

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة سعد دحلب البليدة
Université Saad Dahleb de Blida
كلية الهندسة
Faculté des Sciences de l'Ingénieur



Mémoire pour l'obtention du diplôme de Master
Spécialité : **Chimie Industrielle**
Option : **Génie Chimique**

THEME

**Modélisation de l'isotherme
d'adsorption du Cs dans des sols de la
région de Birine.**

Présent par :

MIL. KACI SMINA

Sous la direction de :

Mr. BOUZIDI ABDELKADER (Chargé de recherche)

Mme. AÏT MESSBAH

Devant le jury composé de :

Président

Examineur

Examineur

Dédicace

*J'ai l'honneur de dédier ce mémoire à mes chers parents **Abdelkader et Mékjoussa**, qui m'ont soutenue durant toutes ces années d'étude, qui ont consacré toute leur vie pour ma réussite.*

A toute ma famille. Un immense merci à ma famille la plus proche qui a toujours été présent et avec laquelle je partage tant de bonheur:

➤ *Ma grande mère fatima.*

➤ *Mes sœurs : Djamila, pour avoir toujours été un exemple*

Nabila, que sans elle je ne puisse pas continuer mes études.

Nadia, pour son aide.

➤ *Mes frères : Sofiane, pour ces encouragements.*

Mon petit frère Aghiless.

➤ *Monsieur A. SARI et sa femme.*

Et comme je le dédie à mes amis: Ilhame, Aïcha, Rahima, Barkahoum, Meriem, Lilia, Nabile, Mohamed, Youcef, Mourad et tous mes collègues à l'université.

Et tous ceux qui m'ont aidé de près ou de loin.

Merci à tous.

Remerciements

*En premier lieu, je remercie **Dieu** tout puissant qui m'a donnée la force de mener à terme ce travail.*

Je tiens très sincèrement à remercier mon promoteur, Mr. A. BOUZIDI Chargé de recherche et chef de laboratoire d'analyses chimiques du Centre de Recherche Nucléaire de Birine, sans ses encouragements et aides je ne serais jamais arrivée à ce stade de ma formation. Il a toujours su me faire confiance et m'apporter l'aide nécessaire, tant sur le plan scientifique que moral.

Je remercie Madame AÏT MESSBAH d'avoir accepté d'être ma Co-promotrice.

Je remercié évidemment le président et les membres du jury d'avoir accepté de juger ce modeste travail.

Je souhaite exprimer mes sincères et respectueuses reconnaissances et remerciements à Monsieur N. BOUCHERIT maître de recherche et chef du département, pour son aide et les moyens qu'il a mis à ma disposition ainsi qu'à tous les membres du département de chimie des réacteurs, aux Messieurs F. ARBAOUI, A. ARRAREM, S. AMZERT, F. REBHI, K. BRAHIMI, M. BELLABAS, A. RADJEM et Mme MERROUCHE W.

Mes remerciements vont aussi à Messieurs A. SARI, K. Saadallah, Madame Louanes et Mademoiselle Nadjet, pour leur aide et encouragement.

Je n'oublie pas de présenter mes remerciements à tous ce qui ont contribué de près ou de loin à la réalisation de ce travail.

SOMMAIRE

Nomenclature & Abréviation	i
Liste des figures	v
Liste des tableaux	viii
Résumé	ix

INTRODUCTION GENERALE	01
-----------------------------	----

CHAPITRE I : *ISOTHERMES D'ADSORPTION*

I.1. Définition	03
I.2. Types d'adsorption	03
I.2.1. Adsorption physique	03
I.2.2. Adsorption chimique	03
I.3. Mécanismes d'adsorption	04
I.3.1 Adsorption par liaison ionique ou échange d'ions	04
I.3.2 Adsorption par liaison hydrogène	04
I.3.3 Adsorption par les forces de Van der Waals	05
I.3.4 Rétention hydrophobe	05
I.4. Paramètres affectant l'adsorption	05
I.4.1 Caractéristiques de la molécule	05
I.4.2. Structure de l'adsorbant	06
I.5. L'équilibre adsorption/désorption	07
I.5.1. Approche thermodynamique	07
I.5.2. Approche expérimentale	09
I.5.3. Modélisation des isothermes d'équilibre mono-soluté	12
I.5.3.1. Modèles mono-soluté à deux paramètres	12
I.5.3.1.1. Modèle de FREUNDLICH	12
I.5.3.1.2. Modèle de LANGMUIR	13
I.5.3.1.3. Modèle de TEMKIN	14
I.5.3.1.4. Modèle d'ELOVICH	15
I.5.3.1.5. Modèle de DUBININ-RADUSHKEVICH	15

I.5.3.2. Modèles mono-soluté à trois paramètres	16
I.5.3.2.1. Modèle de REDLICH-PETERSON	16
I.5.3.2.2. Modèle de TÓTH	17
I.5.3.2.3. Modèle de LANGMUIR-FREUNDLICH	18
I.5.3.2.4. Modèle GENERALISE	18
I.5.3.2.5. Modèle de SIPS ou KOBLE-CORRIGAN	18
I.5.3.2.6. Modèles de RADKE-PRAUSNITZ	19
I.5.3.2.7. Autres modèles à trois paramètres	19
I.5.3.3. Modèles mono-soluté à plus de trois paramètres	20
I.6. Cinétique d'adsorption	20
I.7. Le césium dans les sols	22
I.7.1. Mécanismes de la rétention du césium par les sols	23
I.7.2. Migration du césium dans les sols	24

CHAPITRE II : RESEAUX DE NEURONES ARTIFICIELS

II.1 Introduction	25
II.2. Historique	26
II.3. Neurone biologique	27
II.4. Neurone formel	28
II.4.1. Entrées	29
II.4.2. Poids synaptiques	29
II.4.3. Seuil	30
II.4.4. Fonction d'activation	30
II.4.4.1. Fonction identité	31
II.4.4.2. Fonction linéaire à seuil	31
II.4.4.3. Fonction non linéaire	32
II.4.4.3.1. Fonction sigmoïde logarithmique	32
II.4.4.3.2. Fonction sigmoïde tangentielle	32
II.4.4.4. Fonction Gaussienne	33
II.4.5. Sortie du neurone "S"	33
II.5. Architecture et connexionnisme des réseaux de neurones	33
II.5.1. Réseau de neurones non bouclés	37
II.5.1.1. Perceptron	38

II.5.1.2. Perceptron multicouche (MLP)	39
II.5.2. Réseaux de neurones bouclés	39
II.6. Apprentissage	40
II.6.1. Apprentissage supervisé	41
II.6.2. Apprentissage non supervisé	41
II.7. Règles d'apprentissage	42
II.7.1. Règle d'apprentissage de Hebb	42
II.7.2. Règle d'apprentissage du perceptron de F. Rosenblatt	43
II.7.3. Règle d'apprentissage de Windrow-Hoff (ADALINE)	44
II.7.4. Algorithme de rétropropagation du gradient	46
II.7.4.1. Méthode de Newton	51
II.7.4.2. Méthode de quasi-Newton	52
II.7.4.3. Méthode Levenberg-Marquardt	53
II.8. Théorie de la généralisation	53
II.9. Critère d'arrêt pour l'apprentissage	55
II.10. Procédure de modélisation neuronale	56
II.10.1. Sélection des données	57
II.10.2. Collecte des données	58
II.10.3. Séparation des bases de données	58
II.10.4. Normalisation des données	58
II.10.5. Choix du type de réseau de neurones	59
II.10.6. Choix du nombre de neurones dans la couche cachée	59
II.10.7. Choix de la fonction de transfert	59
II.10.8. Estimation des paramètres ou apprentissage	59
II.10.9. Validation et généralisation	60
II.11. Modélisation à l'aide de réseaux de neurones	60
II.11.1. Modèle « boîte noire »	60
II.11.2. Modèle « boîte grise » ou hybride	61
II.12. Conclusion	62

CHAPITRE III : *PROTOCOLES EXPERIMENTAUX*

III.1. Introduction	63
III.2. Echantillonnage	63

III.3. Procédures expérimentales	63
III.3.1. Isotherme d'adsorption	64
III.3.2. Détermination de K_d	64
III.3.3. Fraction fine	66
III.3.3.1. Préparation des échantillons	67
III.3.3.2. Destruction de la matière organique	67
III.3.3.3. Tamisage	68
III.3.3.4. Dispersion minérale	68
III.3.3.5. Sédimentation et prélèvement de la fraction fine et du limon fin.....	68
III.3.4. Matière organique et carbonates	70
III.3.4.1. Matière organique	70
III.3.4.2. Carbonates	71
III.3.5. Capacité d'échange cationique	71
III.3.6. L'énergie d'activation	71
III.4. Méthodologie de la conception des réseaux de neurones	72
III.4.1. Base de données	73
III.4.1.1. Sélection des entrées et des sorties du modèle neuronal	74
III.4.1.2. Collecte des données	75
III.4.1.3. Mise en forme des données	75
III.4.1.4. Normalisation des données	75
III.4.2. Phase de conception du réseau neuronal	76
III.4.2.1. Choix du type de réseau de neurones	76
III.4.2.2. Choix du nombre des couches cachées	77
III.4.2.3. Choix du nombre de neurones dans la couche cachée	77
III.4.2.4. Choix de la fonction de transfert	77
III.4.2.5. Algorithme d'apprentissage	77
III.4.2.6. Nombre d'itérations (cycles)	78
III.4.3. Phase de finalisation du réseau neuronal	78
III.4.3.1. Apprentissage	78
III.4.3.2. Validation	78
III.4.3.3. Généralisation	78
III.4.3.4. Estimation des paramètres du réseau	79
III.4.3.5. Résultats et discussion	79
III.5. Conclusion	79

CHAPITRE IV : RESULTATS ET DISCUSSIONS

IV.1. Introduction	80
IV.2. Résultats liés aux travaux expérimentaux	80
IV.2.1. Isothermes	80
IV.2.2. Détermination de K_d	81
IV.2.3. Fraction fine	82
IV.2.4. Matière organique et carbonates	82
IV.2.5. Capacité d'échange cationique	83
IV.2.6. L'énergie d'activation	83
IV.3. Résultats liés à la conception	83
IV.3.1. Conception du réseau	83
IV.3.2. Nombre de couches cachées	84
IV.3.3. Nombre de neurones cachés	85
IV.3.4. Effet du nombre d'itération	86
IV.3.5. Le réseau neuronal conçu	86
IV.3.6. Paramètres du réseau	87
IV.4. Résultats et discussion	87
IV.4.1. La première sortie (Q_e)	88
IV.4.2. La deuxième sortie (CEC)	89
IV.4.3. La troisième sortie (E_a)	91
IV.4.4. La Quatrième sortie (K_d)	92
IV.5. Comparaison des isothermes	94
IV.5.1. Sol A	96
IV.5.2. Sol B	98
IV.5.3. Sol C	101
IV.5.4. Sol D	103
IV.6. Conclusion	105
CONCLUSION ET PERSPECTIVES.....	106
REFERENCES BIBLIOGRAPHIQUES	108
ANNEXE A	113
ANNEXE B	119

INTRODUCTION GENERALE

La rétention des polluants particulièrement radioactifs est un processus très complexe qui conjugue des phénomènes dynamiques, tels que la diffusion et la dispersion hydrodynamique et des phénomènes statique, tels que la sorption.

La mobilité et la biodisponibilité des polluants métalliques dans l'environnement dépendent alors de la forme physico-chimique sous laquelle se trouvent ces éléments et aussi des caractéristiques du sol. Par conséquent, il est essentiel de déterminer selon quel mécanisme ces éléments sont associés à la matrice géologique et selon quel mode d'adsorption sont fixés aux sols. L'identification et la détermination de ces formes de rétentions sont souvent sujettes aux tracés des isothermes.

L'isotherme d'adsorption est la courbe liant, à une température fixe, la quantité de substance adsorbée par unité de masse d'adsorbant à la concentration restant en phase fluide après équilibre d'adsorption. Les intérêts de l'isotherme d'adsorption pour un système adsorbant/adsorbat sont multiples. Elle est par exemple, indispensable pour avancer des hypothèses sur le mode d'adsorption. En effet, son allure est représentative de certains phénomènes mis en jeu entre autres : adsorption monocouche ou multicouches, existence d'interactions latérales entre molécules ou non.

De nombreuses modélisations sont utilisées et parfois développées dans la littérature. La plupart de ces modèles sont basés sur des concepts théoriques. La multiplication des hypothèses « simplificatrices » pour la conception de ces modèles leur faire perdre une partie de leur validité. Inscrit ce contexte, ce travail vise la proposition d'une nouvelle technique de modélisation basée sur les réseaux de neurones.

L'objectif du présent mémoire est de développer une nouvelle approche basée sur les réseaux de neurones artificiels pour la modélisation des isothermes de quatre sols de la région Birine.

Cette approche inspirée du traitement de l'information effectué par le cerveau, semble offrir un nouvel espace de modélisation tout à fait original. En effet, les réseaux de neurones artificiels (RNA) suscitent toujours un vif intérêt dans la plupart des domaines d'ingénierie non seulement pour leur attirante capacité d'apprentissage mais aussi pour leur flexibilité et leur bonne performance, par rapport aux approches classiques. Les RNA sont capables d'approximer des relations complexes et fortement non linéaires entre un vecteur de variables d'entrées x et une sortie y .

Le présent mémoire est subdivisé en quatre chapitres. Le premier chapitre est consacré à la présentation des isothermes d'adsorption et aux modèles théoriques. Le deuxième chapitre en revanche est dédié aux réseaux de neurones d'une façon générale : historique, inspiration biologique, neurone formel, réseau multicouches et règles d'apprentissage. Nous avons, bien entendu, mis l'accent sur les réseaux non bouclés type MLP (Multi-Layer Perceptron) du fait qu'ils sont utilisés dans les modélisations statiques ainsi que sur l'algorithme de rétropropagation du gradient qui leur est particulier. Quant au troisième chapitre, on y présente les protocoles expérimentaux des différentes expériences effectuées aux laboratoires. Le quatrième et dernier chapitre est dédié à la présentation des résultats et discussions. En fin, nous clôturons ce travail par une conclusion générale et recommandations.

Liste des tableaux

Tableau I.1: Concentration du césium dans les roches et les sols en $\mu\text{g}/\text{g}^{-1}$ sec	22
Tableau III.1 : Echelle granulométrique des constituants du sol	67
Tableau III.2 : Vitesse de sédimentation	69
Tableau III.3 : Paramètres cinétiques pour les quatre sols	72
Tableau IV.1 : Pourcentage de la fraction fine pour chaque sol.....	82
Tableau IV.2 : Détermination de la matière organique et des carbonates par la PAF.....	82
Tableau IV.3 : Valeurs de la CEC pour chaque sol	83
Tableau IV.4 : Valeurs des énergies d'activation des quatre sols	83
Tableau IV.5 : Structure du réseau neuronal utilisé dans la modélisation des isothermes d'adsorption	87
Tableau IV.6 : Description des modèles étudiés avec leurs équations	95
Tableau IV.7 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol A	97
Tableau IV.8 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol A	98
Tableau IV.9 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol B	99
Tableau IV.10 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol B	101
Tableau IV.11 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol C	102
Tableau IV.12 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol C	103
Tableau IV.13 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol D	103
Tableau IV.14 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol D	104

Liste des figures

Figure I.1: Les quatre types d'isothermes selon GILES et al.	10
Figure I.2: Différents types d'isothermes d'adsorption des gaz selon BET	11
Figure I.3: Différents systèmes de rétention, A: particule non poreuse, B: particules poreuses, C: milieux poreux structuré	22
Figure II.1 : Schématisation d'un neurone biologique	28
Figure II.2 : Mise en correspondance neurone biologique / neurone artificiel	28
Figure II.3 : Neurone formel	29
Figure II.4 : Neurone formel modèle étendu	30
Figure II.5 : Fonction identité	31
Figure II.6 : Fonction linéaire à seuil	31
Figure II.7 : Fonction sigmoïde logarithmique	32
Figure II.8 : Fonction sigmoïde tangentielle	32
Figure II.9 : Fonction Gaussienne	33
Figure II.10 : Couche de M neurones	34
Figure II.11 : Représentation matricielle d'une couche de M neurones	35
Figure II.12 : Représentation matricielle d'un réseau de trois couches	35
Figure II.13 : Élément de retard	36
Figure II.14 : Taxinomie des architectures des réseaux non bouclés et récurrents	37
Figure II.15 Réseau multicouche non bouclé classique	38
Figure II.16 : Principe d'un perceptron	39
Figure II.17 : Un réseau de neurone bouclé à deux entrées contenant une boucle ente les neurones (1) et (2)	40
Figure II.18 : Illustration de l'apprentissage supervisé	41
Figure II.19 : Illustration de l'apprentissage non supervisé	42
Figure II.20 : Séparation des exemples d'apprentissage par la droite de décision	43
Figure II.21 : Sens de propagation du calcul des sorties s_i	47
Figure II.22 : Sens de rétropropagation des gradients δ_i	48
Figure II.23 : Graphe de la surface d'erreur	50
Figure II.24 : Extraction de règles	54
Figure II.25 : Cas d'apprentissage [19] a. Apprentissage correct, b. Sur-apprentissage et c. Sous-apprentissage.....	54

Figure II.26: Evolution des erreurs d'apprentissage et de test au cours du temps	56
Figure II.27 : Diagramme schématique d'un modèle neuronal « boîte noire »	61
Figure II.28 : Diagramme schématique de modèle neuronal hybride d'un réacteur chimique	61
Figure III.1 : Mécanisme de la dispersion minérale de la fraction fine	68
Figure III.2: Méthodologie de calcul des différents paramètres par les réseaux de neurones	73
Figure IV.1 : Isotherme d'adsorption du Cs sur le sol (A)	81
Figure IV.2 : Isotherme d'adsorption du Cs sur le sol (B)	81
Figure IV.3 : Isotherme d'adsorption du Cs sur le sol (C)	81
Figure IV.4 : Isotherme d'adsorption du Cs sur le sol (D)	81
Figure IV.5 : Evaluation du nombre de couches cachées	84
Figure IV.6 : Evaluation du nombre de neurones de la couche cachée	85
Figure IV.7: Evaluation du nombre de cycles d'apprentissage	86
Figure IV.8 : Schéma simplifié de l'architecture du réseau neuronal avec la fonction de transfert globale	87
Figure IV.9 : Résultats d'apprentissage pour les valeurs de la quantité de soluté retenue par le solide (Q_e)	88
Figure IV.10 : Résultats de validation pour les valeurs de la quantité de soluté retenue par le solide (Q_e)	89
Figure IV.11 : Résultats du test de généralisation pour les valeurs de la quantité de soluté retenue par le solide (Q_e)	89
Figure IV.12 : Résultats d'apprentissage pour les valeurs de la capacité d'échange cationique (CEC)	90
Figure IV.13 : Résultats de validation pour les valeurs de la capacité d'échange cationique (CEC)	90
Figure IV.14 : Résultats du test de généralisation pour les valeurs de la capacité d'échange cationique (CEC)	91
Figure IV.15 : Résultats d'apprentissage pour les valeurs de l'énergie d'activation (E_a)	91
Figure IV.16 : Résultats de validation pour les valeurs de l'énergie d'activation (E_a)	92
Figure IV.17 : Résultats du test de généralisation pour les valeurs de l'énergie d'activation (E_a)	92
Figure IV.18 : Résultats d'apprentissage pour les valeurs du coefficient de distribution (K_d)	93
Figure IV.19 : Résultats de validation pour les valeurs du coefficient de distribution (K_d)...	93
Figure IV.20 : Résultats du test de généralisation pour les valeurs du coefficient de	

distribution (K_d)	94
Figure VI.21 : Comparaison de l'isotherme expérimentale avec les modèles (neuronale, Freundlich-Langmuir et Radke-Prausnitz) pour le sol A	96
Figure VI.22 : Comparaison de l'isotherme expérimentale avec les modèles (neuronale, Langmuir, Temkin et Freundlich) pour le sol A	98
Figure VI.23 : Comparaison de l'isotherme expérimentale avec les modèles (neuronale, Freundlich-Langmuir et Radke-Prausnitz) pour le sol B	99
Figure VI.24 : Comparaison de l'isotherme expérimentale avec les modèles (neuronale, Langmuir, Temkin et Freundlich) pour le sol B	100
Figure VI.25 : Comparaison de l'isotherme expérimentale avec les modèles (neuronale, Freundliche-Langmuir et Radke-Prausnitz) pour le sol C	101
Figure VI.26 : Comparaison de l'isotherme expérimentale avec les modèles (neuronale, Langmuir, Temkin et Freundlich) pour le sol C	102
Figure VI.27 : Comparaison de l'isotherme expérimentale avec les modèles (neuronale, Freundlich-Langmuir et Radke-Prausnitz) pour le sol D	103
Figure VI.28 : Comparaison de l'isotherme expérimentale avec les modèles (neuronale, Langmuir, Temkin et Freundlich) pour le sol D	104

Résumé :

Le but de cette étude est la mise en œuvre d'une technique basée sur les réseaux de neurones artificiels de type multicouches (MLP), pour la modélisation des isothermes de quatre types de sols de la région de Birine.

