

UNIVERSITÉ KASDI MARBEH OUARGLA
FACULTÉ DES HYDROCARBURES DES ÉNERGIES RENOUVELABLES ET DES
SCIENCES DE LA TERRE ET DE L'UNIVERS
DÉPARTEMENT DE SCIENCES DE LA TERRE ET DE L'UNIVERS



MÉMOIRE

Présenté pour l'obtention du diplôme de **MAGISTER**

Spécialité : Géologie.

Option : Géologie de Réservoirs.

Présenté par :

AMEUR ZAIMECHE Ouafi

THÈME

**MODÉLISATION ET RECONSTITUTION DES FACIES NON CAROTTÉS A
L'AIDE DES MÉTHODES STATISTIQUES MULTIVARIEES DU RÉSERVOIR
TRIAS ARGILEUX GRESEUX INFÉRIEUR (TAGI) APPLICATION AU CHAMP
DE SIF FATIMA BASSIN DE – BERKINE -**

Soutenu publiquement le : 25 / 09 / 2014. Devant le jury composé de.

Mr. AISSANI Belkacem	Maitre de Conférences UKM Ouargla	Président
Mr. ZEDDOURI Aziez	Maitre de Conférences UKM Ouargla	Directeur de Mémoire
Mr. HACINI Messaoud	Professeur UKM Ouargla	Examineur
Mr. NEDZLI Imed Eddine	Maitre de Conférences UKM Ouargla	Examineur

Dédicace

C'est avec toute mon affection que je dédie ce modeste travail :

*A celui qui m'a donné vie et qui m'as apporté sans cesse son amour, mon
père.*

*A ma très chère mère qui m'a apporté sans cesse amour,
soutien et encouragement ; et qu'elle trouve ici l'expression de mes vives
reconnaisances.*

Mon chère frère Fouzi et ma chère sœur Amani ma petite princesse.

A ma grande mère et mon grand-père.

A mon cousin Omar.

A toutes les familles Ameur Zaimeche et Boutouga.

A tous mes amis:

*Abdesselem, Nasro, Abderrahim, Kawwa, Salaheddine, Sobhi,
Nourddine, Hamza², Mustapha, Walid², Bilel², Mohammed Zighad,
Ahmed *Toche*, Khaoula, Sara, Narimane, Asma, Fadila, Nacira, et à
tous mes autres amis.*

Ameur Zaimeche Ouafi

REMERCIEMENTS

Nous remercions dans un premier temps, notre Créateur ALLAH qui nous a donné la puissance, la patience, la volonté et l'énergie pour terminer ce travail.

Je tiens à remercier mon encadreur de mémoire, Mr. ZEDDOURI Aziez. Je le remercie de m'avoir proposé ce sujet, de m'avoir encadré et de m'avoir consacré une grande partie de son précieux temps pour la réalisation de ce travail. Merci pour vos orientations rigoureuses et vos conseils constructifs.

Un grand merci aux membres de jury qui ont accepté d'examiner ce travail : Mr. AISSANI Belkacem qui m'a fait l'honneur de présider le jury de soutenance. Mr. HACINI Messaoud et Mr. NEDZLI Imed Eddine d'avoir accepté d'examiner ce travail.

Je remercie vivement Mr. HEDDAM Salim, d'université de Skikda pour son aide depuis le commencement du mémoire, pour le soutien, et les orientations enrichissantes surtout dans la partie de programmation.

Je tiens à remercier Mr. AMATA Ammar, mon frère et mon collègue qui m'a encouragé et m'a soutenu je ne peux jamais le remercier assez.

Je ne peux pas oublier Mr KOUADRIA Tayeb mon ex-encadreur de l'ingéniorat, université d'Annaba qu'il trouve ici le témoignage de ma grande reconnaissance.

Je tiens également à remercier, tous les enseignants qui ont participé à ma formation en post graduation, en particulier tous les enseignants de département des Science de la terre et de l'univers de l'Université d'Ouargla.

Une «montagne» de merci pour Mr KECHICHED Rabah mon ami et mon frère, j'ai souvent voulu te remercier pour tout ce que tu avais fait pour moi sans jamais avoir trouvé les mots et le moment opportun pour le faire. Tu m'as transmis avec tant de générosité ta passion pour la géologie, et l'art de la recherche scientifique. Pour tout ça je te remercie de tout mon cœur.

Je ne peux oublier mes frères BELKSIER Mohamed Salah, HADDANE Abdennour, REMITA Abdelatif, SATOUH Adel, BOUSELSAL Boualm, REMITA Abderraouf, FENAZI Bilel, qui font ma deuxième famille. Je voudrais vous remercier également pour m'avoir accueilli si souvent chez vous à bras ouvert et m'avoir si spontanément intégré «Albeer» dans l'équipe, je ne l'oublierai jamais. Je m'estime chanceux d'avoir eu l'occasion d'appartenir à l'équipe, dont l'excellente ambiance ne peut avoir d'égale.

Nous remercions tous ce qui ont contribué de près ou de loin à l'élaboration de ce mémoire.

RESUME

La récupération des carottes n'est pas toujours intégrale, à cause des problèmes techniques et géologiques rencontrés durant l'opération du carottage, à cet effet notre travail fait l'objet de pallier le manque des faciès lithologiques (332.84 m représente la longueur des intervalles non carottés dans 5 puits) au niveau du réservoir du Trias Argileux Gréseux Inférieur (TAGI), bloc 402b, champ de Sif Fatima bassin de Berkine Algérie -.

Deux méthodes de modélisations ont été appliquées sur les données diagraphiques (Gamma Ray, Densité, Potassium et Thorium) combinées à la description détaillée des carottes disponibles, pour reconstituer les faciès non-carottés afin d'établir des logs lithologiques complets.

La première méthode est basée sur la théorie de Réseau de Neurone Artificiel, qui a été inspirée de la biologie, à un modèle de perception multicouches avec un algorithme de (RETROPROPAGATION) à apprentissage supervisé, qui se compose d'une couche d'entrée porte 4 neurones qui sont les variables explicatives (les données diagraphiques), et d'une couche de sortie composée d'un seul neurone qui est la variable à expliquer (le Faciès non carotté), et une seule couche cachée a une fonction de transfert sigmoïdale (TANSIG) de la couche d'entrée vers la couche cachée, et une fonction linéaire (PURELIN) de la couche cachée vers la couche de sortie, et l'algorithme d'activation utilisé est de type (LEVENBERG-MARQUARDT). Le modèle a une tendance d'amélioration et d'adaptation interactivement à chaque fois on change le nombre de neurones dans la couche cachée, jusqu'à ce que les indices de performances numériques (CC, RMSE, MAE, d) et graphique (la fonction coût) soient excellents dans toutes les phases de modélisation, alors on a pris le modèle à 4 neurones dans la couche cachée et de coefficient de corrélation CC de 0,87 pour l'apprentissage, 0,78 pour la validation et 0,92 pour le test.

La deuxième méthode est une technique statistique dite Cluster Analysis qui débrouille le même problème de la non récupération des carottes que la première, mais elle repose sur une technique de calcul qui consiste à diviser un ensemble de données en différents paquets homogènes, dans le sens que les données de chaque sous-ensemble partagent des caractéristiques communes, et faire la déviation du log par faciès, Puis on a calculé l'erreur pour chaque faciès, et attribué le faciès ayant la valeur minimale des erreurs.

Les résultats obtenus des deux méthodes ont été visualisés en logs, et une comparaison quantitative et qualitative a été effectuée. Elle a permis de constater globalement, une correspondance entre les facies des carottes avec celles de la prédiction, ceci avec une concordance des réponses des pics diagraphiques. La comparaison entre les deux méthodes montre que la première s'avère plus performante que la deuxième, et on peut la généraliser sur d'autres caractéristiques géologiques dans le développement du réservoir.

Mots clés: Récupération des carottes, TAGI, Berkine, Modélisations, Diagraphie, Réseau de Neurone Artificiel, Cluster Analysis.

ABSTRACT

The recovery of cores is not always integral, because of the technical and geological problems encountered during coring operation, for this purpose our study is the object to restoration of the missing of the lithological facies (332.84 m are intervals not cored in 5 wells) in the reservoir of Lower Triassic sandstone clay (TAGI), block 402b, field of Sif Fatima, basin of Berkine - Algeria -

Two methods of modeling were used on the logging data (Gamma Ray, Density, Potassium and Thorium) combined with the detailed description of available cores, for rebuilding the not-cored facies in order to display a complete lithological log.

The first method is based on the theory of Artificial Neural Network, which was inspired by biology, with a model of perception multi-layers with an algorithm of (RETRO PROPAGATION) with training supervised, which is composed of a layer of the input contain 4 neurons which are the target variables (logging data), and of a layer of exit made up of one alone neuron which is variable explained it (the Facies not cored), and one hidden layer has a transfer function sigmoidal (TANSIG) from input layer to hidden layer, and (PURELIN) function with from hidden layer to the output layer, and the algorithm of activation used is (LEVENBERG-MARQUARDT). The model has a tendency to improvement and adaptation each time we change the number neurons in the hidden layer, until indices of performance numerical (CC, RMSE, MAE, d) and graphical (cost function) will be excellent with all phases of modeling, then we took the best model had 4 neurons in the hidden layer and of coefficient of correlation CC of 0,87 for the training, 0,78 for the validation and 0,92 for the test.

The second method is a statistical technique known as Cluster Analysis which address the same problem of no recovery of cores like the first, but it rests on a technique of calculation consists in dividing a whole of data into different homogeneous packages, in the sense that the data of each subset have common characteristics, and to make the deviation of the log by facies, Then one calculated the error for each facies, and attribute the facies having the minimal value of the errors.

The results obtained of the two methods were visualized in logs, and a quantitative and qualitative comparison was made. It showed overall correspondence between facies cores with those of the prediction. That with concordance of responses of logging peaks. The comparison between the two methods shows that the first proves to be powerful that the second and we can generalize on other geological characteristics in the development of the reservoir.

Key words: Recovery of cores, TAGI, Berkine, Modeling, Logging, Artificial Neural Network, Cluster Analysis.

الملخص

إن عملية استخراج عينات الصخور ليست دائما كاملة، و ذلك راجع إلى المشاكل الفنية و الجيولوجية المتلقات خلال عملية حفر آبار النفط، ولهذا كان موضوع المذكرة يتناول كيفية تعويض عينات الصخور لـ 332.84 متر في المجالات الغير مستردة لخمسة آبار، في الخزان الترياسي الطيني الحصى السفلي (TAGI) قطع 402b، حقل سيف فطيمة، حوض بركين- الجزائر-

طريقتين من النمذجة استعملت على بيانات من التسجيلات الجيوفيزيائية (أشعة غاما-الكثافة-البوتاسيوم-التوريوم) وضعت جنبا إلى جنب مع الوصف الدقيق للعينات المتاحة، لإعادة بناء المجالات الغير مستردة و التعبير عنها بأعمدة كاملة.

تستند الطريقة الأولى على نظرية الشبكة العصبونية الاصطناعية التي استنبطت من البيولوجيا، باستخدام نموذج من شبكة عصبونية متعددة الطبقات و ذلك بتطبيق خوارزمية الانتشار المعاكس (RETROPROPAGATION)، وقابلة على التدريب الموجه، و التي تتكون من طبقة المدخلات من أربع عصبونات مثلثا المتغيرات (التسجيلات الجيوفيزيائية)، و طبقة الإخراج و التي تشكل من عصبون واحد و هو الناتج المراد الوصول إليه (السحنة الغير مستردة)، في الطبقة الخفية استخدمت فيها عدة دوال للعبور من طبقة إلى أخرى، من طبقة الإدخال إلى الطبقة الخفية طبقت فيها دالة غير خطية (TANSIG) ودالة خطية (PURELIN) من الطبقة الخفية نحو طبقة الإخراج، و وخوارزمية التفعيل هي (LEVENBERG-MARQUARDT) مع إمكانية تحسين النموذج بتغيير عدد العصبونات في الطبقة الخفية، حتى نتحصل على علامات الجودة الحسابية (CC, RMSE, MAE, d) و البيانية (دالة التكلفة) ممتازة في جميع مراحل النمذجة، ثم قمنا بأخذنا أحسن نموذج الذي احتوى على 4 عصبونات في الطبقة الخفية بمعامل ترابط CC في مرحلة التدريب 0,87، في مرحلة التحقق، 0,78 و 0,92 في مرحلة الاختبار .

الطريقة الاحصائية الثانية التي تدعى التحليل العنقودي (CLUSTER ANALYSIS) تعالج نفس المشكل و المرتكزة على تقنية حسابية قاسمتا المعطيات إلى مصفوفات متجانسة تتقاسم نفس الخصائص فيما بينها، للحصول على انحراف السجل للسحنة تم نحسب مقدار الخطأ لكل سحنة و إعطاء كل سحنة اقل نسبة أخطاء.

النتائج المتحصل عليها لكلا الطريقتين تم تمثيلها تخطيطيا ومقارنة كمية و نوعية، بينت لنا التوافق بين الوصف الحقيقي للعينات و التنبؤات المحصل عليها و كذلك توافقهما مع المنحنيات المسجلة، المقارنة بين نتائج الطريقتين أكدت أن الطريقة الأولى أحسن من الطريقة الثانية، مع القدرة على تعميمها على خصائص جيولوجية أخرى في تطوير المخزن.

الكلمات المفتاحية: استخراج عينات الصخور، TAGI، بركين، النمذجة، التسجيلات الجيوفيزيائية، الشبكة العصبونية الاصطناعية، التحليل العنقودي.

SOMMAIRE

RESUME.....	I
ABSTRACT.....	II
المُلخَص.....	III
LISTE DES FIGURES.....	IV
LISTE DES TABLEAUX.....	V
INTRODUCTION GENERALE.....	1

CHAPITRE I : SYNTHÈSE BIBLIOGRAPHIQUE SUR LES METHODES D'INTELLIGENCE ARTIFICIELLE

1. LES RESEAUX DE NEURONES	3
I.1. Les Réseaux de Neurones Biologiques	3
I.2. Les Réseaux de Neurones Artificiels.....	8
I.3 Application des réseaux de neurones dans le domaine de géologie.....	39
2. METHODE DE CLASSIFICATION CLUSTER ANALYSIS.....	44
2.1. Définition.....	44
2.2. Types de Cluster Analysis.....	44
2.3. Application de Cluster Analysis dans le domaine de géologie	44

CHAPITRE II : GEOLOGIE REGIONALE ET LOCALE DE LA ZONE D'ETUDE

1. GEOLOGIE REGIONALE DU BASSIN DE BERKINE.....	48
I.1. Situation Géographique	48
I.2. Cadre Géologique Régionale.....	49
I.3. Evolution tectonique	50
I.4. Géodynamique du Bassin	52
I.5. Stratigraphie	54
I.6. Aspect structural.....	58
I.7. Systèmes pétroliers.....	63
2. GEOLOGIE LOCALE DE CHAMP DE SIF FATIMA.....	64
2.1. Situation Géographique.....	64
2.2. Synthèse Géologique et Géodynamique	64
2.3. Le potentiel pétrolier de la zone d'étude	69

CHAPITRE III : MATERIEL ET METHODES

1. PREPARATION DE LA BASE DE DONNEES	71
1.1. Carottes	71
1.2. Diagraphies	72
1.3. Validation et codage des données.....	75
2. LOGICIELS UTILISEES	78
2.1. Logiciel "EXCEL".....	78
2.2. Logiciel "MATLAB".....	78
2.3. Logiciel "PETREL 2009"	78
3. ELABORATION DU MODELE PAR LES RESEAUX DE NEURONES	79
4. ELABORATION DU MODELE PAR CLUSTER ANALYSIS.....	90

CHAPITRE IV: RESULTATS ET DISCUSSION

1. Reconstitution des facies par la méthode des Réseaux des Neurones Artificiels	96
1.1. Comparaison qualitative	96
1.2. Comparaison quantitative	101
1.3. Synthèse des résultats pour les puits A, B et C.....	104
2. Reconstitution des facies par la méthode de Cluster Analysis.....	106
2.1. Comparaison qualitative	106
2.2. Comparaison quantitative	111
2.3. Synthèse des résultats pour les puits A, B et C.....	113
3. Comparaison entre les réseaux de neurones artificiels et Cluster Analysis	115
CONCLUSION GENERALE.....	117

BIBLIOGRAPHIE

ANNEXES

LISTE DES FIGURES

Fig. 1	- Schéma détaillée d'un neurone biologique.	5
Fig. 2	- Différentes géométries de neurones http://static.howtsu.ffwork.com	7
Fig. 3	- le neurone formel	14
Fig. 4	- fonction de transfert binaire.....	14
Fig. 5	- Mise en correspondance neurone biologique / neurone artificiel	16
Fig. 6	- Modélisation simplifiée d'un neurone Artificiel.....	16
Fig. 7	- Représentation matricielle du modèle d'un neurone artificielle.....	18
Fig. 8	- Définition des couches d'un réseau multicouche	19
Fig. 9	- Réseau à connexions locales	20
Fig. 10	- Réseau à connexions récurrentes.....	20
Fig. 11	- Réseau à connexions complète.....	20
Fig. 12	- couche de S neurones	21
Fig. 13	- Représentation matricielle d'une couche de S Neurones	22
Fig. 14	- Représentation matricielle d'un réseau de trois couches.....	22
Fig. 15	- Schéma bloc de l'apprentissage supervisé	27
Fig. 16	- i le neurone amont, j le neurone aval et w_{ij} le poids de la connexion	28
Fig. 17	- Perceptron à une seule couche avec fonction seuil	31
Fig. 18	- Perceptron à une seule couche avec fonction seuil	32
Fig. 19	- Modèle Mathématique simple d'un PMC.....	34
Fig. 20	- Un réseau de neurone à n entrées	34
Fig. 21	- Schéma détaillée d'un PMC	37
Fig. 22	- Situation géographique (S. Galeazzi et al. 2010),modifié.....	48
Fig. 23	- Carte géologique de l'Algérie (Askri et al., 1995).	49
Fig. 24	- Les majeures évolutions structurales du bassin de Berkine (McKenna et Hedley, 2002).	52
Fig. 25	- Les éléments structuraux du bassin de Berkine (Azoug et al. 2007)	54
Fig. 26	- Log lithostratigraphique du bassin de Berkine (Sonatrach / Anadarko 2012).....	57
Fig. 27	- Coupe géologique régionale du Bassin de Berkine. (S. Galeazzi et al. 2010)	59

Fig. 28 - Sources et migrations dans le bloc 404 (D'après Echikh, 1998)	63
Fig. 29 - Carte de situation de Bloc d'étude (SIF FATIMA) du bassin de Berkine. (WEC, 2007)	64
Fig. 30 - La zonation du TAGI, séquence, lithologies et environnement (P.Tuner et al. 2001).	66
Fig. 31 - présentation de la tectonique du secteur de SIF FATIMA (Rapport Interne Sonatrach 2012).	67
Fig. 32 - Coupe sismique de secteur SIF Fatima (Rapport Interne Sonatrach / Anadarko).	67
Fig. 33 - Carte Structurale au Toit de TAGI réservoir du secteur SIF Fatima (Sonatrach, Anadarko 2012).	68
Fig. 34 - Photos des carottes puits P_A	71
Fig. 35 - Outil de mesure de la radioactivité naturelle (Lagabrielle, 2008)	74
Fig. 36 - Outil de mesure de la densité et Exemple de log gamma-gamma (densité) (Document Schlumberger).....	75
Fig. 37 - Codage des Facies du Réservoirs TAGI.....	76
Fig. 38 - Architecture du réseau de neurones élaboré.....	85
Fig. 39 - Schéma de régression montrant les coefficients de corrélation entre les facies observés et les facies estimés pour des étapes de l'apprentissage, de validation et le test	88
Fig. 40 - Diagrammes de performance des étapes d'apprentissage (ligne bleue), de validation (ligne verte), et du test (ligne rouge). (Fonction coût) La meilleure performance de la validation est égale à 0.35293 et s'arrête à 15 itérations.	88
Fig. 41 - Schéma qui résume les étapes de la méthode « Cluster Analysis »	91
Fig. 42 - Résultat de RNA pour le puits P_A	97
Fig. 43 - Résultat de RNA pour le puits P_B	98
Fig. 44 - Résultat de RNA pour le puits P_C.....	99
Fig. 45 - Résultat de RNA pour le puits P_D.....	100
Fig. 46 - Résultat de RNA pour le puits P_E	101
Fig. 47 - Histogramme de comparaison facies carottes et facies RNA puits P-A.....	102
Fig. 48 - Histogramme de comparaison facies carottes et facies RNA puits P-B	103
Fig. 49 - Histogramme de comparaison facies carottes et facies RNA puits P-C	104
Fig. 50 - Histogramme de comparaison totale entre RNA et la carottes	105
Fig. 51 - Résultat de Cluster Analysis pour le puits P_A	106
Fig. 52 - Résultat de Cluster Analysis pour le puits P_B	107

Fig. 53 - Résultat de Cluster Analysis pour le puits P_C	108
Fig. 54 - Résultat de Cluster Analysis pour le puits P_D	109
Fig. 55 - Résultat de Cluster Analysis pour le puits P_E	110
Fig. 56 - Histogramme de comparaison facies carottes et facies CA puits P-A	111
Fig. 57 - Histogramme de comparaison facies carottes et facies CA puits P-B	112
Fig. 58 - Histogramme de comparaison facies carottes et facies CA puits P-C	113
Fig. 59 - Histogramme de comparaison totale entre Cluster Analysis et les carottes	114
Fig. 60 - Comparaison des résultats de RNA et Cluster Analysis des puits (A, B, C) avec les carottes.....	116

LISTE DES TABLEAUX

Tableau. 1 : Analogie entre les neurones biologiques et artificiels	9
Tableau. 2 : Différents types de fonctions de transfert pour le neurone artificiel	17
Tableau. 3 : Comparaison entre les termes utilisés en statistique et en RNA.....	23
Tableau. 4 : La loi de Hebb	28
Tableau. 5: Illustration du codage des facies	77
Tableau. 6 : Statistique des paramètres étudiés	86
Tableau. 7: Séparation du fichier de base.....	87
Tableau. 8 : Performances du modèle élaboré.....	87
Tableau. 9: Les Paramètres retenus.....	89
Tableau. 10 : Moyennes des logs dans tous les puits carottés.	91
Tableau. 11: Max, Min et Range des logs dans tous les puits carottés	92
Tableau. 12 : Illustration des calculs effectués pour les Argiles.....	93
Tableau. 13 : Illustration des calculs effectués pour les Silts.....	93
Tableau. 14 : Illustration des calculs effectués pour les Grès.....	94
Tableau. 15 : Illustration des calculs effectués pour les Grès Argileux.....	94
Tableau. 16 : Illustration des calculs effectués	95
Tableau. 17 : Comparaison entre RNA et les carottes du puits P-A.....	102
Tableau. 18 : Comparaison entre RNA et les carottes du puits P-B	102
Tableau. 19 : Comparaison entre RNA et les carottes du puits P-C	103
Tableau. 20 : Comparaison de nombre d'échantillons entre RNA et la carotte	105
Tableau. 21 : Comparaison totale entre RNA et les carottes.....	105
Tableau. 22 : Comparaison entre Cluster Analysis et les carottes du puits P-A	111
Tableau. 23 : Comparaison entre Cluster Analysis et les carottes du puits P-B	112
Tableau. 24 : Comparaison entre Cluster Analysis et les carottes du puits P-C	113
Tableau. 25: Comparaison de nombre d'échantillons entre Cluster Analysis et la carotte	114

Tableau. 26 : Comparaison totale entre Cluster Analysis et les carottes	114
Tableau. 27 : Comparaison entre RNA, Cluster Analysis et les carottes Puits-A	115
Tableau. 28 : Comparaison entre RNA, Cluster Analysis et les carottes Puits-B.....	115
Tableau. 29 : Comparaison entre RNA, Cluster Analysis et les carottes Puits-A	115

INTRODUCTION GENERALE

INTRODUCTION GÉNÉRALE

L'industrie pétrolière est un pilier de l'économie mondiale. L'or noir est une source majeure d'énergie. Le Sahara algérien est riche en pétrole. La plate-forme saharienne est une région très vaste et stable qui a été modelée dès le Paléozoïque. Elle est subdivisée en un certain nombre de bassins répartis en trois provinces pétrolières - la province occidentale, la province orientale et la province triasique.

Le Bassin de Berkine est l'un des plus grands champs en Algérie. Cette importance est traduite par les découvertes de plusieurs gisements d'hydrocarbures. Le Bloc de SIF Fatima représente l'un des nouveaux axes de développement dans ce bassin, et contribue au renouvellement des réserves de notre pays en hydrocarbure et de renforcer la production qui connaîtrait, une décroissance ces dernières années (WEC 2007).

Le gisement de SIF Fatima qui appartient aux Bloc 402b couvre une superficie de 103 Km² se situe au Nord-est du bassin de Berkine, à environ 100 Km au sud-ouest du gisement d'El Borma. Ce bloc a été découverte durant des années 1980 par la compagnie National des Hydrocarbures SONATRACH et ses partenaires (WEC 2007).

Ce Bloc a connaît un effort d'exploration et de développement, grâce un énorme programme de forage destructif et carottant. La bonne exploitation des données (sismiques, carottes, diagraphies, ...etc.), va permettre une meilleure caractérisation du réservoir. D'autant plus, l'opération de carottage n'est souvent interrompue par des problèmes techniques ou une mauvaise récupération. Ceci nécessite l'utilisation des techniques permettant la connaissance des niveaux non-carottés et/ou de corriger les erreurs de la description macroscopique.

Dans cette perspective, ce mémoire de fin d'étude présenté en vue de l'obtention de diplôme de magister en Géologie option Géologie de Réservoir, va aborder un sujet qui consiste à exploiter les données diagraphiques et de les coupler avec la description des carottes. L'objectif ciblé repose sur l'utilisation de deux techniques de modélisation des faciès non-carottés, qui sont le réseau de neurone artificiel et Cluster Analysis qui sont choisies pour compléter les informations géologiques manquantes après la non récupération des parties de carottes des cinq puits désignés confidentiellement (P_A, P_B, P_C, P_D, P_E) dans l'application, et voir qu'elle est la méthode la plus performante et qui donne la moindre erreur.

Afin d'arriver à cet objectif, il a été nécessaire d'effectuer les travaux suivantes :

Chapitre I : Ce premier chapitre est subdivisé en deux parties. En premier, nous présenterons succinctement la théorie du réseau de neurones et ses applications dans le domaine de géologie. En deuxième partie, on présente la théorie de Cluster Analysis et ses applications dans le domaine de géologie.

Chapitre II : Comporte une présentation de la région d'étude ; le cadre géologique régional du bassin de Berkine, la géologie locale du champ de Sif Fatima et la caractérisation du réservoir Trias Argileux Gréseux Inférieur TAGI.

Chapitre III : Matériel et Méthodes comporte En premier, nous présenterons la préparation de la base de données en se basant sur la description détaillée des carottes et la combinaison avec les diagraphies. La deuxième partie est consacrée pour l'application des réseaux de neurones pour reconstituer les facies, La troisième partie discute le modèle d'application de Cluster Analysis et la possibilité d'application pour la reconstitution les facies.

Chapitre IV : Résultats et Discussion comporte une analyse qualitative (visualisation des logs), et quantitative (analyse statistique sur les résultats de prédiction), et une comparaison entre les résultats des deux méthodes et voir quelle est la plus performante.

Ce travail s'achève par une conclusion générale présentant les principaux résultats obtenus et une discussion portant sur les perspectives d'étude et recommander la meilleure méthode qui donne les meilleures résultats, et de pouvoir généraliser sur d'autres caractéristiques géologiques pour le développement du réservoir.

Chapitre I

Synthèse Bibliographique sur les Méthodes d'Intelligence Artificielle

Ce premier chapitre est subdivisé en deux parties.

En premier, nous présenterons succinctement la théorie du réseau de neurones et ses applications dans le domaine de géologie.

En deuxième partie, on présente la théorie de Cluster Analysis et ses applications dans le domaine de géologie.

Introduction

L'intelligence artificielle est la construction de programmes informatiques qui s'adonnent des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique. (Marvin et al.1967). Le but d'intelligence artificielle est l'étude de la structure de l'information et de la structure des processus de résolution des problèmes, indépendamment des applications et indépendamment d'une réalisation. (McCarthy, J.1959). Une condition essentielle pour agir intelligemment d'une manière générale est la capacité de produire et de manipuler des structures symboliques. (Allen Newell. 1972). Les principales méthodes utilisées sont Les réseaux de neurones et Cluster Analysis.

1. RESEAUX DE NEURONES ARTIFICIELS

Depuis une cinquantaine d'années, les chercheurs en informatique s'inspirent de la biologie dans le but de réaliser des constructions capables de résoudre divers problèmes complexes. La nature a depuis longtemps forgé le premier processeur (l'A.D.N1), des outils d'optimisation puissants, ainsi que des structures de calcul massivement parallèles possédant de formidables capacités d'adaptation comme notre cerveau. En effet, la centaine de milliards de neurones (connectés chacun à parfois plus de dix mille autres de manière tridimensionnelle) de notre cerveau a de quoi attirer l'attention des informaticiens.

1.1. Les Réseaux de Neurones Biologiques

Ils pèsent environ 1500 grammes et contiennent trois couches successives :

- **le cerveau reptilien**, siège des réflexes instinctifs,
- **l'hippocampe**, siège de l'émotivité et de la sensibilité,
- **le cortex**, spécifique aux mammifères et siège du raisonnement et de la pensée.

Ces couches successives contiennent deux types de cellules :

Les **gliales** constituent le tissu nourricier et la structure du cerveau et représentent environ 99.99% du volume du cerveau.

Les **neurones** dont le nombre, estimé à un nombre variant entre 10-100 milliards, forment un réseau interconnecté complexe (Tarek ,2004).

1.1.1. Une morphologie particulière

Le neurone est une cellule spécialisée dans le traitement et la transmission de l'information, ce qui est reflété par sa morphologie très particulière (Fig.1). Il se compose d'un corps cellulaire, le soma, et d'expansions : les dendrites d'une part, et l'axone d'autre part. Les dendrites forment des ramifications qui rentrent en contact avec d'autres neurones, typiquement de l'ordre de 10000, et dont le rôle est de recevoir des informations, électriques ou chimiques. L'axone est un prolongement de la cellule, typiquement long de quelques millimètres, qui conduit un signal électrique jusqu'à son arborisation terminale, où il peut alors entrer en contact avec les dendrites d'autres neurones. La jonction axone-dendrite est appelée synapse (Romain, 2003).

Si ce schéma s'applique globalement à la majorité des cellules nerveuses, il faut cependant noter qu'il existe une très grande variété de neurones dans le cerveau. Ainsi, certains neurones (peu nombreux) n'ont pas de dendrites, d'autres n'ont pas d'axone (les cellules amacrines). La forme de l'arbre dendritique, qui a certainement une influence déterminante sur la fonction du neurone (Yuste et Tank, 1996 in Romain, 2003) peut varier énormément d'un type de neurone à l'autre.

Il existe 4 modes de communication entre neurones :

- liaison terminaison axonale-dendrite,
- liaison terminaison axonale-corps cellulaire,
- liaison axone-axone,
- liaison dendrite-dendrite.

La synapse est une sorte de bulbe : entre la surface du bulbe (bouton terminal) et celle du nouveau neurone, se trouve un espace dit **espace synaptique**.

Pour que l'impulsion nerveuse qui est un signal de nature électrique, puisse franchir cet espace, des substances électrochimiques entre en action : les **neurotransmetteurs**.

Les **neurotransmetteurs** (les plus connus sont l'acétylcholine et la noradrénine) sont normalement contenus dans le bouton terminal et, une fois libérés, ils modifient les propriétés électriques de neurones suivant, faisant ainsi repartir le signal électrique.

L'action des neurotransmetteurs est extrêmement brève pour empêcher que la stimulation nerveuse ne se prolonge au-delà du temps nécessaire au transfert de l'impulsion nerveuse.

1.1.2. Quelques définitions

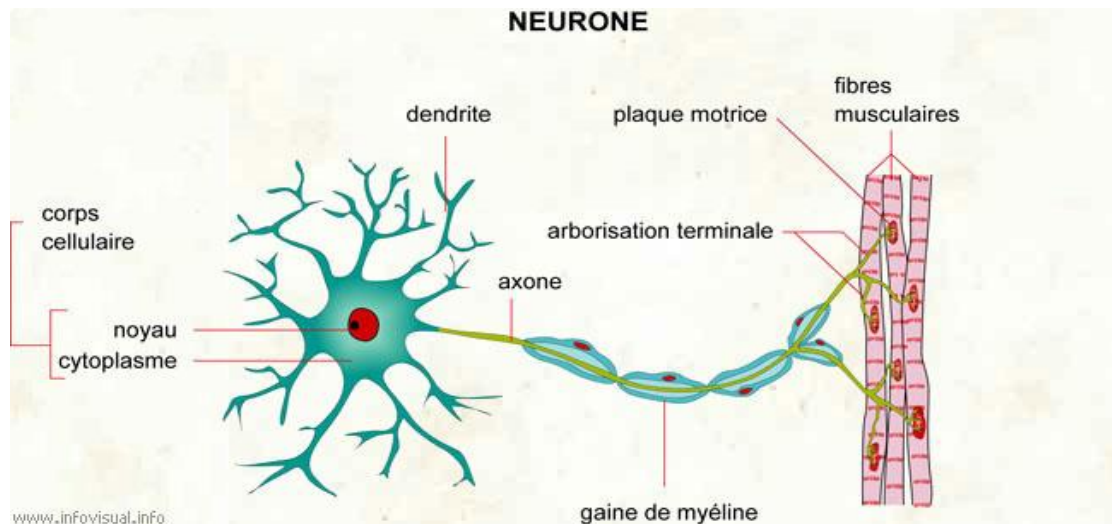


Fig. 1 - Schéma détaillée d'un neurone biologique.

Neurone: Cellule constituée d'un corps cellulaire, d'un axone et de dendrites composant l'unité fonctionnelle du système nerveux.

Dendrite: Extension du cytoplasme de la cellule nerveuse.

Axone: Prolongement du neurone.

Arborisation terminale: Ramification finale.

Plaque motrice: Partie permettant la contraction, le mouvement.

Fibres musculaires: Ensemble de cellules constituant le muscle.

Gaine de myéline: Enveloppe de graisses phosphorées.

Cytoplasme: Gelée contenue dans la cellule.

Noyau: Partie centrale d'une cellule, contenant les chromosomes.

Corps cellulaire: Partie de la cellule nerveuse située entre l'axone et la dendrite.

1.1.3. Principe de base du Neurone Biologique

Dans le cerveau, les neurones sont reliés entre eux par l'intermédiaire d'axones et de dendrites. En première approche. Nous pouvons considérer que ces sortes de filaments sont conducteurs d'électricité et peuvent ainsi véhiculer des messages depuis un neurone vers un autre. Les dendrites représentent les entrées du neurone et son axone sa sortie. Un neurone émet un signal en fonction des signaux qui lui proviennent des autres neurones. On observe en fait au niveau d'un neurone, une intégration des signaux reçus au

cours du temps, c'est à dire une sorte de sommations des signaux. En général, quand la somme dépasse un certain seuil, le neurone émet à son tour un signal électrique.

La notion de synapse explique la transmission des signaux entre un axone et une dendrite. Au niveau de la jonction (c'est à dire de la synapse), il existe un espace vide à travers lequel le signal électrique ne peut pas se propager. La transmission se fait alors par l'intermédiaire de substances chimiques, les neuromédiateurs. Quand un signal arrive au niveau de la synapse, il provoque l'émission de neuromédiateurs qui vont se fixer sur des récepteurs de l'autre côté de l'espace inter-synaptique. Quand suffisamment de molécules se sont fixées, un signal électrique est émis de l'autre côté et on a donc une transmission. En fait, suivant le type de la synapse, l'activité d'un neurone peut renforcer ou diminuer l'activité de ces voisins. On parle ainsi de synapse excitatrice ou inhibitrice.

La géométrie des neurones varie suivant leurs fonctionnalités (et donc selon la partie du système nerveux auquel ils appartiennent). On peut les différencier suivant (Matthieu , 2003).

- ❖ le nombre de neurites (arbres dendritique et axonique), ce qui distingue les neurones monopolaires, bipolaires et multipolaires (Fig. 2).
- ❖ la taille des neurites (longs axones pour les nerfs) et leur arborescence plus ou moins touffue (quelques dizaines de ramifications pour les dendrites de neurones pyramidaux contre plusieurs centaines pour celles des cellules stellaires dans le cortex visuel par exemple).
- ❖ la présence d'« épines » sur les dendrites (sur les cellules pyramidales dans le cortex par exemple), ce qui augmente la surface supportant les récepteurs des synapses et que l'on suppose intervenir dans des phénomènes d'apprentissage et d'auto-adaptation.
- ❖ la myélinisation de l'axone : des cellules gliales de Schwann enroulées autour de l'axone forme une gaine multicouche qui accélère la conduction des PA (présent principalement sur les longs axones formant des nerfs).

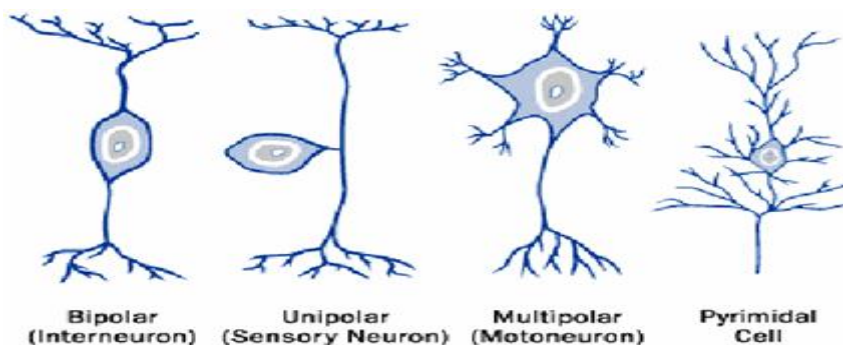


Fig. 2 - Différentes géométries de neurones <http://static.howtsu.ffwork.com>

a) Le potentiel de repos

Lorsqu'un neurone est au repos, on constate qu'il existe une différence de potentiel, entre l'intérieur et l'extérieur de la cellule. On dit alors que le neurone est polarisé. Cela est dû à une différence de concentration ionique entre les milieux intracellulaire et extracellulaire, et à une perméabilité sélective de la membrane à certaines espèces d'ions. Les différences de concentration sont maintenues par des protéines de la membrane appelées pompes à ions, qui transportent des ions à l'encontre de leur gradient de concentration. D'autres protéines, les canaux ioniques, laissent passer spécifiquement certains ions (Romain, 2003).

b) Le potentiel d'action

Lorsque l'on injecte un courant assez faible dans le neurone, son potentiel se modifie en proportion du courant. Cependant, si le courant est suffisamment fort, on constate une montée brutale du potentiel (dépolariation), suivie d'une redescende rapide (hyperpolarisation) : on appelle ce phénomène le potentiel d'action. Cela signifie qu'au-dessus d'un certain seuil, la perméabilité de la membrane change (Romain, 2003).

1.1.4. Simulation

La résolution analytique des équations non-linéaires couplées est un tâche ardue. La préférence se porte sur les modèles compartimentés représentant indépendamment mais très fidèlement des régions distinctes du neurone. En dépit de l'extrême complexité des processus biologiques qui s'accomplissent dans le cerveau humain, des modèles rudimentaires ont été échafaudés (Tarek ; 2004).

1.2. Les Réseaux de Neurones Artificiels

1.2.1. Définitions et Plate-Forme de Comparaison

Les réseaux de neurones sont une des composantes importantes du domaine de l'intelligence artificielle. Les réseaux de neurones « artificiels », pour les différencier des réseaux de neurones biologiques, sont composés de neurones artificiels simples, petites fonctions mathématiques, qui permettent, montés en réseau de former des fonctions complexes très utiles. Par analogie aux neurones biologiques, les neurones artificiels ont pour but de reproduire des raisonnements « intelligents » d'une manière artificielle.

Ces neurones peuvent adopter de certaines qualités habituellement propres au biologique, c'est-à-dire, la généralisation, l'évolutivité, et une certaine forme de déduction. La structure d'un neurone artificiel est en fait copiée sur la structure des neurones biologiques. Les principales structures biologiques des neurones ont toutes leur équivalent artificiel, ceci ayant pour but de reproduire leur fonctionnement de la meilleure façon possible (d'une manière logique, simple et facilement représentable sur informatique).

a) Définitions

Un neurone est une fonction non linéaire, paramétrée, à valeurs bornées. (Dreyfus , 2004) Cette représentation est le reflet de l'inspiration biologique qui a été à l'origine de la première vague d'intérêt pour les neurones formels, dans les années 1940 à 1970 (Mcculloch et Pitts , 1943; Minsky et Papert ,1969).

Un neurone est une minuscule structure qui traite les influx nerveux qui arrivent (*inputs*), chacun selon son importance relative, et qui émet un (et un seul) signal de sortie (*output*). Les neurones artificiels reproduisent le même procédé, recevant chaque signal d'entrée (*input*) tempéré par un poids (*weight*). Ces poids sont aussi appelés poids synaptiques, par analogie (Tableau. 1).

Les intrants pondérés, habituellement (mais pas toujours) sommés, sont ensuite comparés à un seuil d'activation et passés dans la fonction du neurone (habituellement une fonction très simple), qui produit l'extrait (*output*) désiré. Dans les neurones biologiques, les intrants et extrants sont des influx électriques, représentés artificiellement

par des valeurs numériques. Habituellement, les neurones sont reliés ensemble en réseaux, les neurones d'un niveau (les réseaux de neurones sont habituellement bâtis en niveaux hiérarchiques) passant leurs extrants aux neurones du niveau suivant (devenant les intrants des neurones suivants).

D'une manière générale, un neurone constitue un minuscule processeur, capable de gérer une fonction simple.

Tableau. 1 : Analogie entre les neurones biologiques et artificiels

Neurone biologique	Neurone artificiel	Représentation symbolique
Soma	Neurone	
Dendrite	Intrant (<i>Input</i>)	x_i, X (intrant net)
Axone	Extrant (<i>Output</i>)	Y
Synapse	Poids	w_i

Aujourd'hui de nombreux termes sont utilisés dans la littérature pour désigner le domaine des réseaux de neurones artificiels, comme connexionisme ou neuromimétique... Ainsi, les réseaux de neurones artificiels ne désignent que les modèles manipulés ; ce n'est ni un domaine de recherche, ni une discipline scientifique (Touzet, 1992).

Connexionisme et neuromimétique sont tous deux des domaines de recherche à part entière, qui manipulent chacun des modèles de réseaux de neurones artificiels, mais avec des objectifs différents. L'objectif poursuivi par les ingénieurs et chercheurs connexionnistes est d'améliorer les capacités de l'informatique en utilisant des modèles aux composants fortement connectés. Pour leur part, les neuromiméticiens manipulent des modèles de réseaux de neurones artificiels dans l'unique but de vérifier leurs théories biologiques du fonctionnement du système nerveux central (Touzet, 1992).

1.2.2. Historique et Chronologie

a) Historique

Traditionnellement, on fait remonter les travaux sur les réseaux de neurones à ceux de (MC Culloch et Pitts, 1943), les auteurs essaient de comprendre les propriétés du

système nerveux à partir de ses constituants élémentaires : les neurones. Ils sont idéalisés en neurones logiques, car ne connaissent que les réponses binaires (0 ou 1). Ce qui à l'époque était neurologiquement correct/plausible (Touzet, 1992). Chaque neurone est connecté à d'autres neurones, les connexions pouvant être soit inhibitrices, soit excitatrices. Périodiquement, le neurone calcule son activation, si l'activation est supérieure à un certain seuil, le neurone est actif sinon il est passif. Le résultat important est qu'un tel mécanisme permet de calculer n'importe quelle fonction logique, ce qui les rend équivalents aux machines-de-Turing-Universelles.

Un peu plus tard, (Donal Hebb, 1949), psychologue canadien, tentait d'expliquer les effets d'apprentissage, de mémoire et de conditionnement à partir d'un groupe de cellules. Pour expliquer les effets d'apprentissage en fonction de l'expérience, il propose que les cellules apprennent à modifier l'intensité des connexions qui les relient en fonction de leur activité simultanée. C'est la "loi de Hebb" (déjà trouvée par d'autres auteurs), l'approche de Hebb vaut par l'utilisation de cette notion d'un point de vue théorique pour expliquer des phénomènes psychologiques et prédire leurs effets.

(Minsky, 1967) construit ce qui fut le premier modèle réel de réseaux de neurones *Snalr* qui fut l'ancêtre de *Mark I* le premier perceptron créé par Rosenblatt dans les années 1960. Le but du perceptron est d'apprendre des catégorisations perceptives. Dans la même période Widrow, dans le cadre du traitement du signal, analyse une loi d'apprentissage proche de celle du perceptron (loi de Widrow-Hoff) et construit *Adaline* qui sera utilisée en version multiple (*Madaline*) pour apprendre à atténuer le bruit des lignes téléphoniques.

b) Chronologie

- (1943 : **J. MC Culloch et W. Pitts**) laissent leurs noms à une modélisation du neurone biologique (un neurone au comportement binaire). Ceux sont les premiers à montrer que des réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes (tout au moins au niveau théorique).

