

D. I2.2.3. CORRELATION REPORT OF CHARACTERIZATION STUDIES BASED ON INFORMATION FROM GEOPHYSICAL AND TRADITIONAL INVESTIGATIONS

P. Kessouri¹, B. Ndiaye¹, C. Ryckebusch¹,
I. Isunza Manrique²
¹BRGM, ²U. of Liège



SUMMARY

D. I2.2.3. Correlation report of characterization studies based on information from geophysical and traditional investigations.....	1
1 List of figures	3
2 List of tables.....	4
3 Introduction.....	5
3.1 Presentation of the Pompey site	6
3.2 Chemical elements studied.....	6
3.3 Geophysical dataset	7
4 Correlation methods used.....	7
4.1 Principal component analysis.....	7
4.2 Pearson's correlation coefficient matrix.....	8
5 Correlation between chemical elements.....	9
5.1 PCA analysis	9
5.1.1 Selection of the number of PCs to keep.....	9
5.1.2 PCA results.....	10
5.2 Pearson's correlation coefficient matrix analysis	15
6 Correlation between chemical and geophysical datasets	17
6.1 Principal Component analysis.....	17
6.1.1 Selection of the number of PCs to keep.....	17
6.1.2 PCA results.....	17
6.2 Pearson's correlation coefficient matrix.....	19
6.3 Cluster selection for the RAPIDM model	20
7 Conclusions	23
8 References	24

1 LIST OF FIGURES

Figure 1 : Map of the Pompey site showing the location of: a) the different geophysical profiling or mapping measurements and b) the different sampling locations	5
Figure 2 : Cumulative percentage of variance versus the 15th first principal components for the geochemical dataset	9
Figure 3: Scree plot of the 15th first PCs for the geochemical data	9
Figure 4: Contribution of the first 20th chemical components for the first 4 PCs	10
Figure 5: Variable correlation graphs between::a) PC1 and PC2; b) PC3 and PC4, as a function of the contribution of each variables to PC1 and PC2, or PC3 and PC4 respectively. The PCs are represented with their respective percentage of variance (see Table 2), e.g. PC1 represent 42.7% of the variance. All the variables are represented as a function of their contribution (contrib in %) to PC1 and PC2, or PC3 and PC4.	11
Figure 6: Variable correlation graphs between:PC1 and PC2, sorted out using a cluster analysis in a) 2 groups; b) 3 groups; c) 5 groups. The k-means cluster analysis is taking into account the correlations between the variables and all the 31 PCs estimated. The distinction between group 2 and 3 in the 5 groups-analysis can thus not be observed on the PC1 vs PC2 graph, but is clearer on other dimensions.....	12
Figure 7: Individuals correlation graphs between: a) PC1 vs PC2 ; b) PC3 vs PC4 for individuals colored by altitude (in m). The PCs are represented with their respective percentage of variance (see Table 2), e.g. PC1 represent 42.7% of the variance. The number associated for each individual (on top of each point) is the position in the result table of the PCA analysis (see Appendix 1).	13
Figure 8 : a) PC1 vs PC2 ; b) PC3 vs PC4 for individuals colored with borehole number scale (FP1 in blue and FP4 in red). The PCs are represented with their respective percentage of variance (see Table 2), e.g. PC1 represent 42.7% of the variance. The number associated for each individual (on top of each point) is the position in the result table of the PCA analysis (see Appendix 1).	13
Figure 9: a) Bi-plot representing on one graph the PCA analysis on PC1 vs PC2 for both variables (points in grey) and individuals (points colored by altitude); b) Same bi-plot with individual groups identified (A and B) and cluster analysis for variables; c) Variable correlation graphs between:PC1 and PC2, sorted out using a cluster analysis for 2 groups.....	14
Figure 10: a) Bi-plot representing on one graph the PCA analysis on PC1 vs PC2 for both variables (points in grey) and individuals (points colored by altitude); b) Same bi-plot with individual groups identified (A and B) and cluster analysis for variables; c) Variable correlation graphs between:PC1 and PC2, sorted out using a cluster analysis for 3 groups.....	15
Figure 11: Pearson's correlation coefficient matrix for the 31 chemical elements studied.....	16
Figure 12 : Cumulative percentage of variance versus the 15th first principal components for the geochemical dataset	17
Figure 13 : Contribution of the first 20th chemical components for the first 5 PCs (also named Dim here). The geophysical parameters are highlighted in red (rho: electrical resistivity; phi: chargeability; MF: metal factor).....	18
Figure 14: Variable correlation graphs between::a) PC1 (or Dim-1) and PC2 (or Dim-2); b) PC3 (or Dim-3) and PC4 (or Dim-4); c) PC4 (or Dim-4) and PC5 (or Dim-5), as a function of the contribution of each variables to PC1 and PC2, or PC3 and PC4, or PC4 and PC5 respectively. The PCs are represented with their respective percentage of variance (see Table 3), e. g. PC1 represent 38.9% of the variance. All the variables are represented as a function of their contribution (contrib in %) to	

PC1 and PC2, or PC3 and PC4, or to PC4 and PC5. The 3 geophysical variables are highlighted in red on the 3 graphs to enhance their contribution to the different PCs.. 19

Figure 15: Pearson's correlation coefficient matrix for the 31 chemical elements studied and the 3 geophysical parameters: resistivity (ρ), chargeability (M), metal factor (MF). The red square highlights the correlations with the 3 geophysical parameters. 20

Figure 16: a) Bi-plot representing on one graph the chemical PCA analysis on PC1 vs PC2 for both variables (points in grey) and individuals (points colored by selected clusters); b) Same bi-plot with individual groups identified (A and B) and cluster analysis for variables (cluster PCA – see section 5.1.2.3); c) Variable correlation graphs between:PC1 and PC2, sorted out using a cluster analysis (cluster PCA) for 3 groups on the chemical PCA analysis. 21

Figure 17: Geophysical parameters plotted versus altitude for each of the 4 boreholes FP1 to FP4 (3 graphs on top), and versus the 4 identified clusters (3 graphs at the bottom)..... 22

2 LIST OF TABLES

Table 1 : Chemical elements studied in the XRF and pXRF analysis 7

Table 2: Eigen values estimated by the PCA for the 15th first principal components (PC) for the geochemical dataset 9

Table 3 : Eigen values estimated by the PCA for the 15th first principal components (PC) for the geochemical dataset 17

Table 4: Summary of the observations made for each layers of materials, based on the cluster selection 23

3 INTRODUCTION

Pompey is one of the three pilot sites of the NWE-REGENERATIS project. It is a former tailing pond owned by the EPFGE (Etablissement Public Foncier de Grand Est, Public Real-Estate Company of Grand Est region). The site has been chosen for two main reasons: (1) it hosted various activities for iron based alloys production; (2) it was just rehabilitated on surface, and historic documentation and investigations are done with respect of the French legislation and threshold values. One of the interest of this site is that it allows testing the REGENERATIS methodologies developed within WPT1 and WPT2 on a site that has already been remediated.

Objectives of the geophysical investigations

Site works provided access to material to perform lab trials and also allowed on-site geophysical measurements. Geophysical data on site were acquired at two different times: (1) a first electrical profile, completed by magnetic susceptibility measurements in the trench pit were conducted in November 2020; (2) Measurements on 5 electrical profiles, 3 seismic profiles, an electromagnetic map and a magnetic map were led in March 2021. Geophysical methods allow mapping the variations of geophysical properties that are indirectly linked to the physical and chemical nature of the formations in place, over the entire surface of the deposit. These methods provide very valuable and almost continuous information laterally and vertically. However, they are only indirectly linked with the targeted properties, such as concentration of metallic elements, size the metallic particles...

Objectives of the traditional sampling investigations

In addition, two phases of sampling were handled on site: (1) traditional sampling prior to the geophysical investigations in fall 2020, within an already-existing pit, from 0 to 2m deep; (2) targeted sampling after the geophysical investigations in the summer 2021 within 5 selected boreholes, from 0 to 9m deep. Traditional sampling investigations on past metallurgical sites and deposits aims at characterizing the nature, physical and chemical composition of the wastes at a punctual location. The combination of the traditional sampling investigations and the geophysical imaging will allow interpreting in a semi-quantitative approach the geophysical observations, using the chemical and physical laboratory characterizations of the punctual samples.

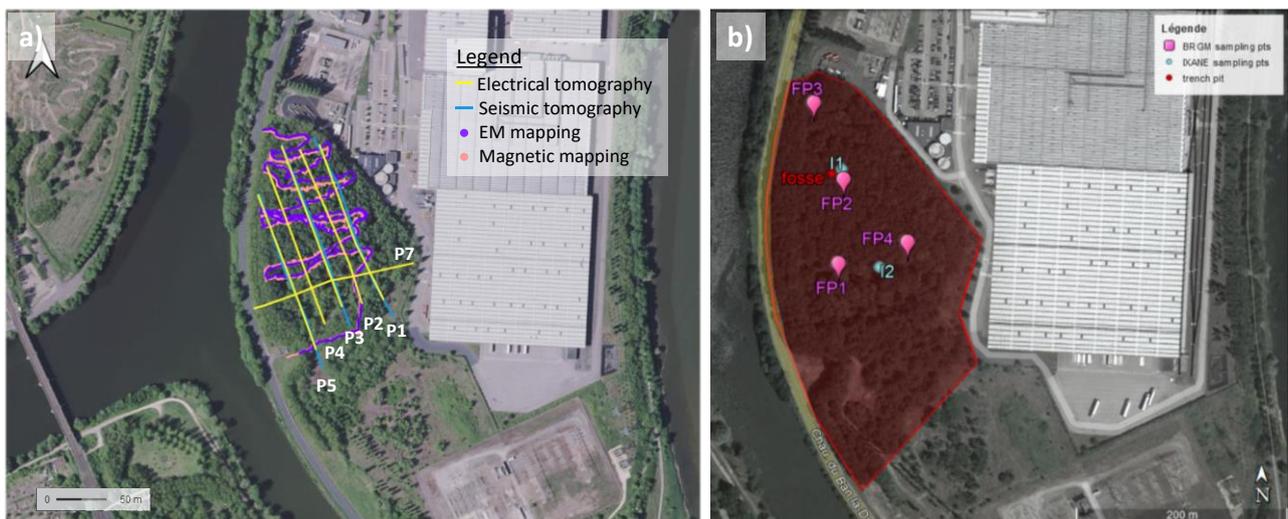


Figure 1 : Map of the Pompey site showing the location of: a) the different geophysical profiling or mapping measurements and b) the different sampling locations

Objectives of the correlation report

This site-specific correlation report is intended to compare the geophysical results to the traditional characterization study (samplings, boreholes, etc.) and results. This correlation results will serve as basis for the establishment of the RAPIDM model for the site.

