1440-0952.2000.00807.x.
- J.-P. Brun, J.-F. Ballard, and C. Le Corre. Identification of Ordovician block-tilting in the Hercynian foldbelt of Central Brittany (France) : Field evidence and computer models. *Journal of Structural Geology*, 13:419–429, 1991.
- C. Bulois, M. Pubellier, N. Chamot-Rooke, and L. Watremez. From orogenic collapse to rifting: A case study of the northern Porcupine Basin, offshore Ireland. *Journal of Structural Geology*, 114:139–162, Sept. 2018. ISSN 01918141. doi: 10.1016/j.jsg.2018.06.021.
- J. Burg, P. Bale, J. Brun, and J. Girardeau. Streching lineations and transport direction in the Ibero-Armorican Arc during the Siluro-Devonian collision. *Geodin. Acta*, 1:71–81, 1987.
- J.-P. Burg and M. Ford. Orogeny through time: An overview. *Geological Society, London, Special Publications*, 121(1):1–17, 1997. ISSN 0305-8719, 2041-4927. doi: 10.1144/GSL.SP.1997.121.01.01.
- F. Cagnard. Carte géologique harmonisée du département du Finistère. Technical Report BRGM/RP-56273-FR, BRGM, Orléans, France, Oct. 2008.
- F. Cagnard. Carte géologique harmonisée du département de Morbihan. Technical Report BRGM/RP-56656-FR, BRGM, Orléans, France, Feb. 2009a.
- F. Cagnard. Carte géologique harmonisée du département de l’Ille-et-Vilaine. Technical Report BRGM/RP-57855-FR, BRGM, Orléans, France, Dec. 2009b.

- R. Capdevila. Les granites varisques du Massif Armoricaïn. *Bull Soc Geol Mineral Bretagne*, 7:1–52, 2010.
- S. Carey. The orocline concept in geotectonics. *Proc. R. Soc. Tasmania*, 89:255–288, 1955.
- S. Carey. The tectonic approach to continental drift. In *Continental Drift, a Symposium*, pages 177–356. Univ. Tasmania, carey, s.w. (ed.) edition, 1958.
- M. Caroff, M. Vidal, A. Bénard, and J.-R. Darboux. A late-Ordovician phreatomagmatic complex in softsubstrate environment : The Crozon volcanic system, Armorican Massif (France). *Journal of Volcanology and Geothermal Research*, 184:351–366, 2009.
- E. J. M. Carranza. Weights of Evidence Modeling of Mineral Potential: A Case Study Using Small Number of Prospects, Abra, Philippines. *Natural Resources Research*, 13(3):173–187, Sept. 2004. ISSN 1520-7439. doi: 10.1023/B:NARR.0000046919.87758.f5.
- E. J. M. Carranza. *Geochemical Anomaly and Mineral Prospectivity Mapping in GIS*, volume 11 of *Handbook of Exploration and Environmental Geochemistry*. Elsevier Science, Amsterdam, elsevier edition, 2008. ISBN 978-0-444-51325-0.
- E. J. M. Carranza. Geocomputation of mineral exploration targets. *Computers & Geosciences*, 37(12):1907–1916, 2011. ISSN 0098-3004.
- E. J. M. Carranza. Natural Resources Research Publications on Geochemical Anomaly and Mineral Potential Mapping, and Introduction to the Special Issue of Papers in These Fields. *Natural Resources Research*, 26(4):379–410, Oct. 2017. ISSN 1520-7439, 1573-8981. doi: 10.1007/s11053-017-9348-1.
- E. J. M. Carranza and M. Hale. Logistic Regression for Geologically Constrained Mapping of Gold Potential, Baguio District, Philippines. *Exploration and Mining Geology*, 10(3):165–175, July 2001. ISSN 0964-1823. doi: 10.2113/0100165.
- E. J. M. Carranza and A. G. Laborte. Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of Random Forests algorithm. *Ore Geology Reviews*, page 11, 2015a.
- E. J. M. Carranza and A. G. Laborte. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers & Geosciences*, 74:60–70, 2015b. doi: 10.1016/j.cageo.2014.10.004.
- E. J. M. Carranza and A. G. Laborte. Data-Driven Predictive Modeling of Mineral Prospectivity Using Random Forests: A Case Study in Catanduanes Island (Philippines). *Natural Resources Research*, 25(1):35–50, Mar. 2016. ISSN 1520-7439, 1573-8981. doi: 10.1007/s11053-015-9268-x.
- E. J. M. Carranza and M. Sadeghi. Predictive mapping of prospectivity and quantitative estimation of undiscovered VMS deposits in Skellefte district (Sweden). *Ore Geology Reviews*, 38(3):219–241, Nov. 2010. ISSN 01691368. doi: 10.1016/j.oregeorev.2010.02.003.
- E. J. M. Carranza, J. C. Mangaoang, and M. Hale. Application of Mineral Exploration Models and GIS to Generate Mineral Potential Maps as Input for Optimum Land-Use Planning in the Philippines. *Natural Resources Research*, 8(2):165–173, June 1999. ISSN 1573-8981. doi: 10.1023/A:1021846820568.
- D. Cassard and G. Bertrand. Mineral Prospectivity Conference : Book of abstracts. In *Mineral Prospectivity : Current Approaches and Future Innovations*, page 120, Orléans, France, Oct. 2017. BRGM.

- D. Cassard, M. Billa, A. Lambert, J.-C. Picot, Y. Husson, J.-L. Lasserre, and C. Delor. Gold predictivity mapping in French Guiana using an expert-guided data-driven approach based on a regional-scale GIS. *Ore Geology Reviews*, 34(3):471–500, Nov. 2008. ISSN 01691368. doi: 10.1016/j.oregeorev.2008.06.001.
- D. Cassard, G. Bertrand, J. Monthel, E. Gloaguen, J. Melleton, Y. Husson, M. Billa, and B. Tourlière. Réévaluation du potentiel français en ressources minérales. Elaboration d’une méthodologie et application à une zone d’étude test (sud-ouest du Massif Central). Technical Report RP-61745-FR, BRGM, Dec. 2012.
- D. Cassard, G. Bertrand, M. Billa, J.-J. Serrano, B. Tourlière, J.-M. Angel, and G. Gaál. ProMine Mineral Databases: New Tools to Assess Primary and Secondary Mineral Resources in Europe. In P. Weihed, editor, *3D, 4D and Predictive Modelling of Major Mineral Belts in Europe*, Mineral Resource Reviews, pages 9–58. Springer International Publishing, Cham, 2015. ISBN 978-3-319-17428-0. doi: 10.1007/978-3-319-17428-0_2.
- J. R. M. Catalán. The Central Iberian arc: Implications for the Iberian Massif. *Geogaceta*, 50(1):7–10, 2011.
- J. R. M. Catalán, R. Arenas, F. D. García, P. G. Cuadra, J. Gómez-Barreiro, J. Abati, P. Castiñeiras, J. Fernández-Suárez, S. S. Martínez, P. Andonaegui, E. G. Clavijo, A. D. Montes, F. J. R. Pascual, and B. V. Aguado. Space and time in the tectonic evolution of the northwestern Iberian Massif: Implications for the Variscan belt. In *Geological Society of America Memoirs*, volume 200, pages 403–423. Geological Society of America, Jan. 2007. ISBN 978-0-8137-1200-0. doi: 10.1130/2007.1200(21).
- M. Cathelineau, M.-C. Boiron, and J. Tuduri. Fluides et genèse des concentrations minérales. *Geosciences*, 13:56–63, 2011. doi: <https://hal.archives-ouvertes.fr/hal-00662471/>.
- P. A. Cawood and G. Suhr. Generation and obduction of ophiolites: Constraints from the Bay of Islands Complex, western Newfoundland. *Tectonics*, 11(4):884–897, Aug. 1992. ISSN 02787407. doi: 10.1029/92TC00471.
- P. A. Cawood, A. Kröner, W. J. Collins, T. M. Kusky, W. D. Mooney, and B. F. Windley. Accretionary orogens through Earth history. *Geological Society, London, Special Publications*, 318(1):1–36, 2009. ISSN 0305-8719, 2041-4927. doi: 10.1144/SP318.1.
- V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):15:1–15:58, July 2009. ISSN 0360-0300. doi: 10.1145/1541880.1541882.
- J. Chantaine, F. Béchenec, and D. Rabu. Notice Explicative de la Carte 1/250 000 du Massif Armoricaïn. Technical Report (Unpublished), 2001a.
- J. Chantaine, E. Egal, D. Thiéblemont, E. Le Goff, C. Guerrot, M. Ballèvre, and P. Guennoc. The Cadomian active margin (North Armorican Massif, France) : A segment of the North Atlantic Panafrican belt. *Tectonophysics*, 331:1–18, 2001b.
- N. Charles, J.-J. Dupuy, P. Christmann, R. Galin, and D. Guillon. Industrie minérale et activité minière en France. *Collection « La Mine en France »*, Tome 1:24, 2017.
- L. Chauris. Les associations paragenetiques dans la metallogenie varisque du massif armoricaïn. *Mineralium Deposita*, 12(3), Nov. 1977. ISSN 0026-4598, 1432-1866. doi: 10.1007/BF00206172.

- L. Chauris. Plate tectonics and ore deposits in Western Europe: The example of the Armorican Massif (France). *Ridge JD (ed) Proc of Fifth Quad IAGOD Symp Snowbird, Utah, USA*, pages 401–413, 1978.
- L. Chauris. Les exploitations minières dans le Massif Armoricain. Déclin ou progrès. *Norvès*, 141:5–32, 1989.
- L. Chauris and J. Guigues. Gites minéraux de la France, vol 1. Massif armoricain. *Mémoire BRGM (Orléans)*, 1:74–96, 1969.
- L. Chauris, E. Houlgatte, C. Laforêt, and P. Picot. Un district antimono-aurifère à gangue quartzo-carbonatée: Le Semnon (Ille-et-Vilaine, Massif Armoricain, France). *Hercynica*, 1:111119, 1985.
- M. Chavent. L'Analyse des Correspondances Multiples (ACM). Technical Report Master MIMSE - 2ème année, Université de Bordeaux, Bordeaux, 2015.
- K. Chen, C. Gumiaux, R. Augier, Y. Chen, Q. Wang, W. Lin, and S. Wang. The Mesozoic palaeorelief of the northern Tian Shan (China). *Terra Nova*, 23(3):195–205, 2011. ISSN 1365-3121. doi: 10.1111/j.1365-3121.2011.00999.x.
- Y. Chen. MRPM: Three visual basic programs for mineral resource potential mapping. *Computers & Geosciences*, 30(9-10):969–983, Nov. 2004. ISSN 00983004. doi: 10.1016/j.cageo.2004.07.002.
- Y. Chen and W. Wu. Mapping mineral prospectivity using an extreme learning machine regression. *Ore Geology Reviews*, 80:200–213, Jan. 2017. ISSN 01691368. doi: 10.1016/j.oregeorev.2016.06.033.
- F. Cheval-Garabedian. *Les minéralisations à Sb et AU tardi-varisques : vers un modèle génétique unifié ? Exemples du MASSIF armoricain et du MASSIF central*. PhD thesis, Université d'Orléans, Institut des Sciences de la Terre d'Orléans, 2019.
- F. Cheval-Garabédian, M. Faure, E. Marcoux, J. Gouin, and M. Picault. The La Bellière gold and antimony district (French Armorican Massif): A two-stage evolution model controlled by Variscan strike-slip tectonic. *Ore Geology Reviews*, 125:103681, Oct. 2020. ISSN 01691368. doi: 10.1016/j.oregeorev.2020.103681.
- B. Choubert. Recherche sur la genèse des chaînes paléozoïque et antécambriennes. *Rev. Géogr. Phys. Géol. Dyn*, 8(1):1–50, 1935.
- B. Cochelin. *Champ de déformation du socle paléozoïque des Pyrénées*. ED SDU2E : Sciences de la Terre et des Planètes Solides, Université Toulouse 3 Paul Sabatier, Toulouse, France, 2016.
- J. Cogné. L'Arc hercynien ibéro-armoricain. *Publ. Inst. Fr. Pétrole*, 1(22):1–23, 1967.
- J. Cogné. Le massif Armoricain et sa place dans la structure des socles ouest européens: l'arc hercynien ibéro-armoricain. In *Histoire Structurale Du Golfe de Gascogne*, volume 1, pages 1–24. Editions Technip, Paris, debyser (ed.) edition, 1971.
- COM(2020) 474 final. Communication de la Commission au Parlement Européen, au Conseil, au Comité économique et social européen et au comité des Régions : Résilience des matières premières critiques: la voie à suivre pour un renforcement de la sécurité et de la durabilité. Technical report, Commission Européenne, Bruxelles, 2020.

- A. Combes, D. Cassard, H. Couto, J. Damiao, P. Ferraz, and P. Urien. Caractérisation structurale des minéralisations aurifères de l'Arénigien dans la région de Valongo (Baixo Douro, Portugal). *Chronique de la Recherche Minière*, (509):3–15, 1992.
- H. Couto and F. S. Borges. Stratiform Sb and Au mineralizations in the Hercynian Dúrico-Beirã area (North Portugal). In J. Mao and F. P. Bierlein, editors, *Mineral Deposit Research: Meeting the Global Challenge*, pages 97–99. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. ISBN 978-3-540-27945-7 978-3-540-27946-4. doi: 10.1007/3-540-27946-6_25.
- H. Couto and G. Roger. Palaeozoic Magmatism Associated with Gold-Antimony-Tin-Tungsten-Lead-Zinc and Silver Mineralization in the Neighbouring of Porto, Northern Portugal. *IOP Conference Series: Earth and Environmental Science*, 95:022054, Dec. 2017. ISSN 1755-1307, 1755-1315. doi: 10.1088/1755-1315/95/2/022054.
- H. Couto, G. Roger, Y. Moëlo, and H. Bril. Le district à antimoine-or Dúrico-Beirão (Portugal): évolution pragénétique et géochimique; implications métallogéniques. *Mineralium Deposita*, (25):69–81, 1990.
- S. F. Cox, M. A. Etheridge, and V. J. Wall. The role of fluids in syntectonic mass transport, and the localization of metamorphic vein-type ore deposits. *Ore Geology Reviews*, 2:65–86, 1986.
- M.-P. Dabard and B. Simon. Discordance des Séries Rouges Initiales sur le socle briovérien : Exemple de la carrière des Landes. *Bulletin de la Société Géologique et Minéralogique de Bretagne*, 8:33–44, 2011.
- M.-P. Dabard, A. Loi, and F. Paris. Relationship between phosphogenesis and sequence architecture: Sequence stratigraphy and biostratigraphy in the Middle Ordovician of the Armorican Massif (NW France). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 248(3-4):339–356, May 2007. ISSN 00310182. doi: 10.1016/j.palaeo.2006.12.011.
- R. D. Dallmeyer and E. M. Garcia, editors. *Pre-Mesozoic Geology of Iberia*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1990. ISBN 978-3-642-83982-5 978-3-642-83980-1. doi: 10.1007/978-3-642-83980-1.
- S. Das and A. Gupta. Multi-criteria decision based geospatial mapping of flood susceptibility and temporal hydro-geomorphic changes in the Subarnarekha basin, India. *Geoscience Frontiers*, 12(5):101206, Sept. 2021. ISSN 16749871. doi: 10.1016/j.gsf.2021.101206.
- M. Daviran, A. Maghsoudi, R. Ghezelbash, and B. Pradhan. A new strategy for spatial predictive mapping of mineral prospectivity: Automated hyperparameter tuning of random forest approach. *Computers & Geosciences*, 148:104688, Mar. 2021. ISSN 00983004. doi: 10.1016/j.cageo.2021.104688.
- A. P. Dempster. A Generalization of Bayesian Inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30(2):205–247, 1968. ISSN 0035-9246.
- J. G. Dennis. International Tectonic Dictionary: English Terminology. *American Association of Petroleum Geologists*, 7:196, Jan. 1967. doi: 10.1306/M7362.
- J. F. Dewey. Evolution of the Appalachian–Caledonian orogen. *Nature*, 222:124–129, 1969.
- J. F. Dewey. Extensional collapse of orogens. *Tectonics*, 7(6):1123–1139, Dec. 1988. ISSN 02787407. doi: 10.1029/TC007i006p01123.
- J. F. Dewey and H. Spall. Pre-Mesozoic plate tectonics. *Geology*, 3:422–424, 1975.

- R. Dias and A. Ribeiro. The Ibero-Armorican Arc: A collision effect against an irregular continent? *Tectonophysics*, 246(1):113–128, June 1995. ISSN 0040-1951. doi: 10.1016/0040-1951(94)00253-6.
- R. Dias, A. Ribeiro, J. Romão, C. Coke, and N. Moreira. A review of the arcuate structures in the Iberian Variscides; constraints and genetic models. *Tectonophysics*, 681:170–194, June 2016. ISSN 00401951. doi: 10.1016/j.tecto.2016.04.011.
- Í. Dias da Silva, J. Gómez-Barreiro, J. R. Martínez Catalán, P. Ayarza, J. Pohl, and E. Martínez. Structural and microstructural analysis of the Retortillo Syncline (Variscan belt, Central Iberia). Implications for the Central Iberian Orocline. *Tectonophysics*, 717:99–115, Oct. 2017. ISSN 00401951. doi: 10.1016/j.tecto.2017.07.015.
- X. Ding, D. E. Harlov, B. Chen, and W. Sun. Fluids, Metals, and Mineral/Ore Deposits. *Geofluids*, 2018:1–6, 2018. ISSN 1468-8115, 1468-8123. doi: 10.1155/2018/1452409.
- J. DiPietro. *Geology and Landscape Evolution*. Elsevier, 2018. ISBN 978-0-12-811191-8.
- J. S. Dramsch. 70 years of machine learning in geoscience in review. In *Advances in Geophysics*, volume 61, pages 1–55. Elsevier, 2020. ISBN 978-0-12-821669-9. doi: 10.1016/bs.agph.2020.08.002.
- D. Dubois, P. Hájek, and H. Prade. Knowledge-Driven versus Data-Driven Logics. *Journal of Logic, Language, and Information*, 9:65–89, 2000.
- J. Durand. Le Grès Armoricaín. Sédimentologie, traces fossiles, milieux de dépôt. *Mémoires du Centre Armoricaín d’Etude Structurale des Socles*, 3:1–189, 1985.
- E. Egal. Carte géologique harmonisée du département des Côtes d’Armor. Technical Report BRGM/RP-58657-FR, BRGM, Orléans, France, June 2010.
- E. Egal, E. Le Goff, P. Guennoc, P. Lebre, D. Thiéblemont, B. Hallégouët, E. Houlgatte, L. Callier, and A. Carn. Notice explicative de la carte géologique de la France au 1:50000, feuille Pontrioux-Etablessur-Mer. Technical Report 204, BRGM, Orléans, 1995.
- T. Euzen. Pétrogenèse des granites de collision post-épaississement. Le cas des granites crustaux et mantelliques du complexe de Pontivy-Rostrenen (Massif Armoricaín, France). *Mémoires Géosciences Rennes*, 51, 1993.
- L. Feltrin. Predictive modelling of prospectivity for Pb–Zn deposits in the Lawn Hill Region, Queensland, Australia. *Ore Geology Reviews*, 34(3):399–427, Nov. 2008. ISSN 01691368. doi: 10.1016/j.oregeorev.2008.05.002.
- R. D. Fernández and M. F. Pereira. Extensional orogenic collapse captured by strike-slip tectonics: Constraints from structural geology and UPb geochronology of the Pinhel shear zone (Variscan orogen, Iberian Massif). *Tectonophysics*, 691:290–310, Nov. 2016. ISSN 00401951. doi: 10.1016/j.tecto.2016.10.023.
- M. Fernandez-Delgado, E. Cernadas, S. Barro, and D. Amorim. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research*, 15:3133–3181, 2014.
- A. Ford, S. G. Hagemann, A. S. Fogliata, J. M. Miller, A. Mol, and P. J. Doyle. Porphyry, epithermal, and orogenic gold prospectivity of Argentina. *Ore Geology Reviews*, 71:655–672, Dec. 2015. ISSN 0169-1368. doi: 10.1016/j.oregeorev.2015.05.013.