- (1949 : **D. Hebb**), physiologiste américain explique le conditionnement chez l'animal par les propriétés des neurones eux-mêmes. Ainsi, un conditionnement de type pavlovien tel que, nourrir tous les jours à la même heure un chien, entraîne chez cet animal la sécrétion de salive à cette heure précise même en l'absence de nourriture. La loi de

modification des propriétés des connexions entre neurones qu'il propose explique en partie ce type de résultats expérimentaux.

-(1956 : **Von Neumann, in Tarek ,2004**), envisageait déjà la possibilité d'un réseau fiable composé de composants non fiables par l'utilisation de la redondance.

-(1962 : **Rosenblatt**), la naissance d'une méthode analytique rigoureuse d'adaptation de poids au sein d'un modèle multi-couches appelée le perceptron.

Rosenblatt démontra la convergence d'un algorithme itératif d'adaptation des poids pour la classe élémentaire des perceptrons mono-couche.

-(1969 : **Minsky et Papert**), Minsky et Papert ont apporté un éclairage théorique sur la structure des problèmes qu'il était impossible de résoudre avec le perceptron.

Problème de OU-exclusif (XOR). Abandonner l'approche connexionniste durant près de 20 ans aux profits des Systèmes experts.

Les années 1970-1986: Les chercheurs relancent les travaux sur les Réseaux de Neurones:

– réalisation d'une mémoire associative, – un formalisme plus riche a été développé par Hopfield (Hopfield, 1982; Hinton et Sejnowski , 1986; Peretto ,1984).

- En 1974 Werbos (Werbos 1974) a inventée l'idée de la rétro-propagation.

Destiné aux réseaux avec des connexions non-bouclées, l'algorithme de rétro propagation a été développé indépendamment en 1985 par Rumelhart (Rumelhart, 1986), Parker (Parker, 1985) et Le Cun (Le Cun ,1985).

c) Les premiers succès

- (**Rosenblatt, 1962**) développe le modèle du Perceptron. Il construit le premier neuro-ordinateur basé sur ce modèle et l'applique au domaine de la reconnaissance de formes. Notons qu'à cet époque les moyens à sa disposition sont limités et c'est une prouesse technologique que de réussir à faire fonctionner correctement cette machine plus de quelques minutes.

- (**Widrow, 1960**), un automaticien, développe le modèle Adaline (Adaptative Linear Element). Dans sa structure, le modèle ressemble au Perceptron, cependant la loi d'apprentissage est différente. Celle-ci est à l'origine de l'algorithme de rétro propagation de gradient très utilisé aujourd'hui avec les Perceptrons multicouches. Widrow a créé dès

cette époque une des premières firmes proposant neuro-ordinateurs et neuro-composants, la "Memistor Corporation". Il est aujourd'hui le président de l'International Neural Network Society (INNS) sur laquelle nous reviendrons au chapitre Informations pratiques.

- (**M. Minsky et S. Papert, 1969**) publient un ouvrage qui met en exergue les limitations théoriques du perceptron. Limitations alors connues, notamment concernant l'impossibilité de traiter par ce modèle des problèmes non linéaires. Ils étendent implicitement ces limitations à tous modèles de réseaux de neurones artificiels. Leur objectif est atteint, il y a abandon financier des recherches dans le domaine (surtout aux U.S.A.), les chercheurs se tournent principalement vers l'IA et les systèmes à bases de règles.

d) L'ombre

- **1967-1982** : Toutes les recherches ne sont, bien sûr, pas interrompues. Elles se poursuivent, mais déguisées, sous le couvert de divers domaines comme : le traitement adaptatif du signal, la reconnaissance de formes, la modélisation en neurobiologie, etc. De grands noms travaillent durant cette période telle : (S. Grossberg, T. Kohonen).

e) Le renouveau

-(**Hopfield, 1982**) est un physicien reconnu à qui l'on doit le renouveau d'intérêt pour les réseaux de neurones artificiels. A cela plusieurs raisons :

Au travers d'un article court, clair et bien écrit, il présente une théorie du fonctionnement et des possibilités des réseaux de neurones. Il faut remarquer la présentation anticonformiste de son article. Alors que les auteurs s'acharnent jusqu'alors à proposer une structure et une loi d'apprentissage, puis à étudier les propriétés émergentes ; J. J. Hopfield fixe préalablement le comportement à atteindre pour son modèle et construit à partir de là, la structure et la loi d'apprentissage correspondant au résultat escompté. Ce modèle est aujourd'hui encore très utilisé pour des problèmes d'optimisation.

D'autre part, entre les mains de ce physicien distingué, la théorie des réseaux de neurones devient respectable. Elle n'est plus l'apanage d'un certain nombre de psychologues et neurobiologistes hors du coup (Touzet, 1992).

1.2.3. Le Modèle Neurophysiologique

Le cerveau se compose d'environ 10^{12} neurones (mille milliards), avec 1000 à 10000 synapses (connexions) par neurone. Nous allons décrire succinctement l'élément de base du système nerveux central : le neurone. L'étape suivante nous conduit à l'étude de petits réseaux de neurones. Ceci nous amène à exposer les propriétés d'habituation, de sensibilisation et surtout à concevoir l'idée d'une modification physique des connexions entre neurones pour supporter ces phénomènes. L'étude du mécanisme de la vision chez l'animal (et l'homme) permet d'appréhender les notions de messages somato-sensoriels, de réduction d'information, d'étages de traitement et de complexification de l'information. (Touzet, 1992).

a) Le Modèle Neurone Formel

C'est dans l'article " A logical Calculus of the ideas immanent in nervous activity paru dans le journal Bulletin of mathematical en 1943 qu'est apparu pour la première fois la notion de neurone formel (McCulloch et Pitts , 1943).

A l'époque, Mc Culloch étudie la strychnisation du cerveau des chimpanzés. Il se convainc que les circuits d'influx nerveux étaient analogues à la propagation logique d'un raisonnement.

- Il en déduit la notion de psychon
- Evénement psychique.
- Tout ou rien
- Implique son antécédent temporel et intervient dans la détermination des psychons subséquents.

Avec l'aide de Pitts, il va présenter le neurone formel.

- Il est binaire : il est soit actif soit inactif.
- chaque neurone possède un seuil de déclenchement fixé.

Le neurone reçoit des entrées de synapse excitatrices, ayant des poids identiques ; le neurone peut aussi recevoir des entrées de synapses inhibitrices, dont l'action est absolue ; c'est à dire que, si la synapse est active, le neurone ne peut pas être actif.

Un quantum de temps est nécessaire à l'intégration des entrées synaptiques, basé sur le délai observé physiologiquement.

Le fonctionnement est simple si aucune synapse inhibitrice n'est active, le neurone ajoute ses entrées synaptiques et vérifie que la somme atteint ou excède son seuil. On dit alors que le neurone devient actif, sinon le neurone est inactif (Fig. 3).

Voyons tout cela en détail sur un exemple. Le neurone reçoit les entrées $x_1, x_2, x_3 \dots x_n$.

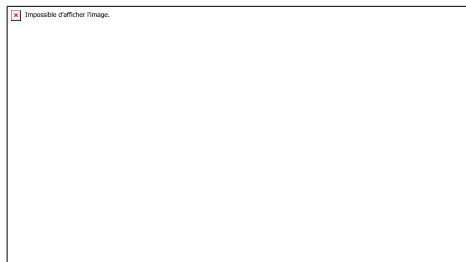


Fig. 3 - le neurone formel.

x_1, x_2, x_3, x_n sont les entrées, ou encore les excitations.

w_1, w_2, w_3 sont appelés les coefficients synaptiques ou poids synaptiques. De nos jours, nous utilisons des poids différents pour $w_1, w_2, w_3 \dots$. Les poids peuvent même être "corrigés" dans le cas de l'apprentissage.

le potentiel p est donné par $p = x_1 * w_1 + x_2 * w_2 + x_3 * w_3 \dots + x_n * w_n$

La fonction de sortie est $s = f(p)$ est la fonction de transfert.

Le neurone créé par McCulloch et Pitts était un neurone utilisant une fonction binaire (Fig. 4). C'est-à-dire si $p > a$ alors $S = 1$ sinon $S = -1$.

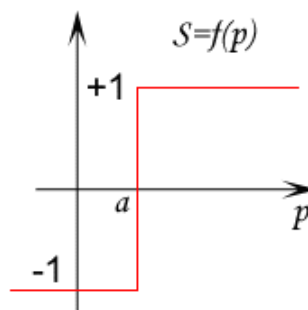


Fig. 4 - fonction de transfert binaire.

De nombreuses autres fonctions sont utilisées en particulier les fonctions sigmoïdes. Nous pouvons ensuite créer un réseau suivant les besoins.

Pour fonctionner, un neurone formel utilise des *entrées* qui sont des grandeurs réelles. Si on "relie" un neurone au monde extérieur par l'intermédiaire de capteurs, il peut réaliser

une simple analyse de ce qu'il "perçoit". Si on représente les valeurs observées par le neurone sous forme d'un vecteur, le neurone réalise alors un découpage de son espace d'entrée (l'espace vectoriel auquel appartient le vecteur d'observation) en deux zones : la zone *d'activité* dont les vecteurs donnent une sortie égale à 1 et la zone *d'inactivité* dont les vecteurs donnent une sortie égale à 0. Comme le calcul effectué est en fait linéaire, la séparation l'est aussi. Les coefficients synaptiques et le seuil définissent l'équation d'un hyperplan qui est la frontière de la séparation entre les deux zones.

Globalement un réseau de neurones formels est un graphe (orienté ou pas) dont les sommets (neurones) ont une capacité (restreinte) à transformer un signal d'entrée en un signal de sortie, (Corsini, 2005). Les connexions sont pondérées et servent à transférer les signaux d'un (ou plusieurs) sommet(s) vers un (ou plusieurs) autre(s). Par ailleurs ces connexions peuvent être équipées de fonction de *retard* permettant ainsi une plus grande souplesse dans le modèle.

Olivier Sarzeaud (Sarzeaud, 1995 in Corsini, 2005) propose une description différente, en considérant qu'un réseau de neurones est un graphe bi-parties, avec deux ensembles de sommets disjoints (d'une part les neurones formels, d'autre part les connexions). Cette approche offre une plus grande puissance formelle au modèle mais complexifie énormément la description, cependant elle semble plus adaptée à une modélisation objet d'un réseau de neurones. On pourra aussi considérer le travail de E. Fiesler (Fiesler, 1994 in Corsini, 2005) à qui propose une taxonomie topologique ainsi qu'une nomenclature uniforme accompagnée d'une proposition de notation.

b) Les Modèles Mathématiques

Les réseaux de neurones biologiques réalisent facilement un certain nombre d'applications telles que la reconnaissance de formes, le traitement du signal, l'apprentissage par l'exemple, la mémorisation, la généralisation. Ces applications sont pourtant, malgré tous les efforts déployés en algorithmique et en intelligence artificielle, à la limite des possibilités actuelles. C'est à partir de l'hypothèse que le comportement intelligent émerge de la structure et du comportement des éléments de base du cerveau que les réseaux de neurones artificiels se sont développés (Touzet, 1992). Les réseaux de neurones artificiels sont des modèles, à ce titre ils peuvent être décrits par leurs composants, leurs variables descriptives et les interactions des composants.

Les réseaux de neurones artificiels ou réseaux connexionnistes sont fondés sur des modèles qui tentent d'expliquer comment les cellules du cerveau et leurs interconnexions parviennent, d'un point de vue global, à exécuter des calculs complexes.

1.2.4. Le Neurone Artificiel

a) Structure

La figure ci-dessous (Fig. 5) montre la structure d'un neurone artificiel. Chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance des neurone amonts. A chacune de ces entrées est associé un poids w (abréviation de weight (poids en anglais) représentatif de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones aval. A chaque connexion est associé un poids (Touzet, 1992).

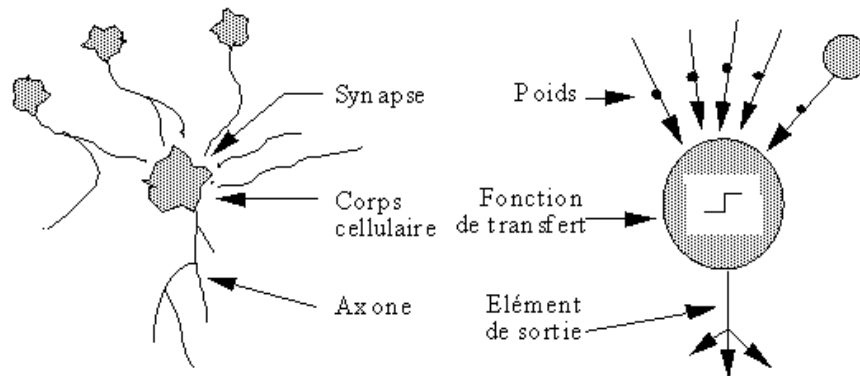


Fig. 5 - Mise en correspondance neurone biologique / neurone artificiel.

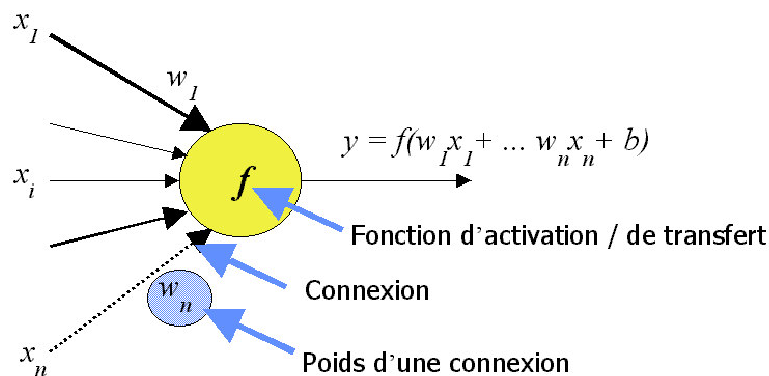


Fig. 6 - Modélisation simplifiée d'un neurone Artificiel.

Pour le neurone d'indice i , les entrées sur celui-ci sont de poids w_{ij} alors que les connexions aval sont de poids w_{ji} .










b) Comportement et fonction de transfert

On distingue deux phases. La première est habituellement le calcul de la somme pondérée des entrées (n) selon l'expression suivante :

$$n = \sum (w_i \cdot e_i) \tag{1}$$

A partir de cette valeur, une fonction de transfert calcule la valeur de l'état du neurone. C'est cette valeur qui sera transmise aux neurones aval. Il existe de nombreuses formes possibles pour la fonction de transfert. Les plus courantes sont présentées sur le (Tableau. 2). On remarquera qu'à la différence des neurones biologiques dont l'état est binaire, la plupart des fonctions de transfert sont continues, offrant une infinité de valeurs possibles comprises dans l'intervalle [0, +1] (ou [-1, +1]).

Tableau. 2 : Différents types de fonctions de transfert pour le neurone artificiel (Mar ,2004).

Nom de la fonction	Relation d'entrée –sortie	Icône
1. Seuil	a = 0 si n < 0 a = 1 si n ≥ 0	
2. Seuil symétrique	a = -1 si n < 0 a = +1 si n ≥ 0	
3. Linéaire	a = n	
4. Linéaire saturé	a = 0 si n < 0 a = n si 0 ≤ n ≤ 1 a = 1 si n > 1	
5. Linéaire saturé symétrique	a = -1 si n < -1 a = n si -1 ≤ n ≤ 1 a = 1 si n > 1	
6. Linéaire positive	a = 0 si n < 0 a = n si n ≥ 0	
7. Sigmoide	a = 1/[1 + exp ⁻ⁿ]	
8. Tangente hyperbolique	a = [e ⁿ - e ⁻ⁿ] / [e ⁿ + e ⁻ⁿ]	
9. Compétitive	a = 0 si n maximum a = 1 si autrement	

Nous constatons que les équations décrivant le comportement des neurones artificiels n'introduisent pas la notion de temps. En effet, et c'est le cas pour la plupart des

modèles actuels de réseaux de neurones, nous avons affaire à des modèles à temps discret, synchrone, dont le comportement des composants ne varie pas dans le temps

c) Variables Descriptives

Ces variables décrivent l'état du système. Dans le cas des réseaux de neurones qui sont des systèmes non autonomes, un sous-ensemble des variables descriptives est constitué par les variables d'entrée, variables dont la valeur est déterminée extérieurement au modèle (Touzet, 1992), (Fig. 7).

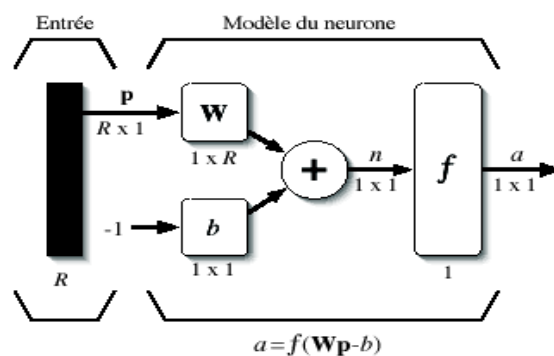


Fig. 7 - Représentation matricielle du modèle d'un neurone artificielle.

Un réseau de neurones se compose de neurones qui sont interconnectés de façon à ce que la sortie d'un neurone puisse être l'entrée d'un ou plusieurs autres neurones. Ensuite il y a des entrées de l'extérieur et des sorties vers l'extérieur (Rumelhart et *al.*, 1988 in Albert, 1992).

(Rumelhart et *al.*) donnent huit composants principaux d'un réseau de neurones (Rumelhart et *al.*, 1986 in Albert, 1992)

- Un *ensemble de neurones*
- Un *état d'activation* pour chaque neurone (actif, inactif, ...)
- Une *fonction de sortie* pour chaque neurone ($f(S)$)
- Un *modèle de connectivité* entre les neurones (chaque neurone est connecté à tous les autres, par exemple)
- Une *règle de propagation* pour propager les valeurs d'entrée à travers le réseau vers les sorties.
- Une *règle d'activation* pour combiner les entrées d'un neurone (très souvent une somme pondérée)
- Une *règle d'apprentissage*

- Un *environnement* d'opération (le système d'exploitation, par exemple)

d) Structure d'Interconnexion

Les connexions entre les neurones qui composent le réseau décrivent la topologie du modèle. Elle peut être quelconque, mais le plus souvent il est possible de distinguer une certaine régularité.

❖ **Réseau multicouche (au singulier)** : les neurones sont arrangés par couche. Il n'y a pas de connexion entre neurones d'une même couche et les connexions ne se font qu'avec les neurones des couches avales (Fig.8). Habituellement, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante et celle-ci seulement. Ceci nous permet d'introduire la notion de sens de parcours de l'information (de l'activation) au sein d'un réseau et donc définir les concepts de neurone d'entrée, neurone de sortie. Par extension, on appelle couche d'entrée l'ensemble des neurones d'entrée, couche de sortie l'ensemble des neurones de sortie. Les couches intermédiaires n'ayant aucun contact avec l'extérieur sont appelés couches cachées.(Touzet,1992)

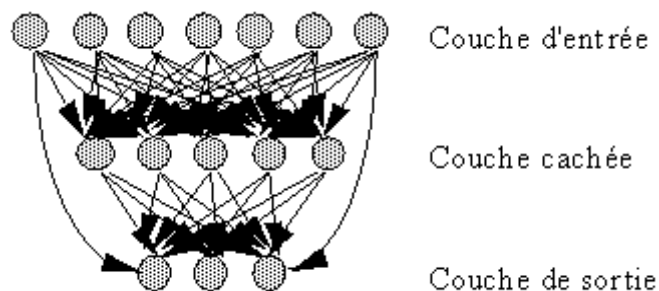


Fig. 8- Définition des couches d'un réseau multicouche.

❖ **Réseau à connexions locales** : Il s'agit d'une structure multicouche, mais qui à l'image de la rétine, conserve une certaine topologie. Chaque neurone entretient des relations avec un nombre réduit et localisé de neurones de la couche avale (Fig. 9) Les connexions sont donc moins nombreuses que dans le cas d'un réseau multicouche classique (Touzet, 1992).

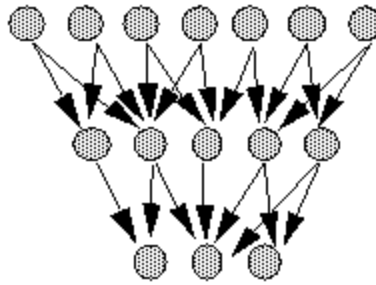


Fig. 9 - Réseau à connexions locales.

❖ **Réseau à connexions récurrentes** : les connexions récurrentes ramènent l'information en arrière par rapport au sens de propagation défini dans un réseau multicouche. Ces connexions sont le plus souvent locales. (Fig. 10) (Touzet, 1992).

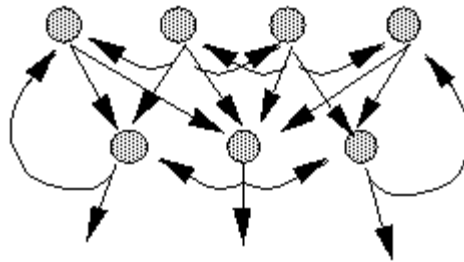


Fig. 10 - Réseau à connexions récurrentes.

❖ **Réseau à connexion complète** : c'est la structure d'interconnexion la plus générale (Fig. 11) Chaque neurone est connecté à tous les neurones du réseau (et à lui-même), (Touzet, 1992).

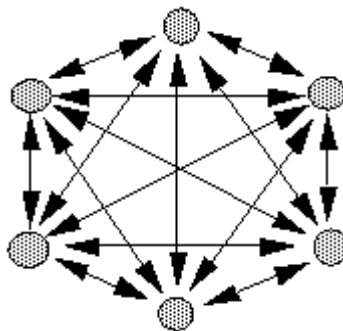


Fig. 11 - Réseau à connexions complète.

Il existe de nombreuses autres topologies possibles, mais elles n'ont pas eu à ce jour la notoriété des quelques unes que nous avons décrites ici.

e) Architecture

L'architecture générale des réseaux de neurones consiste en la représentation des neurones en couches (*layers*) successives, la première représentant la couche d'intrants (*input layer*), appelé souvent la **Rétine** (Coursini,2005), la dernière étant la couche de sortie (*output layer*), les couches intermédiaies étant les couches cachées (*hidden layers*) du réseau. Ces couches sont dites cachées car de l'extérieur du réseau, on ne peut analyser clairement leur fonctionnement. On ne connaît vraiment que les signaux intrants et extrants du réseau. Les neurones de la couche d'intrants ne sont pas vraiment des neurones traitants (*computing neurons*), mais ont pour seule utilité de normaliser l'entrées des signaux ainsi que la distribution des signaux intrants (Fig. 12).

Dans cette architecture normalisée, les couches de neurones sont totalement interconnectées, c'est-à-dire que les neurones d'une couche sont tous reliés à tous les neurones des couches adjacentes. Cette architecture normalisée peut sembler rigide, mais elle permet une représentation correcte de la plupart des réseaux de neurones, tout en permettant l'utilisation d'algorithmes d'entraînement plus généraux.

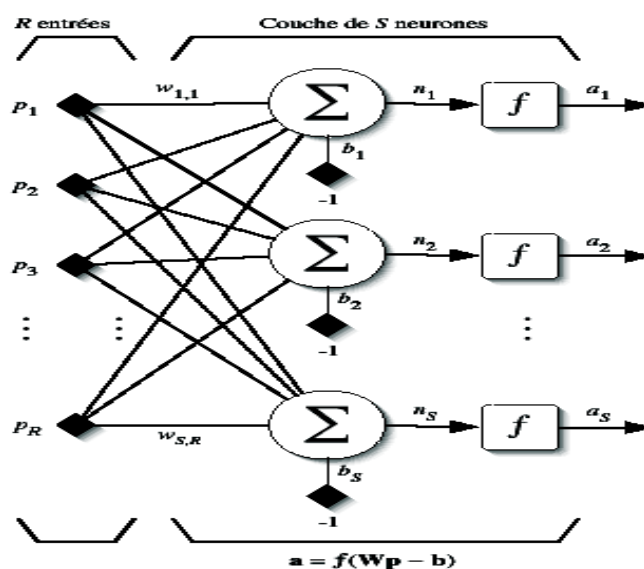


Fig. 12 - couche de S neurones.

D'une manière générale, un réseau de neurones se comporte, d'un point de vue extérieur, comme une fonction S qui traite des données (*inputs*) et produit une réponse correspondante (*output*). Les données entrées peuvent être de n'importe quel type représentable de manière binaire ou numérique. Ces données peuvent également être vues comme des vecteurs, et le réseau de neurones une application vectorielle (Fig. 13).

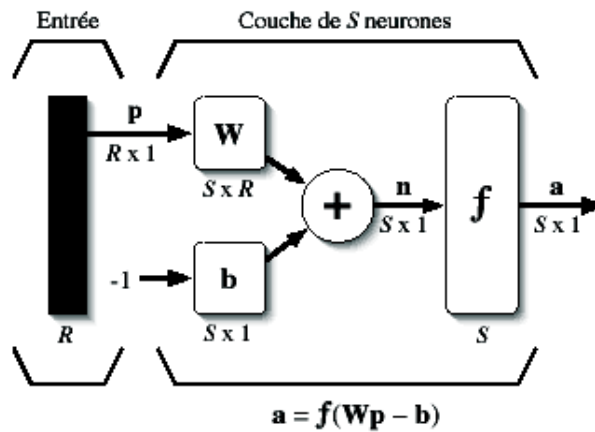


Fig. 13 - Représentation matricielle d'une couche de S Neurones.

Les réseaux de neurones peuvent être soit statiques, soit dynamique, Les réseaux dynamiques ont la capacité de s'actualiser (modifier les poids des liaisons) en tout temps. Ces réseaux possèdent habituellement des algorithmes d'entraînements non-supervisés ce qui permet au réseau en question d'« évoluer » à chaque utilisation. Les réseaux statiques se séparent en deux catégories :

- La première concerne les réseaux entraînés, c'est-à-dire les réseaux qui ont vu leurs poids modifiés par un algorithme d'entraînement, sous la supervision d'un instructeur jusqu'à ce que ces réseaux donnent des résultats satisfaisants aux données d'entraînement.
- La seconde catégorie concerne les réseaux non-évolutifs, comme les réseaux de mémoire. Les poids de ces réseaux sont choisis (et non pas modifiés par un entraînement), pour donner l'effet escompté (Fig. 14).

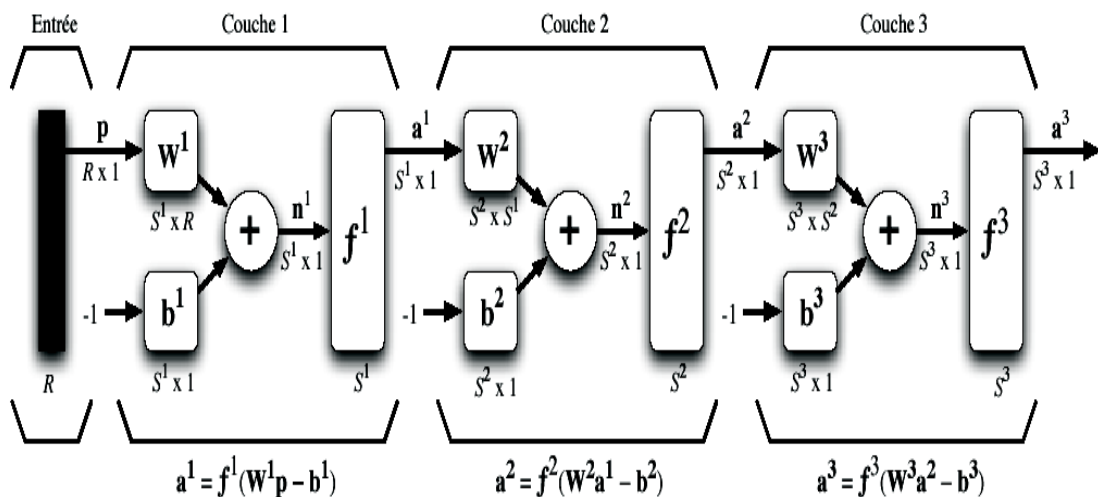


Fig. 14 - Représentation matricielle d'un réseau de trois couches.

f) Relation entre les réseaux de neurones et la statistique

De nombreux auteurs ont remarqué qu'il existait des relations étroites entre les modèles neuronaux et la statistique (voir par exemple dans (Hertz et al, 1991 in Coulibaly et al 1999 ; Hornik, 1989).

Ces relations sont de diverse nature. Dans certains cas, les modèles neuronaux réalisent des tâches traditionnelles de la statistique, comme la classification (perceptron, modèle compétitif), la régression et la prévision (perceptron multicouches). D'autre part, les chercheurs ont proposé des méthodes connexionistes qui permettent de réaliser, par des méthodes itératives, des analyses de données "classiques" telles que l'analyse en composantes principales (Oja, 1982 ; in Coulibaly et al, 1999) et l'analyse discriminante.

Dans ce cadre, on s'intéresse plus particulièrement à la comparaison de différentes méthodes statistiques et des réseaux neuronaux. Ainsi, le seul réseau de neurones utilisé est le réseau multicouche (Tableau 3).

Tableau : 3 - Comparaison entre les termes utilisés en statistique et en RNA.

Réseaux de neurones	Statistiques
1. apprentissage	estimation
2. poids	paramètres
3. connaissance	valeur des paramètres
4. apprentissage supervisé	régression
5. classification	discrimination
6. apprentissage non supervisé	estimation de densité
7. clustering	classification
8. réseau de neurone	modèle
9. grand : 100 000 poids	grand : 50 paramètres
10. ensemble d'apprentissage	échantillon
11. grand : 500 000 exemples	grand : 200 cas

g) Application

Les réseaux de neurones servent aujourd'hui à toutes sortes d'applications dans divers domaines.

Par exemple, on a développée un auto-pilote pour avion, ou encore un système de guidage pour automobile, on a conçu des systèmes de lecture automatique de chèques bancaires et d'adresses postales, on produit des systèmes de traitement du signal pour différentes applications militaires, un système pour la synthèse de la parole, des réseaux sont aussi utilisés pour bâtir des systèmes de vision par ordinateur, pour faire des prévisions sur les marchés monétaires, pour évaluer le risque financier ou en assurance, pour différents processus manufacturiers, pour le diagnostic médical, pour l'exploration pétrolière ou gazière, en robotique, en télécommunication, en prévision hydrologique, en modélisation pluie débit et en particulier la prévision des débits des cours d'eau. Bref, les réseaux de neurones ont aujourd'hui un impact considérable et, il y a fort à parier, que leur importance ira grandissant dans le futur.

On peut en résumé les applications des réseaux de neurones comme suite (Dreyfus, 2004) :

- Une application en reconnaissance de formes : la lecture automatique de codes postaux
- Une application en contrôle non destructif : la détection de défauts dans des rails par courants de Foucault
- Une application en prévision : l'estimation de la probabilité de succès aux élections législatives
- Une application en fouille de données : le filtrage de documents
- Une application en bio-ingénierie : relations structure-activité (QSAR) pour la prédiction de propriétés chimiques de molécules
- Une application en formulation : la prédiction de la température de liquides de verres
- Une application en robotique mobile : le pilotage automatique d'un véhicule autonome.

1.2.5. Processus d'Apprentissage des Réseaux de Neurones

a) Définition

Parmi les propriétés désirables pour un réseau de neurones, la plus fondamentale est sûrement la capacité d'apprendre de son environnement, d'améliorer sa performance à travers un processus d'apprentissage. Mais qu'est-ce donc que l'apprentissage ? Malheureusement, il n'existe pas de définition générale, universellement acceptée, car ce concept touche à trop de notions distinctes qui dépendent du point de vue que l'on adopte.

L'apprentissage est un processus dynamique et itératif permettant de modifier les paramètres d'un réseau en réaction avec les stimuli qu'il reçoit de son environnement. Le type d'apprentissage est déterminé par la manière dont les changements de paramètre surviennent.

Cette définition implique qu'un réseau se doit d'être stimulé par un environnement, qu'il subisse des changements en réaction avec cette stimulation, et que ceux-ci provoquent dans le futur une réponse nouvelle vis-à-vis de l'environnement. Ainsi, le réseau peut s'améliorer avec le temps.

L'objectif de l'apprentissage (Hinton, 1989), est de fournir une méthode au réseau afin qu'il puisse ajuster ces paramètres lorsqu'on lui présente des exemplaires à traiter. Certains réseaux font la différence entre une phase d'entraînement et une phase d'exploitation (cas du Perceptron), ou pas (cas-du-réseau-ART).

b) Type d'Apprentissage

Dans le cas des réseaux de neurones artificiels, on ajoute souvent à la description du modèle l'algorithme d'apprentissage. Le modèle sans apprentissage présente en effet peu d'intérêt. Dans la majorité des algorithmes actuels, les variables modifiées pendant l'apprentissage sont les poids des connexions (Touzet, 1992).

L'apprentissage est la modification des poids du réseau dans l'optique d'accorder la réponse du réseau aux exemples et à l'expérience. Il est souvent impossible de décider a priori des valeurs des poids des connexions d'un réseau pour une application donnée. A l'issue de l'apprentissage, les poids sont fixés : c'est alors la phase d'utilisation. Certains modèles de réseaux sont improprement dénommés à apprentissage permanent. Dans ce

cas il est vrai que l'apprentissage ne s'arrête jamais, cependant on peut toujours distinguer une phase d'apprentissage (en fait de remise à jour du comportement) et une phase d'utilisation. Cette technique permet de conserver au réseau un comportement adapté malgré les fluctuations dans les données d'entrées.

Généralement Il existe trois types d'apprentissages principaux. Ce sont l'apprentissage supervisé, l'apprentissage non-supervisé et l'apprentissage par tentative (graded training en anglais).

Dans le cas de l'apprentissage non-supervisé le réseau décide lui-même quelles sont les bonnes sorties. Cette décision est guidée par un but interne au réseau qui exprime une configuration idéale à atteindre par rapport aux exemples introduits.

Les cartes auto-organisatrices de Kohonen sont un exemple de ce type de réseau (Kohonen, 1984).

L'apprentissage supervisé est le type d'apprentissage le plus utilisé. Pour ce type d'apprentissage la règle la plus utilisée est celle de Widrow-Hoff. D'autres règles d'apprentissage sont par exemple la règle de Hebb, la règle du perceptron, la règle de Grossberg...etc. (Rumelhart et al. 1986).

❖ L'apprentissage supervisé

L'apprentissage dit supervisé est caractérisé par la présence d'un «professeur» qui possède une connaissance approfondie de l'environnement dans lequel évolue le réseau de neurones. En pratique, les connaissances de ce professeur prennent la forme d'un ensemble de Q couples de vecteurs d'entrée et de sortie que nous noterons $\{(p_1, d_1), (p_2, d_2), \dots, (p_Q, d_Q)\}$, où p_i désigne un stimulus (entrée) et d_i la cible pour ce stimulus, c'est-à-dire les sorties désirées du réseau. Chaque couple (p_i, d_i) correspond donc à un cas d'espace de ce que le réseau devrait produire (la cible) pour un stimulus donnée. Pour cette raison, l'apprentissage supervisé est aussi qualifié d'apprentissage par des exemples (Marc, 2004).

L'apprentissage supervisé est illustré d'une manière conceptuelle à la figure (Fig. 15), l'environnement est inconnu du réseau. Celui-ci produit un stimulus p qui est acheminé à la fois au professeur et au réseau. Grâce à ses connaissances intrinsèques, le professeur produit une sortie désirée $d(t)$ pour ce stimulus. On suppose que cette réponse est optimale. Elle est ensuite comparée (par soustraction) avec la sortie du réseau pour

produire un signal d'erreur $e(t)$ qui est ré-injecté dans le réseau pour modifier son comportement via une procédure itérative qui, éventuellement, lui permet de simuler la réponse du professeur. Autrement dit, la connaissance de l'environnement par le professeur est graduellement transféré vers le réseau jusqu'à l'atteinte d'un certain critère d'arrêt. Par la suite, on peut éliminer le professeur et laisser le réseau fonctionner de façon autonome (Marc, 2004).

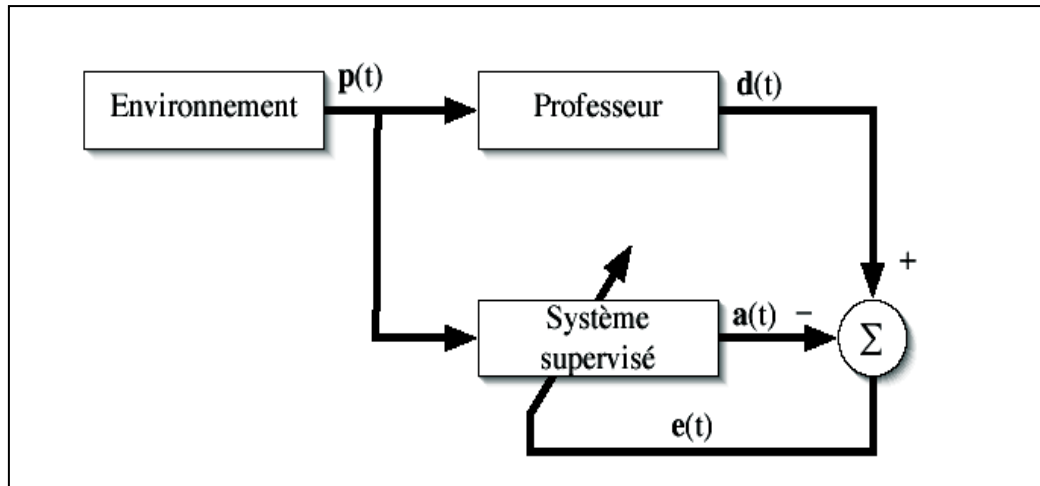


Fig. 15 - Schéma bloc de l'apprentissage supervisé.

❖ L'apprentissage non supervisé

Il n'y a pas là de « professeur », puisque c'est au réseau de découvrir les ressemblances entre les éléments de la base de données, et de les traduire par une proximité dans la « carte » de dimension 2 qu'il doit produire. Les réseaux à apprentissage non supervisé les plus étudiés et utilisés sont les « cartes auto-organisatrices » ou « cartes de Kohonen » (Dreyfus, 2005).

Elle est caractérisée par l'absence complète de professeur, c'est-à-dire qu'on ne dispose ni d'un signal d'erreur, comme dans le cas supervisé, ni d'un indice de satisfaction, comme dans le cas par renforcement. Nous ne disposons donc que d'un environnement qui fournit des stimuli, et d'un réseau qui doit apprendre sans intervention externe. En assimilant les stimuli de l'environnement une description de son état interne, la tâche du réseau est alors de modéliser cet état le mieux possible. Pour y arriver, il importe d'abord de définir une mesure de la qualité pour ce modèle, et de s'en servir par la suite pour optimiser les paramètres libres du réseau, c'est-à-dire ses poids synaptiques.

La fin de l'apprentissage, le réseau a développé une habilité formé des représentations internes des stimuli de l'environnement permettant d'encoder les caractéristiques de ceux-ci et, par conséquent, de créer automatiquement des classes de stimuli similaires.

L'apprentissage non-supervisé s'appuie généralement sur un processus compétitif permettant d'engendrer un modèle où les poids synaptiques des neurones représentent des prototypes de stimuli ». (Dreyfus, 2005).

❖ **La loi de Hebb**, un exemple d'apprentissage non supervisé

La loi de (Donal Hebb, 1949) s'applique aux connexions entre neurones, comme le représente la figure 16.

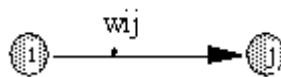


Fig. 16 - i le neurone amont, j le neurone aval et w_{ij} le poids de la connexion.

Elle s'exprime de la façon suivante

"Si 2 cellules sont activées en même temps alors la force de la connexion augmente".

La modification de poids dépend de la coactivation des neurones pré synaptique et post synaptique, ainsi que le montre le (Tableau. 4) x_i et x_j sont respectivement les valeurs d'activation des neurones i et j, S_{wij} (dérivée partielle du poids) correspond à la modification de poids réalisée.

Tableau. 4 - La loi de Hebb.

x_i	x_j	S_{wij}
0	0	0
0	1	0
1	0	0
1	1	+

La loi de Hebb peut être modélisée par les équations suivantes ($w(t+1)$ est le nouveau poids, $w_{ij}(t)$ l'ancien) :

$$w_{ij}(t+1) = w_{ij}(t) + S_{wij}(t) \tag{2}$$

$$S_{wij}(t) = x_i \cdot x_j \tag{3}$$

(la coactivité est modélisée comme le produit des deux valeurs d'activation)

L'algorithme d'apprentissage modifie de façon itérative (petit à petit) les poids pour adapter la réponse obtenue à la réponse désirée. Il s'agit en fait de modifier les poids lorsqu'il y a erreur seulement (Touzet, 1992).

1/ Initialisation des poids et du seuil S à des valeurs (petites) choisies au hasard.

2/ Présentation d'une entrée $E_i = (e_1, \dots, e_n)$ de la base d'apprentissage.

3/ Calcul de la sortie obtenue x pour cette entrée :

$$a = S (w_i \cdot e_i) - S \quad (4)$$

(la valeur de seuil est introduite ici dans le calcul de la somme pondérée)

$$x = \text{signe}(a) \text{ (si } a > 0 \text{ alors } x = +1 \text{ sinon } a = 0 \text{ alors } x = -1 \text{)} \quad (5)$$

4/ Si la sortie x est différente de la sortie désirée d_i pour cet exemple d'entrée E_i alors modification des poids (μ est une constante positive, qui spécifie le pas de modification des poids) :

$$w_{ij}(t+1) = w_{ij}(t) + \mu \cdot (x_i \cdot x_j) \quad (6)$$

5/ Tant que tous les exemples de la base d'apprentissage ne sont pas traités correctement (i.e. modification des poids), retour à l'étape 2.

❖ L'apprentissage Par Renforcement

L'apprentissage par renforcement permet de contourner certaines des limitations de l'apprentissage supervisé. Il consiste en une espèce d'apprentissage supervisé, mais avec un indice de satisfaction scalaire au lieu d'un signal d'erreur vectoriel. Ce type d'apprentissage est inspiré des travaux en psychologie expérimentale de (Marc, 2004).

Dans le contexte des réseaux de Neurones Artificiels, nous pouvons reformuler l'énoncé de la façon suivante :

Lorsqu'une action (décision) prise par le réseau engendre un indice de satisfaction positif, alors la tendance du réseau à prendre cette action doit être renforcée. Autrement, la tendance prendre cette action doit être diminuée.

En pratique, l'usage de l'apprentissage par renforcement est complexe à mettre en oeuvre. Il importe cependant de bien comprendre la différence entre ce type d'apprentissage et l'apprentissage supervisé

L'apprentissage supervisé dispose d'un signal d'erreur qui non seulement permet de calculer un indice de satisfaction (p.ex. l'erreur quadratique moyenne), mais permet aussi d'estimer le gradient local qui indique une direction pour l'adaptation des poids synaptiques. C'est cette information fournie par le professeur qui fait toute la différence. Dans l'apprentissage par renforcement, l'absence de signal d'erreur rend le calcul de ce gradient impossible. Pour estimer le gradient, le réseau est obligé de tenter des actions et d'observer le résultat, pour éventuellement inférer une direction de changement pour les poids synaptiques. Pour ce faire, il s'agit alors d'implanter un processus d'essais et d'erreurs tout en retardant la récompense offerte par l'indice de satisfaction.

L'apprentissage par renforcement, c'est apprendre à agir par essai et erreur. Dans ce paradigme, un agent peut percevoir son état et effectuer des actions. Après chaque action, une récompense numérique est donnée. Le but de l'agent est de maximiser la récompense totale qu'il reçoit au cours du temps.

1.2.6. Perceptron Simple et Perceptron Multicouche

a) Le Perceptron Simple

Mis au point à partir du milieu des années 50 (57-61) par Rosenblatt (Rosenblatt, 1962). Le Perceptron se veut un modèle de l'activité perceptive. Il s'agit d'associer des configurations présentées en entrée à des réponses (figure 17). Il s'agit d'une seule couche de S neurones connectée totalement sur un vecteur-P-de-R-entrée.

L'objectif est d'entraîner le réseau jusqu'à ce que la valeur obtenue en sortie soit égale à la valeur désirée (on est dans le cadre d'un apprentissage supervisé) (Corsini, 2005). Pour cela on utilise la règle de Hebb (du nom de son inspirateur Donald Hebb) qui consiste à renforcer les connexions entre deux cellules si elles répondent similairement et à diminuer cette valeur sinon. On pose:

$$w_{ij} = w_{ij} + \eta (d_i - o_i) s_j \quad (7)$$

ou encore, si on rend le temps explicite dans l'équation

$$w_{ij}(t+1) = w_{ij}(t) + \eta (t) (d_i - o_i(t)) s_j(t) \quad (8)$$

Dans laquelle $o_i(t)$ désigne la valeur calculée (au temps t) par le réseau sur la sortie i , tandis que d_i correspond à la valeur désirée ; $s_j(t)$ correspond au signal transitant sur la connexion de j à i de poids $w_{ij}(t)$; $h(t)$ étant le pas (ou coefficient) d'apprentissage au

temps t . Ce pas sert à moduler l'impact de l'erreur sur la connexion, η le taux d'apprentissage.