Dans ce contexte, nous avons développé un modèle neuronale à quatre sorties. La première sortie concerne la quantité du césium retenue par chaque sol (Q , mg/g), cette grandeur est nécessaire pour le tracé des isothermes. Les trois autres sortie sont : la capacité échange cationique (CEC) qui nous renseigne sur le pouvoir de rétention des sols, l'énergie d'activation qui mesure les barrières énergétique s'opposant à l'adsorption du Cs et le dernier paramètres est le coefficient de distribution K_d qui traduit le taux de sorption.

En fin, nous avons effectué une étude comparative entre le modèle neuronale et les modèles proposés dans la littérature. Cette étude a révélé, d'une part, la précision des résultats obtenus, et d'autre part, l'efficacité du modèle conçu par rapport aux autres modèles.

Mot clés : réseaux de neurones artificiels, isothermes, Birine, CEC, énergie d'activation, coefficient de distribution.

Abstract:

The aim of this study is the implementation of a technique based on artificial neural networks of multilayer type (MLP) to model the isotherms of four types of soils in Birine area.

In this context, we developed a neural model with four outputs. The first output is the amount of cesium retained by each soil (Q , mg / g), this quantity is necessary to plot the isotherms. The other three outputs are: cation exchange capacity (CEC), which provides information on the retention capacity of soils, the activation energy which measures the energy barriers opposing to the Cs adsorption and the last parameter is the distribution coefficient K_d which reflects the rate of sorption.

Finally, we have made a comparative study between the neural model and models proposed by the literature. This study has revealed, in one part, the precision of the results obtained and, in the other part, the efficiency of the designed model compared to the other models.

Key words: artificial neural networks, isotherms, Birine, CEC, activation energy, distribution coefficient.

ملخص:

الغرض من هذه الدراسة هو تطبيق تقنية تقوم على الشبكات العصبية الاصطناعية من نوع متعدد الطبقات (MLP) و ذلك لتقدير الأيسوثرم (isothermes) لأربعة أنواع من التربة في منطقة البيرين.

في هذا السياق ، قمنا بتطوير نموذج عصبي اصطناعي من أربع مخرجات. المخرج الأول يدل على كمية السيزيوم التي احتفظت بها كل التربة (Q mg / g) ، وهذا الكم الضروري لرسم الأيسوثرم. اما باقي الثلاثة مخرج فهي : القدرة لتبادل الأيونات الموجبة (CEC)، الذي يوفر معلومات عن قدرة التربة على التخزين ، وطاقة التنشيط التي تقيس حواجز الطاقة المانعة لعملية ادمصاص السيزيوم والمخرج الأخير هو معامل التوزيع Kd وهو يعكس معدل الامتصاص.

في الأخير ، أجرينا دراسة مقارنة بين النموذج العصبي والنماذج منشورة في بحوث علمية. أثبتت هذه الدراسة من ناحية دقة النتائج، وكذا فعالية نموذج المصمم مقارنة بنماذج الأخرى.

الكلمات المفاتيح : الشبكات العصبية الاصطناعية، البيرين، الأيسوثرم ، CEC، الطاقة التنشيط ، معامل التوزيع.

I.1. Définition

L'adsorption est un procédé de traitement, bien adapté pour éliminer une très grande diversité de composés toxiques dans notre environnement. Elle est essentiellement utilisée pour le traitement de l'eau et de l'air. Au cours de ce processus, les molécules d'un fluide (gaz ou liquide), appelé adsorbat, viennent se fixer sur la surface d'un solide, appelé adsorbant. Ce procédé définit la propriété de certains matériaux de fixer à leur surface des molécules (gaz, ions métalliques, molécules organiques, etc.) d'une manière plus ou moins réversible. Au cours de ce processus, il y aura donc un transfert de matière de la phase aqueuse ou gazeuse vers la surface solide. Le solide acquiert alors des propriétés superficielles (hydrophobie ou hydrophilie) susceptibles de modifier l'état d'équilibre du milieu (dispersion, floculation) [1, 2, 3].

I.2. Types d'adsorption

La nature des liaisons formées ainsi que la quantité d'énergie dégagée lors de la rétention d'une molécule à la surface d'un solide permettent de distinguer deux types d'adsorption : adsorption physique et adsorption chimique [2, 4, 5].

I.2.1. Adsorption physique

Ce type d'adsorption résulte de l'établissement d'un échange de forces de faible énergie entre la surface d'un solide et des molécules à proximité de cette surface. Dans ce cas, la rétention est le résultat des liaisons de nature électrostatique de type Van Der Waals [1].

Du point de vue énergétique, la physisorption se produit aux basses températures avec des énergies inférieures à $40 \text{ kJ}\cdot\text{mol}^{-1}$. Elle correspond à un processus réversible (équilibre dynamique d'adsorption et de désorption) et ne conduit pas à une modification de l'identité chimique de la molécule adsorbée.

Dans le cas d'une telle adsorption, le temps de rétention de la substance adsorbée est court et la surface adsorbante peut être recouverte de multiples couches moléculaires de produit adsorbé [2, 6].

I.2.2. Adsorption chimique

Dans ce cas, l'adsorption est due à la formation d'une liaison chimique, covalente plus permanente, entre l'adsorbat et la surface de l'adsorbant. Par rapport au premier, l'adsorption chimique se distingue par des énergies d'adsorption plus élevées ($> 40 \text{ kJ}\cdot\text{mol}^{-1}$) et par une

fixation irréversible de l'adsorbat sur des sites d'adsorption très spécifiques. Elle est aussi favorisée à température élevée [4, 7].

Dans ce type d'adsorption, le temps de rétention est long et seule la première couche liée à la surface adsorbante est chimiquement adsorbée, les autres couches, dans le cas où elles existent, sont retenues par physisorption.

Il est à noter également que le phénomène d'adsorption constitue la première étape des réactions nécessitant l'emploi d'un catalyseur solide. Ce phénomène peut alors jouer un rôle prédominant dans la cinétique de réaction chimique.

I.3. Mécanismes d'adsorption

Pour mieux qualifier et quantifier la rétention, il convient de s'intéresser aux phénomènes se produisant à l'échelle moléculaire, c'est-à-dire aux mécanismes d'adsorption. Les liaisons composés/adsorbant sont de deux types :

- ✓ Liaisons de fortes énergies ($> 40 \text{ kJ. mol}^{-1}$) : liaisons ioniques et échanges de ligands ;
- ✓ Liaisons de faibles énergies ($< 40 \text{ kJ. mol}^{-1}$) : interactions dipôle-dipôle, liaisons hydrogène, interactions hydrophobes.

Sur la base de ces liaisons, quatre mécanismes principaux peuvent être distingués [8] :

I.3.1 Adsorption par liaison ionique ou échange d'ions

Ce mécanisme ne concerne que les produits sous forme cationique ou ceux qui peuvent le devenir, comme les bases faibles. Il s'agit d'une interaction entre la molécule et les groupements ionisés ou facilement ionisables de l'adsorbant, tels que les groupements phénoliques et carboxyliques présents dans la structure des charbons. Ces liaisons sont le plus souvent peu réactives.

I.3.2 Adsorption par liaison hydrogène

Certains supports peuvent présenter à leur surface de nombreuses fonctions (oxygénées, hydroxyles...) qui peuvent interagir par liaisons hydrogène avec les groupes complémentaires des molécules adsorbat, bien qu'il y ait une forte compétition avec les molécules d'eau pour ces sites d'adsorption.

I.3.3 Adsorption par les forces de Van der Waals

Les forces de Van der Waals sont des attractions dipolaires de faible intensité qui agissent à courte distance. Leur action s'additionne à d'autres types d'interactions plus fortes. Dans le cas de molécules non-ioniques et apolaires, leur participation à la rétention n'est pas négligeable. Si la molécule est assez volumineuse et qu'elle peut s'adapter à la surface de l'adsorbant, ces forces sont alors additives et leur rôle peut devenir important.

I.3.4 Rétention hydrophobe

Ce type de mécanisme concerne en particulier les composés non-ioniques et apolaires. Deux approches peuvent être distinguées :

✓ Certains auteurs proposent un mécanisme d'adsorption hydrophobe indépendant du pH [9]. Les molécules des produits phytosanitaires apolaires interagissent avec les sites hydrophobes du support comme les chaînes aliphatiques, les fractions lipidiques, les dérivés de la lignine (riches en carbone avec peu de groupes fonctionnels). Ces sites sont relativement accessibles car il n'y a quasiment aucune compétition avec les molécules d'eau.

✓ D'autres auteurs préfèrent ne pas parler d'adsorption mais plutôt d'une partition [10, 11]. Ils assimilent la surface de l'adsorbant à un solvant liquide non-miscible à l'eau.

I.4. Paramètres affectant l'adsorption

Un grand nombre de paramètres et de propriétés, du support et du substrat, peuvent influencer le processus d'adsorption et notamment la capacité et la cinétique de rétention d'une substance sur un support [12, 7]. Il s'agit des paramètres suivants :

I.4.1 Caractéristiques de la molécule

a) Taille des molécules adsorbés : la disposition des molécules sur la surface du matériau peut fortement affecter le processus d'adsorption. Nous citons comme exemple la fixation, sur un support et à la verticale, des acides et des alcools à longue chaîne carbonée par l'intermédiaire de leur groupement carboxylique (-COOH) et hydroxyle (-OH) respectivement dans le cas d'un recouvrement élevé (forte concentration du substrat). Ceci conduit à une fixation forte du substrat [13].

En effet, d'un point de vue purement mécanique, il faut que la taille de la molécule soit inférieure au diamètre du pore d'adsorbant pour que celle-ci puisse diffuser rapidement dans le volume poreux et atteindre le site d'adsorption.

b) Solubilité : plusieurs chercheurs ont montré que les constantes d'adsorption ont tendance à être plus importants quand la solubilité du composé diminue. D'autres chercheurs, présentent un certain nombre de relations entre la constante d'adsorption et diverses propriétés de la molécule organique.

c) $p K_a$: un certain nombre de produits organiques sont caractérisés par des propriétés d'acides faibles ou de bases faibles. Le pH conditionne donc la forme sous laquelle se trouve la molécule (ionisée ou neutre). Ainsi, ce ne sont pas les mêmes éléments de la matrice qui interviendront dans le mécanisme d'adsorption. La majorité des études, montrent que la rétention est maximale lorsque le pH est égal au pKa [14, 10, 11, 15]. Toutefois, il ne faut pas confondre le pH de l'eau et le pH à la surface des sédiments. En général, ce dernier est inférieur d'environ deux unités, car il dépend de la qualité de groupements carboxyliques et phénoliques [9]. Le pH reste cependant un facteur limitant du milieu puisqu'il ne peut pas être facilement modifié.

d) Polarité et polarisabilité de la molécule adsorbée : l'adsorption va être fortement influencée par les dimensions du composé à piéger (surface, volume), mais aussi par les groupements fonctionnels de la molécule (alcools, aldéhydes, cétones, acides carboxyliques, amines, soufre, halogène...) induisant des effets de polarisabilité plus ou moins marqués.

Des études importantes ont été entreprises afin de déterminer des relations quantitatives entre la structure moléculaire et les paramètres d'adsorption (capacité d'adsorption, énergie d'interaction) [16, 17]. La forte réactivité de certaines molécules peut donner lieu à des réactions d'oxydation à la surface de l'adsorbant qui joue alors le rôle de catalyseur. Un mélange de composés impliquera donc une compétition d'adsorption entre les divers composés et réduire les capacités unitaires d'adsorption.

I.4.2. Structure de l'adsorbant

Elle joue un rôle déterminant dans la fixation du substrat, par la taille de ses particules. En effet plus la taille est petite, plus la surface de contact sera grande. Ce qui va alors augmenter son pouvoir de rétention. En outre, cette surface s'accroît lorsque le support est poreux.

Les solides dans les milieux naturels (argiles, silice..) possèdent des surfaces spécifiques variables avec l'état physico-chimique du milieu aqueux (pH, nature des cations liés, saturation de la surface par les molécules organiques...). Ainsi certaines argiles comme les bentonites (Montmorillonite par exemple) ont une surface, accessible à la plupart des

molécules, variant de 40 à 800 m².g⁻¹. Leur capacité d'adsorption est très variable mais constitue le paramètre essentiel de la régulation des échanges et de la mobilité des éléments dans le milieu naturel [18].

Les adsorbants industriels (essentiellement les charbons actifs) développent des surfaces spécifiques énormes (de 600 à environ 1200 m².g⁻¹) caractéristique d'une très forte microporosité [19]. D'autres adsorbants comme les hydroxydes métalliques formés au cours de la coagulation-floculation développent eux aussi de très grandes surfaces dont l'extension est étroitement dépendante du pH [20]. Dans ce cas la capacité d'adsorption de la molécule à éliminer est proportionnelle à la surface spécifique de l'adsorbant. Plus la surface est importante, plus grande sera la quantité de molécules adsorbées. Il convient d'utiliser un adsorbant majoritairement microporeux (diamètre de pore < 2 nm) pour la plupart des applications en traitement de l'air afin d'obtenir une bonne adsorption.

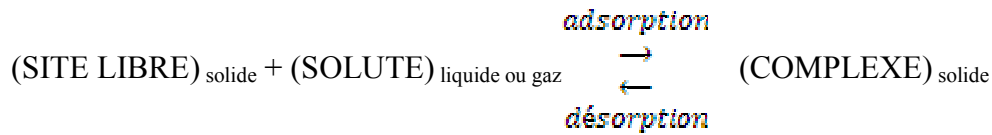
Les caractéristiques de la molécule et la structure de l'adsorbant peuvent affecter directement:

- ✓ La nature de la liaison adsorbat-adsorbant, c'est-à-dire de l'énergie libre d'interaction G entre les sites d'adsorption et la partie de la molécule en contact avec la surface ;
- ✓ La capacité d'échange cationique (CEC) : Il n'existe pas de règle générale qui lie la nature du cation à son pouvoir adsorbant mais il est à noter que l'adsorption dépend de l'électronégativité du cation compensateur ;
- ✓ D'autres facteurs, comme la température qui définit la nature de l'adsorption, peuvent également influencer la capacité d'adsorption [21, 22]. Du fait de l'exothermicité de la réaction d'adsorption, celle-ci est moins efficace lorsque la température augmente. Il est toujours intéressant d'effectuer l'opération d'adsorption à une température la plus basse possible, ce qui permet d'augmenter les capacités d'adsorption. Le travail à des températures inférieures à 40 °C est préconisé.

I.5. L'équilibre adsorption/désorption

I.5.1. Approche thermodynamique

La sorption d'un composé chimique (organique ou minéral) à la surface d'un solide (sol, charbon actif, zéolithe, oxyde métallique, résidus divers) est un équilibre entre une réaction d'adsorption et la réaction inverse de désorption, qui peut être schématisé par :



Par ailleurs, la sorption peut être caractérisée, par une constante d'équilibre standard (K^0), par des constantes de vitesse d'adsorption (k_{ads} ou k_1) et de désorption ($k_{\text{dés}}$ ou k_{-1}).

Comme toute réaction chimique, la variation d'enthalpie libre en cours de réaction (ΔG) est la différence entre l'enthalpie libre des produits (G_{complexe}) diminuée de la somme des enthalpies libres des réactifs ($G_{\text{site libre}} + G_{\text{soluté}}$), aux coefficients stœchiométriques près.

Elle peut être exprimée par la relation suivante :

$$\Delta G = \Delta G^0 + RT \cdot \ln (a_{\text{complexe}} / a_{\text{site libre}} \cdot a_{\text{soluté}})$$

ΔG^0 = variation d'enthalpie libre standard

$$R = 8,314 \text{ J} \cdot \text{mol} \cdot \text{K}^{-1}$$

T = température absolue (K)

a = activité ($\text{mol} \cdot \text{L}^{-1}$)

ΔG prend une valeur négative en début de réaction et devient nulle quand l'équilibre est atteint. ΔG^0 peut avoir l'expression suivante :

$$\Delta G^0 = - RT \cdot \ln K^0 \quad (\text{I.1})$$

Quand $\Delta G^0 < 0$ (ce qui implique $K^0 > 1$), la réaction est spontanée.

A l'équilibre, la variation d'enthalpie libre standard ΔG^0 est encore la différence (équation I.2) entre les énergies d'activation des réactions d'adsorption (E_1) et de désorption (E_{-1}). L'énergie d'activation est elle-même reliée à la constante de vitesse (k) de la réaction par la loi d'ARRHENIUS (équation I.3).

$$\Delta G^0 = E_1 - E_{-1} \quad (\text{I.2})$$

$$\text{et } k = D \cdot \exp\left(-\frac{E_a}{RT}\right) \quad (\text{I.3})$$

D, coefficient de fréquence de collision

Les équations (I.1), (I.2) et (I.3) permettent d'établir que :

$$\Delta G^0 = RT \ln \left(\frac{k_1}{k_{-1}} \right) \quad (\text{I.4})$$

et par la suite,

$$K^D = \frac{K_1}{K_{-1}} \quad (I.5)$$

I.5.2. Approche expérimentale

Au plan expérimental, il est très fréquent dans les études de réaction de sorption, de relier la concentration en complexes de surface formés à l'équilibre (quantité de soluté retenue par le solide, q_e , en mol.kg⁻¹ ou kg.kg⁻¹ ou unités dérivés) à la concentration en équilibre en soluté dans la phase liquide ou gazeuse (C_e , en mol.L⁻¹ ou kg.kg⁻¹ ou unités dérivées).

La relation obtenue est appelée « isotherme d'équilibre » à condition que l'expérience soit effectuée à une température constante.

$$q_e = \left(\frac{V}{m}\right) \cdot (C_0 - C_e) + q_0 \quad (I.6)$$

V: volume de solution,

m: masse de solide adsorbant,

C₀: concentration initiale en soluté en phase liquide ou gazeuse,

q₀: concentration en soluté initialement présente sur le solide (généralement nul ou négligeable).

Plusieurs auteurs, dont GILES et al. (1974), ont proposé une classification des isothermes de sorption basée sur leur forme et sur leur pente initiale [23]. Les revues de HINZ (2001) [25] et, plus récemment, de LIMOUSIN et al. (2007) présentent une synthèse plus ou moins détaillée de cette classification [24]. Parmi les formes d'isothermes, ces auteurs distinguent (Fig. I-1) :

- la forme H, dite de « haute affinité »,
- la forme L, dite de « Langmuir »,
- la forme C, dite de « partition constante »,
- la forme S, dite « sigmoïdale ».

