



Application des méthodes d'analyses statistiques multivariées à la délimitation des anomalies de Sidi Chennane (Multivariate Statistical Analysis tool for the interpretation of geoelectrical data: application to Sterile Bodies in the Sidi Chennane phosphatic deposit (Morocco))

Khadija Baba^{*}, Lahcen Bahi, Latifa Ouadif, Choukri Cherradi

(3GIE Laboratory, Mohammadia Engineering School, University Mohammed V, Rabat, Morocco)

Received 10 Oct 2013, Revised 15 Mar 2014, Accepted 15 Mar 2014

* Corresponding author: baba.khadija@gmail.com;

Abstract

The geophysical data carried out in Sidi Chennane phosphatic deposit in Oulad Abdoun basin, were subject to statistical analysis: a simple descriptive analysis followed by a multi-statistical analysis: principal component analysis (PCA) and a hierarchical cluster analysis (HCA). Our database consists of 41 Schlumberger vertical soundings with the electrode spacing AB ranges from 6m to 300m. The principal component analysis provided us with important information on the dispersion of electrical soundings (individuals) on the bases of the apparent resistivity values and AB which highlights the geoelectric mapping of study area. This classification is naturally related to geological features, known through the well drillings data in the area. The multi-statistical analysis proved reliable in the interpretation of a large number of electrical surveys to have an overall view of the distribution of apparent resistivity in the whole area.

Keywords: Sidi Chennane, phosphates, anomalies, geoelectric survey, multivariate statistics.

Résumé

Le gisement de Sidi Chennane est l'une des parcelles de phosphate du bassin sédimentaire des Oulad Abdoun. La série phosphatée est formée d'une intercalation régulière de niveaux phosphatés et marno-calcaires. Certains gisements de phosphate dans ce bassin montrent des corps stériles qualifiés de dérangements. Les données de la prospection géophysique effectuée dans le gisement ont fait l'objet d'une analyse statistique : une simple analyse descriptive suivie par une étude statistique multidirectionnelle (l'analyse en composantes principales (ACP) et la classification hiérarchique ascendante (CHA)). Notre base de données est constituée d'une série de 41 sondages électriques verticaux type Schlumberger dont les écartements des électrodes AB varient de 6m à 300m. L'analyse en composantes principales nous a fourni les principales informations sur la dispersion des sondages électriques (Individus) en fonction des différentes valeurs des résistivités apparentes et des AB ce qui a permis de bien mettre en évidence la cartographie géoélectrique de la région d'étude. Cette classification est naturellement liée aux caractéristiques géologiques, connues à travers les données de terrain effectués dans la zone. L'analyse statistique multidirectionnelle s'est révélée fiable quant à l'interprétation d'un nombre important de sondages électriques afin d'avoir une vision globale de la distribution des résistivités apparentes dans toute la zone.

Mots-clés : Sidi Chennane, phosphates, dérangement, prospection géo-électrique, statistique multivariée.

1. Introduction

Le sous-sol marocain recèle plus des trois quarts des réserves mondiales en phosphate. Ces réserves existent dans trois bassins sédimentaires dont le plus important est celui des Ouled Abdoun, au Centre du Maroc. Dans ce bassin, la couverture sédimentaire de structure tabulaire s'échelonne du Cénomaniens au Lutétien [1,2]. La partie supérieure de cette couverture correspond à la série phosphatée, formée d'une succession de niveaux phosphatés et d'intercalaires stériles marno-calcaires sur plus de 50m de puissance [3, 4, 5]. Toutefois, en plusieurs endroits, cette succession verticale et régulière de niveaux phosphatés n'existe plus et tous les termes constituant la série phosphatée se trouvent mélangés et transformés pour donner lieu, le plus souvent, à des corps stériles, formés exclusivement de calcaires silicifiés ou, parfois, à des poches remplies de blocs de calcaires noyés dans une matrice argileuse [6,7]. Ces corps, qualifiés de dérangement sont masqués par des terres végétales et du recouvrement quaternaire et ne peuvent donc être cartographiés à partir de la surface.