- M. Fouquet. *Les districts antimonifères de Quimper et du Cap Sizun (Finistère) leur place dans la métallogénie de l'antimoine dans le Massif Armoricain (France)*. PhD thesis, Université de Clermont-Ferrand, Clermont-Ferrand, France, 1980.
- W. Franke, L. R. M. Cocks, and T. H. Torsvik. The Palaeozoic Variscan oceans revisited. *Gondwana Research*, 48:257–284, Aug. 2017. ISSN 1342937X. doi: 10.1016/j.gr.2017.03.005.
- G. Fu, Q. Lü, J. Yan, C. G. Farquharson, G. Qi, K. Zhang, Y. Zhang, H. Wang, and F. Luo. 3D mineral prospectivity modeling based on machine learning: A case study of the Zhuxi tungsten deposit in northeastern Jiangxi Province, South China. *Ore Geology Reviews*, 131:104010, Apr. 2021. ISSN 01691368. doi: 10.1016/j.oregeorev.2021.104010.
- C. Galerne, M. Caroff, J. Rolet, and B. Le Gall. Magma – sediment mingling in an Ordovician rift basin : The Plouérec-Plourivo half-graben, Armorican Massif, France. *Journal of Volcanology and Geothermal Research*, 155:164–178, 2006.
- Y. Gao, Z. Zhang, Y. Xiong, and R. Zuo. Mapping mineral prospectivity for Cu polymetallic mineralization in southwest Fujian Province, China. *Ore Geology Reviews*, 75:16–28, June 2016. ISSN 01691368. doi: 10.1016/j.oregeorev.2015.12.005.
- D. Gapais and C. Le Corre. Is the Hercynien belt of Brittany a major shear zone? *Nature*, 288(5791): 574–576, Dec. 1980. ISSN 0028-0836, 1476-4687. doi: 10.1038/288574a0.
- D. Gapais, J.-L. Lagarde, C. Le Corre, C. Audren, P. Jegouzo, A. Casas Sainz, and J. Van Den Driessche. La zone de cisaillement de Quiberon : Témoin d’extension de la chaîne varisque en Bretagne méridionale au Carbonifère. *La zone de cisaillement de Quiberon : témoin d’extension de la chaîne varisque en Bretagne méridionale au Carbonifère*, 316(8):1123–1129, 1993. ISSN 0764-4450.
- D. Gapais, J.-P. Brun, C. Gumiaux, F. Cagnard, G. Ruffet, and C. Le Carlier De Veslud. Extensional tectonics in the Hercynian Armorican belt (France). An overview. *Bulletin de la Société Géologique de France*, 186(2-3):117–129, Mar. 2015. ISSN 0037-9409. doi: 10.2113/gssgfbull.186.2-3.117.
- P. Gentilhomme, Y. Deschamps, J. Mroz, and P. Vadala. Synthèse et bilan de l’Inventaire minier métropolitain, volet de réévaluation économique. Technical Report BRGM report RP-51422-FR, Orléans, 2001.
- R. Genuer and J.-M. Poggi. Arbres CART et Forêts aléatoires, Importance et sélection de variables. *HAL*, page 45, Jan. 2017. doi: hal-01387654v2f.
- G. K. Gilbert. Lake Bonneville. Report 1, U.S. Government Printing Office, Washington, D.C., 1890.
- E. Gloaguen and B. Tourlière. Revalorisation du potentiel minier français. Méthodes prédictives appliquées à la province à antimoine du Massif Armoricain. Technical Report RP-65534-FR, BRGM, Orléans, France, Nov. 2016a.
- E. Gloaguen and B. Tourlière. Revalorisation du potentiel minier français : Le district à antimoine-or du Semnon (Ille-et-Vilaine, France). Document Public BRGM/RP-66200-FR, BRGM, Orléans, France, 2016b.
- E. Gloaguen, B. Tourlière, and A. Audion-Serrand. Revalorisation du potentiel minier français: Méthodes prédictives appliquées à la province à étain du Massif Armoricain. Rapport final. Technical Report BRGM/RP-67620-FR, BRGM, Orléans, France, 2018.

- A. Gorichon and M. Kerjean. Reconnaissance par travaux miniers des indices d'antimoine de Quimper (PER de l'Odet). Résultats au 31 mai 1979. Technical Report 79-RDM-035-FE, BRGM, Orléans, France, 1979.
- J. Granek and E. Haber. Advanced Geoscience Targeting via Focused Machine Learning Applied to the QUEST Project Dataset, British Columbia. page 10, 2015.
- M. Greenacre and J. Blasius. *Multiple Correspondence Analysis and Related Methods*. Statistics in the Social and Behavioral Sciences. Chapman and Hall/CRC, June 2006. ISBN 978-1-58488-628-0.
- D. Groves, R. Goldfarb, M. Gebre-Mariam, S. Hagemann, and F. Robert. Orogenic gold deposits: A proposed classification in the context of their crustal distribution and relationship to other gold deposit types. *Ore Geology Reviews*, 13(1-5):7–27, Apr. 1998. ISSN 01691368. doi: 10.1016/S0169-1368(97)00012-7.
- D. I. Groves and F. P. Bierlein. Geodynamic settings of mineral deposit systems. *Journal of the Geological Society*, 164(1):19–30, Jan. 2007. ISSN 0016-7649, 2041-479X. doi: 10.1144/0016-76492006-065.
- C. Guerrot, J.-Y. Calvez, J. Bonjour, J. Chantraine, J.-J. Chauvel, L. Dupret, and D. Rabu. Le Briovérien de Bretagne centrale et occidentale : Nouvelles données radiométriques. *Comptes Rendus de l'Académie des Sciences de Paris II*, 315:1741–1746, 1992.
- P. Guiollard. *L'industrie minière de l'Antimoine et du Tungstène. Emergence, prospérité et disparition des exploitations de France métropolitaine aux XIXe et XXe siècles*. PhD thesis, Université de Haute-Alsace Mulhouse-Colmar, Biarritz, 2009.
- C. Gumiaux. *Modélisation du cisaillement hercynien de Bretagne centrale: déformation crustale et implications lithosphériques*. PhD thesis, Sciences de la Matière, Université de Rennes, Rennes, France, Apr. 2003.
- C. Gumiaux, J. P. Brun, and D. Gapais. Strain removal within the Hercynian Shear Belt of Central Brittany (western France): Methodology and tectonic implications. *Geological Society, London, Special Publications*, 224(1):287–305, 2004a. ISSN 0305-8719, 2041-4927. doi: 10.1144/GSL.SP.2004.224.01.18.
- C. Gumiaux, D. Gapais, J. Brun, J. Chantraine, and G. Ruffet. Tectonic history of the Hercynian Armorican Shear belt (Brittany, France). *Geodinamica Acta*, 17(4):289–307, Aug. 2004b. ISSN 0985-3111, 1778-3593. doi: 10.3166/ga.17.289-307.
- P. Gumiel. *Metalogenia de los Yacimientos de Antimonio de la Península Iberica*. PhD thesis, Universidad de Salamanca, Departamento de Cristalografía y Mineralogía, Salamanca, 1982.
- P. Gumiel and A. Arribas. Antimony deposits in the Iberian Peninsula. *Economic Geology*, 82(6): 1453–1463, Oct. 1987. ISSN 1554-0774, 0361-0128. doi: 10.2113/gsecongeo.82.6.1453.
- P. Gumiel, A. Arribas, and J. Saavedra. Geología y Metalogenia del Yacimiento de Estibina-Scheelita de "San Antonio" Albuquerque (Badajoz). *Studia Geologica*, X:61–93, 1976.
- X. Guo, Z. Wu, M. He, X. Meng, X. Jin, N. Qiu, and J. Zhang. Adsorption of antimony onto iron oxyhydroxides: Adsorption behavior and surface structure. *Journal of Hazardous Materials*, 276:339–345, July 2014. ISSN 0304-3894. doi: 10.1016/j.jhazmat.2014.05.025.

- G. Gutiérrez-Alonso, S. Johnston, A. Weil, D. Pastor-Galán, and J. Fernández-Suárez. Buckling an orogen: The Cantabrian Orocline. *GSA Today*, 22(7):4–9, July 2012. ISSN 10525173. doi: 10.1130/GSATG141A.1.
- G. Gutiérrez-Alonso, A. S. Collins, J. Fernández-Suárez, D. Pastor-Galán, E. González-Clavijo, F. Jourdan, A. B. Weil, and S. T. Johnston. Dating of lithospheric buckling: $^{40}\text{Ar}/^{39}\text{Ar}$ ages of syn-orocline strike-slip shear zones in northwestern Iberia. *Tectonophysics*, 643:44–54, 2015.
- D. Harris and G. Pan. Mineral Favorability Mapping: A Comparison of Artificial Neural Networks, Logistic Regression, and Discriminant Analysis. *Natural Resources Research*, 8(2):93–109, 1999.
- D. Harris, L. Zurcher, M. Stanley, J. Marlow, and G. Pan. A Comparative Analysis of Favorability Mappings by Weights of Evidence, Probabilistic Neural Networks, Discriminant Analysis, and Logistic Regression. *Natural Resources Research*, 12(4):241–255, Dec. 2003. ISSN 1520-7439. doi: 10.1023/B:NARR.0000007804.27450.e8.
- J. Harris, E. Grunsky, P. Behnia, and D. Corrigan. Data- and knowledge-driven mineral prospectivity maps for Canada’s North. *Ore Geology Reviews*, 71:788–803, Dec. 2015. ISSN 01691368. doi: 10.1016/j.oregeorev.2015.01.004.
- J. R. Harris, L. Wilkinson, K. Heather, S. Fumerton, M. A. Bernier, J. Ayer, and R. Dahn. Application of GIS Processing Techniques for Producing Mineral Prospectivity Maps—A Case Study: Mesothermal Au in the Swayze Greenstone Belt, Ontario, Canada. *Natural Resources Research*, 10(2):91–124, 2001.
- É. Hartai, N. Németh, and J. Földessy. EGS-relevant review of metallogenesis. Technical Report CHPM2030 Deliverable D1.1, University of Miskolc, Hungary, Dec. 2016.
- T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York, New York, NY, 2009. ISBN 978-0-387-84857-0 978-0-387-84858-7. doi: 10.1007/978-0-387-84858-7.
- J. He, Y. Wei, T. Zhai, and H. Li. Antimony-based materials as promising anodes for rechargeable lithium-ion and sodium-ion batteries. *Materials Chemistry Frontiers*, 2(3):437–455, Mar. 2018. ISSN 2052-1537. doi: 10.1039/C7QM00480J.
- J. W. Hedenquist and J. B. Lowenstern. The role of magmas in the formation of hydrothermal ore deposits. *Nature*, 370(6490):519–527, Aug. 1994. ISSN 0028-0836, 1476-4687. doi: 10.1038/370519a0.
- C. A. Heinrich. Fluid-Fluid Interactions in Magmatic-Hydrothermal Ore Formation. *Reviews in Mineralogy and Geochemistry*, 65(1):363–387, July 2007. ISSN 1529-6466. doi: 10.2138/rmg.2007.65.11.
- M. L. C. M. Henckens, P. P. J. Driessen, and E. Worrell. How can we adapt to geological scarcity of antimony? Investigation of antimony’s substitutability and of other measures to achieve a sustainable use. *Resources, Conservation and Recycling*, 108:54–62, Mar. 2016. ISSN 0921-3449. doi: 10.1016/j.resconrec.2016.01.012.
- R. Holdsworth, M. Stewart, J. Imber, and R. Strachan. The structure and rheological evolution of reactivated continental fault zones: A review and case study. In *Continental Reactivation and Reworking*, volume 194 of *Spec. Publ.*, pages 115–137. Geol. Soc. Lond., miller, j.a., holdsworth, r.e., buick, i.s., hand, m. edition, 2001.

- D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant. *Applied Logistic Regression*. Wiley Series in Probability and Statistics. Wiley, first edition, Mar. 2013. ISBN 978-0-470-58247-3 978-1-118-54838-7. doi: 10.1002/9781118548387.
- J. M. Hronsky and O. P. Kreuzer. Applying spatial prospectivity mapping to exploration targeting: Fundamental practical issues and suggested solutions for the future. *Ore Geology Reviews*, 107:647–653, Apr. 2019. ISSN 01691368. doi: 10.1016/j.oregeorev.2019.03.016.
- J. M. A. Hronsky and D. I. Groves. Science of targeting: Definition, strategies, targeting and performance measurement. *Australian Journal of Earth Sciences*, 55(1):3–12, Feb. 2008. ISSN 0812-0099. doi: 10.1080/08120090701581356.
- C. Huber, O. Bachmann, J.-L. Vigneresse, J. Dufek, and A. Parmigiani. A physical model for metal extraction and transport in shallow magmatic systems: METALS EXTRACTION AND TRANSPORT. *Geochemistry, Geophysics, Geosystems*, 13(8):n/a–n/a, Aug. 2012. ISSN 15252027. doi: 10.1029/2012GC004042.
- F. Husson, J. Josse, and J. Pages. Principal component methods - hierarchical clustering - partitional clustering: Why would we need to choose for visualizing data? Technical Report, Agrocampus ouest, Rennes, Sept. 2010.
- M. A. Jadamec, D. L. Turcotte, and P. Howell. Analytic models for orogenic collapse. *Tectonophysics*, 435(1-4):1–12, May 2007. ISSN 00401951. doi: 10.1016/j.tecto.2007.01.007.
- X. Jiang, S. Liang, X. He, A. D. Ziegler, P. Lin, M. Pan, D. Wang, J. Zou, D. Hao, G. Mao, Y. Zeng, J. Yin, L. Feng, C. Miao, E. F. Wood, and Z. Zeng. Rapid and large-scale mapping of flood inundation via integrating spaceborne synthetic aperture radar imagery with unsupervised deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 178:36–50, Aug. 2021. ISSN 09242716. doi: 10.1016/j.isprsjprs.2021.05.019.
- I. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag, New York, 2002. ISBN 978-0-387-95442-4.
- T. Juteau, P. Nonnotte, I. Jegou, M. Lamour, T. Naour, and J. Cotten. Le volcanisme sous-marin d’âge ordovicien supérieur de la presqu’île de Crozon (Finistère). Etude des processus de bréchification. *Bulletin de la Société Géologique et Minéralogique de Bretagne*, 4:1–67, 2007.
- A. Karbalaee Ramezanali, F. Feizi, A. Jafarirad, and M. Lotfi. Application of Best-Worst method and Additive Ratio Assessment in mineral prospectivity mapping: A case study of vein-type copper mineralization in the Kuhsiah-e-Urmak Area, Iran. *Ore Geology Reviews*, 117:103268, Feb. 2020. ISSN 01691368. doi: 10.1016/j.oregeorev.2019.103268.
- R. Kerrich. Fluid infiltration into fault zones: Chemical, isotopic, and mechanical effects. *Pure and Applied Geophysics PAGEOPH*, 124(1-2):225–268, 1986. ISSN 0033-4553, 1420-9136. doi: 10.1007/BF00875727.
- M. Keykhay-Hosseinpour, A.-H. Kohsary, A. Hossein-Morshedy, and A. Porwal. A machine learning-based approach to exploration targeting of porphyry Cu-Au deposits in the Dehsalm district, eastern Iran. *Ore Geology Reviews*, 116:103234, Jan. 2020. ISSN 01691368. doi: 10.1016/j.oregeorev.2019.103234.

- J. Klus, P. Pořízka, D. Prochazka, P. Mikysek, J. Novotný, K. Novotný, M. Slobodník, and J. Kaiser. Application of self-organizing maps to the study of U-Zr-Ti-Nb distribution in sandstone-hosted uranium ores. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 131:66–73, May 2017. ISSN 05848547. doi: 10.1016/j.sab.2017.03.008.
- C. M. Knox-Robinson and D. I. Groves. Gold prospectivity mapping using a Geographic Information System (GIS) with examples from the Ylgarn Block of Western Australia. *Chronique de la Recherche Minière*, (529), 1997.
- M. Köhler, D. Hanelli, S. Schaefer, A. Barth, A. Knobloch, P. Hielscher, J. Cardoso-Fernandes, A. Lima, and A. C. Teodoro. Lithium Potential Mapping Using Artificial Neural Networks: A Case Study from Central Portugal. *Minerals*, 11(10):1046, Sept. 2021. ISSN 2075-163X. doi: 10.3390/min11101046.
- T. Kohonen. Essentials of the self-organizing map. *Neural Networks*, 37:52–65, Jan. 2013. ISSN 08936080. doi: 10.1016/j.neunet.2012.09.018.
- M. Kuhn and K. Johnson. *Applied Predictive Modeling*. Springer New York, New York, NY, 2013. ISBN 978-1-4614-6848-6 978-1-4614-6849-3. doi: 10.1007/978-1-4614-6849-3.
- B. Lamy-Chappuis, C. A. Heinrich, T. Driesner, and P. Weis. Mechanisms and patterns of magmatic fluid transport in cooling hydrous intrusions. *Earth and Planetary Science Letters*, 535:116111, Apr. 2020. ISSN 0012821X. doi: 10.1016/j.epsl.2020.116111.
- G. N. Lance and W. T. Williams. A General Theory of Classificatory Sorting Strategies: 1. Hierarchical Systems. *The Computer Journal*, 9(4):373–380, Feb. 1967. ISSN 0010-4620, 1460-2067. doi: 10.1093/comjnl/9.4.373.
- G. Launay. *Hydrodynamiques Des Systèmes Minéralisés Péri-Granitiques : Étude Du Gisement à W-Sn-(Cu) de Panasqueira (Portugal)*. PhD thesis, Université d’Orléans, 2018.
- G. Launay, S. Sizaret, P. Lach, J. Melleton, É. Gloaguen, and M. Poujol. Genetic relationship between greisenization and Sn-W mineralizations in vein and greisen deposits: Insights from the Panasqueira deposit (Portugal). *Bulletin de la Société Géologique de France*, 192(1):2, Feb. 2021. ISSN 0037-9409. doi: 10.1051/bsgf/2020046.
- C. Le Corre. *Approche quantitative des processus synschisteux. L'exemple du segment Hercynien de Bretagne Centrale*. Thèse d’état, Université de Rennes 1, Rennes, France, 1978.
- S. Le Dien and J. Pagès. Analyse factorielle multiple hiérarchique. *Revue de statistique appliquée*, 2(51): 47–73, 2003.
- Y. Le Fur, A. Allon, R. Biron, M. Lequertier, and M. Roussel. La découverte du gisement d’antimoine des Brouzils en Vendée (Massif Armoricaïn, France). Historique des travaux, description du gisement et projet d’exploitation. *Chronique de la Recherche Minière*, 492:5–18, 1988.
- J. Le Gall. Pétrogenèse des magmas andésitiques et ignimbritiques et leur signification dans l’évolution géodynamique cadomienne. Reconstitution des dynamismes éruptifs d’une province paléovolcanique : L’exemple du graben cambrien du Maine (Est du Massif Armoricaïn). *Mémoires de Géosciences Rennes*, 52:370 p., 1993.