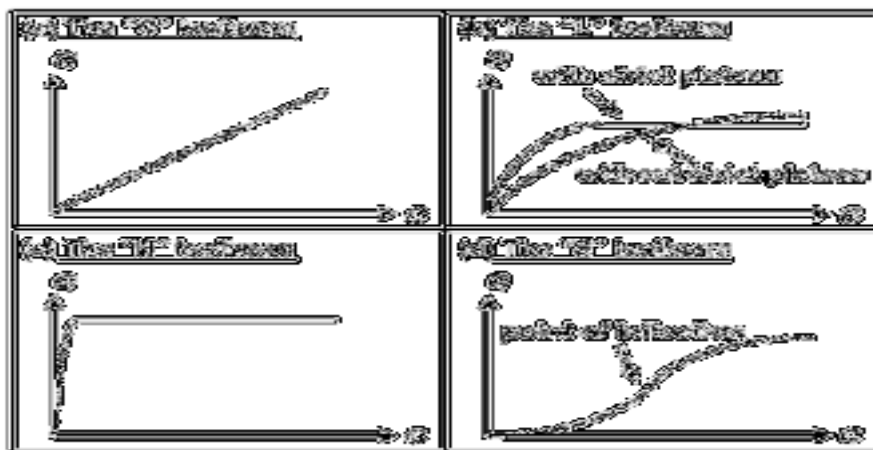


Figure I.1 : Les quatre types d'isothermes selon GILES et al. [23, 24]

Les formes « L » et « H » sont les plus observées, notamment dans le cas de l'adsorption de composés organiques en solution aqueuse sur des charbons actifs.

La forme « H » est un cas particulier de la forme « L », où la pente initiale est très élevée. C'est le cas quand le soluté exhibe une forte affinité pour le solide.

La forme « L » correspond plutôt aux faibles concentrations en soluté dans l'eau. L'isotherme est de forme convexe, ce qui suggère une saturation progressive du solide. Quand C_e tend vers zéro, la pente de l'isotherme est constante.

La forme « S » présente un point d'inflexion révélateur d'au moins deux mécanismes d'adsorption. C'est le cas, par exemple, quand une première couche de soluté est d'abord adsorbée puis quand l'adsorption d'une ou plusieurs couches supplémentaires devient favorisée.

La forme « C » est une droite passant par zéro ce qui signifie que le rapport q_e/C_e (appelé coefficient de distribution K_d) est constant. C'est souvent le cas de l'adsorption sur les argiles.

D'après HINZ (2001), une autre interprétation possible des données expérimentales est de tracer le coefficient de distribution $K_d = q_e/C_e$ en fonction de q_e , en échelle linéaire ou logarithmique, notamment quand les expérimentations ont été effectuées sur plusieurs ordres de grandeur de concentration.

Le tracé de K_d vs q_e exhibe :

- une pente négative presque constante (sauf pour les fortes valeurs de q_e) pour les isothermes de type « L »,

- une pente négative décroissante pour le type « H »,
- une horizontale pour le type « C »,
- une pente positive pour le type « S » à faibles concentrations.

Toujours d'après HINZ (2001), les modélisations dites de « Langmuir », « Langmuir – 2 sites », « Tóth » et « Redlich-Peterson » sont celles qui décriraient le mieux les données de l'isotherme de type « L », alors que le type « S » serait mieux décrit par les modélisations dites de « Freundlich », « Langmuir-Freundlich » et « Freundlich généralisé » [25].

Brunauer, Emmet et Teller ont classés les isothermes en 6 types : [26]

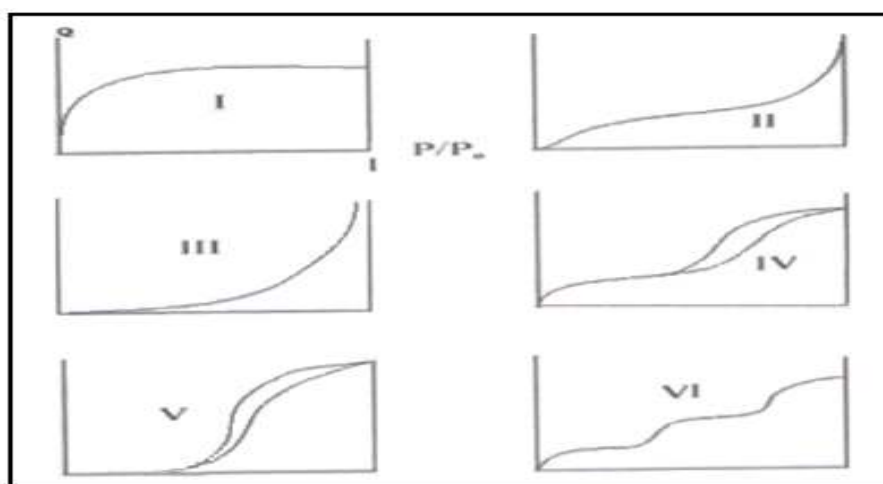


Figure I.2 : Différents types d'isothermes d'adsorption des gaz selon BET, [26]

Type I : Typique d'une adsorption en monocouche, ou correspondant au remplissage des micropores avec saturation lorsque le volume à disposition est rempli.

Type II : Il correspond en général à l'adsorption multicouche sur des surfaces ouvertes, cependant ce type d'isotherme peut résulter d'une somme d'isotherme de type I et de type III (remplissage de micropores suivi d'une adsorption multicouche sur une surface externe).

Type III : Il reflète un manque d'affinité entre l'adsorbât et l'adsorbant, et des interactions adsorbât-adsorbant relativement faibles.

Type IV : Il peut résulter de la combinaison d'une isotherme de type I (adsorption forte mais limitée) et de type V.

Type V : Il reflète aussi une forte interaction entre les adsorbats.

Type VI : Il présente des marches caractéristiques d'une adsorption multicouche sur une surface non poreuse très homogène.

I.5.3. Modélisation des isothermes d'équilibre mono-soluté

De très nombreuses publications sont parues depuis le début du 20^{ème} siècle et sont toujours produites. Il serait utopique d'en faire une étude bibliographique exhaustive. Seules quelques synthèses et travaux récents ont été pris en compte dans ce travail.

De nombreuses modélisations sont utilisées et parfois développées dans la littérature. La plupart de ces modèles sont présentés ci-après. On distingue :

- les modèles à deux paramètres (Freundlich, Langmuir, Temkin, Elovich et Dubinin-Radushkevich) ;
- les modèles à trois paramètres (Langmuir-Freundlich, Langmuir généralisé, Tóth, Sips, Redlich-Peterson, Fritz-Schluender, Radke-Prausnitz, Koble-Corridan, Jossens, Kislev, Hill-de Boer, Fowler Guggenheim);
- les modèles à quatre paramètres ou plus (par exemple dans le travail de HAMDAOUI et NAFFRECHOUX) [27];
- les modèles numériques de type régression linéaire multi-paramètres (MLR) ou non linéaire
- les réseaux de neurones artificiels (ANN) [28].

I.5.3.1. Modèles mono-soluté à deux paramètres

I.5.3.1.1. Modèle de FREUNDLICH

Le modèle simple et empirique de Freundlich est le plus communément utilisé [29, 30]. On considère qu'il s'applique à de nombreux cas, notamment dans le cas de l'adsorption multicouche avec possibles interactions entre les molécules adsorbées.

$$q_e = K_F \cdot C_e^n \quad (I.7)$$

La forme exploitée la plus courante est le tracé en échelle logarithmique des variations de q_e en fonction de C_e :

$$\log q_e = \log K_F + n \log C_e \quad (I.8)$$

Une autre exploitation possible des résultats par l'isotherme de Freundlich consiste à tracer en échelle logarithmique les variations du coefficient de distribution K_d en fonction de q_e :

$$\log K_d = \left(\frac{1}{n}\right) \log K_F + \left[\frac{(n-1)}{n}\right] (\log q_e) \quad (I.9)$$

K_F est une constante qui est relative à la capacité d'adsorption. Comme C_e est souvent exprimée en mg.L^{-1} et q_e en mg.g^{-1} , l'unité de K_F est $\text{mg}^{(1-n)}.\text{L}^n.\text{g}^{-1}$. En accord avec HALSEY [31], la relation entre K_F et la capacité maximum d'adsorption (q_m) est :

$$K_F = \frac{q_m}{C_e^n} \quad (I.10)$$

La constante n (adimensionnelle) donne une indication sur l'intensité de l'adsorption. Il est généralement admis que des faibles valeurs de n ($0,1 < n < 0,5$) sont caractéristiques d'une bonne adsorption, alors que des valeurs plus élevées révèlent une adsorption modérée ($0,5 < n < 1$) ou faible ($n > 1$) [32]. La constante « n » est très souvent remplacée par « $1/n$ » ou facteur d'hétérogénéité. Il faut noter que si n (ou $1/n$) tend vers 1, l'isotherme devient linéaire donc de type C.

I.5.3.1.2. Modèle de LANGMUIR

Le second modèle le plus communément utilisé est celui de Langmuir [33]. Les hypothèses initiales sont que le solide adsorbant présente une capacité d'adsorption limitée (q_m), que tous les sites actifs soient identiques, qu'ils ne peuvent complexer qu'une seule molécule de soluté (adsorption monocouche) et qu'il n'y a pas d'interactions entre les molécules adsorbées. Il s'agit dans ce cas d'une simple application de la loi d'action de masse conduisant à la constante thermodynamique d'équilibre K^0 définie ci-dessus (I.1).

$$K^0 = (a_{\text{complexes}})_{\text{éq.}} / (a_{\text{sites libres}})_{\text{éq.}} \cdot (a_{\text{soluté}})_{\text{éq.}}$$

En remplaçant les activités par des concentrations et K^0 par K_L , l'équation devient :

$$K_L = \frac{q_e}{C_e(q_m - q_e)}$$

$$\frac{q_e}{q_m} = \theta = \frac{K_L \cdot C_e}{(1 + K_L \cdot C_e)} \quad (I.11)$$

K_L , constante d'équilibre de Langmuir

θ , taux de recouvrement.

Des développements de l'équation 11 conduisent à des formes linéaires de l'isotherme de Langmuir. Parmi les cinq formes citées dans la littérature [32], deux d'entre elles sont très couramment utilisées :

$$\frac{1}{q_e} = \left(\frac{1}{q_m} \right) + \left(\frac{1}{K_L \cdot C_e \cdot q_m} \right) \quad \text{« Langmuir I »} \quad (\text{I.12})$$

$$\frac{C_e}{q_e} = \left(\frac{1}{K_L \cdot q_m} \right) + \left(\frac{C_e}{q_m} \right) \quad \text{« Langmuir II »} \quad (\text{I.13})$$

Quand q_e et q_m sont exprimées en mg.g^{-1} et C_e en mg.L^{-1} , la constante K_L est exprimée en L.mg^{-1} . Notons que K_L est souvent appelée « b » ou encore « L » et que certains auteurs définissent le rapport R_L (équation 14) comme une grandeur sans unité indiquant si l'adsorption d'autant plus favorable que R_L tend vers zéro ($R_L \rightarrow 0$) et d'autant plus défavorable que R_L tend vers un ($R_L \rightarrow 1$) HALL [34].

$$R_L = \frac{1}{1 + K_L \cdot C_e} \quad (\text{I.14})$$

I.5.3.1.3. Modèle de TEMKIN

Le modèle de Temkin repose sur l'hypothèse que, lors de l'adsorption en phase gazeuse, la chaleur d'adsorption due aux interactions avec l'adsorbant décroît linéairement avec le taux de recouvrement θ [35]. C'est une application de la relation de Gibbs pour les adsorbants dont la surface est considérée comme énergétiquement homogène [36].

Plusieurs auteurs proposent d'utiliser ce modèle en phase liquide, en traçant q_e ou θ en fonction de $\ln C_e$. [25, 37, 38, 24, 39, 32, 40]

$$\frac{q_e}{q_{\max}} = \theta = \left(\frac{RT}{\Delta Q} \right) \ln(K_T \cdot C_e) \quad (\text{I.15})$$

$$R = 8,314 \text{ J.mol}^{-1} \cdot \text{K}^{-1}$$

T, température absolue (en K)

ΔQ , variation d'énergie d'adsorption (en J.mol^{-1})

K_T , constante de Temkin (en L.mg^{-1})

Ce tracé permet de déterminer $B_T = \frac{q_m RT}{\Delta Q}$ puis en introduisant une valeur de q_m (par exemple issue de l'application de Langmuir), de calculer la variation d'énergie d'adsorption ΔQ .

I.5.3.1.4. Modèle d'ELOVICH

Le modèle d'Elovich est basé sur un développement cinétique faisant l'hypothèse que les sites d'adsorption augmentent exponentiellement avec l'adsorption, ce qui implique une adsorption multi-couches [41]. Quand il est utilisé (ex : HAMDAOUI et NAFFRECHOUX) [32], il est exprimé par la relation :

$$\frac{q_e}{q_m} = \theta = K_E \cdot C_E \cdot \exp\left(-\frac{q_e}{q_m}\right) \quad (I.16)$$

K_E , constante d'Elovich (en $L \cdot mg^{-1}$)

Le tracé de $\ln\left(\frac{q_e}{C_E}\right)$ vs q_e permet d'atteindre q_m et K_E .

$$\ln\left(\frac{q_e}{C_E}\right) = -\left(\frac{q_e}{q_m}\right) + \ln(K_E q_m) \quad (I.17)$$

I.5.3.1.5. Modèle de DUBININ-RADUSHKEVICH

Le modèle de Dubinin-Radushkevich (cité par KENNEDY et al.) [42] ne fait pas l'hypothèse d'une surface homogène ou d'un potentiel d'adsorption constant, comme le modèle de Langmuir. Sa théorie de remplissage du volume des micropores repose sur le fait que le potentiel d'adsorption est variable et que l'enthalpie libre d'adsorption est reliée au degré de remplissage des pores. L'isotherme de Dubinin-Radushkevich est donnée par l'équation :

$$\frac{q_e}{q_{mDR}} = \exp(-\beta \varepsilon^2) \quad (I.18)$$

q_{mDR} , capacité maximum d'adsorption dans les micropores

β , constante reliée à l'énergie d'adsorption par $E = \frac{1}{(\beta)^{0.5}}$

ε , potentiel de Polanyi = $RT \ln\left(\frac{C_E}{C_s}\right)$

C_s , solubilité du soluté.

L'équation de Dubinin-Radushkevich devient alors :

$$\ln q_e = \ln q_{mDR} - \left(\frac{RT}{E}\right)^2 \cdot \left(\ln\left(\frac{C_e}{C_0}\right)\right)^2 \quad (I.19)$$

Le tracé de $\ln q_e$ vs $\ln\left(\frac{C_e}{C_0}\right)^2$ permet d'atteindre q_{mDR} et E .

I.5.3.2. Modèles mono-soluté à trois paramètres

Vu que la plupart de ces modèles étant dérivés des modèles simples de Langmuir et de Freundlich, il a été choisi, dans ce travail bibliographique, de symboliser les trois paramètres par « q_m », « K_L » et « n » pour simplifier la comparaison entre les différents modèles. Certaines revues de synthèse l'ont d'ailleurs également fait [25, 24].

I.5.3.2.1. Modèle de REDLICH-PETERSON

C'est le modèle monosoluté à trois paramètres qui est le plus cité et le plus utilisé dans la littérature parce que annoncé comme pouvant s'appliquer sur une large gamme de concentration [25,43, 44, 45, 45, 24, 27, 46, 47,49 ,48]. C'est un modèle empirique combinant les paramètres des équations de Langmuir et de Freundlich.

Dans la publication initiale [49], le modèle est appliqué à l'adsorption en phase gazeuse. Par analogie, son expression en phase liquide est de la forme [45, 47, 46, 27]:

$$q_e = \frac{A C_e}{[1 + K_L (C_e)^n]} \quad (I.20)$$

A et K_L , constantes de Redlich-Peterson,

D'après HINZ et LIMOUSIN et al. [25 et 24], ce modèle peut s'écrire encore :

$$\frac{q_e}{q_m} = \theta = \frac{K_L C_e}{1 + (K_L C_e)^n} \quad (I.21)$$

K_L , constante de Langmuir (en $L \cdot mg^{-1}$) correspondant à A/q_m du modèle précédent
 n , coefficient (de type "n" de FREUNDLICH) compris entre 0 et 1, correspondant à β

$$K = K_L^n \text{ (ou } \beta)$$

Une linéarisation possible est :

$$\frac{C_e}{q_e} = \frac{1}{K_L q_m} + \frac{K_L^{(n-1)}}{q_m} (C_e)^n \quad (I.22)$$

Une autre forme possible de linéarisation est d'introduire les valeurs de K_L et de q_m de Langmuir et de tracer :

$$\ln \left[\left(\frac{q_m K_L C_e}{q_e} - 1 \right) \right] = n \cdot \ln (K_L \cdot C_e)$$

$$\ln \left[\left(\frac{q_m K_L C_e}{q_e} - 1 \right) \right] = n \cdot \ln (C_e) + \ln K_L$$

On peut noter que dans certaines conditions, notamment pour les fortes concentrations de soluté en phase liquide, l'expression de Redlich-Peterson devient comparable à celle de Freundlich.

$$(K_L C_e)^n \gg 1$$

$$q_e = q_m \cdot K_L (C_e)^{(1-n)}$$

Si $n = 0$, l'isotherme devient linéaire (de type C).

I.5.3.2.2. Modèle de TÓTH

C'est également un modèle très souvent cité et utilisé [25, 38, 24, 27, 50, 51, 47, 49, 48].

Le modèle de Tóth a été établi pour l'adsorption en phase gazeuse (TÓTH) [52] à partir de l'isotherme de Langmuir, mais en considérant que la surface de l'adsorbant n'est pas énergétiquement homogène. Ce modèle présente donc un intérêt particulier puisqu'il considère que la surface de l'adsorbant est hétérogène. En phase liquide, il est généralement utilisé comme une adaptation du modèle de Langmuir, proche du modèle empirique de Redlich-Peterson (I.18).

$$\frac{q_e}{q_m} = \theta = \frac{K_L C_e}{[1 + (K_L C_e)^n]^{\frac{1}{n}}} \quad (I.23)$$

D'après RUDZINSKI et PLAZINSKI [51], une linéarisation possible de l'isotherme de Tóth est :

$$\left(\frac{C_e}{q_e} \right)^n = \left(\frac{1}{q_m K_L} \right)^n + \left(\frac{1}{q_m} \right)^n \cdot (C_e)^n \quad (I.24)$$

Le tracé de $(C_e/q_e)^n$ vs $(C_e)^n$ doit être linéaire si le modèle est vérifié.