We will first study the correlation between the different chemical elements for the samples taken on 4 different locations of the site, at various depths from 0 to 9m deep. This will allow us to identify the main chemical elements present at each depth of the settling pond.

We will then study the correlation between the geochemical results and the geophysical dataset in order to identify to which chemical elements the electrical resistivity and chargeability are the most sensitive. We chose to focus on the electrical measurements as they are the ones describing the most the settling pond both laterally and vertically, and they will also be used to implement the RAPIM model.

3.1 PRESENTATION OF THE POMPEY SITE

The Pompey site is a former tailing pond from the iron and steel complex of Pompey-Frouard-Custines, located 10 km North from Nancy (see deliverable DI2.1.1 and Huot, 2013). The steel complex was active from 1870 to 1986. It is renowned for producing cast iron and special steels, such as ferromanganese (ferro-alloy rich in manganese). The last blast furnace of the Pompey-Frouard-Custines iron and steel complex was stopped in 1986. Over time, a forest ecosystem developed on the former tailing pond. The dike delimiting the site was planted with a curtain of black locust trees in 1997. The rest of the pond gradually got covered with diversified deciduous vegetation, more or less dense depending on the area.

The geological substratum of the former tailing pond consists of the Lias marl formations (at 181 m NGF), which are covered by alluvium from the two rivers, composed of coarse siliceous materials (sands, gravel and pebbles) at the base over 3 to 6 m surmounted by finer materials (sands, silts and clays) on 1 to 3 m. These alluvial formations were locally exploited and backfilled with waste rock and iron and steel by-products.

The depth of the deposits in the basin is estimated at around 10 m. The surface of the former pond is estimated to 26 000 m², for a total estimated volume of wastes equal to 260 000 m³.

The waters of the alluvial table would circulate from the channeled Moselle towards the Meurthe, whose level is lower. The piezometric levels measured in 2002 are 187.5 m NGF upstream (South-West of the island) and 184 m NGF downstream (East of the island), the basin surface being at 197 m NGF (ANTEA, 2002).

3.2 CHEMICAL ELEMENTS STUDIED

We studied the correlation between pairs of chemical elements (average content) found after the pXRF analysis in all samples. In order to facilitate the description in the rest of the report, the associated symbol will be used to describe each chemical element. The symbols for each of the chemical elements studied are listed in Table 1. There are 31 chemical elements studied.

Table 1 : Chemical elements studied in the XRF and pXRF analysis

Symbol	Element	Symbol	Element
Fe	Iron	Cu	Copper
Si	Silicon	Sr	Strontium
Ca	Calcium	K	Potassium
Al	Aluminum	V	Vanadium
Mg	Magnesium	P	Phosphorus
Mn	Manganese	S	Sulfur
Ni	Nickel	Ti	Titanium
Cr	Chromium	Zn	Zinc

3.3 GEOPHYSICAL DATASET

The geophysical dataset used in this correlation report is describing 3 different geophysical parameters:

- **Electrical resistivity (rho in ohm.m):** describes the ability of the underground layers studied to resist the flow of current applied. It is linked with the movement of free charges in the layers such as ions in the electrolyte or metallic particles
- **Chargeability (M or phi in mV/V):** describes the capacity of the underground layers studied to store charges, i.e. to act like a capacitor. It is linked with the movement of bound charges in the layers, mostly at the interface between different phases of the medium (e.g. electrical double layer around the solid grains, polarization of metallic particles...)
- **Metal Factor (MF in 1/(ohm.m)):** is a parameter that summarizes the variations observed in terms of electrical resistivity and phase shift. It can be expressed as a ratio between chargeability M and resistivity rho. Here it is taken as: $MF=2000.M/\rho$

The dataset presented here is extracted from the inverted sections presented in report DI2.2.1. We extracted the results from the inversion at cells close to the location of the boreholes and for the different depths at which the samples were taken. For each depth/altitude, we take the mean of the different inversion cells the closest to each sample location. This extraction allows to directly compare the geophysical dataset with the chemical elements analysis made on the samples.

4 CORRELATION METHODS USED

4.1 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set. The goal is to increase the interpretability of the dataset and enable the visualization of multidimensional data. To reduce the dimensionality of the dataset, the data are linearly transformed into a new coordinate system containing the smallest possible number of artificial variables, called principal components, to describe (most of) the data. It is commonly used in the study of the variability of large geochemical datasets (e.g. Grunsky and Smee, 1999; Dempster *et al.*, 2013). The amount of calculated eigenvalues is equal to the number of variables in our analysis (Reimann *et al.*, 2011), which in our case is equal to the number of chemical elements with (or without) the geophysical parameters added (inverted electrical resistivity and chargeability).

Normalization

In order to be able to compare the initial variables between one another, a normalization is needed in order to: (1) remove units from variables by making them dimensionless, (2) standardize the range of values for each variable. The standardization consists of removing the mean and scaling the data so that the variance is equal to one, i.e., subtract the mean of the data and divide by standard deviation (e.g. Davis, 1973).

Selection of the number of principal components

A common approach in helping to choose the number of principal components (PCs) to keep is:

- **The Elbow criterion (or scree plot criterion):** A scree plot is a simple line segment plot that represents the eigenvalues (y-axis) for each individual PC (x-axis). The first component usually explains much of the variability, the next few components explain a moderate amount, and the latter components only explain a small fraction of the overall variability. The scree plot criterion looks for the “elbow” in the curve and selects all components just before the line flattens out. In other terms, the “elbow” is the point beyond which the remaining eigenvalues are all relatively small and of comparable size (Jolliffe 2002, Peres-Neto, Jackson, and Somers (2005))

If the scree plot curve doesn't show a clear “elbow”, other approaches to help reduce the number of PCs are:

- **The Kaiser criterion (or eigenvalue criterion 1):** Only the PCs whose eigenvalues are greater than 1 are retained. An eigenvalue > 1 indicates that PCs account for more variance than accounted by one of the original variables in standardized data. Indeed they represent the majority of the variance in the dataset.
- **Eigenvalue criterion 2:** Another criteria based on eigenvalues analysis is to limit the number of component to that number that accounts for a certain fraction of the total variance, e.g. 70% of the total variance explained.

If too many principal components (more than 3) are kept after applying the different reduction criteria, PCA might not be the best way to visualize the dataset. Instead, other dimension reduction techniques can be considered, such as t-SNE and MDS

4.2 PEARSON'S CORRELATION COEFFICIENT MATRIX

A Bivariate relationship describes a correlation between two variables. There are two primary methods to compute a bivariate relationship between two variables:

- **Pearson's method:** parametric correlation test
- **Spearman and Kendall's methods:** rank-based correlation analysis

The Pearson correlation method is usually used as a primary check for the relationship between two variables. The coefficient of correlation, is a measure of the strength of the linear relationship between two variables. It is computed as follow:

$$r = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

With σ_x and σ_y the respective standard deviation of x and y

The correlation ranges between -1 and 1. A value of near or equal to 0 implies little or no linear relationship between the 2 variables. In contrast, the closer comes to 1 or -1, the stronger the linear relationship is between the 2 variables.

Here we will use the same correlation method than for the other pilote sites: Pearson's correlation coefficient (see DI3.2.3) in order to be able to compare the results between the different sites.

5 CORRELATION BETWEEN CHEMICAL ELEMENTS

5.1 PCA ANALYSIS

5.1.1 Selection of the number of PCs to keep

After calculating the eigenvalues for the 31 principal components, corresponding to the 31 chemical elements that are evaluated, the different selection criteria were applied (see section 4.1 and below). The results for the first 15 PCAs are presented in Table 2.

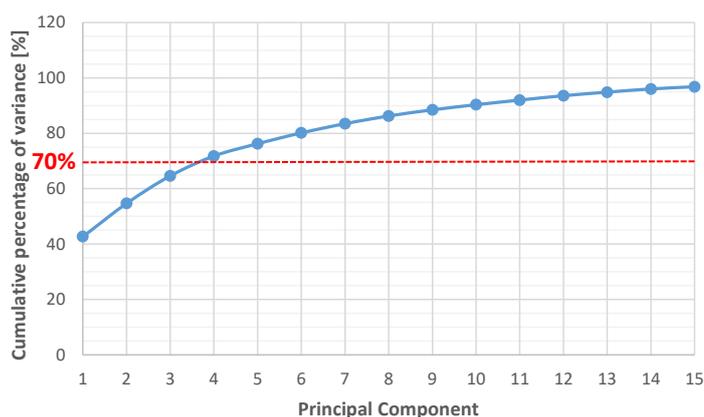
Kaiser's criterion: we decided to keep only the PCs whose eigenvalues are greater than 2. In our case, we will retain the **first 4 principal components** (see Table 2).

Eigenvalue criterion 2: 70% of the total variance is explained by the **first 3 principal components** (see Figure 2).