- F. Le Hebel, S. Fourcade, D. Gapais, C. Marignac, R. Capdevila, and F. Martineau. Fluid-assisted spreading of thickened continental crust: Preliminary data from the Variscan belt of South Brittany (France). *Journal of Geochemical Exploration*, 69–70:561–564, June 2000. ISSN 03756742. doi: 10.1016/S0375-6742(00)00121-7.
- F. Le Hebel, D. Gapais, S. Fourcade, and R. Capdevila. Fluid-assisted large strains in a crustal-scale décollement (Hercynian Belt of South Brittany, France). *Geological Society, London, Special Publications*, 200(1):85–101, 2002. ISSN 0305-8719, 2041-4927. doi: 10.1144/GSL.SP.2001.200.01.06.
- P. Lecumberri-Sanchez, R. Vieira, C. Heinrich, F. Pinto, and M. Wille. Fluid-rock interaction is decisive for the formation of tungsten deposits. *Geology*, 45(7):579–582, July 2017. ISSN 0091-7613, 1943-2682. doi: 10.1130/G38974.1.
- J. Lefort. Iberian–Armorican arc and Hercynian orogeny in Western Europe. *Geology*, 7:384–388, 1979.
- B. Legaz, M. Unguru, L. Mancini, C. Latunussa, T. Hamor, F. Ardente, F. Mathieux, V. Nita, C. Torres de Matos, B. Plazzotta, B. Bonollo, G. Blengini, F. Pasimeni, D. Wittmer, U. Eynard, E. Garbossa, D. Pennington, and C. Ciupagea. 3rd Raw Materials Scoreboard. Technical Report ISBN 978-92-76-23795-2, European Innovation Partnership on Raw Materials, Luxembourg: Publications Office of the European Union, 2021.
- E. P. Leite and C. R. de Souza Filho. Self organizing maps applied to mineral potential mapping in the serra leste region, Carajás mineral province, state of Pará, Brazil. *Revista Brasileira de Geofísica*, 28(3):397–409, Sept. 2010. ISSN 0102-261X. doi: 10.1590/S0102-261X2010000300006.
- J. L. Lescuyer, J. M. Leistel, E. Marcoux, J. P. Milési, and D. Thiéblemont. Late Devonian–Early Carboniferous peak sulphide mineralization in the Western Hercynides. *Mineralium Deposita*, 33(1-2): 208–220, Dec. 1997. ISSN 0026-4598, 1432-1866. doi: 10.1007/s001260050141.
- J. Lever, M. Krzywinski, and N. Altman. Model selection and overfitting. *Nature Methods*, 13(9):703–704, Sept. 2016. ISSN 1548-7105. doi: 10.1038/nmeth.3968.
- X. Li, F. Yuan, M. Zhang, C. Jia, S. M. Jowitt, A. Ord, T. Zheng, X. Hu, and Y. Li. Three-dimensional mineral prospectivity modeling for targeting of concealed mineralization within the Zhonggu iron ore-field, Ningwu Basin, China. *Ore Geology Reviews*, 71:633–654, Dec. 2015. ISSN 01691368. doi: 10.1016/j.oregeorev.2015.06.001.
- X. Li, F. Yuan, M. Zhang, S. M. Jowitt, A. Ord, T. Zhou, and W. Dai. 3D computational simulation-based mineral prospectivity modeling for exploration for concealed Fe–Cu skarn-type mineralization within the Yueshan orefield, Anqing district, Anhui Province, China. *Ore Geology Reviews*, 105:1–17, Feb. 2019a. ISSN 01691368. doi: 10.1016/j.oregeorev.2018.12.003.
- X.-P. Li, H.-P. Schertl, and J. Reinhardt. Preface: Metamorphism and Orogenic Belts—Response from Micro- to Macro-Scale. *Journal of Earth Science*, 30(6):1075–1083, Dec. 2019b. ISSN 1674-487X, 1867-111X. doi: 10.1007/s12583-019-1269-y.
- L. Liu, J. Lu, C. Tao, S. Liao, C. Su, N. Huang, and X. Xu. Fuzzy Forest Machine Learning Predictive Model for Mineral Prospectivity: A Case Study on Southwest Indian Ridge 48.7°E–50.5°E. *Natural Resources Research*, 31(1):99–116, Feb. 2022a. ISSN 1520-7439, 1573-8981. doi: 10.1007/s11053-021-09996-1.

- Y. Liu, Q. Cheng, Q. Xia, and X. Wang. Mineral potential mapping for tungsten polymetallic deposits in the Nanling metallogenic belt, South China. *Journal of Earth Science*, 25(4):689–700, Aug. 2014. ISSN 1674-487X, 1867-111X. doi: 10.1007/s12583-014-0466-y.
- Y. Liu, E. J. M. Carranza, and Q. Xia. Developments in Quantitative Assessment and Modeling of Mineral Resource Potential: An Overview. *Natural Resources Research*, 31(4):1825–1840, Aug. 2022b. ISSN 1520-7439, 1573-8981. doi: 10.1007/s11053-022-10075-2.
- D. London. Ore-forming processes within granitic pegmatites. *Ore Geology Reviews*, 101:349–383, Oct. 2018. ISSN 01691368. doi: 10.1016/j.oregeorev.2018.04.020.
- F. J. López-Moro, M. López-Plaza, G. Gutiérrez-Alonso, J. Fernández-Suárez, A. López-Carmona, M. Hofmann, and R. L. Romer. Crustal melting and recycling: Geochronology and sources of Variscan syn-kinematic anatectic granitoids of the Tormes Dome (Central Iberian Zone). A U–Pb LA-ICP-MS study. *International Journal of Earth Sciences*, 107(3):985–1004, Apr. 2018. ISSN 1437-3254, 1437-3262. doi: 10.1007/s00531-017-1483-8.
- W. Luo, J.-J. Gaumet, and L.-Q. Mai. Antimony-based intermetallic compounds for lithium-ion and sodium-ion batteries: Synthesis, construction and application. *Rare Metals*, 36(5):321–338, May 2017. ISSN 1867-7185. doi: 10.1007/s12598-017-0899-4.
- Y. Ma, J. Zhao, Y. Sui, S. Liao, and Z. Zhang. Application of Knowledge-Driven Methods for Mineral Prospectivity Mapping of Polymetallic Sulfide Deposits in the Southwest Indian Ridge between 46° and 52°E. *Minerals*, 10(11):970, Oct. 2020. ISSN 2075-163X. doi: 10.3390/min10110970.
- A. A. Madani. Knowledge-driven GIS modeling technique for gold exploration, Bulghah gold mine area, Saudi Arabia. *The Egyptian Journal of Remote Sensing and Space Science*, 14(2):91–97, Dec. 2011. ISSN 11109823. doi: 10.1016/j.ejrs.2011.10.001.
- F. Maepa, R. S. Smith, and A. Tessema. Support vector machine and artificial neural network modelling of orogenic gold prospectivity mapping in the Swayze greenstone belt, Ontario, Canada. *Ore Geology Reviews*, 130:103968, Mar. 2021. ISSN 01691368. doi: 10.1016/j.oregeorev.2020.103968.
- X. Mao, J. Ren, Z. Liu, J. Chen, L. Tang, H. Deng, R. C. Bayless, B. Yang, M. Wang, and C. Liu. Three-dimensional prospectivity modeling of the Jiaojia-type gold deposit, Jiaodong Peninsula, Eastern China: A case study of the Dayingezhuang deposit. *Journal of Geochemical Exploration*, 203:27–44, Aug. 2019. ISSN 03756742. doi: 10.1016/j.gexplo.2019.04.002.
- E. Marcoux and L. Chauris. Metallogeny of the Armorican Massif. In *Pre-Mesozoic Geology in France and Related Areas*, pages 243–264. Springer-Verlag Berlin Heidelberg, 1994. doi: 10.1007/978-3-642-84915-2_23.
- E. Marcoux and M. Jébrak. *Géologie des ressources minérales*. Ministère des ressources naturelles et de la faune, géologie quebec edition, 2008. ISBN 978-2-551-23737-1.
- E. Marcoux, R. Serment, and A. Allon. Les gites d’antimoine de Vendée (Massif Armoricaïn, France). *Chronique de la Recherche Minière*, 476:3–30, 1984.
- F. O. Marques, A. Mateus, and C. Tassinari. The Late-Variscan fault network in central–northern Portugal (NW Iberia): A re-evaluation. *Tectonophysics*, 359(3):255–270, Dec. 2002. ISSN 0040-1951. doi: 10.1016/S0040-1951(02)00514-0.

- J. R. Martínez Catalán. The Central Iberian arc, an orocline centered in the Iberian Massif and some implications for the Variscan belt. *International Journal of Earth Sciences*, 101(5):1299–1314, July 2012. ISSN 1437-3254, 1437-3262. doi: 10.1007/s00531-011-0715-6.
- J. R. Martínez Catalán, K. Schulmann, and J.-F. Ghienne. The Mid-Variscan Allochthon: Keys from correlation, partial retrodeformation and plate-tectonic reconstruction to unlock the geometry of a non-cylindrical belt. *Earth-Science Reviews*, 220:103700, Sept. 2021. ISSN 0012-8252. doi: 10.1016/j.earscirev.2021.103700.
- L. P. Martins. Mineral Resources of Portugal 2012. Technical report, Direcção Geral de Energia e Geologia, Lisboa, 2012.
- A. Mateus and F. Noronha. Epigenetic ore-forming systems in Central-Iberian Zone; products of an evolving orogenic framework in meso- to late-variscan times. *Ciências Geológicas: Ensino, Investigação e sua História*, 2:47–61, 2010.
- P. Matte. Accretionary history and crustal evolution of the Variscan belt in Western Europe. *Tectonophysics*, 196:309–337, 1991.
- P. Matte. The Variscan collage and orogeny (480-290 Ma) and the tectonic definition of the Armorica microplate: A review. *Terra Nova*, 13(2):122–128, Dec. 2001. ISSN 0954-4879, 1365-3121. doi: 10.1046/j.1365-3121.2001.00327.x.
- P. Matte. Variscan thrust nappes, detachments, and strike-slip faults in the French Massif Central: Interpretation of the lineations. In *Geological Society of America Memoirs*, volume 200, pages 391–402. Geological Society of America, Jan. 2007. ISBN 978-0-8137-1200-0. doi: 10.1130/2007.1200(20).
- R. B. McCammon, J. M. Botbol, R. Sinding-Larsen, and R. W. Bowen. Characteristic analysis—1981: Final program and a possible discovery. *Journal of the International Association for Mathematical Geology*, 15(1):59–83, Feb. 1983. ISSN 1573-8868. doi: 10.1007/BF01030076.
- R. B. McCammon, J. M. Botbol, R. Sinding-Larson, and R. W. Bowen. The new characteristic analysis (NCHARAN) program. U.S. Geological Survey Bulletin 1621, U.S. Government Publishing Office, 1984.
- T. C. McCuaig, S. Beresford, and J. Hronsky. Translating the mineral systems approach into an effective exploration targeting system. *Ore Geology Reviews*, 38(3):128–138, Nov. 2010. ISSN 0169-1368. doi: 10.1016/j.oregeorev.2010.05.008.
- W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, Dec. 1943. ISSN 1522-9602. doi: 10.1007/BF02478259.
- G. McKay and J. R. Harris. Comparison of the Data-Driven Random Forests Model and a Knowledge-Driven Method for Mineral Prospectivity Mapping: A Case Study for Gold Deposits Around the Huritz Group and Nueltin Suite, Nunavut, Canada. *Natural Resources Research*, 25(2):125–143, June 2016. ISSN 1520-7439, 1573-8981. doi: 10.1007/s11053-015-9274-z.
- M. McMillan, J. Fohring, E. Haber, and J. Granek. Orogenic gold prospectivity mapping using machine learning. *ASEG Extended Abstracts*, 2019(1):1–4, Dec. 2019. ISSN 2202-0586. doi: 10.1080/22020586.2019.12073020.

- R. Montorio, F. Pérez-Cabello, D. Borini Alves, and A. García-Martín. Unitemporal approach to fire severity mapping using multispectral synthetic databases and Random Forests. *Remote Sensing of Environment*, 249:112025, Nov. 2020. ISSN 00344257. doi: 10.1016/j.rse.2020.112025.
- W. M. Moon and P. An. Integration of Geophysical, Geological and Remote Sensing Data Using Fuzzy Set Theory. *Geoinformatics*, 2(2):171–176, 1991. ISSN 0388-502X, 1347-541X. doi: 10.6010/geoinformatics1990.2.2_171.
- B. Moyroud, J. Féraud, M. Bornuat, and A. Ziserman. Ressources minières françaises. Tome 6, les gisements d’Antimoine (Situation en 1979). Technical Report BRGM report RR-41430-FR-06, 1979.
- M. Munoz, P. Courjault-Radé, and F. Tollon. The massive stibnite veins of the French Palaeozoic basement: A metallogenic marker of Late Variscan brittle extension. *Terra Nova*, 4(2):171–177, Mar. 1992. ISSN 0954-4879, 1365-3121. doi: 10.1111/j.1365-3121.1992.tb00468.x.
- F. Murtagh, P. Legendre, and W. Place. Ward’s Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. *Journal of Classification*, 31(3):274–295, Dec. 2011.
- S. Nawar and A. Mouazen. Comparison between Random Forests, Artificial Neural Networks and Gradient Boosted Machines Methods of On-Line Vis-NIR Spectroscopy Measurements of Soil Total Nitrogen and Total Carbon. *Sensors*, 17(10):2428, Oct. 2017. ISSN 1424-8220. doi: 10.3390/s17102428.
- A. Neiva, P. Andráš, and J. Ramos. Antimony quartz and antimony–gold quartz veins from northern Portugal. *Ore Geology Reviews*, 34(4):533–546, Dec. 2008. ISSN 01691368. doi: 10.1016/j.oregeorev.2008.03.004.
- S. H. Nielsen, F. Cunningham, R. Hay, G. Partington, and M. Stokes. 3D prospectivity modelling of orogenic gold in the Marymia Inlier, Western Australia. *Ore Geology Reviews*, 71:578–591, Dec. 2015. ISSN 01691368. doi: 10.1016/j.oregeorev.2015.02.001.
- C. Noblet. *La sédimentation arénacée de l’Ordovicien inférieur au Nord du Gondwana : la Formation du Grès Armoricaïn et ses équivalents*. Thèse Doctorat 3ème cycle, Univ. Rennes, Rennes, France, 1984.
- V. Nykänen, I. Lahti, T. Niiranen, and K. Korhonen. Receiver operating characteristics (ROC) as validation tool for prospectivity models — A magmatic Ni–Cu case study from the Central Lapland Greenstone Belt, Northern Finland. *Ore Geology Reviews*, 71:853–860, Dec. 2015. ISSN 01691368. doi: 10.1016/j.oregeorev.2014.09.007.
- H.-J. Oh and S. Lee. Application of Artificial Neural Network for Gold–Silver Deposits Potential Mapping: A Case Study of Korea. *Natural Resources Research*, 19(2):103–124, June 2010. ISSN 1520-7439, 1573-8981. doi: 10.1007/s11053-010-9112-2.
- J. Orcel. La métallogénie. *Bulletin de Minéralogie*, 77(1):573–597, 1954. doi: 10.3406/bulmi.1954.4910.
- L. Ortega and E. Vindel. Evolution of ore-forming fluids associated with late Hercynian antimony deposits in Central/Western Spain: Case study of Mari Rosa and El Juncalón. *European Journal of Mineralogy*, 7(3):655–674, May 1995. ISSN 0935-1221. doi: 10.1127/ejm/7/3/0655.
- L. Ortega, R. Oyarúun, and M. Gallego. The Mari Rosa late Hercynian Sb–Au deposit, western Spain: Geology and geochemistry of the mineralizing processes. *Mineralium Deposita*, 31(3):172–187, Mar. 1996. ISSN 0026-4598, 1432-1866. doi: 10.1007/BF00204025.

- J. Pagès. Analyse factorielle de données mixtes. *Revue de statistique appliquée*, 52(4):93–111, 2004.
- F. J. Palero-Fernández, A. Martin-Izard, M. Zarzalejos Prieto, and L. Mansilla-Plaza. Geological context and plumbotectonic evolution of the giant Almadén Mercury Deposit. *Ore Geology Reviews*, 64:71–88, Jan. 2015. ISSN 0169-1368. doi: 10.1016/j.oregeorev.2014.06.013.
- J.-L. Paquette, M. Ballèvre, J.-J. Peucat, and G. Cornen. From opening to subduction of an oceanic domain constrained by LA-ICP-MS U-Pb zircon dating (Variscan belt, Southern Armorican Massif, France). *Lithos*, 294–295:418–437, Dec. 2017. ISSN 00244937. doi: 10.1016/j.lithos.2017.10.005.
- M. Parsa. A data augmentation approach to XGboost-based mineral potential mapping: An example of carbonate-hosted Zn Pb mineral systems of Western Iran. *Journal of Geochemical Exploration*, 228: 106811, Sept. 2021. ISSN 03756742. doi: 10.1016/j.gexplo.2021.106811.
- M. Parsa, E. J. M. Carranza, and B. Ahmadi. Deep GMDH Neural Networks for Predictive Mapping of Mineral Prospectivity in Terrains Hosting Few but Large Mineral Deposits. *Natural Resources Research*, 31(1):37–50, Feb. 2022. ISSN 1520-7439, 1573-8981. doi: 10.1007/s11053-021-09984-5.
- G. Partington. Developing models using GIS to assess geological and economic risk: An example from VMS copper gold mineral exploration in Oman. *Ore Geology Reviews*, 38:197–207, 2010. doi: 10.1016/j.oregeorev.2010.02.002.
- D. Pastor-Galán, T. Groenewegen, D. Brouwer, W. Krijgsman, and M. Dekkers. One or two oroclinal in the Variscan orogen of Iberia? Implications for Pangea amalgamation. *Geology*, 43:527–530, 2015.
- P. Patinha, A. Pereira, L. Quental, A. Filipe, D. P. S. de Oliveira, and C. Lopes. SIORMINP : Plataforma de gestão e disponibilização de dados harmonizados dos Recursos Minerais Portugueses. *Comunicações Geológicas*, 107(Especial I):61–63, Aug. 2020. ISSN 0873-948X. doi: 10.34637/hmd5-7w43.
- E. Pereira, M. Iglésias, and A. Ribeiro, M. A. Leucogranitos-“stockscheider” e o controlo estrutural da mineralização na mina de Montesinho-Bragança. *Comunicações dos Serviços Geológicos de Portugal*, 70:11–22, 1984.
- E. Pereira, A. Ribeiro, and C. Meireles. Variscan Shear zones and control of Sn-W, Au, U mineralizations in the Central-Iberian Zone in Portugal. *Coruña*, 18:89–119, 1993.
- A. Pérez-Estaún, F. Bastida, J. Alonso, J. Marquínez, J. Aller, J. Álvarez-Marrón, A. Marcos, and J. Pulgar. A thin-skinned tectonic model for an arcuate fold and thrust belt: The Cantabrian Zone (Variscan Ibero-Armorican Arc). *Tectonics*, 7:517–537, 1988. doi: 10.1029/tc007i003p00517.
- A. Pérez-Estaún, J. Martínez Catalán, and F. Bastida. Crustal thickening and deformation sequence in the footwall to the suture of the Variscan belt of northwest Spain. *Tectonophysics*, 191:243–253, 1991.
- H. Perroud and N. Bonhommet. Palaeomagnetism of the Ibero-Armorican arc and the Hercynian orogeny in Western Europe. *Nature*, 292:445–448, 1981.
- J.-J. Peucat, B. Auvray, Y. Hirbec, and J.-Y. Calvez. Granites et cisaillements hercyniens dans le Nord du Massif Armoricaín; géochronologie Rb-Sr. *Bulletin de la Société Géologique de France*, S7-XXVI (6):1365–1373, Jan. 1984. ISSN 1777-5817, 0037-9409. doi: 10.2113/gssgfbull.S7-XXVI.6.1365.
- A. Pochon. *Magmatisme Mafique et Minéralisations à Sb-Au Dans Le Domaine Centre Armoricaín: Contrôles Spatio-Temporels et Implications Métallogéniques*. PhD thesis, Université de Rennes, 2017.