I.5.3.2.3. Modèle de LANGMUIR-FREUNDLICH

$$\frac{q_e}{q_m} = \Theta = \frac{(K_L C_e)^n}{[1+(K_L C_e)^n]} \quad (I.25)$$

$$\text{Ou } \frac{q_m}{q_e} = \frac{1}{(K_L C_e)^n} + 1 \quad (I.26)$$

A noter que ce modèle est cité par LIMOUSIN et al. [25] sous une forme différente:

$$\frac{q_e}{q_m} = \Theta = \frac{K_L C_e^n}{1+(K_L C_e)^n} \quad (I.25 \text{ bis})$$

$$\text{Ou bien, } \frac{q_m}{q_e} = \frac{1}{K_L C_e^n} + \frac{1}{K_L} \quad (I.26 \text{ bis})$$

I.5.3.2.4. Modèle GENERALISE

Trois appellations sont utilisées pour ce modèle à trois paramètres : « Généralisé » (ex. : GIMBERT et al.) [40], « Freundlich généralisé » (ex. : HINZ) [25] et « Langmuir généralisé » (ex. : LIMOUSIN et al.) [24]. Son expression est de la forme :

$$\frac{q_e}{q_m} = \Theta = \left[\frac{K_L C_e}{1+K_L C_e} \right]^n \quad (I.27)$$

$$\text{Ou } \left(\frac{q_m}{q_e} \right)^{1/n} = \left(\frac{1}{K_L C_e} \right) + 1 \quad (I.28)$$

I.5.3.2.5. Modèle de SIPS ou KOBLE-CORRIGAN

Le modèle de Sips [53] est quant à lui peu cité [54, 27], parfois sous l'appellation « Koble-Corrigan » [45]. Il est de la forme suivante et généralement mieux validé quand $n > 1$:

$$\frac{q_e}{q_m} = \Theta = \frac{K_L C_e^n}{(1+K_L C_e^n)} \quad (I.29)$$

$$\text{Ou bien, } \frac{q_e}{q_m} = \left(\frac{1}{K_L C_e^n} \right) + 1 \quad (I.30)$$

Il est possible de retrouver son expression en appliquant la loi d'action de masse, comme pour l'expression de Langmuir dont il est issu, mais en considérant que n molécules de soluté sont adsorbées par site, ce qui revient à écrire que la stœchiométrie de la réaction d'adsorption est de « n » molécules de soluté pour un site libre sur l'adsorbant. Cette interprétation permet de comprendre pourquoi la valeur de n est plutôt supérieure à 1.

I.5.3.2.6. Modèles de RADKE-PRAUSNITZ

D'après HAMDAOUI et NAFFRECHOUX [27], ces modèles sont de forme comparable à ceux cités précédemment [55]. Si on assimile les constantes de Radke-Prausnitz à q_m , K_L et n , afin de permettre une meilleure comparaison avec les autres modèles, les expressions citées par ces auteurs sont :

$$\text{Modèle R-P 1} \quad \frac{q_e}{q_m} = \theta = \frac{K_L C_e}{(1 + K_L C_e)^n} \quad (\text{I.31})$$

$$\text{Modèle R-P 2} \quad \frac{q_e}{q_m} = \theta = \frac{K_L C_e}{1 + K_L C_e^n} \quad (\text{I.32})$$

$$\text{Modèle R-P 3} \quad \frac{q_e}{q_m} = \theta = \frac{K_L C_e^n}{1 + K_L C_e^{(n-1)}} \quad (\text{I.33})$$

La forme 1 est la seule citée par d'autres auteurs (ex. : KUMAR et PORKODI) [56].

I.5.3.2.7. Autres modèles à trois paramètres

Cités également par HAMDAOUI et NAFFRECHOUX [27], plusieurs autres modèles à 3 paramètres sont parfois utilisés pour valider les résultats des isothermes. Il s'agit des modèles :

- Fritz et Schluender [57]

$$\frac{q_e}{q_m} = \theta = \frac{K_L C_e}{1 + q_m C_e^n} \quad (\text{I.34})$$

- Jossens [58]

$$C_e = \left(\frac{q_e}{H} \right) \exp (F \cdot q_e^p) \quad (\text{I.35})$$

Avec H , F et p , constantes de Jossens

- Kiselev [59]

$$K_L C_e = \frac{\theta}{(1-\theta)(1+K_L \theta)} \quad (I.36)$$

- Hill-de Boer [60]

$$K_{FG} C_e = \left(\frac{\theta}{1-\theta} \right) \exp\left(\frac{\theta}{1-\theta} - \frac{K_2 \theta}{RT} \right) \quad (I.37)$$

- Fowler-Guggenheim [61]

$$K_{FG} C_e = \frac{\theta}{(1-\theta)^2} \exp\left(\frac{2\theta W}{RT} \right) \quad (I.38)$$

I.5.3.3. Modèles mono-soluté à plus de trois paramètres

D'autres modèles à plus de trois paramètres peuvent être trouvés dans la littérature (ex. HAMDAOUI et NAFFRECHOUX) [27]. Un seul est présenté ici, il s'agit d'une variante du modèle de Langmuir, appelée « Langmuir-2 sites ».

Comme l'expression simple de Langmuir, ce modèle repose également sur la loi d'action de masse avec deux types de sites de réactivité différente.

En effet, si les deux types de sites libres (site A et site B) sont chacun caractérisés par des constantes d'équilibre de Langmuir k_{LA} et k_{LB} et des fractions présentes f_A et f_B , l'application de la loi d'action de masse (avec $q_{eA} + q_{eB} = q_e$) conduit à l'expression suivante :

$$\frac{q_e}{q_m} = \theta = \left[\frac{f_A K_{LA} C_e}{(1+K_{LA} C_e)} + \frac{f_B K_{LB} C_e}{(1+K_{LB} C_e)} \right] \quad (I.39)$$

Cette expression est citée par certains auteurs (HINZ, LIMOUSIN et al.) [25, 24].

I.6. Cinétique d'adsorption

Toutes les études d'adsorption et de désorption de solutés reposent sur la détermination des concentrations des substances étudiées dans une phase liquide. Or l'interprétation dépend de la nature du système étudié. Une représentation simple peut se faire par trois systèmes modèles.

a- Système 1

Il correspond à une suspension d'un matériau solide adsorbant non poreux dans une solution contenant un soluté. Si la solution est agitée mécaniquement, on peut admettre que les molécules de soluté se trouvent très rapidement au voisinage de la surface à proximité des

sites d'adsorption. La vitesse de la réaction d'adsorption est déterminée par celle avec laquelle les molécules atteignent la surface par diffusion moléculaire à travers le film liquide qui entoure les particules. Dans ce cas, elle est probablement très grande, du moins pour une adsorption de type physique. Pour une adsorption de type chimique, l'existence d'une énergie d'activation peut être à l'origine de vitesse de réaction plus petite.

Dans ce système l'adsorption est très rapide et peut être considérée comme instantanée. Cette situation est rencontrée avec des suspensions de particules de kaolinite, d'oxydes métalliques bien cristallisés et de tout matériau non poreux qui ne possède que des surfaces externes.

b- Système 2

Il représente aussi une suspension mais avec des particules poreuses. Même avec une agitation mécanique, les surfaces adsorbantes ne sont pas également toutes accessibles. Les surfaces situées à la périphérie des particules le sont directement comme le cas du système 1, contrairement aux surfaces des parois des pores qui ne sont accessibles que par la diffusion des molécules. Dans ce cas, Il peut être constaté que la vitesse avec laquelle la concentration de la solution environnant les particules varie, dépend des vitesses de deux phénomènes: l'adsorption proprement dite et la diffusion dans les pores intra particulaires.

Cette situation assez fréquente est rencontrée avec les particules de minéraux argileux constituées de quasi cristaux et possédant des espaces inter foliaires accessibles, à tous les agrégats du sol, petits ou grands, mais aussi avec les particules organiques et en particulier les agrégats moléculaires de substances humiques.

c- Système 3

Ce modèle est un exemple des milieux poreux structurés où la situation est très complexe car il existe schématiquement deux phases liquides, l'une immobile dans les micropores et l'autre mobile dans les mésopores et macropores. La solution qui est au contact de la phase liquide immobile a une concentration qui est déterminée par le processus de convection-dispersion dans la phase mobile. Par conséquent, la vitesse d'adsorption sur les parois des micropores est déterminée, non seulement par la vitesse du phénomène d'adsorption mais aussi par la diffusion dans les micropores et par la convection-dispersion dans les méso et macropores.

Il est donc difficile de déduire des informations sur les vitesses d'adsorption à partir des vitesses de variation de la concentration de la phase liquide mobile qui est seule, accessible à la mesure.

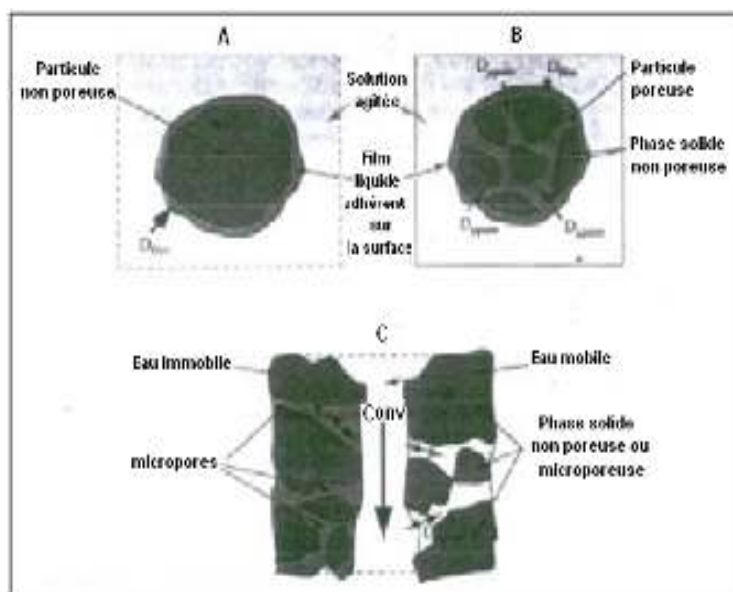


Figure I.3 : différents systèmes de rétention

A: particule non poreuse, B: particules poreuses, C: milieux poreux structuré

I.7. Le césium dans les sols

Le césium est le plus rare des métaux alcalins présents dans l'écorce terrestre. L'isotope stable, ^{133}Cs , existe naturellement dans deux minerais : la pollucite tectosilicate de formule $(\text{Cs}, \text{Na}) \text{AlSi}_2\text{O}_6 \cdot n\text{H}_2\text{O}$, ou la rhodizite ; les concentrations du césium varient de 0.3 à 25 $\mu\text{g/g}^{-1}$ dans les sols (tableau I. 1). Dans les végétaux terrestres, les concentrations présentent une amplitude très large allant de moins de $8.8 \cdot 10^{-5}$ à 89 $\mu\text{g/g}^{-1}\text{sec}$ [62, 63].

Tableau I.1 : Concentration du césium dans les roches et les sols en $\mu\text{g/g}^{-1}\text{sec}$, selon Coughtrey P.J. et al. [62]

Matériaux	Concentrations ($\mu\text{g/g}$)
Roches magmatiques (écorce terrestre)	2.7 à 3
Roches ignées ; granites	1
Roches sédimentaires ; schistes	4 à 5
Sables ; calcaires	0.5
Sols (général)	5 (0.3 à 25)
Sols français	2 à 3 (valeur IPSN)

La présence des radio-césiums dans l'environnement est liée à l'activité humaine avec trois sources principales :

- Les rejets contrôlés des installations nucléaires (réacteurs, usines de retraitement) qui sont en majorité des rejets liquides, d'importance limitée sur l'environnement terrestre (si ce n'est qu'au niveau local).
- Les tirs d'armes nucléaires : on considère qu'environ $1.3 \cdot 10^{18}$ Bq de ^{137}Cs ont été relâchés dans l'atmosphère depuis 1945.

Après apport en surface du sol par les eaux d'irrigation, les dépôts atmosphériques secs ou humides, la décomposition de végétaux contaminés, le césium est soumis à un ensemble de processus qui vont déterminer sa mobilité dans le sol (fixation et migration) et sa biodisponibilité (aptitude à être adsorbé par les racines des plantes).

De part sa structure physique, sa composition chimique et ses composantes biologiques, le sol, plus que tout autre facteur environnemental, va jouer un rôle récepteur actif sur le devenir du césium dans les écosystèmes terrestres.

I.7.1. Mécanismes de la rétention du césium par les sols

Il s'agit des processus physiques et chimiques qui se déroulent à l'interface sol-solution du sol et qui déterminent l'équilibre de sorption-désorption d'une substance chimique dans un sol. Cet équilibre est la résultante de plusieurs phénomènes dont les plus importants sont l'adsorption physique, l'adsorption chimique et, dans le cas du césium, l'échange d'ions.

- Adsorption physique consiste en une rétention, sur la surface des particules minérales, de l'élément présent dans la solution du sol. Ce mode de fixation qui met en jeu des liaisons faibles (forces de Van der Waals) est réversible. C'est un phénomène qui survient de façon pratiquement instantanée.

- Adsorption chimique se réalise par l'intermédiaire de liaisons ioniques ou covalentes. La cinétique de sa mise en place est lente. Mais de part la force des types de liaisons mis en jeu, ce mode de fixation est fort et parfois irréversible.

- Echange d'ions est un mécanisme qui prime comme mode de rétention du césium par les sols [64, 65]. Les ions Cs^+ présents dans la solution du sol vont s'échanger avec des cations sorbes, en particulier sur les fractions les plus fines du sol qui sont constituées d'éléments chargés électriquement et qui ont une importante surface spécifique.

Les principaux constituants du sol impliqués dans ces phénomènes d'échange d'ions sont les minéraux argileux, certains composés organiques ainsi que les oxydes et les hydroxydes de silice, d'aluminium et de fer. Ces composants entrent dans la constitution du complexe absorbant qui joue un rôle régulateur de la composition ionique de la solution du sol en libérant ou en stockant des éléments.

I.7.2. Migration du césium dans les sols

Les transferts du césium dans les sols sont essentiellement liés aux mouvements de l'eau du sol ; ces transferts sont engendrés par :

- Le déplacement de la solution dans laquelle l'élément est soit sous forme ionique (phénomène de lixiviation), soit parfois sous forme colloïdale (complexes organique ou organométallique) ;
- Le déplacement de particules minérales, organiques ou organominérale du sol sur lesquelles le césium peut être retenu (phénomène de lessivage).

Du fait que le césium est rapidement et très fortement retenu par les sols de part la capacité des minéraux argileux à le fixer de façon énergique, sa vitesse de migration en profondeur est très faible. La vitesse de migration du césium est déterminée en grande partie par les propriétés physiques (perméabilité et capacité de rétention en eau) et chimiques du sol (pH et nature du complexe absorbant).

II.1 Introduction

Un réseau de neurones (ou Artificial Neural Network en anglais) est un modèle de calcul utilisé en intelligence artificielle, et dont la conception est schématiquement inspirée du fonctionnement de vrais neurones (humains ou non), pour prendre des décisions s'appuyant davantage sur la perception que sur le raisonnement. Le cerveau humain contient environ 100 milliards de neurones. Ces neurones nous permettent entre autre, de lire un texte tout en maintenant une respiration régulière permettant d'oxygéner notre sang, en actionnant notre cœur qui assure une circulation efficace de ce sang pour nourrir nos cellules, etc. Ils nous permettent même de comprendre certaines idées [66]. Chacun de ces neurones est par ailleurs fort complexe. Essentiellement, il s'agit de tissu vivant et de chimie. Les neurophysiciens commencent à peine à comprendre quelques uns de leurs mécanismes internes. On croit en général que leur différentes fonctions neuronales, y compris celle de la mémoire sont stockées au niveau des connexions (synapses) entre les neurones [67].

En voulant prendre sa source d'inspiration dans le monde de la biologie, en n'en gardant que les principes directeurs, les chercheurs ont inventé un grand nombre d'algorithmes différents. Si aujourd'hui, le modèle de neurone que l'ingénieur considérait comme biologiquement "plausible" est devenu très approximatif pour le biologiste, la biologie reste pourtant une source d'inspiration inépuisable, ainsi qu'en témoigne l'importante littérature qui lui est dédiée. Si l'on admet qu'un réseau de neurones multicouches comportant un nombre de neurones et de couches cachées suffisant peut approximer n'importe quelle fonction continue avec une précision arbitraire, il est rare que ce nombre soit compatible avec la quantité de données disponibles dans le cas d'une application réelle. C'est en essayant de s'affranchir des contraintes pesant sur les données que l'on a assisté au développement d'un grand nombre de méthodes baptisées "neuronales" [68]. La dimension des données influence également le type de solution envisagée pour un problème. Parce que les problèmes de classification sont d'autant plus difficiles que la dimension de l'espace est grande (du fait d'une lacunarité statistiquement croissante), le comportement de différents types d'algorithmes pourra être fort différent d'un cas à l'autre [69].

L'approche neuronale, très en vogue, sert aujourd'hui à toutes sortes d'application dans divers domaines. Par exemple, un autopilote pour avion, ou encore un système de guidage pour automobile, systèmes de lecture automatique de chèques bancaires et d'adresses postales, 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 la diagnostic médical, pour l'exploration pétrolière ou gazière, en robotique, en télécommunication, et bien d'autres. 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 surtout dans le domaine nucléaire à savoir leur intégration dans les systèmes de surveillance et diagnostique des réacteurs de puissance pour la prévention des accidents graves [70, 71, 72, 73, 74].

Nous aborderons ici les principales architectures des réseaux de neurones. Il ne s'agit pas de les étudier toutes, car elles sont trop nombreuses, mais plutôt d'en comprendre les mécanismes internes fondamentaux et de savoir comment et quand les utiliser.

II.2. Historique

Les recherches menées dans le domaine du connexionnisme ont démarré avec la présentation en 1943 par W. MCCulloch et W. Pitts d'un modèle simplifié de neurone biologique communément appelé neurone formel [75]. Ils montrèrent également théoriquement que des réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes.

En 1949, D. Hebb initie, dans son ouvrage "The Organization of Behavior", la notion d'apprentissage [76]. Deux neurones entrant en activité simultanément vont être associés (c'est-à-dire que leurs contacts synaptiques vont être renforcés). On parle de loi de Hebb et d'associationnisme.

En 1958, F. Rosenblatt développe le modèle du Perceptron [77]. C'est un réseau de neurones inspiré du système visuel. Il possède deux couches de neurones : une couche de perception (sert à recueillir les entrées) et une couche de décision. C'est le premier modèle pour lequel un processus d'apprentissage a pu être défini.

S'inspirant du perceptron, Widrow et Hoff, développent, dans la même période, le modèle de l'Adaline (Adaptive Linear Element) [78]. Ce dernier sera, par la suite, le modèle de base des réseaux de neurones multi-couches.

En 1969, Les recherches sur les réseaux de neurones ont été pratiquement abandonnées lorsque M. Minsky et S. Papert ont publié leur livre « Perceptrons » (1969) et démontré les

limites théoriques du perceptron, en particulier, l'impossibilité de traiter les problèmes non linéaires par ce modèle [79].

En 1982, Hopfield développe un modèle qui utilise des réseaux totalement connectés basés sur la règle de Hebb pour définir les notions d'attracteurs et de mémoire associative [80].

En 1984 c'est la découverte des cartes de Kohonen avec un algorithme non supervisé basé sur l'auto-organisation et suivi une année plus tard par la machine de Boltzman (1985) [81].

Une révolution survient alors dans le domaine des réseaux de neurones artificiels : une nouvelle génération de réseaux de neurones, capables de traiter avec succès des phénomènes non-linéaires : le perceptron multicouche ne possède pas les défauts mis en évidence par Minsky [79]. Proposé pour la première fois par Werbos, le Perceptron Multi-Couche apparaît en 1986 introduit par Rumelhart [82], et, simultanément, sous une appellation voisine, chez Le Cun (1985) [83]. Ces systèmes reposent sur la rétropropagation du gradient de l'erreur dans des systèmes à plusieurs couches, chacune de type Adaline de Bernard Widrow [78], proche du Perceptron de Rumelhart [82].

De nos jours, l'utilisation des réseaux de neurones dans divers domaines ne cesse de croître. Les applications en sont multiples et variées.

II.3. Neurone biologique

Les cellules nerveuses, appelées neurones, sont les éléments de base du système nerveux central. Les neurones possèdent de nombreux points communs dans leur organisation générale et leur système biologique, avec les autres cellules. Ils présentent cependant des caractéristiques qui leur sont propres et se trouvent dans cinq fonctions spécialisées qu'ils assurent [84] :

- Recevoir des signaux en provenance de neurones voisins,
- Intégrer ces signaux,
- Engendrer un flux nerveux,
- Conduire le flux nerveux et le transmettre à un autre neurone capable de le recevoir.