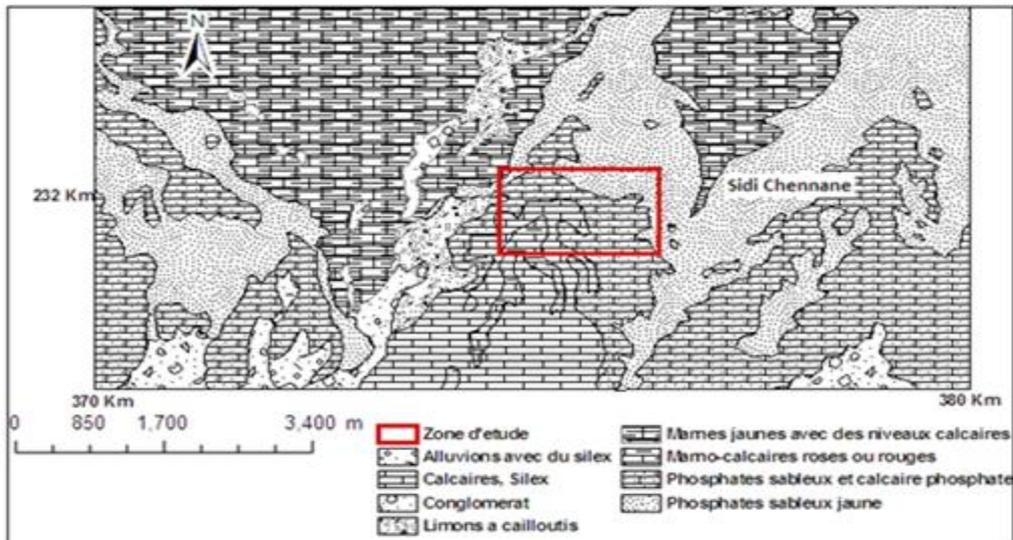


Figure1: situation géographique et géologie de la zone d'étude [8]

L'objectif de cette étude est l'analyse statistiques multidimensionnelles (analyse en composantes principales et classification hiérarchique ascendante) des sondages électriques verticaux afin d'obtenir une synthèse de l'ensemble des données de résistivités apparentes permettant d'avoir une vision globale des caractéristiques électriques moyennes des formations constituant la zone d'étude et en déduire un critère de définition et de délimitation des zones dérangées.

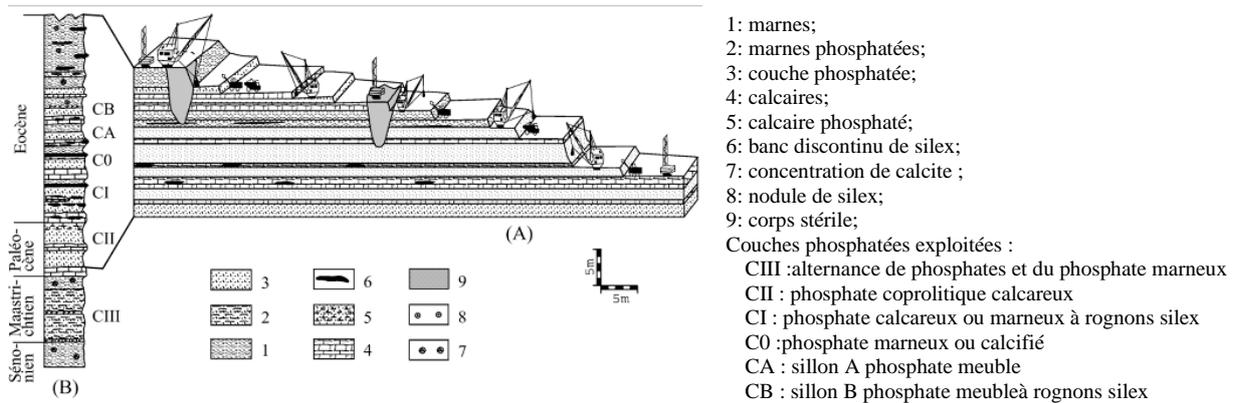


Figure2 : Coupe et log stratigraphique de de la série phosphatée de Sidi Chennane [9].

2. Matériels et méthodes

Au Maroc, la prospection géoélectrique est parmi les techniques les plus utilisées pour la caractérisation des terrains de subsurface [10, 11]. Pour couvrir toutes les zones pouvant être perturbées, nous avons réalisé, au cours de la prospection géophysique sur une parcelle de 50 ha, 41 sondages électriques verticaux en utilisant le Syscal-R2 (IRIS Instruments). Le dispositif utilisé est un dispositif Schlumberger centré de longueur de ligne AB = 6m, AB = 8m, AB = 10m, AB = 14m, AB = 20m, AB = 30m, AB = 40m, AB = 60m, AB = 80m, AB = 100m, AB = 140m, AB = 200m, et AB = 300m, afin de cartographier la distribution spatiale des inclusions stériles.