- A. Pochon, D. Gapais, E. Gloaguen, C. Gumiaux, Y. Branquet, F. Cagnard, and G. Martelet. Antimony deposits in the Variscan Armorican belt, a link with mafic intrusives? *Terra Nova*, 28(2):138–145, Apr. 2016. ISSN 09544879. doi: 10.1111/ter.12201.
- A. Pochon, E. Gloaguen, Y. Branquet, M. Poujol, G. Ruffet, M.-C. Boiron, P. Boulvais, C. Gumiaux, F. Cagnard, F. Gouazou, and D. Gapais. Variscan Sb-Au mineralization in Central Brittany (France): A new metallogenic model derived from the Le Semnon district. *Ore Geology Reviews*, 97:109–142, June 2018. ISSN 01691368. doi: 10.1016/j.oregeorev.2018.04.016.
- A. Pochon, Y. Branquet, E. Gloaguen, G. Ruffet, M. Poujol, P. Boulvais, C. Gumiaux, F. Cagnard, J.-M. Baele, I. Kéré, and D. Gapais. A Sb $\pm$ Au mineralizing peak at 360 Ma in the Variscan belt. *BSGF - Earth Sciences Bulletin*, 190:4, 2019. ISSN 0037-9409, 1777-5817. doi: 10.1051/bsgf/2019004.
- M. I. Ponce de Leon and P. Choukroune. Shear zones in the iberian arc. *Journal of Structural Geology*, 2(1):63–68, Jan. 1980. ISSN 0191-8141. doi: 10.1016/0191-8141(80)90035-8.
- A. Porwal, E. J. M. Carranza, and M. Hale. Knowledge-Driven and Data-Driven Fuzzy Models for Predictive Mineral Potential Mapping. *Natural Resources Research*, 12(1):1–25, Mar. 2003a. ISSN 1573-8981. doi: 10.1023/A:1022693220894.
- A. Porwal, E. J. M. Carranza, and M. Hale. Artificial Neural Networks for Mineral-Potential Mapping: A Case Study from Aravalli Province, Western India. *Natural Resources Research*, 12(3):155–171, Sept. 2003b. ISSN 1573-8981. doi: 10.1023/A:1025171803637.
- A. Porwal, E. Carranza, and M. Hale. Bayesian network classifiers for mineral potential mapping. *Computers & Geosciences*, 32(1):1–16, Feb. 2006. ISSN 00983004. doi: 10.1016/j.cageo.2005.03.018.
- A. Porwal, I. González-Álvarez, V. Markwitz, T. McCuaig, and A. Mamuse. Weights-of-evidence and logistic regression modeling of magmatic nickel sulfide prospectivity in the Yilgarn Craton, Western Australia. *Ore Geology Reviews*, 38(3):184–196, Nov. 2010. ISSN 01691368. doi: 10.1016/j.oregeorev.2010.04.002.
- A. Porwal, R. D. Das, B. Chaudhary, I. Gonzalez-Alvarez, and O. Kreuzer. Fuzzy inference systems for prospectivity modeling of mineral systems and a case-study for prospectivity mapping of surficial Uranium in Yeelirrie Area, Western Australia. *Ore Geology Reviews*, 71:839–852, Dec. 2015. ISSN 01691368. doi: 10.1016/j.oregeorev.2014.10.016.
- E. M. G. Prado, C. R. de Souza Filho, E. J. M. Carranza, and J. G. Motta. Modeling of Cu-Au prospectivity in the Carajás mineral province (Brazil) through machine learning: Dealing with imbalanced training data. *Ore Geology Reviews*, 124:103611, Sept. 2020. ISSN 01691368. doi: 10.1016/j.oregeorev.2020.103611.
- C. Quesada and J. T. Oliveira, editors. *The Geology of Iberia: A Geodynamic Approach: Volume 2: The Variscan Cycle*. Regional Geology Reviews. Springer International Publishing, Cham, 2019. ISBN 978-3-030-10518-1 978-3-030-10519-8. doi: 10.1007/978-3-030-10519-8.
- T. Raimondo, A. S. Collins, M. Hand, A. Walker-Hallam, R. H. Smithies, P. M. Evins, and H. M. Howard. The anatomy of a deep intracontinental orogen: ANATOMY OF DEEP INTRACONTINENTAL OROGEN. *Tectonics*, 29(4):n/a–n/a, Aug. 2010. ISSN 02787407. doi: 10.1029/2009TC002504.

- M. Rajabi, A. Mansourian, P. Pilesjö, F. Hedefalk, R. Groth, and A. Bazmani. Comparing Knowledge-Driven and Data-Driven Modeling methods for susceptibility mapping in spatial epidemiology : A case study in Visceral Leishmaniasis. In *17th AGILE International Conference on Geographic Information Science, 2014*, pages 1–5. Association of Geographic Information Laboratories for Europe, 2014.
- M. Reuter and A. van Schaik. Resource Efficient Metal and Material Recycling. In *REWAS 2013 : Enabling Materials Resource Sustainability*, The Minerals, Metals & Materials Series, pages 332–340. Springer, Cham, kvithyld a. et al. (eds) edition, 2013. ISBN 978-3-319-48763-2.
- P. Rey, O. Vanderhaeghe, and C. Teyssier. Gravitational collapse of the continental crust: Definition, regimes and modes. *Tectonophysics*, 342(3-4):435–449, Dec. 2001. ISSN 00401951. doi: 10.1016/S0040-1951(01)00174-3.
- C. Rham. La classification hiérarchique ascendante selon la méthode des voisins réciproques. *Les cahiers de l'analyse des données*, 5(2):135–144, 1980.
- S. Riahi, A. Bahroudi, M. Abedi, S. Aslani, and G.-R. Elyasi. Integration of airborne geophysics and satellite imagery data for exploration targeting in porphyry Cu systems: Chahargonbad district, Iran. *Geophysical Prospecting*, 69(5):1116–1137, 2021. ISSN 1365-2478. doi: 10.1111/1365-2478.13092.
- A. Ribeiro. *Contribution à l'étude Tectonique de Trás-os-Montes Oriental*. Number 24 in Memória / Serviços Geológicos de Portugal. Nova Série. Servicos geológicos de portuga edition, 1974.
- A. Ribeiro. Evolução geodinâmica de Portugal: os ciclos ante-mesozóicos. In *Geologia de Portugal*, volume 1, page 807. Escolar Editora., Lisboa, 2013. ISBN 978-972-592-364-1.
- A. Ribeiro, M. T. Antunes, R. B. Rocha, A. F. Soares, G. Zbyzewski, F. M. Almeida, D. Carvalho, and J. H. Monteiro. *Introduction à La Géologie Générale Du Portugal*. Lisboa, servicos geológicos de portuga edition, 1979.
- A. Ribeiro, J. Munhá, R. Dias, A. Mateus, E. Pereira, L. Ribeiro, P. Fonseca, A. Araújo, T. Oliveira, J. Romão, H. Chaminé, C. Coke, and J. Pedro. Geodynamic evolution of the SW Europe Variscides. *Tectonics*, 26(6):n/a–n/a, Dec. 2007. ISSN 02787407. doi: 10.1029/2006TC002058.
- A. Ribeiro, R. Dias, A. Mateus, J. Relvas, A. Pinto, and J. Romão. Geodynamics of Iberia, supercontinent cycles and metallogenic implications. *Comunicações Geológicas*, 101(Especial I):293–296, 2014. ISSN 0873-948X.
- M. Robardet, J. Bonjour, F. Paris, and P. Racheboeuf. Ordovician, Silurian, and Devonian of the Medio-North-Armorican Domain. In *Pre-Mesozoic Geology in France and Related Areas*, IGCP-Project 233 (Terranes In The Circum-Atlantic Paleozoic Orogens). Springer, Berlin, Heidelberg, chantraine j., rolet j., santallier d.s., piqué a., keppie j.d. edition, 1994. ISBN 978-3-642-84917-6.
- L. Rodríguez Fernández and T. Oliveira. Mapa Geológico de la Península Ibérica, Baleares y Canarias a escala 1:1.000.000, 2015.
- V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71:804–818, Dec. 2015. ISSN 01691368. doi: 10.1016/j.oregeorev.2015.01.001.

- D. Rogge, N. Halden, and C. Beaumont-Smith. Application of data integration for shear-hosted Au potential modelling: Lynn Lake Greenstone Belt, Northwestern Manitoba, Canada. *Canadian Journal of Remote Sensing*, 29(4):458–471, 2006. doi: 10.5589/m03-013.
- P. Rolet and J.-J. Seguin. Traitement de donnees multivariables (Application aux sciences de la Terre) - Tome 1. Technical Report BRGM/86-DT-005-ISA, BRGM, Orléans, France, 1986a.
- P. Rolet and J.-J. Seguin. Traitement de donnees multivariables (Application aux sciences de la Terre) - Tome 2. Technical Report BRGM/86-DT-005-ISA, BRGM, Orléans, France, 1986b.
- R. Roy, D. Cassard, P. Cobbold, E. Rossello, M. Billa, L. Bailly, and A. Lips. 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(3-4):260–286, Nov. 2006. ISSN 01691368. doi: 10.1016/j.oregeorev.2005.10.002.
- S. Roy, P. Upton, and D. Craw. Gold in the hills: Patterns of placer gold accumulation under dynamic tectonic and climatic conditions. *Mineralium Deposita*, 53(7):981–995, Oct. 2018. ISSN 0026-4598, 1432-1866. doi: 10.1007/s00126-017-0789-6.
- F. J. Rubio Pascual, R. Arenas, J. R. Martínez Catalán, L. R. Rodríguez Fernández, and J. R. Wijbrans. Thickening and exhumation of the Variscan roots in the Iberian Central System: Tectonothermal processes and $40\text{Ar}/^{39}\text{Ar}$ ages. *Tectonophysics*, 587:207–221, Mar. 2013. ISSN 00401951. doi: 10.1016/j.tecto.2012.10.005.
- F. J. Rubio Pascual, A. López-Carmona, and R. Arenas. Thickening vs. extension in the Variscan belt: P–T modelling in the Central Iberian autochthon. *Tectonophysics*, 681:144–158, June 2016. ISSN 00401951. doi: 10.1016/j.tecto.2016.02.033.
- M. Sadeghi, G. Bertrand, D. De Oliveira, N. Arvanitidis, S. Decrée, H. Gautneb, E. Gloaguen, T. Törmänen, H. Reginiussen, H. Sievers, L. Quental, M. J. Ferreira, and A. Wittenberg. Prospectivity mapping of critical raw material at the continental scale- a part of the FRAME project. In *Mineral Exploration Symposium*, pages 1–3, Virtual Event,, 2020. European Association of Geoscientists & Engineers. doi: 10.3997/2214-4609.202089003.
- S. Samson and R. D’Lemos. U-Pb geochronology and Sm-Nd isotopic composition of Proterozoic gneisses, Channel Islands, UK. *Journal of the Geological Society of London*, 155:609–618, 1998.
- E. Sawyer. Melt segregation in the continental crust. *Geology*, 22:1019–1022, 1994.
- H. Schaeben. Comparison of Mathematical Methods of Potential Modeling. *Mathematical Geosciences*, 44(1):101–129, Jan. 2012. ISSN 1874-8961, 1874-8953. doi: 10.1007/s11004-011-9373-2.
- E. Seedorff, J. H. Dilles, J. M. Proffett, M. T. Einaudi, L. Zurcher, W. J. A. Stavast, D. A. Johnson, and M. D. Barton. *Porphyry Deposits Characteristics and Origin of Hypogene Features*. Society of Economic Geologists, 2005. ISBN 978-1-887483-01-8. doi: 10.5382/AV100.10.
- R. Serment. La mine d’antimoine et or de La Lucette (Mayenne, France). *Chronique de la Recherche Minière*, 442:35–42, 1978.
- R. Serment. La mine d’antimoine et or de la Rochettréjoux (Vendée). *Chronique de la Recherche Minière*, 456:91–97, 1980.

- G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, New Jersey, 1976. ISBN 978-0-691-10042-5.
- D. Silva, A. Lima, E. Gloaguen, F. Noronha, and S. Deveaud. Spatial Geostatistical Analysis Applied To The Barroso-Alvão Rare-Elements Pegmatite Field (Northern Portugal). *Frontiers in Information Systems*, 1:67–101, 2018.
- Í. D. D. Silva. Evolução tectono-térmica de um sector da Zona Centro Ibérica na região de Palaçoulo (leste do Maciço de Morais, NE Portugal). 11:5, 2010.
- D. A. Singer and R. Kouda. Application of a feedforward neural network in the search for Kuroko deposits in the Hokuroku district, Japan. *Mathematical Geology*, 28(8):1017–1023, Nov. 1996. ISSN 0882-8121, 1573-8868. doi: 10.1007/BF02068587.
- A. Skabar. Mineral Potential Mapping Using Bayesian Learning for Multilayer Perceptrons. *Mathematical Geology*, 39(5):439–451, Sept. 2007. ISSN 0882-8121, 1573-8868. doi: 10.1007/s11004-007-9106-8.
- R. G. Skirrow, J. Murr, A. Schofield, D. L. Huston, S. van der Wielen, K. Czarnota, R. Coghlan, L. M. Highet, D. Connolly, M. Doublier, and J. Duan. Mapping iron oxide Cu-Au (IOCG) mineral potential in Australia using a knowledge-driven mineral systems-based approach. *Ore Geology Reviews*, 113: 103011, Oct. 2019. ISSN 01691368. doi: 10.1016/j.oregeorev.2019.103011.
- C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1):25, Dec. 2007. ISSN 1471-2105. doi: 10.1186/1471-2105-8-25.
- C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1):307, Dec. 2008. ISSN 1471-2105. doi: 10.1186/1471-2105-9-307.
- P. Suire, M.-P. Dabard, and J.-J. Chauvel. Nouvelles données sur les séries rouges nord-armoricaines : étude du bassin ordovicien de Bréhec. *Comptes-Rendus de l'Académie des Sciences de Paris II*, 312: 721–727, 1991.
- Q. Sun, W.-J. Li, and Z.-W. Fu. A novel anode material of antimony nitride for rechargeable lithium batteries. *Solid State Sciences*, 12(3):397–403, Mar. 2010. ISSN 1293-2558. doi: 10.1016/j.solidstatesciences.2009.12.003.
- T. Sun, F. Chen, L. Zhong, W. Liu, and Y. Wang. GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. *Ore Geology Reviews*, 109:26–49, June 2019. ISSN 01691368. doi: 10.1016/j.oregeorev.2019.04.003.
- T. Sun, H. Li, K. Wu, F. Chen, Z. Zhu, and Z. Hu. Data-Driven Predictive Modelling of Mineral Prospectivity Using Machine Learning and Deep Learning Methods: A Case Study from Southern Jiangxi Province, China. *Minerals*, 10(2):102, Jan. 2020. ISSN 2075-163X. doi: 10.3390/min10020102.
- T. Tabaei, M. Esfahani, P. Rasekh, and A. Esna-ashari. Mineral prospectivity mapping in GIS using fuzzy logic integration in Khondab area, western Markazi province, Iran. *Journal of Tethys*, 5(4):367–379, 2017.

- M. H. Tangestani and F. Moore. The use of Dempster–Shafer model and GIS in integration of geoscientific data for porphyry copper potential mapping, north of Shahr-e-Babak, Iran. *International Journal of Applied Earth Observation and Geoinformation*, 4(1):65–74, Aug. 2002. ISSN 03032434. doi: 10.1016/S0303-2434(02)00008-9.
- R. Tartèse and P. Boulvais. Differentiation of peraluminous leucogranites “en route” to the surface. *Lithos*, 114(3-4):353–368, Feb. 2010. ISSN 00244937. doi: 10.1016/j.lithos.2009.09.011.
- C. Thiart and M. de Wit. Linking Spatial Statistics to GIS: Exploring potential gold and tin models of Africa. *South African Journal of Geology*, 103(3-4):215–230, Dec. 2000. ISSN 1012-0750. doi: 10.2113/1030215.
- F. Tornos, C. M. C. Inverno, C. Casquet, A. Mateus, G. Ortiz, and V. Oliveira. The metallogenic evolution of the Ossa-Morena Zone. *Journal of Iberian Geology*, 30:143–181, 2004.
- J. Torppa, V. Nykänen, and F. Molnár. Unsupervised clustering and empirical fuzzy memberships for mineral prospectivity modelling. *Ore Geology Reviews*, 107:58–71, Apr. 2019. ISSN 01691368. doi: 10.1016/j.oregeorev.2019.02.007.
- B. Tourlière, E. Pakyuz-Charrier, D. Cassard, L. Barbanson, and C. Gumiaux. Cell Based Associations: A procedure for considering scarce and mixed mineral occurrences in predictive mapping. *Computers & Geosciences*, 78:53–62, May 2015. ISSN 00983004. doi: 10.1016/j.cageo.2015.01.012.
- T. T. Tuyen, A. Jaafari, H. P. H. Yen, T. Nguyen-Thoi, T. V. Phong, H. D. Nguyen, H. Van Le, T. T. M. Phuong, S. H. Nguyen, I. Prakash, and B. T. Pham. Mapping forest fire susceptibility using spatially explicit ensemble models based on the locally weighted learning algorithm. *Ecological Informatics*, 63: 101292, July 2021. ISSN 15749541. doi: 10.1016/j.ecoinf.2021.101292.
- O. Vanderhaeghe and C. Teyssier. Partial-melting and flow of orogens. *Tectonophysics*, 342:451–472, 2001.
- Y. Vernhet. Carte géologique harmonisée du département de la Mayenne. Technical Report BRGM/RP - 58050- FR, BRGM, Orléans, France, Jan. 2010.
- G. Vic and B. Tourlière. Revalorisation du potentiel minier français Revue et cartographie prédictive pour W, Au et Ge dans les Pyrénées. Technical Report BRGM/RP-66278-FR, BRGM, Nov. 2016.
- M. Vidal, M.-P. Dabard, R. Gourvenec, A. Le Hérissey, A. Loi, F. Paris, Y. Plusquellec, and P. Racheboeuf. Le Paléozoïque de la presqu’île de Crozon, Massif armoricain (France). *Géologie de la France 2011*, 1:3–45, 2011.
- J. Viruete, R. Arenas, and J. Catalán. Tectonothermal evolution associated with Variscan crustal extension in the Tormes Gneiss Dome (NW Salamanca, Iberian Massif, Spain). *Tectonophysics*, 238(1-4): 117–138, Nov. 1994. ISSN 00401951. doi: 10.1016/0040-1951(94)90052-3.
- J. F. von Raumer, G. M. Stampfli, and F. Bussy. Gondwana-derived microcontinents — the constituents of the Variscan and Alpine collisional orogens. *Tectonophysics*, 365(1-4):7–22, Apr. 2003. ISSN 00401951. doi: 10.1016/S0040-1951(03)00015-5.
- C. Wang, J. Chen, and Y. Ouyang. Determination of Predictive Variables in Mineral Prospectivity Mapping Using Supervised and Unsupervised Methods. *Natural Resources Research*, 31(4):2081–2102, Aug. 2022. ISSN 1573-8981. doi: 10.1007/s11053-021-09982-7.