Donc le neurone biologique est une cellule vivante spécialisée dans le traitement des signaux électriques. Les neurones sont reliés entre eux par des liaisons appelées axones. Ces axones vont eux-mêmes jouer un rôle important dans le comportement logique de l'ensemble.

Ces axones conduisent les signaux électriques de la sortie d'un neurone vers l'entrée (synapse) d'un autre neurone.

Les neurones font une sommation des signaux reçus en entrée et en fonction du résultat obtenu vont fournir un courant en sortie. La structure d'un neurone se compose de trois parties (Figure II.1) :

- Le corps cellulaire : ou encore la somma, est la cellule d'activité nerveuse, au centre du neurone.
- L'axone : attaché au somma qui est électriquement actif, ce dernier conduit l'impulsion conduite par le neurone.
- Dendrites : électriquement passives, elles reçoivent les impulsions d'autres neurones.

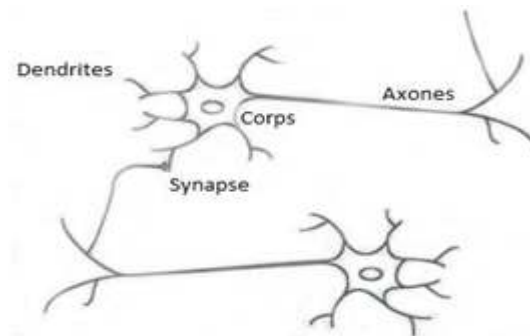


Figure II.1 : Schématisation d'un neurone biologique [68]

II.4. Neurone formel

La modélisation consiste à mettre en œuvre un système de réseaux neuronaux sous un aspect non pas biologique mais artificiel, cela suppose que d'après le principe biologique on aura une correspondance pour chaque élément composant le neurone biologique, donc une modélisation pour chacun d'entre eux (Figure II.2).



Figure II.2 : Mise en correspondance neurone biologique / neurone artificiel [69]

Les réseaux de neurones formels sont à l'origine d'une tentative de modélisation mathématique du cerveau humain. Le premier neurone formel a été modélisé par W. McCulloch et W. Pitts, en 1943 [75, 85]. Un neurone formel ou artificiel est un opérateur mathématique très simple. Un neurone possède des entrées qui peuvent être les sorties d'autres neurones, ou des entrées de signaux extérieures, et une sortie. La valeur de la sortie résulte du calcul de la somme des entrées pondérées par des coefficients (dits poids de connexions ou poids synaptiques) et du calcul d'une fonction non linéaire (dite fonction d'activation) de cette somme pondérée (Figure II.3). Un neurone formel peut être défini par les cinq éléments suivants [85] :

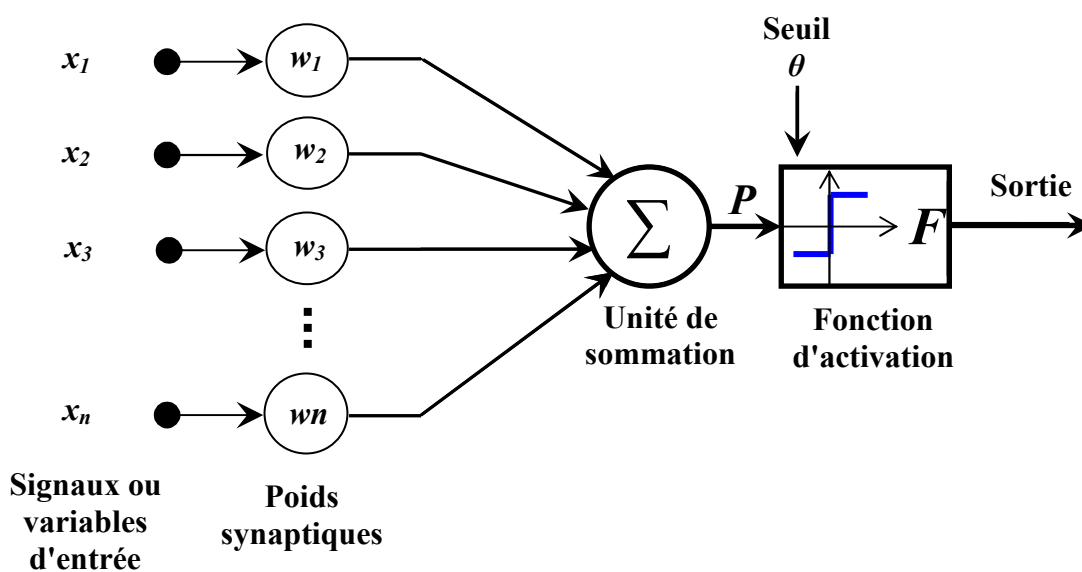


Figure II.3 : Neurone formel [86]

II.4.1. Entrées

Typiquement, les valeurs d'entrée sont des stimuli externes en provenance de l'environnement ou des sorties d'autres neurones artificiels. Elles peuvent être discrétisées grâce à un ensemble tel que $\{0,1\}$ ou bien être des valeurs réelles [87].

II.4.2. Poids synaptiques

Les paramètres w_n (figure II.3) sont souvent désignés sous le nom de «poids» ou, en raison de l'inspiration biologique des réseaux de neurones, «poids synaptiques». Ils sont des valeurs réelles qui déterminent la contribution de chaque entrée. Un algorithme d'apprentissage permet de déterminer le meilleur ensemble de poids pour le problème

considéré. Trouver l'ensemble optimal est souvent un compromis entre temps de calcul, minimisation de l'erreur et conservation de la capacité de généralisation pour le réseau.

La première chose que fait un neurone formel est de calculer la somme pondérée de ses entrées. Conformément à l'usage (également inspiré par la biologie), cette somme pondérée est appelée potentiel, à laquelle s'ajoute souvent un terme constant ou «biais» :

$$P = b + \sum_{i=1}^{n-1} w_i x_i \quad (\text{II.1})$$

II.4.3. Seuil

Le seuil est un nombre réel qui est soustrait de la somme pondérée des valeurs d'entrée. Pour des raisons de simplicité, le seuil peut être vu comme une entrée supplémentaire associée à un poids, où ($w = \theta$) et ($x = -1$), tel qu'il est illustré par la figure suivante :

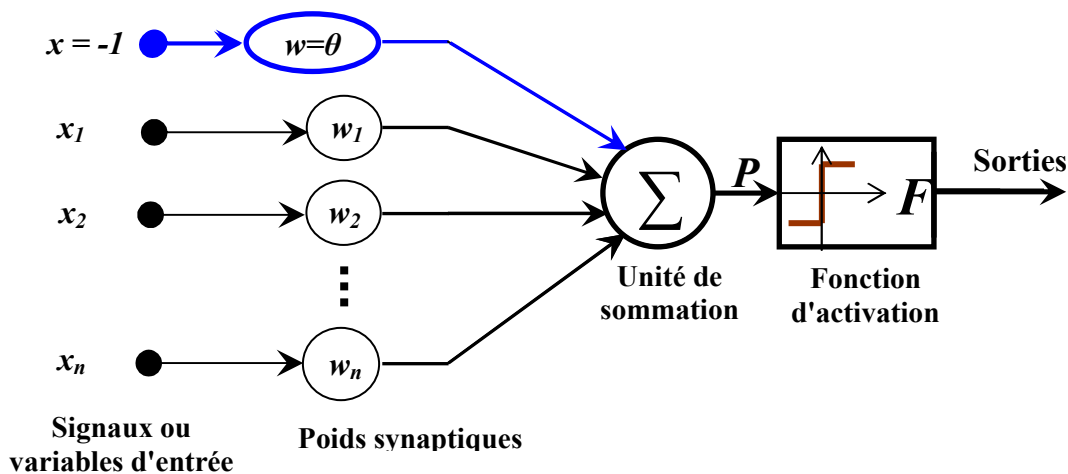


Figure II.4 : Neurone formel modèle étendu [86]

II.4.4. Fonction d'activation

La fonction de transfert ou d'activation 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. A 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 $[-1, +1]$.

Nous décrivons en ce qui suit, les formes les plus courantes des fonctions de transfert [85] :

II.4.4.1. Fonction identité

La fonction identité dite aussi linéaire dont l'expression est la suivante ;

$$S = F\left(P = \sum_{i=1}^n w_i x_i\right) \quad (\text{II.2})$$

Admet des valeurs non bornées pour les états, ce qui peut entraîner des débordements lors des simulations. Elle est employée souvent pour les neurones de sortie dans les réseaux multicouches. *Kohonen* l'a également utilisé pour construire son modèle de mémoires associatives [82].

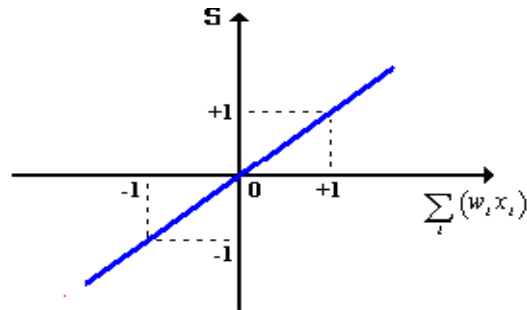


Figure II.5 : Fonction identité

II.4.4.2. Fonction linéaire à seuil

Cette fonction, qui est aussi appelée linéaire par morceaux, représente un compromis entre la fonction linéaire et la fonction seuil : entre ses deux bornes de saturation, elle confère au neurone une gamme de réponses possibles. En modulant la pente de linéarité, on affecte la plage de réponses du neurone (figure II.6).

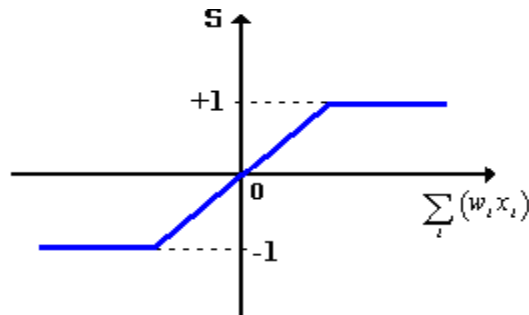


Figure II.6 : Fonction linéaire à seuil

II.4.4.3. Fonction non linéaire

II.4.4.3.1. Fonction sigmoïde logarithmique

La fonction sigmoïde est l'équivalent continu de la fonction linéaire bornée. Vu ses nombreux avantages (continuité et différentiabilité), ce type de fonction est généralement employé dans le réseau de neurones multicouches (figure II.7). L'équation qui définit la sigmoïde logarithmique est :

$$F(P) = \frac{1}{1 + e^{-P}} \quad (\text{II.3})$$

Pour $0 < F(P) < +1$

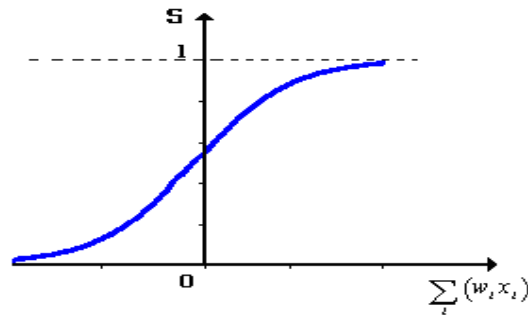


Figure II.7 : Fonction sigmoïde logarithmique

II.4.4.3.2. Fonction sigmoïde tangentielle

La fonction sigmoïde tangentielle ou tangente hyperbolique est fort utilisée dans les réseaux multicouches, surtout dans les neurones de la couche cachée (figure II.8). L'équation qui définit l'état du neurone est :

$$F(P) = \frac{1 - e^{-2P}}{1 + e^{-2P}} \quad (\text{II.4})$$

Pour $-1 < F(P) < +1$

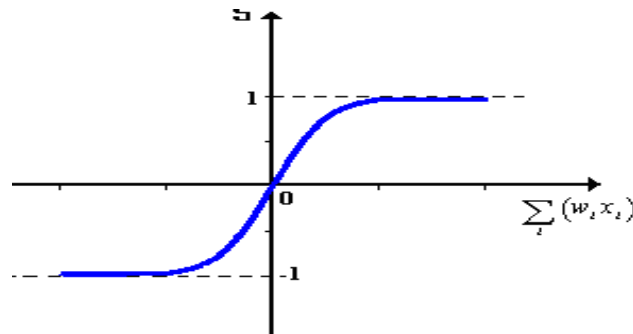


Figure II.8 : Fonction sigmoïde tangentielle

II.4.4.4. Fonction Gaussienne

Cette fonction n'est pas monotone croissante, elle est souvent employé dans les réseaux RBF (Radial Basis Function).

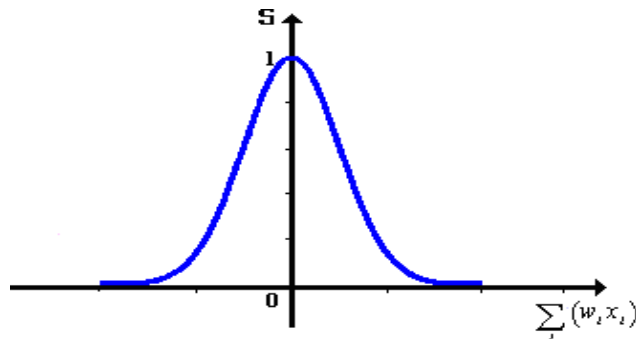


Figure II.9 : Fonction Gaussienne

Dans une étude comparative de fonctions d'activation [88], la fonction sigmoïde s'est révélée meilleure en termes de capacité d'approximation. *H.N.Mhaskar* et *C.A.Micchelli* [89] ont établi une relation entre la précision de l'approximation, le nombre d'unités cachées dans un réseau à une seule couche cachée et la régularité de la fonction d'activation. D'après leur théorème, plus la fonction d'activation est régulière, plus la précision de l'approximation est bonne. Cependant, *Fombellida* et al. [90] ont constaté que l'utilisation de fonctions d'activation non monotones peut présenter certains avantages : accélération de l'apprentissage, réduction du nombre d'unités cachées, etc.

II.4.5. Sortie du neurone "S"

La sortie du neurone ni d'autre que l'application de la fonction de transfert à la somme pondérée de ces entrées selon l'équation suivante :

$$S = F\left(\sum_{i=1}^n w_i x_i - \theta\right) \quad (\text{II.5})$$

La sortie calculée, qui peut être discrète ou réelle selon la fonction d'activation utilisée, passe aux autres neurones ou directement à l'environnement extérieur.

II.5. Architecture et connexionnisme des réseaux de neurones

Par définition, un réseau de neurones est un maillage de plusieurs neurones, généralement organisés en couches. Pour construire une couche de M neurones, il s'agit simplement de les assembler comme à la figure ci-dessous [91] :

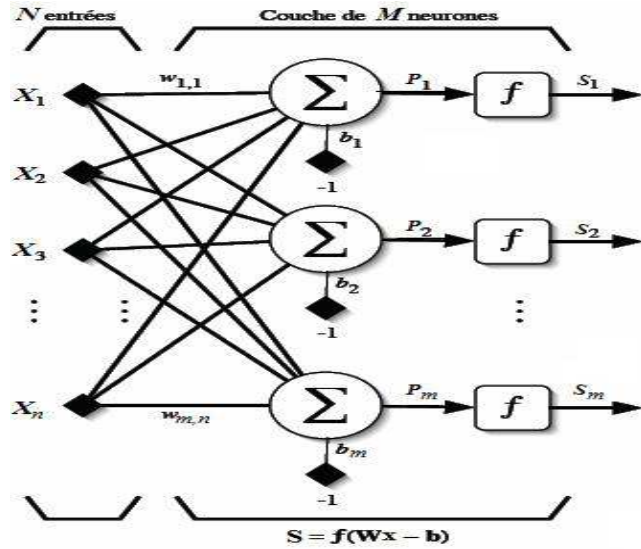


Figure II.10 : Couche de M neurones [92]

Les M neurones d'une même couche sont tous branchés aux N entrées. Nous disons alors que la couche est "totalement connectée". Un poids $W_{i,j}$ est associé à chacune des connexions. Nous noterons toujours le premier indice par i et le deuxième par j (jamais l'inverse). Le premier indice (rangée) désigne toujours le numéro de neurone sur la couche, alors que le deuxième indice (colonne) spécifie le numéro de l'entrée. Ainsi, désigne le poids de la connexion qui relie le neurone i à son entrée j . L'ensemble des poids d'une couche forme donc une matrice W de dimension :

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,n} \\ w_{2,1} & w_{2,2} & w_{2,n} \\ \vdots & \vdots & \vdots \\ w_{m,1} & w_{m,2} & w_{m,n} \end{bmatrix} \quad (\text{II.6})$$

Il faut bien sûr prendre en compte que nous n'avons pas nécessairement $M=N$ dans le cas général (les nombres de neurones et d'entrées sont indépendants). Si nous considérons que les M neurones forment un vecteur de neurones, alors nous pouvons créer les vecteurs :

$$\vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \quad \vec{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{pmatrix} \quad \vec{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{pmatrix} \quad (\text{II.7})$$

Ceci nous amène à la représentation simplifiée illustrée ci-dessous :

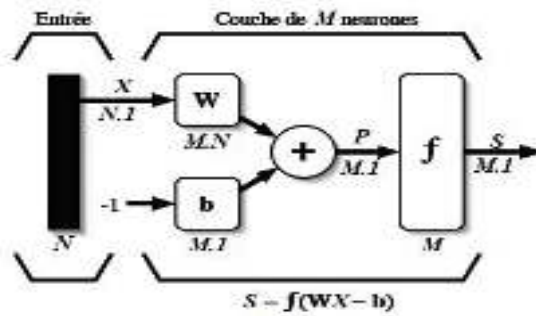


Figure II.11 : Représentation matricielle d'une couche de M neurones [92]

Pour construire un réseau de neurones (ou PMC pour "Perceptron Multicouches"), il ne suffit plus que de combiner des couches comme ci-dessous :

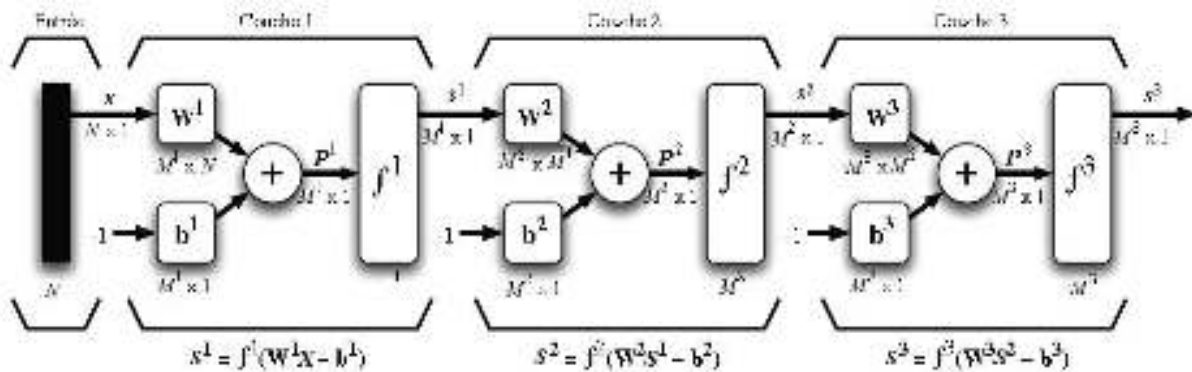


Figure II.12 : Représentation matricielle d'un réseau de trois couches [92]

Cet exemple comporte N entrées et trois couches de neurones comptant respectivement M^1 , M^2 , M^3 neurones. Dans le cas général, de nouveau ces nombres ne sont pas nécessairement égaux. Chaque couche possède aussi sa propre matrice de poids W^k , où k désigne l'indice de couche. Dans le contexte des vecteurs et des matrices relatives à une couche, nous emploierons toujours un exposant pour désigner cet indice. Ainsi, \vec{b}^k , \vec{p}^k , \vec{s}^k les vecteurs sont aussi associés à la couche k . Il importe de remarquer dans cet exemple que les couches qui suivent la première ont comme entrée la sortie de la couche précédente. Ainsi, nous pouvons enfilet autant de couches que nous voulons, du moins en théorie. Nous pouvons fixer un nombre quelconque de neurones sur chaque couche. En pratique, nous verrons plus tard qu'il n'est cependant pas souhaitable d'utiliser trop de neurones. Notons aussi que rien ne nous empêche de changer de fonction de transfert d'une couche à l'autre. Ainsi, dans le cas général nous n'avons pas nécessairement $f^1 = f^2 = f^3$.