Table 2: Eigen values estimated by the PCA for the 15th first principal components (PC) for the geochemical dataset

	eigenvalue	percentage of variance [%]	cumulative percentage of variance [%]
PC 1	13.23	42.66	42.66
PC 2	3.72	12.01	54.67
PC 3	3.08	9.94	64.61
PC 4	2.22	7.17	71.78
PC 5	1.37	4.43	76.22
PC 6	1.21	3.91	80.13
PC 7	1.03	3.32	83.45
PC 8	0.86	2.78	86.24
PC 9	0.69	2.21	88.45
PC 10	0.58	1.86	90.30
PC 11	0.52	1.67	91.98
PC 12	0.49	1.58	93.56
PC 13	0.40	1.28	94.83
PC 14	0.36	1.16	95.99
PC 15	0.25	0.79	96.78

Figure 2 : Cumulative percentage of variance versus the 15th first principal components for the geochemical dataset



Elbow criterion: The strongest slope break is observed between PC4 and PC5. We thus decided to place the elbow between these two PCs. According to this criteria, we will retain the **first 4 principal components**.

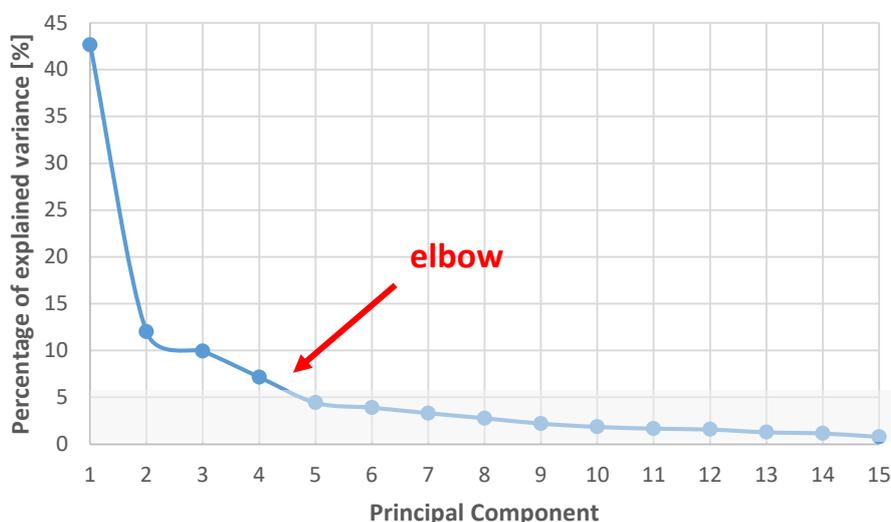


Figure 3: Scree plot of the 15th first PCs for the geochemical data

According to the 3 selection criteria, 3 or 4 principal components are describing the majority of the variability of the dataset. In the next section, we will study the first 4 principal components.

5.1.2 PCA results

5.1.2.1 Contributions of the different chemical component to the PCs

We look at the contributions of the different chemical elements to the first 4 PCs (see Figure 4). The red dashed line on the graphs indicates the expected average contribution. If the contribution of the variables were uniform, the expected value would be $1/\text{length}(\text{variables}) = 1/31 = 3.2\%$. For a given component, a variable with a contribution larger than this cutoff could be considered as important in contributing to the component.

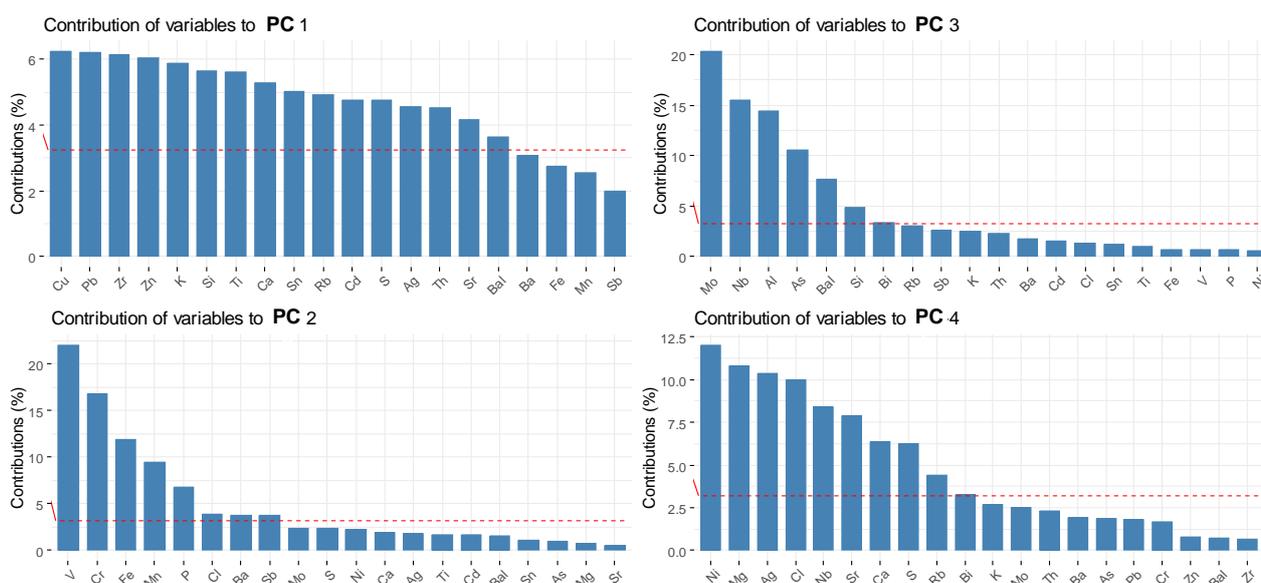


Figure 4: Contribution of the first 20th chemical components for the first 4 PCs

According to the results in Figure 4, the main contributions for each PC are:

- **PC1:** Cu, Pb, Zr, K, Si, Ti, Ca, Sn, Rb, Cd, S, Ag, Th, Sr and Bal
- **PC2:** V, Cr, Fe, Mn, P, Cl, Ba, Sb
- **PC3:** Mo, Nb, Al, As, Bal, Si, Bi
- **PC4:** Ni, Mg, Ag, Cl, Nb, Sr, Ca, S, Rb, Bi

These results are only taking into account the dynamics of the variations for each chemical component, and not their concentrations that might be small for some (e.g. Mo, Nb, V...).

5.1.2.2 PCA results for variables (chemical elements)

The term **variable** refers here to the different chemical components that are studied, while the **observations or individuals** refers to the results obtained for each sample measured.

The correlation between a variable and a principal component (PC) is used as the coordinates of the variable on the principal component. The representation of variables differs from that of observations (also called individuals): observations are represented by their projections, but variables are represented by their correlations (Abdi and Williams, 2010).

The graphs presented in Figure 5 are known as the **variable correlation graphs**. They show the relationships between all the variables and the PCs studied (PC1 and PC2 for Figure 5 a); and PC3 and PC4 for Figure 5 b)). It can be interpreted as follows:

- Positively correlated variables are grouped together.
- Negatively correlated variables are positioned on opposite sides of the origin of the graph (opposing quadrants).
- The distance between the variables and the origin measures the quality of representation of the variables. Variables that are far from the origin are well represented by the studied PCs.

The contributions of the variables in the definition of a given main axis are expressed as a percentage (also see Figure 4).

- The variables correlated with PC1 (i.e., Dim.1) and PC2 (i.e., Dim.2) are the most important to explain the variability in the dataset.
- Variables that are not correlated with an axis or that are correlated with the last axes are low-input variables and can be removed to simplify the overall analysis.

Figure 5 is another way of representing the same results than Figure 4, but for two different components on the same graph, and in a much more graphic way.

Regarding the contribution to PC1 and PC2, several variables are contributing equally to the PCs. Two main groups can be identified:

- **Positively correlated variables:** Cu, Pb, Ca, Sn, Cd, S, Ag; V, Fe, Mn, Cr
- **Negatively correlated variables:** Zr, K, Si, Ti, Rb

Regarding the contribution to PC3 and PC4, less variables are contributing to the two PCs. The two main negatively correlated variables are Mo and Nb.

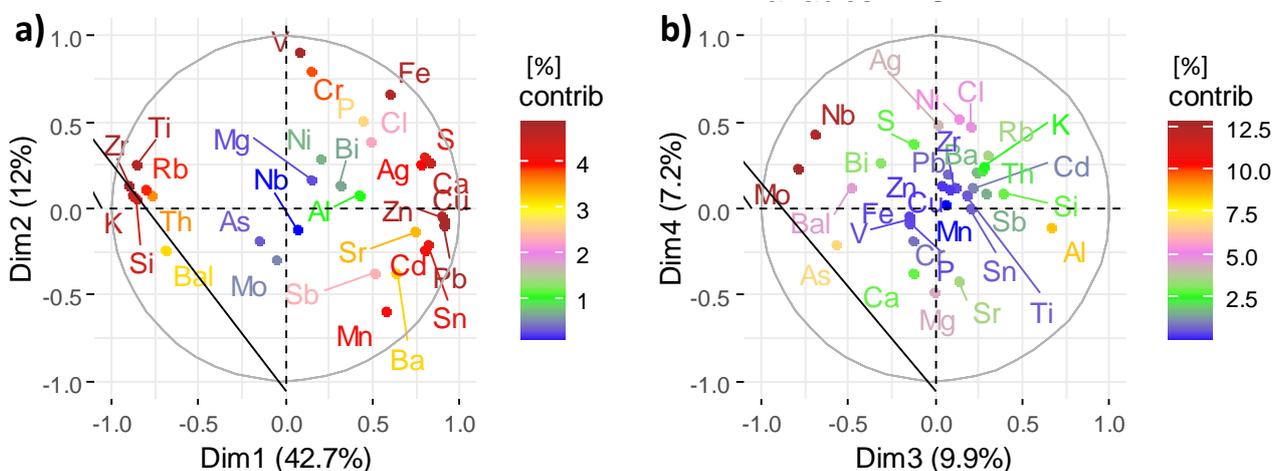


Figure 5: Variable correlation graphs between::a) PC1 and PC2; b) PC3 and PC4, as a function of the contribution of each variables to PC1 and PC2, or PC3 and PC4 respectively. The PCs are represented with their respective percentage of variance (see Table 2), e.g. PC1 represent 42.7% of the variance. All the variables are represented as a function of their contribution (contrib in %) to PC1 and PC2, or PC3 and PC4.