Il est clair que la sortie du réseau peut prendre seulement deux valeurs distinctes selon le niveau d'activation du neurone : -1 lorsque ce dernier est strictement inférieur à 0 ; +1 dans le cas contraire.

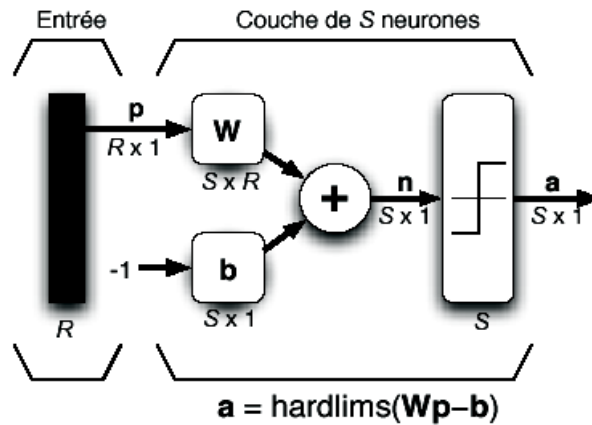


Fig. 17 - Perceptron à une seule couche avec fonction seuil.

❖ Algorithme de traitement

✓ Initialisation

Mettre les poids initiaux w_1, w_2, \dots, w_n ainsi que le seuil θ à des valeurs aléatoires de l'intervalle $[-0,5, 0,5]$. Mettre le taux d'apprentissage à une petite valeur positive.

✓ Activation

Activer le perceptron en appliquant les intrants $x_1(p), x_2(p), \dots, x_n(p)$ et l'extrant désiré $Y_d(p)$. Calculer l'extrant actuel à l'itération $p = 1$.

$$Y(p) = \text{étage} \left[\sum_{i=1}^n x_i(p)w_i(p) - \theta \right] \quad (9)$$

où n est le nombre de signaux intrants, et étage est la fonction d'activation par étage.

✓ Entraînement des poids

Mettre à jour les poids du perceptron

$$w_i(p+1) = w_i + \Delta w_i \quad (10)$$

où $\Delta w_i(p)$ est la correction de poids à l'itération p .

La correction de poids est calculée par la loi delta :

$$\Delta w_i = \alpha \times x_i(p) \times e(p) \tag{11}$$

où α est le taux d'apprentissage et $e(p)$ l'erreur à l'itération p (la différence entre l'extrant désiré et l'extrant actuel du perceptron)

✓ **Itération**

Augmenter p de 1, retourner à l'étape 2 et répéter le procédé jusqu'à convergence.

❖ **Adaline et Widrow-Hoff**

À la section précédente nous avons traité le cas du perceptron simple où les neurones utilisent une fonction de transfert de type 'seuil'.

Nous allons maintenant considérer la même architecture de réseau à une seule couche mais avec cette fois ci une fonction de transfert linéaire comme présenté à la figure. (Fig.18). Ce réseau s'appelle « ADALINE » [en anglais ADAPTATIVE LINEAR NEURON] à cause de sa fonction de transfert linéaire. IL souffre des mêmes limitations que le perceptron simple : il ne peut résoudre que des problèmes linéairement séparables. Cependant son algorithme d'apprentissage, la règle du moindre carrée (*Least Mean Square*) est beaucoup plus puissante que la règle du perceptron original, car bien que cette dernière soit assurée de converger vers une solution, si celle ci existe. En revanche la règle LMS minimise l'erreur quadratique moyenne, de sorte que la frontière de décision à tendance à se retrouver aussi loin que possible des prototypes. (Marc, 2004).

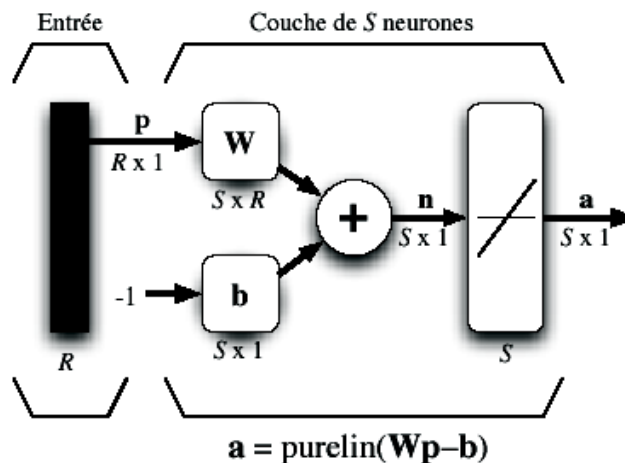


Fig. 18 - Réseau ADLINE.

b) Perceptron multicouche et rétro propagation de l'erreur

❖ Définition

Les perceptrons Multicouches sont des réseaux dont les neurones sont organisés en couches successives, les connexions sont toujours dirigées des couches inférieures vers les couches supérieures et les neurones d'une même couche ne sont pas interconnectés. Un neurone ne peut donc transmettre son état qu'à un neurone situé dans une couche postérieure à la sienne (Hornik, K. 1989).

Choisir l'architecture d'un PMC consiste à :

1. Etape 01 : fixer le nombre de couches cachées.
2. Etape 02 : déterminer le nombre de neurones par couches cachées
3. Etape 03 : Choisir la fonction d'activation
4. Etape 04 : choisir l'apprentissage.

Le perceptron multicouche est un réseau comportant L couches, chaque neurone d'une couche étant totalement connecté aux neurones de la couche suivante. L'algorithme d'apprentissage par rétro-propagation du gradient de l'erreur est un algorithme itératif qui a pour objectif de trouver le poids des connexions minimisant l'erreur quadratique moyenne commise par le réseau sur l'ensemble d'apprentissage. Cette minimisation par une méthode du gradient conduit à l'algorithme d'apprentissage de rétro-propagation (Lippmann, 1987). Cet algorithme, qui présente l'avantage d'exister, reste discuté dans la mesure où sa convergence n'est pas prouvée. Son utilisation peut conduire à des blocages dans un minimum local de la surface d'erreur. Son efficacité dépend, en effet, d'un grand nombre de paramètres que doit fixer l'utilisateur : le pas du gradient, les paramètres des fonctions sigmoïdes des automates, l'architecture du réseau ; nombre de couches, nombre de neurones par couche ..., l'initialisation des poids.

❖ Modèle Mathématique du Perceptron Multicouches

On dispose d'une variable quantitative ou d'une variable qualitative (notée y) à q modalités que l'on doit prédire à partir de p variables x_1, x_2, x_p, \dots prédictives. On dispose par ailleurs de n individus ou observations (échantillon d'apprentissage) décrits par les p variables x_1, x_2, x_p, \dots et pour lesquels on connaît les valeurs de y . On suppose que la couche d'entrée est formée de p entrées, auxquelles seront appliquées des coefficients appelés les poids synaptiques w_{jm} (Fig. 19).

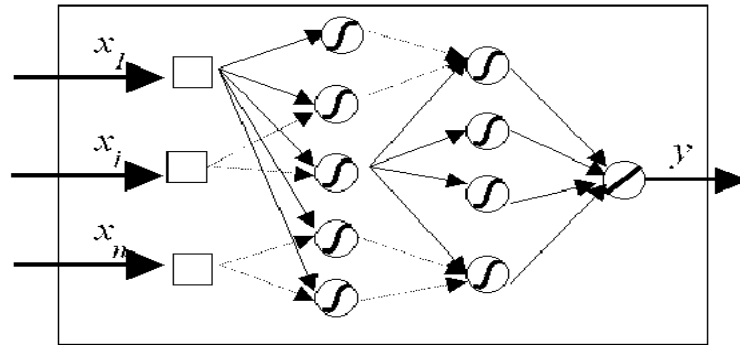


Fig. 19 - Modèle Mathématique simple d'un PMC.

De plus, il existe un terme constant en entrée qui, pour des raisons pratiques, prend la valeur 1. La couche cachée comprend c neurones qui seront chacun activés par une intégration (en général fonction monotone de la somme) des p signaux en provenance de la couche d'entrée [On voit souvent des textes qui mentionnent, outre la couche cachée et la couche de sortie, une « couche d'entrée », voire des « neurones d'entrée ». Cette expression est trompeuse, car les entrées (représentées par des carrés sur la figure (Fig.20) ne sont pas des neurones : elles ne réalisent aucun traitement de l'information, puisqu'elles ne font que transmettre les valeurs des variables] (Dreyfus,2004).

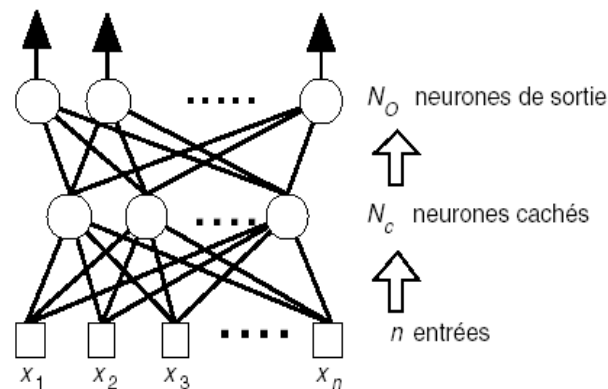


Fig. 20 - un réseau de neurone à n entrées, une couche de N_c neurones cachés et N_o neurones de sortie.

La même opération a lieu pour les q éléments de la couche de sortie mettant en jeu les poids synaptiques. Il existe aussi une connexion directe de l'entrée constante à la sortie.

L'introduction de la constante d'entrée unitaire, connectée à chaque neurone situé dans la couche cachée ainsi qu'à chaque sortie, évite d'introduire séparément ce que les

informaticiens appellent un biais pour chaque unité. Les biais deviennent simplement parties intégrantes de la série de poids (les paramètres).

La fonction d'activation f peut être paramétrée de manière quelconque. Deux types de paramétrages sont fréquemment utilisés (Dreyfus, 2004).

1. Les paramètres sont attachés aux entrées du neurone : la sortie du neurone est une fonction non linéaire d'une combinaison des entrées $\{x_i\}$ pondérées par les paramètres $\{w_i\}$, qui sont alors souvent désignés sous le nom de « poids » ou, en raison de l'inspiration biologique des réseaux de neurones, « poids synaptiques ». Conformément à l'usage (également inspiré par la biologie), cette combinaison linéaire sera appelée « potentiel ». Le potentiel v le plus fréquemment utilisé est la somme pondérée, à laquelle s'ajoute un terme constant ou « biais » :

$$v = w_0 + \sum_{i=1}^{n-1} w_i x_i \quad (12)$$

La fonction f est appelée fonction d'activation il est recommandé d'utiliser Pour f une fonction « sigmoïde » (c'est-à-dire une fonction en forme de «s») symétrique par rapport à l'origine, telle que la tangente hyperbolique ou la fonction Arctangente. Ainsi, dans la très grande majorité des applications, la sortie d'un neurone a pour équation [tangente hyperbolique] ou sigmoïde, (Voir tableau. 2).

$$Y = SIG \left[w_0 + \sum_{i=1}^{n-1} w_i X_i \right] \quad (13)$$

$$y = th \left[x_0 + \sum_{i=1}^{n-1} w_i x_i \right] \quad (14)$$

2. Les paramètres sont attachés à la non-linéarité du neurone (Dreyfus, 2004): ils interviennent directement dans la fonction f ; cette dernière peut être une *fonction radiale* ou RBF (en anglais *Radial Basis Function*), ou encore une ondelette ; les premières tirent leur origine de la théorie de l'approximation (Powell, 1987), les secondes de la théorie du signal (M allat, 1989).

Par exemple, la sortie d'un neurone RBF à non-linéarité gaussienne a pour équation :

$$y = \exp \left[- \frac{\sum_{i=1}^n (x_i - w_i)^2}{2w_{n+1}^2} \right] \quad (15)$$

où les paramètres w_i , $i = 1$ à n sont les coordonnées du centre de la gaussienne, et w_{n+1} est son écart-type.

❖ La rétro propagation du gradient

La méthode de modification des poids est très simple avec l'algorithme de Rosenblatt, mais il implique quelques limitations d'apprentissage. Dans le cas de perceptrons multicouches, comme on ne sait pas les extrants (outputs) désirés des couches cachées, mais seulement de la dernière couche, il faut propager la responsabilité des erreurs de la dernière couche à la première, dans le sens contraire de l'exécution de réseau, d'où le nom rétro-propagation. De plus, les perceptrons multicouches utilisent des neurones munis de la fonction d'activation sigmoïde, celle-ci permettant les nuances nécessaires à une bonne utilisation de la rétro-propagation. Voici les principales lignes de la méthode de rétro-propagation :

- Calcul de l'erreur :

$$e_k(p) = y_{d,k} - y_k \quad (16)$$

l'erreur du neurone k est la différence entre la valeur de l'extrant désiré et la valeur actuelle de l'extrant du neurone k

- Correction de poids (par la loi delta):

$$\Delta w_{jk}(p) = \alpha \times y_j(p) \times \delta_k(p) \quad (17)$$

La correction du poids du lien du neurone j au neurone k , où $\delta_k(p)$ est le gradient d'erreur du neurone k à l'itération p .

- Gradient d'erreur (pour les neurones de la couche de sortie) :

$$\delta_k(p) = F'[X_k(p)] \times e_k(p) = y_k(p) \times [1 - y_k(p)] \times e_k(p) \quad (18)$$

La deuxième partie est obtenue par la dérivée de la fonction sigmoïde appliquée à l'intrant net du neurone k à l'itération p .

- Gradient d'erreur (pour les neurones d'une couche cachée) :

$$\delta_j(p) = y_j(p) \times [1 - y_j(p)] \times \sum_{k=1}^l \delta_k(p) w_{j,k}(p) \quad (19)$$

Où l est le nombre de neurones sur la couche suivante (Fig. 21).

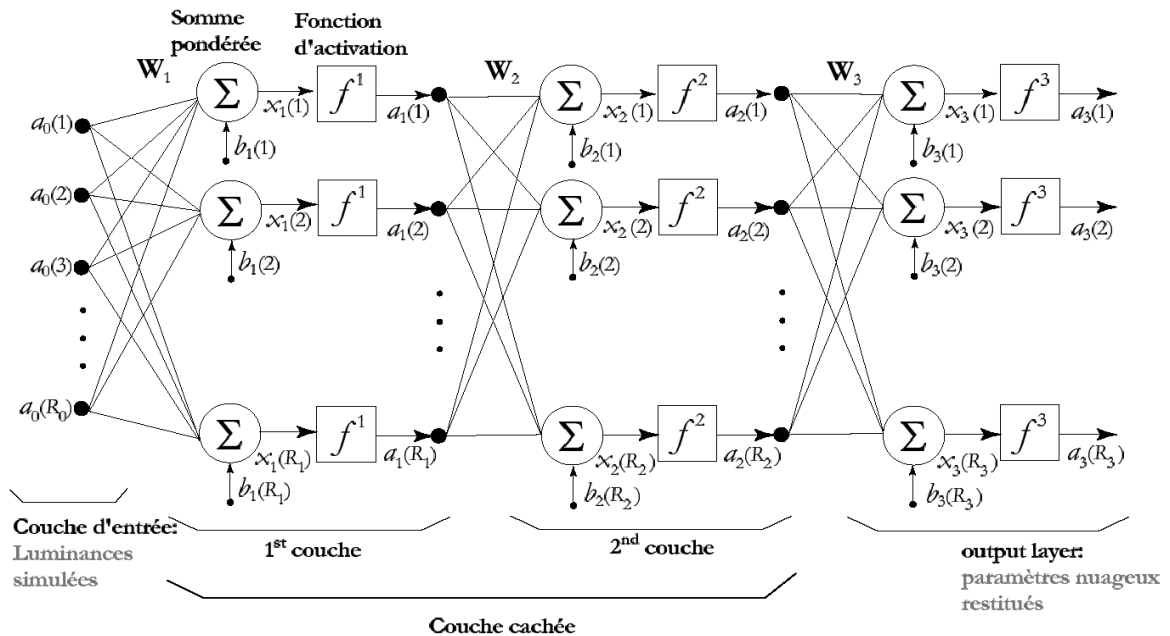


Fig. 21- Schéma détaillée d'un PMC.

✓ **Algorithme de traitement**

➤ **Initialisation**

Mettre tous les poids et les seuils d'activation du réseau à des valeurs aléatoires uniformément distribuées dans un petit intervalle :

$$\left(-\frac{2,4}{F_i}, \frac{2,4}{F_i} \right), \text{ où } F_i \text{ est le nombre total de signaux intrants du neurone } i \text{ dans le réseau.}$$

Cette initialisation est faite un neurone à la fois.

Mettre la valeur du taux d'apprentissage à une petite valeur positive.

➤ **Activation**

Activer le réseau de neurones en appliquant les signaux intrants et les extrants désirés. Calculer les signaux sortant des neurones des couches successives, de la première couche cachée à la couche de sortie.

➤ **Entraînement des poids**

Mettre à jour les poids du réseau en propageant les erreurs dans le sens inverse :

Calculer les gradients d'erreurs pour les neurones des couches successives (ordre inverse), de la couche de sortie à la première couche cachée.

Calculer les corrections de poids pour chaque lien et mettre à jour les poids.

➤ **Itération**

Incrémenter p , retourner à l'étape 2 et répéter le procédé jusqu'à ce que le critère d'erreur soit atteint.

c) Propriété fondamentale des réseaux de neurones

Propriété fondamentale des réseaux de neurones non bouclés à apprentissage supervisé l'approximation parcimonieuse (Dreyfus, 2004) : Les réseaux de Neurones à Apprentissage Supervisé sont des Approximateurs Universels :

Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire (Hornik et *al.* 1989; Hornik, 1991 in Dreyfus, 2004).

1.3. Application des réseaux de neurones dans le domaine de géologie

Les domaines des sciences de la terre a connu un grand développement d'utilisation de l'intelligence artificielle, dans l'ingénierie et la recherche, tel que les réseaux de neurones artificiels RNA, surtout dans le domaine pétrolier grâce à la performance et la précision de ces résultats, on introduit ce présente mémoire par des résumés de quelques principaux travaux qui traitent des thématiques similaires à notre problématique, par différentes types et techniques des réseaux de neurones artificiels.

-Article intitulé: *Prediction of natural fracture porosity from well log data by means of fuzzy ranking and an artificial neural network in Hassi Messaoud oil field, Algeria.* (Zerrouki, A.A. et al., 2014). Cet article fait l'objet d'estimation de la porosité secondaire (de fracturation) à partir des logs de (densité, neutron porosité, et le sonique) et les caractéristiques de la boue (la densité de fluide, le temps de transition et la saturation de fluide).si l'absence l'un de ces paramètres, l'estimation de la porosité secondaire est impossible. Le problème rencontré dans le champ d'huile de Hessi Messaoud est l'absence du temps de transition dans plusieurs puits, qui provoque une difficulté de calculer la porosité de fracturation. A cet effet la méthodologie de classification floue et les réseaux de neurones artificiels sont alors proposés pour rédimier ce problème. Les données utilisé sont des log conventionnels (la résistivité profonde, la densité, neutron porosité et gamma ray).la classification floue est utilisée pour classer les log (input) pour un degré d'influence acceptable avec les out put de ANN, les résultats obtenus confirment de toutes les data utilisés par ANN sont importants et on peut pas les négligés.la structure du ANN qui a utilisée est de type retro propagation, cette méthodologie peut servir pour prédire la porosité de fracturation dans le cas d'absence du log sonique ou les caractéristiques de boue de puits.

-Article intitulé: *Fracture density estimation from core and conventional well logs data using artificial neural networks: The Cambro-Ordovician reservoir of Mesdar oil field, Algeria.* (Samy Zazoun, R. 2013). Cet article présent un modèle capable de prédire la densité des fractures à partir des puits conventionnels calibrés avec les carottes en utilisant le réseau des neurones artificiels. Cette technique est testée dans le réservoir détritique du Cambro-Ordovician dans le champ d'huile de Mesdar. 70 carottes (2120.14 m) du 17 puits ont été étudiées en détail, 7 algorithmes d'apprentissage ont été faites et 8 architectures du réseau ont été testées.la meilleure est en 4 couches [6-16-3-1], avec 6 input (Gamma ray,

Sonique, Caliper, Neutron porosité, densité logs et la profondeur), avec 2 couches cachées la première est de 16 neurones, et la deuxième avec 3 neurones et un seul neurone dans la couche de sortie (la densité des fractures). Les résultats sont basés sur 8094 données des 13 puits avoir une excellente prédiction. Le coefficient de détermination entre les valeurs calculées et les valeurs mesurées montre une bonne corrélation de 0.848. Alors le modèle peut être généralisé avec les puits fracturés naturels voisins.

-Article intitulé: *Methodology of organic-rich shale lithofacies identification and prediction: A case study from Marcellus Shale in the Appalachian basin.* (**Guochang, W. et al., 2012**). Le succès du gaz de schiste qui connut le nord de l'Amérique a augmenté l'intérêt des gisements non conventionnels. Les deux facteurs qui contrôlent l'importance de ce dernier, la fracturation hydraulique qui dépend de la géo-mécanique des roches réservoirs et le gaz naturel est sa relation avec la matière organique (TOC). Alors il est nécessaire de construire un modèle 3D à partir de la minéralogie, la géo-mécanique et (TOC), pour l'optimisation des sections horizontales, 8 inputs sont utilisés pour reconstituer le modèle. Basant sur les méthodes l'analyse discriminante, de logique floue et le réseau de neurones artificiels.

La lithofaciès des schistes sont très importantes quantitativement et qualitativement pour prédire le TOC, à partir de trois paramètres : le pourcentage d'argiles, le rapport quartz carbonate et le TOT.

-Article intitulé: *The application of improved NeuroEvolution of Augmenting Topologies neural network in Marcellus Shale lithofacies prediction.* (**Guochang, W. et al., 2013**). Cet article est une étude complémentaire sur le même champ de Marcellus Shale de l'article précédente, qui a fait appel au d'autre type NeuroEvolution of Augmenting Topologies (NEAT), qui est une idée nouvelle dans l'architecture du réseau des neurones. Ils ont réussi grâce à l'efficacité de cette technique dans 3 aspects : ils ont introduit 2 nouveaux attributs de nœuds de réseau. Nœud de position et de répétition de connexion (RCC) pour augmenter l'efficacité. Ils ont aussi développé la dimension de la population, de la petite valeur à la grande valeur. Ils ont combiné la sélection des variables et moduler le réseau de neurone à une sélection automatiquement des variables, et optimiser la topologie de chaque classification binaire, cet développement est testé et vérifié.

-Article intitulé: *Artificial neural network modeling and cluster analysis for organic facies and burial history estimation using well log data: A case study of the South Pars Gas Field, Persian Gulf, Iran.* (**Alizadeh, B. et al., 2012**). Ils ont utilisé dans cet article l'intelligence artificiel et les statistiques pour extraire les facies organiques cachés à partir des réponses des logs dans un puits Bassin de Sud Giant, champ de Pars Gas , Persian Golf, Iran. Les données utilisés sont :GR, SGR, CGR, THOR, POTA, NPHI et DT logs sont appliqués dans le modèle, il y a une relation entre les diagraphies et le total carbone (TOC), donc le réseau de neurone artificiel est alors utilisé pour construire une log de (TOC),le coefficient de corrélation (R^2) entre les valeurs mesures et calculées TOC égale à 89%,la performance du modèle est mesurés par Mean Squared Error function, ne dépasse pas 0.0073. L'utilisation de la méthode de Cluster Analysis qui est une classification binaire hiérarchique du arbre de la colonne TOC, a montré qui existé 5 organiques facies ont une géochimie similaire. Après ils ont généralisé le modèle ANN pour toute les puits du champ, chaque facies créé est corrélé par histoire de la courbe de pyrolyse, pour monter dans chaque puits le temps et la profondeur de la génération du huile ou du gaz.pour pouvoir savoir la roche mère.la formation peut être identifié par while well logging operations (LWD).

-Article intitulé: *Neural network prediction of carbonate lithofacies from well logs, Big Bow and Sand Arroyo Creek fields, Southwest Kansas.* (**Lianshuang, Qi., et al 2006**). Les réservoirs carbonatés de Kansas en USA, ont une faible récupération des hydrocarbures, à cause de l'hétérogénéité, cette étude est initiée de faire une relation entre les logs de diagraphie et la description des carottes, par la suite elle a été appliquée la prédiction des lithofaciès dans 90 puits non carottés. Dans 10 puits une seule couche cachée du réseau de neurones faite par les logs et la description des textures des carbonates, ils ont utilisé un no linéaire relation entre les lithofaciès des carottes et les logs. Le modèle du réseau de neurones est optimisé par la sélection des variables et le cross validation automatique. Les résultats ont été utilisés pour construire les coupe entres les puits (cross-sections).Le réseau de neurones est applicable dans d'autres réservoirs complexes, avec la géométrie des faciès et la distribution sont les clés pour contrôler l'hétérogénéité et la distribution des propriétés de roches.

-Article intitulé: *Functional networks as a new data mining predictive paradigm to predict permeability in a carbonate reservoir.* (**El-Sebakhy, E.A. et al., 2012**). Dans cet

article ils ont focalisé sur la prédiction d'un paramètre pétrophysique très important, qui est la perméabilité qui pose un problème dans l'ingénierie de réservoirs, due à la méthode directe de mesure sur les carottes dans le laboratoire qui toujours ne couvre pas toute le réservoir, cependant les chercheurs utilisent les outils statistique de régression, les réseaux de neurones artificiels, et le logique floue, pour estimer la perméabilité et la porosité dans les puits. Malheureusement à cause de l'incertitude et le non précision, la modélisation est moins exacte devant l'expérience dans le laboratoire. Cet article fait l'objet d'une originelle architecture de réseau de neurones fonctionnel pour prédire la perméabilité en utilisant les logs de puits du réservoir carbonaté. Les résultats obtenus s'avèrent performants pour cette méthode.

-Article intitulé: *Evolving neural network using real coded genetic algorithm for permeability estimation of the reservoir.* (**Irani, R. et al., 2011**). Dans ce travail ils ont examiné comment le réseau de neurones évolué avec Génétique Algorithme (GA) est performé la fiabilité et la précision du réseau de neurones artificiel. Cette stratégie est appliqué pour prédire la perméabilité du réservoir de Mansuri Bangestan localisé en Ahwaz, Iran utilisant les logs gamma ray, densité, neutron porosité, sonique, porosité totale. la méthode consiste sur une hybridation entre Génétique Algorithme et le réseau de neurone artificiel (GA-ANN). l'algorithme proposé est l'algorithme combiné du gradient du retro propagation local (BP). La comparaison de la stratégie d'hybridation GA-ANN et ANN entre les valeurs mesurés et calculées de l'apprentissage et la validation phases montre une bonne corrélation.

-Article intitulé: *Development of artificial neural network models for predicting water saturation and fluid distribution.* (**Al-Bulushi, N. et al., 2009**). Dans cet article ils ont développé un réseau de neurones artificiel adapté pour la prédiction de la saturation en eau à partir des logs de diagraphie. Deux réservoirs gréseux du moyen orient sont investigués, en premier temps le model de ANN est testé dans la formation de Haradh à Oman, utilisant les logs et les carottes. Deuxième le moèle ANN est utilisé pour la saturation dans les réservoirs gréseux complexe.

Le modèle est basé sur 3 couches du réseau est donne des résultats fiable de la saturation en eau avec un mean square error (RMSE) environ de 0.025, et coefficient de corrélation égale 0.91. Le modèle ANN s'avère plus performant que les statistiques classiques telles que la régression linière multiple. Ils ont utilisé (gamma ray, neutron, densité, résistivité,

et la capacité d'échange) comme inputs. Le ANN est utilisé pour la prédiction de la saturation des hydrocarbures dans la formation du Gharif donne des bonne résultats.

Le modèle du réseau de neurones appliqué s'avère un robuste pour la prédiction de la saturation dans d'autres champs qui ont des formations similaires

- Article intitulé: *Traitement des diagraphies à l'aide d'un réseau de neurones du type « carte auto-organisatrice » : application à l'étude lithologique de la couche silteuse de Marcoule (Gard, France)*. **(Briqueu, L. et al., 2002)**. Cette étude est faite par la mise en œuvre d'un réseau de neurones non supervisé « carte auto-organisatrice » se révèle être une méthode statistique de classification automatique des données, bien adaptée à l'interprétation lithologique des diagraphies. Au cours de la première phase du traitement, à partir de l'exploration systématique du jeu de données, la carte permet de déterminer le nombre de faciès ainsi que les transitions discontinues, entre chacun des bancs identifiés. Dans une seconde phase, l'indexation de la carte par l'apport d'une connaissance géologique, même très fragmentaire, résultant de l'observation de quelques portions de carottes permet de déterminer un log stratigraphique quasiment continu.

-Article intitulé: *Use of neural networks in log's data processing: prediction and rebuilding of lithologic facies*. **(Frayssinet, D. et al., 2000)**. Ils ont appliqué la méthode de perception multi-couches (MLP) qui est une méthode de régression non linéaire, pour résoudre le problème des données manquantes lors de l'acquisition des diagraphies, basant sur les logs des puits voisins pour pallier le manque d'information, ils ont utilisé trois logs ; PEF (photoelectric effect), RHOB (relative density), and GR (gamma ray), pour reconstituer le Log de NPHI (neutronic porosity measured in percentage of porosity). Dans la seconde partie ils ont utilisé la même méthode de la classification automatique type « carte auto-organisatrice » comme le précédent article pour reconstituer les faciès.

-Article intitulé: *Identification of lithofacies using Kohonen self-organizing maps*. **(Hsien-Cheng, C. et al., 2002)**. Cet article discute presque les mêmes thématiques que les 2 articles précédentes.

Après cette recherche bibliographique, il apparait clairement que les méthodes de l'intelligence artificielle ont une relation étroite avec la recherche et le développement dans le domaine pétrolier.

2. METHODE DE CLASSIFICATION CLUSTER ANALYSIS

2.1. Définition

L'analyse de regroupement est une technique qui permet d'identifier des regroupements d'individus (ou d'objets) qui se ressemblent ou partagent des attributs communs. On retrouve cette technique d'analyse sous une très grande variété de noms, incluant l'analyse typologique, l'analyse de regroupement, l'analyse classificatoire, l'analyse taxonomique, etc. Cette grande variété de noms s'explique probablement par la diversité des disciplines qui utilisent cette méthode d'analyse: sciences économiques, marketing, biologie, psychologie et les sciences de la terre. (Baillargeon, J. 2004)

2.2. Types de Cluster Analysis

Il existe plusieurs types de Cluster Analysis parmi lesquelles on cite : (Gallegos, M.T, 2005).

- Groupes disjoints : placer chaque objet dans un et un seul groupe.
- Grappes hiérarchiques : organisées de telle sorte qu'un cluster peut être entièrement contenu dans un autre groupe, mais aucun autre type de chevauchement entre les groupes.
- Grappes floues sont définies par une probabilité de grade ou d'appartenance de chaque objet dans chaque grappe.
- UPGMA (Unweighted Pair Group Method) Méthode des paires des groupes non pondérée qui a été appliquée dans notre étude.

2.3. Application de Cluster Analysis dans le domaine de Géologie

Les articles cités ci-dessous présentent des exemples d'application des techniques de « cluster analysis » dans le domaine de la géologie. Ces études montrent l'intérêt et l'utilité de ladite méthode dans la résolution de nombreuses problématiques liées aux ressources souterraines.

Article intitulé: *Use of Cluster Analysis method in log's data processing: prediction and rebuilding of lithologic facies.* (**Ameur Zaimeche, O. et al .2014**). Cet article est une partie du mémoire de magister, à pour but de reconstituer les facies non carottés à l'aide de la méthode de Cluster Analysis sur les données de diagraphies et la description

détaillée de la carottes, pour pallier le manque d'information lithologique dans les intervalles non carottés.

-Article intitulé: *Well log cluster analysis and electrofacies classification: a probabilistic approach for integrating log with mineralogical data.* (**T. Euzen. et al., 2005**). Le but est de comprendre la minéralogie du réservoir et quantifier l'effet de la matrice sur les paramètres pétrophysiques, l'identification de la variation de la minéralogie et son rôle dans l'analyse séquentielles et la corrélation. Ce travail est fait sur des données de carottes et des cuttings, en utilisant une approche innovatrice visant à bien identifier la typologie des électro-faciès par l'intégration des mesures de minéralogie quantitative des carottes et des cuttings avec des données de diagraphies.

-Article intitulé: *Use of the Cluster Analysis and Artificial Neural Network Technology for Log Data Interpretation.* (**Anatoly, V. 2011**). Méthodes de l'analyse de cluster et des réseaux de neurones artificiels mise en œuvre par le logiciel « Techlog » utilisé pour le traitement et l'interprétation des données sur les puits du champ de pétrole et de gaz à condensats de Verkoshansky. Il a été démontré que la clustérisation des données permet d'améliorer significativement la fiabilité et l'exactitude de la détermination des lithotypes ainsi que la porosité et la perméabilité des roches.

-Article intitulé: *The utility of Cluster Analysis in determining sedimentary facies.* (**Muayyad, H., et al., 1981**). Dans cet article ils ont appliqué la méthode de cluster Analysis pour déterminer les micros facies sédimentaires en se basant sur 98 lames minces à partir de 150 échantillons de carottes des carbonates, dans 13 puits répartis dans le Grand Burgan gisement de pétrole, en Koweït.

-Article intitulé: *Hierarchical cluster analysis of hydrochemical data as a tool for assessing the evolution and dynamics of groundwater across the Ethiopian rift.* (**Tenalem A., et al., 2009**). Cet article montre le modèle de l'évolution et les flux hydrochimiques des eaux souterraines analysés dans un complexe fermé (bassin de rift éthiopien) sur la base de plusieurs variables physico-chimiques. L'analyse a été appliquée sur des échantillons d'eau prélevés à partir de plus de 120 puits et les eaux de surface. Cinq zones d'eaux souterraines avec différents faciès hydrochimiques ont été identifiées par l'application de cette technique.

-Article intitulé: *Application of Multivariate Cluster Analysis in Log-facies Determination and Reservoir Zonation, Case Study of Marun Field, South of Iran.* (Tavakoli, V., et al 2006). Cette étude a pour but la détermination des log-facies sur la formation Asmari du champs de Maroun en Iran, par la programmation d'un nouveau algorithme avec le logiciel MATLAB. L'étude est réalisée sur les logs diagraphiques d'un puits dans la formation d'Asmari entièrement non carotté. Les log-facies prédits sont corrélés avec les lithofaciès définies sur les carottes et les lames minces. La fiabilité de cette technique dans la détermination des log-facies est confirmée par les puits voisins.

-Article intitulé: *an application of cluster analysis as a method of determining biofacies.* (Ames, F., et al., 1968). Dans cet article, les auteurs ont appliqué la méthode de Cluster Analysis pour étudier les populations des foraminifères dans la côte centrale du Texas. Le but de cette étude est d'évaluer l'utilité de la méthode et de comparer les résultats obtenus avec d'autres techniques pour la détermination des bio-facies.

-Article intitulé: *Late Quaternary marginal marine palaeoenvironments of northern Australia as inferred from Cluster Analysis of coccolith assemblages.* (M.J.J. Couapel, et al., 2007). L'application de la méthode de cluster Analysis sur des assemblages de coccolites Quaternaire supérieur dans une carotte de sédiments (MD972132) recueillies dans le golfe de Carpentarie au nord de l'Australie. L'analyse de cluster définit six assemblages d'espèces, et six groupes d'échantillons (en fonction de la profondeur), certains avec des sous-groupes, en vue de reconstituer la paléogéographie.

-Article intitulé: *Deciphering magma mixing: The application of cluster analysis to the mineral chemistry of crystal populations.* (Cortés, J.A., et al., 2007). Cet article montre une typologie cristalochimique d'un mélange de magma, Une technique de classification utilisée pour regrouper les données dans de nombreux domaines, est développée ici comme un outil pour étudier un mélange de magmas et les populations de cristaux mixtes dans les roches volcaniques. La méthode est basée sur la quantification du degré de similitude chimique entre les populations de données de chimie minérale, ce qui permet l'identification de groupes discrets. Dans le but d'appliquer la technique du problème particulier des populations de cristaux mixtes, dans lequel un mélange de magmas a été identifié entre basaltique et andésitique, hébergés dans la chambre magmatique. Les résultats sont en accord avec les interprétations antérieures de mélange de magmas, qui valide l'utilisation de la technique d'analyse de cluster dans le contexte de

magma, et ouvre la possibilité de généraliser cette méthodologie à d'autres aspects de la pétrologie ignée.

-Article intitulé: *Hydrogeochemistry of volcanic hydrogeology based on cluster analysis of Mount Ciremai, West Java, Indonesia.* (**Irawan, D.E., et al., 2009**). L'analyse hydrogéochimique a été effectuée sur 119 sites appartenant à un système hydrogéologique volcanique du mont Ciremai à l'ouest de Java en Indonésie. L'analyse de cluster sur 14 paramètres a extrait trois groupes. Groupe 1 (112 sources) se distingue par des températures normales, faible total de solide dissous (TDS), CE, et de fortes concentrations des bicarbonates. Groupe 2 (cinq sources) à modérément haute température, TDS, conductivité électrique (CE), et une forte concentration de chlorure. Groupe 3 (deux ressorts) présente une haute température, haute TDS anormal, CE, et la concentration de chlorure. Trois systèmes hydrogéologiques ont été repérés sur la base des 3 pôles. Le 1^{er} réseau est développé en aquifère libre peu profonde, avec la dominance des bicarbonates (4,2 me / L) d'eau météorique. Le 2^{ème} système est prédominant dans les processus de mélange des eaux souterraines dans l'aquifère libre ; et souterraine chaude des aquifères profonds. Le 3ème système est principalement dominé par le flux des eaux souterraines de la formation profonde.

Après cette recherche bibliographique, il apparait clairement que les méthodes de classifications ont aussi une relation étroite avec la recherche et le développement et trouvent beaucoup d'applications dans le domaine géologie et pétrolier.

Chapitre II

Géologie Régionale et Locale de la zone d'étude

Ce deuxième chapitre est subdivisé en deux parties.

En premier, nous présenterons succinctement la géologie régionale du bassin de Berkine qui est de type intracratonique.

En deuxième partie, nous aborderons la géologie locale du champ de Sif Fatma qui fait partie de ce dernier et les caractéristiques du réservoir Trias Argileux Gréseux Inférieur qui fait l'objet de notre étude.

1. GEOLOGIE REGIONALE DU BASSIN DE BERKINE

Le bassin de Berkine est une région très importante pour la recherche pétrolière. Les études géologiques et géophysiques effectuées dans la région (Berkine) ont été principalement consacrées au TAGI.

1.1. Situation Géographique

Le bassin de Berkine (ex GHADAMES) se situe dans l'Erg oriental, du Sahara algérien entre les latitudes 29° et 34° Nord et les longitudes 5° et 10° Est, (Fig. 22), il est limité :

- A l'Est par, le bassin de Syrte entre la Tunisie et la Libye, avec une série de failles de direction NW-SE.
- Du sud par le bassin d'Illizi.
- A l'ouest par le mole d'Amghuid El Biod Hassi Messaoud.
- Au nord par le bourrelet d'Ain-Roumana et la route de Dahar.

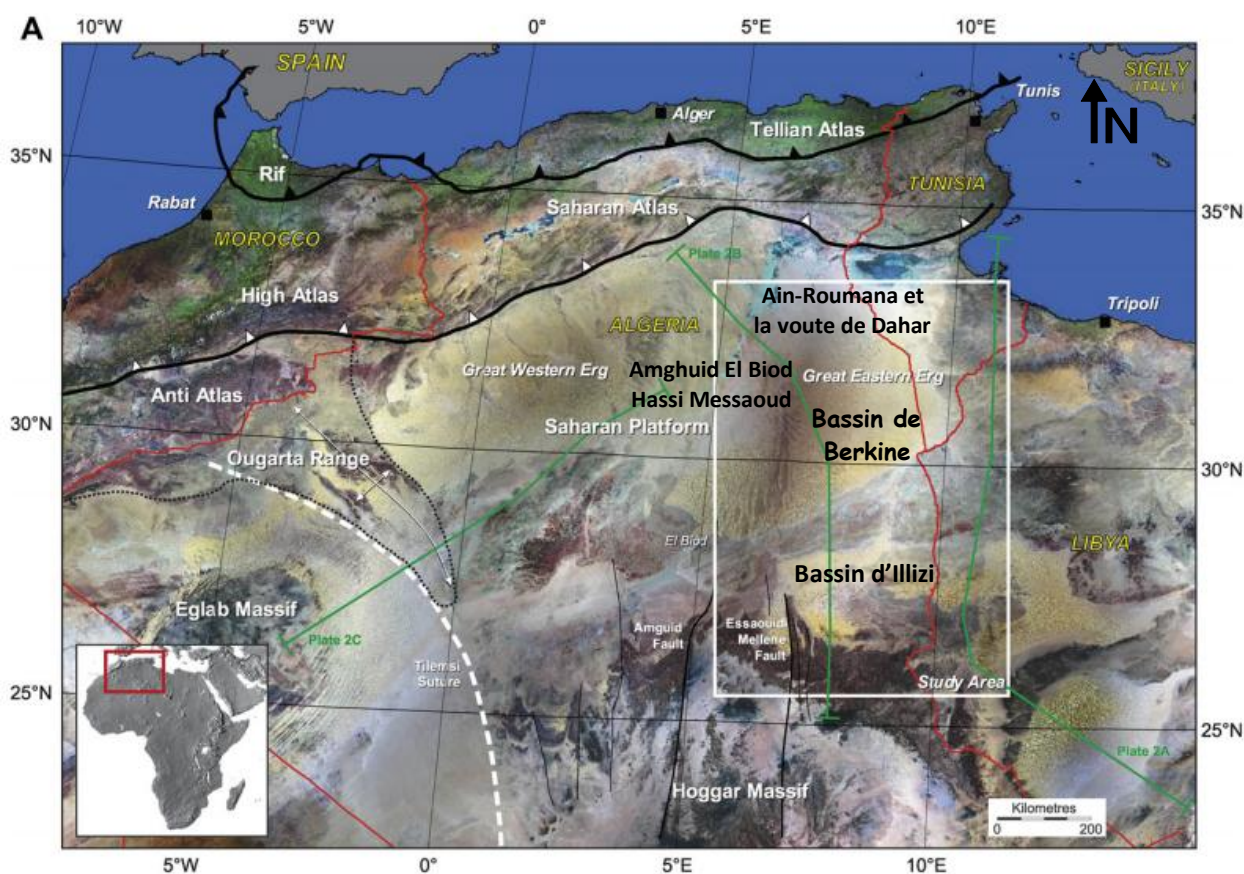


Fig. 22 -Situation géographique (S. Galeazzi et al. 2010), modifiée.

1.2. Cadre Géologique Régionale

De point de vue géologique, le bassin de BERKINE est de type intracratonique (WEC.2007), il se situe dans la partie Nord Est de la plate-forme saharienne. Il est défini comme étant un bassin de plate-forme à substratum infracambrien. Ce bassin occupe une aire totale de près de 300.000 Km², mais seulement 103.000 Km² sont situés dans le territoire Algérien. Le socle qu'on peut rencontrer à une profondeur (côte absolue) de 7000m à est constitué par des roches cristallines, métamorphiques, et volcaniques. (Fig. 23).

Au-dessus, du socle repos en discordance une série sédimentaire comprenant des termes, allant du cambrien à l'actuel.