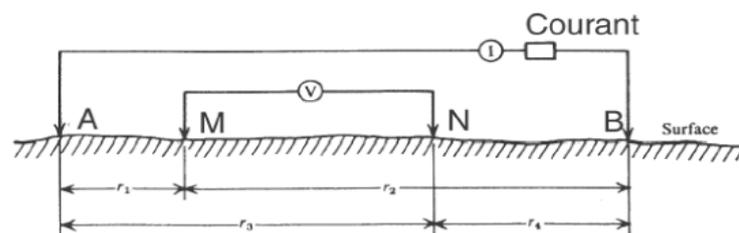


Figure 3: Principe de base du dispositif de Schlumberger

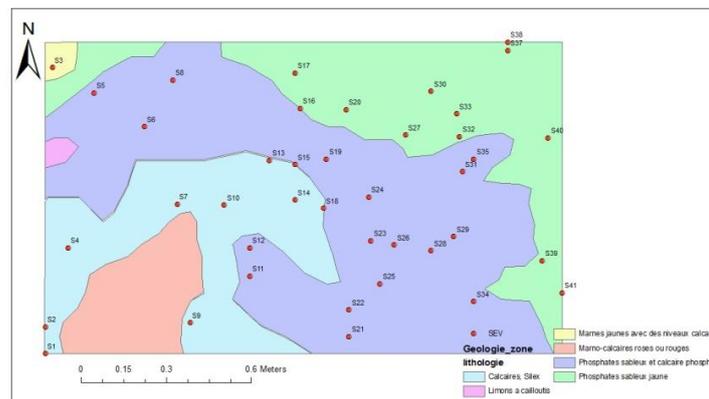


Figure 4: localisation des sondages électriques dans la zone d'étude

La méthodologie adoptée dans cette étude est basée sur une combinaison de l'Analyse en Composantes Principales (ACP) et de la Classification Hiérarchique Ascendante (CHA). Des analyses en composantes principales (ACP) des variables résistivité géoélectrique ont déjà été utilisées avec succès par d'autres auteurs [10, 11, 12, 13].

3. Résultats et discussion

3.1. Analyse en composantes principales des SEV

L'analyse en composantes principales (ACP) consiste en la représentation d'un nuage de points qui correspond à une matrice de données avec N individus et P variables dans un sous espace de l'espace à P dimensions absorbant le maximum de la variance totale du nuage, choisi de façon à optimiser un certain critère. Intuitivement, on cherchera le sous-espace donnant la meilleure visualisation possible de notre nuage de points; ceci nous amène à chercher une rotation du système d'axes initial (les variables) permettant de mieux voir notre nuage [14, 15, 16]. Dans notre cas on aura à projeter le nuage des 41 individus (SEV) dans un espace à 13 dimensions (nombre de variables soit de résistivités apparentes pour chaque distance inter-électrodes d'injection AB).

Pour une meilleure visualisation de nos données, leur analyse consiste principalement à établir les relations existant entre les observations, entre les variables et entre les observations et les variables et les transformer par la suite en nouvelles variables dites composantes principales, les axes engendrés par ces variables sont appelés axes principaux. Ces axes principaux sont ceux de l'ellipsoïde d'une distribution normale multidimensionnelle.

La représentation dans un espace de dimension réduite, permet de mettre en évidence d'éventuelles structures au sein des données. Pour cela, nous recherchons les sous-espaces dans lesquels la projection du nuage déforme le moins possible le nuage initial. Plusieurs critères permettent d'obtenir les composantes, le critère de l'inertie est le plus ancien dont le principe consiste à considérer chaque ligne du tableau de données comme un point dans un espace à P dimensions, pour visualiser le positionnement des individus les uns par rapport aux autres, il faut projeter le nuage de points de l'espace à P dimensions sur un espace de plus faible dimension [16].

La représentation dans l'espace des variables se base donc sur la représentation de chaque ligne de la matrice des données qui correspond aux valeurs des résistivités apparentes pour les différents écartements AB, on obtient ainsi pour chaque sondage, un vecteur dans un espace à 13 dimensions qui a pour origine le centre de gravité (moyenne des valeurs).

Tableau 1 : Coefficients de corrélation des 13 variables des 41 sondages électriques

	AB=6	AB=8	AB=10	AB=14	AB=20	AB=30	AB=40	AB=60	AB=80	AB=100	AB=140	AB=200	AB=300
AB=6	1	0,972	0,936	0,909	0,794	0,686	0,605	0,557	0,458	0,407	0,242	0,217	0,252
AB=8	0,972	1	0,975	0,947	0,853	0,778	0,688	0,621	0,511	0,427	0,201	0,157	0,191
AB=10	0,936	0,975	1	0,973	0,883	0,837	0,755	0,678	0,564	0,465	0,201	0,143	0,199
AB=14	0,909	0,947	0,973	1	0,893	0,875	0,792	0,692	0,558	0,449	0,165	0,092	0,144
AB=20	0,794	0,853	0,883	0,893	1	0,867	0,845	0,753	0,674	0,539	0,189	0,091	0,131
AB=30	0,686	0,778	0,837	0,875	0,867	1	0,965	0,876	0,741	0,575	0,185	0,046	0,063
AB=40	0,605	0,688	0,755	0,792	0,845	0,965	1	0,944	0,851	0,687	0,255	0,101	0,115
AB=60	0,557	0,621	0,678	0,692	0,753	0,876	0,944	1	0,949	0,828	0,448	0,264	0,251

AB=80	0,458	0,511	0,564	0,558	0,674	0,741	0,851	0,949	1	0,927	0,572	0,384	0,355
AB=100	0,407	0,427	0,465	0,449	0,539	0,575	0,687	0,828	0,927	1	0,785	0,623	0,572
AB=140	0,242	0,201	0,201	0,165	0,189	0,185	0,255	0,448	0,572	0,785	1	0,915	0,838
AB=200	0,217	0,157	0,143	0,092	0,091	0,046	0,101	0,264	0,384	0,623	0,915	1	0,965
AB=300	0,252	0,191	0,199	0,144	0,131	0,063	0,115	0,251	0,355	0,572	0,838	0,965	1

Dans cette matrice des corrélations présentée dans le tableau n°1, nous lisons la force des associations linéaires entre les variables prises deux à deux.