- G. Wang, S. Zhang, C. Yan, Y. Song, Y. Sun, D. Li, and F. Xu. Mineral potential targeting and resource assessment based on 3D geological modeling in Luanchuan region, China. *Computers & Geosciences*, 37(12):1976–1988, Dec. 2011. ISSN 00983004. doi: 10.1016/j.cageo.2011.05.007.
- K. Wang, N. Li, L. Bagas, S. Li, X. Song, and Y. Cong. GIS-based prospectivity-mapping based on geochemical multivariate analysis technology: A case study of MVT Pb–Zn deposits in the Huanyuan-Fenghuang district, northwestern Hunan Province, China. *Ore Geology Reviews*, 91:1130–1146, Dec. 2017. ISSN 01691368. doi: 10.1016/j.oregeorev.2017.09.015.
- Y. Wang, Z. Fang, H. Hong, R. Costache, and X. Tang. Flood susceptibility mapping by integrating frequency ratio and index of entropy with multilayer perceptron and classification and regression tree. *Journal of Environmental Management*, 289:112449, July 2021. ISSN 03014797. doi: 10.1016/j.jenvman.2021.112449.
- J.-H. Ward. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301):236–244, Mar. 1963.
- J. Wilson. Did the Atlantic close and then re-open? *Nature*, 211:676–681, 1966.
- K. Xiao, N. Li, A. Porwal, E.-J. Holden, L. Bagas, and Y. Lu. GIS-based 3D prospectivity mapping: A case study of Jiama copper-polymetallic deposit in Tibet, China. *Ore Geology Reviews*, 71:611–632, Dec. 2015. ISSN 01691368. doi: 10.1016/j.oregeorev.2015.03.001.
- Y. Xiong and R. Zuo. Effects of misclassification costs on mapping mineral prospectivity. *Ore Geology Reviews*, 82:1–9, Apr. 2017. ISSN 01691368. doi: 10.1016/j.oregeorev.2016.11.014.
- Y. Xiong and R. Zuo. A positive and unlabeled learning algorithm for mineral prospectivity mapping. *Computers & Geosciences*, 147:104667, Feb. 2021. ISSN 00983004. doi: 10.1016/j.cageo.2020.104667.
- Y. Xiong, R. Zuo, and E. J. M. Carranza. Mapping mineral prospectivity through big data analytics and a deep learning algorithm. *Ore Geology Reviews*, 102:811–817, Nov. 2018. ISSN 01691368. doi: 10.1016/j.oregeorev.2018.10.006.
- D. Yadolah. Chi-Square Distance. In *The Concise Encyclopedia of Statistics*, page 622. Springer, New York, 2008. ISBN 978-0-387-31742-7.
- B. W. D. Yardley and J. S. Cleverley. The role of metamorphic fluids in the formation of ore deposits. *Geological Society, London, Special Publications*, 393(1):117–134, 2015. ISSN 0305-8719, 2041-4927. doi: 10.1144/SP393.5.
- C. M. Yeomans, M. Middleton, R. K. Shail, S. Grebby, and P. A. Lusty. Integrated Object-Based Image Analysis for semi-automated geological lineament detection in southwest England. *Computers & Geosciences*, 123:137–148, Feb. 2019. ISSN 00983004. doi: 10.1016/j.cageo.2018.11.005.
- W. J. Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950. ISSN 1097-0142. doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.
- M. Yousefi and E. J. M. Carranza. Fuzzification of continuous-value spatial evidence for mineral prospectivity mapping. *Computers & Geosciences*, 74:97–109, Jan. 2015. ISSN 00983004. doi: 10.1016/j.cageo.2014.10.014.

- F. Yuan, X. Li, M. Zhang, S. M. Jowitt, C. Jia, T. Zheng, and T. Zhou. Three-dimensional weights of evidence-based prospectivity modeling: A case study of the Baixiangshan mining area, Ningwu Basin, Middle and Lower Yangtze Metallogenic Belt, China. *Journal of Geochemical Exploration*, 145:82–97, Oct. 2014. ISSN 03756742. doi: 10.1016/j.gexplo.2014.05.012.
- L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, June 1965. ISSN 0019-9958. doi: 10.1016/S0019-9958(65)90241-X.
- S. Zhang, K. Xiao, E. J. M. Carranza, and F. Yang. Maximum Entropy and Random Forest Modeling of Mineral Potential: Analysis of Gold Prospectivity in the Hezuo–Meiwu District, West Qinling Orogen, China. *Natural Resources Research*, 28(3):645–664, July 2019. ISSN 1520-7439, 1573-8981. doi: 10.1007/s11053-018-9425-0.
- S. Zhang, E. J. M. Carranza, K. Xiao, H. Wei, F. Yang, Z. Chen, N. Li, and J. Xiang. Mineral Prospectivity Mapping based on Isolation Forest and Random Forest: Implication for the Existence of Spatial Signature of Mineralization in Outliers. *Natural Resources Research*, 31(4):1981–1999, Aug. 2022. ISSN 1573-8981. doi: 10.1007/s11053-021-09872-y.
- Y. Zhang, P. Tiño, A. Leonardis, and K. Tang. A Survey on Neural Network Interpretability. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 5(5):726–742, Oct. 2021. ISSN 2471-285X. doi: 10.1109/TETCI.2021.3100641.
- Z. Zhang, R. Zuo, and Y. Xiong. A comparative study of fuzzy weights of evidence and random forests for mapping mineral prospectivity for skarn-type Fe deposits in the southwestern Fujian metallogenic belt, China. *Science China Earth Sciences*, 59(3):556–572, Mar. 2016. ISSN 1674-7313, 1869-1897. doi: 10.1007/s11430-015-5178-3.
- J. Zhao, S. Chen, R. Zuo, and E. J. M. Carranza. Mapping complexity of spatial distribution of faults using fractal and multifractal models: Vectoring towards exploration targets. *Computers & Geosciences*, 37(12):1958–1966, Dec. 2011. ISSN 0098-3004. doi: 10.1016/j.cageo.2011.04.007.
- M. Ziaii, E. J. M. Carranza, and M. Ziaei. Application of geochemical zonality coefficients in mineral prospectivity mapping. *Computers & Geosciences*, 37(12):1935–1945, Dec. 2011. ISSN 0098-3004. doi: 10.1016/j.cageo.2011.05.009.
- F. Zoller, K. Peters, P. M. Zehetmaier, P. Zeller, M. Döblinger, T. Bein, Z. Sofer, and D. Fattakhova-Rohlfing. Making Ultrafast High-Capacity Anodes for Lithium-Ion Batteries via Antimony Doping of Nanosized Tin Oxide/Graphene Composites. *Advanced Functional Materials*, 28(23):1706529, 2018. ISSN 1616-3028. doi: 10.1002/adfm.201706529.
- R. Zuo and E. J. M. Carranza. Support vector machine: A tool for mapping mineral prospectivity. *Computers & Geosciences*, 37(12):1967–1975, Dec. 2011. ISSN 00983004. doi: 10.1016/j.cageo.2010.09.014.
- R. Zuo, Z. Zhang, D. Zhang, E. J. M. Carranza, and H. Wang. Evaluation of uncertainty in mineral prospectivity mapping due to missing evidence: A case study with skarn-type Fe deposits in Southwestern Fujian Province, China. *Ore Geology Reviews*, 71:502–515, Dec. 2015. ISSN 01691368. doi: 10.1016/j.oregeorev.2014.09.024.

Appendix A

Overlapping simulations

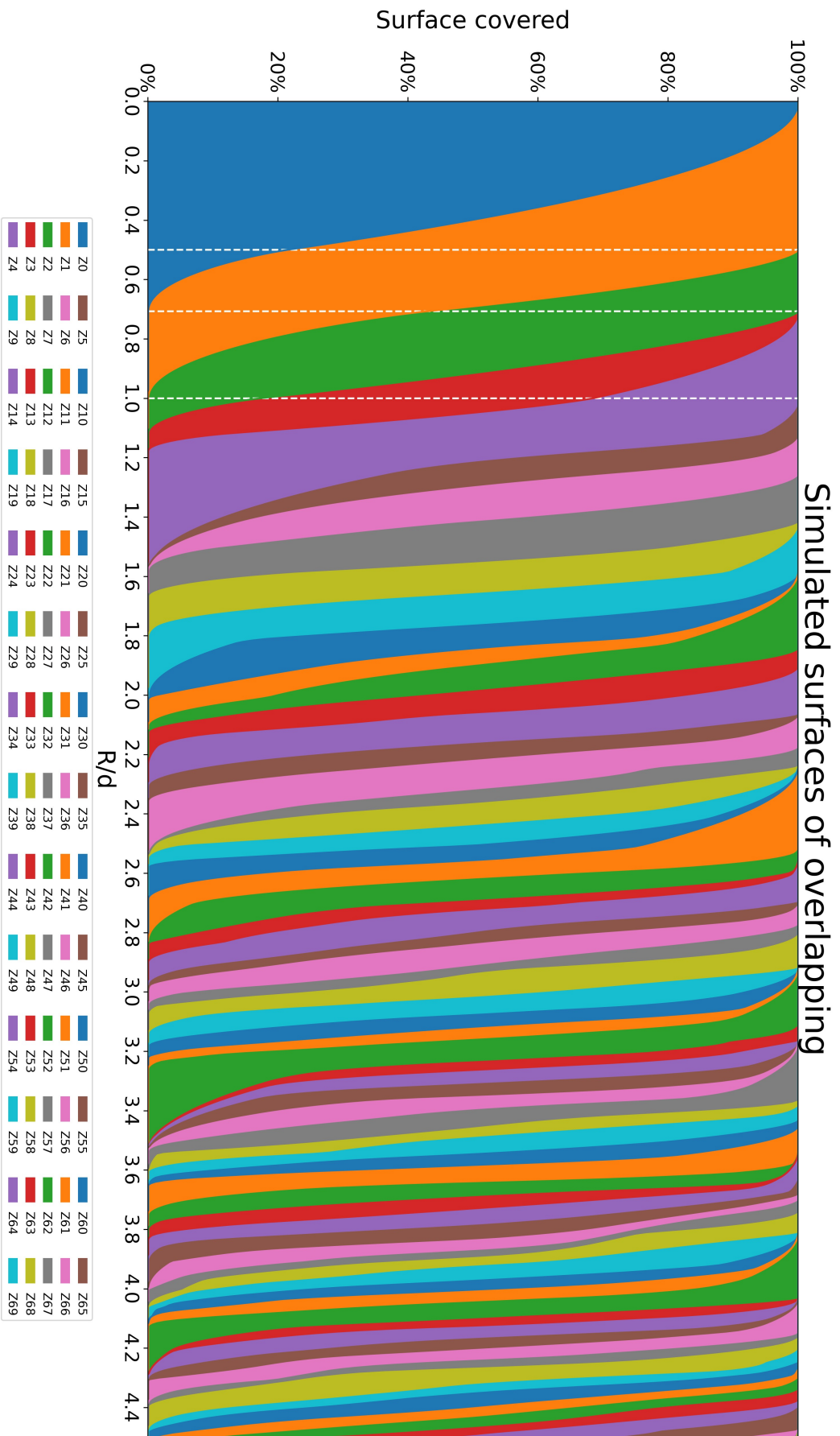


Figure A.1: Simulated surfaces of overlapping. The colors used are cyclical. 1. Threshold of 0.5, the neighbouring search correspond to an inscribed circle. 2. Threshold of 0.71 ($\frac{d\sqrt{2}}{2}$), all of the studied surface is cover by at least 1 disc. 3. Threshold of 1, the neighbouring search correspond to a circumscribed circle.

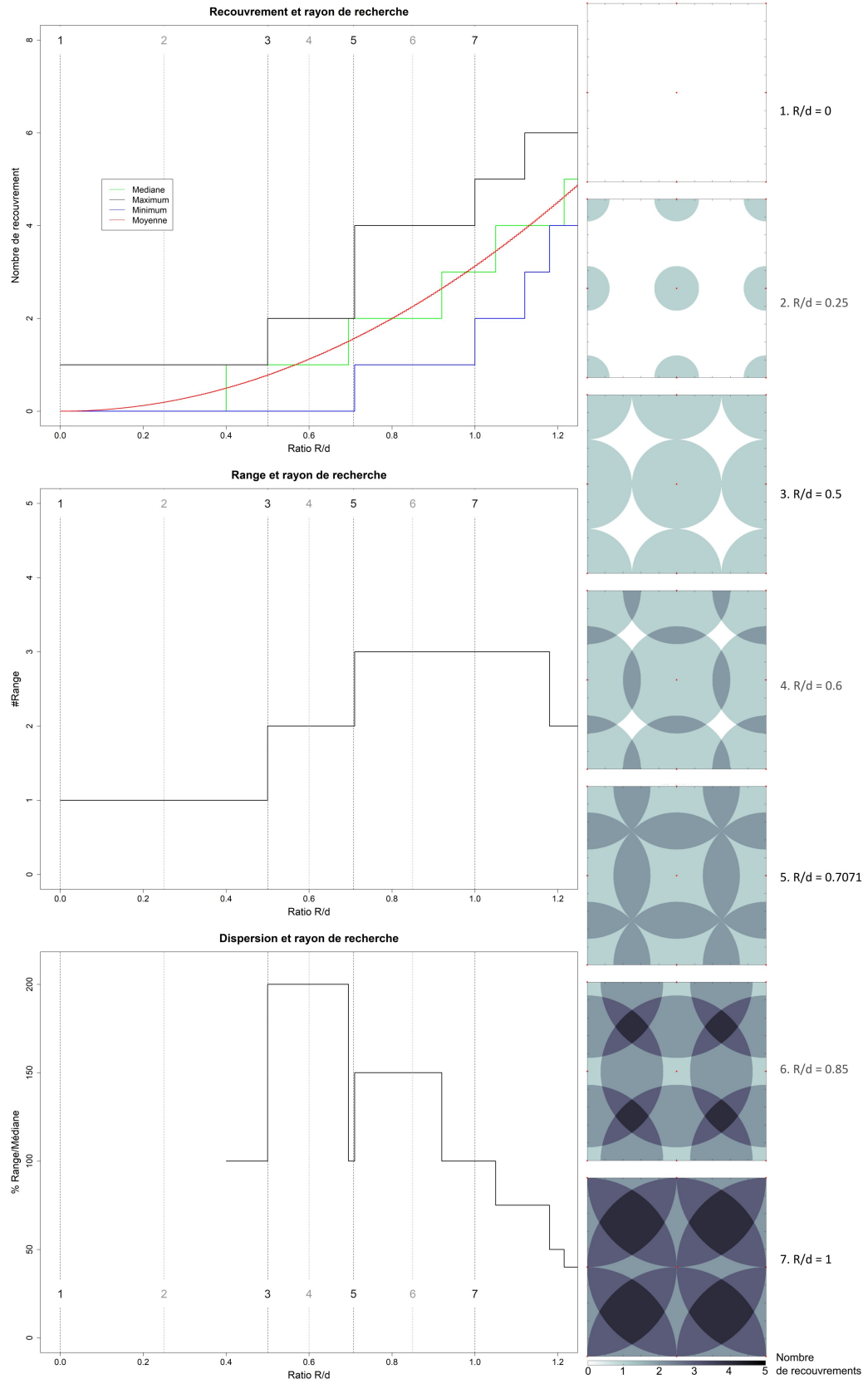


Figure A.2: Overlapping computation results from a R/d of 0 to 1, with a step of 0.005. On the top left are the *Maximum*, *Minimum*, *Median* and *Average* overlapping, followed under by the *Range* value and at the bottom left the *Dispersion* values. On the right are various results of the numerical simulation, depending of the R/d ratio.

Recouvrement et rayon de recherche

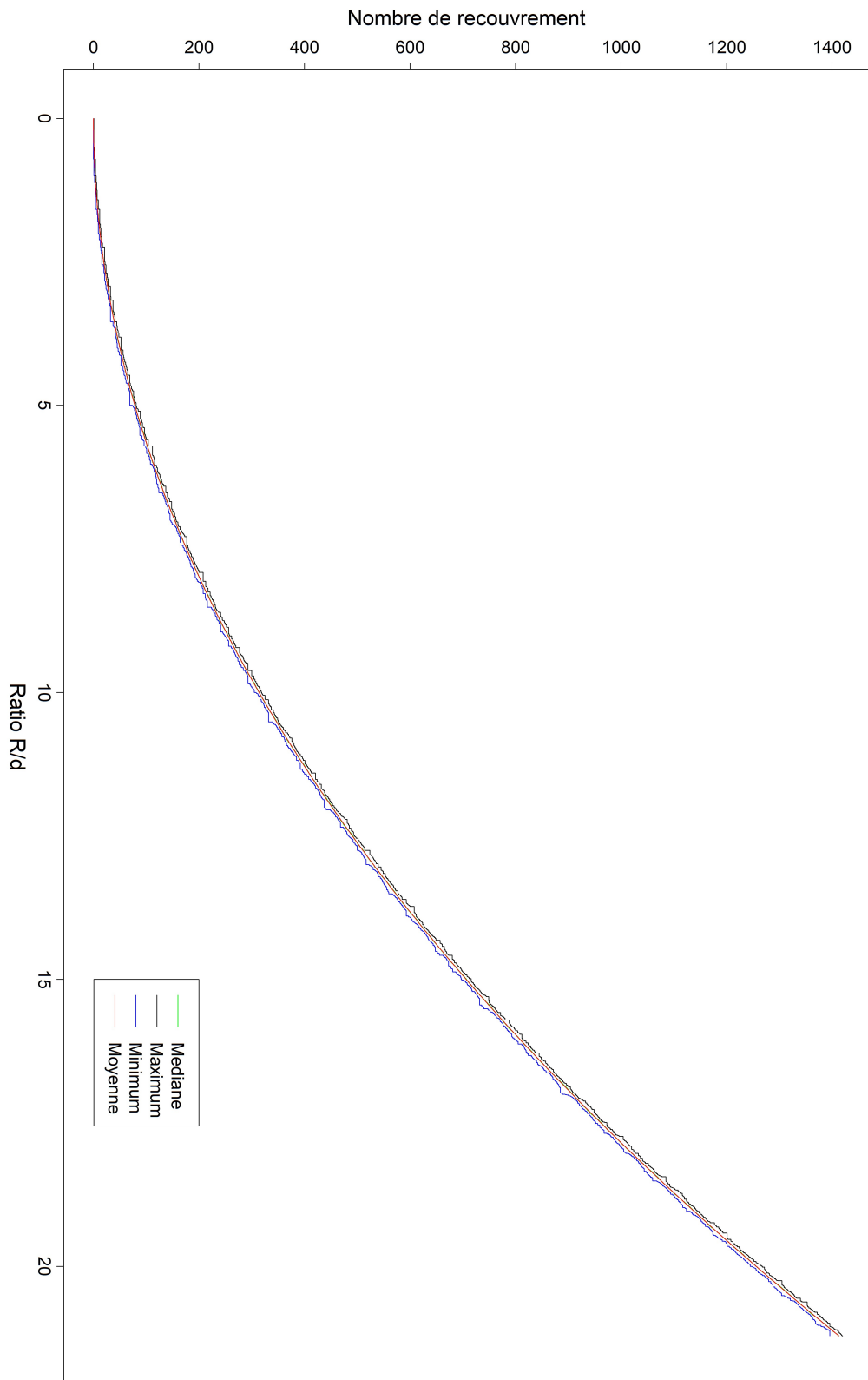


Figure A.3: Statistics from the overlapping computation results for a R/d of 0 to 22, with a step of 0.005.