Groups can be sorted out when looking at the correlations between the variables in all the 31 principal components. We used the k-means classification algorithm to classify variables into groups using the PCA results on all the 31st PCs estimated. The clusters returned by the k-means algorithm are then used to color the variables in the variable correlation graphs. We decided to fix the number of groups estimated with the k-means algorithm following 3 different observations:

- **2 groups (see Figure 6 a)):** From the results presented in Figure 5 a), two obvious groups can be extracted from the data:
 1. positively correlated variables with PC1;
 2. negatively correlated variables with PC1
- **3 groups (see Figure 6 b)):** the 1st group of Figure 6 a) includes a wide range of chemical elements correlated with each other. It is reasonable to expect layers within the settling pond with less chemical element variety. When using the clustering algorithm for 3 groups, cluster 1 gets splitted into 2 distinct groups:
 1. positively correlated variable with PC2;
 2. negatively correlated variables with PC2
- **5 groups (see Figure 6 c)):** According to the geophysical interpretation (see DI2.2.2), 4 to 5 different layers are identified within the entire settling pond. We thus tried to split the results into 5 different groups to see if some links could be made between the geophysical interpretation and the PCA results on the geochemical samples.

These results will be applied on the individuals analysis (see section 5.1.2.3).

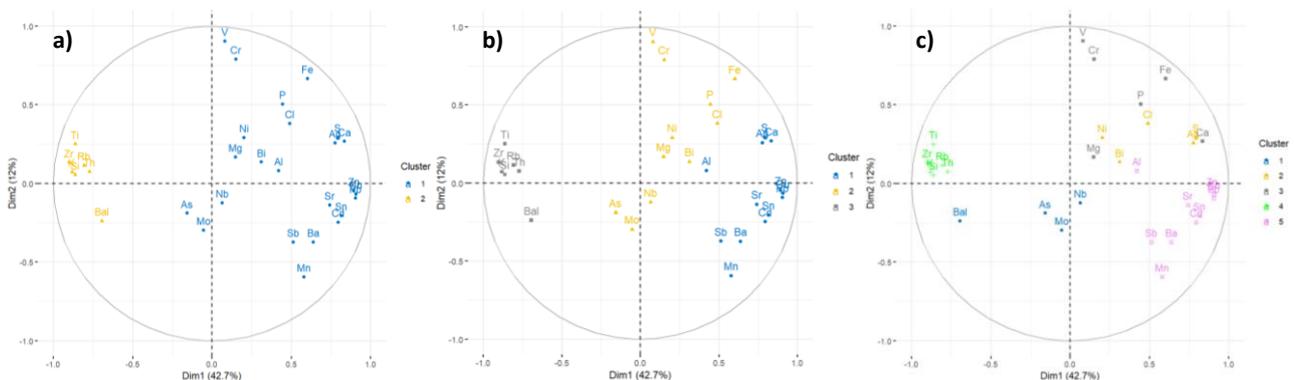


Figure 6: Variable correlation graphs between:PC1 and PC2, sorted out using a cluster analysis in a) 2 groups; b) 3 groups; c) 5 groups. The k-means cluster analysis is taking into account the correlations between the variables and all the 31 PCs estimated. The distinction between group 2 and 3 in the 5 groups-analysis can thus not be observed on the PC1 vs PC2 graph, but is clearer on other dimensions.

5.1.2.3 PCA results for individuals (samples collected)

We then study the correlations observed between the PCs and the different individuals, i.e. between the different samples collected at various locations and depths. The individuals are represented by their projections on the different PCs.

We chose to color the individuals by altitude (Figure 7) and by borehole number (Figure 8) to try to distinguish spatial and vertical variations of chemical composition within the settling pond.

Link with altitude

Figure 7 presents the individuals correlation graphs colored by altitude for PC1 vs PC2 (Figure 7 a)) and for PC3 vs PC4 (Figure 7 a)).

For PC1 vs PC2, a clear distinction can be made between:

- **Altitudes ≥ 191 m:** most of the points are positively correlated with PC1. The highest altitudes are also negatively correlated with PC2(group B).
- **Altitudes ≤ 191 m:** most of the points are strongly negatively correlated with PC1 (group A)

For PC3 vs PC4, no clear trend can be observed versus altitude. We observe that group A previously identified is very weakly correlated to PC3 and PC4.

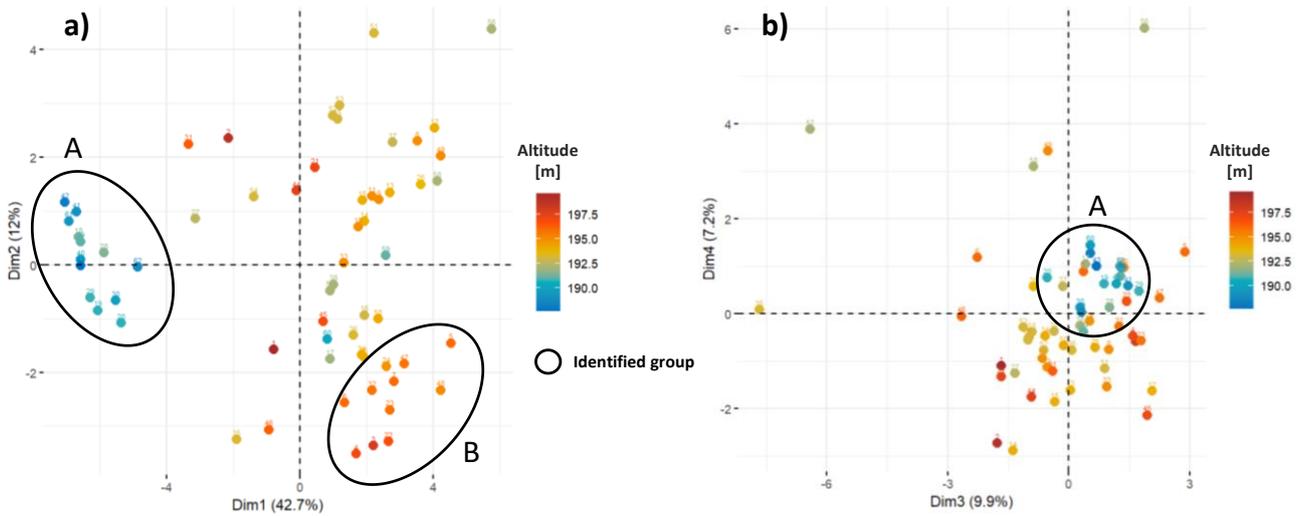


Figure 7: Individuals correlation graphs between: a) PC1 vs PC2 ; b) PC3 vs PC4 for individuals colored by altitude (in m). The PCs are represented with their respective percentage of variance (see Table 2), e.g. PC1 represent 42.7% of the variance. The number associated for each individual (on top of each point) is the position in the result table of the PCA analysis (see Appendix 1).

Link with location of the borehole

Figure 8 presents the individuals correlation graphs colored by borehole number for PC1 vs PC2 (Figure 8 a)) and for PC3 vs PC4 (Figure 8 a)). We observe that there are no clear distinctions between the different borehole locations. For each of the groups previously identified, the 4 boreholes are represented. It seems that there is no clear distinction between the 4 boreholes in terms of chemical composition.

Based on these observations, the settling pound main variations in terms of chemical composition are vertical and not lateral.

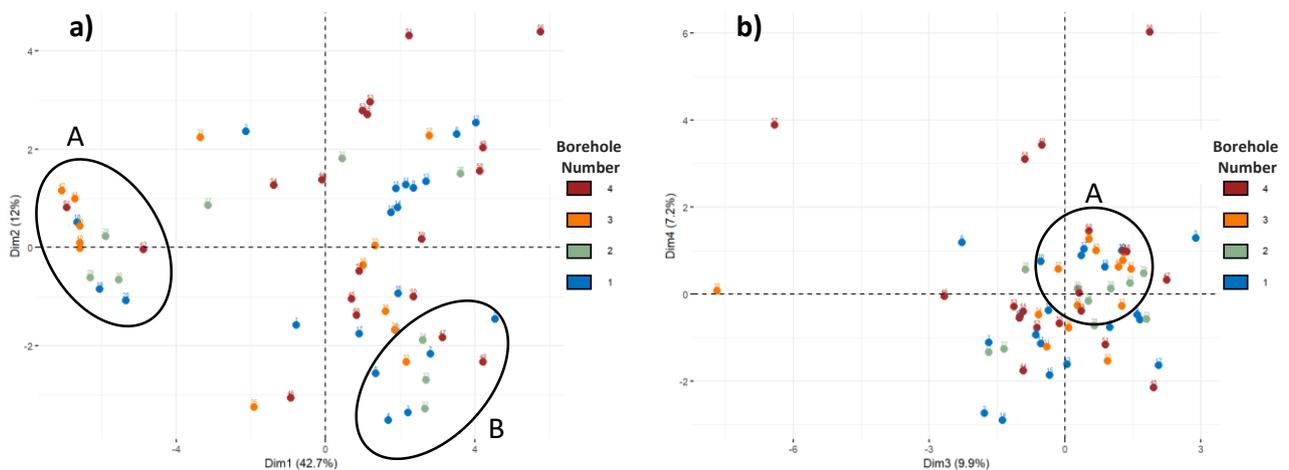


Figure 8 : a) PC1 vs PC2 ; b) PC3 vs PC4 for individuals colored with borehole number scale (FP1 in blue and FP4 in red). The PCs are represented with their respective percentage of variance (see Table 2), e.g. PC1 represent 42.7% of the variance. The number associated for each individual (on top of each point) is the position in the result table of the PCA analysis (see Appendix 1).