La décroissance des coefficients de corrélation avec l'augmentation des valeurs des AB traduit la baisse de l'influence des terrains superficiels sur les valeurs des résistivités apparentes mesurées et montre bien qu'il existe une succession verticale de terrains de nature différente.

Nous obtenons une première composante principale très dominante, expliquant près de 62,34% de la variance totale. Avec seulement les trois premiers axes factoriels, nous avons près de 95,3% de l'inertie (ou de la dispersion) de tout le nuage de points.

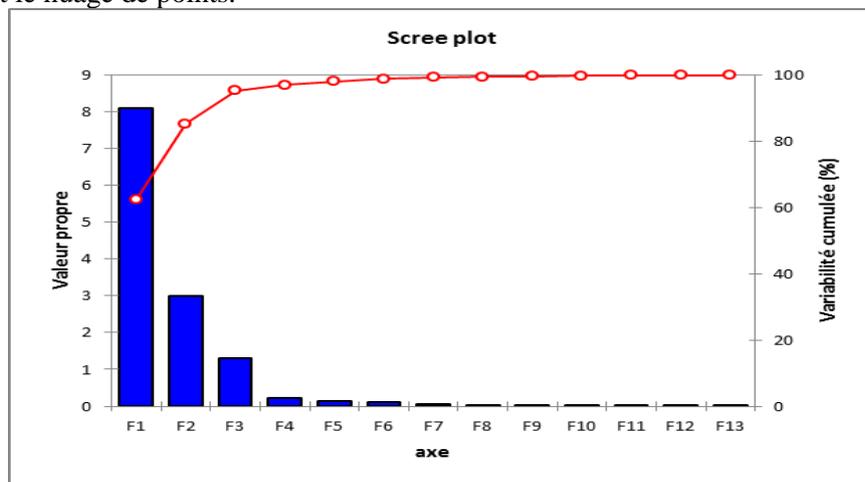


Figure 5: Les valeurs propres de l'analyse

Plan factoriel 1-2 :

Le premier plan factoriel avec ces deux premières composantes principales absorbe 85,36 % de la variance totale, ou ce qui revient au même, de l'inertie totale.

La carte des variables (Figure 6) présente la projection des variables sur les nouveaux axes. Le cercle des corrélations représenté par le plan F1-F2, montre que, les résistivités apparentes obtenues avec les petits écartements sont corrélées positivement à l'axe F1 (premier facteur): les résistivités apparentes pour les écartements AB= 60 sont les variables qui contribuent le plus fortement à la définition de l'axe F1.

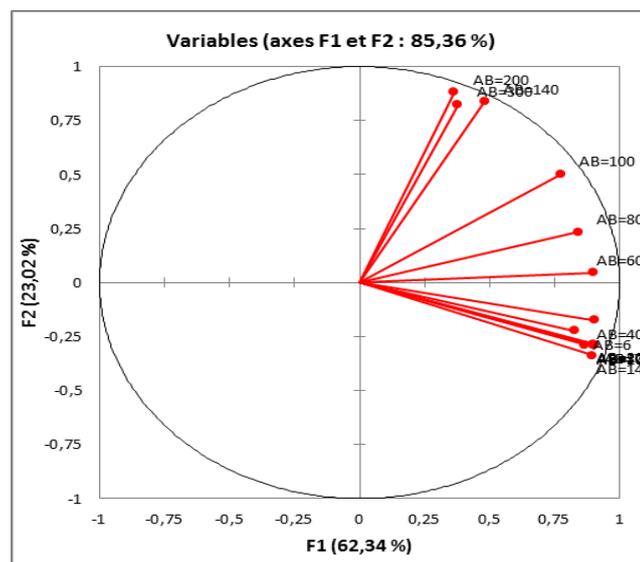


Figure 6: Cercles des corrélations et carte des variables des SEV sur le premier plan factoriel des 13 variables.

La deuxième composante principale oppose les résistivités apparentes des écartements relativement grands (AB=60, AB=80, AB=100, AB=140, AB=200, AB=300) aux résistivités apparentes des petits écartements (AB=6, AB=8, AB=10, AB=14, AB=20, AB=30, AB=40) il en résulte que les sondages électriques présentant une composante positive sur l'axe de cette deuxième composante présenteront les plus faibles valeurs de résistivité apparente pour les grands écartements et, a contrario, des valeurs de résistivités apparentes plus grandes pour les petits écartements.