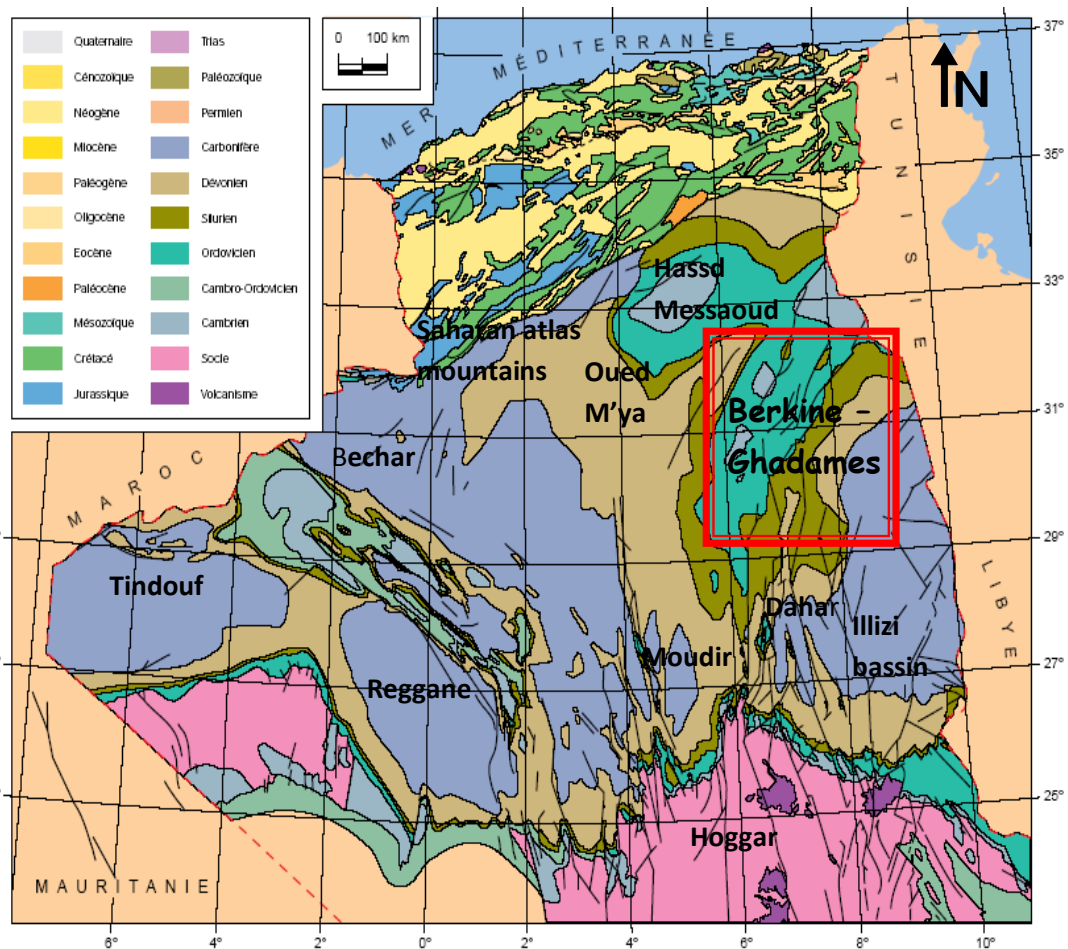


Fig. 23 - Carte géologique de l'Algérie (Askri et al., 1995).

1. 3. Evolution tectonique

Le contexte structural actuel est le résultat d'une succession de plusieurs phases tectonique depuis le paléozoïque.

1.3.1. Phase Panafricaine

C'est une phase précambrienne caractérisée par des accidents subméridiens verticaux, dues à une collision.

1.3.2. Distension Cambro-Ordovicienne

C'est une période d'érosion importante qui a nivelé les structures et les reliefs, les variations d'épaisseur, et de faciès de part et d'autre d'accidents N-S, sont visible au niveau de série Cambro-Ordovicien indiquant l'instabilité tectonique accompagnant leurs dépôt, entre la partie Ouest africaine rigide et la partie Est moins rigide

1.3.3. Compression taconique

C'est une phase de compression le long des failles N-S de la distension précédente donnant naissance à l'architecture en bassin actuel.

1.3.4. Distension Silurienne

C'est la phase de distension suite à la fonte d'une calotte glacière qui a engendré des dépôts d'argiles noirs.

1.3.5. Compression Calédonienne

C'est une phase de soulèvement général, suivi d'une période détritique, elle a lieu entre le silurien et le dévonien.

1.3.6. Phase Dévonienne elle est caractérisée :

- **Au Dévonien inférieur:** par des variations d'épaisseur et de faciès préexistants le long des axes structuraux subméridien qui indiquent une phase distensive.
- **Au Dévonien moyen et supérieur :** par une période d'érosion appelé discordance Frasnienne.

1.3.7. Discordance Hercynienne

Elle est caractérisée par l'arrêt de la sédimentation carbonifère et des dépôts lagunaires (Argile rouge et Gypse). Cette phase tectonique a touché surtout le nord du bassin de Berkine, ou l'érosion de l'étage carbonifère est presque entière.

La discordance hercynienne a permis de mettre en contact avec le Trias et les roches mère du Silurien et du Frasnien. Les mouvements de cette phase ont joué un rôle important dans la structuration des bassins de la plateforme saharienne.

1.3.8. Phase Autrichienne

C'est une phase de compression E-W qui a causé une inversion structural, les failles normales créés pendant la séquence trias lias se sont inversé durant cette phase.

1.3.9. Phase Pyrénéenne

C'est une phase de direction NW-SE qui a donné naissance à de nombreux anticlinaux (Fig. 24)

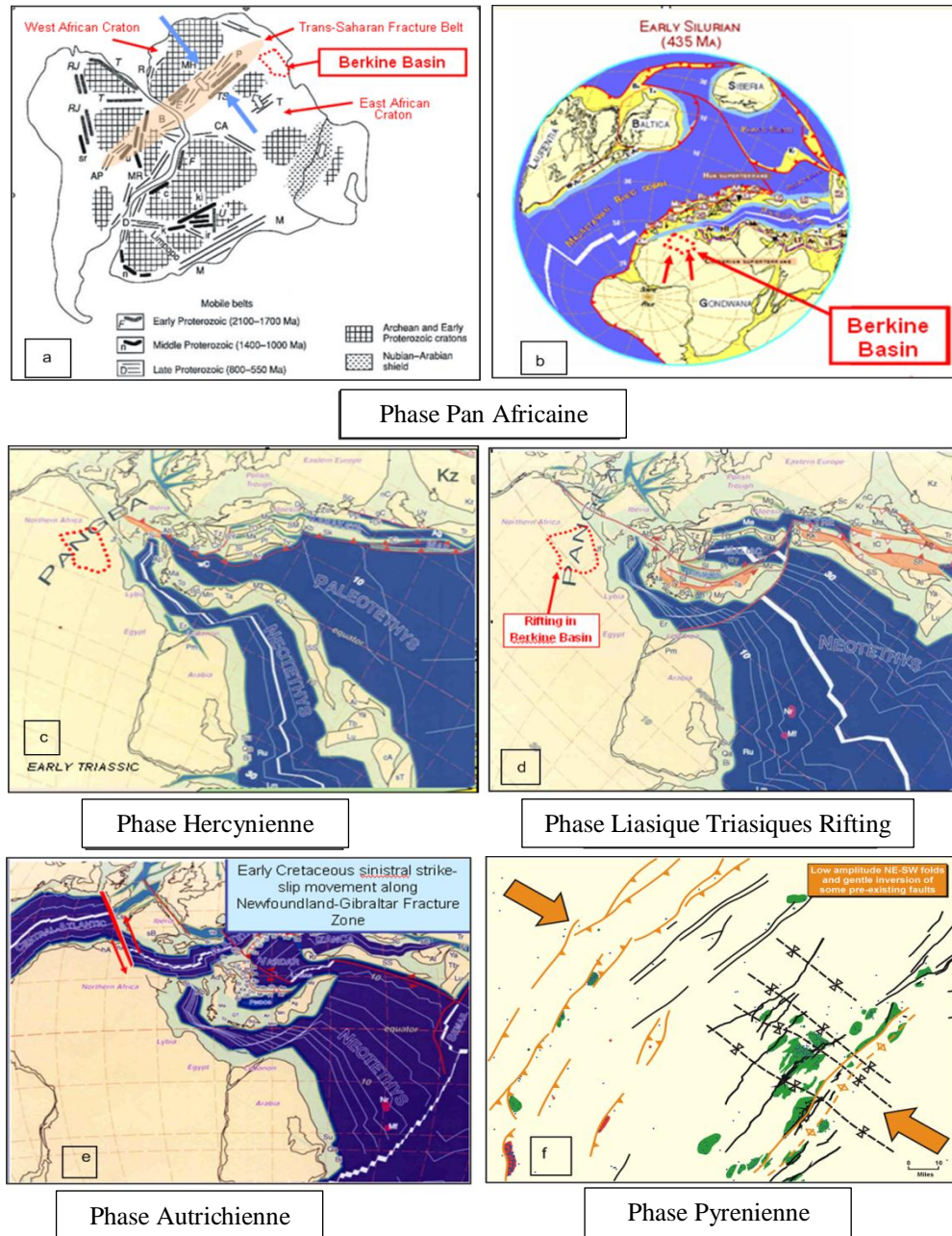


Fig. 24 – Les majeures évolutions structurales du bassin de Berkine (McKenna et Hedley, 2002).

1.4. Géodynamique du bassin

Le bassin est divisé en trois zones à pétrole: la dépression triasique sud-est, la dépression de Dahar et la dépression de Berkine (Fig. 25).

1.4.1. La dépression Triasique Sud-Est

Cette dépression est localisée sur la cuvette saline de Dorbane et de ses

prolongements au Sud-Ouest. Elle représente une zone entre Emguid Hassi Messaoud jusqu'à l'Ouest et l'Est de bassin du Berkine. Elle est marquée par le développement des réservoirs paléozoïques et des TAGS et TAGI, dans la majorité des champs découverts (Rhourde Nouss, Gassi Touil, Gassi El Adem, Hassi Chergui, Brides, Rhourde Adra, et Rhourde Chouf). Ces champs sont généralement associés aux réservoirs triasiques et aux quartzites de l'Hamra.

Le pétrole, le gaz et le condensât sont associés à de grande structure reliée par des systèmes de faille inverse à grands rejets.

1.4.2. La dépression de Dahar

Cette dépression mésozoïque est très étendue. C'est un monoclinal plongeant vers le nord. Elle est continue au Nord jusqu'à la cuvette de Melrhir. Elle représente la limite Nord du bassin de Berkine. De plus, cette zone est marquée par une érosion hercynienne et taconienne. Le développement structural à une tendance Nord Est – Sud-Ouest. La séquence est imposée sur les séries anciennes du silurien jusqu'au Combro-Ordovicien.

1.4.3. La dépression de Berkine

Cette dépression est marquée par des dunes épaisses qui couvrent l'erg Est. Le bassin de Berkine est une dépression très vaste qui est datée du Paléozoïque et dans lequel le socle est situé à plus de 7000m de profondeur. Sa partie centrale fut affectée à peine par l'érosion Hercynienne. Les séries terminales du Carbonifère sont préservées

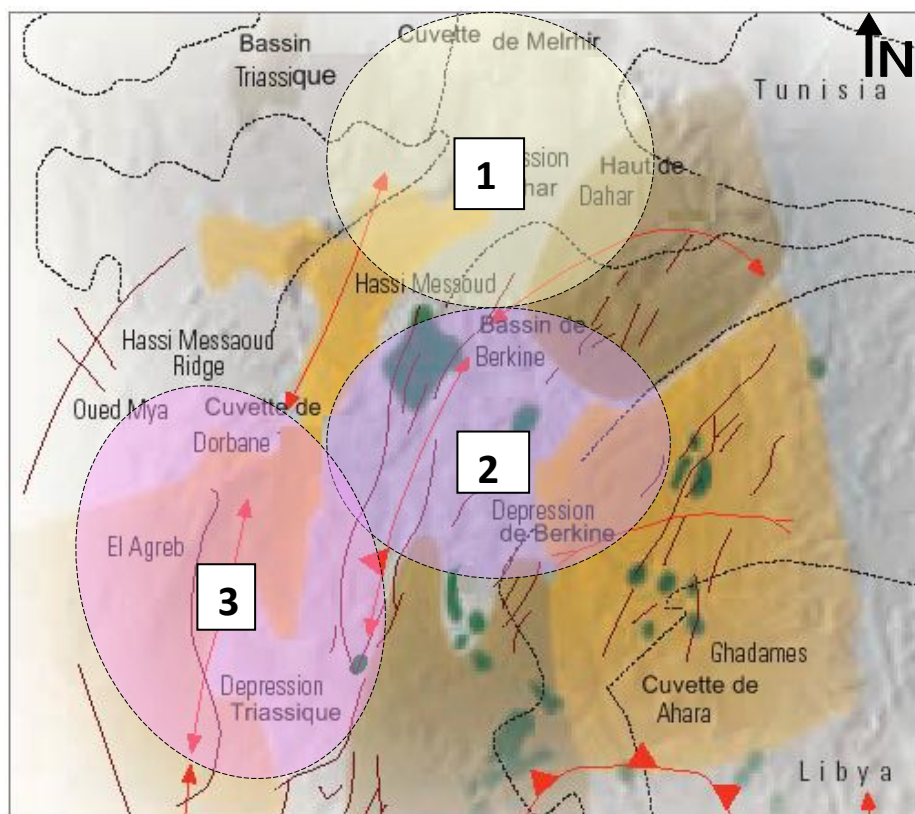


Fig. 25 - Les éléments structuraux du bassin de Berkine (Azoug et al. 2007) :
 1- Dépression de Dahar ; 2- Dépression de Berkine ; 3- Dépression Triassique.

1.5. Stratigraphie

La série stratigraphique de la région est essentiellement composée de dépôts mésozoïques, reposant en discordance sur le Paléozoïque. Enfin, un faible épandage détritique d'âge tertiaire repose en discontinuité sur le Mésozoïque, (Fig. 26)

1.5.1. Le Socle

Le socle composé de roches précambriennes sur lesquelles reposent en discordance les formations paléozoïques.

1.5.2. Le Paléozoïque

Celui-ci est constitué par :

a) Le Cambrien

Ces dépôts reposent sur un socle précambrien, représenté par des grès et des

quartzites à passées conglomératiques, avec une épaisseur moyenne de 300m.

b) L'Ordovicien

Qui repose sur le cambrien est composé d'une alternance des grès et l'argile :

- Les argiles d'El Gassi.
- Les grès d'El Atchane.
- Les quartzites de Hamra.
- Les grès d'Ouargla.
- Les argiles d'Azzel.
- Les grès d'Oued Saret.
- Les argiles micro conglomératiques.
- Les grès de Ramade.

c) Le Silurien

Reposant sur l'ordovicien et composée d'argiles noires riches en matière organique qui forme une importante roche mère, celle-ci est surmontée par un important dépôt argilo-gréseux.

d) Le Dévonien : il est constitué de :

- **Dévonien inférieur:** il est composé d'une alternance argilo-gréseuse.

- **Dévonien moyen**

Il est caractérisé par un dépôt argileux qui comporte des intercalations de niveaux carbonaté.

- **Dévonien supérieur**

Il est représenté par des argiles à passe de marnes de calcaire et de dolomites, au sommet on trouve des grès et des silts.

1.5.2. Le Carbonifère : Il est constitué par:

- **Carbonifère inférieur**

Qui débute par des argiles à passés des siltstones et de calcaire. Au sommet, on distingue une alternance d'argiles de siltstones et de grès avec des passées des roches carbonatées.

➤ **Carbonifère supérieur**

Qui est représenté à la base par une alternance des grès et des argiles sauf dans la partie centrale où on observe des carbonates.

1.5.3. Le Mésozoïque

a) Le Trias

➤ **Le Trias Argilo- Gréseux inférieur (TAGI)**

Il est constitué de grès gris blanc à blanc et gris vert, fin à moyen parfois grossier subarrondi à subanguleux, siliceux à silico-argileux, localement pyriteux, friables à moyennement consolidé avec intercalations d'argile brun rouge parfois gris-vert tendre à indurée silteuse. des Traces de pyrite.

➤ **Le Trias Carbonaté**

Il est constitué d'argile verte à gris verdâtre et brun rouge, tendre à indurée, silteuse localement dolomitique avec passées de dolomie blanc beige, tendre et de grès gris blanc à gris vert, très fin à fin, silico-argileux, moyennement dur.

➤ **Le Trias Argilo- Gréseux supérieur (TAGS)**

Il est constitué de grès gris blanc, très fin, siliceux à silico-argileux, dur avec intercalations d'argile brun rouge et gris vert, tendre à indurée, silteuse.

➤ **Le Trias Argileux**

Il est constitué d'argile brun rouge et brunâtre, rarement gris vert, tendre à indurée silteuse avec traces d'anhydrite blanche, tendre, pulvérulente

1.5.3.2. Le Jurassique

Il est constitué par le Lias, le Dogger lagunaire, le Dogger argileux et le Malm

1.5.4. Le Cénozoïque

a) Le Mio-Pliocène

Il est constitué de sable blanc à translucide, jaunâtre localement rougeâtre, rarement rosâtre fin à moyen, parfois grossier subarrondi à arrondi avec intercalations d'argile brun rouge sableuse, et présence de banc de calcaire argileux.

b) Le Quaternaire : Ce sont des sables représentant des placages dunaires.

ETAGE	AGES	LITHO	DESCRIPTION	EPAIS (m)
QUAT.	QUATERNAIRE		Sable blanc	38-185
TERT.	Mio-Pliocène		Sable blanc avec passé de Calcaire gris	138-185
CRETACE	Sénonien Carbonaté		calcaire gris-blanchâtre, passé Dolomie de Calcaire dolomitique,d'argile et anhydrite	88-120,5
	Sénonien Anhydritique		Sel blanc, passées d' andyrite	329,5-331
	Sénonien Salifère		sel blanc, avec passées d'Argile	152-160
	Turonien		calcaire gris blanc, passées d'Argile et Dolomie	67
	Cénomanién		Argile brun-rouge, calcaire et dolomie	231-240
	Albien		Grés gris, Argile brun, traces de pyrite	90-109
	Aptien		Dolomie blanc, calcaire gris-clair	25-28
	Barrémien		Grés gris, et Argile, traces de dolomie	231-324
	Néocomien		Argile versicolore, et calcaire et Grés gris	247,5-280
JURASSIQUE	Malm		Argile brun-rouge, fine passées de Grés	211-245
	Dogger Argileux		Argile brun-rouge et fine passées de Grés	122-152
	Dogger lagunaire		Argile grise, passées de calcaire et dolomie	123-138
	Lias Anhydritique		Anhydrite massive et Argile grise	156,5-165
	Lias Salifère		sel massif, fines passées d'Argile gris	60-64
	Lias "HB"		calcaire dolomitique, passées d'Argile	19-23
	Lias S1+S2		sel massif avec intercalations d'anhydrite massive et Argile grise, tendre.	142-223,5
	Lias S3		sel massif avec fines passées d'Argile grise	94-125,5
	Lias Argileux		Argile brun rouge avec passée de sel masif	25-32
TRIAS	Trias (S4)		Argile brun rouge avec passée de sel masif	30-50
	Trias Argileux		Argile brun-rouge, trace d'anhydrite	19-33
	Trias Carbonaté		Argile verte a grise, passé de dolomie grise microcristalline, présence de Grés gris-blanc	76-86,5
	T.A.G.I		Grés blanc à gris brun , intercalé d'Argile brune trace de pyrite	77-100
DEV. SUP	Strunien F2		Grés blanc à gris beige avec passées d'argile	66-100
	Famennien		Argile gris-foncé, trace de calcaire argileux et de grés argileux	

Fig. 26 - Log lithostratigraphique du bassin de Berkine (Sonatrach / Anadarko 2012).

1.6. Aspect structural

Le bassin de Berkine est caractérisé par deux domaines principaux qui sont :

- **Domaine ouest:** c'est un domaine très structuré, où les accidents présentent des rejets importants et les structures sont de grande amplitude.

- **Domaine centre :** c'est un domaine, où les structures sont relativement moins prononcées avec une remontée monoclinale vers le nord au niveau de la voûte de Dahar, pour les formations paléozoïques. Les failles qui résultent de la tectonique cassante, ayant affectée les deux domaines sont multidirectionnelles, Les études structurales ont montré que les accidents NE-SW seraient dues principalement aux mouvements distensifs à la fin Trias / début Jurassique. Les failles régionales ont défini des trends structuraux majeurs sur lesquels viennent se greffer les pièges et gisements d'huile et de gaz dans le bassin. Les principales orientations observées sont :

1.6.1. Direction Subméridienne

Les failles subméridiennes se caractérisent par de forts rejets, pouvant atteindre 2000m comme dans le cas de la faille de Ramade et 500m et plus dans la région de Rhourde Nouss, avec des structures de grandes amplitudes (la structure de Rhourde Nouss atteint 900m de fermeture structurale).

1.6.2. Direction Nord Est - Sud-Ouest

Les failles NE-SW sont caractérisées par une composante de décrochement importante avec ramification dans la partie supérieure, les structures associées sont des blocs basculés ou des structures 'en fleur', cette direction interfère avec la direction subméridienne. Sur la majorité des rejeux tectoniques à caractère inverse (faille inverse) s'appuient des structures anticlinales allongées formant des pièges à hydrocarbures dans le TAGI et le TAGS (Fig. 27).

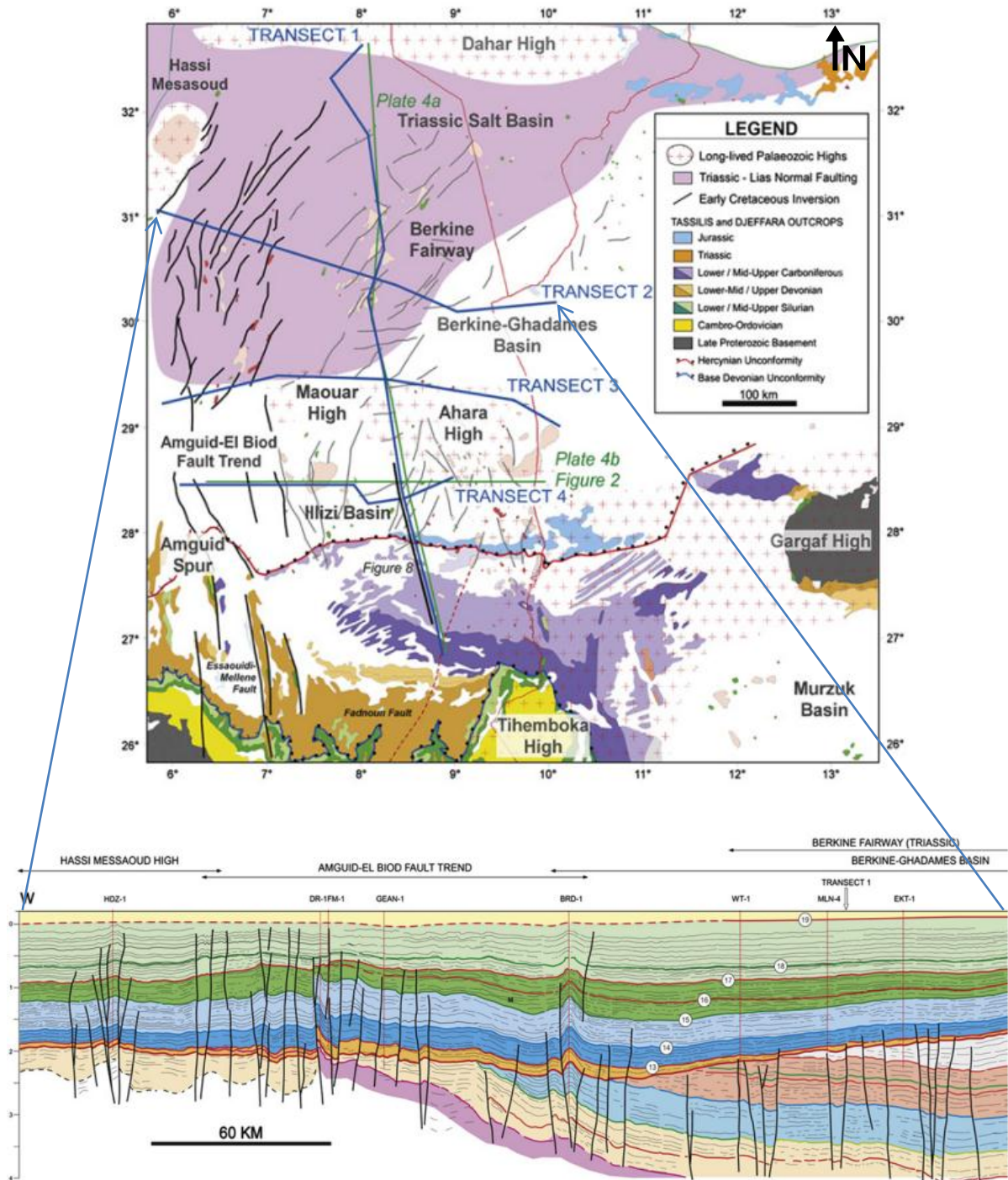


Fig. 27 - Coupe géologique régionale du Bassin de Berkine. (S. Galeazzi et al. 2010).

1.7. Système pétrolier

Une étude de synthèse interne montre que le système pétrolier du bassin de Berkine est constitué par les éléments suivant :

1.7.1. La roche mère

Les argiles à Graptolites du Silurien basal et les argiles du Dévonien supérieur (Frasnien) qui représentent les deux principales roches mères du bassin de Berkine existent dans le bloc404, elles ont généré un important volume d'hydrocarbures depuis le Kimméridgien-fin début Crétacé.

La Frasnien, qui est érodé durant la phase hercynien à l'ouest du bloc 404 présente une épaisseur de 100m dans la région. Il est actuellement dans la fenêtre d'huile mature-début gaz condensât.

1.7.2. La roche couverture

Dans le Carbonifère et le Dévonien, les couvertures sont représentées par les argiles intraformationnelles. Sur les bordures de bassin, la couverture du réservoir TAGI est formée par des argiles et des évaporites du Trias carbonaté, Trias S4 ainsi que le Lias S3. Les réservoirs F1a/F1b, F2 et Upper Rhoude Lakhrouf (RKF) sont couverts par des argiles sus -jacentes de même âge. La couverture des réservoirs gréseux de l'Upper RKF présente cependant un grand risque quand ces réservoirs affleurent directement sous la discordance hercynienne.

Les passages du TAGI basal avec de mauvaises caractéristiques représentent localement une couverture verticale associée au Trias carbonaté. La couverture des réservoirs de l'Ordovicien et du Cambrien est assurée respectivement par des argiles du Silurien et de l'intra-Ordovicien.

Dans le Cambrien, la couverture est assurée par les argiles d'El Gassi. Latéralement, les rejets verticaux des failles associées aux axes structuraux assurent les fermetures.

1.7.3. Les roches réservoirs

Les principaux réservoirs du bassin sont ceux du Trias, du Carbonifère, du Dévonien inférieur, de l'Ordovicien et du Cambrien.

a) TAGS (Trias Argilo Gréseux Supérieur)

Le TAGS est présent dans la dépression SE triasique, au sud-ouest du bassin. Il est constitué de séquences de chenaux fluviaux et deltaïques caractérisées principalement par un grain moyen à grossier témoignant de la proximité des sources d'apport

b) Trias Carbonaté Trias intermédiaire

L'épisode de formation du rift Trias carbonaté-Trias intermédiaire est caractérisé par des faciès argileux, généralement dolomitiques, et de séquences gréseuses du Trias intermédiaire bien développées dans la dépression SE triasique. Ses épaisseurs sont variables, influencées par les rejets en distension des failles majeures.

c) TAGI (Trias Argilo-Gréseux Inférieur)

Le TAGI constitue la série basale du Mésozoïque. Il est caractérisé par des dépôts de type fluvial (parfois éoliens), se développant sur l'ensemble du bassin sur une épaisseur totale pouvant varier de 65 à 80m. La distribution des épaisseurs est contrôlée par les failles et par la paléo-topographie de la région.

d) Carbonifère

D'âge Strunien à Viséen, les réservoirs du Carbonifère se présentent sous la forme d'intervalles gréseux compris dans des séquences argilo-gréseuses liées à un environnement de dépôt marin peu profond. Le Carbonifère basal présente une extension restreinte à la partie centrale et aux bordures occidentales du bassin de Berkine

c) Dévonien inférieur

Les réservoirs du Dévonien inférieur sont représentés par deux séquences essentielles du Gedinnien, où prédomine une série de grès massifs de type fluvial (post-calédonien) d'une épaisseur moyenne de l'ordre de 200 m, et du Siegénien transgressif qui présente des passages gréseux, correspondant à des réservoirs de type cordon littoral et deltaïque.

d) Ordovicien

(Gaz et huile à Hamra et Rhourde Nouss, huile à Nezla) et dans la dépression de Dahar (huile sur le gisement de Semhari) sur les extensions des axes structuraux de Rhourde El Baguel. Ces réservoirs, d'une épaisseur moyenne de 250 m, évoluent progressivement depuis la dépression SE triasique en se biseautant en direction du nord-est vers la région de Touggourt- Semhari. Les réservoirs quartzitiques sont essentiellement liés à l'existence de fracturation (production par fracturation).

e) Cambrien

Les réservoirs du Cambrien sont représentés par quatre unités Ri, Ra, R2 et R3. Les meilleurs réservoirs sont rencontrés au niveau des grès du Ri et du Ra, qui sont généralement quartzitiques. Les grès du Cambrien sont productifs d'huile à Rhourde El Baguel, Ain Romana et Damrane. Les réservoirs cambro-ordoviciens reconnus à ce jour se situent sur les bordures nord et ouest du bassin de Berkine.

1.7.4. La migration (Fig. 28)

Les voies de migration et les systèmes de charge des pièges dans la région sont complexes et difficiles à établir avec précision pour chaque cas, mais néanmoins il est reconnu que les hydrocarbures se sont mis en place à la faveur de deux mécanismes :

- Verticalement à la faveur des failles.
- A travers les formations gréseuses ayant servi de drains.

Ce dernier processus est à envisager dans le cas du remplissage du TAGI par les hydrocarbures ayant migré à travers les unités gréseuses du Carbonifère qui se biseautent sous la discordance hercynienne, puis suivent cette dernière en amont pendage.

1.7.5. Les pièges

Les pièges sont de deux types (WEC. 2007) : structuraux et stratigraphiques. Les pièges structuraux sont des pièges liés aux phases compressives hercyniennes et autrichiennes. Dans les anticlinaux simples, l'exemple le plus marquant est la structure de RKF, qui est polyphasée et non faillée. L'étanchéité des réservoirs triasiques est assurée par les argiles du Trias carbonaté-Trias intermédiaire et les évaporites du Trias-Lias. Les rejets des failles assurent dans la plupart des pièges une étanchéité latérale

Les anticlinaux faillés sont des structures contre faille du type d'Ourhoud et Rhourde Nous. Les pièges stratigraphiques sont des biseaux sous discordance, contre faille, ou des lentilles gréseuses associées aux extensions des faciès du Trias et du Carbonifère. Ces pièges sont principalement mésozoïques (Trias) et localement paléozoïques (Dévonien inférieur et Carbonifère).

Les pièges sous discordance se trouvent dans la partie nord-ouest du bassin où le Dévonien inférieur se biseautent sous les roches éruptives ou sous le Trias. Les lentilles gréseuses du Trias carbonaté peuvent former des pièges stratigraphiques locaux si elles se trouvent sur le chemin de migration des hydrocarbures. (Fig. 28)

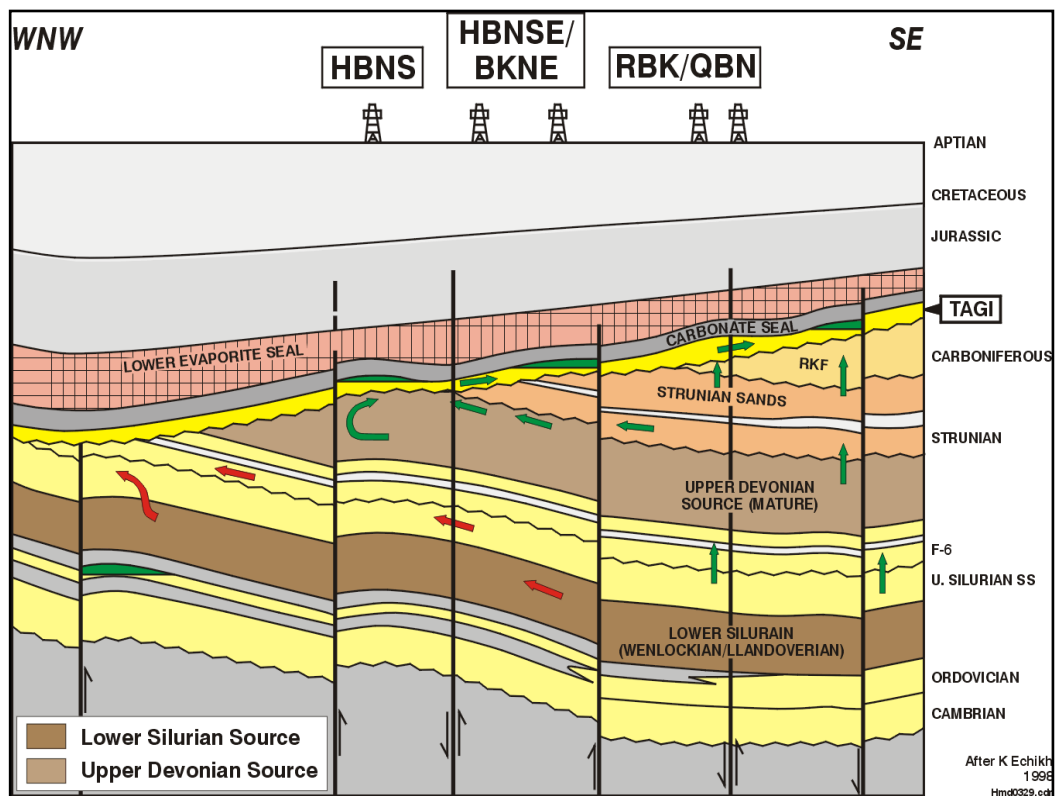


Fig. 28 -Sources et migrations dans le bloc 404 (D'après Echikh, 1998).

2. GEOLOGIE LOCALE DE CHAMP DE SIF FATIMA

2.1. Situation géographique

Le secteur de SIF FATIMA (bloc 402) est situé dans la partie Nord-est du bassin de BERKINE, à environ 100Km au sud-ouest du gisement d'El Borma. Ce bloc (402) couvre une superficie de 103Km² (Fig. 29).

Géographiquement, la zone d'étude est située entre les latitudes 31°-32° Nord et longitudes 8°-9° Est.

2.2. Synthèse Géologique et Géodynamique

La région de SIF-FATIMA est située dans le bassin de BERKINE limitée:

- Au Nord, par Bire Rabaa Nord.
- Au Sud, par Rhoude Debdaba.
- A l'ouest, par Bir Berkine.
- A l'Est, par Bir SIF Fatima.

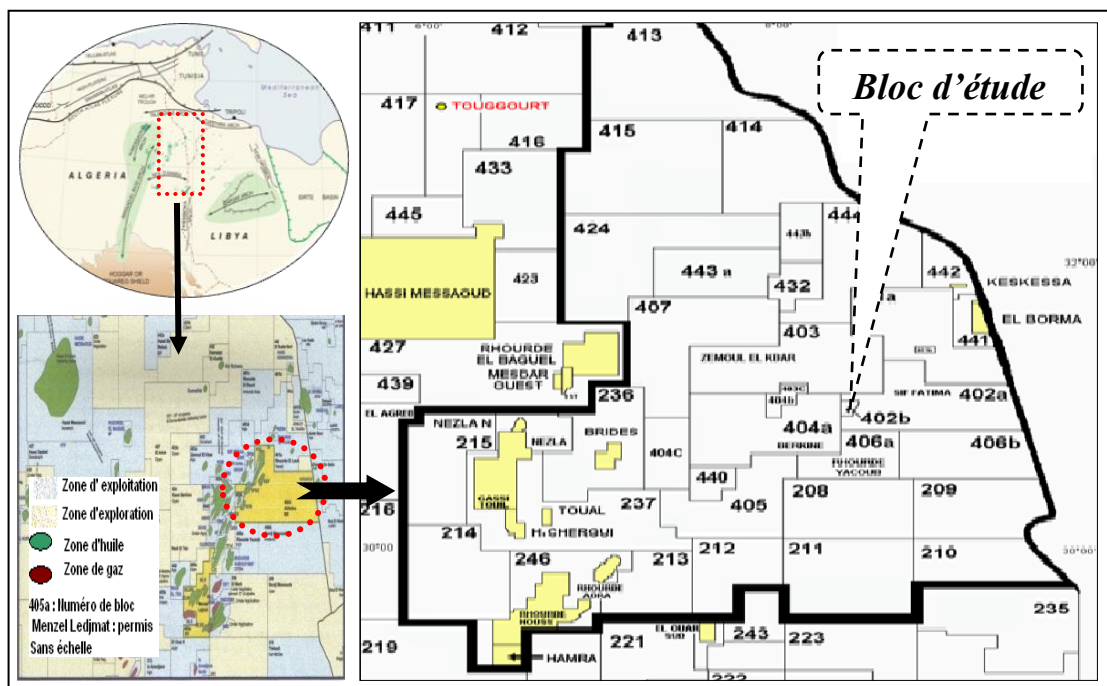


Fig. 29 - Carte de situation de Bloc d'étude (SIF FATIMA) du bassin de Berkine. (WEC, 2007).

2.2.1. Stratigraphie

Il est constitué par des dépôts mésozoïques, reposant en discordance majeure sur des terrains d'âge paléozoïque, traversés partiellement par un épandage détritique daté tertiaire et surmonte en discontinuité avec le crétacé qui est couvert de sable dunaire attribué au quaternaire.

2.2.2. Sédimentologie

La structure et l'organisation du TAGI à SIF-FATIMA indiquent que la séquence est d'origine fluviatile, trois facies reconnus sont généralement associés en succession verticale dans l'ordre suivant lequel ils sont décrits:

- a) **La séquence 1** : à base discordante, c'est un intervalle fluviatile, rempli de paléo-reliefs sur la discordance hercynienne.
- b) **La séquence 2** : grano-décroissante à la base, puis elle devient grano-croissante vers le sommet, composée de grés fluviatiles et des argiles de la plaine d'inondation, avec de fins éléments de crevasses et de paléo-sols.
- c) **La séquence 3** : à base érodée, cette séquence est de type fluvio-lacustre, caractérisée par des grés fluviatiles associés à des grés et des argiles lacustres ou de la plaine d'inondation. Cette séquence représente le réservoir principal d'hydrocarbures et elle est subdivisée en deux sections principales **3A** et **3B**. La base de **3B** est distinguée par les incisions et le dépôt des grés de chenal.
- d) **La séquence 4** : c'est une plaine côtière et marine, peu profonde, comportant des argiles, des évaporites de type Sebkha ainsi que des compartiments gréseux et carbonatés

e) Interprétation des séquences

La subdivision séquentielle et la répartition des principaux milieux de dépôt de TAGI sont illustrées simultanément sur les figures suivantes (Fig. 30).

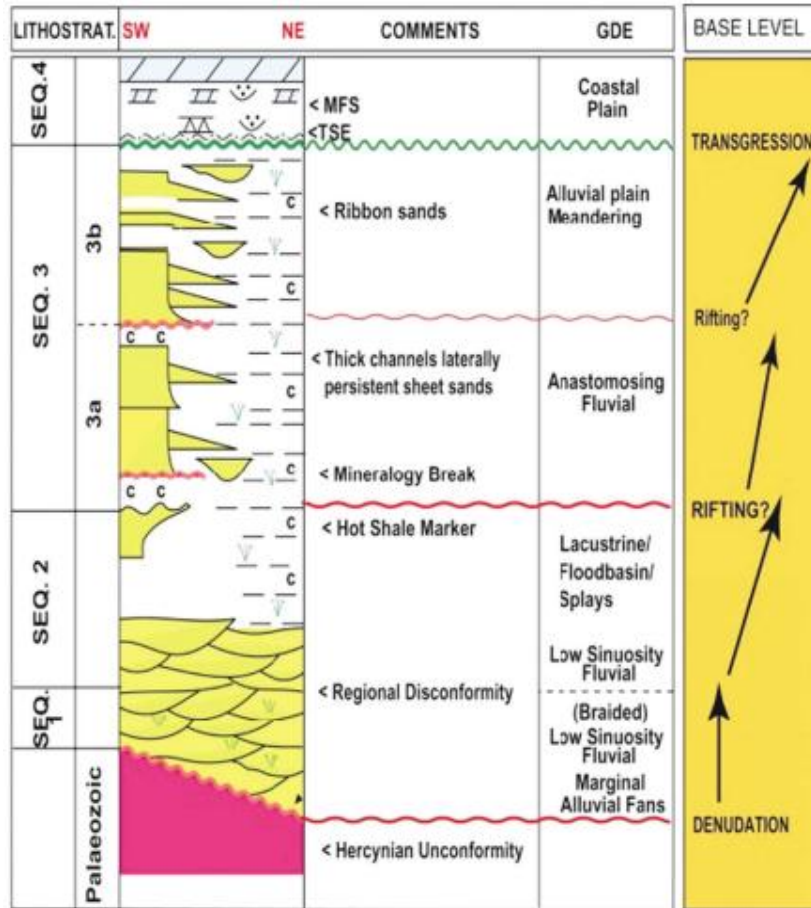


Fig. 30 - La zonation du TAGI, séquence, lithologies et environnement (P.Tuner et al. 2001).

2.2. 3. Tectonique locale

Le secteur de SIF FATIMA est défini en générale par une tectonique cassante relativement intense, les failles majeures Ourhoud-SIF Fatima et les failles secondaires sont pratiquement de même direction. NE-SO, (Fig. 31).

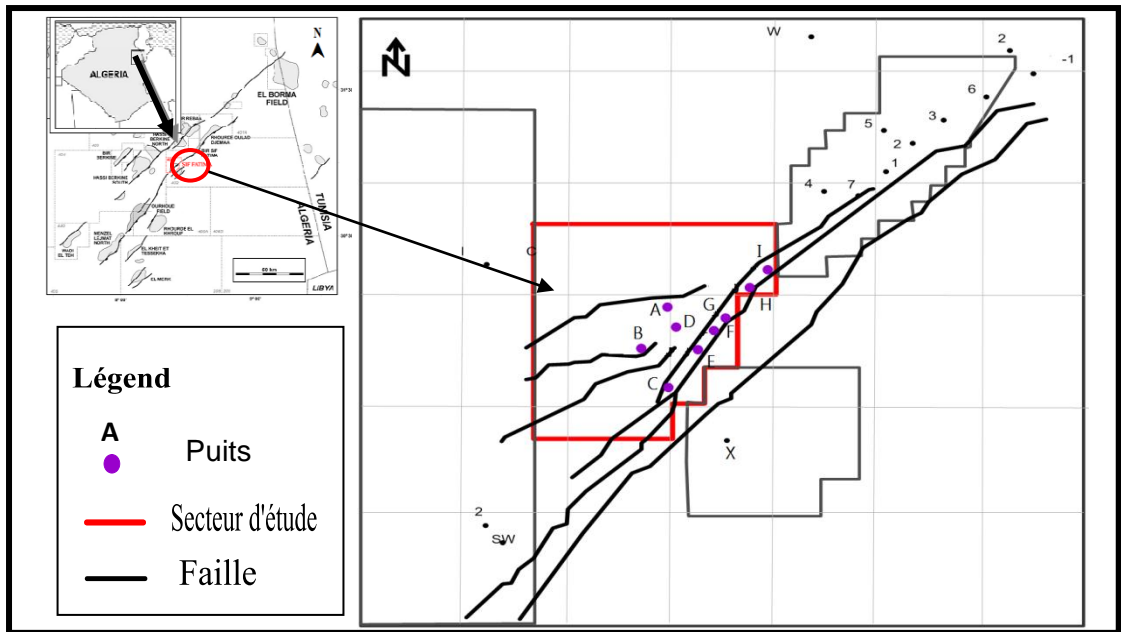


Fig. 31 - présentation de la tectonique du secteur de SIF FATIMA (Rapport Interne Sonatrach 2012).

À partir d'analyse des données sismiques on a: Les marqueurs sismiques représentés par le toit du S4 et TC et à priori TAGI ont été interprétés. Les premiers marqueurs constituent les réflecteurs les plus continus et énergétiques. (Fig. 32).

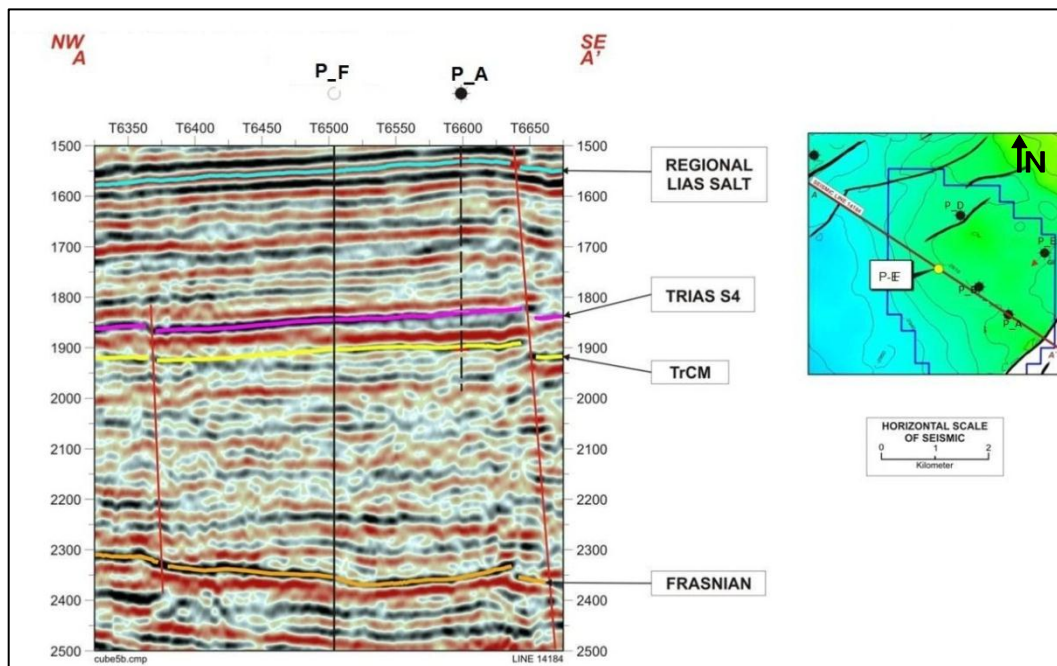


Fig. 32- Coupe sismique de secteur SIF Fatima. (Rapport Interne Sonatrach / Anadarko).