Comparison with variable classifications

To go further in the PCA analysis, we used a bi-plot to represent on one graph the PCA analysis on PC1 vs PC2 for both variables (points in grey) and individuals (points colored by altitude), combined with the cluster analysis shown in 5.1.2.2. The 5 groups separation used in one the clustering analysis is not used here since there could not be found much correlations between the variables and individuals groups.

In Figure 10, the 2-groups clustering of variables is compared to the individual analysis, colored by altitudes.

We note that group A of samples with altitudes lower than 191 m corresponds well with cluster n°2. Cluster n°2 corresponds to the negatively correlated chemical elements with PC1: Zr, K, Si, Ti, Rb. This group of samples has a distinct chemical composition than the rest of the samples for the 4 borehole locations. It could correspond to the natural alluvia layer on which the settling pond was built.

Most of the rest of the samples, along with group B, are included within cluster n°1. They correspond to the layers of techno-soil above 191 m. The variables included within cluster n°1 are too many to be able to draw conclusions in terms of chemical composition of the techno-soils.

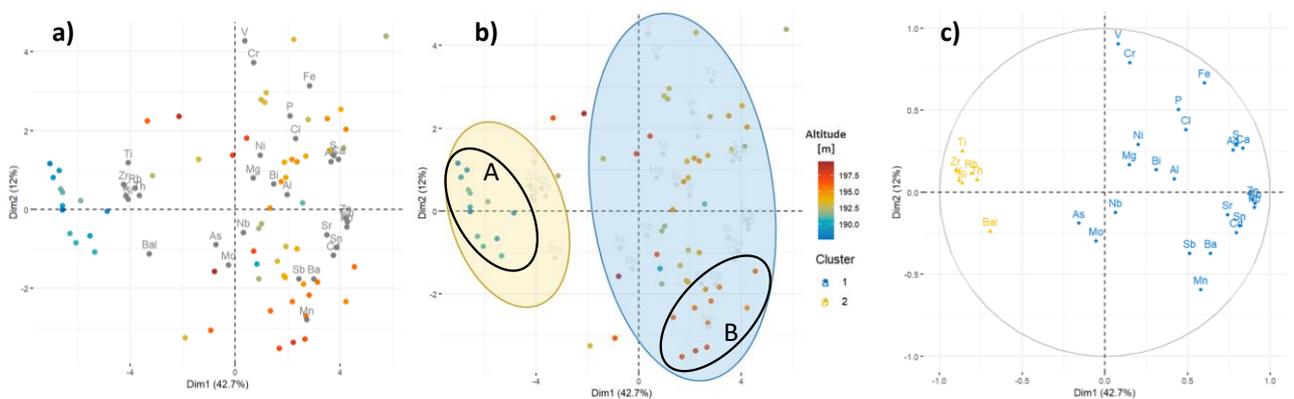


Figure 9: a) Bi-plot representing on one graph the PCA analysis on PC1 vs PC2 for both variables (points in grey) and individuals (points colored by altitude); b) Same bi-plot with individual groups identified (A and B) and cluster analysis for variables; c) Variable correlation graphs between:PC1 and PC2, sorted out using a cluster analysis for 2 groups.

In Figure 10, the 3-groups clustering of variables is compared to the individual analysis, colored by altitudes. Several observations can be made based on this analysis:

- **Cluster n°1** intersect with the group B identified in the PCA results for individuals. Cluster n°1 includes chemical elements such as Zn, Pb, Cu, Mn, as well as Ca and Al. They intersect with group B of the samples that have the highest altitudes. These techno-soil layers were placed within the pond the latest, or even after the closure of the pond.
- **Cluster n°2** corresponds mostly to samples with medium altitudes, at the center of the settling pond. Cluster n°2 includes chemical elements such as Fe, Cl; Mg and As. With the 3-group clustering, we note that Fe and Mn are not placed in the same groups, indicating that they are probably not present in their alloy forms.
- **Cluster n°3** corresponds to cluster n°2 of the 2-groups clustering. This group of samples has a distinct chemical composition than the rest of the samples for the 4 borehole locations. It could correspond to the natural alluvia layer on which the settling pond was built.

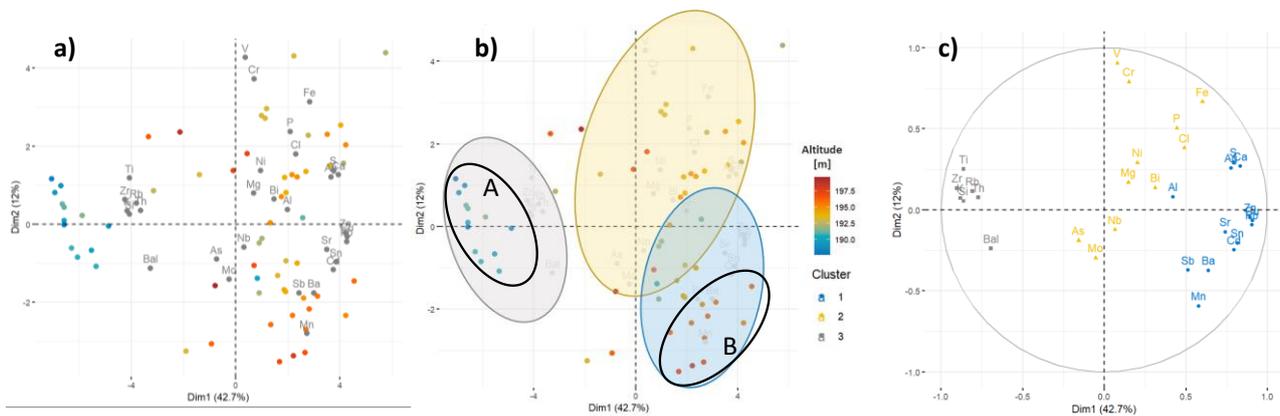


Figure 10: a) Bi-plot representing on one graph the PCA analysis on PC1 vs PC2 for both variables (points in grey) and individuals (points colored by altitude); b) Same bi-plot with individual groups identified (A and B) and cluster analysis for variables; c) Variable correlation graphs between:PC1 and PC2, sorted out using a cluster analysis for 3 groups.

The 3-group clustering analysis is in good coherence with the PCA results for individual. We note some coherence between the chemical element clusters identified and the altitude of the samples. 3 layers could thus be identified within the pond based on the PCA analysis. These results could be used for the establishment of the RAPIDM model for the site

5.2 PEARSON'S CORRELATION COEFFICIENT MATRIX ANALYSIS

In order to compare the correlation analysis between the different pilote site, we also run a Pearson's correlation coefficient matrix analysis.

Figure 11 is showing the Pearson's correlation coefficient matrix for the 31 chemical elements studied. In the correlation matrix the strong and positive relations are presented in blue ([1;0.8] and [0.8;0.6]) while the strong and negative relations are shown in beige ([-1;-0.8] and [-0.8;-0.6]). Interesting relationships can be observed 2 by 2.

Positive correlations:

- Zn-Pb-Cu and S-Ag and Cd-Sn-Ba. All these elements are part of the cluster n°1 identified using the PCA analysis
- Mn is only strongly correlated with Ba
- Fe is not very strongly correlated with any other element, except Ag
- Si-K-Ti-Rb-Th-Zr. All these elements are part of the cluster n°3 identified using the PCA analysis

Negative correlations:

- Elements belonging to separate clusters such as Zr and Cu with Cu, Pb and Zn correlate negatively

Using Pearson's correlation coefficient matrix highlights the same correlations between chemical elements than the ones identified using PCA analysis. As the correlation is only 2-by-2, it is a weaker tool to distinguish clusters of chemical elements that are bound together.

6 CORRELATION BETWEEN CHEMICAL AND GEOPHYSICAL DATASETS

6.1 PRINCIPAL COMPONENT ANALYSIS

6.1.1 Selection of the number of PCs to keep

After calculating the eigenvalues for the 34 principal components, corresponding to the 31 chemical elements, plus the three geophysical parameters (resistivity: rho; chargeability: phi or M; metal factor: MF) that are evaluated, the different selection criteria were applied (see section 4.1 and below). The results for the first 15 PCAs are presented in Table 3.

Kaiser's criterion: we decided to keep only the PCs whose eigenvalues are greater than 2. In our case, we will retain the **first 5 principal components** (see Table 3).

Eigenvalue criterion 2: 70% of the total variance is explained by the **first 4 principal components** (see Figure 12).

According to the 2 selection criteria used above, 4 or 5 principal components are describing the majority of the variability of the dataset. In the next section, we will study the first 5 principal components.