Le premier facteur qui absorbe 62,34 % de l'information globale est corrélé positivement à toutes les variables, les coefficients de corrélation croissent avec les écartements AB allant de 6 à 40 m pour décroître ensuite pour l'écartement AB de 60 à 300 m. Le deuxième facteur présente une corrélation négative décroissante lorsque les écartements AB croissent de 6 à 40 m et une corrélation positive croissante lorsqu'ils croissent de 60 à 300 m.

Plan factoriel 1-3 :

Le deuxième plan factoriel F1-F3, absorbe 72,29 % de l'information globale, il est donc important d'analyser l'apport de ce troisième facteur.

Sur le cercle de corrélation dans le plan factoriel 1-3 (Figure 7), la structure du nuage de points le long de l'axe de la première composante est naturellement identique à celle qui a été décrite précédemment.

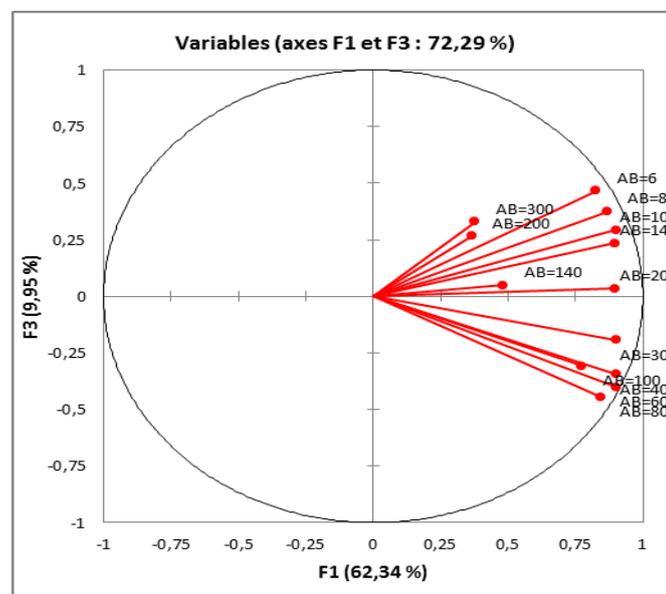


Figure 7: Cercles des corrélations et carte des variables des SEV sur le deuxième plan factoriel des 13 variables.

Il est à noter que les variables résistivités apparentes de l'écartement AB=20 ne s'écartent pas de cet axe alors que ce n'est pas le cas des autres variables. La troisième composante principale oppose les résistivités apparentes des écartements AB=30, AB=40, AB=60, AB=80, AB=100 aux résistivités apparentes des écartements AB=6, AB=8, AB=10, AB=14, AB=20, AB=140, AB=200, AB=300.

Dans le plan principal des individus, l'axe F1 présente toujours les sondages électriques qui présentent des fortes valeurs de résistivités apparentes pour des écartements AB=20 vers la droite et vers la gauche les sondages électriques qui présentent des faibles valeurs de résistivités apparentes pour le même écartement.

L'analyse en composantes principales nous a fourni les principales informations sur la dispersion des sondages électriques (Individus) en fonction des différentes valeurs des résistivités apparentes par écartements des électrodes d'injection de courant.

3.2. Classification et typologie des sondages électriques

Pour mieux comprendre la relation entre les variables, nous avons utilisé la méthode de la classification ascendante hiérarchique. C'est un outil puissant pour déterminer des groupes statistiques naturels ou des structures dans les données. Cette méthode se base sur la construction d'une typologie ou des partitions des sondages électriques en classes relativement les plus homogènes. C'est un système de classification qui utilise

la distance euclidienne entre les individus ici les SEV et le centre de gravité (moyenne)[17], ce type de distance est le plus couramment utilisé, Il s'agit simplement d'une distance géométrique dans un espace multidimensionnel. Elle se calcule ainsi :

$$distance(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Les niveaux de similarité à laquelle les observations sont fusionnées sont utilisés pour construire un dendrogramme (figure 9)[17]. La distance euclidienne est représentée sur l'axe vertical du dendrogramme.