À partir La corrélation sismique qui a été faite sur les horizons sismiques suivants : Aptien, Malm, Lias salifère, S4, Trias carbonaté (T C), Trias Argileux Gréseux Inférieur (T.A.G.I), base du Frasnien (fig.16) et Les cartes structurales TAGI (fig.17) font ressortir les traits suivants :

Un réseau de failles d'orientation NNE/SSW dont la principale faille est connue sous le nom de faille d'Ourhoud, et une autre faille, antithétique de RDB, relativement proche située beaucoup plus à l'Est et à regard Ouest, formant avec cette principale faille un horst très étroit, constituant une terrasse intermédiaire entre le panneau de SF et la zone de grabben ouest. (Fig. 33)

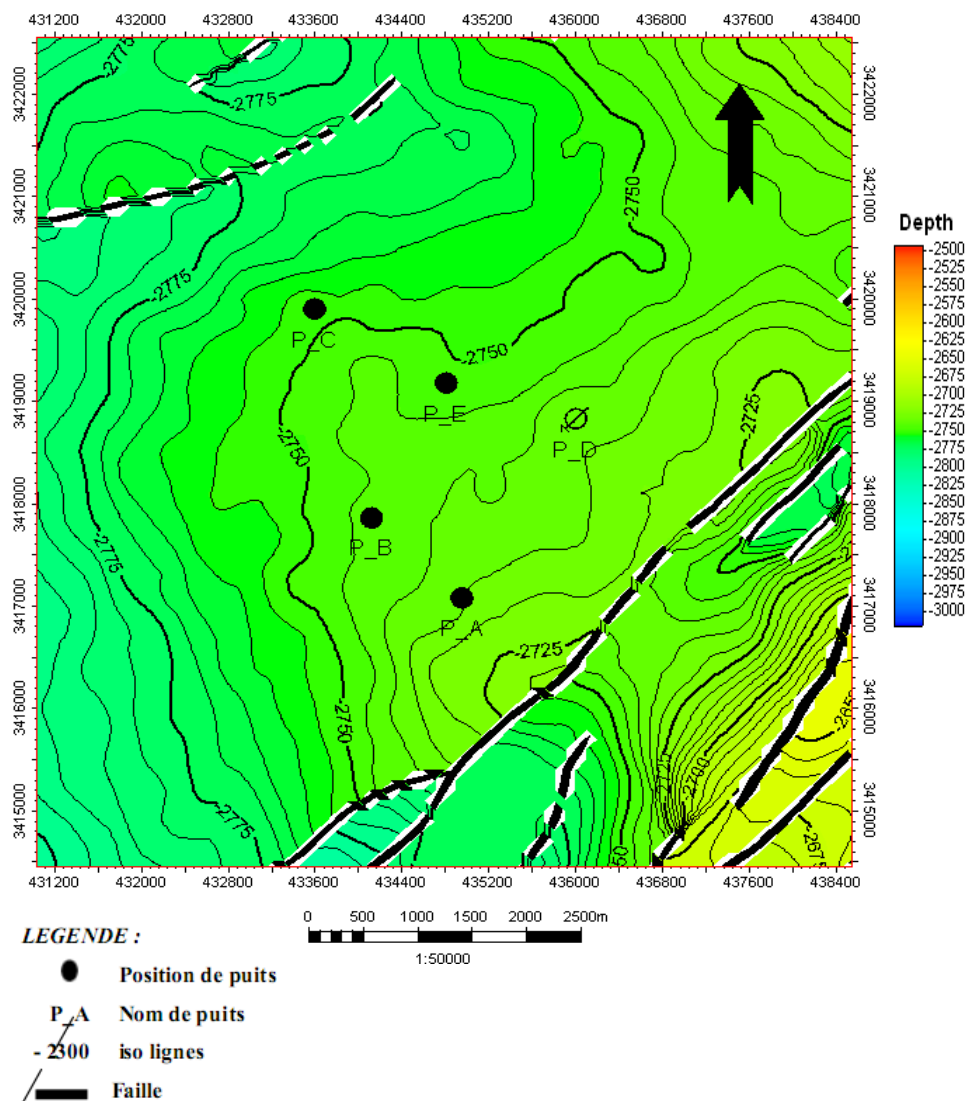


Fig. 33- Carte Structurale au Toit de TAGI réservoir du secteur SIF Fatima. (Sonatrach, Anadarko 2012).

2. 3. Le potentiel pétrolier de la zone d'étude

2. 3.1. Roche mère: Les principales roches mères de la région sont:

- Les argiles du frasnien.
- Les argiles du dévonien qui sont fortement productives .

2.3.2. Roche réservoir: Le principal réservoir dans la région d'étude est le Trias Argileux Gréseux Inferieur (TAGI) qui est subdivisé en trois niveaux qui sont de bas en haut.

a) TAGI basal : composé par:

- Des Argile brun rouge et gris vert à vert, parfois gris clair indurée légèrement silteuse localement micacée .
- Des grés blancs ,fin à moyen parfois blanc verdâtre, et friable à moyennement consolidé sub-anguleux à sub-arrondi pyriteux par endroits.
- grés gris-blanc a gris –clair moyen à grossier, sub-arrondi avec passe millimétrique d'argile verdâtre, indurée, silteuse et traces de pyrite.

b) TAGI moyen : composé par:

- Grés beige fin à moyen bien classe sub-anguleux à sub arrondi, parfois verdâtre très friable, fin siliceux à silico-argileux.
- Grés blanc à beige, sub-anguleux à sub-arrondi friable, micacé, siliceuse et pyriteux avec fines passées d'argile verte.

c) TAGI supérieur: Compose d'une alternance de grés blanc à beiges parfois verdâtre, Sub- anguleux à sub –arrondi, silico argileux, micacé; pyriteux moyennement dur à friable avec niveaux de galets argileux.

2.3.3. Roche couverture : est assurée par:

- Le Trias carbonaté.
- Les argiles radioactives du Trias argilo gréseux supérieur.

Conclusion

La région SIF FATIMA (bloc 402) est située dans l'erg orientale du Sahara Algérienne (bassin de BERKINE), Ce secteur couvre une superficie de 103Km². Il est limitée au Nord par Bir Rabaa Nord; au Sud par Rhoude Debdaba; à l'ouest par Bir Berkine; et à l'est par Bir SIF Fatima. Les formations sédimentaires sont représentées par des dépôts d'âge mésozoïque, reposant en discordance majeure sur des terrains d'âge paléozoïque, qui sont traversées partiellement, par un épandage détritique. Cet épandage est daté du tertiaire et surmonté en discontinuité avec le crétacé. L'ensemble est couvert par des sables dunaires attribués au mio-pliocène et au quaternaire.

Le secteur de SIF Fatima est caractérisé en générale, par une tectonique cassante relativement intense. La direction des failles majeures représentées par Ourhoud-SIF Fatima (NNE/SSW) et les faille secondaire ayant pratiquement la même direction. L'ensemble forme avec, une autre faille antithétique de Rhoude Debdaba (de l'Est à regard Ouest), un horst très étroit, constituant une terrasse intermédiaire entre le panneau de SF et la zone de grabben ouest RDB.

Le réservoir Trias Argileux Gréseux Inferieur (TAGI) est l'objectif principal de la structure de SIF Fatima. Cette structure est caractérisée par une séquence verticale d'origine fluviatile. Les découvertes d'huiles dans ce réservoir (TAGI) ont permis de mettre en évidence un gisement multicouche associé à des Grès fluviatiles présentant des caractéristiques pétrophysiques adéquates. Le bloc SIF Fatima ayant un autre réservoir objectif secondaire c'est le Trias carbonaté (TC).

Chapitre III

Matériel et Méthodes

Ce chapitre est subdivisé en trois parties.

En premier, nous présenterons la préparation de la base de données en se basant sur la description détaillée des carottes et la combinaison avec les diagraphies.

La deuxième partie est consacrée pour l'application des réseaux de neurones pour reconstituer les facies,

La troisième partie discute le modèle d'application de Cluster Analysis et la possibilité d'application pour la reconstitution les facies.

1. PREPARATION DE LA BASE DE DONNEES

La préparation de la base de données est basée sur le codage des observations sur carottes en chiffres selon les faciès disponibles et le calage avec les diagraphies suivant le même pas d'échantillonnage qui est 0,1524 m. pour les intervalles carottés.

1.1. Carottes

En forage, les déblais remontés par la circulation ne sont pas toujours représentatifs. Pour le géologue, les renseignements sont incomplets. Ces déblais peuvent provenir de niveaux différents (cas de retombées) et la vitesse de remontée des déblais est incertaine. Dans l'industrie de l'huile et du gaz, carotter peut être défini comme "couper et enlever un échantillon cylindrique de roche de la paroi du puits." (In Grondin, 2004). La prise d'un échantillon massif (carotte) important fournit au géologue : une certitude sur la côte et le terrain, la valeur du pendage, une appréciation de la dureté des terrains (important également pour le foreur), La possibilité de faire des mesures précises, de porosité, perméabilité, La possibilité de reconnaître avec certitude l'âge du terrain par l'étude des fossiles (micropaléontologie). Généralement, Il existe deux types de carottage à savoir: Le carottage axial: les carottes axiales sont obtenues en utilisant des carottiers. Le carottage latéral: le système de carottage latéral par percussion utilise une charge explosive cylindrique tirée dans les parois du trou pour obtenir un échantillon de roche (Fig. 34).



Fig. 34 -Photos des carottes puits P_A.

1.2. Diagraphies

La diagraphie est l'enregistrement continu d'une caractéristique des formations traversées par un sondage en fonction de la profondeur (Serra, O., 1979).

Les diagraphies différées sont enregistrées pendant l'arrêt du forage (les diagraphies instantanées sont celles qui sont enregistrées pendant le forage proprement dit). Les diagraphies sont des instruments essentiels dans l'évaluation des gisements.

Les buts principaux des diagraphies différées sont les suivants:

- La reconnaissance des réservoirs : lithologie, porosité, saturation (eau-huile-gaz). On n'obtient pas de valeurs de perméabilité.
- La connaissance du puits : diamètre, inclinaison, cimentation des cuvelages, liaison couche-trou (perforation).

Pour réaliser cet enregistrement, on utilise en particulier un camion d'enregistrement (ou cabine en off-shore), un treuil motorisé, un câble électrique et une sonde:

a. La sonde

C'est ensemble électronique avec émetteurs (électriques ou nucléaires), récepteurs et amplificateurs. Le rayon d'investigation est de l'ordre du mètre en général.

b. L'enregistrement

Le signal du récepteur est contrôlé, étalonné et enregistré. L'enregistrement s'appelle un log. Souvent les outils sont combinés (enregistrement simultané) et de toute façon les enregistrements des différents outils utilisés sont recalés les uns par rapport aux autres.

c. L'interprétation

Un micro-ordinateur permet une première interprétation rapide, sur camion, ce qui peut orienter le géologue dans ses prises de décision (autres diagraphies, carottage, essais).

Les diagraphies sont donc très utiles pour faire des corrélations de puits à puits et donnent des indications très précieuses sur les variations lithologiques. Les principales caractéristiques enregistrées sont: Potentiel spontané, Résistivité, Radioactivité naturelle (rayons gamma) et provoquée (neutrons), vitesse du son (log sonique), Diamètre du sondage, sa déviation, le pendage des couches.

Les diagraphies sont utilisées dans ce mémoire pour la prédiction des faciès non-carottés, Ces diagraphies sont décrites ci-dessous :

1.2.1. Les rayons gamma naturelle et le spectral gamma ray

Cet outil mesure la radioactivité naturelle des formations. Les seuls éléments radioactifs ayant une concentration notable dans les matériaux naturels sont le potassium, l'uranium et le thorium. Pour les formations sédimentaires on enregistre une radioactivité importante dans:

- Les formations argileuses qui renferment du potassium (illite spécialement).
- Les sels de potassium.
- Les formations riches en matière organique peuvent concentrer l'uranium.
- Les formations détritiques contenant des feldspaths (potassium) ou enrichies en minéraux lourds.

Donc, il permet de mettre en évidence les formations argileuses normalement compactées et sous-compactées, de calculer le pourcentage d'argile contenu dans une formation et de faire des corrélations de puits à puits.

Le gamma ray est fortement perturbé par l'utilisation de boue au chlorure de potassium et par un trou cavé. Il existe deux types de mesures:

1.2.1.1. La radioactivité naturelle GR:

Elle est sensible à l'énergie totale de radiation des éléments radioactifs. Les plus prépondérants sont le ^{40}K , le ^{232}Th et le ^{238}U . La réponse de l'outil GR est, par conséquent, fonction de la concentration et de la densité respective de ces éléments radioactifs. Elle est exprimée par la relation:

$$\text{GR} = A \cdot V \cdot \rho / \rho_b$$

Avec:

- GR: radioactivité globale mesurée.
- ρ : densité de l'élément radioactif.
- ρ_b : densité globale de la formation.
- V: Pourcentage de l'élément dans la formation.
- A: facteur de proportionnalité correspondant à l'activité gamma.

1.2.1.2. Le spectral gamma ray

Il permet l'enregistrement, en plus, de l'énergie totale, les concentrations spécifiques des pics des éléments radioactifs ^{40}K , ^{232}Th et ^{238}U , ainsi, que l'énergie de radiation du Th+K. Cinq courbes sont mesurées. Il s'agit du SGR (énergie totale en API), CGR (énergie de Th+K en API), U (concentration en uranium en ppm), Th (concentration en thorium en ppm) et en fin, K (concentration en potassium en %) (Fig. 35).

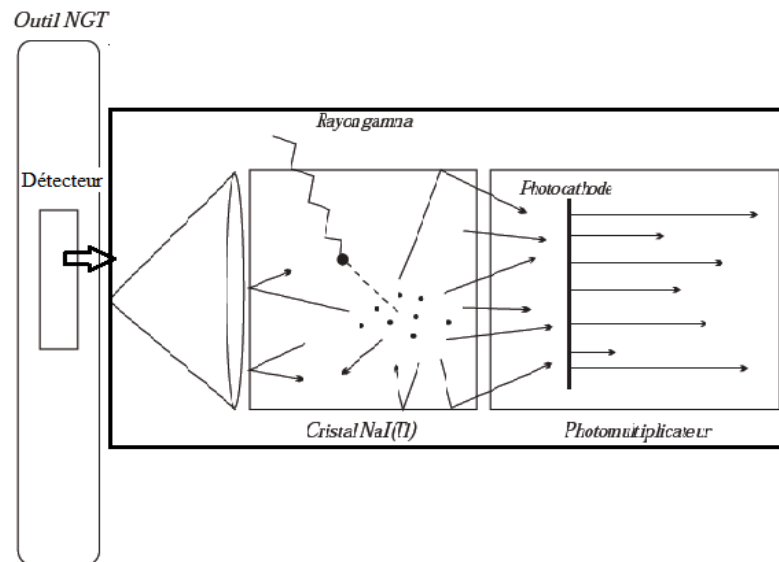


Fig. 35 - Outil de mesure de la radioactivité naturelle (Lagabrielle, 2008).

1.2.1.3. La densité (gamma-gamma γ - γ)

Ces diagraphies utilisent des sources radioactives pour l'émission de rayonnement gamma. Celles-ci peuvent être ^{60}Co ou ^{137}Cs . Les rayons gamma émis rentrent en collision avec les électrons des atomes des formations traversées. La perte d'énergie à chaque interaction passe par trois stades d'énergie qui sont : la production paire, l'effet photoélectrique et Compton. Ce sont ces deux derniers qui intéressent les diagraphies de densité (Fig.36). En effet dans le domaine des faibles énergies, caractérisé par l'effet photoélectrique, les diagraphies de densité enregistrent le coefficient d'absorption photoélectrique P_e (barns/e-). Il est exprimé, dans le cas d'un élément simple, par la relation : $P_e = (Z/10)^{3.6}$, Pour une substance moléculaire la relation devient : $P_e = \sum A_i Z_i P_i / \sum A_i Z_i$. Avec A_i , Z_i et P_i sont, respectivement, la masse atomique, le numéro atomique et le facteur photoélectrique des éléments simples constituant la molécule.

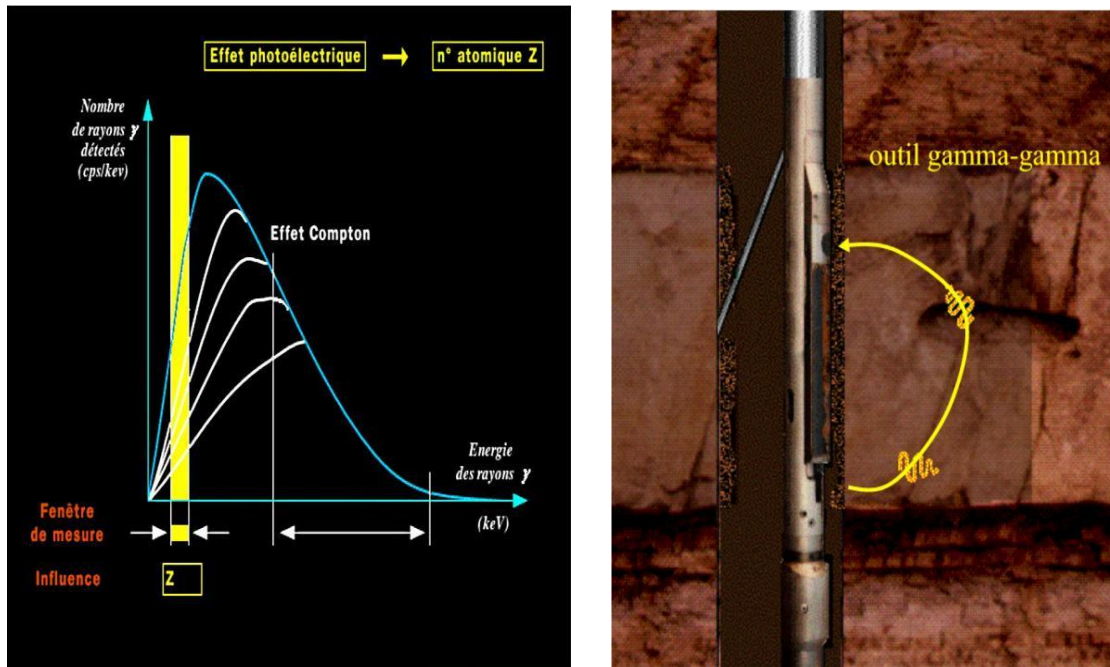


Fig. 36 - Outil de mesure de la densité et Exemple de log gamma-gamma (densité)
(Document Schlumberger).

1. 3. Validation et codage des données

Concernant la validation des données, il s'agit dans un premier temps de vérifier la validité des logs (échelle des valeurs, présence de bruit ou de pics anomalies qui correspondent généralement à des zones cavées ou à des fonds de trous) puis de vérifier le calage en profondeur de ces logs.

Le recalage des carottes (cotes sondeurs) par rapport aux cotes diagraphiques (cotes logs) est indispensable à une bonne calibration des électrofaciès. Plusieurs tests ont été effectués afin de choisir les logs à utiliser pour la détermination des électrofaciès parmi lesquels:- GR, NPHI, RHOB,- GR, NPHI, RHOB, DT, et - GR, NPHI, RHOB, DT, THOR, POTA.

Il est donc nécessaire d'effectuer au préalable des tests visant à déterminer le choix des logs à utiliser. Les meilleurs résultats ont été obtenus avec deux jeux de logs: « GR, RHOB, POTA, THOR ». Nous avons pu identifier et discriminer ces faciès d'un point de vue lithologique uniquement, et regrouper ces faciès en quatre faciès deux faciès non réservoirs : (Argile et Silt) et deux faciès réservoirs: (Grès et Grès-Argileux).

Des codes ont été attribués de 0 à 3, pour simplifier la nomenclature de ces faciès, définis comme suit: (Fig. 37)


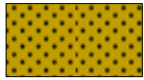


	Argiles	Code = 0
	Silts	Code = 1
	Grès	Code = 2
	Grès Argileux	Code = 3

Fig. 37 - Codage des Facies du Réservoirs TAGI.

Une fois les descriptions de carotte homogénéisées, les facies reconnus ont été regroupés en 4 principaux facies, des codes ont été attribués afin de mieux les discriminer.

On a élaboré une banque de données provenant de la combinaison des digraphies nucléaires (Gamma Ray, Thorium, Potassium, et Densité) avec la description et le codage des facies (Argiles (0), Silts (1), Grès (2), et Grès Argileux (3)) dans les 5 puits (P_A, P_B, P_C, P-D, P-E), (Tableau. 5).

Tableau. 5 : illustration du codage des facies.

DEPTH.M	GR(A)	RHOB(B)	THOR(C)	POTA(D)	FACIES CAROTTE
3008.1983	139.5226	2.5149	8.8933	5.0200	0
3008.5031	140.7935	2.5994	8.916	5.0555	0
3008.6555	135.8285	2.6049	8.6119	5.0331	1
3008.9603	120.9756	2.5304	7.9106	4.9461	0
3009.1127	126.0833	2.5217	7.9145	5.0242	0
3009.2651	138.0849	2.5213	8.1457	5.1348	0
3009.4175	135.8914	2.543	7.975	5.1045	0
3009.5699	130.7323	2.5632	7.8935	5.1021	1
3009.7223	122.7028	2.5599	7.1232	4.7162	0
3010.0271	118.9685	2.5192	6.2386	4.2687	1
3010.4843	105.3975	2.5366	5.5401	4.1142	1
3010.7891	95.1815	2.5206	5.3184	4.0747	2
3010.9415	91.288	2.5318	5.4333	4.0816	2
3011.0939	93.3739	2.5364	5.1698	4.0171	Intervalle non carottée
3011.2463	94.4379	2.5305	4.8969	3.9578	
3011.3987	93.2152	2.508	4.6474	3.5629	
3011.8559	62.1277	2.4299	2.9624	3.0769	
3012.0083	61.0556	2.4311	3.5374	3.0881	
3012.1607	65.4611	2.4222	3.847	3.0742	
3012.3131	71.1185	2.4043	4.1769	3.0713	
3012.6179	72.1212	2.4087	4.1649	2.9403	
3012.7703	68.9092	2.4299	4.155	2.8907	
3013.3799	63.6058	2.3789	3.6102	2.4353	
3013.5323	59.5139	2.3647	3.6816	2.3466	

On dispose finalement un fichier de données de toutes les puits (3426 observations), avec 1242 observations pour les intervalles carottés, (enregistrements de digraphies et codage des facies), et pour les puits et les intervalles non carottés 2184 observations, seulement les enregistrements des diagraphies et (Missing Data) qui sont le but de travail pour les reconstituer). Alors on a 189.2 m sont des intervalles carottés et 332.8 m sont des intervalles non carottés.

2. LOGICIELS UTILISES

2.1. Logiciel "EXCEL"

Ce logiciel qui est un tableur a été utilisé essentiellement pour la saisie et la gestion des fichiers des données. Les fichiers de données diagraphie de tout le réservoir ont été faits au préalable à l'aide de ce logiciel. En générale, il a été utilisé pour certains calculs (méthode de Cluster Analysis) et aussi pour qu'ils soient bien adapter à tout traitement statistique et se mettre en liaison avec Matlab.

2.2. Logiciel "MATLAB"

MATLAB est une application scientifique interactive orientée au calcul vectoriel et matriciel avec une puissante librairie de visualisation. On peut résoudre des problèmes de calcul très complexes d'une façon simple et rapide comparée aux langages de programmation traditionnels. Le nom MATLAB est dérivé de l'anglais MATrix LABoratory Ce logiciel comporte une Toolbox relatifs aux réseaux de neurones artificielles qui a été utilisée dans la programmation.

2.3. Logiciel "PETREL 2009"

Petrel est une application complète de simulation développé par Schlumberger. Entièrement intégré avec les outils géologiques et de technologie, le logiciel tient compte d'une interprétation sismique 2D et 3D rapide. Prélever vos données sismiques directement dans le modèle du réservoir, Ce logiciel a été utilisé dans notre travail pour la visualisation de résultats obtenus des deux méthodes utilisés.

3.4. ELABORATION DU MODELE PAR LES RESEAUX DE NEURONES

3.4.1. Les Phases

Elles se subdivisent en sept étapes :

1. La collecte des données.
2. L'analyse des données.
3. La séparation des bases de données (apprentissage, validation, test).
4. Le choix de l'architecture du réseau de neurones.
5. L'apprentissage.
6. La validation.
7. Le test.

a) Collecte des données

L'objectif de cette étape est de recueillir des données, à la fois pour développer et valider le modèle. Le but est de rassembler un nombre de données suffisant pour constituer une base représentative nécessaire à la modélisation.

b) Analyse des données

Celle-ci consiste à détecter d'éventuelles valeurs douteuses de la base de données recueillies qui peuvent avoir une répercussion sur les résultats de la modélisation.

c) Séparation de la base de données

La base de données est scindée en trois parties, la première servant comme base d'apprentissage), et la seconde utilisée comme base de validation (performances du modèle) et la troisième pour (test). Il n'y a pas de règle pour déterminer ce partage de manière quantitatif. Il résulte souvent d'un compromis tenant compte du nombre de données dont on dispose et du temps imparti pour effectuer l'apprentissage. Chaque base doit cependant satisfaire aux contraintes de représentativité de chaque classe de données, et doit généralement refléter la distribution réelle, c'est à dire la probabilité d'occurrence des diverses classes.

d) Choix de l'architecture du réseau

Il existe un grand nombre de types de réseaux de neurones, avec pour chacun des avantages et des inconvénients. Le choix de l'architecture du réseau peut dépendre :

- ❖ De la tâche à effectuer (classification, association, contrôle de processus, ...)
- ❖ De la nature des données (données présentant des variations au cours du temps).

Ce choix est aussi fonction de la maîtrise, ou de la connaissance que l'on a de certains Réseaux, ou encore du temps dont on dispose pour tester une architecture prétendue plus performante.

e) Apprentissage du réseau de neurones

La plupart des modèles de réseaux de neurones requièrent un apprentissage. Plusieurs types d'apprentissages peuvent être adaptés à un même type de réseau de neurones. Les critères de choix sont souvent la rapidité de convergence, ou les performances de généralisation.

f) Validation

Une fois le réseau de neurones entraîné, il est nécessaire de le tester sur une base de données différente de celle utilisée pour l'apprentissage. Ce test permet d'apprécier les performances du modèle élaboré. Si ces performances ne sont pas satisfaisantes, il faudra soit modifier l'architecture du réseau, soit modifier la base d'apprentissage. La démarche expérimentale exige cependant une étape de validation avant toute utilisation. Elle est en effet nécessaire, pour vérifier la fiabilité de la représentation interne du réseau. Elle consiste en fait à simuler l'utilisation du réseau.

Cette validation est effectuée à l'aide de couples entrées-sorties, autres que ceux utilisés lors de l'apprentissage. Évidemment, le réseau étant par définition un outil d'interpolation, les valeurs des entrées et des sorties de ces couples, tout comme celles des entrées fournies lors de l'utilisation, appartiennent aux mêmes intervalles que celles des expériences qui ont servies à l'apprentissage.

La validation consiste alors à fournir au réseau de neurones les entrées de ces couples, à partir desquelles il calcule (grâce aux poids définis à l'apprentissage) les sorties correspondantes. Ce qui permet de constater l'erreur commise par le réseau de neurones.

Cette étape de validation est particulièrement utilisée lors de l'ajustement d'un réseau de neurones à la situation à modéliser, puisqu'il quantifie l'erreur commise. Elle est donc systématiquement réalisée à chaque test effectué lors de la détermination de l'architecture optimale et des autres paramètres (fonction d'activation par exemple); Il s'agit d'une évaluation de l'erreur de généralisation. La différence entre l'erreur d'apprentissage et l'erreur de généralisation représente une mesure de la qualité de l'apprentissage effectué.

L'erreur de généralisation dépend avant tout de trois paramètres : le nombre d'exemples utilisé pour l'apprentissage, la complexité du problème sous-jacent et l'architecture du réseau.

Les approches statistiques de la généralisation sont l'un des domaines d'investigation majeur pour optimiser les performances de l'apprentissage des réseaux de neurones. On constate qu'une amélioration de la généralisation peut être observée sous deux aspects imbriqués :

- ❖ Si la taille du réseau est fixée, quel est l'ensemble d'exemples d'apprentissage qui donnera la meilleure généralisation?
- ❖ Si le nombre des exemples est fixé, comment choisir le réseau pour avoir la meilleure généralisation?

g) Test

En réalité il n'y a pas grande différence entre la validation et le test puisque il réalise la même opération, Seulement la phase de validation est utilisée aussi pour éviter le problème de sur apprentissage (Overfitting) et augmenter la capacité de généralisation du modèle. Pendant l'arrêt de l'apprentissage du modèle, on constate que l'erreur totale sur la partie validation cesse de se dégrader, donc à ce point-là l'apprentissage est stoppé ce qui s'appelle l'Early Stopping en Anglais.

3.4.2. Critères de validation.

Une fois le modèle construit se pose alors la question s'il rend fidèlement compte de l'échantillon. Pour cela, la modélisation nécessite des critères de validation sur la base desquels, le concepteur juge les performances finales du modèle qui sera retenu. Certains sont numériques et d'autres graphiques. Willmott CJ (1981,1982).in I. Ladlani et al. (2012).

a) Les critères numériques

❖ Le coefficient de détermination (R^2)

Celui-ci évalue le degré d'association entre deux variables et juge de la qualité de l'ajustement des valeurs observées et calculées :

$$R^2 = \left[\frac{\frac{1}{N} \sum_{i=1}^N (Y_{iobs} - \bar{Y}_{obs})(Y_{ical} - \bar{Y}_{cal})}{\sigma_{obs} \cdot \sigma_{cal}} \right]^2 \quad (1)$$

Avec :

Y_{iobs} : Valeur observée.

Y_{ical} : Valeur calculée par le modèle.

\bar{Y}_{obs} : Moyenne de la variable observé.

\bar{Y}_{cal} : Moyenne de la variable calculée.

N : est le nombre de points.

Un coefficient de détermination proche de l'unité signifie que le modèle élaboré est très efficace.

❖ Erreur Quadratique Moyenne, Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (Y_{iobs} - Y_{isim})^2} \quad (2)$$

Quand le RMSE est proche de zéro signifie que le modèle est très bon.

❖ **Erreur Absolue Moyenne, Mean Absolute Error (MAE)**

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_{iobs} - Y_{isim}| \quad (3)$$

Lorsque le MAE est proche de zéro signifie que le modèle est très bon.

❖ **Coefficient de Corrélacion (CC)**

$$CC = \frac{\frac{1}{N} \sum (Y_{iobs} - Y_{isim})(Y_{iobs} - Y_{isim})}{\sqrt{\frac{1}{N} \sum (Y_{iobs} - Y_{isim})^2} \sqrt{\frac{1}{N} \sum (Y_{iobs} - Y_{isim})^2}} \quad (4)$$

Un CC proche de 1 signifie que le modèle est très bon.

❖ **Les critères graphiques**

L'analyse graphique est indispensable et primordiale, cela est obtenu en portant sur un graphique les valeurs observées et celles calculées. la fonction coût ou l'arrêt précoce (*Early Stopping*) pour éviter le sur apprentissage (Morgan, et Boulard 1990).

3.4.3. Méthodologie de Programmation

L'élaboration de tous types de modèle mathématique nécessite obligatoirement, le recours aux programmes mathématiques, capable de prendre en large le volume important des informations à traiter.

Le domaine de réseaux de neurones artificiels est l'un des domaines qui exige le recours au programme informatique.

Dans le cadre de notre travail nous avons utilisé le logiciel MATLAB, développé par la société américaine MATWORK.

3.4.4. Algorithmes

1. Ouverture et lecture des fichiers de données.
2. Définition des paramètres du réseau (nombre de neurones, taux d'apprentissage, bornes des poids maximum et minimum, nombre d'itérations).
3. Initialisation aléatoire des poids.
4. Propagation en avant et calcul de l'erreur totale.
5. Rétro propagation en arrière.
6. Arrêt de l'apprentissage (critère de convergence atteint).
7. Vérification des performances du modèle (critères graphiques, critères numériques (RMSE, MAE, CC et le d)).
8. Modèle de reconstitution des Missing Data.

On présentera ainsi tour-à-tour tous les motifs de la base d'apprentissage, jusqu'à convergence de l'algorithme.

3.4.5. Applications

a) la Base de données.

On a élaboré une banque de données provenant de la combinaison des digraphies nucléaires (Gamma Ray, Thorium, Potassium, et Densité) avec la description et le codage des facies (Argiles (0), Silts (1), Grès (2), et Grès Argileux (3) dans les trois puits carottés (P_A, P_B, P_C). Alors on a retenu huit paramètres comme variables d'entrées les quatre digraphies sont considérés des entrés (des variables explicatifs), les quatre facies sont pris pour faire l'apprentissage.

On dispose finalement un fichier de données de toutes les puits (3426 observations), pour les puits carottés 1242 observations des 4 entrés (réponses de digraphies et codage des facies, pour les puits et les intervalles non carottés 2184 observations

b) L'architecture du réseau de neurone

La couche d'entrée est constituée de quatre variables qui sont les diagraphies (Gamma Ray, Thorium, Potassium, et la Densité). Une seule couche cachée, constituée de 13 neurones avec une fonction de transfert sigmoïdale. La couche de sortie est constituée d'un seul neurone qui est le Facies qu'on le cherche dans les intervalles non carottés (Fig. 38).

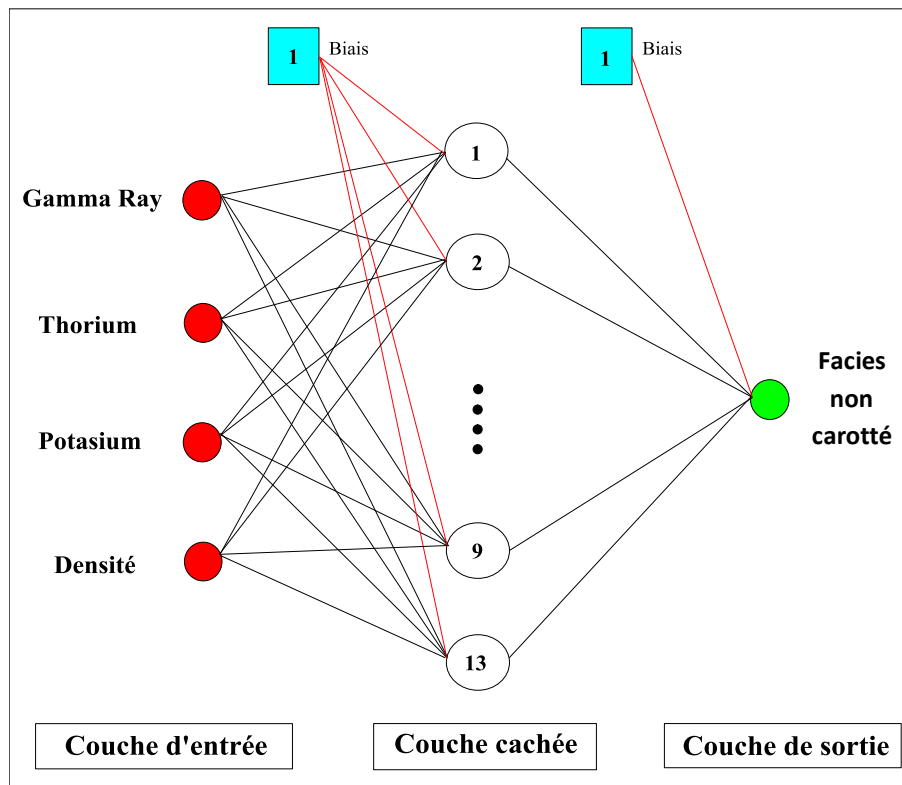


Fig. 38 - Architecture du réseau de neurones élaboré.

c) L'algorithme du rétropropagation du gradient

C'est une méthode qui permet de calculer le gradient de l'erreur pour chaque neurone d'un réseau de neurones, de la dernière couche vers la première. De façon abusive (Rumelhart et Hinton 1986).

d) Statistique des paramètres étudiés

Nous avons procédé à l'analyse statistique des paramètres retenus dans le cadre de cette étude, les résultats obtenus sont résumés dans le tableau suivant :

Tableau. 6 : Statistique des paramètres étudiés

	GR(A)	RHOB(B)	THOR(C)	POTA(D)
Moyenne	81.4866	2.5398	6.9686	2.3319
Erreur-type	1.5608	0.0046	0.1951	0.0440
Médiane	71.7729	2.5395	5.2769	2.2797
Écart-type	31.6040	0.0938	3.9513	0.8917
Variance	998.8144	0.0088	15.6126	0.7951
Minimum	13.1673	2.3604	0.7851	0.0714
Maximum	140.8081	2.7203	15.3682	5.1348
Somme	33409.5027	1041.3347	2857.1118	956.0851

e) Application des réseaux de neurones

L'élaboration d'un modèle neuronale passe à la fois par le choix des paramètres d'entrée mais aussi par l'optimisation de l'architecture du réseau en lui-même. Ceci signifie qu'il faut trouver à la fois la meilleure répartition des données entre l'apprentissage, le test et la validation, et le nombre de neurones constituant la couche d'entrée. L'optimisation à la fois du type de fonction non linéaire et de la géométrie du réseau de neurones a été faite après de nombreux essais empiriques. Nous avons fait varier le nombre de neurones au niveau de la couche cachée de 1 à 20, et on a arrêté à 13 neurones. Chaque réseau a été entraîné avec la fonction sigmoïdale. La couche de sortie est constituée d'un seul neurone qui est le facies à laquelle on a appliqué la fonction linéaire. Parmi les différentes configurations testées du réseau, nous avons retenu celle produisant la plus faible racine carrée de l'erreur quadratique moyenne (RMSE), ainsi que le meilleur coefficient de détermination R^2 pour la partie validation. mean absolute error MAE.

Le fichier d'entrée (constitué de 1242 observations) a été scindé en trois, plus la partie Missing Data (2184).

Tableau. 7 : Séparation du fichier de base.

Fichier	Observations
Total	1242
Apprentissage	742
Validation	250
Test	250
Missing Data	2184

3.4.6. Amélioration du réseau

Des tentatives d'améliorations ont été réalisées on ajoute à chaque fois un autre neurone dans la couche cachée, et on a suivi l'évolution des performances du modèle élaboré. Les résultats trouvés sont les suivants :

Tableau. 8 : Performances du modèle élaboré.

Nbr de neurones	Apprentissage			Validation			Test		
	CC	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE
1	0.84	0.52	0.28	0.76	0.57	0.30	0.86	0.50	0.27
2	0.46	0.87	0.75	0.07	0.91	0.85	0.12	0.97	0.91
3	0.84	0.52	0.30	0.76	0.58	0.35	0.86	0.51	0.30
4	0.87	0.48	0.27	0.78	0.55	0.30	0.92	0.39	0.23
5	0.84	0.54	0.37	0.74	0.60	0.37	0.85	0.55	0.40
6	0.83	0.55	0.31	0.74	0.59	0.34	0.84	0.55	0.33
7	0.84	0.53	0.32	0.76	0.58	0.35	0.86	0.51	0.31
8	0.85	0.51	0.30	0.76	0.58	0.33	0.88	0.47	0.29
9	0.85	0.51	0.32	0.74	0.59	0.38	0.86	0.50	0.33
10	0.86	0.50	0.28	0.76	0.58	0.33	0.88	0.47	0.27
11	0.86	0.50	0.30	0.76	0.58	0.36	0.88	0.47	0.30
12	0.87	0.49	0.27	0.76	0.58	0.34	0.87	0.48	0.29
13	0.84	0.53	0.37	0.74	0.59	0.36	0.85	0.55	0.40

Les performances du modèle changent à chaque ajout d'un neurone dans la couche cachée. Les meilleurs résultats sont obtenus à l'ajout de quatre neurones. Les indices numériques de ce modèle sont : un coefficient de corrélation (CC) de 0,87 pour l'apprentissage, 0,78 pour la validation et 0,92 pour le test.

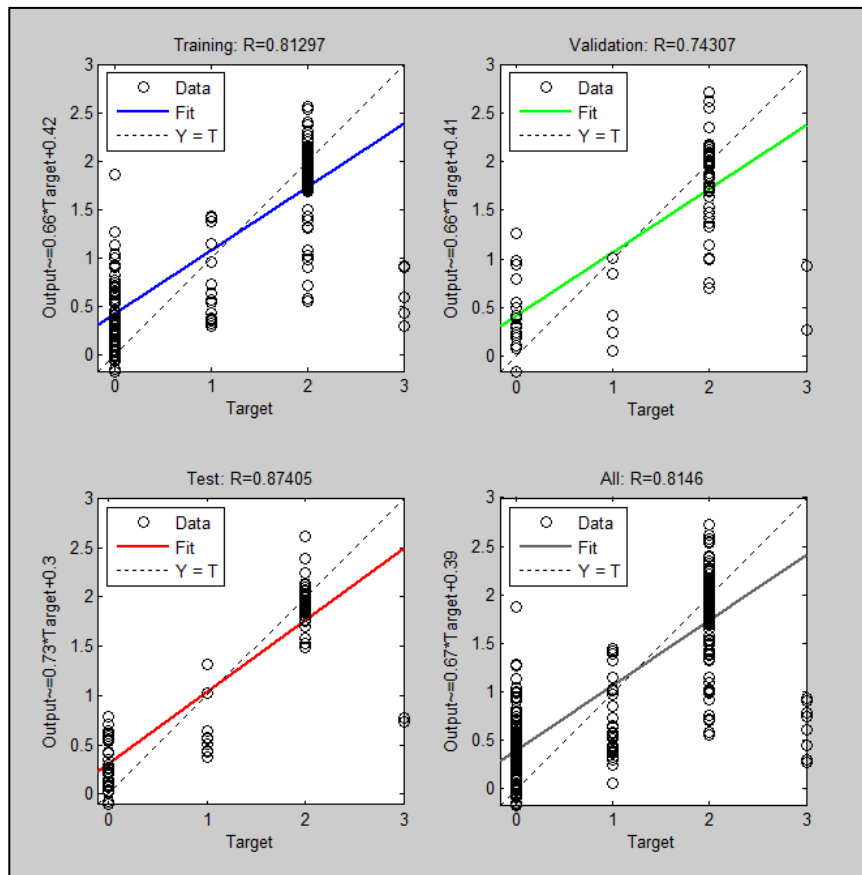


Fig. 39- Schéma de régression montrant les coefficients de corrélation entre les faciès observés et les faciès estimés pour des étapes de l'apprentissage, de validation et le test.

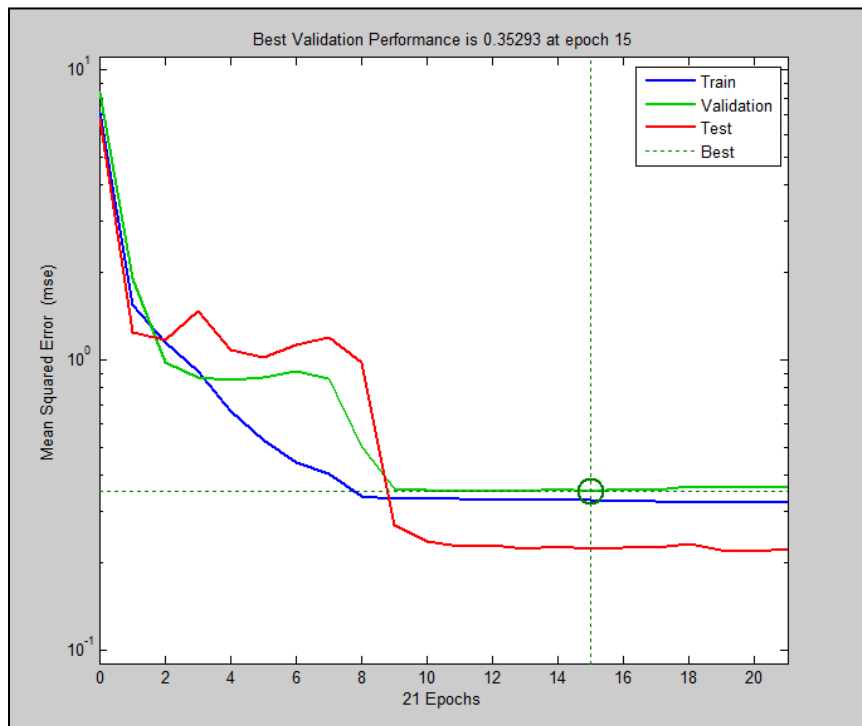


Fig. 40 - Diagrammes de performance des étapes d'apprentissage (ligne bleue), de validation (ligne verte), et du test (ligne rouge). (Fonction coût) La meilleure performance de la validation est égale à 0.35293 et s'arrête à 15 itérations.