Principal Component	eigenvalue	percentage of variance [%]	cumulative percentage of variance [%]
PC 1	13.23	38.91	38.91
PC 2	3.76	11.05	49.96
PC 3	3.09	9.09	59.05
PC 4	2.47	7.27	66.32
PC 5	2.21	6.50	72.82
PC 6	1.37	4.03	76.85
PC 7	1.37	4.03	80.88
PC 8	1.10	3.24	84.12
PC 9	0.92	2.72	86.84
PC 10	0.67	1.97	88.81
PC 11	0.60	1.77	90.58
PC 12	0.56	1.64	92.22
PC 13	0.45	1.32	93.54
PC 14	0.43	1.25	94.79
PC 15	0.35	1.03	95.82

Table 3 : Eigen values estimated by the PCA for the 15th first principal components (PC) for the geochemical dataset

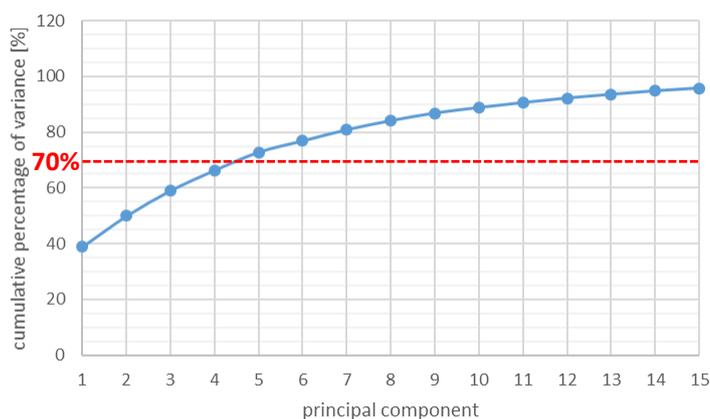


Figure 12 : Cumulative percentage of variance versus the 15th first principal components for the geochemical dataset

6.1.2 PCA results

6.1.2.1 Contributions of the different variables to main PCs

We look at the contributions of the different chemical elements to the first 5 PCs (see Figure 13Figure 4). The red dashed line on the graphs indicates the expected average contribution. If the contribution of the variables were uniform, the expected value would be $1/\text{length}(\text{variables}) = 1/34 = 2.9\%$. For a given component, a variable with a contribution larger than this cutoff could be considered as important in contributing to the component.

The results are very similar to the results with only the 31 chemical elements (see section 5.1.2.1) for the first 3 principal components. They change only for PC4, where the contribution of rho (resistivity) is dominant and for PC5, where the contribution of MF (metal factor) and phi (chargeability) are dominant.

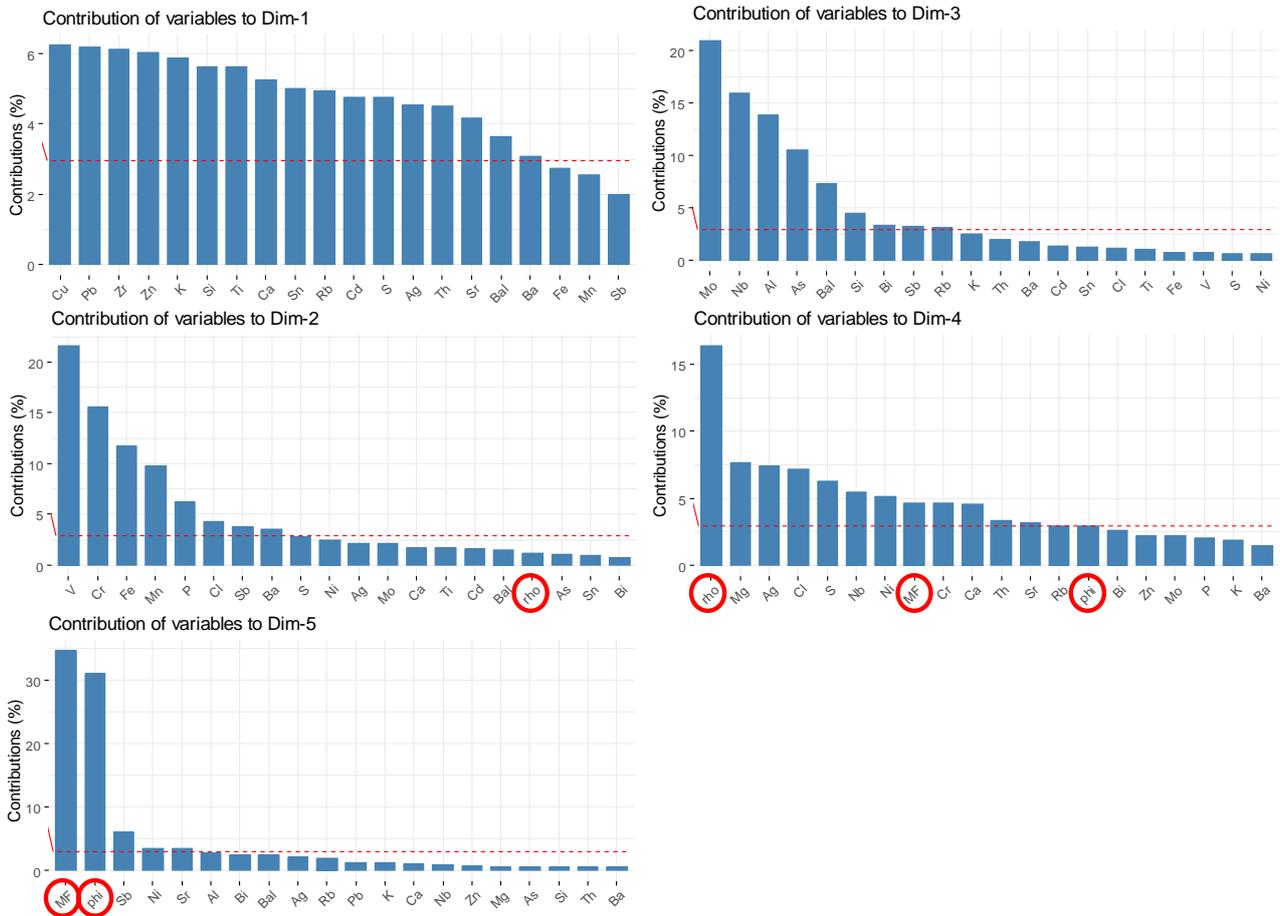


Figure 13 : Contribution of the first 20th chemical components for the first 5 PCs (also named Dim here). The geophysical parameters are highlighted in red (rho: electrical resistivity; phi: chargeability; MF: metal factor).

6.1.2.2 Change in PCA results for variables

The new variable correlation graphs are shown in Figure 14. The same observations than in 5.1.2.2 can be made for the first two components. The contributions of the geophysical variables are small for PC1 and PC2. For PC3, PC4 and PC5, the following observations can be made for each geophysical parameters:

- **Rho – electrical resistivity:**

The contribution of rho to PC3 and PC4 is above average (7%). It is negatively correlated with PC4. It is isolated compared to the other chemical elements. No cluster can be intuitively built around rho.

The contribution of rho to PC4 and PC5 close to 11%. It is negatively correlated with PC4. It is isolated compared to the other chemical elements. No cluster can be intuitively built around rho.

- **M – chargeability and MF – metal factor:**

M and MF are closely correlated with each other. The 2 variables are located very close to each other in the variable correlation graphs.

The contributions of M and MF to PC3 and PC4 are very low, below the threshold level estimated in section 6.1.2.1 (2.9%)

The contributions of M and MF to PC4 and PC5 are above 15%. They are positively correlated with PC5. The 2 points are isolated compared to the other chemical elements. No cluster can be intuitively built around M and MF.

between the chemical elements and the geophysical parameters. The geophysical parameters can thus not be correlated with specific chemical elements. The cluster analysis studied for the chemical components alone will be used to separate the different layers observed in the field when the depth is varying.

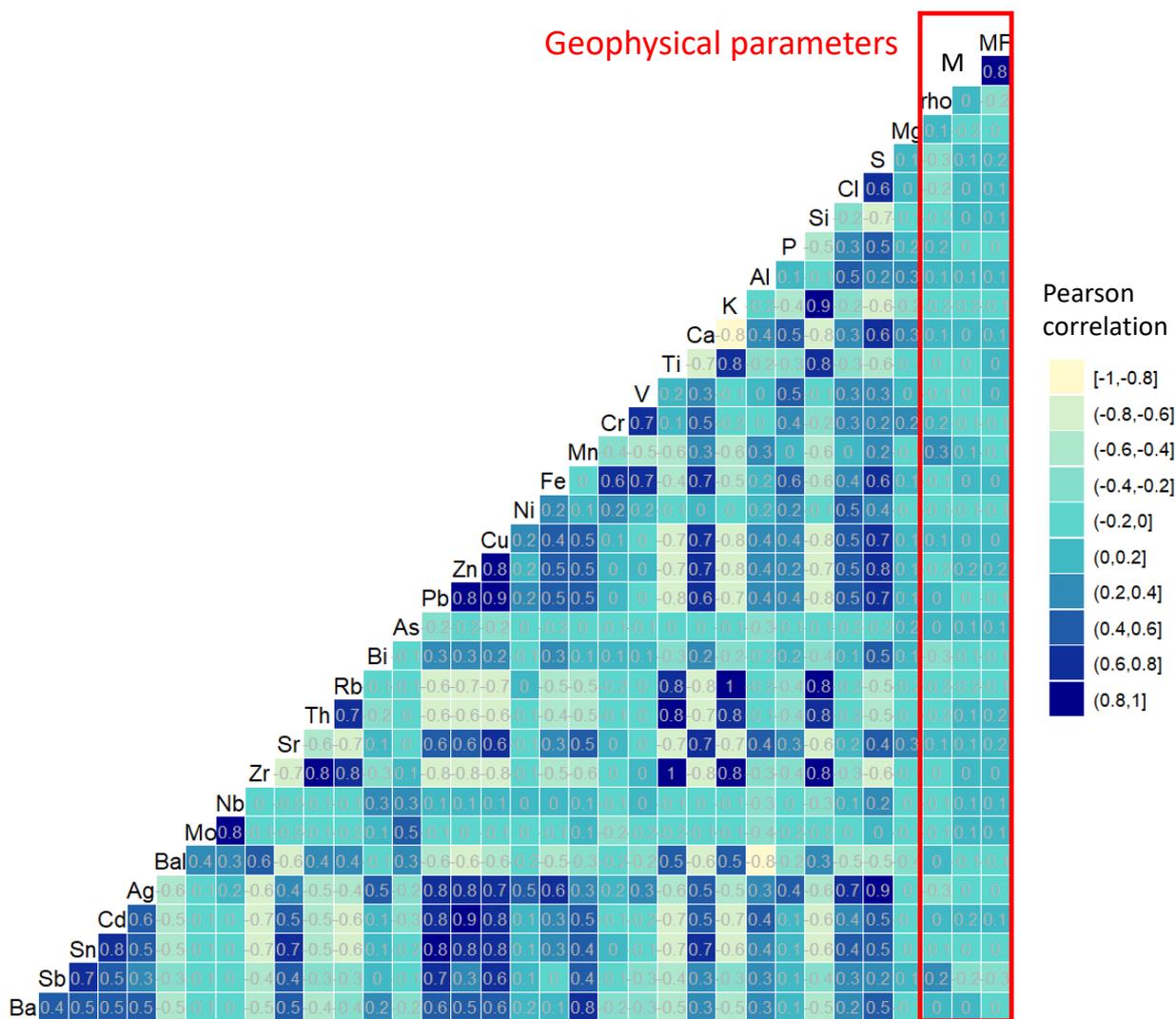


Figure 15: Pearson's correlation coefficient matrix for the 31 chemical elements studied and the 3 geophysical parameters: resistivity (rho), chargeability (M), metal factor (MF). The red square highlights the correlations with the 3 geophysical parameters.