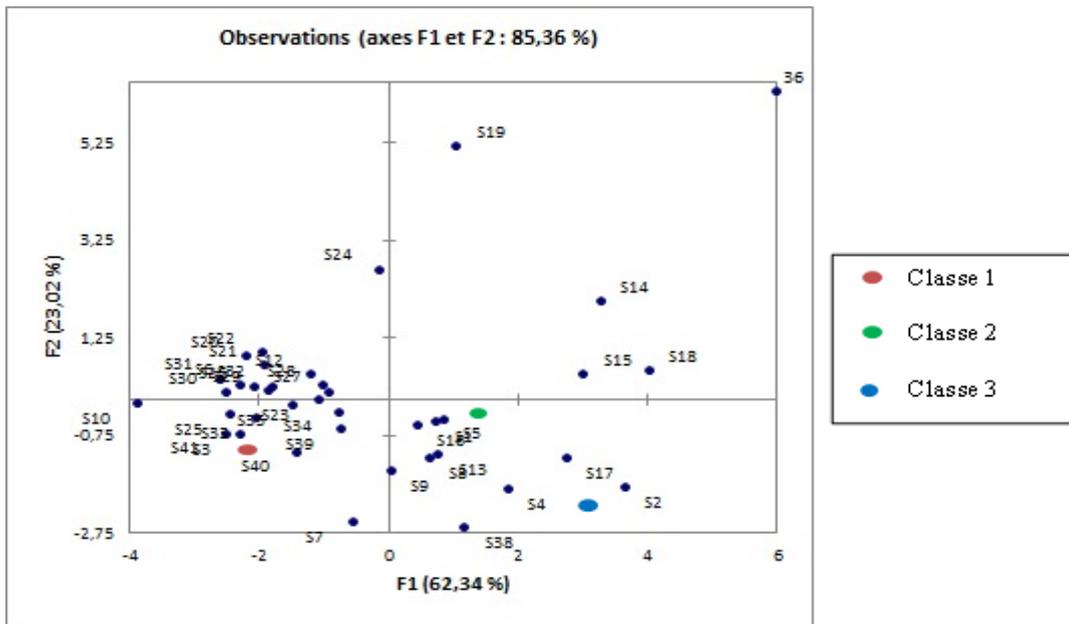


Figure 8 : Cartes de projection du nuage des sondages électriques sur le premier plan factoriel

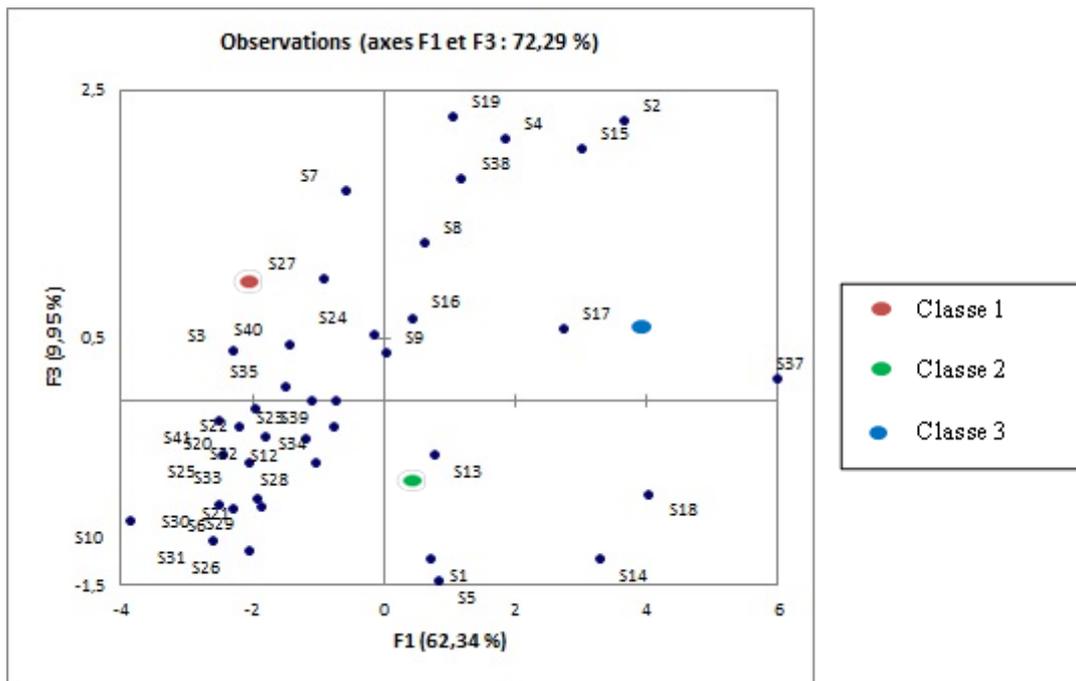


Figure 9: Cartes de projection du nuage des sondages électriques sur le deuxième plan factoriel

L'analyse des résultats de la classification des 41 SEV (figures 8 et 9), fait ressortir 3 grandes classes de sondages, ces classes se distinguent nettement sur le plan des deux premières composantes principales qui absorbent 85,36% de la variance totale (figures 8 et 10). On peut distinguer nettement les sondages de la classe 1 des sondages des classes 2 et 3 de par le signe de leurs composantes sur l'axe F1, négatives ou positives. En effet, en comparaison avec les sondages des classes 2 et 3, Les valeurs des résistivités apparentes pour les sondages de la classe 1 sont les plus fortes pour les faibles écartements et les plus faibles pour les grands écartements.