La figure (40) montre la fonction coût (l'arrêt précoce), qui mesure l'écart entre la sortie du modèle (fonction réalisée par le réseau de neurones) et la sortie désirée. La fonction de coût est une fonction scalaire qui dépend du vecteur de paramètres du modèle, et des individus de l'ensemble d'apprentissage ; dans le cas des réseaux de neurones, le vecteur de paramètres est constitué par les poids du réseau. Plus la valeur de la fonction de coût est petite, plus le modèle reproduit fidèlement les observations utilisées pour l'apprentissage. Les différents algorithmes d'apprentissage cherchent donc à trouver le point, dans l'espace des paramètres, pour lequel la fonction de coût est minimale. Et au-delà le modèle se diverge et se produit le surapprentissage.

Les paramètres retenus par le modèle élaboré sont présentés dans le tableau suivant :

Tableau. 9 : Les Paramètres retenus.

N° Neurone	Poids de la couche d'entrée					Neurone de sortie	
	Gamma Ray	Densité	Thorium	Potassium	biais	poids	biais
1	2.95	-36.08	-3.82	25.26	4.60	0.17	1.99
2	-5.50	-47.03	3.11	-0.05	35.25	0.69	
3	-9.86	87.74	2.45	-0.93	-20.27	-0.49	
4	0.59	23.84	-12.31	-13.89	6.41	1.83	
5	8.46	21.89	-1.99	-3.76	-18.02	-0.89	
6	0.19	147.48	0.52	0.37	-53.98	0.34	
7	7.69	-50.16	0.69	12.15	11.37	-1.08	
8	-5.37	93.57	-17.78	-0.82	-18.28	0.39	
9	4.96	60.07	-3.39	-1.54	-32.76	-2.40	
10	15.35	-38.26	15.84	-1.80	25.22	-3.79	
11	2.95	-36.08	-3.82	25.26	0.98	3.06	
12	-5.50	-47.03	3.11	-0.05	62.93	0.41	
13	-9.86	87.74	2.45	-0.93	44.39	-0.71	

3.4.7. Code source du modèle réalisé (Voir l'annexe N° 1)

Les résultats obtenus ont été visualisés par le logiciel PETREL. Voir le chapitre Résultats et Discussion.

4. ELABORATION DU MODELE PAR CLUSTER ANALYSIS

4.2. Méthodologie

Comme mentionné précédemment, la méthode "Cluster Analysis" se base sur l'utilisation des diagraphies dans la discrimination des facies, les facies reconnus sont dits : électrofaciès. Un électrofaciès correspond donc à un ensemble d'échantillons ayant des réponses diagraphiques similaires. La détermination des électrofaciès est donc fortement dépendante des logs choisis, de l'intervalle stratigraphique. Il est aussi nécessaire d'effectuer au préalable des tests.

Déterminer le choix des logs à utiliser qui sont : GR, RHOB, POTA et THOR.

La méthode Cluster Analysis est un partitionnement de données (data cluster ING en anglais) qui est une des méthodes statistiques d'analyse des données. Elle vise à diviser un ensemble de données en différents « paquets » homogènes, en ce sens que les données de chaque sous-ensemble partagent des caractéristiques communes. (Moura, 1985).

Nous avons pu identifier et discriminer ces facies d'un point de vu lithologique uniquement, et regrouper ces facies en 4 facies: (02 facies non réservoirs : Argile et Silt). (02 facies réservoirs: Grès et Grès-Argileux).

Des codes ont été attribués de 0 à 3, pour simplifier la nomenclature de ces facies, définis comme.

Pour ce qui est de la démarche, dans un premier temps, il a fallu déterminer pour chaque facies, les différences entre la valeur moyenne des logs (VA, VB, VC, VD) et leurs valeurs actuelles à une profondeur définie (LA, LB, LC, LD). On obtient de Ce fait un "déviations log" par facies.

Ensuite, nous avons calculé la valeur absolue de cette différence et nous avons normalisée en utilisant le range pour chaque variable. La "déviations log" prend alors des valeurs entre [0, 1], permettant d'établir des comparaisons.

Puis nous avons calculé l'erreur pour chaque facies, qui correspond à la moyenne des quatre "déviations log" et attribué le facies ayant la valeur minimale des erreurs pour chaque Profondeur (Fig. 41).

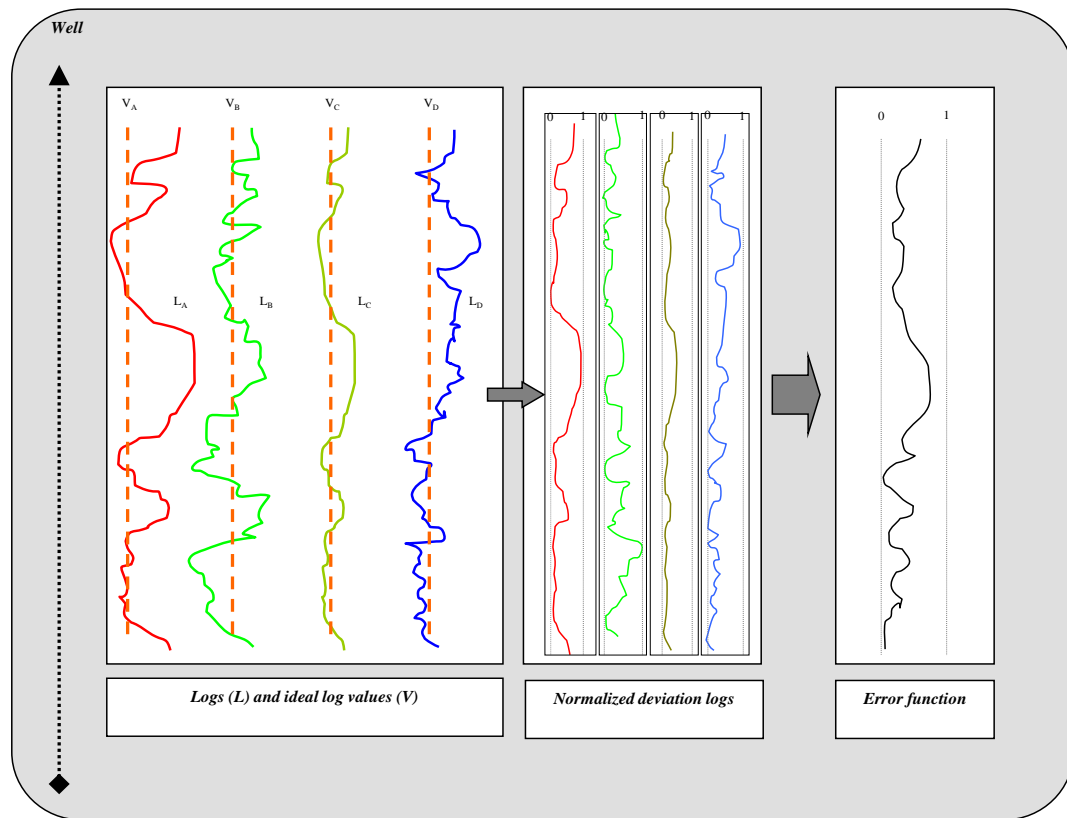


Fig. 41 - Schéma qui résume les étapes de la méthode « Cluster Analysis »

La prédiction des faciès a été effectuée à partir de la combinaison des logs : GR, RHOB, POTA et THOR des puits P_A, P_B, P_C, P_D, P_E. (Ameur Zaimeche, O. 2012, 2014).

Les valeurs moyennes de chaque log conventionnel utilisé ont été calculées pour chaque faciès sur Excel par le calcul de la moyenne de toutes les valeurs pour chaque diagraphie de chaque puits. Les paramètres moyens de chaque électrofaciès ont été reportés dans le (Tableau.10)

Tableau. 10: Moyennes des logs dans tous les puits carottés.

FACIES	CODE	GR (A)	RHO (B)	THOR (C)	POTA (D)
Argile	0	107.82	2.60	10.32	2.94
Silt	1	103.22	2.61	7.98	3.53
Grès	2	61.12	2.50	4.87	1.96
Grès argileux	3	108.96	2.59	10.16	3.32

La valeur maximale, minimale et le range se calculent en tenant compte de tous les valeurs à travers tous les puits.

Tableau. 11 : Max, Min et Range des logs dans tous les puits carottés.

	GR (A)	RHO (B)	THOR (C)	POTA (D)
MAX	174.17	2.87	18.41	5.58
MIN	13.17	2.24	0.50	0.03
RANGE	161.01	0.63	17.91	5.55

Tableaux 3&4 Les valeurs moyennes de tous les logs avec les valeurs maximales, minimales et les ranges pour le modèle

Excel calcule un plot 1D qui permet d'identifier des ensembles de points (ou pics statistiques) qui ont les mêmes caractéristiques logs. Chaque pic correspond à un ensemble multidimensionnel de points (plusieurs logs, puits et intervalles). Les pics les plus représentatifs sont sélectionnés comme des échantillons d'apprentissage ("training samples").

Ces échantillons déterminés sur les logs sont simultanément calibrés avec les faciès issus des descriptions de carottes. Ils ne sont conservés que s'ils sont bien définis en termes de lithologie. Les paramètres moyens de chaque électrofaciès ont été reportés dans les tableaux suivant :

Tableau. 12: illustration des calculs effectués pour les Argiles.

MODELE					
FACIES	CODE	GR (A)	RHOB (B)	THOR (C)	POTA (D)
Argile	0	107.82	2.60	10.32	2.94

Range	161.0061	0.6313	17.9122	5.5526														
DEPTH.M	GR(A)	RHOB(B)	THOR(C)	POTA(D)	(A)-(A)I	ABS[D(A)]	N(A)	(B)-(B)I	ABS[D(B)]	N(B)	(C)-(C)I	ABS[D(C)]	N(C)	(D)-(D)I	ABS[D(D)]	N(D)	error(f)0	
2997.9875	114.5871	2.5575	8.6128	4.991	6.77	6.77	0.04	-0.040467616	0.04	0.06	-1.706069395	1.71	0.10	2.048050534	2.05	0.37	0.143	
2998.1399	115.5988	2.5629	8.5227	4.9707	7.78	7.78	0.05	-0.035067616	0.04	0.06	-1.796169395	1.80	0.10	2.027750534	2.03	0.37	0.142	
2998.2923	120.4777	2.5603	8.6468	4.9962	12.66	12.66	0.08	-0.037667616	0.04	0.06	-1.672069395	1.67	0.09	2.053250534	2.05	0.37	0.150	
2998.4447	118.0204	2.5568	8.9964	5.0999	10.20	10.20	0.06	-0.041167616	0.04	0.07	-1.322469395	1.32	0.07	2.156950534	2.16	0.39	0.148	
2998.5971	118.2508	2.56	9.046	5.0924	10.43	10.43	0.06	-0.037967616	0.04	0.06	-1.272869395	1.27	0.07	2.149450534	2.15	0.39	0.146	

Tableau. 13 : illustration des calculs effectués pour les Silts.

MODELE					
FACIES	CODE	GR (A)	RHOB (B)	THOR (C)	POTA (D)
Silt	1	103.22	2.61	7.98	3.53

Range	161.006	0.63	17.912	5.553														
DEPTH.M	GR(A)	RHOB(B)	THOR(C)	POTA (D)	(A)-(A)I	ABS[D(A)]	N(A)	(B)-(B)I	ABS[D(B)]	N(B)	(C)-(C)I	ABS[D(C)]	N(C)	(D)-(D)I	ABS[D(D)]	N(D)	error(f)1	
2997.9875	114.587	2.5575	8.6128	4.991	6.77	6.77	0.04	-0.040468	0.04	0.06	-1.706069	1.71	0.10	2.048051	2.05	0.37	0.143	
2998.1399	115.599	2.5629	8.5227	4.9707	7.78	7.78	0.05	-0.035068	0.04	0.06	-1.796169	1.80	0.10	2.027751	2.03	0.37	0.142	
2998.2923	120.478	2.5603	8.6468	4.9962	12.66	12.66	0.08	-0.037668	0.04	0.06	-1.672069	1.67	0.09	2.053251	2.05	0.37	0.150	
2998.4447	118.02	2.5568	8.9964	5.0999	10.20	10.20	0.06	-0.041168	0.04	0.07	-1.322469	1.32	0.07	2.156951	2.16	0.39	0.148	
2998.5971	118.251	2.56	9.046	5.0924	10.43	10.43	0.06	-0.037968	0.04	0.06	-1.272869	1.27	0.07	2.149451	2.15	0.39	0.146	

Tableau. 14 : illustration des calculs effectués pour les Grès.

MODELE					
FACIES	CODE	GR (A)	RHOB (B)	THOR (C)	POTA (D)
Grès	2	61.12	2.499	4.867	1.963

Range	161.006	0.63	17.912	5.553														
DEPTH.M	GR(A)	RHOB(B)	THOR(C)	POTA (D)		(A)-(A)I	ABS[D(A)]	N(A)	(B)-(B)I	ABS[D(B)]	N(B)	(C)-(C)I	ABS[D(C)]	N(C)	(D)-(D)I	ABS[D(D)]	N(D)	error(f)l
2997.9875	114.587	2.5575	8.6128	4.991		6.77	6.77	0.04	-0.040468	0.04	0.06	-1.706069	1.71	0.10	2.048051	2.05	0.37	0.143
2998.1399	115.599	2.5629	8.5227	4.9707		7.78	7.78	0.05	-0.035068	0.04	0.06	-1.796169	1.80	0.10	2.027751	2.03	0.37	0.142
2998.2923	120.478	2.5603	8.6468	4.9962		12.66	12.66	0.08	-0.037668	0.04	0.06	-1.672069	1.67	0.09	2.053251	2.05	0.37	0.150
2998.4447	118.02	2.5568	8.9964	5.0999		10.20	10.20	0.06	-0.041168	0.04	0.07	-1.322469	1.32	0.07	2.156951	2.16	0.39	0.148
2998.5971	118.251	2.56	9.046	5.0924		10.43	10.43	0.06	-0.037968	0.04	0.06	-1.272869	1.27	0.07	2.149451	2.15	0.39	0.146

Tableau. 15 : illustration des calculs effectués pour les Grès Argileux.

MODELE					
FACIES	CODE	GR (A)	RHOB (B)	THOR (C)	POTA (D)
Grès-Argileux	3	108.96	2.59	10.16	3.32

Range	161.006	0.63	17.912	5.553														
DEPTH.M	GR(A)	RHOB(B)	THOR(C)	POTA (D)		(A)-(A)I	ABS[D(A)]	N(A)	(B)-(B)I	ABS[D(B)]	N(B)	(C)-(C)I	ABS[D(C)]	N(C)	(D)-(D)I	ABS[D(D)]	N(D)	error(f)l
2997.9875	114.587	2.5575	8.6128	4.991		6.77	6.77	0.04	-0.040468	0.04	0.06	-1.706069	1.71	0.10	2.048051	2.05	0.37	0.143
2998.1399	115.599	2.5629	8.5227	4.9707		7.78	7.78	0.05	-0.035068	0.04	0.06	-1.796169	1.80	0.10	2.027751	2.03	0.37	0.142
2998.2923	120.478	2.5603	8.6468	4.9962		12.66	12.66	0.08	-0.037668	0.04	0.06	-1.672069	1.67	0.09	2.053251	2.05	0.37	0.150
2998.4447	118.02	2.5568	8.9964	5.0999		10.20	10.20	0.06	-0.041168	0.04	0.07	-1.322469	1.32	0.07	2.156951	2.16	0.39	0.148
2998.5971	118.251	2.56	9.046	5.0924		10.43	10.43	0.06	-0.037968	0.04	0.06	-1.272869	1.27	0.07	2.149451	2.15	0.39	0.146

Tableau. 16 : illustration des calculs effectués.

DEPTH.M	error(f)0	error(f)1	error(f)2	error(f)3	MIN(error)	FACIES
2997.9875	0.220	0.173	0.293	0.197	0.173	1
2998.1399	0.215	0.167	0.276	0.191	0.167	1
2998.2923	0.227	0.180	0.288	0.204	0.180	1
2998.4447	0.239	0.192	0.295	0.216	0.192	1
2998.5971	0.258	0.210	0.306	0.234	0.210	1
2998.7495	0.208	0.256	0.303	0.232	0.208	0
2998.9019	0.189	0.237	0.282	0.213	0.189	0

(X)-(X) : la différence entre le log et la moyenne du log dans le facies

ABS [D(X)] : la valeur absolue de la différence (A)-(A)

N(X) : ABS [D(A)] divisé par le Range.

Error (f) i : la moyenne de N(X) dans le facies (i)

MIN (error) : le minimum des erreurs Error(f) i

FACIES est défini par la formule : Si (MIN (error)=error(f)0,0, Si

(MIN(error)=error(f)1,1 , Si (MIN (error)=error(f)2,2, Si

(MIN (error)=error(f)3,3,"ERROR"))

Les résultats ont été discutés dans le chapitre suivant. (Résultats et discussion).

Conclusion

Ce chapitre a permis de décrire l'ensemble des données, matériel et logiciels utilisées pour l'élaboration de ce mémoire. En effet, la description macroscopique passe nécessairement par l'opération de préparation des carottes. La diagraphie a pour but de reconnaître et confirmation des caractéristiques des réservoirs. L'étape de la validation des données de la prédiction dépend de ces logs diagraphique et la description détaillée, c'est pour cela, il a été nécessaire d'effectuer des tests visant à déterminer le choix des logs à utiliser.

Le fichier final est fait l'objet des deux méthodes utilisées (les réseaux de neurones Artificiels et Cluster Analysis) pour reconstituer les facies non carottés et non récupérés.

Chapitre IV

Résultats et Discussion

Ce chapitre présente les résultats obtenus par les deux méthodes de modélisation et une comparaison qualitative et quantitative a été faite pour recommander la méthode la plus proche de la carotte, les réseaux de neurones artificiels s'avèrent très performantes devant Cluster Analysis, on peut la généraliser sur d'autres caractéristiques géologiques.

1. RECONSTITUTION DES FACIES PAR LA METHODE DES RESEAUX DES NEURONES ARTIFICIELS

L'application de la méthode de RNA a permis de reconstituer les logs lithologiques des puits carottés et de prédire les niveaux non carottés

1.1. Comparaison qualitative

La visualisation des résultats de la prédiction a été faite à l'aide de l'établissement des logs des facies de prédiction, en utilisant la commande Log discret du logiciel PETREL (2009). Ceci, a permet de faire une comparaison entre le log prédiction et celui des carottes. On a aussi visualisé les logs de prédiction des puits non-carottés et confirmer les résultats avec les diagraphies.

1.1.1. Les puits carottés

- Puits P-A

On constate dans ce puits, que on a déterminé le faciès à l'aide de RNA est globalement, le même que le faciès décrits en carotte. Il a été relevé également, les mêmes limites entres les bancs avec une bonne correspondance entre les pics de Gamma Ray. Cependant, la prédiction a permis de mettre en évidence un niveau de silt, or, la description macroscopique a été rapportée en tant que grès. Ce niveau est bien détecté avec la réponse de Gamma ray (Fig. 42).

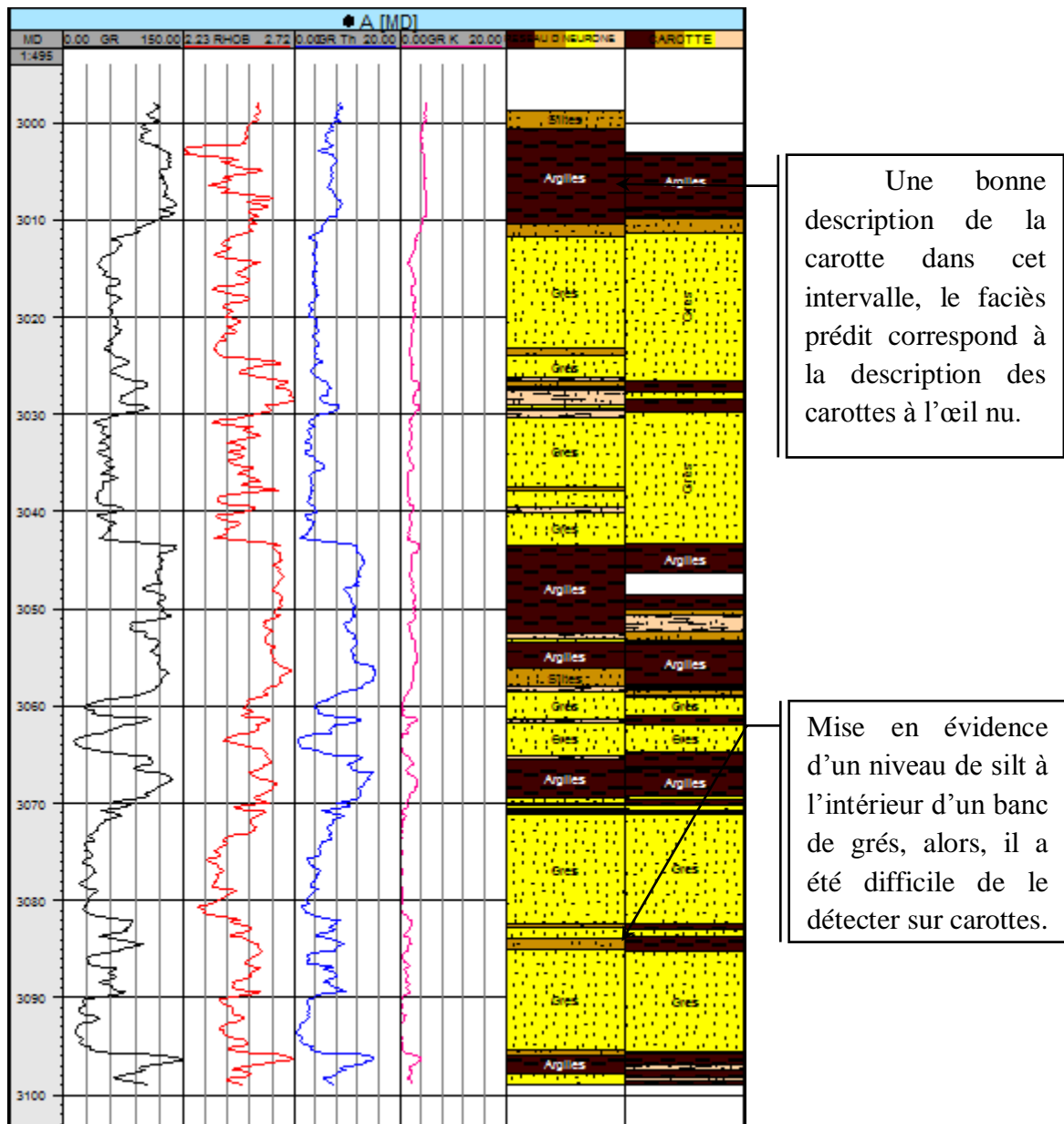


Fig. 42 - Résultat des RNA pour le puits P_A.

- Puits P-B

Dans le puits B, on constate qu'il y a une bonne correspondance entre les faciès prédit par RNA d'une part et la description des carottes et les pics diagraphiques d'autre part. L'intervalle non carottée entre 3073m et 3110m a été reconstitué par les réseaux des neurones. les résultats sont en parfaite concordance avec les logs diagraphiques (Fig. 43).

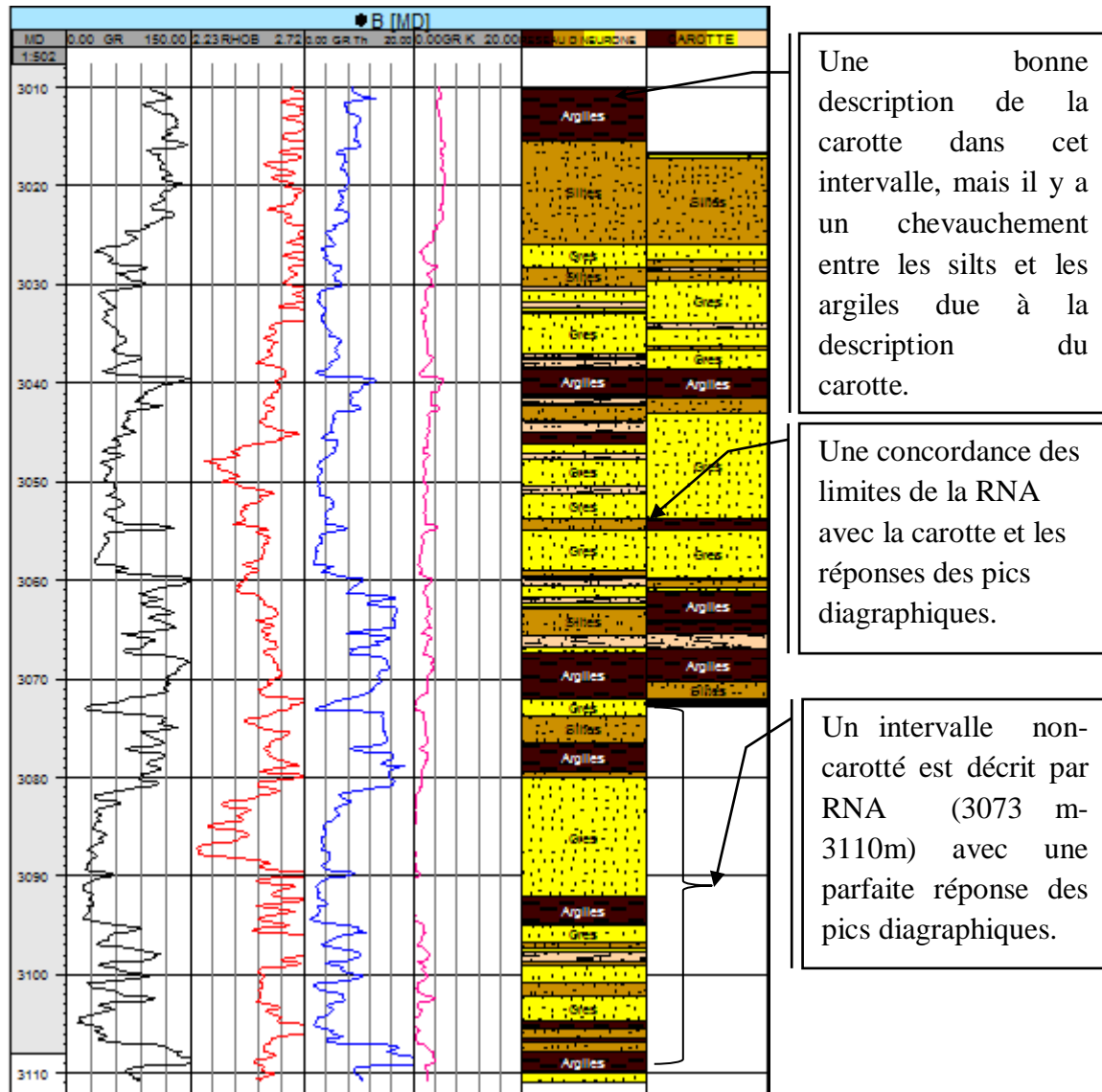


Fig. 43 - Résultat des RNA pour le puits P_B.

- Puits P-C

Ce puits présente une absence des carottes à l'intervalle (3046m – 3110m). Cette lacune a été reconnue par le log RNA. Ce dernier se présente en parfaite harmonie avec les pics dans ce niveau. Une concordance entre les limites de RNA et les carottes a été remarquée (Fig. 44)

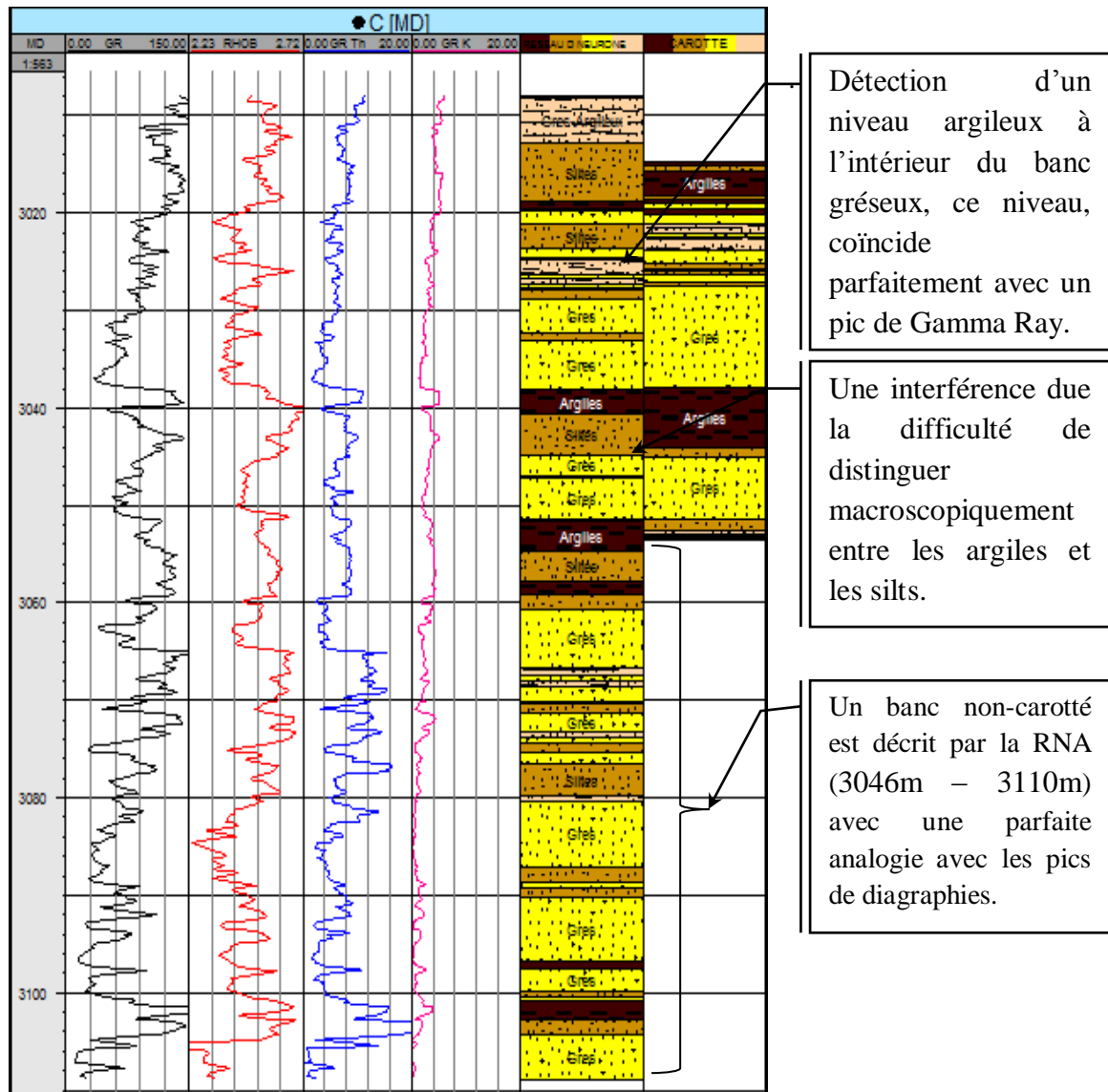


Fig. 44 - Résultat des RNA pour le puits P_C.

1.1.2. Les puits non carottés

- Puits P-D

Ce puits qui n'a pas été carotté représente un véritable test pour le modèle des RNA. Le log lithologique a été entièrement reconstitué. Les résultats obtenus montrent que les faciès prédits et les pics des diagraphies présentent une grande ressemblance (Fig. 45).

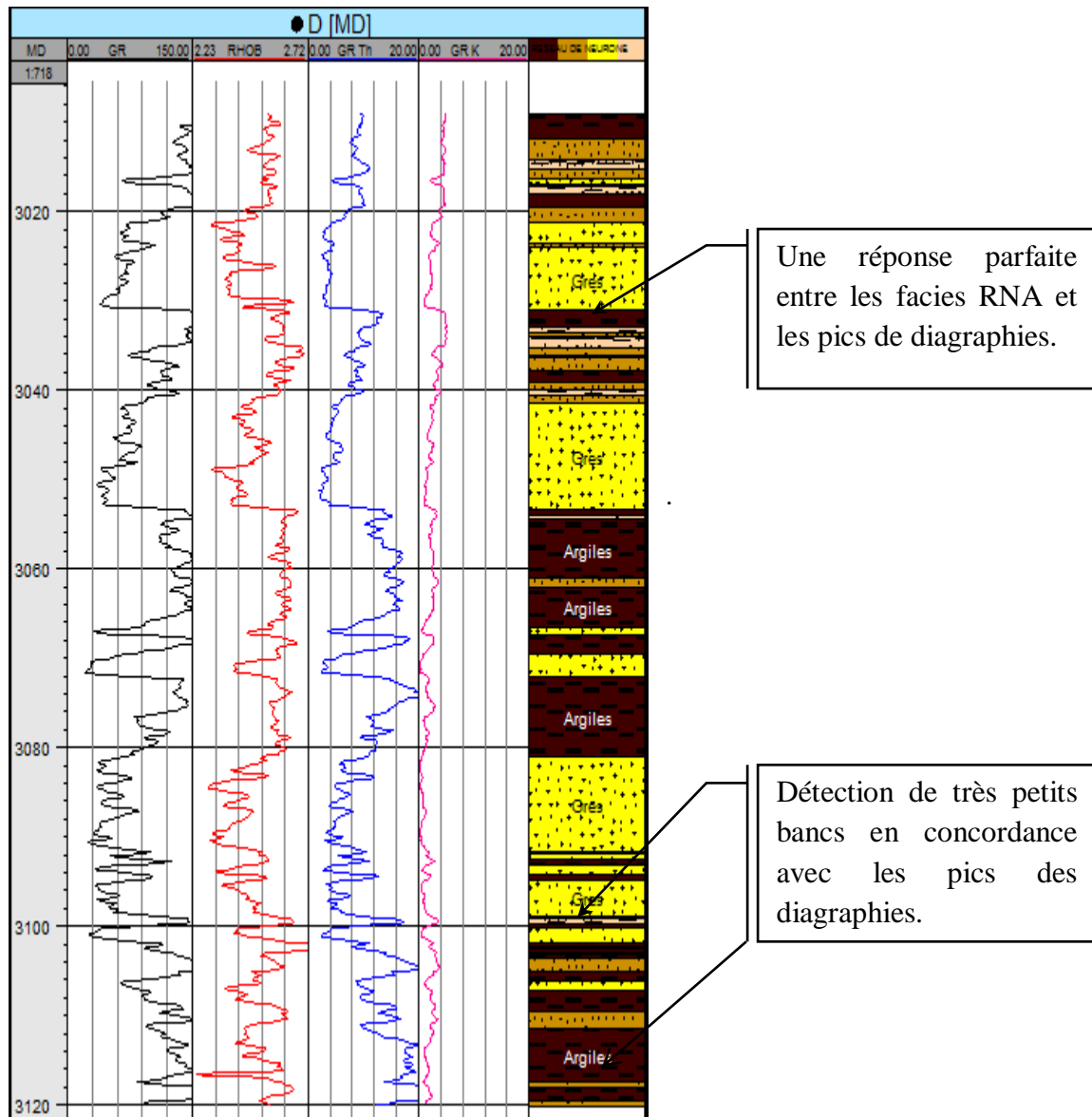


Fig. 45 - Résultat des RNA pour le puits P_D.

- Puits P-E

A l'aide des facies RNA, ce puits a été reconnu, car il n'a pas fait l'objet de l'opération de carottage. Une parfaite concordance entre les faciès RNA avec les réponses de pics des logs diagraphiques (Fig. 46).

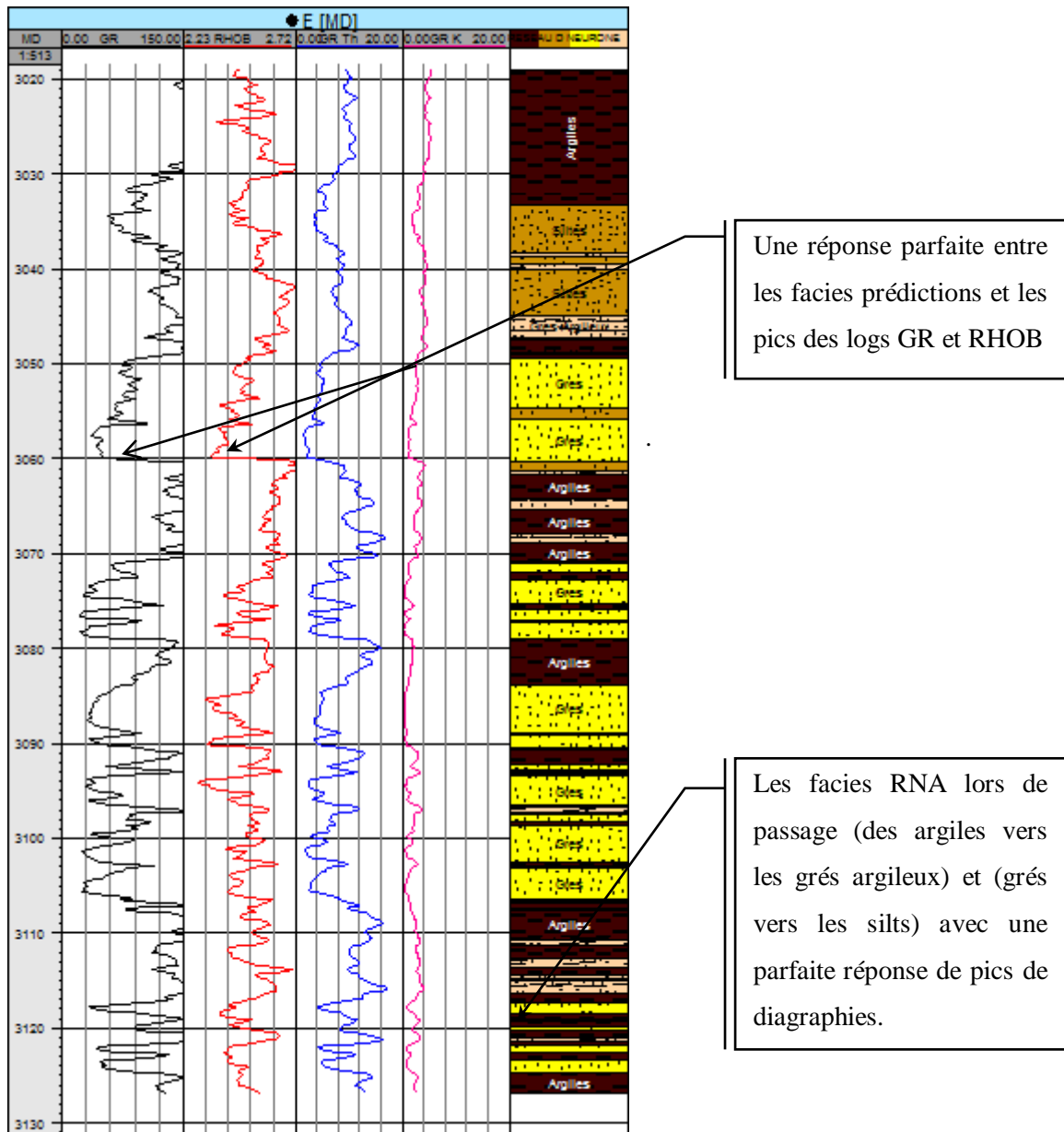


Fig. 46 - Résultat des RNA pour le puits P_E.

1.2. Comparaison Quantitative

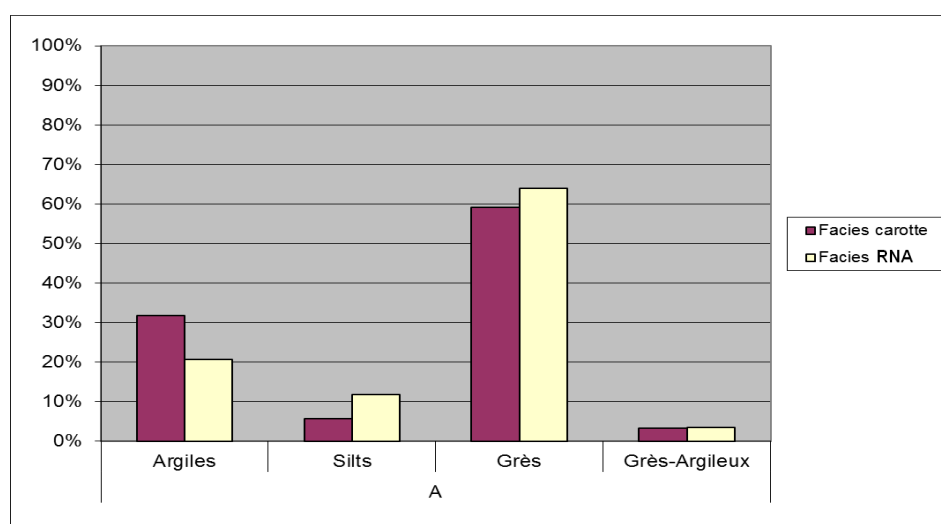
Pour chaque puits, on a calculé la différence entre le pourcentage des faciès déterminés par les réseaux de neurones (faciès RNA) celui déterminé par la description des carottes (faciès carotte)

1.2.1. Puits P-A

La prédiction a donné des résultats très proches de la description des carottes pour ce puits. Elle a augmenté le pourcentage des silts et elle a diminué le pourcentage des argiles. Mais pour le grès il y a une différence de 5%. Pour les grès argileux la différence est négligeable. (Tableau. 17).

Tableau. 17 : Comparaison entre RNA et les carottes du puits P-A.

		MODELE			
P-A	Facies	Code	Facies Carottes	Facies (RNA)	Différence
	Argiles	0	32%	21%	-11%
	Silts	1	6%	12%	6%
	Grès	2	59%	64%	5%
	Grès-Argileux	3	3%	4%	1%

**Fig. 47-** Histogramme de comparaison facies carottes et facies RNA puits P-A.

1.2.2. Puits P-B

Il y a une augmentation du pourcentage de grès-argileux et une diminution de pourcentage des argiles dans le facies RNA. Pour les silts la différence est négligeable.

Tableau. 18 : Comparaison entre RNA et les carottes pour le puits P-B.

		MODELE			
P-B	Facies	Code	Facies Carottes	Facies (RNA)	Différence
	Argiles	0	21%	16%	-6%
	Silts	1	27%	27%	1%
	Grès	2	48%	49%	2%
	Grès-Argileux	3	5%	8%	3%

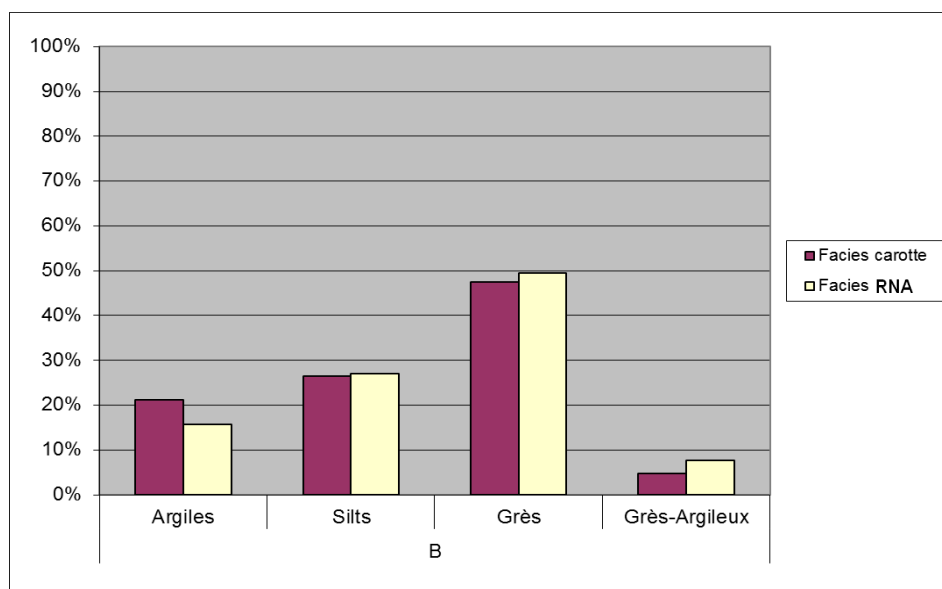


Fig. 48 - Histogramme de comparaison facies carottes et facies RNA puits P-B.

1.2.3. Puits P-C

Dans ce puits la différence entre la prédiction et la carotte est importante, concernant surtout les argiles et les silts, que le logiciel a généré un pourcentage de silts qui n'a pas été décrit sur la carotte et il a diminué en parallèle le pourcentage des argiles. On a aussi une petite différence entre les grès (10%) et le grès argileux (4%).

Tableau. 19 : Comparaison entre RNA et les carottes pour le puits P-C.