La dernière couche est nommée "couche de sortie". Les couches qui précèdent la couche de sortie sont nommées "couches cachées". Le réseau de la figure III.8 possède donc deux couches cachées et une couche de sortie.

Les réseaux multicouches sont beaucoup plus puissants que les réseaux simples à une seule couche bien évidemment. En utilisant deux couches, à condition d'employer une fonction d'activation sigmoïde sur la couche cachée, nous pouvons "entraîner" un réseau à produire une approximation de la plupart des fonctions, avec une précision arbitraire. Sauf dans de rares cas, les réseaux de neurones formels exploitent deux ou trois couches. Entraîner un réseau de neurones signifie modifier la valeur de ses poids et de ses biais pour qu'il réalise la fonction d'entrée sortie (I/O).

Finalement, la figure II.13 illustre le dernier élément de construction que nous emploierons pour bâtir des réseaux dit "récurrents". Il s'agit d'un registre à décalage qui permet d'introduire un retard dans une donnée que l'on veut acheminer dans un réseau. La sortie retardée $S(t)$ prend la valeur de l'entrée u au temps $t-1$. Cet élément de retard présuppose que l'on peut initialiser la sortie au temps $t = 0$ avec la valeur $s(0)$. Cette condition initiale est indiquée à la figure II.13 par une flèche qui entre par le bas de l'élément.

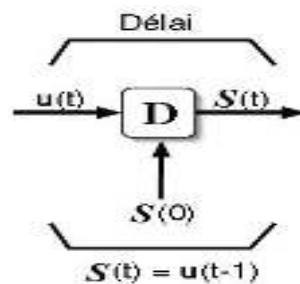


Figure II.13 : Élément de retard [27]

Selon le mode de connexion des neurones, On distingue les réseaux suivants (Figure II.14) [93]:

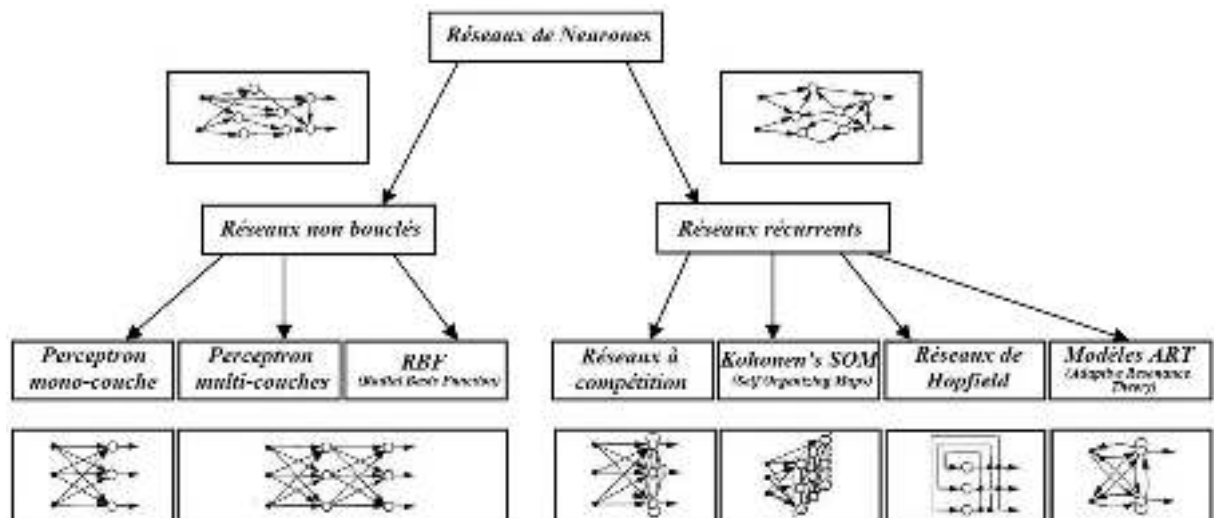


Figure II.14 : Taxinomie des architectures des réseaux non bouclés et récurrents

Selon le mode de connexion des neurones, On distingue les réseaux suivants :

II.5.1. Réseau de neurones non bouclés

La définition de ce type de réseaux selon G. Dreyfus [94] est :

"Un réseau de neurones non bouclé réalise une (ou plusieurs) fonction algébriques de ses entrées, par composition des fonctions réalisées par chacun des neurones".

Un réseau de neurones non bouclé est donc représenté graphiquement par un ensemble de neurones connectés entre eux, l'information circulant des entrées vers les sorties sans retour en arrière : si l'on se déplace dans le réseau, à partir d'un neurone quelconque, en suivant les connexions, on ne peut pas revenir au neurone de départ. Les neurones qui effectuent le dernier calcul de la composition des fonctions sont les neurones de sortie ; ceux qui effectuent des calculs intermédiaires sont les neurones cachés.

Il convient d'insister sur le fait que le temps ne joue aucun rôle fonctionnel dans le réseau de neurones non bouclé : si les entrées sont constantes, les sorties le sont également. Le temps nécessaire pour le calcul de la fonction réalisée par chaque neurone est négligeable et, fonctionnellement, on peut considérer ce calcul comme instantané. Pour cette raison, les réseaux non bouclés sont souvent appelés "réseau statique", par opposition aux réseaux bouclés ou dynamique que nous introduirons plus loin.

Le domaine d'application par excellence des réseaux non bouclé, dont l'apprentissage est souvent supervisé, est la modélisation statique type boîte noire. En effet, si on dispose

d'un ensemble de variables mesurées et d'un ensemble de mesures d'une grandeur relative à un processus de nature quelconque (physique, chimique, économique, financier...) et en supposant qu'il existe une relation (suffisamment régulière) entre le vecteur des variables et la grandeur en question, alors il est possible d'approximer cette relation par un réseau de neurone non bouclé.

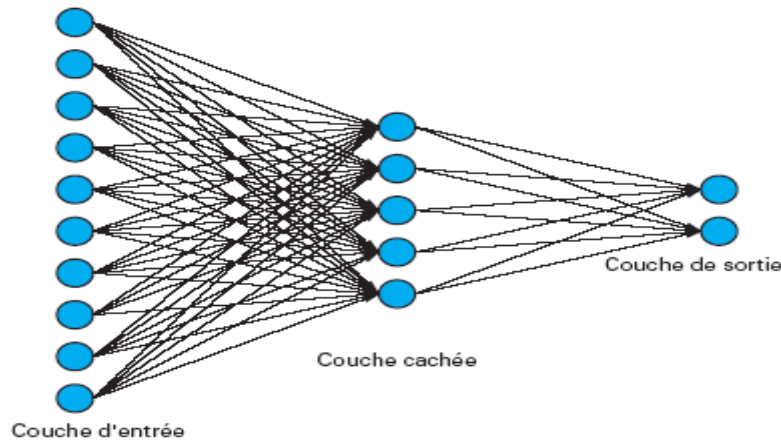


Figure II.15 Réseau multicouche non bouclé classique

Selon le nombre de couches dans un réseau et la fonction d'activation, on distingue :

II.5.1.1. Perceptron

C'est l'un des premiers réseaux de neurones, conçu en 1958 par F.Rosenblatt [95]. Il est linéaire et représente l'unité élémentaire d'un réseau de neurones.

Le perceptron est inspiré du système visuel. La première couche ou couche d'entrée représente la rétine, et la couche de sortie est équivalente aux cellules de décision.

Un Perceptron calcule à partir d'un vecteur de n entrées réelles $[x_1 \ x_2 \dots \ x_n]$, une combinaison linéaire de ces entrées ; le résultat P de cette combinaison est ensuite utilisé comme argument d'une fonction F d'une variable pour délivrer la valeur de sortie (Figure II.16). On a donc la relation :

$$s = F(P) = F\left(b + \sum_{i=1}^n x_i w_i\right) \quad (\text{II.8})$$

Dans laquelle les $(w_i) \ i \leq n$ appelés poids représentent la contribution des entrées $(x_i) \ i \leq n$ sur la sortie du perceptron. La fonction F est appelée fonction d'activation du neurone.

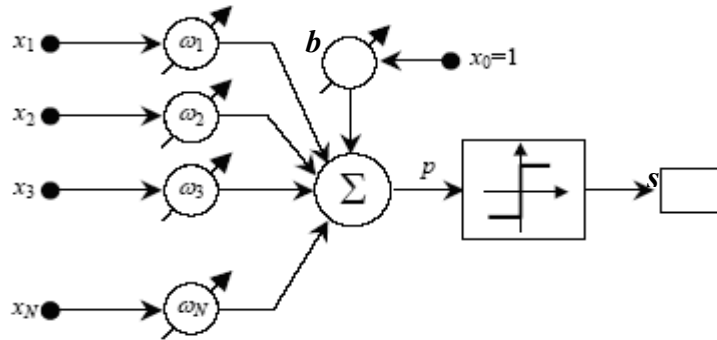


Figure II.16 : Principe d'un perceptron

II.5.1.2. Perceptron multicouche (MLP)

Les réseaux neurones multicouches contiennent en plus des couches d'entrée et de sortie, un certain nombre (supérieur ou égal à un) de couches intermédiaires ou cachées dont les neurones ont une fonction d'activation sigmoïde. Les liens n'existent évidemment qu'entre les cellules d'une couche avec les cellules de la couche suivante (réseau non bouclé). De telles architectures sont habituellement désignées dans la littérature sous le nom "Multi-Layer Perceptron" MLP (figure II.15) [94].

Les MLP possédant un seul neurone de sortie linéaire, peuvent approcher toute fonction bornée suffisamment régulière, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables [96]. Néanmoins, cette propriété, qui n'est qu'un théorème d'existence et ne donne pas la méthode de trouver les paramètres du réseau, n'est pas spécifique aux réseaux de neurones. C'est la propriété de parcimonie qui leur est particulière et fait tout leur intérêt.

Les MLP sont, en effet, des approximateurs parcimonieux, c'est-à-dire ; que le nombre des paramètres de pondération ou poids, pour une précision donnée, croît linéairement avec le nombre des variables d'entrées du modèle. La parcimonie, qui est une propriété fondamentale des MLP, est donc d'autant plus importante que le nombre des entrées d'un modèle est grand [94].

II.5.2. Réseaux de neurones bouclés

Un réseau de neurone est dit bouclé ou récurrent lorsque des sorties de neurones d'une couche sont connectées aux entrées d'une couche d'indice inférieur. Autrement dit, lorsqu'on se déplace dans le réseau en suivant le sens des connexions, il est possible de trouver au moins un chemin qui revient à son point de départ. La sortie d'un neurone du réseau peut

donc être fonction d'elle-même ; cela n'est évidemment concevable que si la notion de temps est explicitement prise en considération. Raison pour la quelle ce type de réseaux sont appelés dynamique.

Ainsi, à chaque connexion d'un réseau de neurones bouclé est attaché, outre un poids comme pour les réseaux non bouclé, un retard, multiple entier de l'unité de temps choisie [94]. Une grandeur, à un instant donné, ne pouvant pas être fonction de sa propre valeur au même instant, tout cycle du graphe du réseau doit avoir un retard non nul.

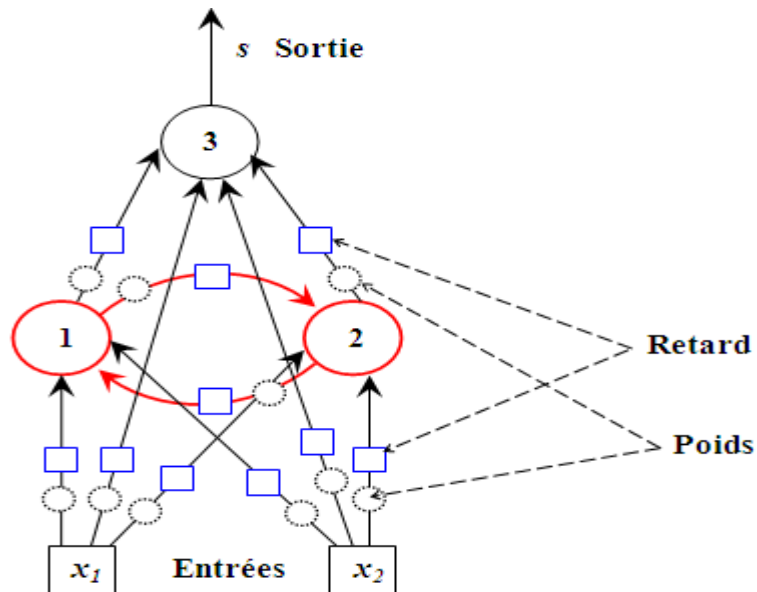


Figure II.17 : Un réseau de neurone bouclé à deux entrées contenant une boucle ente les neurones (1) et (2).

Les réseaux de neurones bouclés sont mise en œuvre essentiellement pour concevoir des modelés dynamiques. Ils peuvent être utilisés à titre d'exemples ; comme modèle de simulation pour prévoir l'évolution d'un processus quelconque dans le temps ou comme prédicteur à très court terme (une période d'échantillonnage) afin de l'intégrer à un système de commande [94].

Selon la complexité de son connexionnisme, on distingue :

II.6. Apprentissage

On appelle apprentissage des réseaux de neurones la procédure qui consiste à estimer les paramètres (poids) des neurones du réseau, afin que celui-ci remplisse au mieux la tâche qui lui est affectée [94]. L'apprentissage est une phase du développement du réseau de neurones durant laquelle on calcule les poids des neurones de telle manière que les sorties du

réseau soient aussi proches que possible des sorties désirées. L'apprentissage fait appel à des exemples de comportement du processus à modéliser [97].

On distingue deux types d'apprentissage : l'apprentissage supervisé et l'apprentissage non supervisé, selon qu'on fournit ou non au modèle les sorties désirées :

II.6.1. Apprentissage supervisé

L'apprentissage supervisé admet l'existence d'un expert (ou professeur) qui possède une connaissance innée de l'environnement. Le rôle de l'expert est de fournir les informations relatives à l'environnement nécessaires à l'apprentissage du réseau, sous forme d'un ensemble d'exemples composés de stimuli auxquels sont associées des réponses désirées.

L'apprentissage supervisé exige donc que les exemples ou les données soient constitués de couples de valeurs du type : (valeur d'entrée, valeur de sortie désirée). Tout le problème de l'apprentissage supervisé consiste, étant donné un vecteur d'apprentissage de n couples (x, d) $i = 1, 2, \dots, n$, à déterminer le vecteur des poids des neurones capables de prédire le même vecteur de sortie à partir du même vecteur d'entrée.

Un apprentissage supervisé force donc le réseau à converger vers un état final précis, en même temps qu'on lui présente un motif (données).



Figure II.18 : Illustration de l'apprentissage supervisé

II.6.2. Apprentissage non supervisé

Contrairement à l'apprentissage supervisé, lors d'un apprentissage non-supervisé, le réseau est laissé libre de converger vers n'importe quel état final lorsqu'on lui présente un motif. Il n'y a donc pas de professeur qui dirige le réseau vers une situation souhaitée, c'est au réseau de découvrir les ressemblances entre les éléments de la base de données, et de les traduire par des réponses adéquates.

L'apprentissage est qualifié donc de non supervisé lorsque seules les valeurs d'entrée sont disponibles. Dans ce cas, les exemples présentés à l'entrée provoquent une auto adaptation du réseau afin de produire des valeurs de sortie qui soient proches en réponse pour des valeurs d'entrées similaires.

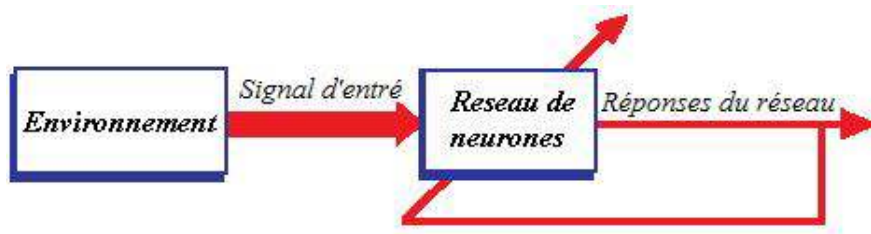


Figure II.19 : Illustration de l'apprentissage non supervisé

II.7. Règles d'apprentissage

II.7.1. Règle d'apprentissage de Hebb

Les réseaux neuronaux se sont développés grâce à leur capacité d'apprentissage. Le psychologue Hebb fût le premier à présenter un apprentissage à partir des neurones artificiels. "Lorsqu'un axone de la cellule A est suffisamment proche pour pouvoir exciter la cellule B et qu'il prend part de manière répétitive ou persistente à cette excitation, alors on doit trouver soit un phénomène de croissance, soit un changement métabolique dans l'une ou l'autre des cellules tel que l'efficacité de la cellule A pour exciter B doit être accrue." [76].

En d'autres termes, la synapse entre la cellule A et la cellule B doit faciliter la transmission de l'influx nerveux lorsque les deux sont excités en même temps, et au contraire si les cellules A et B ne sont pas activés sur la même excitation, la synapse doit inhiber cette transmission.