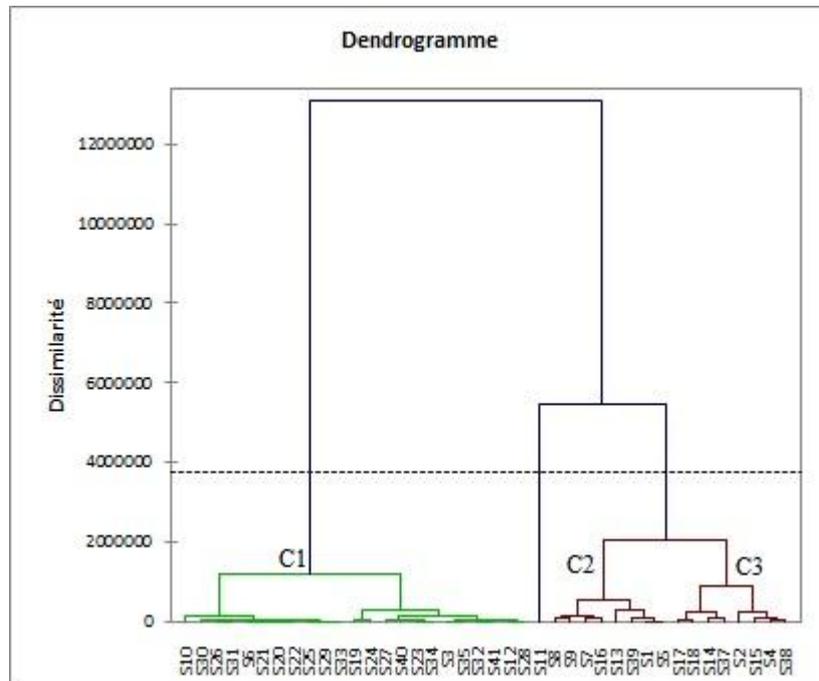


Figure 10: arbre hiérarchique issu de la classification hiérarchisée réalisée à partir des données de SEV

La classe 1 est formée de 23 Sondages (S3, S6, S10, S12, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29, S30, S31, S32, S33, S34, S35, S40 et S41) (figures 8, 9 et 10) caractérisés par des valeurs de résistivités apparentes les plus fortes en subsurface et faibles en profondeur.

Si l'on tient compte cette fois-ci des composantes sur le deuxième et troisième axes principaux (figures 8 et 9), on peut distinguer clairement les classes 2 et 3, cette dissimilarité est due à l'influence des faibles valeurs des résistivités apparentes pour les faibles écartements AB (6 à 14) et les fortes valeurs de résistivités apparentes pour les écartements AB=80m et AB=100m pour la classe 2 à l'inverse des sondages de la classe 3 qui se caractérisent par les fortes valeurs de résistivités pour les faibles écartements contre de faibles valeurs de résistivités apparentes pour les écartements AB=80m et AB=100m.

La classe 2 est formée des sondages S1, S5, S7, S8, S9, S13 et S16, les sondages de la classe 3 sont S2, S4, S14, S15, S17, S18, S37, et S38

A partir d'observation de terrain et selon la position des sondages par rapport aux affleurements de la série phosphatée, les sondages peuvent être scindés en trois groupes: les sondages du groupe I (S1, S2, S11, S12, S13, S14, S15, S16, S17, S18, S19, S23, et S24) qui doivent normalement représenter les couches s'étalant depuis le Sénonien jusqu'au Lutétien moyen, les sondages du groupe II (S7, S8, S9, S10, S21, S26, S27, S30, S31, S32 et S33) comprennent la série phosphatées sans la dalle à thersités et enfin les sondages du groupe III (S3, S4, S5, S6, S22, S25, S28, S34, S35, S36, S37, S38, S39, S40 et S41) qui regroupent les SEV implantées sur les terrains de l'ensemble 3 allant du sillon B à la couche 3 inférieure (couche CB et couche CIII respectivement de la figure 2).

De ce fait, la combinaison de ces observations avec les résultats de l'analyse statistique multivariée des sondages électriques permettrait de dire que :

- 1- les fortes résistivités en subsurface des sondages de la classe 1 sont dues soit i- à la présence de la dalle à thersités, observée sur place (S12, S19, S23 et S24) ii- à la présence de zones rocheuses (S34 et S40) iii- à la forte teneur en calcaire et silex du sillon B (S3, S6, S22, S25, S28, S35 et S41) iv- ou alors à la présence de dérangement proche de la surface (S10, S20, S21, S26, S27, S30, S31, S32 et S33) ;

- 2- la classe 2 est constituée de sondages probablement dérangés;
- 3- la classe 3 est composée de sondages sains dont la forte résistivité en surface est due soit à la présence de phosphate riche en silex.