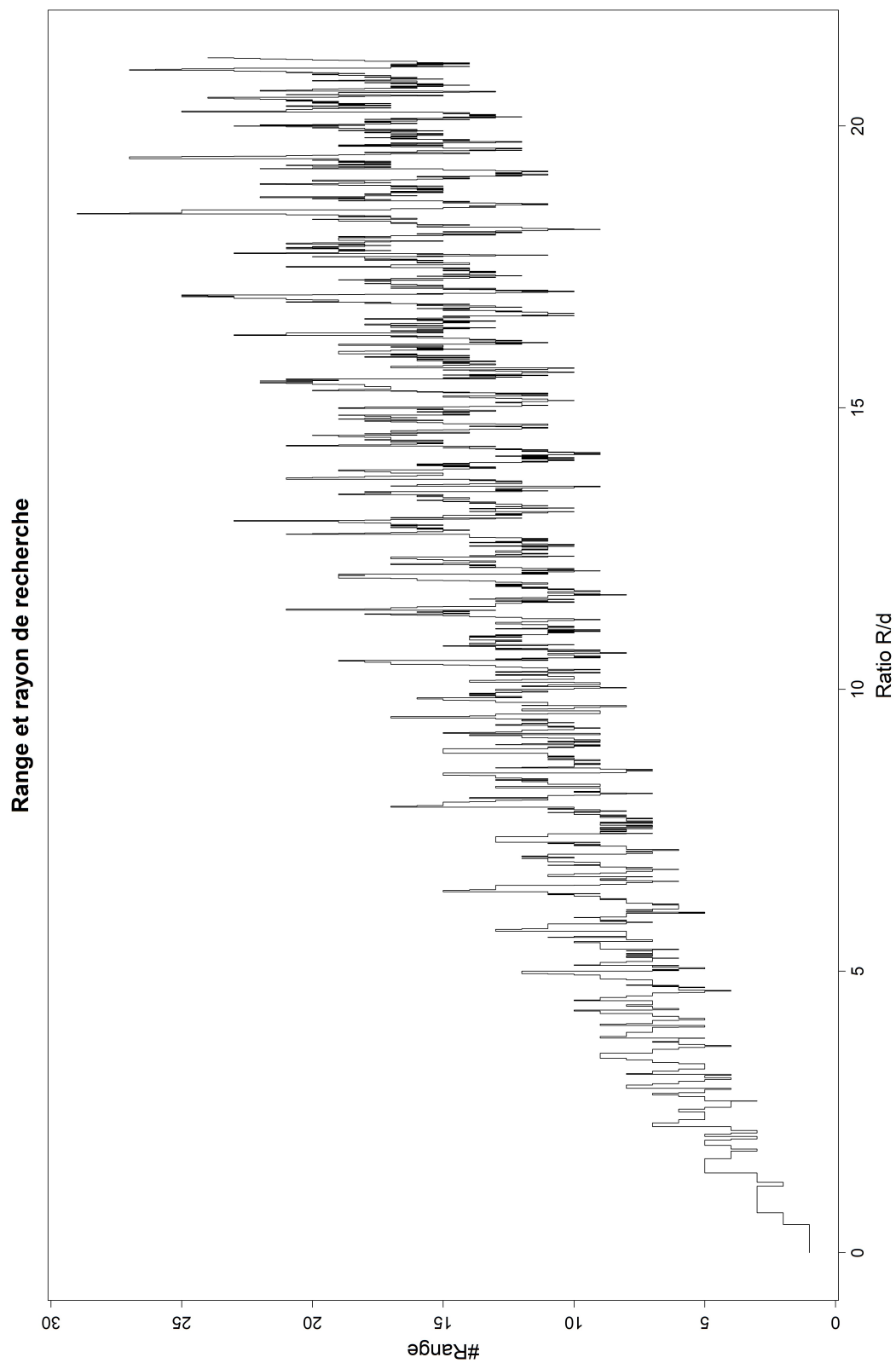


Figure A.4: Range value from the overlapping computation results for a R/d of 0 to 22, with a step of 0.005. $Range = Maximum - Minimum$

Dispersion et rayon de recherche

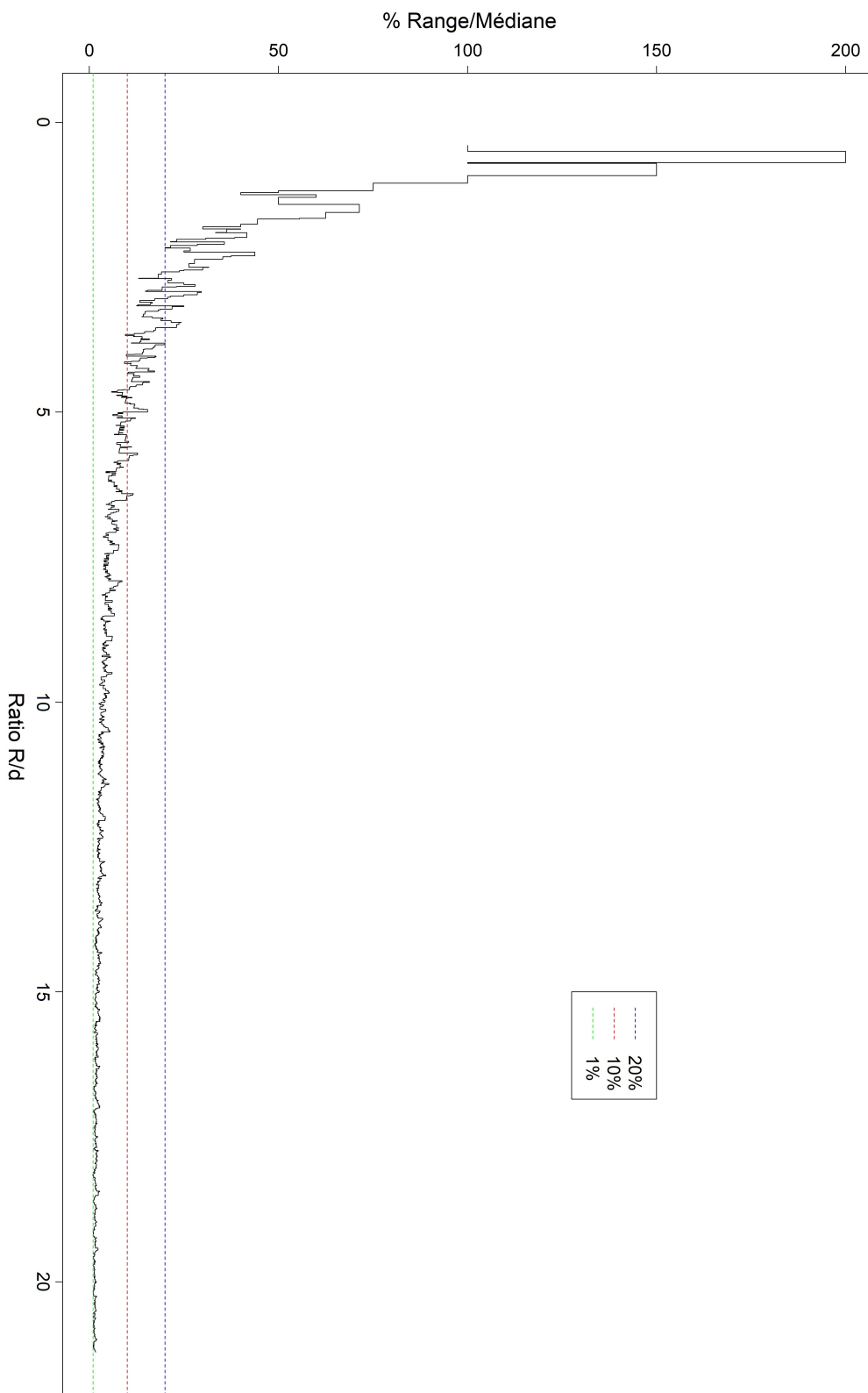


Figure A.5: Range values from the overlapping simulation, from a R/d of 0 to 2 (step of 0.005). $Dispersion = \frac{Maximum - Minimum}{Median} = \frac{Range}{Median}$

Appendix B

Lithological datasets details for the Armorican Massif (50K) and Central Iberian Zone (1M)

a.

Age stratigraphique	Intervalle de datation	Code
Carbonifère	359-299Ma	7
Devonien	416-359Ma	6
Silurien	443-416Ma	5
Ordovicien	488-443Ma	4
Cambrien	542-488Ma	3
Néoprotérozoïque	100-542Ma	2
Paléoprotérozoïque	2500-1600Ma	1

b.

Roches magmatiques		Filons		Roches métamorphiques		Formations diverses du socle		Formations sédimentaires et quaternaires	
Nom	Code	Nom	Code	Nom	Code	Nom	Code	Nom	Code
Albitites	ALBT	aprites	FAPLI	Amphibolites	AMPH	Andésites	ANDE	Alluvions	ALLU
Diorites	DIOR	barytine	FBAR	Gneiss	GNEI	Anorthosites	ANOR	Colluvions	COLL
Dolérites	DOLE	dacite	FDACI	Cornéennes	KORN		CFER	Formations crétacées	CRET
Gabbros	GABB	dolérite	FDOLE	Leptynites	LEPT	Calcaires	CALC	Formations éocènes	EOC
Granodiorites	GDIO	microgranite	FMCGR	Micaschistes	MCSH	Grès	GRES	Formations jurassiques	JUR
Granite	GRAN	microdiorite	FMDIO	Migmatites	MIGM	Greywackes	GRWK	formations miocènes	MIOC
Microgranites	MCGR	microsyénite	FMSYE	Orthogneiss	ORTG	Volcanites acides	VACI	Formations oligocènes	OLIG
Monzogranites	MZGR	quartz	FQTZ	Schistes	SCSH	Volcanites basiques	VBAS	Formations quaternaires	QUAT
Ultra basites	UBAS					Volcano-détritique	VDET	Formations anthropiques	X
						Volcanites	VOLC	Lacs, rivières	HYDRO

Figure B.1: Lithostratigraphic code for the 1:50,000 Harmonization of the Armorican Massif by [Gloaguen and Tourlière, 2016b]. a. Numerical code and correspondence to stratigraphic system. b. Text code and correspondence to lithological formations.

Lithological formations of the Central Iberian Zone, from the geological map of the Iberian Peninsula (1M)

LABEL	NAME
138	Lutitas, areniscas, conglomerados, yesos y a veces niveles calcareos. Facies Garum (Gr Tremp y equivalentes)
133	Areniscas y lutitas
132	Dolomías, calizas, calcarenitas, brechas dolomíticas de aspecto masivo, areniscas calcáreas, margas, calizas bioclásticas, a veces con arenas en la base
131	Calizas, areniscas, margas y dolomías; magmatismo básico
130	Calizas bien estratificadas, margas y dolomías de aspecto masivo, a veces con arenas en la base
128	Areniscas y conglomerados cuarcíticos de tonalidades blanco-amarillentas. Facies Utrillas
127	Calizas, calizas con orbitolinas, calizas margosas, margas, areniscas y lutitas
125	Calizas, calizas con toucasías, calizas detróticas, margas y areniscas. Complejo Urgoniano y otros
121	Dolomías masivas, calizas, dolomías y margas
116	Calizas, margas y dolomías
115	Calizas, dolomías y margas
114	Calizas, calizas margosas, calizas nodulosas, calizas oolíticas, calizas arrecifales. Gr. Turia y otros
113	Calizas margosas, calizas tableadas, calizas bioclásticas, margas, margocalizas y areniscas. Gr. Ablanquejo, Chelva y otros
112	Carniolas, brechas dolomíticas, dolomías y calizas. Gr. Renales y equivalentes
111	Areniscas, arcillas y evaporitas con intercalaciones carbonatadas, vulcanismo básico
109	Lutitas, areniscas rojas, calizas, dolomías y yesos
108	Lutitas rojas y verdes con yesos, a veces con intercalaciones de areniscas. Facies Keuper
107	Dolomías, calizas y margas. Facies Muschelkalk
106	Conglomerados, areniscas y arcillas rojas. Facies Buntsandstein
87	Rocas volcánicas. Andesitas, basaltos, dacitas y riolitas
86	Areniscas, lutitas, conglomerados, vulcanitas y calizas. Facies Autuniense, Saxoniense y otras
85	Areniscas, conglomerados, pizarras y capas de carbón. Cuencas del Bierzo, Cifera-Matallana y otras
83	Areniscas, conglomerados, lutitas y capas de carbón. Cuenca de Peñarroya y otras
76v	Vulcanitas
76	Pizarras, areniscas y calizas
70	Cuarcitas y pizarras. Fm San Vitero y otras
65	Esquistos negros, lúditas, cuarcitas y calizas
64	Pizarras, grauwacas y calizas. Fm. Terena, Sobrado y otras
63	Pizarras negras, lúditas, cuarcitas y calizas
60	Pizarras, areniscas con cantos, cuarcitas, calizas y vulcanitas básicas
59	Pizarras y areniscas
58	Pizarras, areniscas, cuarcitas, calizas y ocasionalmente conglomerados. Gr Agüeira y otros
57	Pizarras negras. Fms. Luarca, Róo y otras
56	Cuarcitas - Cuarcita Armoricana

55	Esquistos, pizarras, cuarcitas y conglomerados. SERIE Púrpura y otras
54	Gneises glandulares, vulcanitas, porfiroides ácidos y rocas vulcanosedimentarias. Olló de Sapo, Canigó y otros
53	Cuarcitas y pizarras. SERIE de Los Cabos y otras
52	Esquistos, areniscas, basaltos y riolitas. Fms: Fatuquedo, Ossa y otras
48	Calizas, dolomías, mármoles, conglomerados y vulcanitas ácidas. Fm Pedroches
47e	Esquistos moscovíticos. Fm Cándana-Herrería y otras
47	Areniscas, pizarras, calizas y vulcanitas. Fm Cándana-Herrería y otras
46	Pizarras y vulcanitas. Fm Malcocinado
45	Esquistos, grauwacas y niveles carbonatados. Alcudiense superior
44	Esquistos, grauwacas, cuarcitas y conglomerados. Complejo esquistograuwáquico
43	Esquistos, grauwacas y vulcanitas. Pizarras del Narcea. Serie de Villaba y otras
42mg	Migmaitas
42gn	Gneises
42	Esquistos, pizarras, grauwacas y cuarcitas negras. SERIE negra
41	Gneises blastomilonóticos, migmaitas, anfibolitas, eclogitas y granulitas básicas
40	Paragneises, anfibolitas y metadioritas
39	Gneises, ortogneises, granulitas anfibolitizadas y gabrodioritas
38	Metagrauwacas y filitas. Filitas de Fontes
37	Paragneises con intercalaciones de eclogitas o granulitas. Gneises bandeados de Chimparra
36	Pizarras, esquistos y paragneises. Gneis de Cariño, Esquistos de Árdenes y otros
35	Ortogneises. Ortogneises de Corredoira, Gneis de Lagoa y otros
34	Metagabros, anfibolitas, granulitas básicas, eclogitas y rocas ultrabásicas. Gabro de Montecastelo, Anfibolitas de Fornás, Fm. Bacariza y otra
33	Rocas ultrabásicas
32	Metagrabos, anfibolitas, metasedimentos y rocas ultrabásicas. Unidades de Bazar, Purrido, Moeche, Complejo de Morais-Braganza y otros
31	Rocas ultrabásicas
30	Grauwacas, filitas y vulcanitas básicas
29v	Vulcanitas peralcalinas
29	Esquistos y vulcanitas
28	Filitas, cuarcitas y vulcanitas
27	Esquistos, grauwacas, paragneises y vulcanitas básicas. Esquistos de Cean y otros
26	Ortogneises con intercalaciones de eclogitas. Gneises de Santiago, Gneis de Aguaiada y otros
25	Anfibolitas, Anfibolitas de Cambre y otras
24	Filitas y cuarzovacas con intercalaciones de cuarcita y mármoles
23	Esquistos gráftosos, líditas, cuarcitas, grauwacas y vulcanitas
22	Micaesquistos, metagrauwacas, cuarcitas milonóticas y anfibolitas
21	Gneises con granates, ortogneises y migmaitas

17	Granitoides biotíticos
16	Granitos de dos micas
15	Granitos biotíticos con cordierita
14	Dioritas y gabros
13	Granitoides biotíticos
12	Granitos de dos micas
10	Granitoides biotíticos y rocas básicas relacionadas
9	Granitos y leucogranitos de dos micas
8	Granitoides inhomogéneos y Complejos migmatíticos
7	Rocas básicas y ultrabásicas
6	Monzogranitos biotíticos y granófiros
5	Granitoides alcalinos
4	Rocas básicas y ultrabásicas
3	Granitoides calcoalcalinos y peraluminico y/o ortogneises glandulares
2	Rocas ácidas e intermedias
1	Rocas básicas y ultrabásicas
e	Diques doleríticos

Figure B.2: Lithological formations of the Central Iberian Zone (CIZ) from the 1:1,000,000 geological map of the Iberian Peninsula [[Rodríguez Fernández and Oliveira, 2015](#)].

Appendix C

Applications: HCPC clusters

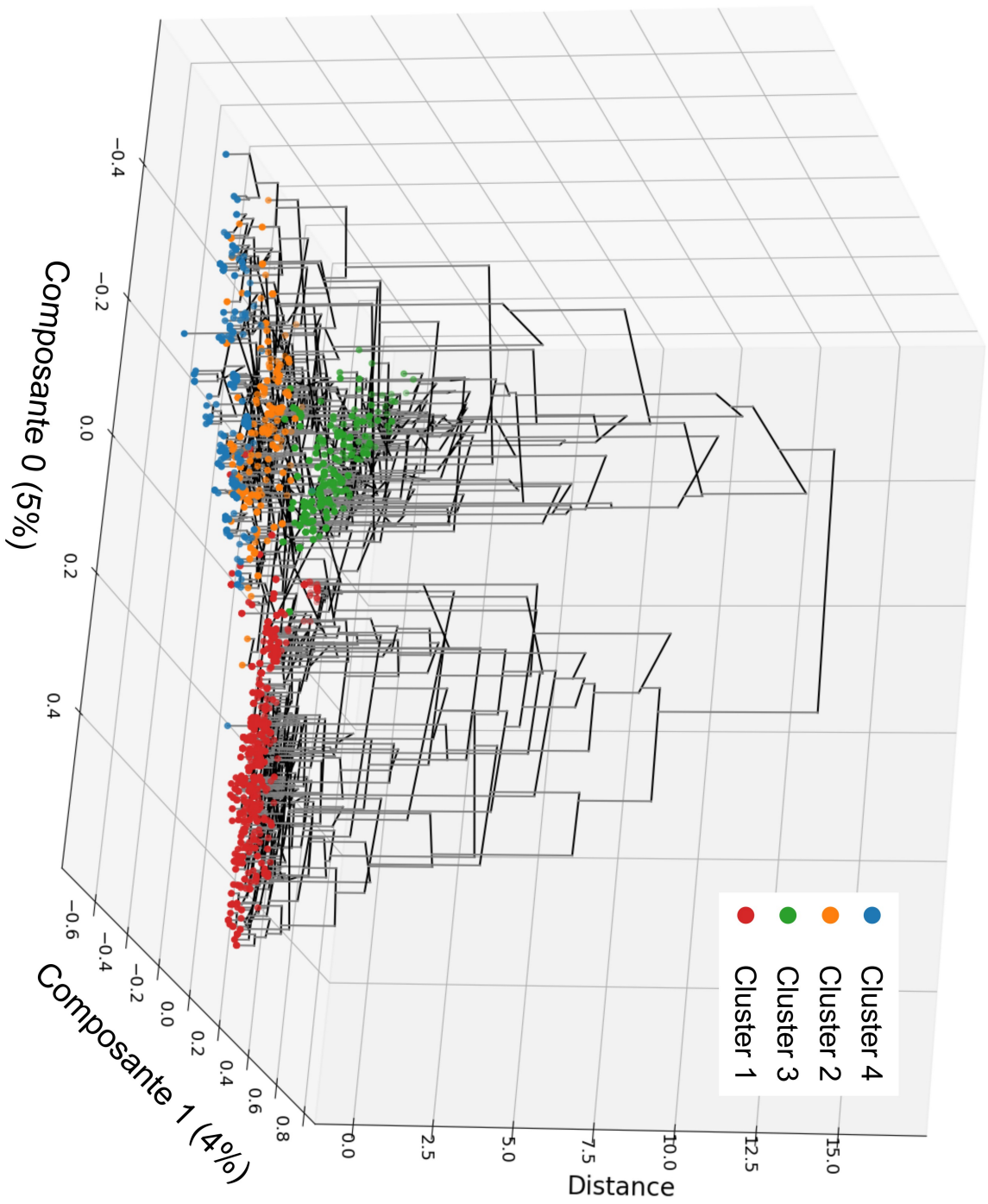


Figure C.1: 3D HCPC classification tree of the cells of the Armorican Massif DBA matrix.