La règle de Hebb est utilisée dans un apprentissage supervisé. Son algorithme modifie de façon itérative les poids pour adapter la réponse obtenue à la réponse désirée. Il comprend les points suivants [98] :

1. Présentation des matrices d'entrée X et des vecteurs de sortie désirée D de la base d'apprentissage;
2. Initialisation des poids w_i et du seuil θ à des valeurs petites, choisies au hasard, ainsi que μ le pas de modification des poids;

3. Calcul de la sortie S obtenue pour cette entrée :

$$S = F(x, w) = F\left(\sum_i (w_i x_i) - \theta\right) \quad (\text{II.9})$$

$$F(x, w) = \begin{cases} +1 & \text{si } \left(\sum_i (w_i x_i) - \theta\right) > 0 \\ -1 & \text{si } \quad \quad \quad \text{non} \end{cases}$$

(II.10)

4. Si la sortie S est différente de la sortie désirée D alors modification des poids :

$$w_i^{t+1} = w_i^t + \mu(x_i \cdot d_i) \quad (\text{II.11})$$

5. Tant que tous les exemples de la base d'apprentissage ne sont pas traités correctement, retour à l'étape 3.

Le perceptron est donc capable de réaliser une partition de son espace d'entrée en deux classes (1 et 2), selon la valeur de sa sortie (+1 ou -1). La séparation entre ces deux classes étant donnée par un hyperplan (figure II.20). L'équation de la droite séparatrice appelée encore droite de décision est [98]:

$$\sum_i (w_i x_i) - \theta = 0 \quad (\text{II.12})$$

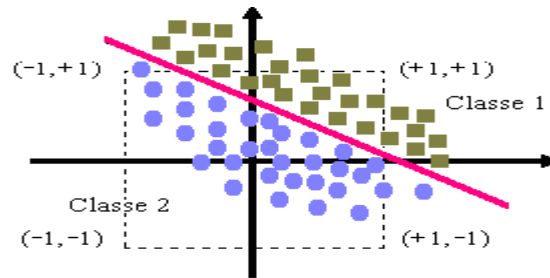


Figure II.20 : Séparation des exemples d'apprentissage par la droite de décision

II.7.2. Règle d'apprentissage du perceptron de F. Rosenblatt

Elle consiste à présenter au réseau une série d'exemples d'apprentissage, puis à minimiser l'erreur entre la sortie désirée et la sortie effectivement obtenue. L'algorithme est le suivant [98]:

1. Présentation des matrices des vecteurs d'entrée X et de la sortie désirée D de la base d'apprentissage;

2. Initialisation des poids w_i et du seuil θ à des valeurs petites choisies au hasard, ainsi que μ le pas de modification des poids et ε l'erreur admissible;
3. Calcul de la sortie S obtenue pour cette entrée :

$$S = F(x, w) = F\left(\sum_i (w_i x_i) - \theta\right) \quad (\text{II.13})$$

$$F(x, w) = \begin{cases} +1 & \text{si } \left(\sum_i (w_i x_i) - \theta\right) > 0 \\ -1 & \text{si } \quad \quad \quad \text{non} \end{cases} \quad (\text{II.14})$$

4. Calcul de l'erreur :

$$e = d - s \quad (\text{II.15})$$

5. Si $e < \varepsilon$ le calcul s'arrête;
6. Si non, on ajuste les poids synaptiques :

$$w_i^{t+1} = w_i^t + \mu[(d_i - s_i)x_i] \quad (\text{II.16})$$

7. Tant que tous les exemples de la base d'apprentissage ne sont pas traités correctement, retour à l'étape 3.

II.7.3. Règle d'apprentissage de Widrow-Hoff (ADALINE)

Le modèle de l'adaline (ADaptive LINar Element) a été présenté par B. Widrow, [99] chercheur américain à Stanford, dans les années 1960. Dans la règle de Widrow-hoff (appelé aussi règle delta), l'apprentissage est réalisé par itération. Les poids à la phase initiale sont générés aléatoirement. L'idée est de remplacer la fonction seuil d'activation dans les cellules de sortie par la fonction identité (figure II.20). L'avantage de cette approche est de ne plus être restreinte à une réponse binaire, et d'offrir tout l'éventail de valeurs possibles. Cependant, l'apprentissage va être défini comme la minimisation de l'erreur globale du système. La fonction d'erreur ou fonction de coût repose sur la méthode des moindres carrés [100]:

$$E = \frac{1}{2} \sum_i (d_i - s_i)^2 \quad (\text{II.17})$$

L'appellation delta vient du fait qu'on exprime la modification d'une connexion par :

$$w_i^{t+1} = w_i^t + \Delta w_i \quad (\text{II.18})$$

Avec :

$$\Delta w_i = -\mu \frac{\partial E}{\partial w_i} \quad (\text{II.19})$$

$$\frac{\partial E}{\partial w_i} = \frac{\partial}{\partial w_i} \left[\frac{1}{2} \sum_i (d_i - s_i)^2 \right] \quad (\text{II.20})$$

$$\frac{\partial E}{\partial w_i} = -(d_i - s_i) \frac{\partial s_i}{\partial w_i} \quad (\text{II.21})$$

$$\frac{\partial E}{\partial w_i} = -(d_i - s_i) x_i \quad (\text{II.22})$$

D'où

$$w_i = w_i + \mu (d_i - s_i) x_i \quad (\text{II.23})$$

Son algorithme d'apprentissage peut être décrit par les points suivants [100] :

1. Présentation des matrices des vecteurs d'entrée X et de la sortie désirée D de la base d'apprentissage;
2. Initialisation des paramètres initiaux;
3. Calcul de la sortie S

$$S = F(P) = P \quad (\text{II.24})$$

Avec

$$P = \sum_i w_i x_i - \theta \quad (\text{II.25})$$

4. Phase de test : calcul de l'erreur quadratique E

$$E = \frac{1}{2} \sum_i (d_i - s_i)^2 \quad (\text{II.26})$$

5. Si $E < \varepsilon$ (erreur admissible), le calcul s'arrête
6. Si non, on ajuste les poids synaptiques :

$$w_i^{t+1} = w_i^t + \mu (d_i - s_i) x_i \quad (\text{II.27})$$

7. Retour à l'étape 3.

II.7.4. Algorithme de rétropropagation du gradient

L'algorithme de rétropropagation du gradient est certainement à la base des premiers succès des réseaux de neurones multicouches. Il figure parmi les algorithmes d'apprentissage les plus utilisés.

Il s'agit toujours de minimiser l'erreur quadratique, ce qui est assez simple quand on utilise une fonction d'activation dérivable (la sigmoïde par exemple). On propage la modification des poids de la couche de sortie jusqu'à la couche d'entrée [101].

L'apprentissage par rétropropagation du gradient d'un perceptron multicouche peut donc se décomposer principalement en quatre phases [101] :

- la propagation du signal de la couche d'entrée jusqu'à la couche de sortie,
- le calcul de l'erreur en sortie,
- la rétropropagation de l'erreur;
- correction des poids synaptiques.

Avant tout apprentissage, il est indispensable de normaliser et de centrer toutes les variables d'entrées. Il faut aussi, qu'au lancement de l'apprentissage, les valeurs initiales des poids doivent être différentes de zéro pour que l'algorithme de rétropropagation puisse fonctionner. D'autre part, l'utilisation de valeurs élevées peut provoquer un phénomène de saturation prématurée qui contribue à diminuer la vitesse de convergence de l'apprentissage [102, 103]. La fonction de coût la plus utilisée dont on cherche à évaluer le gradient est de la forme [104] :

$$E = \sum_i (d_i - s_i)^2 \quad (\text{II.28})$$

Elle exprime l'erreur ou la distance quadratique entre la sortie réelle du modèle s_i et la sortie désirée d_i . Cette fonction a tendance à amplifier les erreurs les plus importantes. Par conséquent, au cours de l'apprentissage, la mise à jour des poids est largement déterminée par la correction des grandes erreurs, ce qui est recherché en général [105].

L'apprentissage comporte une première phase de calcul dans le sens direct (figure II.21), où chaque neurone effectue la somme pondérée de ses entrées et applique en suite la

fonction d'activation F (fonction dérivable), pour obtenir la mise à jour du neurone. L'équation suivante correspond à cette mise à jour :

$$s_i = F\left(P_i = \sum_{j=0}^n w_{ij} x_j\right) \tag{II.29}$$

Avec :

- P_i : le potentiel post-synaptique du neurone i ;
- x_j : l'état du neurone de la couche précédente;
- w_{ij} : le poids de la connexion du neurone j vers le neurone i .

Cette phase, dite de propagation, permet de calculer la sortie du réseau en fonction de l'entrée.

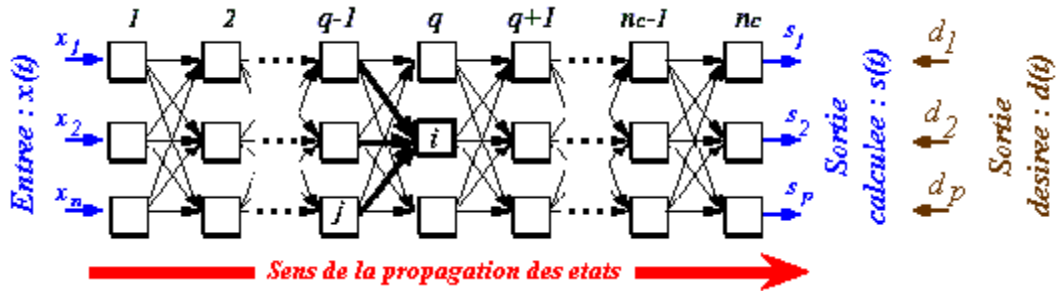


Figure II.21 : Sens de propagation du calcul des sorties s_i

Pour pouvoir minimiser la fonction de coût, on doit calculer son gradient pour tous les poids de la manière suivante [101] :

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial P_i} \frac{\partial P_i}{\partial w_{ij}} = \frac{\partial E}{\partial P_i} x_j \tag{II.30}$$

Le gradient δ_i sera ensuite :

$$\delta_i = -\frac{\partial E}{\partial P_i} \tag{II.31}$$

On distingue alors deux cas, suivant que le neurone d'indice i est un neurone de sortie ou non. Dans le cas de la couche de sortie, le gradient attaché aux cellules de sortie est alors obtenu par l'équation suivante [101] :

$$\delta_i = -\frac{\partial E}{\partial P_i} = -\frac{\partial E}{\partial s_i} \frac{\partial s_i}{\partial P_i} = -\frac{\partial E}{\partial s_i} \frac{\partial F(P_i)}{\partial P_i} = 2(d_i - s_i)F'(P_i) \tag{II.32}$$

Car seul s_i dépend de P_i et $s_i = F(P_i)$.

Pour les neurones des couches cachées, l'ordre de calcul des gradients est l'inverse de celui utilisé pour la mise à jour des états dans le réseau. Il s'effectue d'une manière récursive en menant les calculs depuis la (ou les) sortie(s) du réseau vers ses entrées (figure II.22); on parle alors de rétropropagation. L'expression du gradient prend la forme suivante [101] :

$$\delta_i = -\frac{\partial E}{\partial P_i} = -\sum_l \frac{\partial E}{\partial P_l} \frac{\partial P_l}{\partial P_i} = \sum_l \delta_l \frac{\partial P_l}{\partial P_i} = \sum_l \delta_l \frac{\partial P_l}{\partial s_i} \frac{\partial s_i}{\partial P_i} \tag{II.33}$$

Soit encore :

$$\delta_i = F'(P_i) \sum_l \delta_l w_{li} \tag{II.34}$$

Où δ_l : le gradient du neurone l de la couche suivante (en aval) de la cellule i .

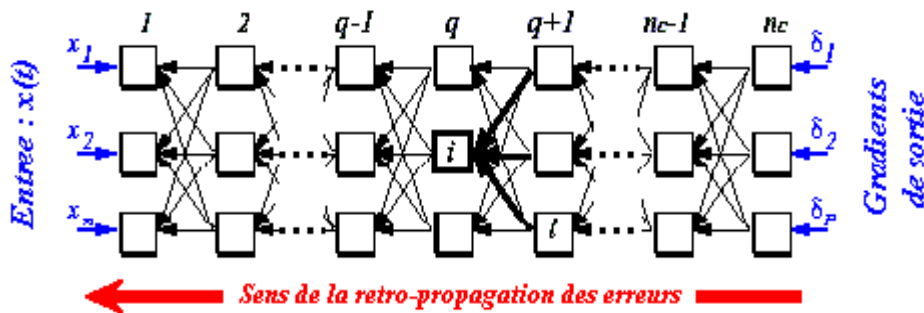


Figure II.22 : Sens de rétropropagation des gradients δ_i .

Dans le cas de l'algorithme de gradient total, ils s'agit d'un apprentissage non adaptatif c'est-à-dire tous les exemples de la base d'apprentissage sont présentés successivement au réseau, la modification des poids n'intervient qu'après présentation de tous les exemples. Par opposition au gradient stochastique, qui s'effectue au moyen d'un apprentissage adaptatif, la modification des poids est effectuée au fur et à mesure que les données ou les exemples sont disponibles [106].

Enfin, selon l'équation générale de modification des poids w_{ij} [103] :

$$w_{ij}^{t+1} = w_{ij}^t + \Delta w_{ij}^t \tag{II.35}$$

$$\Delta w_{ij} = -\mu \frac{\partial E}{\partial w_{ij}} = \mu \delta_i x_j \quad (\text{II.36})$$

D'où :

$$w_{ij}^{t+1} = w_{ij}^t + \mu \delta_i x_j \quad (\text{II.37})$$

μ : est un petit nombre positif qui représente le pas de déplacement en direction du minimum le plus proche (pas d'apprentissage ou pas du gradient).

La méthode d'évaluation des paramètres w du modèle (l'équation (II.37)), dite méthode du gradient simple, est commode, mais elle présente de nombreux inconvénients:

- si le pas du gradient est trop petit, la décroissance du coût est très lente ; si le pas est trop grand, le coût peut augmenter ou osciller ;
- au voisinage d'un minimum de la fonction du coût, le gradient de cette dernière tend vers zéro, donc l'évolution du vecteur des coefficients devient très lente ; il en va de même si la fonction de coût présente des plateaux où sa pente est très faible ; ces plateaux peuvent être près d'un minimum, et, dans la pratique, il est impossible de savoir si une évolution très lente du gradient est due au fait que l'on est au voisinage d'un minimum, ou que l'on se trouve sur un plateau de la fonction du coût ;
- si la courbure de la fonction du coût varie beaucoup, la direction du gradient peut être très différente de la direction qui mènerait vers un minimum.

Un autre problème surgit fréquemment lors de l'apprentissage des réseaux de neurones est que certains neurones sont écartés lors de la phase d'apprentissage vu que leur vecteur de poids ne subit pas de modification au cours de l'apprentissage ce qui rend la contribution du neurone au quel sont associés dans l'ensemble du réseau, insignifiante. Ces neurones sont souvent désignés de neurones morts ou inactifs.

Pour résoudre ce problème un nombre appelé biais est souvent ajouté à chaque neurone du réseau. Le résultat est que le biais des neurones qui ne répondent pas fréquemment, augmente au cours de l'apprentissage pour permettre à ces derniers de contribuer d'une manière significative dans l'élaboration du modèle. Par contre, le biais des neurones très actifs diminue. Ce terme joue, en effet, le rôle de régulateur.

La rétropropagation du gradient permet donc de minimiser la fonction de coût $E(w_{ij})$. Toutefois, cette méthode ne garantit pas l'obtention du minimum global de la fonction (blocage dans un minimum local). Dans ce cas la minimisation de l'erreur n'est pas optimale.

Cette méthode consiste en fait à se placer en un point et à se laisser descendre le long des pentes de la courbe $Z = E(w_{ij})$, tel qu'elle est illustrée par la figure suivante :

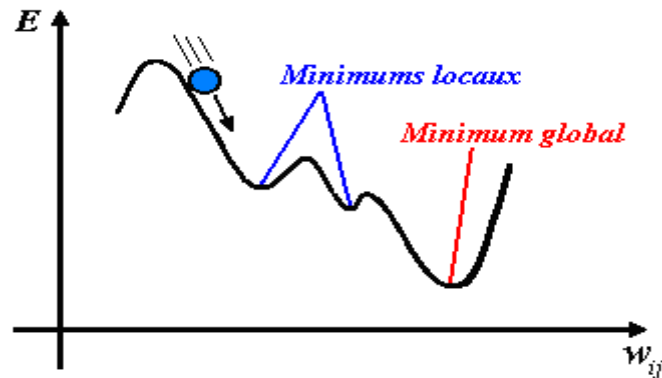


Figure II.23 : Graphe de la surface d'erreur

En résumé, l'algorithme de rétropropagation du gradient s'articule autour des points suivants :

1. Initialisation des poids à des petites valeurs aléatoires;
2. Présentation d'une entrée X_k et de la sortie désirée d_k ;
3. Calcul de la sortie actuelle par propagation à travers les couches en utilisant la formule (II.29);
4. Accumulation des erreurs en sortie, en utilisant la formule (II.19);
5. Rétropropagation du gradient de l'erreur depuis la dernière couche vers la première couche :
 - Pour chaque cellule de sortie en utilisant la formule (II.32);
 - Pour chaque cellule cachée en utilisant la formule (II.34);
6. Mise à jour des poids en utilisant la formule (II.37);
7. Répéter les étapes de 2 à 6 pour l'entrée suivante, jusqu'à ce que l'erreur en sortie soit inférieure à la limite fixée ou que le nombre maximum d'itérations soit atteint.

Ils existent d'autres variantes pour la modification des paramètres d'un modèle (gradient du second ordre, BFGS, Levenberg-Marquardt...) dont nous présentons respectivement les principes ci-après.

II.7.4.1. Méthode de Newton

Pour trouver une solution qui minimise la fonction de coût, la méthode de rétropropagation du gradient utilise uniquement l'information donnée par la dérivée de premier ordre. Il est connu en optimisation que l'utilisation des informations données par les dérivées d'ordre supérieur peut accélérer le processus de recherche d'une solution optimale. De nombreux travaux ont tenté d'étendre cette approche aux techniques adaptatives dans le cadre des réseaux de neurones [107].

Les méthodes du second ordre opèrent une recherche itérative d'un point minimum w^* à partir d'approximations locales quadratiques de la fonction de coût E .

Supposons l'adaptation ($\Delta w^t = w^{t+1} - w^t$) des paramètres du réseau relativement faible à chaque itération. Il est alors possible d'estimer la valeur $E(w^{t+1})$ de la fonction de coût E à l'instant $(t+1)$, par un développement en série de *Taylor* multivariable, limité à l'ordre 2 au voisinage du point w_t :

$$E(w^{t+1}) = E(w^t + \Delta w^t) \cong E(w^t) + \nabla_{(w)} E(w^t) \Delta w^t + \frac{1}{2} \Delta w^t H(w^t) \Delta w^t \quad (\text{II.38})$$

Où $H(w^t) = \nabla_{(w)}^2 E(w^t)$ représente la matrice Hessienne des dérivées partielles du second ordre de la fonction de coût E , évaluée au point w_t . Les éléments de cette matrice sont définis par :

$$H(w)_{ij} = \frac{\partial^2 E(w)}{\partial w_i \partial w_j} \quad (\text{II.39})$$

$$H(w) = \begin{bmatrix} \frac{\partial^2 E}{\partial w_1^2} & \frac{\partial^2 E}{\partial w_1 \partial w_2} & \dots & \frac{\partial^2 E}{\partial w_1 \partial w_n} \\ \frac{\partial^2 E}{\partial w_2 \partial w_1} & \frac{\partial^2 E}{\partial w_2^2} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial w_n \partial w_1} & \frac{\partial^2 E}{\partial w_n \partial w_2} & \dots & \frac{\partial^2 E}{\partial w_n^2} \end{bmatrix} \quad (\text{II.40})$$

Un point minimum w^* de la fonction de coût est approximativement localisé, à partir de l'expression approchée (I.36), par annulation du vecteur gradient $\nabla_{(\Delta w)} E(w^{t+1})$ selon :

$$\nabla_{(\Delta w)} E(w_{t+1}) = \nabla_{(w)} E(w_t) + H(w^t) \Delta w_t = 0 \quad (\text{II.41})$$

Soit :

$$\Delta w^t = -H(w^t)^{-1} \nabla_{(w)} E(w^t) \quad (\text{II.42})$$

La règle d'adaptation des paramètres du réseau du second ordre se déduit alors de l'expression (II.39) suivant :

$$w^{t+1} = w^t - \mu^t A \quad (\text{II.43})$$

$A = -H(w^t)^{-1} \nabla_{(w)} E(w^t)$: représente la direction de minimisation du second ordre empruntée à l'instant t , et μ^t constitue le pas de la descente (coefficient d'adaptation positif $0 < \mu^t < 1$).

La méthode d'optimisation paramétrique décrite par la règle d'adaptation (II.43) est connue sous la dénomination de méthode d'optimisation de Newton. Elle peut localiser le minimum d'une fonction de coût quadratique en une unique itération (en ce cas le pas de la descente est constant fixé à un $\mu^t = 1$).

En pratique, cette méthode est peu employée car elle nécessite le calcul du Hessien et surtout de son inverse à chaque itération qui est complexe. On lui préfère une méthode plus économique dite de quasi-Newton.

II.7.4.2. Méthode de quasi-Newton

La méthode de quasi-Newton [107] consiste à approcher l'inverse du Hessien plutôt que de calculer sa valeur exacte. L'inverse du Hessien est approché par une matrice M^k définie positive modifiée à chaque itération. La suite des matrices $\{M^k\}$ est construite de manière à converger vers l'inverse du Hessien lorsque la fonction de coût est quadratique. La modification des paramètres est donnée par :

$$w^{t+1} = w^t - \mu^t M^t \nabla E^t \quad (\text{II.44})$$

A la première itération, la matrice M^0 est prise égale à la matrice identité. Cette suite est construite grâce à la méthode dite BFGS, développée indépendamment par *C.G.Broyden*, *R.Fletcher*, *D.Goldfarb* et *D.F.Shanno* [108], dont la vitesse de convergence est beaucoup plus grande que celle de la méthode du gradient.