		MODELE			
P-C	Facies	Code	Facies Carottes	Facies (RNA)	Différence
	Argiles	0	27%	17%	10%
	Silts	1	11%	22%	11%
	Grès	2	55%	50%	-5%
	Grès-Argileux	3	7%	11%	4%

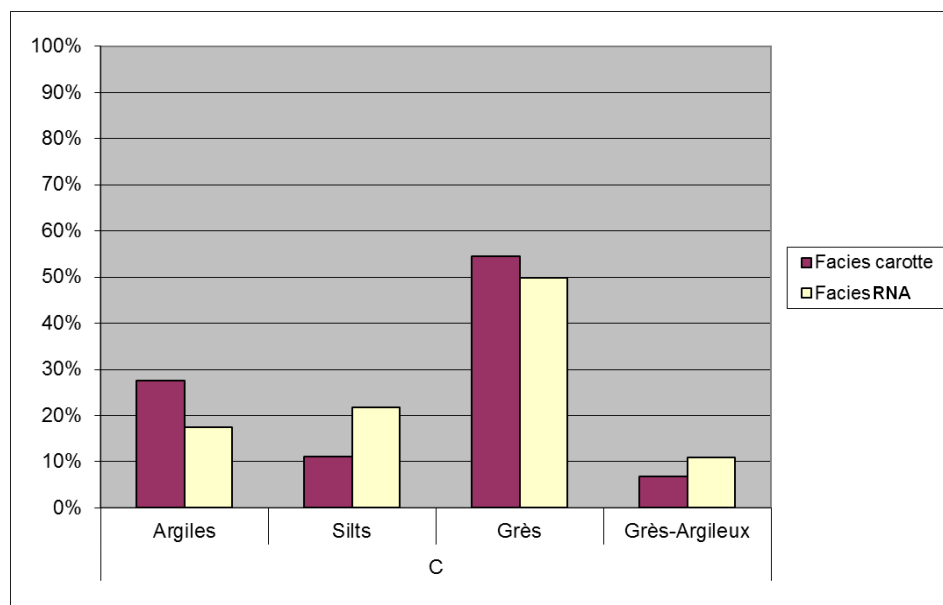


Fig. 49 - Histogramme de comparaison facies carottes et facies RNA puits P-C.

1.3. Synthèse des résultats pour les puits A, B et C

En totalité, pour les grès et les argiles, la différence entre la prédiction et la carotte est négligeable. Pour les silts et les grès-argileux la différence est de 10% et 2% respectivement.

Les erreurs obtenus lors de la prédiction sont due au plusieurs facteurs :

- Les descriptions des carottes été faites par plusieurs auteurs incluant la possibilité de l'existence d'erreurs dans les descriptions.
- Les descriptions ont été simplifiées en quatre faciès principaux;
- La prédiction est basée sur un traitement statistique des données sur Excel ;
- Utilisation d'une combinaison de logs (modèles) dans la prédiction (GR, RHOB, THOR, POTA).

Il y a une confusion entre les silts et les argiles liée à plusieurs facteurs : la possibilité de l'existence d'erreurs dans les descriptions ; les teneurs des éléments radioactives dans les formations (U, Th, K) et le pas d'échantillonnage.

Tableau. 20: Comparaison de nombre d'échantillons entre RNA et les carottes.

	Facies	CODE	Facies carotte	Facies RNA
			Nbr Des échantillons	Nbr d'échantillons
TOTAL	Argiles	0	344	271
	Silts	1	159	245
	Grès	2	678	634
	Grès-Argileux	3	55	86
			1236	1236

Tableau. 21 : Comparaison totale entre RNA et les carottes.

	Facies	CODE	Facies Carotte	Facies RNA	Différence
TOTAL	Argiles	0	28%	22%	6%
	Silts	1	13%	18%	5%
	Grès	2	55%	53%	2%
	Grès-Argileux	3	4%	7%	3%

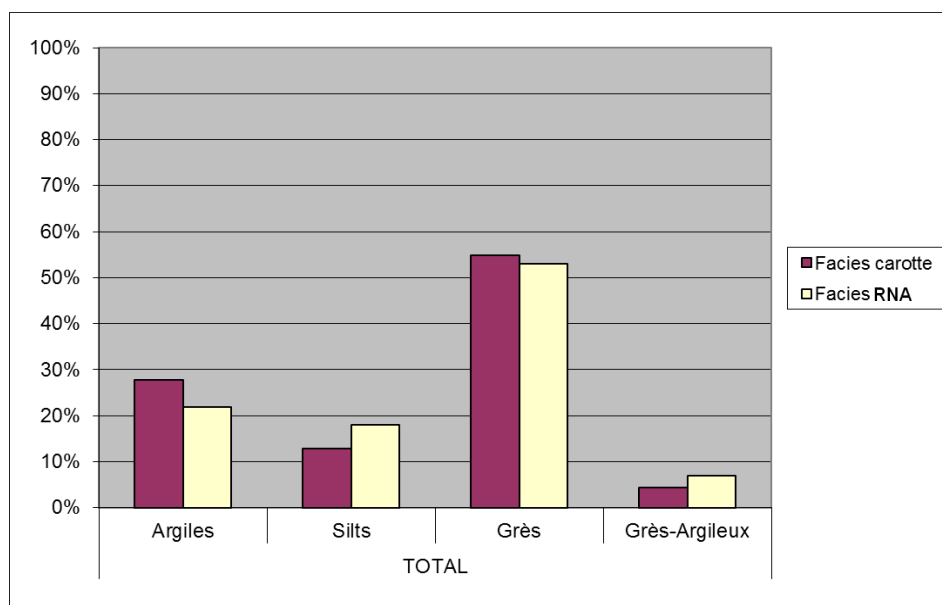


Fig. 50 - Histogramme de comparaison totale entre facies RNA et les carottes.

2. RECONSTITUTION DES FACIES PAR LA METHODE DE CLUSTER ANALYSIS

2.1. Comparaison qualitative

2.1.1. Les puits carottés

- Puits P-A

On constate que la détermination du faciès à l'aide de Cluster Analysis est globalement la même que le faciès décrit sur carotte. Seulement qu'il y a un chevauchement entre les argiles et les silts (Fig. 51).

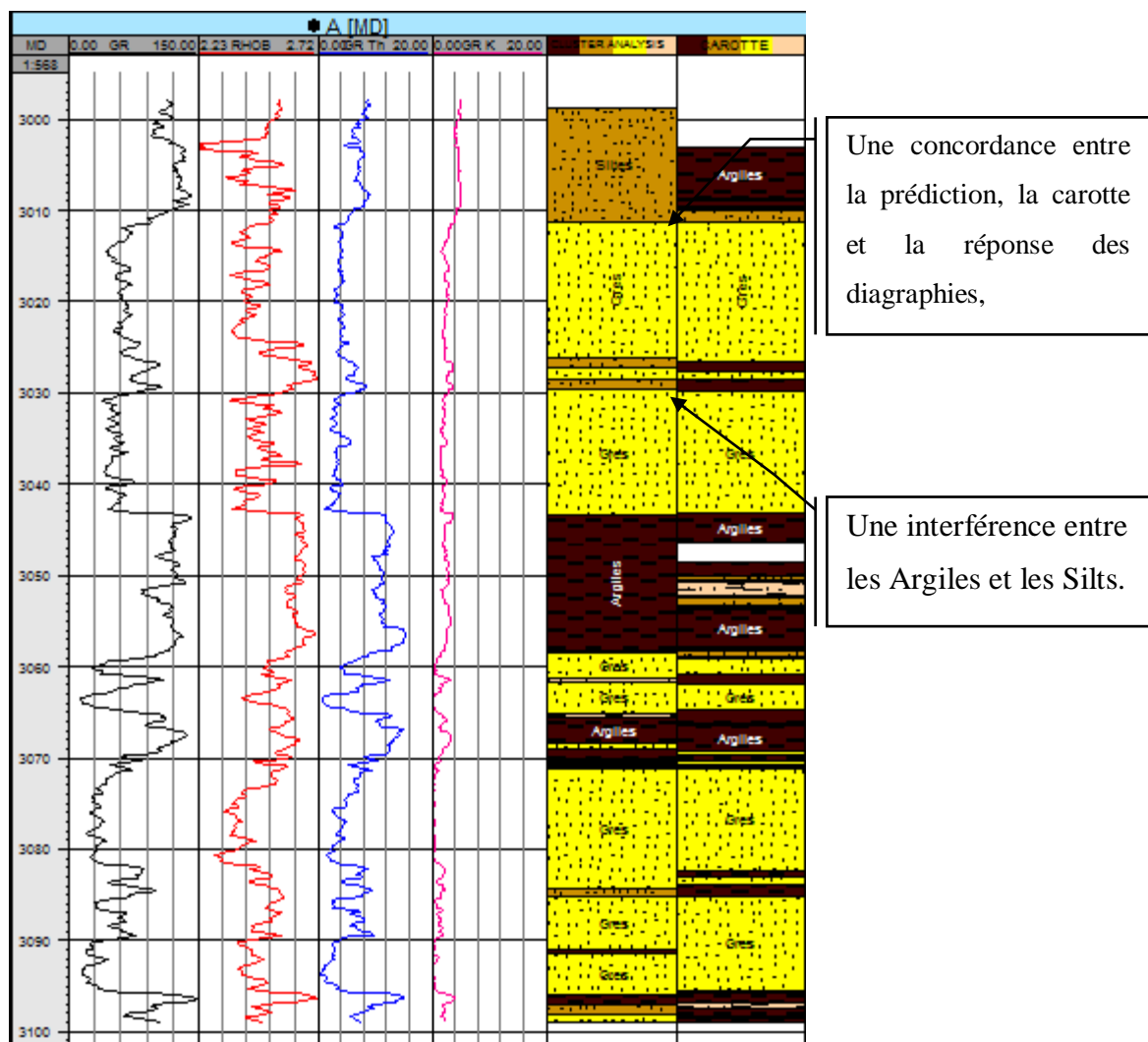


Fig. 51 - Résultat de Cluster Analysis pour le puits P_A.

- Puits P-B

On constate que les faciès de la prédiction sont pratiquement les mêmes que ceux des carottes. Les limites des bancs et la lithologie coïncide avec les pics des diagraphies. La prédiction a permis de distinguer entre silts et les argiles, or, il a été difficile de les distinguer macroscopiquement. A partir de la profondeur 3080m, un intervalle non carotté a été décrit à l'aide de Cluster Analysis (Fig. 52).

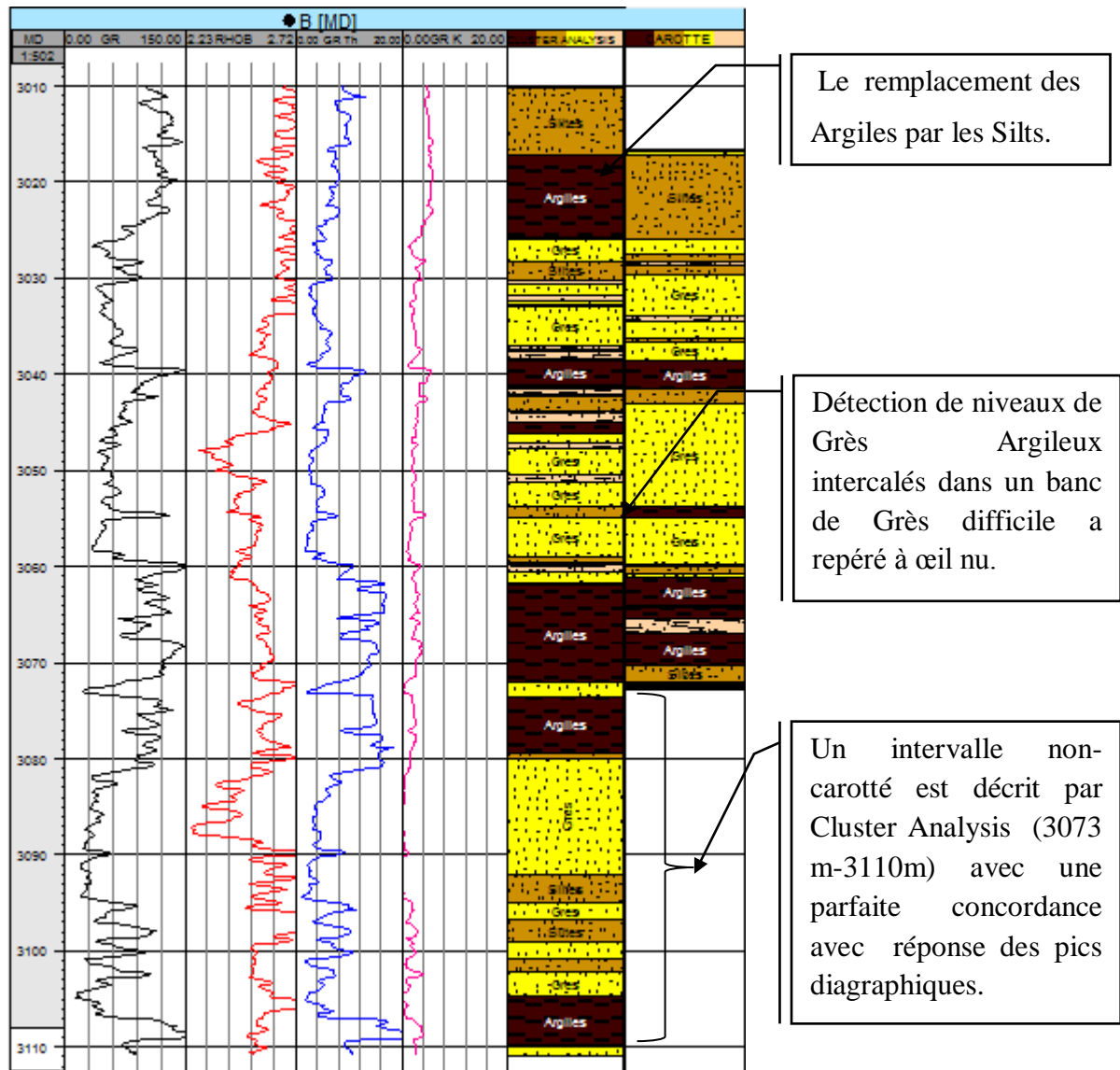


Fig. 52 - Résultat de Cluster Analysis pour le puits P_B.

- Puits P-C

Il a été constaté une absence des carottes à l'intervalle (3060m – 3100m). Cette lacune a été reconnue par le log prédiction. Ce dernier présente une parfaite réponse des pics dans ce niveau. Une concordance des limites de la prédiction et des carottes a été remarquée (Fig. 53).

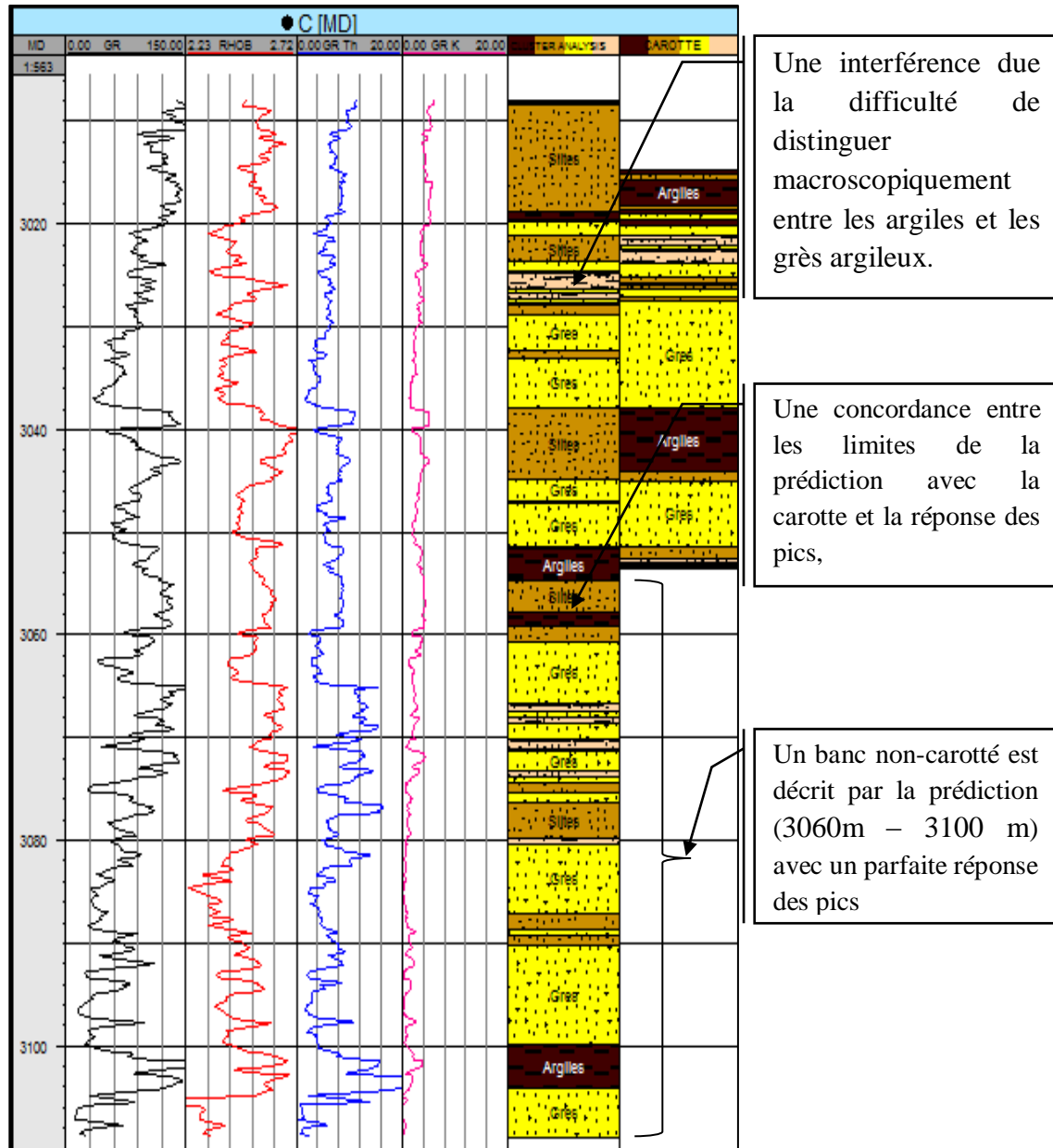


Fig. 53 - Résultat de Cluster Analysis pour le puits P_C.

2.1.2. Les puits non carottés

- Puits P-D

La reconstitution du log de ce puits non carotté par la méthode Cluster Analysis a donné un bon résultat diagraphiques (surtout gamma ray et densité). Il a été détecté de très petits bancs et concordant avec les pics des logs (Fig. 54).

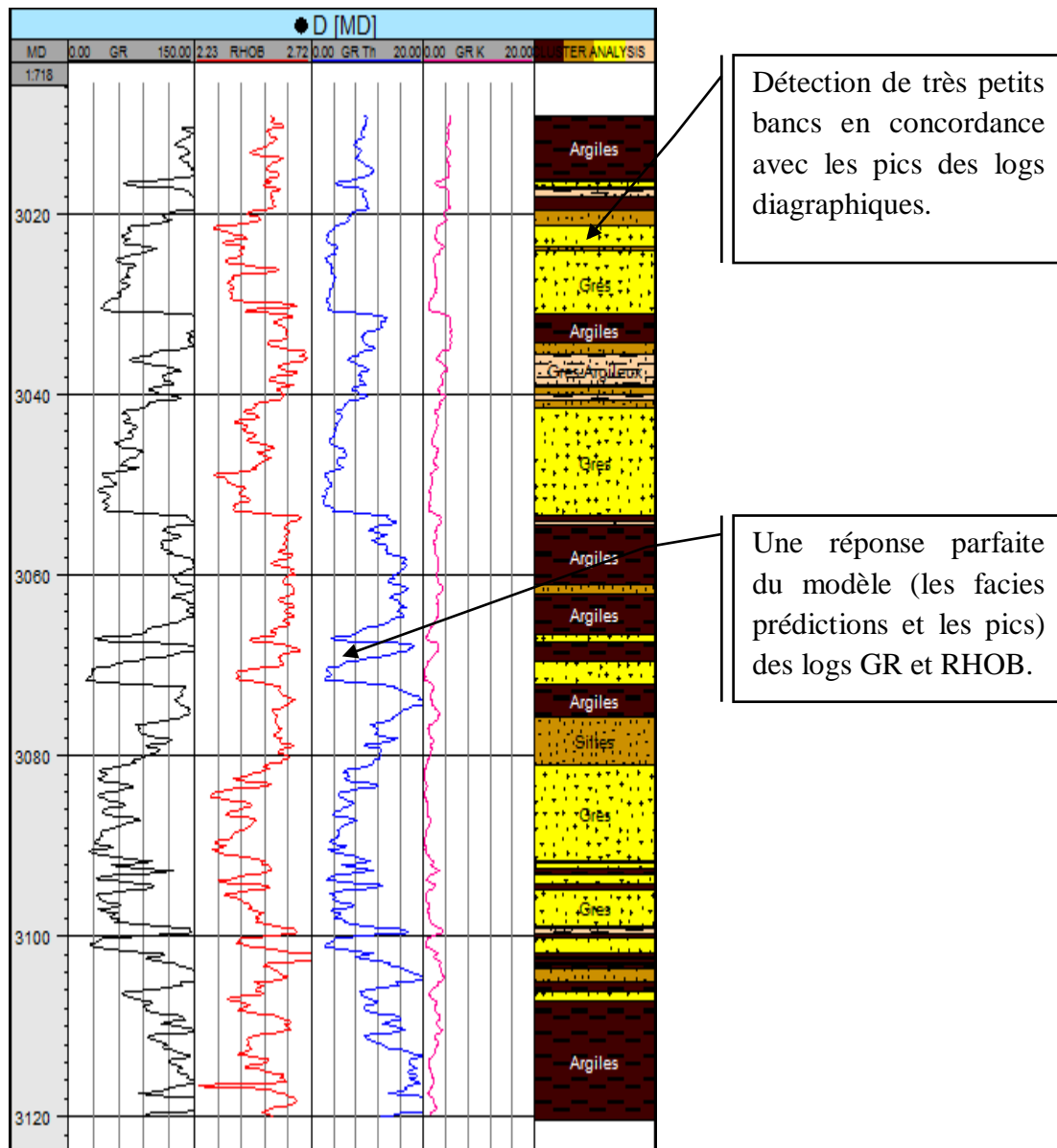


Fig. 54 - Résultat de Cluster Analysis pour le puits P_D.

- Puits P-E

A l'aide des faciès prédictions, ce puits a été entièrement reconstruit, car il n'a pas été l'objet de l'opération de carottage. Une parfaite concordance entre les faciès prédictions avec les de pics des logs (Fig. 55).

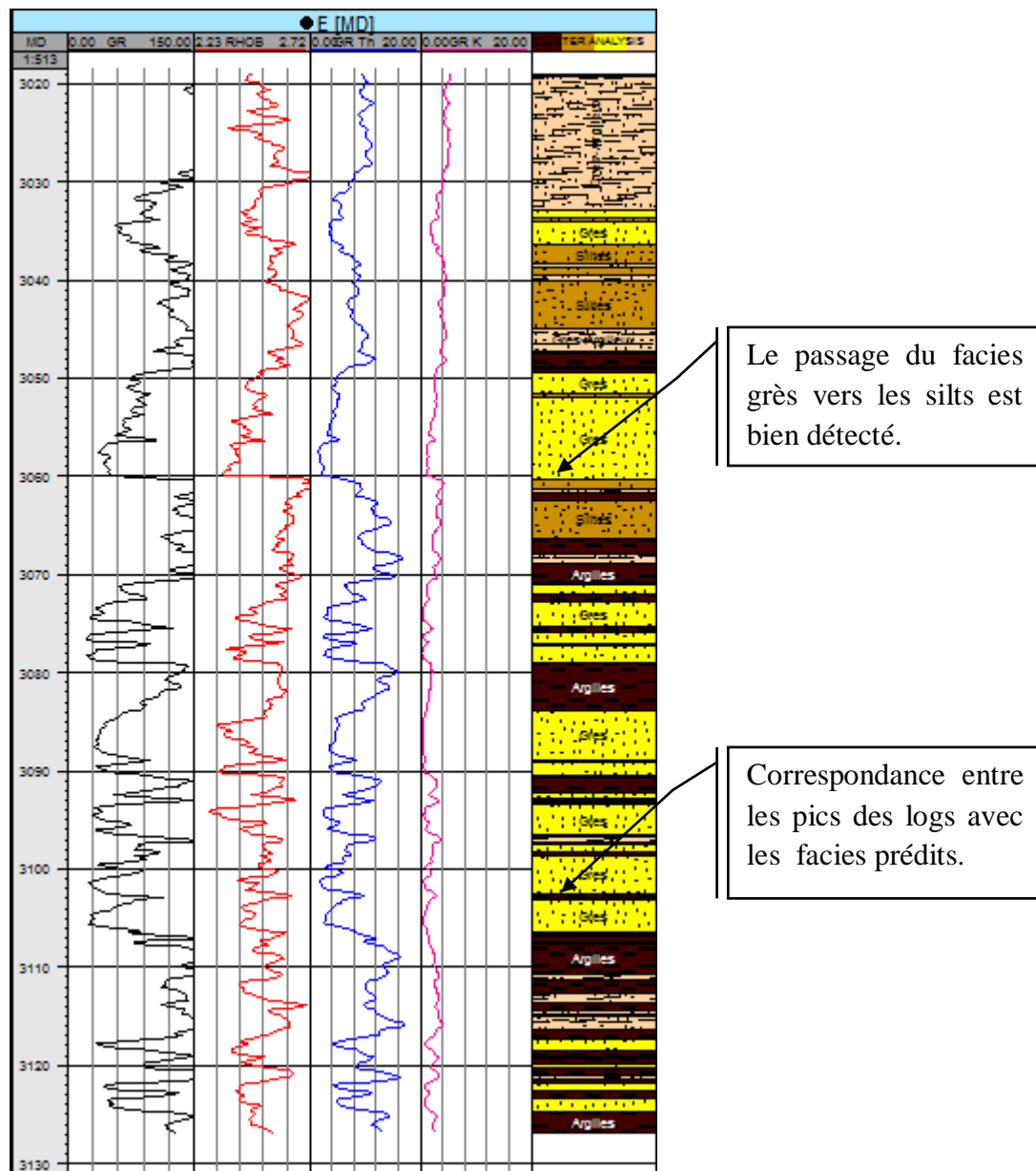


Fig. 55 - Résultat de Cluster Analysis pour le puits P_E.

2.2. Comparaison quantitative

Pour chaque puits, on a calculé la différence entre le pourcentage des faciès prédits par la méthode Cluster Analysis (faciès CA) et celui décrit sur carotte (faciès carotte)

La somme des pourcentages des quatre faciès (Argiles, Silts, Grès et Grès-Argileux) pour faciès carotte ou bien pour faciès CA égale à 100%. Les résultats ont été exprimés sous forme de tableau suivi d'un histogramme.

2.2.1. Puits P-A

La méthode Cluster Analysis a donné des résultats relativement différents de la description des carottes. La prédiction a augmenté le pourcentage des silts et a diminué le pourcentage des argiles ; pour le grès il y a une différence de 7% ; Pour les grès argileux la différence est négligeable (Fig. 56).

Tableau. 22 : Comparaison entre Cluster Analysis et les carottes du puits P-A.

		MODELE			
P-A	Faciès	Code	Faciès Carottes	Faciès (CA)	Différence
	Argiles	0	32%	23%	8%
	Silts	1	6%	22%	16%
	Grès	2	59%	52%	7%
	Grès-Argileux	3	3%	3%	0%

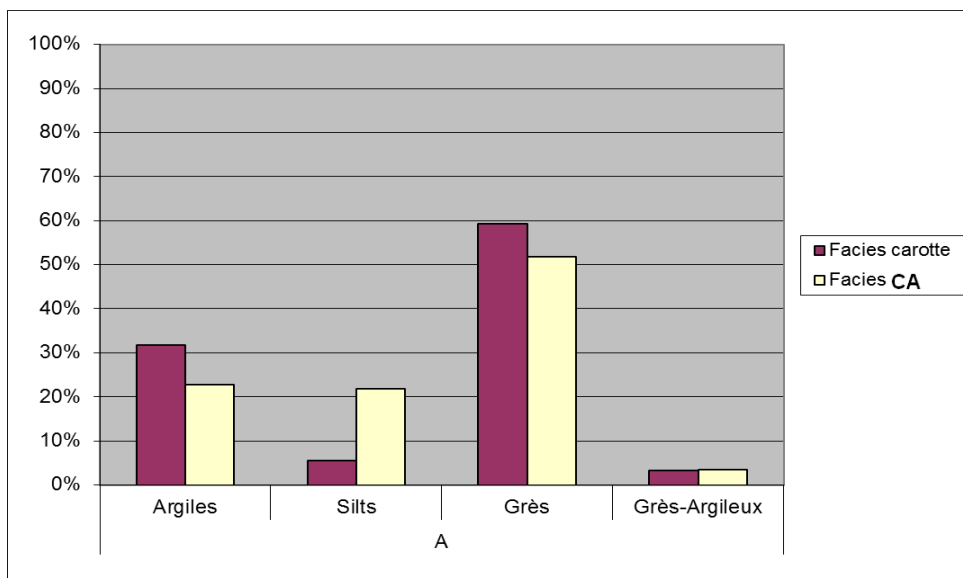


Fig. 56 - Histogramme de comparaison faciès carottes et faciès CA puits P-A.

2.2.2. Puits P-B

Il y a une augmentation du pourcentage de grès-argileux et les argiles et une diminution de pourcentage des grès dans le faciès CA. Pour les silts la différence est de 7%. (Fig. 57).

Tableau. 23 : Comparaison entre Cluster Analysis et les carottes pour le puits P-B.

		MODELE			
P-B	Faciès	Code	Faciès Carottes	Faciès (CA)	Différence
	Argiles	0	21%	30%	9%
	Silts	1	27%	20%	7%
	Grès	2	47%	40%	7%
	Grès-Argileux	3	5%	10%	5%

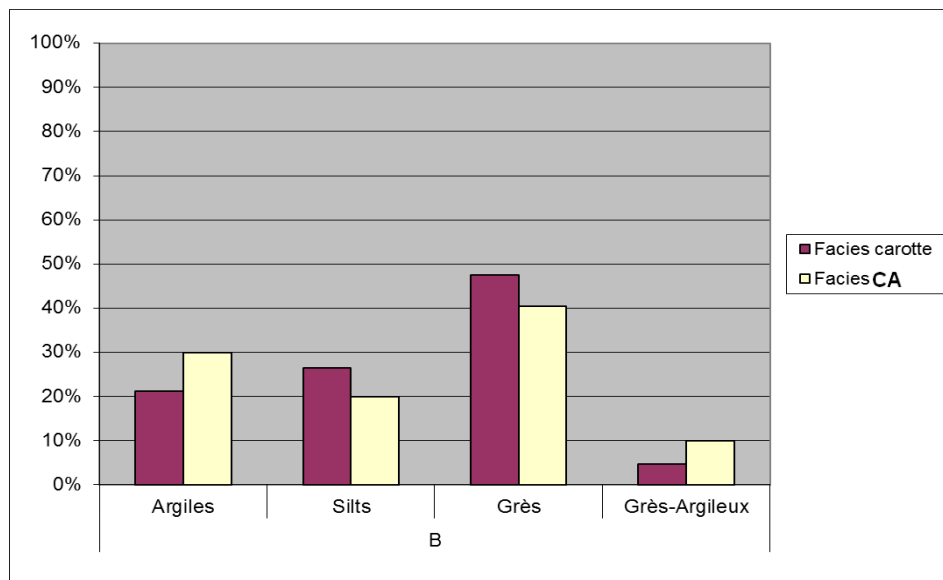


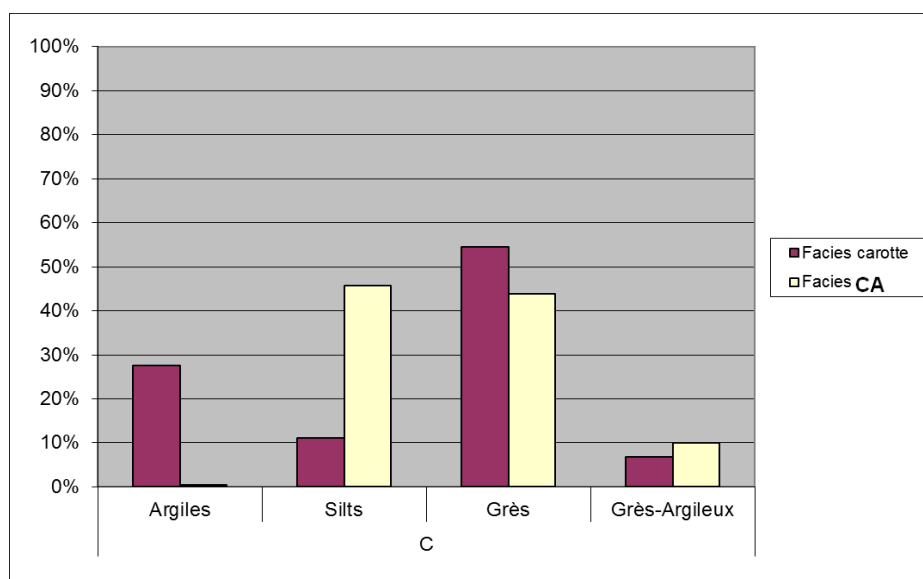
Fig. 57 - Histogramme de comparaison faciès carottes et faciès CA puits P-B.

2.2.3. Puits P-C

Dans ce puits la différence entre Cluster Analysis et la carotte est très grande, concernant surtout les argiles 27% et les silts 35%, que le logiciel a généré un pourcentage de silts qui n'a pas été décrit sur la carotte et il a diminué en parallèle le pourcentage des argiles. Aussi on remarque une différence entre les grès (11%) et le grès argileux (3%), (Fig. 58).

Tableau. 24 : Comparaison entre Cluster Analysis et les carottes pour le puits P-C.

		MODELE			
P-C	Facies	Code	Facies Carottes	Facies (CA)	Différence
	Argiles	0	27%	0%	-27%
	Silts	1	11%	46%	35%
	Grès	2	55%	44%	-11%
	Grès-Argileux	3	7%	10%	3%

**Fig. 58 -** Histogramme de comparaison facies carottes et facies CA puits P-C.

2.3. Synthèse des résultats pour les puits A, B, et C

En totalité, pour les grès et les argiles, la différence entre la prédiction et la carotte est négligeable. Pour les argiles et les silts la différence est de 13% et 10% respectivement (Tableau. 26).

Tableau. 25 : Comparaison de nombre d'échantillons entre Cluster Analysis et les carottes.

	Facies	CODE	Facies carotte	Facies CA
			Nbr Des échantillons	Nbr d'échantillons
TOTAL	Argiles	0	344	22
	Silts	1	159	1
	Grès	2	678	29
	Grès-Argileux	3	55	8
			1236	60
			1236	1236

Tableau. 26 : Comparaison totale entre Cluster Analysis et les carottes.

	Facies	CODE	Facies Carotte	Facies CA	Différence
TOTAL	Argiles	0	28%	15%	-13%
	Silts	1	13%	23%	10%
	Grès	2	55%	56%	1%
	Grès-Argileux	3	4%	6%	2%

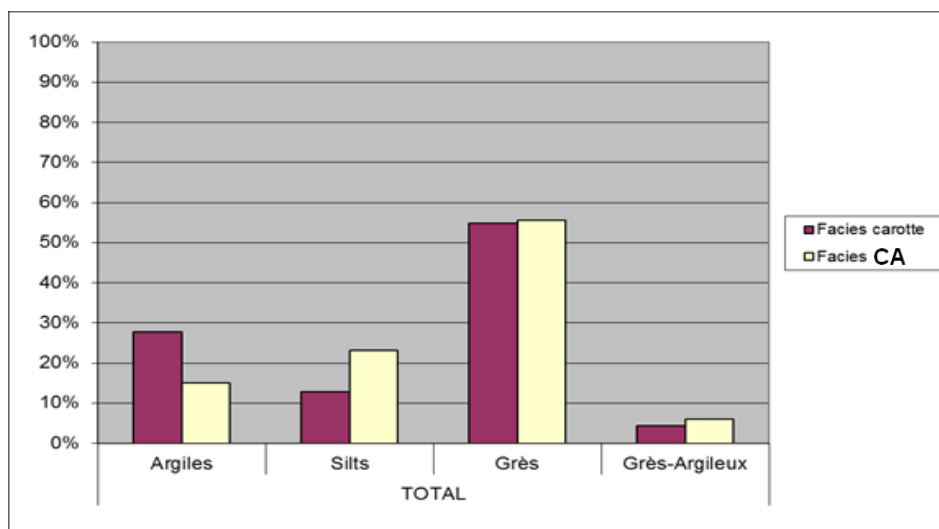


Fig. 59 - Histogramme de comparaison totale entre Cluster Analysis et les carottes.

3. COMPARAISON ENTRE LES RESULTATS DES RESEAUX DE NEURONE ET CLUSTER ANALYSIS

Après la visualisation des logs il apparaît clairement que les réseaux de neurones donnent des meilleurs résultats de reconstitution des faciès non carottés devant Cluster Analysis. Les histogrammes montrent la comparaison entre les deux méthodes et la figure ci-dessous qui montre la comparaison entre le faciès carotte, cluster Analysis et réseau de neurone.

Tableau. 27: Comparaison entre RNA, Cluster Analysis et les carottes Puits-A.

P-A	Faciès	Code	Faciès Carottes	Faciès (RNA)	Faciès (CA)
	Argiles	0	32%	21%	23%
	Silts	1	6%	12%	22%
	Grès	2	59%	64%	52%
	Grès-Argileux	3	3%	4%	3%

Tableau. 28: Comparaison entre RNA, Cluster Analysis et les carottes Puits-B.

P-B	Faciès	Code	Faciès Carottes	Faciès (RNA)	Faciès (CA)
	Argiles	0	21%	16%	30%
	Silts	1	27%	27%	20%
	Grès	2	48%	49%	40%
	Grès-Argileux	3	5%	8%	10%

Tableau. 29: Comparaison entre RNA, Cluster Analysis et les carottes Puits-C.

P-C	Faciès	Code	Faciès Carottes	Faciès (RNA)	Faciès (CA)
	Argiles	0	27%	20%	0%
	Silts	1	11%	26%	46%
	Grès	2	55%	42%	44%
	Grès-Argileux	3	7%	12%	10%

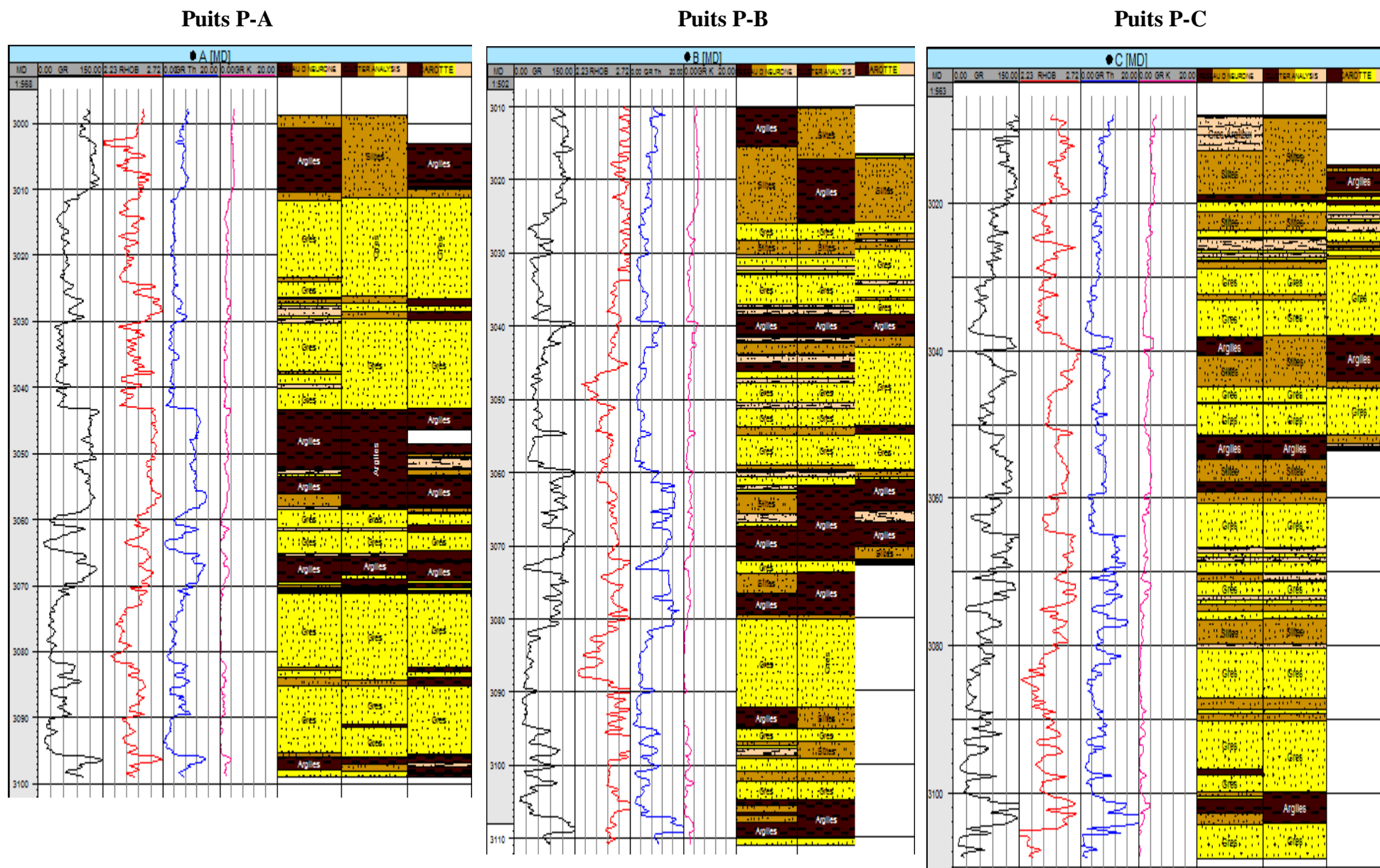


Fig. 60 - Comparaison des résultats de RNA et Cluster Analysis des puits (A, B, C) avec les carottes.

CONCLUSION GENERALE

CONCLUSION GENERALE

La région SIF Fatima (Bloc 402b) est située dans l'erg oriental du Sahara Algérienne (Bassin de Berkine). Ce bloc est caractérisé par une tectonique cassante relativement intense. Cette tectonique se manifeste par la présence d'une faille principale d'Ourhoud-Sif Fatima orientée NNE/SSW et des failles secondaires ayant la même direction que la principale. Le réservoir essentiel du secteur est le Trias Argileux Gréseux Inferieur (TAGI). Les roches réservoirs constituent les séries basales du Mésozoïque. Elles sont marquées par les dépôts de nature fluviatiles qui se développent le long du bassin. Il a deux principales séquences :

- Le TAGI inférieur: L'extension et l'épaisseur sont contrôlés par la paléogéographie hercynienne.
- Le TAGI supérieurs et moyen où les formations des chenaux et les méandres prédominent. Les séquences terminales du TAGI supérieur dans la partie centrale et vers le Nord-Est du bassin de Berkine sont marquées par le début d'une transgression marine à la base des calcaires Triasiques.

Le réservoir (TAGI) représente l'objet de la présente étude, de modélisation et reconstitution des facies non carottés. En effet, deux méthodes de l'intelligence artificielle (le réseau de neurone artificiel et cluster Analysis) ont été appliquées à l'aide des données des diagraphies en particulier (GR, RHOB, POTA et THOR) et la description détaillée des carottes. Pour cela un codage des carottes a été effectué chaque 15 cm (argile = 0 ; silt = 1 ; grés = 2 et grès argileux = 3). Une base de données a été établie comportant 189,2m d'intervalles carottés et 332,8m d'intervalles non carottés en plus des données des diagraphies. Cette base a fait l'objet de modélisation par les deux méthodes afin de déterminer les électro- facies.

L'application de ces méthodes, nous a permis de définir 332.84 m de faciès lithologiques dans les puits non carottés ou partiellement carottés,. Ces méthodes ont été d'avantage utilisé pour corriger les interférences liées à la description macroscopique des carottes. En effet, les géologues trouvent toujours, des difficultés à distinguer entre les facies qui se ressemblent en aspect extérieur (argile-silt et grès-grès argileux). Pour cela, les puits carottés ont été également fait l'objet d'une prédiction. Ceci a permis de mettre

en évidence les niveaux de faibles épaisseurs qui présentent de bonnes réponses avec les paramètres diagraphiques.

La première méthode est basée sur la théorie de Réseau de Neurone Artificiel, qui a été inspirée de la biologie, à un modèle de perception multicouches avec une algorithme de RETROPROPAGATION à apprentissage supervisé, qui se compose d'une couche d'entrée de 4 neurones qui sont les variables explicatifs (GR, RHOB, THO, POT), et d'une couche de sortie composée d'un seule neurone qui est le variable a expliqué (le Facies prédit), et une seule couche cachée. La fonction de transfert de la couche d'entrée vers la couche cachée est sigmoïdale (TANSIG). cependant la fonction qui relie la couche cachée à la couche de sortie est de type linéaire (PURLIN). l'algorithme d'activation utilisé est de LEVENBERG-MARQUARDT. Le modèle a une tendance d'amélioration et d'adaptation interactivement à chaque fois on change le nombre de neurones dans la couche cachée, jusqu'au ce que les indices de performances numériques (CC, RMSE, MAE, d) et graphique (la fonction coût) seront excellents à toutes les phases. Le modèle retenu composé de 4 neurones dans la couche cachée a donné un coefficient de corrélation CC de 0,87 pour l'apprentissage, 0,78 pour la validation et 0,92 pour le test.