CLUSTER 1							
Lithology	Cells	Cells in AOI	Frequency	Cluster contribution in AOI	Frequency in AOI	Cells in DBA	Frequency in DBA
GRES4	1 497	1 902	87,75%	78,71%	56,84%	17 556	35,92%
Faults	1 435	2 703	84,11%	53,09%	80,78%	30 335	62,07%
SCSH4	1 270	1 602	74,44%	79,28%	47,88%	12 847	26,29%
SCSH2	804	1 340	47,13%	60,00%	40,05%	18 940	38,76%
SCSH5	745	796	43,67%	93,59%	23,79%	9 548	19,54%
GRES6	708	747	41,50%	94,78%	22,33%	5 686	11,63%
FQTZ	689	1 582	40,39%	43,55%	47,28%	21 287	43,56%
CALC6	542	654	31,77%	82,87%	19,55%	2 619	5,36%
FDOLE	489	747	28,66%	65,46%	22,33%	12 020	24,60%
GRES5	475	505	27,84%	94,06%	15,09%	4 523	9,25%
SCSH6	403	462	23,62%	87,23%	13,81%	4 012	8,21%
VACI4	363	474	21,28%	76,58%	14,17%	3 656	7,48%
SCSH7	357	360	20,93%	99,17%	10,76%	3 321	6,80%
GRWK2	282	290	16,53%	97,24%	8,67%	4 868	9,96%
GRWK7	263	266	15,42%	98,87%	7,95%	2 739	5,60%
KORN	249	514	14,60%	48,44%	15,36%	8 436	17,26%
GRES2	248	329	14,54%	75,38%	9,83%	8 537	17,47%
VBAS4	200	213	11,72%	93,90%	6,37%	1 599	3,27%
VACI7	182	182	10,67%	100,00%	5,44%	1 797	3,68%
GDIO2	161	161	9,44%	100,00%	4,81%	5 053	10,34%
CALC7	131	131	7,68%	100,00%	3,92%	1 113	2,28%
VDET4	119	119	6,98%	100,00%	3,56%	812	1,66%
GRES7	109	325	6,39%	33,54%	9,71%	1 989	4,07%
MC SH2	99	944	5,80%	10,49%	28,21%	11 794	24,13%
MCGR7	80	80	4,69%	100,00%	2,39%	713	1,46%
VDET7	68	102	3,99%	66,67%	3,05%	1 134	2,32%
FMDIO	60	60	3,52%	100,00%	1,79%	1 733	3,55%
GRAN7	59	807	3,46%	7,31%	24,12%	16 317	33,39%
VOLC7	53	53	3,11%	100,00%	1,58%	365	0,75%
GRAN2	46	76	2,70%	60,53%	2,27%	3 571	7,31%
GRWK6	45	45	2,64%	100,00%	1,34%	592	1,21%
FMCGR	45	79	2,64%	56,96%	2,36%	6 762	13,84%
VBAS3	34	34	1,99%	100,00%	1,02%	49	0,10%
GRWK3	29	498	1,70%	5,82%	14,88%	1 485	3,04%
CALC5	26	26	1,52%	100,00%	0,78%	118	0,24%
VACI2	24	24	1,41%	100,00%	0,72%	1 771	3,62%
VDET5	15	15	0,88%	100,00%	0,45%	283	0,58%
DOLE6	11	26	0,64%	42,31%	0,78%	654	1,34%
DOLE7	10	10	0,59%	100,00%	0,30%	454	0,93%
VBAS6	10	186	0,59%	5,38%	5,56%	1 246	2,55%
VBAS2	7	7	0,41%	100,00%	0,21%	1 509	3,09%
DOLE2	3	152	0,18%	1,97%	4,54%	1 311	2,68%
GNEI2	3	254	0,18%	1,18%	7,59%	3 726	7,62%
UBAST	3	341	0,18%	0,88%	10,19%	3 305	6,76%
ORTG4	3	446	0,18%	0,67%	13,33%	5 852	11,97%
MC SH5	2	179	0,12%	1,12%	5,35%	1 351	2,76%
AMPH4	2	338	0,12%	0,59%	10,10%	1 740	3,56%
GRAN4	1	168	0,06%	0,60%	5,02%	2 448	5,01%
AMPH2	1	195	0,06%	0,51%	5,83%	2 914	5,96%
SCSH3	1	346	0,06%	0,29%	10,34%	1 351	2,76%

Figure C.2: Attributes of the lithological formations within the Cluster 1 of the Armorican Massif.

CLUSTER 2							
Lithology	Cells	Cells in AOI	Frequency	Cluster contribution in AOI	Frequency in AOI	Cells in DBA	Frequency in DBA
Faults	358	2 703	75,21%	13,24%	80,78%	30 335	62,07%
ORTG4	340	446	71,43%	76,23%	13,33%	5 852	11,97%
UBAST	306	341	64,29%	89,74%	10,19%	3 305	6,76%
AMPH4	303	338	63,66%	89,64%	10,10%	1 740	3,56%
MCSH4	262	288	55,04%	90,97%	8,61%	4 776	9,77%
GRES4	194	1 902	40,76%	10,20%	56,84%	17 556	35,92%
GNEI2	187	254	39,29%	73,62%	7,59%	3 726	7,62%
MCSH5	177	179	37,18%	98,88%	5,35%	1 351	2,76%
MCSH2	177	944	37,18%	18,75%	28,21%	11 794	24,13%
SCSH4	143	1 602	30,04%	8,93%	47,88%	12 847	26,29%
GRWK3	139	498	29,20%	27,91%	14,88%	1 485	3,04%
FQTZ	116	1 582	24,37%	7,33%	47,28%	21 287	43,56%
KORN	110	514	23,11%	21,40%	15,36%	8 436	17,26%
VACI4	87	474	18,28%	18,35%	14,17%	3 656	7,48%
VBAS6	79	186	16,60%	42,47%	5,56%	1 246	2,55%
GNEI4	76	80	15,97%	95,00%	2,39%	3 359	6,87%
LEPT4	72	181	15,13%	39,78%	5,41%	1 701	3,48%
CALC6	71	654	14,92%	10,86%	19,55%	2 619	5,36%
GRES7	61	325	12,82%	18,77%	9,71%	1 989	4,07%
SCSH6	49	462	10,29%	10,61%	13,81%	4 012	8,21%
GRES2	47	329	9,87%	14,29%	9,83%	8 537	17,47%
GRES6	38	747	7,98%	5,09%	22,33%	5 686	11,63%
AMPH6	37	37	7,77%	100,00%	1,11%	740	1,51%
SCSH5	34	796	7,14%	4,27%	23,79%	9 548	19,54%
GRWK4	29	29	6,09%	100,00%	0,87%	255	0,52%
MCSH3	28	255	5,88%	10,98%	7,62%	866	1,77%
LEPT6	27	27	5,67%	100,00%	0,81%	118	0,24%
SCSH3	25	346	5,25%	7,23%	10,34%	1 351	2,76%
FMCGR	20	79	4,20%	25,32%	2,36%	6 762	13,84%
FDOLE	16	747	3,36%	2,14%	22,33%	12 020	24,60%
AMPH2	13	195	2,73%	6,67%	5,83%	2 914	5,96%
VBAS4	10	213	2,10%	4,69%	6,37%	1 599	3,27%
SCSH2	9	1 340	1,89%	0,67%	40,05%	18 940	38,76%
VDET3	7	195	1,47%	3,59%	5,83%	662	1,35%
MIGM7	7	213	1,47%	3,29%	6,37%	3 155	6,46%
GABB6	5	5	1,05%	100,00%	0,15%	451	0,92%
FPEGM	5	68	1,05%	7,35%	2,03%	3 252	6,65%
GRWK7	3	266	0,63%	1,13%	7,95%	2 739	5,60%
SCSH7	2	360	0,42%	0,56%	10,76%	3 321	6,80%

Figure C.3: Attributes of the lithological formations within the Cluster 2 of the Armorican Massif.

CLUSTER 3							
Lithology	Cells	Cells in AOI	Frequency	Cluster contribution in AOI	Frequency in AOI	Cells in DBA	Frequency in DBA
GRAN7	707	807	87,94%	87,61%	24,12%	16 317	33,39%
Faults	692	2 703	86,07%	25,60%	80,78%	30 335	62,07%
MCSH2	664	944	82,59%	70,34%	28,21%	11 794	24,13%
SCSH2	526	1 340	65,42%	39,25%	40,05%	18 940	38,76%
FQTZ	436	1 582	54,23%	27,56%	47,28%	21 287	43,56%
MIGM7	189	213	23,51%	88,73%	6,37%	3 155	6,46%
GRAN4	166	168	20,65%	98,81%	5,02%	2 448	5,01%
GRES7	153	325	19,03%	47,08%	9,71%	1 989	4,07%
FDOL	153	747	19,03%	20,48%	22,33%	12 020	24,60%
DOL2	149	152	18,53%	98,03%	4,54%	1 311	2,68%
ORTG4	95	446	11,82%	21,30%	13,33%	5 852	11,97%
KORN	94	514	11,69%	18,29%	15,36%	8 436	17,26%
GABB4	85	85	10,57%	100,00%	2,54%	554	1,13%
GDIO7	76	76	9,45%	100,00%	2,27%	2 018	4,13%
DIOR4	75	75	9,33%	100,00%	2,24%	570	1,17%
AMPH2	66	195	8,21%	33,85%	5,83%	2 914	5,96%
GNEI2	63	254	7,84%	24,80%	7,59%	3 726	7,62%
SCSH4	57	1 602	7,09%	3,56%	47,88%	12 847	26,29%
GRES4	54	1 902	6,72%	2,84%	56,84%	17 556	35,92%
GABB2	48	48	5,97%	100,00%	1,43%	1 810	3,70%
FPEGM	47	68	5,85%	69,12%	2,03%	3 252	6,65%
MCGR4	41	41	5,10%	100,00%	1,23%	85	0,17%
VDET7	34	102	4,23%	33,33%	3,05%	1 134	2,32%
AMPH4	33	338	4,10%	9,76%	10,10%	1 740	3,56%
MIGM2	30	30	3,73%	100,00%	0,90%	1 703	3,48%
GRAN2	30	76	3,73%	39,47%	2,27%	3 571	7,31%
ORTG7	27	27	3,36%	100,00%	0,81%	598	1,22%
MCSH4	26	288	3,23%	9,03%	8,61%	4 776	9,77%
UBAST	25	341	3,11%	7,33%	10,19%	3 305	6,76%
GDIO4	14	14	1,74%	100,00%	0,42%	314	0,64%
DOL6	13	26	1,62%	50,00%	0,78%	654	1,34%
SCSH6	10	462	1,24%	2,16%	13,81%	4 012	8,21%
GRES2	7	329	0,87%	2,13%	9,83%	8 537	17,47%
GNEI4	4	80	0,50%	5,00%	2,39%	3 359	6,87%
SCSH5	4	796	0,50%	0,50%	23,79%	9 548	19,54%
VBAS6	3	186	0,37%	1,61%	5,56%	1 246	2,55%
VBAS4	3	213	0,37%	1,41%	6,37%	1 599	3,27%
GRES5	3	505	0,37%	0,59%	15,09%	4 523	9,25%
GRAN6	2	2	0,25%	100,00%	0,06%	421	0,86%
GABB7	1	1	0,12%	100,00%	0,03%	772	1,58%

Figure C.4: Attributes of the lithological formations within the Cluster 3 of the Armorican Massif.

CLUSTER 4							
Lithology	Cells	Cells in AOI	Frequency	Cluster contribution in AOI	Frequency in AOI	Cells in DBA	Frequency in DBA
FQTZ	341	1 582	94,72%	21,55%	47,28%	21 287	43,56%
GRWK3	330	498	91,67%	66,27%	14,88%	1 485	3,04%
SCSH3	320	346	88,89%	92,49%	10,34%	1 351	2,76%
MCSH3	227	255	63,06%	89,02%	7,62%	866	1,77%
Faults	218	2 703	60,56%	8,07%	80,78%	30 335	62,07%
VDET3	188	195	52,22%	96,41%	5,83%	662	1,35%
GRES4	157	1 902	43,61%	8,25%	56,84%	17 556	35,92%
SCSH4	132	1 602	36,67%	8,24%	47,88%	12 847	26,29%
AMPH2	115	195	31,94%	58,97%	5,83%	2 914	5,96%
LEPT4	109	181	30,28%	60,22%	5,41%	1 701	3,48%
VBAS6	94	186	26,11%	50,54%	5,56%	1 246	2,55%
FDOLE	89	747	24,72%	11,91%	22,33%	12 020	24,60%
MZGR7	61	61	16,94%	100,00%	1,82%	4 371	8,94%
KORN	61	514	16,94%	11,87%	15,36%	8 436	17,26%
CALC6	41	654	11,39%	6,27%	19,55%	2 619	5,36%
GRAN7	41	807	11,39%	5,08%	24,12%	16 317	33,39%
VACI3	29	29	8,06%	100,00%	0,87%	1 334	2,73%
GRES2	27	329	7,50%	8,21%	9,83%	8 537	17,47%
GRES5	27	505	7,50%	5,35%	15,09%	4 523	9,25%
VACI4	24	474	6,67%	5,06%	14,17%	3 656	7,48%
MIGM7	17	213	4,72%	7,98%	6,37%	3 155	6,46%
FPEGM	16	68	4,44%	23,53%	2,03%	3 252	6,65%
FMCGR	14	79	3,89%	17,72%	2,36%	6 762	13,84%
SCSH5	13	796	3,61%	1,63%	23,79%	9 548	19,54%
GRES3	8	8	2,22%	100,00%	0,24%	1 365	2,79%
GRWK2	8	290	2,22%	2,76%	8,67%	4 868	9,96%
ORTG4	8	446	2,22%	1,79%	13,33%	5 852	11,97%
DIOR7	7	7	1,94%	100,00%	0,21%	1 058	2,16%
UBAST	7	341	1,94%	2,05%	10,19%	3 305	6,76%
MCSH2	4	944	1,11%	0,42%	28,21%	11 794	24,13%
DOLE6	2	26	0,56%	7,69%	0,78%	654	1,34%
GRES7	2	325	0,56%	0,62%	9,71%	1 989	4,07%
GRAN4	1	168	0,28%	0,60%	5,02%	2 448	5,01%
GNEI2	1	254	0,28%	0,39%	7,59%	3 726	7,62%
SCSH7	1	360	0,28%	0,28%	10,76%	3 321	6,80%
GRES6	1	747	0,28%	0,13%	22,33%	5 686	11,63%
SCSH2	1	1 340	0,28%	0,07%	40,05%	18 940	38,76%

Figure C.5: Attributes of the lithological formations within the Cluster 4 of the Armorican Massif.

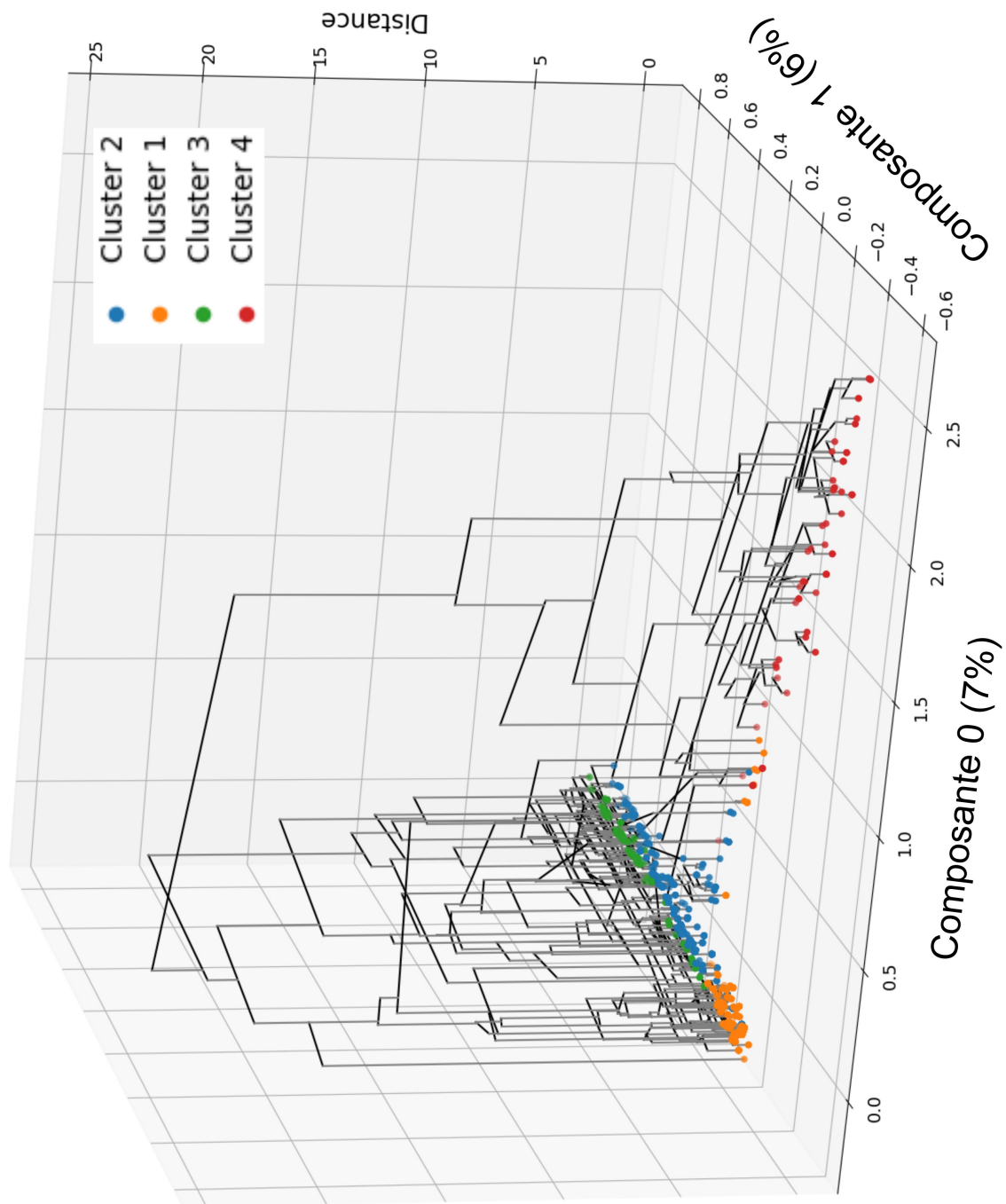


Figure C.6: 3D HCPC classification tree of the cells of the Central Iberian Zone DBA matrix.