6.3 CLUSTER SELECTION FOR THE RAPDIM MODEL

We will use the results of the PCA on the chemical components (see part 5.1.2.3) to separate the datasets into clusters linked with different altitudes. This analysis will help determine various layers with different chemical composition that will be used for the RAPDIM analysis.

We chose to separate the dataset into 4 clusters (see Figure 17). We use 2 clusters extracted from the PCA analysis of the chemical elements (cluster n°1 and cluster n°3) (see Figure 16), and that also corresponds to 2 different altitude ranges. We add two other clusters (cluster n°2 and cluster n°4) in altitudes in between cluster n°1 and cluster n°3.

From the chemical PCA analysis, we can identify the chemical components that are dominant for each cluster (see Figure 16):

- **Cluster n°1:** corresponds to group A of the individual analysis (see section 5.1.2.3), and partially to cluster n°1 of the variable analysis (see section 5.1.2.2). It is mostly correlated with chemical elements such as **Zn, Cu and Pb, as well as Mn**. These elements are cited because they have the highest concentrations of the group.
- **Cluster n°3:** corresponds to group B of the individual analysis (see section 5.1.2.3), and partially to cluster n°3 of the variable analysis (see section 5.1.2.2). It is mostly correlated with chemical elements such as **Si and K**. These elements are cited because they have the highest concentrations of the group.
- **Cluster n°2:** is not as constrained within the bi-plot of PC1 vs PC2. Most of the individual points are located in the positive dial for PC1 and PC2. They can thus be associated with chemical elements of cluster n°2 identified with the variable PCA analysis. One of the main chemical element contributing to this cluster is **Fe**.
- **Cluster n°4:** is scattered in the bi-plot. No conclusions can be drawn in terms of chemical composition for this cluster.

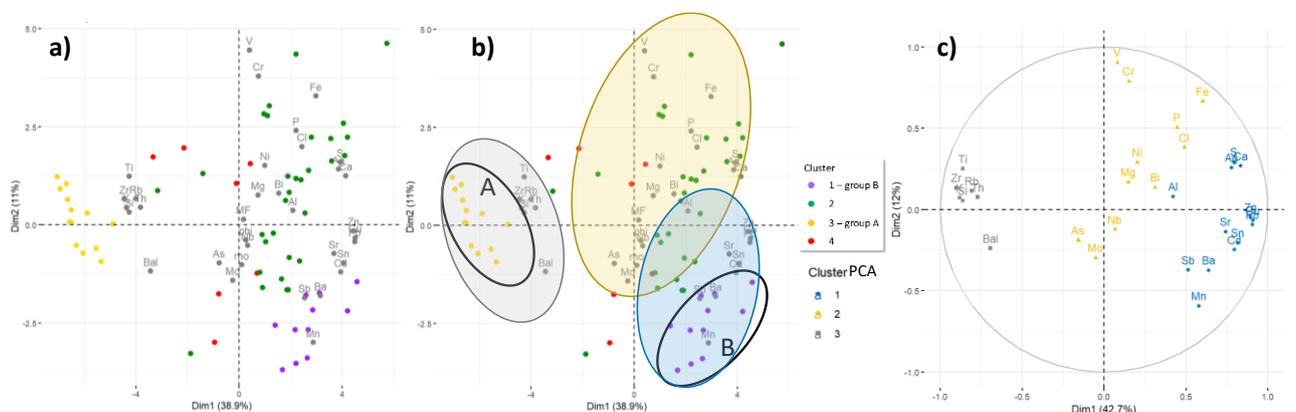


Figure 16: a) Bi-plot representing on one graph the chemical PCA analysis on PC1 vs PC2 for both variables (points in grey) and individuals (points colored by selected clusters); b) Same bi-plot with individual groups identified (A and B) and cluster analysis for variables (cluster PCA – see section 5.1.2.3); c) Variable correlation graphs between:PC1 and PC2, sorted out using a cluster analysis (cluster PCA) for 3 groups on the chemical PCA analysis.

In terms of geophysical parameters variations, the identified clusters present the following characteristics, from the highest to the lowest altitudes (also see Table 4):

- **From 200 to 196 m of altitude:** cluster n°4 corresponds to the highest resistivities of the profile (mostly between 200 and 600 ohm.m), average chargeabilities (mostly between 7 and 20 mV/V) and low metal factor (mostly between 30 and 90 (ohm.m)⁻¹). Most of the points are extracted from borehole FP1, where a mound could be observed at the surface. This cluster could thus correspond to a layer of wastes that were placed here after the closure of the settling pond, with unknown origin. Its composition can thus be very variable, and explains the scattering of the points within the bi-plot results (see Figure 16)
- **From 198 to 194 m of altitude:** cluster n°1 represents a transition zone for the geophysical parameters variations. The resistivity decreases from 300 ohm.m to 30 ohm.m. The chargeability is very scattered, ranging from 1 to 100 mV/V. The metal factor increases above 100 (ohm.m)⁻¹. This zone is mostly correlated with chemical elements such as Zn, Cu and Pb, as well as Mn, and not as much with Fe. It could correspond to a period when the processes and/or the raw components used in the plant productions changed, using more alloys with higher concentrations of Zn, Cu, Pb and Mn. The historical data study indicates

several changes in the processes used to produce steel, and in the suppliers of the iron-ores used (see BRGM, 2022).

- **From 195 to 191 m of altitude:** cluster n°2 represent a zone where: (1) the resistivity is low, ranging between 10 and 100 ohm.m; (2) the chargeability is scattered between 1 and 100 mV/V; (3) the metal factor is high, ranging from 200 to 2000 (ohm.m)⁻¹. This layer corresponds to the base of the settling pond. Fe is one of the major chemical elements contributing to the chemical composition of the layer.
- **From 191 to 187 m:** cluster n°3 has a very similar response than cluster n°2. We can observe that: (1) the resistivity is increasing slightly; (2) the chargeability is decreasing at least for FP1 and FP3; (3) the metal factor is slightly decreasing. We don't observe a major change for the geophysical parameters. It is the case for the chemical compounds. This cluster has major contributions of very different chemical elements, e.g. Si and K, compared to the other clusters. It is interpreted as the first layer of natural geological formation. This interpretation is consistent with the drillers observations that stopped drilling when the samples changed visually. For the geophysical parameters, the sharp transition is at lower altitudes (see DI2.2.2). This variation can be explained by the fact that the chemical composition analysis only focuses on the solid grains, while the geophysical parameters also take into account the changes in the electrolyte. Electrolyte with higher ionic strength at the bottom of the pond could explain the geophysical observations. The geophysical data will thus not allow to clearly distinguish the physical bottom of the settling pond in terms of solid particles.

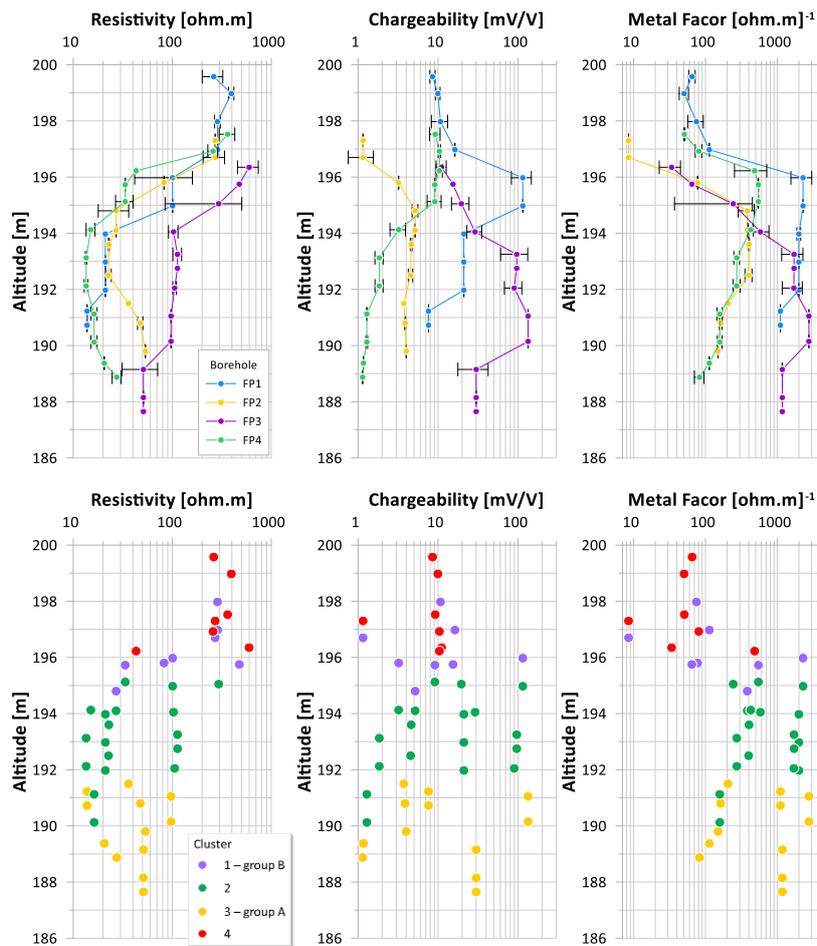


Figure 17: Geophysical parameters plotted versus altitude for each of the 4 boreholes FP1 to FP4 (3 graphs on top), and versus the 4 identified clusters (3 graphs at the bottom)