Conclusion

Dans le cadre de ce travail et dans le but de caractériser spatialement les anomalies de Sidi Chennane, une analyse statistique multivariée (analyse descriptive suivie par une étude statistique multidimensionnelle: ACP et CHA) des résistivités apparentes des sondages électriques verticaux a été réalisée. L'analyse statistique multivariée s'est révélée fiable quant à l'interprétation d'un nombre important de sondages électriques afin d'avoir une vision globale de la distribution des résistivités apparentes dans toute la zone. En effet, les résultats de l'analyse statistique multivariée en combinaison avec les observations de terrain ont permis de redéfinir les trois grands ensembles verticalement superposés à savoir i- un premier ensemble avec des valeurs de résistivités apparentes très élevées qui caractérisent bien les formations d'argiles à phosphates, série phosphatée et marne, ii- un deuxième ensemble qui présente des résistivités moyennes dues à la présence de phosphates riches en silex, de la série phosphatée et des marnes, iii- et un dernier ensemble avec des résistivités apparentes moyennement élevées qui reflète la présence de la série phosphatée dérangée et des formations marneuses.

References

1. Azmany M., Évolution des faciès sur le gisement des Ouled Abdoun, Mines, Géologie et Énergie, Rabat 44 (1979) 35–38.
2. Michard A., Saddiqi O., Chalouan A., *Continental evolution: the geology of Morocco: structure and stratigraphy*, (2008) Springer
3. Yans J. et al. First carbon isotope chemostratigraphy of the Ouled Abdoun phosphate Basin, Morocco; implications for dating and evolution of earliest African placental mammals, *Gondwana Research* 25 (2014) 257–269
4. Kocsis L. et al. Comprehensive stable isotope investigation of marine biogenic apatite from the late Cretaceous–early Eocene phosphate series of Morocco. *Palaeogeography, Palaeoclimatology, Palaeoecology* 394 (2014) 74–88
5. GHEERBRANT E. et al. Les localités à mammifères des carrières de Grand Daoui, bassin des Ouled Abdoun, Maroc, Yprésien : premier état des lieux. *Bull. Soc. géol. Fr.* 174 (2003), 3, 279-293
6. Boujo A. About shape and development of sterile bodies in phosphatic deposits. *C. R. Geoscience* 334 (2002) 1113, 1114
7. El Assel N et al. A Ground Penetrating Radar and Electrical Resistivity Tomography Prospection for Detecting Sterile Bodies in the Phosphatic Bearing of Sidi Chennane (Morocco). *International Journal of Geosciences* 2 (2011) 406-413.
8. Baba K. et al. Mapping Sterile Bodies in the Sidi Chennane phosphatic deposit (Morocco) using geoelectrical Investigations, *Int. J. Eng. Res. Appl. (IJERA)*, 2 (5) (2012) 2132-2136
9. Kchikach A. et al. Cartographie de corps stériles sous couverture quaternaire par méthode de résistivités électriques dans le gisement phosphaté de Sidi Chennane (Maroc), *Geosciences*, 334 (2002) 379-386
10. Essahlaoui A. et al. Caractérisation des réservoirs d'eau souterraine à l'aide de l'analyse géostatistique des sondages électriques au sud du bassin de Saïss (Maroc). *Journal of hydrocarbons mines and environmental research*, 3, (2) (2012) 76-82
11. Harmouzi O. et al. Interprétation de sondages électriques verticaux par combinaison de méthodes statistique, géostatistique et d'inversion : application au bassin de Saïss (Maroc). *Africa Geoscience Review*, 16 (2) (2009). 119-140
12. Oualaouss Z. et al. Inversion de sondages électriques verticaux par la méthode des essais particuliers : application à l'étude du biseau d'eau salée de la région du Sahel Doukkala (Maroc). *Africa Geoscience Review*, 16 (2) (2009) 103-117
13. Riss J.A Methodology for Fast Reinterpretation of VES Historical Data. Application to the Saïss Basin (Morocco). *Geophysics*, 76 (6) (2011) B225-B236
14. Ahoussi KE et al. Application des méthodes d'analyses statistiques multivariées à l'étude de l'origine des métaux lourds (Cu²⁺, Mn²⁺, Zn²⁺ et Pb²⁺) dans les eaux des nappes phréatiques de la ville d'Abidjan. *International Journal of Biological and Chemical Sciences*, 4(5) (2010) 1753-1765.
15. Kassou A. et al. Modélisation numérique de la minéralisation polymétallifère du gisement de Tighza (Maroc central) par l'outil SIG et la géostatistique. *Afrique Science: Revue Internationale des Sciences et Technologie*, 7(1) (2011)
16. Ding S. On the Application of PCA Technique to Fault Diagnosis. *Tsinghua Science & Technology* 15 (2) (2010) 138–144.
17. Chen K. et al. Multivariate statistical evaluation of trace elements in groundwater in a coastal area in Shenzhen, China, *Environmental Pollution*, 147(3), (2007), 771-780.
18. Telesca L. Principal component analysis of geoelectrical signals measured in the seismically active area of Basilicata Region (southern Italy). *Natural Hazards and Earth System Sciences* 4 (2004) 663–667.