CLUSTER 1							
Lithology	Cells in Cluster 1	Cells in AOI	Frequency in cluster	Cluster contribution in AOI	Frequency in AOI	Cells in DBA	Frequency in DBA
57 Pizarras negras, Fms. Luanca, Róo y otras	2348	2430	94,22%	96,63%	49,30%	8959	20,59%
58 Pizarras, areniscas, cuarcitas, calizas y ocasionalmente conglomerados. Gr Agüeira y otros	2008	2088	80,58%	96,17%	42,36%	5032	11,57%
55 Esquistos, pizarras, cuarcitas y conglomerados. SERIE Púrpura y otras	1896	2032	76,08%	93,31%	41,23%	6336	14,56%
65 Esquistos negros, lilitas, cuarcitas y calizas	1826	1960	73,27%	93,16%	39,76%	6101	14,02%
Faults	1592	2548	63,88%	62,48%	51,69%	16303	37,47%
44 Esquistos, grauwacas, cuarcitas y conglomerados. Complejo esquistó-grauváquico	1360	3182	54,57%	42,74%	64,56%	15243	35,04%
45 Esquistos, grauwacas y niveles carbonatados. Alcuense superior	780	830	31,30%	93,98%	16,84%	5185	11,92%
76 Pizarras, areniscas y calizas	302	505	12,12%	59,80%	10,25%	4038	9,28%
53 Cuarcitas y pizarras. SERIE de Los Cabos y otras	300	327	12,04%	91,74%	6,63%	4630	10,64%
56 Cuarcitas - Cuarcita Armoricana	295	324	11,84%	91,05%	6,57%	4181	9,61%
70 Cuarcitas y pizarras. Fm San Vitero y otras	274	335	11,00%	81,79%	6,80%	1483	3,41%
64 Pizarras, grauwacas y calizas. Fm. Terena, Sobrado y otras	59	319	2,37%	18,50%	6,47%	1075	2,47%
47 Areniscas, pizarras, calizas y vulcanitas. Fm Cándana-Herrero y otras	52	240	2,09%	21,67%	4,87%	4095	9,41%
60 Pizarras, areniscas con cantos, cuarcitas, calizas y vulcanitas básicas	26	849	1,04%	3,06%	17,22%	6190	14,23%
2 Rocas ácidas e intermedias	26	34	1,04%	76,47%	0,69%	670	1,54%
42 Esquistos, pizarras, grauwacas y cuarcitas negras. SERIE negra	18	125	0,72%	14,40%	2,54%	1830	4,21%
46 Pizarras y vulcanitas. Fm Malcochinado	18	62	0,72%	29,03%	1,26%	701	1,61%
1 Rocas básicas y ultrabásicas	13	13	0,52%	100,00%	0,26%	413	0,95%
42mg Migmatitas	6	6	0,24%	100,00%	0,12%	21	0,05%
13 Granitoides biotíticos	4	490	0,16%	0,82%	9,94%	9699	22,29%
10 Granitoides biotíticos y rocas básicas relacionadas	3	220	0,12%	1,36%	4,46%	5723	13,15%
41 Gneises blastomiloníticos, migmatitas, anfíbolitas, eclogitas y granulitas básicas	1	10	0,04%	10,00%	0,20%	837	1,92%
106 Conglomerados, areniscas y arcillas rojas. Facies Buntsandstein	1	1	0,04%	100,00%	0,02%	2259	5,19%
107 Dolomías, calizas y margas. Facies Muschelkalk	1	1	0,04%	100,00%	0,02%	774	1,78%
128 Areniscas y conglomerados cuarcíticos de tonalidades blanco-amarillentas. Facies Utrillas	1	1	0,04%	100,00%	0,02%	707	1,63%
130 Calizas bien estratificadas, margas y dolomías de aspecto masivo, a veces con arenas en la base	1	1	0,04%	100,00%	0,02%	828	1,90%
132 Dolomías, calizas, calcarenitas, brechas dolomíticas de aspecto masivo, areniscas calcáreas, margas, calizas bioclásticas, a veces con arenas en la base	1	1	0,04%	100,00%	0,02%	1423	3,27%

Figure C.7: Attributes of the lithological formations within the Cluster 1 of the Central Iberian Zone.

CLUSTER 2						
Lithology	Cells in Cluster 1	Cells in AOI	Frequency in cluster	Cluster contribution in AOI	Frequency in AOI	Cells in DBA
44 Esquistos, grauwacas, cuarcitas y conglomerados. Complejo esquistoso-grauváquico	1822	3182	92,96%	57,26%	64,56%	15243
15 Granitos biotíticos con cordierita	940	940	47,96%	100,00%	19,07%	3581
60 Pizarras, areniscas con cantos, cuarcitas, calizas y vulcanitas básicas	727	849	37,09%	85,63%	17,22%	6190
Faults	568	2548	28,98%	22,29%	51,69%	16303
13 Granitoides biotíticos	486	490	24,80%	99,18%	9,94%	9699
63 Pizarras negras, lilitas, cuarcitas y calizas	344	404	17,55%	85,15%	8,20%	1752
85 Areniscas, conglomerados, pizarras y capas de carbón. Cuencas del Bierzo, Ciñera-Matallana y otras	306	306	15,61%	100,00%	6,21%	509
64 Pizarras, grauwacas y calizas. Fm. Terena, Sobrado y otras	258	319	13,16%	80,88%	6,47%	1075
9 Granitos y leucogranitos de dos micas	231	388	11,79%	59,54%	7,87%	11843
10 Granitoides biotíticos y rocas básicas relacionadas	176	220	8,98%	80,00%	4,46%	5723
55 Esquistos, pizarras, cuarcitas y conglomerados. SERIE Púrpura y otras	136	2032	6,94%	6,69%	41,23%	6336
76 Pizarras, areniscas y calizas	119	505	6,07%	23,56%	10,25%	4038
47 Areniscas, pizarras, calizas y vulcanitas. Fm Cándana-Herrería y otras	111	240	5,66%	46,25%	4,87%	4095
36 Pizarras, esquistos y paragneises. Gneis de Cariño, Esquistos de Á	107	116	5,46%	92,24%	2,35%	1120
3 Granitoides calcoalcalinos y peraluminico y/o ortogneises glandulares	81	82	4,13%	98,78%	1,66%	4454
65 Esquistos negros, lilitas, cuarcitas y calizas	73	1960	3,72%	3,72%	39,76%	6101
17 Granitoides biotíticos	66	76	3,37%	86,84%	1,54%	4909
35 Ortogneises. Ortogneises de Corredoiras, Gneis de Lagoa y otros	62	65	3,16%	95,38%	1,32%	676
70 Cuarcitas y pizarras. Fm San Vitero y otras	60	335	3,06%	17,91%	6,80%	1483
57 Pizarras negras. Fms. Luear, Río y otras	53	2430	2,70%	2,18%	49,30%	8959
45 Esquistos, grauwacas y niveles carbonatados. Alcludense superior	50	830	2,55%	6,02%	16,84%	5185
58 Pizarras, areniscas, cuarcitas, calizas y ocasionalmente conglomerados. Gr Agüeira y otros	49	2088	2,50%	2,35%	42,36%	5032
12 Granitos de dos micas	47	92	2,40%	51,09%	1,87%	2697
8 Granitoides inhomogéneos y Complejos migmatíticos	38	38	1,94%	100,00%	0,77%	4896
42 Esquistos, pizarras, grauwacas y cuarcitas negras. SERIE negra	28	125	1,43%	22,40%	2,54%	1830
53 Cuarcitas y pizarras. SERIE de Los Cabos y otras	12	327	0,61%	3,67%	6,63%	4630
23 Esquistos grafitosos, lilitas, cuarcitas, grauwacas y vulcanitas	11	325	0,56%	3,38%	6,59%	5077
83 Areniscas, conglomerados, lilitas y capas de carbón. Cuenca de Peñarroya y otras	10	91	0,51%	10,99%	1,85%	365
48 Calizas, dolomías, mármoles, conglomerados y vulcanitas ácidas. Fm Pedroches	9	21	0,46%	42,86%	0,43%	610
41 Gneises blastomiloníticos, migmatitas, anfíbolitas, eclogitas y granulitas básicas	7	10	0,36%	70,00%	0,20%	837
56 Cuarcitas - Cuarcita Armoricana	6	324	0,31%	1,85%	6,57%	4181
131 Calizas, areniscas, margas y dolomías; magmatismo básico	5	5	0,26%	100,00%	0,10%	283
133 Areniscas y lilitas	1	1	0,05%	100,00%	0,02%	236
2 Rocas ácidas e intermedias	1	34	0,05%	2,94%	0,69%	670
46 Pizarras y vulcanitas. Fm Malcocinado	1	62	0,05%	1,61%	1,26%	701

Figure C.8: Attributes of the lithological formations within the Cluster 2 of the Central Iberian Zone.

CLUSTER 3							
Lithology	Cells in Cluster 1	Cells in AOI	Frequency in cluster	Cluster contribution in AOI	Frequency in AOI	Cells in DBA	Frequency in DBA
29 Esquistos y vulcanitas	360	360	93,99%	100,00%	7,30%	1247	2,87%
28 Filitas, cuarcitas y vulcanitas	355	355	92,69%	100,00%	7,20%	983	2,26%
23 Esquistos grafitosos, lilitas, cuarcitas, gravacas y vulcanitas	314	325	81,98%	96,62%	6,59%	5077	11,67%
Faults	300	2548	78,33%	11,77%	51,69%	16303	37,47%
9 Granitos y leucogranitos de dos micas	157	388	40,99%	40,46%	7,87%	11843	27,22%
30 Gravacas, filitas y vulcanitas básicas	151	151	39,43%	100,00%	3,06%	533	1,23%
24 Filitas y cuarzovas con intercalaciones de cuarcita y mármoles	112	112	29,24%	100,00%	2,27%	1468	3,37%
32 Metagrabos, anfibolitas, metasedimentos y rocas ultrabásicas. Unidades de Bazar, Purrido, Moeche, Complejo de Morais-Braganza y otros	84	84	21,93%	100,00%	1,70%	1089	2,50%
63 Pizarras negras, lilitas, cuarcitas y calizas	60	404	15,67%	14,85%	8,20%	1752	4,03%
29v Vulcanitas peralcalinas	57	57	14,88%	100,00%	1,16%	71	0,16%
60 Pizarras, areniscas con cantos, cuarcitas, calizas y vulcanitas básicas	49	849	12,79%	5,77%	17,22%	6190	14,23%
12 Granitos de dos micas	45	92	11,75%	48,91%	1,87%	2697	6,20%
58 Pizarras, areniscas, cuarcitas, calizas y ocasionalmente conglomerados. Gr Agüeira y otros	31	2088	8,09%	1,48%	42,36%	5032	11,57%
57 Pizarras negras. Fms. Luarca, Río y otras	29	2430	7,57%	1,19%	49,30%	8959	20,59%
56 Cuarcitas - Cuarcita Armoricana	23	324	6,01%	7,10%	6,57%	4181	9,61%
65 Esquistos negros, lilitas, cuarcitas y calizas	15	1960	3,92%	0,77%	39,76%	6101	14,02%
53 Cuarcitas y pizarras. SERIE de Los Cabos y otras	15	327	3,92%	4,59%	6,63%	4630	10,64%
16 Granitos de dos micas	14	14	3,66%	100,00%	0,28%	197	0,45%
36 Pizarras, esquistos y paragneises. Gneis de Cariño, Esquistos de Árdenes y otros	9	116	2,35%	7,76%	2,35%	1120	2,57%
10 Granitoides biotíticos y rocas básicas relacionadas	4	220	1,04%	1,82%	4,46%	5723	13,15%
35 Ortogneises. Ortogneises de Corredoiras, Gneis de Lagoa y otros	3	65	0,78%	4,62%	1,32%	676	1,55%
31 Rocas ultrabásicas	1	1	0,26%	100,00%	0,02%	403	0,93%

Figure C.9: Attributes of the lithological formations within the Cluster 3 of the Central Iberian Zone.

CLUSTER 4							
Lithology	Cells in Cluster 1	Cells in AOI	Frequency in cluster	Cluster contribution in AOI	Frequency in AOI	Cells in DBA	Frequency in DBA
Faults	88	2548	93,62%	3,45%	51,69%	16303	37,47%
76 Pizarras, areniscas y calizas	84	505	89,36%	16,63%	10,25%	4038	9,28%
83 Areniscas, conglomerados, lutitas y capas de carbón. Cuenca de Peñarroya y otras	81	91	86,17%	89,01%	1,85%	365	0,84%
42 Esquistos, pizarras, grauwacas y cuarcitas negras. SERIE negra	79	125	84,04%	63,20%	2,54%	1830	4,21%
76v Vulcanitas	78	78	82,98%	100,00%	1,58%	820	1,88%
47 Areniscas, pizarras, calizas y vulcanitas. Fm Cándana-Herrería y otras	77	240	81,91%	32,08%	4,87%	4095	9,41%
42gn Gneises	68	68	72,34%	100,00%	1,38%	287	0,66%
47e Esquistos moscovíticos. Fm Cándana-Herrería y otras	62	62	65,96%	100,00%	1,26%	306	0,70%
60 Pizarras, areniscas con cantos, cuarcitas, calizas y vulcanitas básicas	47	849	50,00%	5,54%	17,22%	6190	14,23%
65 Esquistos negros, lilitas, cuarcitas y calizas	46	1960	48,94%	2,35%	39,76%	6101	14,02%
46 Pizarras y vulcanitas. Fm Malcocinado	43	62	45,74%	69,35%	1,26%	701	1,61%
7 Rocas básicas y ultrabásicas	41	41	43,62%	100,00%	0,83%	451	1,04%
10 Granitoides biotíticos y rocas básicas relacionadas	37	220	39,36%	16,82%	4,46%	5723	13,15%
6 Monzogranitos biotíticos y granófiros	17	17	18,09%	100,00%	0,34%	281	0,65%
48 Calizas, dolomías, mármoles, conglomerados y vulcanitas ácidas. Fm Pedroches	12	21	12,77%	57,14%	0,43%	610	1,40%
52 Esquistos, areniscas, basaltos y riolitas. Fms. Fatuquedo, Ossa y otras	11	11	11,70%	100,00%	0,22%	196	0,45%
17 Granitoides biotíticos	10	76	10,64%	13,16%	1,54%	4909	11,28%
2 Rocas ácidas e intermedias	7	34	7,45%	20,59%	0,69%	670	1,54%
5 Granitoides alcalinos	3	3	3,19%	100,00%	0,06%	419	0,96%
64 Pizarras, grauwacas y calizas. Fm. Terena, Sobrado y otras	2	319	2,13%	0,63%	6,47%	1075	2,47%
41 Gneises blastomiloníticos, migmatitas, anfíbolitas, eclogitas y granulitas básicas	2	10	2,13%	20,00%	0,20%	837	1,92%
70 Cuarcitas y pizarras. Fm San Vitero y otras	1	335	1,06%	0,30%	6,80%	1483	3,41%
3 Granitoides calcoalcalinos y peraluminico y/o ortogneises glandulares	1	82	1,06%	1,22%	1,66%	4454	10,24%

Figure C.10: Attributes of the lithological formations within the Cluster 4 of the Central Iberian Zone.

Mise en évidence des environnements géologiques minéralisés par une nouvelle approche de cartographie prédictive Data-driven

Les différentes approches de cartographie prédictive minérale ont pour objectif commun de délimiter des zones susceptibles d'abriter des minéralisations à travers un domaine d'étude donné. Dans ce cadre, les approches de prédictivité minérale dites « Data-driven » visent à identifier ces zones favorables en utilisant des caractéristiques communément associées aux gisements connus. D'un autre côté, les méthodes de clustering par apprentissage automatique sont capables de généraliser des environnements et des caractéristiques géologiques complexes. En combinant ces méthodes, avec l'utilisation d'un système d'information géographique (SIG) basé sur des analyses spatiales et statistiques, il est alors possible de mettre en évidence les relations entre les minéralisations connues et les caractéristiques géologiques environnantes. Cependant, les données cartographiques utilisées pour la cartographie prédictive Data-driven nécessitent systématiquement un pré-traitement et une homogénéisation, leurs résolutions spatiales et leurs précisions pouvant varier. L'utilisation de cartes géologiques peut également conduire à une mauvaise estimation de la proportion relative d'objets géologiques en fonction de leurs configurations cartographiques, les cartes affichant par exemple des structures 3D sur une surface 2D. Ainsi, une approche basée sur l'application d'une méthode de pré-traitement, combinée à des méthodes d'apprentissage automatique pour la modélisation prédictive et le clustering est développée dans cette thèse, afin d'automatiser l'identification des relations entre minéralisations et objets géologiques s.l. Cette méthode, appelée méthode DBA / RF / HCPC, permet (i) d'identifier des associations spatiales locales d'entités cartographiques au sein de la base de données et en utilisant la *Disc-Based Association* (DBA), (ii) de quantifier les chances que chaque association locale soit potentiellement minéralisée, aussi qualifiées de favorabilité, en se basant sur un modèle prédictif de *Forêt Aléatoire* (Random Forest - RF) et (iii) de discrétiser les zones les plus favorables en fonction des associations contenues, en utilisant la *Classification Hiérarchique sur Composantes Principales* (HCPC). L'utilisation d'une approche Data-driven permet de se libérer autant que possible de l'influence des modèles métallogéniques déjà établis pour tester l'efficacité de processus automatisés et également pouvoir mettre de nouveaux modèles ou des modèles alternatifs en évidence. En effet, cette approche statistique permet d'identifier un ou plusieurs environnements géologiques spécifiques associés à la minéralisation dans une zone d'étude, pouvant être interprétée comme des métallogènes distincts des gisements. Dans les zones bien étudiées, les relations spatiales statistiques trouvées par le modèle peuvent être comparées aux métallogènes et modèles métallogéniques connus et être utilisées pour (i) préciser ces métallogènes (i.e. pondérer leurs caractéristiques constitutives) et (ii) éventuellement même mettre en évidence de nouveaux métallogènes. Cette approche promet également d'être très efficace pour l'exploration de domaines géologiques moins étudiés, délimitant des zones favorables et donnant des indices sur les environnements géologiques à explorer. Cette nouvelle méthodologie est ici appliquée dans un cas d'étude sur l'Arc Ibéro-Armoricain (IAA, chaîne Varisque d'Europe occidentale) et les relations entre les structures Varisques s.l. et les minéralisations d'antimoine. Cette nouvelle approche Data-driven, appliquée à des cas naturels, devrait permettre d'améliorer la cartographie prédictive minérale ainsi que la reconnaissance automatique de nouveaux métallogènes, en fournissant de nouvelles informations sur les processus génétiques à l'origine des gisements.

Mots clés : Cartographie prédictive minérale, Apprentissage machine, Data-driven, Forêts Aléatoires, Antimoine, Chaîne Varisque

Highlighting mineralized geological environments through a new Data-driven predictive mapping approach

Mineral Prospectivity Mapping (MPM) aims at outlining areas with the highest likelihood for hosting mineralization. In that frame, Data-driven MPM approaches aims at identifying such favorable areas based on features associated to known or studied mineral deposits. Meanwhile, machine learning clustering methods are able to perform a generalization of complex geological environments and features. By combining those methods, with the use of Geographic Information System (GIS) based on spatial and statistical processes, it should be possible to highlight the relationships between known mineralization and the geological surrounding features. However, the input cartographic datasets of Data-driven MPM approach requires pre-processing and homogenization, as they can be of various spatial resolutions and precision, while the use of geological maps can also lead to a misestimation of the relative proportion of geological features, depending on their structural configurations, as maps display 3D structures on a 2D surface. As such, a Data-driven approach, based on the application of a pre-processing method combined with machine learning predictive modeling and clustering methods is developed in this thesis to automate the identification of the relationships between mineralization occurrences and geological features s.l. This method, referred to as the DBA / RF / HCPC method, allows (i) identifying spatial associations of features from the input dataset with the *Disc-Based Association* (DBA), (ii) quantifying the likelihood of each local association to contain the targeted commodity based on a *Random Forest* (RF) predictive model and (iii) discretizing the area with the highest likelihood depending on the associations it contains using the *Hierarchical Clustering on Principal Component* (HCPC). The focus on a Data-driven method allows avoiding as far as possible the influence of known and established metallogenic models to test the effectiveness of "fully" automated processes. Indeed, this approach allows the identification of one or several specific geological environments associated to mineralization within a study area which can easily be interpreted as one or several distinct metallogenes for the deposits. In well studied areas, the spatial and statistical relationships found by the model can be compared to the known metallogenes and metallogenic models and be used to (i) precise those metallogenes (i.e. quantifying weights for its constituting features) and (ii) eventually highlight new metallogenes even. Such approach also promises to be very effective for exploration within less studied geological domains giving keys on one to several geological environments to explore, in addition to the delineation of favorable areas. This newly developed methodology is applied as benchmark to the Ibero-Armorican Arc (IAA, Western European Variscan Range) and the relationships between the Variscan structures s.l. and antimony mineralization. We infer this new Data-driven approach applied to natural cases will improve prospectivity mapping and automatic recognition of new metallogenes while giving new insights on the genetic processes forming deposits.

Keywords: Mineral Prospectivity Mapping, Machine learning, Data-driven, Random forest, Antimony, Variscan belt