Table 4: Summary of the observations made for each layers of materials, based on the cluster selection

Altitudes [m]	Group n°	Chemical composition	Geophysical parameter variations	Interpretation
195 – 193.3	4	scattered	<ul style="list-style-type: none"> - High rho - Average M - Low MF 	<p>Anthropic wastes placed after the closure of the settling pond</p> <p>Present mostly for FP1 and FP4</p>
193.9 – 191.4	1	Main contributions: Zn, Cu and Pb, Mn	<p>Transition zone:</p> <ul style="list-style-type: none"> - Decrease of rho - Scattered M - Increase of MF 	Settling pond layer n°1
193.2 – 187.4	2	Main contribution: Fe	<ul style="list-style-type: none"> - Low rho - Scattered M - High MF 	Settling pond layer n°2
<p>Very variable limit:</p> <p>~186.5 for FP1 and FP4</p> <p>~189 for FP2 and FP3</p>	3	Main contributions: Si and K	<ul style="list-style-type: none"> - Low rho (slightly increasing) - Scattered M (decreasing) - High MF (slightly decreasing) 	Natural alluvia with high ionic strength electrolyte?

7 CONCLUSIONS

Two different statistical analysis were used to identify correlations between: (1) the chemical elements, and (2) the chemical elements and the geophysical parameters.

The Pearson's correlation coefficient analysis allows correlating variables 2-by-2. It is interesting because it shows positive and negative correlations. It is not very well suited to analyze datasets with a lot of different variables. Here we have 31 variables if we only consider the chemical elements analyzed, and 34 variables if we add the 3 geophysical parameters.

The principal components analysis (PCA) presents the advantage of reducing the number of dimensions considered in the correlation analysis. It allows identifying clusters both for the variables (=chemical elements and geophysical parameters) projected on the new calculated dimensions (principal components or PCs), and for the individuals (=samples).

Using the PCA on the 31 chemical elements studied, we were able to identify cluster of variables correlated with each other. We splitted the variables into 3 different clusters depending on their contributions to the calculated PCs. We then linked these clusters with groups identified within the individual analysis: group A corresponding to samples taken at the lowest altitudes, and group B corresponding to samples taken at medium altitudes, within the settling pond.

Using PCA to link the geophysical parameters to specific chemical elements was not successful. Indeed, the geophysical parameters contributed to independent PCs. We thus decided to use the PCA analysis on the chemical elements alone to distinguish layers with various compositions, and link these layers to geophysical parameters variations. The results are presented in Table 4. The results of these correlation analysis will be used to build the RAPIDM model for the Pompey site.

8 REFERENCES

Papers, books and reports

- Abdi, H. and Williams, L.J., 2010. Tukey's honestly significant difference (HSD) test. *Encyclopedia of research design*, 3(1), pp.1-5.
- ANTEA, 2001a. Site du Ban-la-Dame à Frouard : Etude technique pour la mise en place d'une couverture étanche et pour l'aménagement paysager, octobre 2001 – A24803/A
- ANTEA, 2001b. Site du Ban-la-Dame à Custines-Frouard : Diagnostic du bassin de décantation Sud, novembre 2001 – A25070/A
- ANTEA, 2002a. Projet d'extension de Délipapier sur le site de Ban-la-Dame à Frouard : Reconnaissance de l'extension de la limite Sud de l'ancien bassin de décantation, février 2002 – A26295/A
- ANTEA, 2002b. Investigations complémentaires sur le site du Ban-la-Dame à Custines-Frouard, octobre 2002 – A27912/A
- ANTEA, 2003. Site du Ban-la-Dame à Custines-Frouard : Evaluation Simplifiée des Risques, mars 2003 – A29674/A
- ANTEA, 2004. Ancien bassin de décantation du Ban-la-Dame à Custines/Frouard : Evaluation Détaillée des Risques pour la Santé : Scénario type promenade, mai 2004 – A32151/B
- ArcelorMittal archives - file 167/092 Moselle pollution
- Berrar, J.C. 2011. Le fer en Lorraine. Ed. Alan Sutton, collection Mémoire en images, Saint-Cyr-sur-Loire, France, 127 p.
- BRGM, Geological map of Nancy 1/50 000, BRGM Publications.
- Davis. 1973. Statistics and data analysis in geology (with Fortran programs by RJ Samson). New York and London (John Wiley and Sons). xiii+ 550 pp.
- Deliverable DT1.3.1. Alamar et al. 2020. Benchmark report on geophysics and non intrusive investigation techniques for quantification of the potential resources located on a PMSD, 29 p.
- Dempster, M., Dunlop, P., Scheib, A. and Cooper, M., 2013. Principal component analysis of the geochemistry of soil developed on till in Northern Ireland. *Journal of Maps*, 9(3), pp.373-389.
- FONDASOL, Juin 1993. Pôle d'Activités de Pompey – Frouard – Custines - Ban la Dame à FROUARD – Elargissement du chemin d'accès à la sous station EDF 63 :20 kv - Rapport d'Etude - MN ; 93/089.
- Geindre, L. 1966. Pompey sous l'Avant-Garde. 80, impr. Fetzer Raon-l'Etape, France, 112 p.
- Gerber, G. 2005. Pompey. Ed. Alan Sutton. Mémoires en images. 127 p.
- GISFI, 2014. Note concernant les activités de recherche menées sur le bassin de décantation sud (Frouard) – GISFI – Février 2014
- Grunsky, E.C. and Smee, B.W., 1999. The differentiation of soil types and mineralization from multi-element geochemistry using multivariate methods and digital topography. *Journal of Geochemical Exploration*, 67(1-3), pp.287-299.
- Günther, T., 2021a. python/pybert/manager/ertmanager.py · master · resistivity-net / BERT · GitLab. In : *Gitlab* [en ligne]. [Consulté le 21 juillet 2021]. Available at : <https://gitlab.com/resistivity-net/bert-/blob/master/python/pybert/manager/ertmanager.py>.
- Günther, T., 2021b. python/pybert/tdip/mipmodelling.py · master · resistivity-net / BERT · GitLab. In : *GitLab* [en ligne]. [Consulté le 21 juillet 2021]. Available at : <https://gitlab.com/resistivity-net/bert-/blob/master/python/pybert/tdip/mipmodelling.py>.

Huot, H., 2013. Formation, fonctionnement et évolution d'un Technosol sur des boues sidérurgiques (Doctoral dissertation, Université de Lorraine).

IRSID. 1984. « Analyse du contexte géologique et hydrogéologique des zones des dépôts de déchets sidérurgiques de l'Usine de Pompey - étude de l'influence des dépôts sur les eaux superficielles et souterraines ». MPFE-RC 1043.

Jolliffe, I.T., 2002. *Principal component analysis for special types of data* (pp. 338-372). Springer New York.

Parfitt, R. L., et A. D. Wilson. 1985. « Estimation of allophane and halloysite in three sequences of volcanic soils, New Zealand ». *Catena Suppl* 7: 1-8.

Peres-Neto, P.R., Jackson, D.A. and Somers, K.M., 2005. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, 49(4), pp.974-997.

Reimann, C., Filzmoser, P., Garrett, R. and Dutter, R., 2011. *Statistical data analysis explained: applied environmental statistics with R*. John Wiley & Sons.

Rideau, E. 1956. Essor et problèmes d'une région française. Houillères et sidérurgie de Moselle. Les éditions ouvrières, Paris. 244 p.

Rogé, M. 1982. L'Eau et la sidérurgie en Lorraine ». Thèse 3ème cycle Géographie et Aménagement, Nancy.

Schwartz, C., L. Florentin, D. Charpentier, S. Muzika, et J.L. Morel. 2001. « Le pédologue en milieux industriels et urbains. I. Sols d'une friche industrielle. » *Etude et Gestion des Sols* 8 (2): 135-148.

SEMACO, 2011. Parc d'Activités de Nancy Pompey – Dossier de servitudes pour le bassin de décantation sud - Site Ban la Dame (Frouard 54) – SEMACO 21 Nov 2011.

SNAP, 1943 - Espace Archives ArcelorMittal (plan 514-724)).

Truffaut, E. 1989. « Ferromanganèse élaboré au haut fourneau ». *Techniques de l'ingénieur Elaboration et recyclage métaux* (ref. article : M2445).

Truffaut, E. 2004. « La fabrication du ferro-manganèse au haut-fourneau en France, 1875-2003. Naissance, vie et mort d'un procédé industriel. » <http://soleildacier.ouvaton.org/savoir/fr/DNA02-S2-W.pdf>.

Internet websites :

Atelier Mémoire Ouvrière : website of a group of enthusiasts who bring te story of the Carreau minier from Val de Fer at Neuves-Maisons to life, <http://amo.fjep.pagesperso-orange.fr/>

BASOL: database on polluted or potentially polluted sites and soils calling for action by public authorities, for preventive or curative purposes, <http://basol.developpement-durable.gouv.fr/>

Community of communes of the Pompey basin website: <http://www.bassinpompey.fr/>

Géoportail: website of the National Institute of Geographic and Forest Information (IGN), <http://www.geoportail.gouv.fr>

Infoterre : website gathering all available BRGM data: geological maps from 1: 1,000,000 to 1: 50,000, files from the Basement Data Bank and geological logs, maps of natural and industrial risks, data on groundwater, etc., <https://infoterre.brgm.fr/>