La deuxième méthode est une technique statistique dite Cluster Analysis qui débrouille le même problème de la non récupération des carottes, mais elle repose sur une technique de calcul qui consiste à diviser un ensemble de données en différents paquets homogènes, dans le sens que les données de chaque sous-ensemble partagent des caractéristiques communes, et faire la déviation du log par facies, Puis on a calculé l'erreur pour chaque facies, et attribué le facies ayant la valeur minimale des erreurs.

Les résultats obtenus ont été visualisés en logs, et une comparaison quantitative et qualitative a été effectuée. Elle a permis de constater globalement, une correspondance entre les facies des carottes avec celles de la prédiction, ceci avec une concordance avec les pics diagraphiques.

La comparaison quantitative et qualitative des résultats de prédiction a été effectuée. Elle a permis de constater une correspondance plus ou moins significative entre les facies des carottes et celles de la prédiction, en accord avec les réponses des pics diagraphiques (Gamma ray en particulier).

Ces méthodes présentent les avantages suivants :

- Le gain économique car ils sont moins coûteuses par rapport l'opération de carottage.
- La simulation des faciès présent l'intérêt quant au gain du temps.

Ces méthodes sont comme toutes méthodes de simulation présentent des inconvénients parmi lesquels :

- La mauvaise description des carottes va influencer le codage des électrofaciès et provoquer des erreurs au stade des résultats de prédiction.
- La description et le codage est limité en quatre (4) faciès, or, l'existence de plusieurs faciès rend la prédiction difficile par exemple (le chlorite est décrite en argile).
- Les meilleurs résultats ont été acquis suite à l'utilisation de la diagraphie (GR, RHOB, THOR, POTA). L'absence d'un log va entrainer des erreurs à la prédiction malgré la possibilité de faire des substitutions avec autres logs.

La comparaison entre les deux méthodes montre que la première s'avère plus performante que la deuxième, et on peut la généraliser sur d'autres caractéristiques géologiques pour le développement du réservoir.

PERSPECTIVES

Ce travail nous a permis de constater que les méthodes d'intelligence artificielle sont très utiles dans le domaine pétrolier et peuvent valoriser d'avantage les données disponibles.

Cette constatation nous incite à employer d'autres méthodes tel que les méthodes de classification (carte auto-organisatrice), la logique floue et d'autres méthodes hybrides, afin d'une meilleure reconnaissance des potentialités en hydrocarbures de notre sous-sol.

BIBLIOGRAPHIE

BIBLIOGRAPHIE

Albert, M. (1992). Les réseaux de neurones comme outils d'aide à la décision floue Rapport de D.E.A, Ingénierie Informatique, I.N.S.A de Lyon, 64 p.

Al-Bulushi, N., Peter R, K., Martin, J., Kraaijveld, M. (2009). Development of artificial neural network models for predicting water saturation and fluid distribution. Journal of Petroleum Science and Engineering 68 (2009) 197–208.

Alizadeh, B., Najjari, S., Kadkhodaie-Ilkhchi, A. (2012). Artificial neural network modeling and cluster analysis for organic facies and burial history estimation using well log data: A case study of the South Pars Gas Field, Persian Gulf, Iran. Computers & Geosciences 45 (2012) 261–269.

Ames, F., Mello and Martin, Buzas, A. (1968). An application of Cluster Analysis as a method of determining biofacies Journal of Paleontology, V. 42, NO. 3, P. 747-758, 6 TEXT-FIGS., May 1968.

Ameur Zaimeche, O., Zeddouri, A., Kouadria, T., Kechiched, R., Belaksier, M S. (2014). Use of Cluster Analysis method in log's data processing: prediction and rebuilding of lithologic facies. International Conference on Environmental Science and Geoscience (ESG '14) Venice, Italy, March 15-17, 2014 98 – 101. <http://www.europment.org/library/2014/venice/bypaper/ENVIR/ENVIR-16.pdf>

Ameur Zaimeche, O. (2012). Prédiction des facies non carottés par la méthode de Cluster Analysis et contribution à l'étude du réservoir TAGI du Bassin de Berkine SW. Mémoire d'ingénieur, université Badji Mokhtar, Annaba. 76 pages.

Anatoly, V., Chashkova, Valery M. Kiselevb. (2011). Use of the Cluster Analysis and Artificial Neural Network Technology for Log Data Interpretation. Journal of Siberian Federal University. Engineering & Technologies 4 (2011 4) 453-462.

Anderberg, M.R. (1973). Cluster analysis for applications. Academic Press, NewYork.

Askri, H., et al. (1995) - Petroleum Geology of Algeria; Well Evaluation Conference- WEC Algeria. 93 pages.

Azoug, Y., et al. (2007) - Well Evaluation Conference- WEC Algeria 2007 - Sonatrach-Schlumberger. Edited by Schlumberger; Produced by Lynx Consulting, Inc., Houston, TX, USA. 489 pages.

Baillargeon, J. (2004). Application et interprétation des techniques statistiques avancées SRP-6018 http://www.uqtr.ca/cours/srp-6018/c_srp6018_2004.htm

Briqueu, L., Gottlib-Zeh, S., Ramadan, M., Brulhet, J. (2002) - Traitement des diagraphies à l'aide d'un réseau de neurones du type « carte auto-organisatrice » :

application à l'étude lithologique de la couche silteuse de Marcoule (Gard, France). C. R. Geoscience 334 (2002) 331–337.

Carney, S., Peffer, J. W., Pink, A., Drumheller, R. E., Jasek, J. H., Hadjali, M., Bouazza, L. (1999) – The Hassi Berkine South (HBNS) Field, Block 404, Berkine basin, Algeria, American Association of petroleum geologists, Annual Meeting, Exposed abstracts, San Antonio, Texas, United states, p.A21.

Corsini, M.M. (2005): Introduction aux réseaux de neurones Université Victor Segalen France.

Cortés, J.A. et al. (2007). Deciphering magma mixing: The application of cluster analysis to the mineral chemistry of crystal populations. Journal of Volcanology and Geothermal Research 165 (2007) 163–188.

Coulibaly, P., Francois, A., Bernard, B. Al (1999) : Prévion Hydrologique par Réseaux de Neurones Artificiels –état de l'art .Can.J.Civ.Eng. 26 PP 293-304.

D.E. Irawan et al. (2009). Hydrogeochemistry of volcanic hydrogeology based on cluster analysis of Mount Ciremai, West Java, Indonesia. Journal of Hydrology 376 (2009) 221–234.

Dreyfus, G. (2004) : Réseaux de neurones ; Méthodologie et Application EYROLLES, 2004 pp 01-65.

Echikh, K. (1998) – Geology and hydrocarbon occurrences in Ghadames basin, Algeria, Tunisia, Libya. In D. S. MacGregor, R. T. J. Moody & D. D. Clark-Lowes (Eds.), Petroleum geology of North Africa. Geological Society of London special publication, 132. 109- 130.

El-Sebakhy, E.A. Asparouhov, O., Abdulraheemb, A.A., Al-Majed, A.A., Donghui, W., Latinski, K., Raharja, I. (2012). Functional networks as a new data mining predictive paradigm to predict permeability in a carbonate reservoir. Expert Systems with Applications 39 (2012) 10359–10375.

Fiesler, E. (1994). Neural network classification and formalization. FTP Available at ftp.cis.ohio-state.edu in pub/neuroprose.

Frayssinet, D., Thiria, S., Badran, F., Briqueu, L. (2000) -Use of neural networks in log's data processing: prediction and rebuilding of lithologic facies. Petrophysics meets Geophysics - Paris, France, 6 - 8 November 2000.

Galeazzi, S., Point, O., Haddadi, N., Mather, J., Druesne, D. (2010). Regional geology and petroleum systems of the Illizi–Berkine area of the Algerian Saharan Platform: An overview. Marine and Petroleum Geology 27 (2010) 143–178.

Gallegos, M.T., Ritter, G. (2005). A robust method for cluster analysis. *Annals of Statistics* 33, 347-380.

Grondin, F. (2004)- le carottage :division forage département de formation .40 p.

Guochang, W., Timothy R, C. (2012). Methodology of organic-rich shale lithofacies identification and prediction: A case study from Marcellus Shale in the Appalachian basin. *Computers & Geosciences* 49 (2012) 151–163.

Guochang, W., Guojian, C., Timothy, R., (2013) .The application of improved NeuroEvolution of Augmenting Topologies neural network in Marcellus Shale lithofacies prediction. *Computers & Geosciences* 54 (2013) 50–65.

Hertz,J. , Krogh, A. (1991).Introduction to the theory of neural computation. Addison-Wesley, Reading, Massachusetts.

Hinton, G.E. (1989). Deterministic Boltzmann machine learning performs steepest descent in weight-space. *Neural Computation*, 1 :143–150.

Hinton, G.E et Sejnowski, T.J. (1986). Learning and relearning in Boltzman machines .In *Parallel Distributed Processing* 1,chap .7.

Hopfield, J.J.(1982).Neural Network and physical systems with emergent Collective abilities. *Proceedings of the National Academy Of Science , USA* 81 , PP.3088-3092.

Hornik, K. (1989). Stinchcombe, M., & White, H. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2 (5),pp. 359-366, 1989.

Hsien-Cheng, C., David C, K., Hui-Chuan, C. (2002). Identification of lithofacies using Kohonen self-organizing maps. *Computers & Geosciences* 28 (2002) 223–229.

Kohonen, T. (1984). Self organization and associative memory, 2nd ed, Springer Verlag, 1984.

I, Ladlani., L, Houichi., L, Djemili., S, Heddami.,K, Belouz. Modeling (2012). daily reference evapotranspiration (ET₀) in the north of Algeria using generalized regression neural networks (GRNN) and radial basis function neural networks (RBFNN): a comparative study. *Meteorol Atmos Phys* (2012) 118:163–178.

Irani, R., Nasimi, R. (2011). Evolving neural network using real coded genetic algorithm for permeability estimation of the reservoir. *Expert Systems with Applications* 38 (2011) 9862–9866.

Le Cun , Y. (1985). Une procédure d'apprentissage pour réseau à seuil assymétrique. In *Cognitiva 85: A la frontière de l'intelligence Artificielle des Sciences de la Connaissance des Neurosciences*, CESTA, Paris, pp. 599-604.

- Lianshuang, Qi., Timothy, R. (2006).** Neural network prediction of carbonate lithofacies from well logs, Big Bow and Sand Arroyo Creek fields, Southwest Kansas. *Computers & Geosciences* 32 (2006) 947–964.
- Lippmann, R.P. (1987).** An introduction to computing with neural nets. *IEEE Acoustics, Speech and Signal Processing Magazine* , 4 : 4-22.
- Mallat, S.(1989).** A theory for multiresolution signal decomposition the wavelet transform, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 674-693.
- Manly, B.F.J. (2005),** *Multivariate Statistical Methods: A primer*, Third edition, Chapman and Hall.
- Marc , P.(2004).** Réseaux de Neurones Université de l'AVAL.
- Matthieu , G .(2003).** Entraînement de Réseaux de Neurones Récurent à Pulses Appliqués à la Modélisation d'un Tissu Neuronal Biologique .Thèse Maîtrise en Science Appliqué Génie Electrique ECOLE POLYTECHNIQUE DE MONTREAL 200 pp 14-20.
- Merabet, K., Moumen, MA., Sekkache, A.(2009).**Modélisation de l'évapotranspiration par la technique des réseaux de neurones, mémoire d'ingénieur, Université 20 Aout Skikda, 92p.
- McCarthy, J. (1959).** Programs with Common Sense. In *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, 756-91. London: Her Majesty's Stationery Office.
- McCulloch, W. S et Pitts ,W. (1943).** A logical calculus of the ideas immanent in Nervous activity . *Bull Math .Biophysics* , 5;113 –115.
- McKenna, S., Hedley, R. (2002).** The structural evolution of the Berkine-Ghadames Basin(Groupement Berkine Anadarko). Rapport interne – Anadarko Petroleum Corporation.
- Minsky, M.L , (1967).** *Computation : Finite and Infinite Machines* .Englewood Cliffs : Prentice –Hall .
- Minsky, M.L et Papert , S. A (1969).** *Perceptrons*. CAMBRIDGE : MIT press.
- M.J.J. Couapel et al. (2007).** Late Quaternary marginal marine palaeoenvironments of northern Australia as inferred from cluster analysis of coccolith assemblages. *Marine Micropaleontology* 65 (2007) 213–231.
- Morgan, N., Boulard, H., (1990).**Generalization and parameter estimation in feedforward nets: some experments.pages 630-637, 1990.

MOURA, C. A. V. (1985). Aplicação de Tratamento Estatístico Multivariante em Dados Geoquímicos de Solo no Mapeamento Geológico na Província de Carajás (Alvo 2 – Corpo 4). *Revista Brasileira de Geociências*, v. 15, n. 3, p. 241-248, 1985.

Muayyad, H., Khaiwka, M.I., EL-sayed, A., Alia, A. (1981). The utility of Cluster Analysis in determining sedimentary facies. *Sedimentary Geology*, 30 (1981) 245—253

Newell, A., & Simon, H. A. (1972). Human problem solving (Vol. 104, No. 9). Englewood Cliffs, NJ: Prentice-Hall.

OJA, E. (1982). A simplified neuron model as a principal component analyser. *Journal of Mathematical Biology*, 15 : 267–273.

Parker ,D.B.(1985). Learning logic. Technical report. Center of Computational Research in Economics and Management Science. MIT,Cambridge, MA, 1985.

Peretto ,P. (1984). Collective properties of neural networks : a statical physics Approach. *Biological Cybernetics* 50 , PP.51-62.

Powell, M. J.D. (1987). Radial basis functions for multivariable interpolation : a review, *Algorithms for approximation*, p. 143-167.

Romain , B . (2003). Modèles Impulsionnels de Réseaux de Neurones Biologiques Thèse Doctorat , Paris 2003 pp 03-10.

Rosenblatt, F. (1962). Principales of Neurodynamics NEW YORK : Spartan.

Rossi, C., Kalin, O., Arribas, J., Tortosa, A. (2002) – Diagenesis, provenance and reservoir quality of Triassic TAG sandstones from Ourhoud field, Berkine (Ghadames) Basin, Algeria . *Marine and Petroleum Geology* 19 (2002) 117-145.

Rumelhurt, D.E. (1986). Learning internal representations by error propagation. *Parallel Distributed Processing Explorations in the Microstructure of Cognition*. Eds. D. E. Rumelhart and J.L. McClelland, Cambridge, MA, MIT Press, Bradford Books, vol. 1, pp. 318-362,1986.

Samy Zazoun, R. (2013). Fracture density estimation from core and conventional well logs data using artificial neural networks: The Cambro-Ordovician reservoir of Mesdar oil field, Algeria. *Journal of African Earth Sciences* 83 (2013) 55–73.

Sarzeaud , O.(1995). Les réseaux de neurones : contribution à une théorie. *Ouest Éditions*, 1995. ISBN 2-808261-94-4.

Scott, A.j., Wheller, D.A. (1999) – Regional trends in Triassia fluvial-lacustrine cycle, Berkine Basin, Algeria, American association of petroleum of petroleum Geologists Annual Meeting Expanded Abstracts, San Antonio, Texas, United States, p. 1126.

Serra, O. (1985) -Diagraphies Différés base de l'interprétation ,Mémoire7 Tome2. Etudes et productions Schlumberger, Montrouge.

Serra,O. (1979) - Diagraphies Différés base de l'interprétation, Mémoire1 Tome1. Services techniques Schlumberger, Paris.

Sonatrach. (2002) – Bilan des activités et perspectives – Rapport interne - Sonatrach– DirectionGénérale - Chemin Djenane El Malik, Hydra, Alger, Algérie.

Sonatrach, Anadarko. (2012). Rapport interne.

Tarek, A. (2004). Introductions au systèmes Adaptatifs à base d'apprentissage Laboratoire A2SI-ESIEE-Paris.

Tenalem, A., Shimeles, F., Frank, W., Molla, D., Stefan, W. (2009). Hierarchical cluster analysis of hydrochemical data as a tool for assessing the evolution and dynamics of groundwater across the Ethiopian rift. International Journal of Physical Sciences Vol. 4 (2), pp. 076-090, February, 2009.

T. Euzen. (2005). well log cluster analysis and electrofacies classification: a probabilistic approach for integrating log with mineralogical data. GeoConvention, IFP Technologies (Canada) Inc.

Touzet, (1992). Les Réseaux de Neurones Artificiels – Introduction au Connexionisme. COURS- EXERCICES- et TRAVAUX PRATIQUES.

Turner, P., Pilling, D., Walker, D., Binnie, J., Sabaou, N., (2001). Sequence stratigraphy and sedimentology of the late Triassic TAGI (Blocks 401 – 402. Berkine, Algeria), Marine Petroleum Geology 18, 959 – 981.

Von Neumann, J .(1956). Probabilistics Logics and the synthesis of reliable Organisms from unreliable components. In Automata Studies, Eds.C.E.Shannon et J.McCarthy Princeton University Press, PP 43-98.

WEC (2007). Well Evaluation Conference.

Widrow, B. (1960). In IRE Western Electric Show and Convention Record, pp. 96-104, 1960.

Willmott CJ (1981). On the validation of model. Phys Geogr 2:184–194.

Willmott CJ (1982). Some comments on the evaluation of model performance. Bull Am Meteorol Soc 63:1309–1313.

Yuste, R et Tank, D. W. (1996). Dendritic integration in mammalian neurons, a century after Cajal, Neuron, 16, pp. 701–16.

Zerrouki, A.A., Aïfa, T., Baddari, K. (2014).Prediction of natural fracture porosity from well log data by means of fuzzy ranking and an artificial neural network in Hassi Messaoud oil field, Algeria. J. Petrol. Sci. Eng. (2014).

ANNEXES

ANNEXE N° : 1

CODE SOURCE DU PROGRAMME REALISE

```
%Listing du programme
close all; %Fermeture de tous les fenêtres windows
clear all; %mise à jour de tous les variables

[LES_DONNEES_Berkine1,TEXT]=xlsread('OUAFI1_LOG.xls','TRAINING','H3
:L248');
[LES_DONNEES_Berkine2,TEXT]=xlsread('OUAFI1_LOG.xls','VALIDATION','H2
:L84');
[LES_DONNEES_Berkine3,TEXT]=xlsread('OUAFI1_LOG.xls','TESTING','H2
:L84');
[LES_DONNEES_Berkine4,TEXT]=xlsread('OUAFI1_LOG.xls','ALLDATA','H3
:L412');
[LES_DONNEES_Berkine5,TEXT]=xlsread('OUAFI1_LOG.xls','ALLDATA',
'H413:K598');

A1= LES_DONNEES_Berkine1;
A2= LES_DONNEES_Berkine2;
A3= LES_DONNEES_Berkine3;
A4= LES_DONNEES_Berkine4;
A5= LES_DONNEES_Berkine5;

[L1,C1]= size (A1)
[L2,C2]= size (A2)
[L3,C3]= size (A3)
[L4,C4]= size (A4)
[L5,C5]= size (A5)

%C : LES COLONNE %L : LES LIGNES

P1(:,1:5)=A1(:,1:5);% Partie apprentissage
P2(:,1:5)=A2(:,1:5);% Partie validation
P3(:,1:5)=A3(:,1:5);% Partie TEST
P4(:,1:5)=A4(:,1:5);% Partie ALLDATA
P5(:,1:4)=A5(:,1:4);% Partie MISSINGDATA

T1=P1(:,5)' ; % la partie output du réseau de neurone
P1=P1(:,[1 2 3 4])' ; % la partie input du réseau de neurone

T2=P2(:,5)' ; % la partie output du réseau de neurone
P2=P2(:,[1 2 3 4])' ; % la partie input du réseau de neurone

T3=P3(:,5)' ; % la partie output du réseau de neurone
P3=P3(:,[1 2 3 4])' ; % la partie input du réseau de neurone

T4=P4(:,5)' ; % la partie output du réseau de neurone
P4=P4(:,[1 2 3 4])' ; % la partie input du réseau de neurone

P5=P5(:,[1 2 3 4])' ; % la partie input du réseau de neurone
trainV.P=P1;
trainV.T=T1;
```

```

valV.P =P2;
valV.T =T2;
testV.P =P3;
testV.T =T3;

NH=13; % Nombre de neurones dans l'unique couche cachée
NS=1 ; % nombre de neurones dans la couche de sortie

pr = minmax(trainV.P);
net=newff(pr, [NH NS], {'logsig', 'purelin'}, 'trainlm');
net=init(net);

[net, tr] = train(net,trainV.P,trainV.T, [], [], valV,testV);

%CALCUL DES SORTIES POUR LA PARTIE APPRENTISSAGE

an      = sim(net,P1);
Y1      = an;
Y11     = T1;

%CALCUL DES SORTIES POUR LA PARTIE VALIDATION

vn      = sim(net,P2);
Y2      = vn;
Y22     = T2;

%CALCUL DES SORTIES POUR LA PARTIE TEST

tn      = sim(net,P3);
Y3      = tn;
Y33     = T3;

%CALCUL DES SORTIES POUR LA PARTIE MISSING DATA

Mn      = sim(net,P5);
Y4      = Mn ;

[Y1' Y11'];
[Y2' Y22'];
[Y3' Y33'];
[Y4'   ];

XLSWRITE('OUAFI1_LOG.xls', [Y11',Y1']      , 'MLPNN' , 'A3'); %TRAINING
XLSWRITE('OUAFI1_LOG.xls', [Y22',Y2']      , 'MLPNN' , 'D3'); %VALIDATION
XLSWRITE('OUAFI1_LOG.xls', [Y33',Y3']      , 'MLPNN' , 'G3'); %TESTING
XLSWRITE('OUAFI1_LOG.xls', [Y4'   ]       , 'MLPNN' , 'M3'); %MISSINGDATA

XLSWRITE('OUAFI1_LOG.xls', [Y11',Y1']      , 'MLPNN' , 'J3' ); %ALL DATA
XLSWRITE('OUAFI1_LOG.xls', [Y22',Y2']      , 'MLPNN' , 'J331'); %ALL DATA
XLSWRITE('OUAFI1_LOG.xls', [Y33',Y3']      , 'MLPNN' , 'J413'); %ALL DATA

%MATRICE DES POIDS DE LA COUCHE D ENTRREE VERS LA COUCHE CACHEE

fprintf('\n matrice des poids couche entree vers la couche cachée: \n');
X1=net.IW{1,1}

```

```

%[X1'];
%MATRICE DES POIDS DE LA COUCHE CACHEE VERS LA COUCHE DE SORTIE

fprintf('\n matrice des poids couche cachée vers la couche de sortie:
\n');
X2=net.LW{2,1}

%[X2'];
%BIAS DE LA COUCHE CACHEE

fprintf('\n les biais de la couche cachée: \n');
X3=net.b{1}

%[X3'];
%BIAS DE LA COUCHE DE SORTIE

fprintf('\n le biais de la couche de sortie: \n');
X4=net.b{2}
%[X4'];

XLSWRITE('OUAFI1_LOG.xls',[X1'],'LES PARAMETRES','B3');
XLSWRITE('OUAFI1_LOG.xls',[X2'],'LES PARAMETRES','B16');
XLSWRITE('OUAFI1_LOG.xls',[X3'],'LES PARAMETRES','B18');
XLSWRITE('OUAFI1_LOG.xls',[X4'],'LES PARAMETRES','B20');

% POUR APPLICATION ET reconstituer les faciès non carottés (des données
manquantes : missing Data).

XLSWRITE('OUAFI1_LOG.xls',[X1],'APPLICATION','C7');
XLSWRITE('OUAFI1_LOG.xls',[X2'],'APPLICATION','C26');
XLSWRITE('OUAFI1_LOG.xls',[X3'],'APPLICATION','C45');
XLSWRITE('OUAFI1_LOG.xls',[X4'],'APPLICATION','C65');

```




**Recent Advances in Environmental Science
and Geoscience**

**Proceedings of the 2014 International Conference on
Environmental Science and Geoscience (ESG '14)**

Venice, Italy, March 15-17, 2014

Edited by

Vladimir Sokolov

Aida Bulucea

Energy, Environmental and Structural Engineering Series - 23

RECENT ADVANCES in ENVIRONMENTAL SCIENCE and GEOSCIENCE

**Proceedings of the 2014 International Conference on Environmental
Science and Geoscience (ESG '14)**

**Venice, Italy
March 15-17, 2014**

Copyright © 2014, by the editors

All the copyright of the present book belongs to the editors. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the editors.

All papers of the present volume were peer reviewed by no less than two independent reviewers. Acceptance was granted when both reviewers' recommendations were positive.

Series: Energy, Environmental and Structural Engineering Series - 23

ISSN: 2227-4359

ISBN: 978-1-61804-224-8

Table of Contents

<u>Keynote Lecture 1: On the Distinguished Role of the Mittag-Leffler and Wright Functions in Fractional Calculus</u>	11
<i>Francesco Mainardi</i>	
<u>Keynote Lecture 2: Latest Advances in Neuroinformatics and Fuzzy Systems</u>	12
<i>Yingxu Wang</i>	
<u>Keynote Lecture 3: Recent Advances and Future Trends on Atomic Engineering of III-V Semiconductor for Quantum Devices from Deep UV (200nm) up to THZ (300 microns)</u>	14
<i>Manijeh Razeghi</i>	
<u>An Assessment of Self-Organizing Maps and k-means Clustering Approaches for Atmospheric Circulation Classification</u>	17
<i>Despina Deligiorgi, Kostas Philippopoulos, Georgios Kouroupetroglou</i>	
<u>An Environmental Monitoring Plan Related to the Laying of Marine Power Cables: The Case Study of SAPEI Project</u>	24
<i>O. Nonnis, C. Maggi, T. Bacci, A. Izzì, F. Massara, F. S. Rende, B. Trabucco, M. Gabellini</i>	
<u>Consequence of the Overexploitation of Groundwater Sands Aquifer of Biskra -South East Algerian-</u>	30
<i>N. Sedrati, L. Djabri</i>	
<u>Environmental Micropollutants in Livestock Products from Susa Valley in Piedmont</u>	33
<i>R. Desiato, E. Baioni, V. Filipello, G. Cuttica, B. Sparagna, P. Berruto, M. C. Abete, S. Bertolini, C. Maurella, M. Caramelli, L. Chiavacci, G. Ru</i>	
<u>Life Cycle External Cost of Green Electricity: The Case of Greek Power Plants</u>	38
<i>Dimitrios A. Georgakellos, Eleni A. Didaskalou</i>	
<u>The Effectiveness of Electrocoagulation in Wastewater Treatment</u>	43
<i>Demetra Tsigalou, Maria Psaroudi, Christos Tourikas, Efthymios Lytras</i>	
<u>Research of Monthly Precipitation Trends in Libya and Slovakia</u>	46
<i>M. Zeleňáková, P. Purcz, I. Gargar, H. Hlavatá</i>	
<u>Possibilities to Reuse Batteries for Reduction of Waste Volume</u>	51
<i>Bata R., Bačinová I.</i>	
<u>Meeting Rural Water Supply Challenges: Lessons from a South African Groundwater Supply Scheme</u>	56
<i>Jeremiah Mutamba</i>	
<u>The Effect of Landscape Elements on the Satisfaction of the Low Income People in Egypt</u>	62
<i>Rania Rushdy Moussa</i>	

<u>Prediction of Atterberg Limits via ANN and ANFIS: A Comparison</u>	69
<i>Muhammad Mukhlisin, Aini Sharina Binti Abd Rahman</i>	
<u>Using Updated DurovPwin Program for Hydro-Chemical Data Processing: Case Study of Al-Ula Area, Saudi Arabia</u>	75
<i>Abdelaziz M. Al-Bassam, Ahmad R. Khalil, Osama M. Kassem</i>	
<u>The Analysis of the Possible Use of Harvested Rainwater in Real Conditions at the University Campus of Kosice</u>	82
<i>Gabriel Markovič, Daniela Kaposztásová, Zuzana Vranayová</i>	
<u>The Problem of Prediction of Temperature Changes within the Industrial Dumps</u>	89
<i>Blanka Filipova</i>	
<u>Assessing <i>Chenopodium album</i> L. Potential for Phytoremediation of Lead-Polluted Soils</u>	95
<i>Nazli Alipour, Mehdi Homae, Safoora Asadi Kapourchal, Mahboobeh Mazhari</i>	
<u>Use of Cluster Analysis Method in Log's Data Processing: Prediction and Rebuilding of Lithologic Facies</u>	98
<i>Ameur Zaimeche Ouafi, Zeddouri Aziez, Kouadria Taib, Kechiched Rabah, Belaksir Mohammed Salah</i>	
<u>Authors Index</u>	102

Use of Cluster Analysis method in log's data processing: prediction and rebuilding of lithologic facies

Ameur Zaimeche Ouafi¹, Zeddouri Aziez¹, Kouadria Taib², Kechiched Rabah¹, Belaksir Mohammed Salah¹.

1 . Kasdi Merbah Ouargla University, Faculty of Hydrocarbons, Renewable Energy and Earth Sciences and the Universe, Ouargla 30 000 Algeria.

2 . Annaba University, Faculty of Earth Sciences, Annaba 23 000 Algeria.

ouafigeology@gmail.com

Abstract— when a core is missing in a drilling hole, geologists hope to deduce it from others cores available in another part of the hole or in a neighboring, in order to define the lithologic facies. This paper presents a Cluster Analysis method uses to predict the no coring facies from other available cores. This method, based on algorithm which is recognition of facies after coding electro-facies using Wireline logging data especially (GR, ROHB., THOR., POTH) and detailed description of cores [1].

Calculations are made using Matlab and Excel. The prediction and rebuilding results are similar to the real facies in the core. These results showed a significant correlation between prediction, cores facies and corresponding responses of logging peaks (especially Gamma ray). This method has considerable economic benefits permitting a gain of time money and energy compared to coring operation [1].

Keywords— Missing Core, Cluster Analysis, Wireline logging, prediction, Matlab, Excel, Economic

I. INTRODUCTION

In order to investigate the subsoil's composition, geologists should realize cores, but the recuperation is not always completed depending on technical feasibility and geological problems encountered during the coring operation, especially in old wells that are partially or completely non-cored. In fact, our work aim is to overcome the lack of geological information; so sensors were taken down in the drilling holes for measuring physical properties of the rocks. We call logs the result of these measures. We applied this process to deduce non-coring interval in the Lower Triassic Clay Sandstone reservoir (TAGI) in Sif Fatima field in the basin of Berkine, (SE Algeria). Usually geologists use the correlation existing between the different available cores and the no coring interval, to deduce the value they need [3], we propose Cluster Analysis method for helping geologists in predict and rebuilding the lithologic facies.

II. MATERIALS AND METHODS

A. Preparation of core samples

Drilling, cuttings put together by traffic are not always representative For the geologist, the information is incomplete. These cuttings can be from different levels (case of impact) and the ascent of the cuttings is uncertain. In the industry of oil and gas, coring can be defined as "cut and remove a cylindrical sample of rock from the walls of the hole" (Grondin, 2004),

We could identify and discriminate these lithological facies of electro-facies, we give to each facies a code were presented 0-3 to simplify the nomenclature of these facies. Clay code = 0. Silt, code = 1. Sandstone code = 2, Sandstone with clay code = 3

B. Wireline Logging

Well logs or Wireline logs are continuous recordings of well depth versus different petrophysical characteristics of the rocks through which the well is drilled. There are many types of well logs, depending upon the characteristics of the rock being measured. [2]. Logging surveys taken after the casing is lowered are usually categorized as cased hole logs Logging surveys taken before the hole is cased are called open hole logs. We interest with data logging open hole especially (Gamma Ray total, Thorium. Potassium and Density) because they give information on lithology. [7]

1. Total Gamma Ray

The Gamma Ray log is a measurement of the formation's natural radioactivity; his emission is produced by three radioactive series found in the Earth's crust, Potassium, Uranium, Thorium. [7]

2. spectral Gamma ray

Gamma Rays are bursts of high-energy electromagnetic waves that are emitted spontaneously by some radioactive elements. [7] Nearly all the Gamma Radiation encountered on Earth is emitted by:

-Radioactive potassium Isotope (K40) with half-life 1.3×10^9 .

-Uranium 238 (U238) with half-life of 4.4×10^9 .

-Thorium 232 (Th232) with half-life of 1.4×10^{10} .

3. Gamma-Gamma (Density)

Measuring the number of gamma rays and their energy levels at a given distance from the source, the electron density (RHOB).of the formation can be predicted. Is known as litho density tool. [7]. In addition to measuring bulk density, it measures the photoelectric factor (Pe) which is an indicator of formation lithology.

C. Cluster Analysis method

Cluster Analysis method is based on the use of logs in the discrimination of facies, facies are recognized electrofacies. An electrofacies therefore corresponds to a set of samples having similar log responses. Determining electrofacies therefore is highly dependent on selected logs from the stratigraphic interval and or geographical area. Cluster Analysis is a method of partitioning data which is a statistical data analysis method. Its purpose is to divide a set of different data "packets" homogeneous [4,5], in that the data in each subset share common characteristics.

It is also necessary to first perform tests to determine the choice of logs to use, determine the number of studies to be conducted based on stratigraphic or geographies necessary In terms of the process, at first, it was necessary to determine for each facies, the differences between the average value of log (VA, VB, VC, VD) and current values at a defined depth (LA, LB, LC, LD). This fact is obtained from a "log deviation" the facies [8].

Then, we calculated the absolute value of this difference and we normalized using the range to each variable. The "log error" takes values between [0, 1], allowing for comparisons.

Then we calculated the error for each facies, which corresponds to the average of the four "log error" and attributed the facies with the minimum of errors for each depth [1].

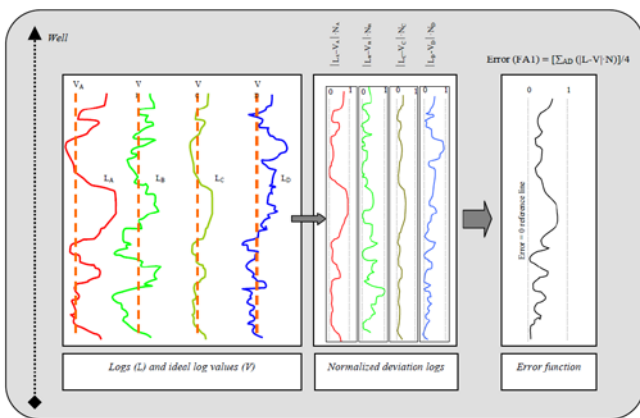


Fig. 1- Diagram summarizes the steps of the Cluster Analysis method.

D. SOFTWARES

1. Matlab

In our work we used the MATLAB software, developed by the American company Matwork. This is very powerful software.

2. EXCEL

This software is a spreadsheet was used primarily for entering and managing data files, it has been used for some calculations,

III. RESULTS AND DISCUSSION

First we calculated the average values of each conventional log, were calculated for each facies the average of all values for each logging in all wells.

Table 1. Averages of logs in all cored wells

Facies	CODE	GR (A)	RHOB (B)	THO (C)	POTA (D)
Clay	0	108,8	2,55	10,86	2,83
Silt	1	91,29	2,57	10,98	2,30
Sandstone	2	64,26	2,43	6,77	1,39
Sandstone and Clay	3	84,98	2,54	7,01	3,29

Then the Maximum, minimum and the range are calculated taking into account all values through all wells maximum and minimum extent of logs in all wells cored values.

Table 1. Maximum, minimum and the range all cored wells

	GR (A)	RHOB (B)	THOR (C)	POTA (D)
MAX	210,29	3,11	23,84	6,10
MIN	13,08	2,12	1,78	0,06
Range	197,21	0,99	22,06	6,04

The most representative peaks are selected as learning samples, these samples determined on the logs are calibrated simultaneously with facies descriptions from core.

- Qualitative visualization of results of prediction

The visualization of the results of the prediction was made with the establishment of logs facies prediction, using the Log discrete control software Petrel (Schlumberger). This allows for a comparison between a prediction and the log of this core. Thus, view logs prediction of non-cored wells and confirm with logging, especially (GR and RHOB). The prediction results of each well in comparison with the core descriptions.

The result of prediction in all the wells was allowed to see the correspondence between a large prediction facies and core facies, there was a decrease in percentage of clays in parallel increase in silts. The difference is negligible in the sandstone.

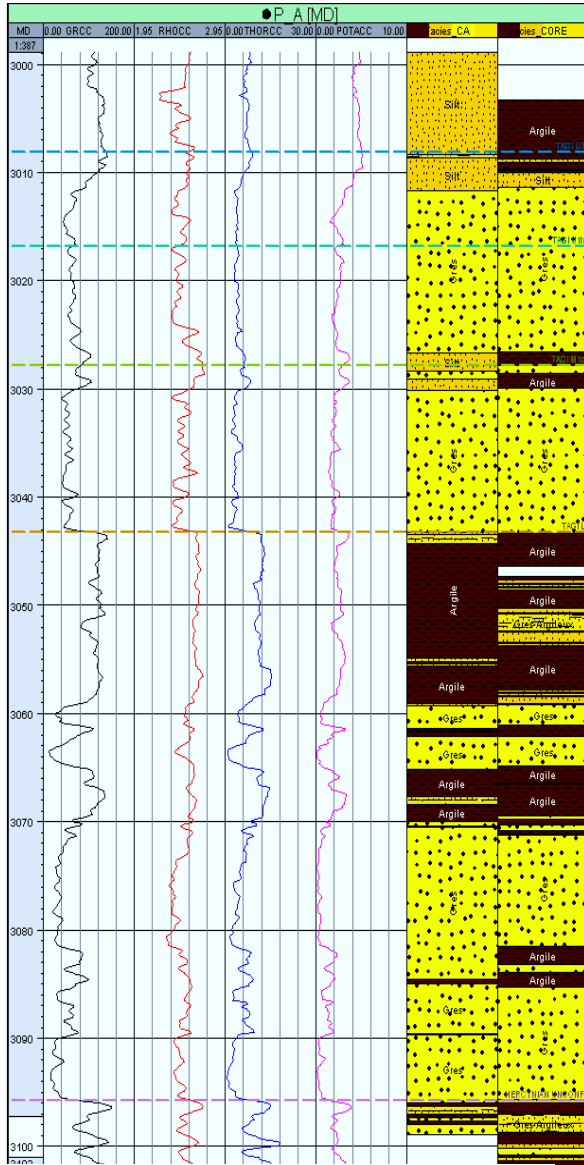


Fig 2. Comparison between results of Prediction and core facies well A

We see that well A , a total lack of carrots by substituting facies prediction. There is a perfect response prediction and facies between peaks of logs (gamma ray, density). It was detected very small benches and matching the peaks of logs determining the facies using prediction is generally the same as the facies described in carrot. It was also found the same entered limits benches with the same answers peaks log.

A good description of the core in this interval, facies prediction matches the description core with human eyes.

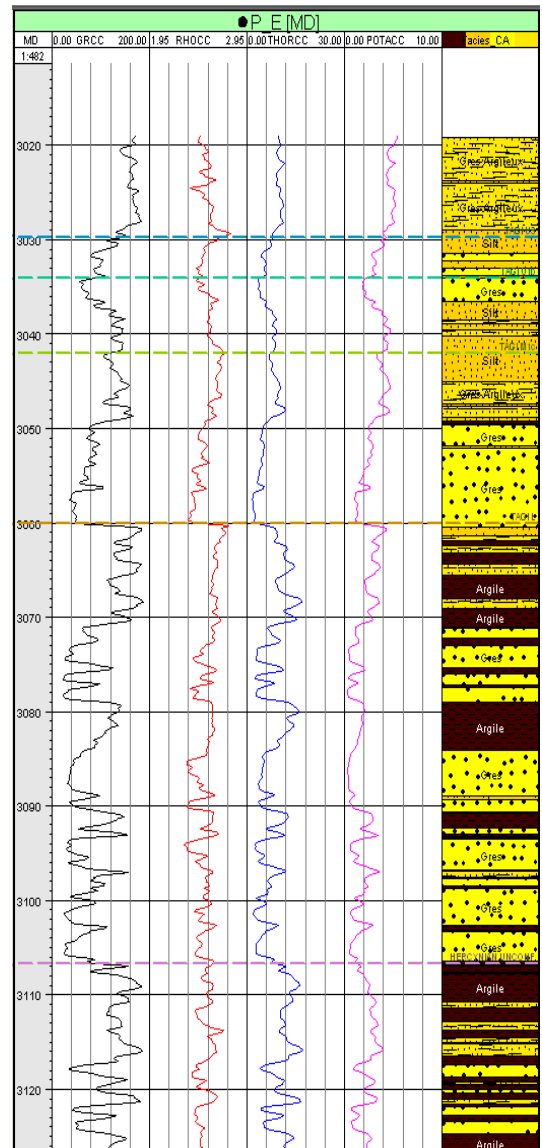


Fig 3.Rebuilding of non-coring well E

We take example to radiuses the results of prediction by the well E is a no coring well and we can get the facies corresponding and we look a response with all logs specially Gamma ray

Correspondence between the response peaks logs with facies predictions.

Apart from very thin levels facies consistent with the peak of gamma ray.so we have a virtual core.

- Quantitative comparison between prediction and core

The quantitative comparison was made for each cores well. The difference between the percentage of log facies and facies of the core was calculated and this, for each facies. The results were expressed as a histogram table monitoring

Table 3. Complete comparison between the prediction and core

Facies	CODE	Core Facies	Prediction Facies
Clays	0	28%	15%
Siltstone	1	13%	23%
Grainstone	2	55%	56%
Granstone with Clays	3	4%	6%

Prediction gave results very close to the description of carrots in this well. However, the percentage of silts prediction is more detailed description that, while the percentage of clays decreased. This observation is explained by the interference of description between clay and silt. The percentage of sandstone is of benefit increased by 12%.it is explained by the intuition of the Geologist when he describe the core, the siltstone are sometimes difficult to distingue from the clay.

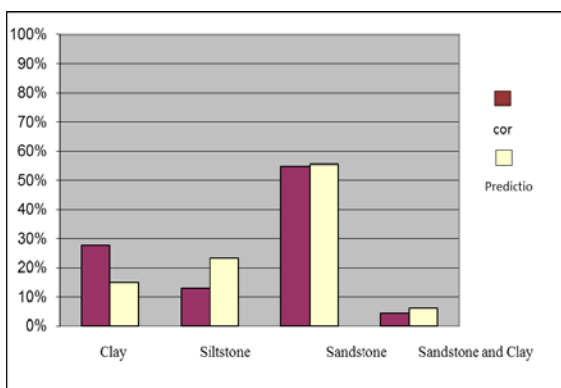


Fig 4. Histogram comparison between the prediction facies core in all coring wells

The result of prediction in all the wells was allowed to see the correspondance between a large facies and facies carrots log (prediction), there was a decrease in percentage of clays in parallel increase in silts. The difference is negligible in the sandstone reservoir facies; it is of the order of one percent (1%).

IV. CONCLUSION

Cluster analysis method was performed for the prediction of non-cored facies and to correct the interference due to difficulties of the macroscopic description. Qualitative and quantitative interpretation were made using a comparison between facies and facies prediction carrot logs and histograms, this comparison has revealed a similarity between

the facies prediction and core especially for the sandstone facies, except clay passages merge with silts. Indeed, the prediction has to distinguish between facies are difficult to differentiate macroscopically (silt and clayey sandstone), and with a perfect match peaks logs logging.

It is less expensive compared with the coring operation. And a considerable gain of time. The simulation-facies present interest as to gain time and help in making the decision. This method is like any simulation method has drawbacks among which: The bad description of carrots will influence the calculation of electro-facies and cause errors at the stage of results (prediction).

We doing also another method for prediction using an unsupervised self-organizing neural network to compare with Cluster Analysis how in two method give a good results with minimal error .

REFERENCES

- [1] Ameur Zaimeche, O. (2012): Prédiction des facies non carottés par la méthode de Cluster Analysis et contribution à l'étude du réservoir TAGI du Bassin de Berkine SW, Mémoire de fin D'étude, UNIVERSITÉ BADJI MOKHTAR-ANNABA, W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
- [2] Chapellier, D. (1998): cours online de géophysique, *Université de Lausanne, Institute français du pétrole*.
- [3] Dominique Frayssinet use of neural networks in log's data processing: prediction and rebuilding of lithologic facies
- [4] J.C. Bezdek, R. Ehrlich, W. Full, FCM: the Fuzzy c-Means clustering algorithm, *Comput. Geosci.* 10 (1984) 191–203.
- [5] Louis Briqueua, Traitement des diagraphies à l'aide d'un réseau de neurones du type « carte auto-organisatrice » : application à l'étude lithologique de la couche silteuse de Marcoule (Gard, France
- [6] Rabaute A. (1999): Obtenir une représentation en continu de la lithologie et de la minéralogie. Exemples d'applications du traitement statistique de données de diagraphie aux structures sédimentaires en régime de convergence de plaques (Log ODP 134, 156 et 160). Thèse de doctorat, Université de Montpellier II. Mémoires géosciences – Montpellier.
- [7] SERRA, O. (1985): -Diagraphies Différés base de l'interprétation, Mémoire7 Tome2. Etudes et productions Schlumberger, Montrouge
- [8] They, P. P. (1991): Log data acquisition and quality control. Editions Technip, Paris.