II.7.4.3. Méthode Levenberg-Marquardt

La méthode de *Levenberg-Marquardt* [109] consiste à modifier les paramètres selon la relation suivante :

$$w^{t+1} = w^t - (H^t + \eta^t I)^{-1} \nabla E^t \quad (\text{II.45})$$

Avec I : matrice identité

Cette méthode est particulièrement astucieuse car elle s'adapte d'elle-même à la forme de la fonction de coût. Elle effectue un compromis entre la direction du gradient et la direction donnée par la méthode de Newton. En effet, si η^t est grand, on reconnaît la méthode du gradient simple (dans ce cas la valeur du pas est donnée par $1/\eta^t$) et si η^t est petit, la modification des paramètres correspond à celle de la méthode de Newton [108].

II.8. Théorie de la généralisation

La généralisation concerne la tâche accomplie par le réseau une fois son apprentissage achevé [110]. Elle peut être évaluée en testant le réseau sur des données qui n'ont pas servi à l'apprentissage. Elle est influencée essentiellement par quatre facteurs : la complexité du problème, l'algorithme d'apprentissage, la complexité de l'échantillon (le nombre d'exemples et la manière dont ils représentent le problème) et enfin la complexité du réseau (nombre de poids). La complexité du problème est déterminée en partie par sa nature même : on peut parler de "complexité intrinsèque". Par ailleurs, l'algorithme d'apprentissage influe sur la généralisation par son aptitude à trouver un minimum local assez profond, sinon le minimum global [84].

La complexité de l'échantillon est définie par la figure II.24. Supposons que l'on cherche à extraire une règle à partir d'exemples, la zone U représente l'ensemble des couples entrée-sortie possibles. La partie R représente la région dans laquelle la règle qu'on cherche à extraire est valable. On dit que R est consistant avec la règle. On sélectionne un certain nombre d'exemples pour l'apprentissage du réseau, qui forment l'ensemble d'apprentissage Z_A . On sélectionne un autre ensemble distinct Z_T pour tester le réseau et estimer sa généralisation. Les deux ensembles Z_A et Z_T doivent être représentatifs de la région R .

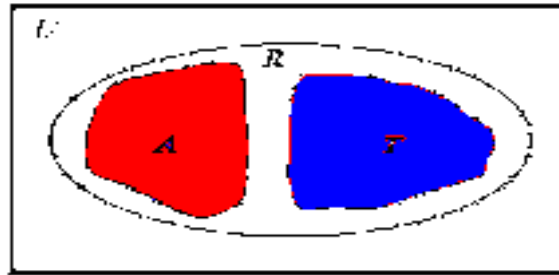


Figure II.24 : Extraction de règles

L'autre facteur influant sur la généralisation est la complexité du réseau. Pour bien cerner cet aspect, on fait l'analogie avec un problème de régression polynomiale classique. Supposons que l'on dispose d'un nuage de points issus d'une fonction g d'une variable réelle inconnue. Les exemples à notre disposition sont des couples (x_i, y_i) bruités de la forme [84] :

$$y_i = g(x_i) + \varepsilon_i \quad (\text{II.46})$$

Où les ε_i sont des réalisations d'une variable aléatoire. L'objectif est de modéliser la fonction g par un modèle polynomial en utilisant les exemples d'apprentissage.

Les courbes données par la figure II.25 représentent la modélisation de g par trois modèles qui diffèrent par leur nombre de paramètres. On peut constater que le modèle ayant très peu de paramètres n'a pas assez de flexibilité pour réaliser un apprentissage correct des exemples d'apprentissage. Les erreurs d'apprentissage et de test sont toutes deux importantes : c'est la situation de sous-apprentissage (figure II.25.c). En revanche, le modèle constitué de nombreux paramètres, lisse parfaitement les exemples d'apprentissage. Il commet donc une erreur faible sur ces données, mais probablement une erreur plus importante sur les données de test. C'est la situation de sur-apprentissage (figure II.25.b). Finalement, le modèle possédant un nombre de paramètres modérés réalise un bon compromis entre précision d'apprentissage et bonne généralisation (figure II.25.a).

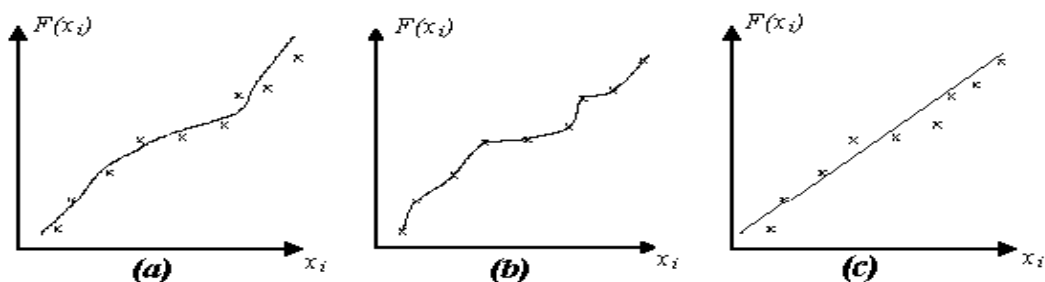


Figure II.25 : Cas d'apprentissage [19] a. Apprentissage correct, b. Sur-apprentissage et c. Sous-apprentissage.

Le problème de la généralisation est souvent vu sous deux perspectives différentes. Dans la première, la taille du réseau est fixée (en accord avec la complexité du problème) et la question est : combien d'exemples d'apprentissage sont nécessaires pour atteindre une bonne généralisation ? Cette perspective est intéressante dans les applications où l'on a la possibilité d'acquérir autant d'exemples que l'on veut. Dans le second cas, le nombre d'exemples d'apprentissage est fixé et la question est : quelle taille du réseau donne la meilleure généralisation pour ces données ? On est conduit à adopter ce point de vue lorsque l'on est limité dans la possibilité d'acquérir des données d'apprentissage : il importe alors de déterminer quelle est la taille permettant au réseau de décrire au mieux les données en notre possession [84].

Cependant une manière d'évaluer la faculté de généraliser la représentation construite par le réseau de neurones à toutes les données, y compris celles n'appartenant pas à l'ensemble d'apprentissage, consiste à mesurer les performances du réseau sur des données représentatives du problème non apprises. Il s'agit d'une évaluation de l'erreur de généralisation. La différence entre l'erreur d'apprentissage et l'erreur de généralisation (figure II.26) représente une mesure de la qualité de l'apprentissage effectué [111].

II.9. Critère d'arrêt pour l'apprentissage

L'une des méthodes qui permet de contrôler la capacité d'un réseau de neurones consiste à arrêter l'apprentissage "à temps". De ce fait différents critères permettant de décider quand stopper l'algorithme d'apprentissage ont été développés :

- quand l'erreur d'apprentissage a atteint un seuil fixé;
- après un nombre fixé de cycles d'apprentissage;
- quand une estimation de l'erreur de généralisation est minimum.

Les méthodes qui évaluent l'erreur de généralisation sont presque toutes basées sur la partition de l'ensemble des données qu'on possède en plusieurs sous-ensembles. Par exemple un ensemble utilisé pour l'apprentissage et un ensemble de validation. L'ensemble de validation est utilisé pour contrôler et mesurer la généralisation du réseau. Pendant l'apprentissage l'erreur d'apprentissage décroît continuellement, tandis que sur l'ensemble de validation elle ne diminue que jusqu'à un certain point au-delà duquel elle augmente [110]. A partir de ce point le réseau apprend par cœur les données de l'ensemble d'apprentissage et l'apprentissage doit être stoppé (figure II.26). Ces deux ensembles servent à déterminer

l'architecture la plus appropriée pour le problème à traiter : pour différentes architectures (nombre de neurones cachés variable), on contrôle l'erreur de validation et on choisit l'architecture pour laquelle elle est minimale.

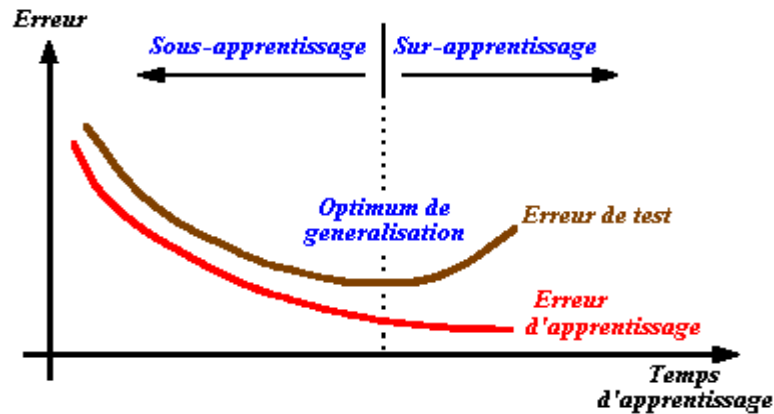


Figure II.26: Evolution des erreurs d'apprentissage et de test au cours du temps [111].

Parmi les méthodes utilisant cette méthodologie on peut citer les méthodes appelées "split-samples", "Cross-validation" :

- La technique nommée "split-samples" réserve un troisième ensemble de données appelé ensemble de test, pour tester le réseau sur des données qui n'ont jamais été utilisées ni pour l'apprentissage ni pour la validation.
- Pour la cross-validation l'ensemble des données de départ est découpé en k parties de taille égale. Le réseau est entraîné k fois, chaque fois en utilisant $(k-1)$ parties pour l'apprentissage et la dernière pour la validation et le calcul des erreurs commises en généralisation. L'ouvrage de *R. Tibshirani* [112] donne une comparaison de ces méthodes.

II.10. Procédure de modélisation neuronale

Le modèle boîte noire constitue la forme la plus primitive de modèle mathématique : il est réalisé uniquement à partir de données expérimentales ou d'observations ; il peut avoir une valeur prédictive, dans un certain domaine de validité, mais il n'a aucune valeur explicative [94]. A l'opposé, un modèle de connaissance, que l'on pourrait qualifier de boîte blanche, est issu de l'analyse des phénomènes physiques, chimiques, biologiques, etc., dont résulte la grandeur que l'on cherche à modéliser. Ces phénomènes sont mis en équation à l'aide des connaissances théoriques disponibles au moment de l'élaboration du modèle. La démarche scientifique s'efforce, dans la mesure du possible, de construire des modèles de connaissance :

l'élaboration d'un modèle de connaissance nécessite l'existence d'une théorie, alors qu'un modèle boîte noire requiert seulement la disponibilité de mesures.

Entre la boîte noire et le modèle de connaissance se situe le modèle semi- empirique, ou modèle boîte grise, qui contient à la fois des équations résultant d'une théorie, et des équations purement empiriques, résultant d'une modélisation de type boîte noire [94].

Les réseaux de neurones constituent le plus souvent des modèles boîte noire, néanmoins dans certaines conditions, ils peuvent être utilisés avec profit dans le cadre de modèles boîte grises.

Nous examinerons en bref, exclusivement, dans ce qui suit la conception d'un modèle statique type boîte noire, réalisé à l'aide de réseaux de neurones non bouclés (MLP) du fait qu'il a un lien direct avec le thème de ce travail.

L'objectif de la modélisation boîte noire est de trouver, à partir des données ou des mesures disponibles, une relation déterministe, si elle existe, entre les variables d'entrées du modèle x et la grandeur à modéliser y .

La procédure de développement d'un réseau de neurone qui répond à une modélisation statique peut se résumer aux points suivants:

1. Sélection des données
2. Collecte des données
3. Séparation des bases de données,
4. Normalisation des données,
5. Choix du type de réseau de neurones,
6. Choix du nombre de neurones dans la couche cachée
7. Choix de la fonction de transfert
8. Estimation des paramètres ou apprentissage,
9. Validation et généralisation.

II.10.1. Sélection des données

Lorsqu'on conçoit un modèle à partir de mesures, il est indispensable que l'ensemble des variables d'entrée soit aussi réduit que possible ; en effet, à chaque entrée du modèle correspond un nombre de paramètres égal au nombre de neurones cachés. Il est donc nécessaire

- que la représentation des entrées soit aussi compacte que possible,

- que la représentation des entrées contienne tous les facteurs pertinents par rapport à la grandeur à modéliser, mais seulement ceux-ci. En effet, la présence de variables qui ont, sur la grandeur à modéliser, moins d'influence que les perturbations ou le bruit, entraîne celle de paramètres superflus et introduit des variations non significatives sur les variables d'entrées, qui constituent autant de sources d'erreur pour le modèle [94].

Avant la conception d'un modèle neuronale, il convient donc de procéder à une sélection des entrées. Cette sélection revêt deux aspects :

- réduction de la dimension de l'espace de représentation des variables du modèle,
- élimination des entrées qui n'ont qu'une faible influence sur la sortie.

II.10.2. Collecte des données

Une fois les entrées pertinentes du modèle sont bien connues, l'objectif de cette étape est de les recueillir, à la fois pour développer (l'apprentissage) le réseau de neurones et pour le tester. La base de donnée reconstituée doit être représentative des données susceptibles d'intervenir en phase d'utilisation du système neuronal.

II.10.3. Séparation des bases de données

Afin de développer un modèle neuronal, il est nécessaire de disposer d'au moins deux bases de données : une base pour effectuer l'apprentissage et une autre pour valider le réseau obtenu et déterminer ses performances. On peut utiliser une troisième base de données, si elle est disponible, pour tester le réseau. Il n'y a pas de règle précise pour déterminer le partage de manière quantitative des données. Il résulte souvent d'un compromis tenant compte du nombre de données dont on dispose, l'essentiel, est que la base de données consacrées à l'apprentissage doit être la plus importante. Il est important de souligner que les sous-ensembles formés à partir de la base de données de départ devront être représentatifs de celle-ci.

II.10.4. Normalisation des données

Avant tout apprentissage, il est indispensable de normaliser et de centrer toutes les variables d'entrée. En effet, si des entrées ont des grandeurs très différentes, celles qui sont petites n'ont pas d'influence sur l'apprentissage. Il existe plusieurs approches de

normalisation des données dont le choix dépend de la nature des données et de la fonction à approximer.

II.10.5. Choix du type de réseau de neurones

Comme il s'agit d'une modélisation statique type boîte noire pour approcher une fonction bornée et continue, la meilleure approche neuronale envisageable est celle des MLP à une seule couche cachée.

II.10.6. Choix du nombre de neurones dans la couche cachée

Il n'existe pas de règle permettant de prévoir à l'avance le nombre de neurones dans la couche cachée. On pourrait penser qu'en plaçant un grand nombre de neurones dans la couche cachée, on obtiendrait systématiquement de meilleurs résultats. Mais il convient de prendre garde d'une part au temps de calcul (plus le nombre de neurones placés dans la couche cachée est important, plus le temps de calcul est long), d'autre part au risque de sur apprentissage (diminue la capacité de généralisation du réseau) ou de sous apprentissage lorsque leur nombre est insuffisant.

D'autre part, nous savons que le nombre des paramètres croît linéairement avec le nombre des variables d'entrées du modèle (parcimonie). Il va falloir donc trouver un compromis entre la parcimonie et le sur ou le sous apprentissage.

II.10.7. Choix de la fonction de transfert

La fonction de transfert neuronale, que l'on appelle aussi fonction d'activation, est la fonction mathématique appliquée à la sortie de chaque neurone. Là encore, il n'existe pas de méthode permettant de prévoir a priori quelle sera la fonction qui donnera les meilleurs résultats. Mais pour les MLP, il faut des fonctions de type sigmoïde pour les neurones de la couche cachée et généralement une fonction linéaire pour le neurone de la couche de sortie.

II.10.8. Estimation des paramètres ou apprentissage

Pour les modélisations statiques au moyen des MLP, il s'agit de faire un apprentissage supervisé.

Le critère d'arrêt de l'apprentissage est souvent calculé à partir d'une fonction de coût, caractérisant l'écart entre les valeurs de sortie obtenues et les valeurs désirées. Ce qui permet un arrêt adéquat de l'apprentissage pour obtenir de bonnes performances de généralisation.

L'obtention des paramètres architecturaux optimaux se fera par comparaison des performances obtenues pour différentes architectures de réseaux de neurones [113].

II.10.9. Validation et généralisation

La validation consiste à tester le réseau après ou au cours de l'apprentissage avec des données proches des données utilisées lors de la phase d'apprentissage. Les résultats de ce test permettront au modèle d'éviter les situations de sous ou de sur apprentissage. En revanche la généralisation s'effectue avec des données tout à fait nouvelles et différentes de la base de données de départ. Elle permet de savoir si le réseau a développé une capacité de généralisation assez performante.

II.11. Modélisation à l'aide de réseaux de neurones

Deux principales stratégies de modélisation qui emploient des réseaux de neurones peuvent être distinguées: la première appelée l'approche par boîte noire, quand le processus entier est représenté avec réseau neuronal approprié, et l'approche hybride qui est une combinaison de la modélisation traditionnelle du processus avec un réseau neuronal qui représente les phénomènes moins connus du processus.

II.11.1. Modèle « boîte noire »

Le terme de « boîte noire » s'oppose aux termes de « modèle de connaissance » ou « modèle de comportement interne » qui désignent un modèle mathématique établi à partir d'une analyse physique (ou chimique, physico-chimique, économique, etc.) du processus que l'on étudie. Ce modèle peut contenir un nombre limité de paramètres ajustables, qui possèdent une signification physique. Nous verrons par la suite que les réseaux de neurones peuvent être utilisés pour l'élaboration de modèle « boîte grise », intermédiaire entre les modèles « boîtes noires » et les modèles de connaissance.

Le modèle « boîte noire » (figure II.27) constitue la forme la plus primitive de modèle mathématique : il est réalisé uniquement à partir de données expérimentales ou d'observations ; il peut avoir une valeur prédictive, dans un certain domaine de validité, mais il n'a aucune

valeur explicative. Ainsi, le modèle de l'univers selon Ptolémée était un modèle « boîte noire » : il ne donnait aucune explication de la marche des astres, mais il permettait de la prédire avec toute la précision souhaitable au regard des instruments de mesure disponibles à l'époque.

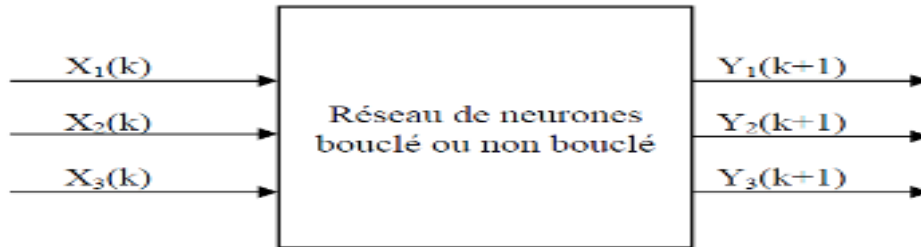


Figure II.27 : Diagramme schématisé d'un modèle neuronal « boîte noire ».

Entre la boîte noire et le modèle de connaissance se situe le modèle semi-physique, ou modèle « boîte grise » (appelé aussi modèle hybride), qui contient à la fois des équations résultant d'une théorie, et des équations purement empiriques, résultant d'une modélisation de type « boîte noire ».

II.11.2. Modèle « boîte grise » ou hybride

Lorsque des connaissances, exprimables sous forme d'équations, sont disponibles, mais insuffisantes pour concevoir un modèle de connaissance satisfaisant, on peut avoir recours à une modélisation "boîte grise" (ou modélisation semi-physique) qui prend en considération à la fois les connaissances et les mesures. Une telle démarche peut concilier les avantages de l'intelligibilité d'un modèle de connaissance avec la souplesse d'un modèle comportant des paramètres ajustables.

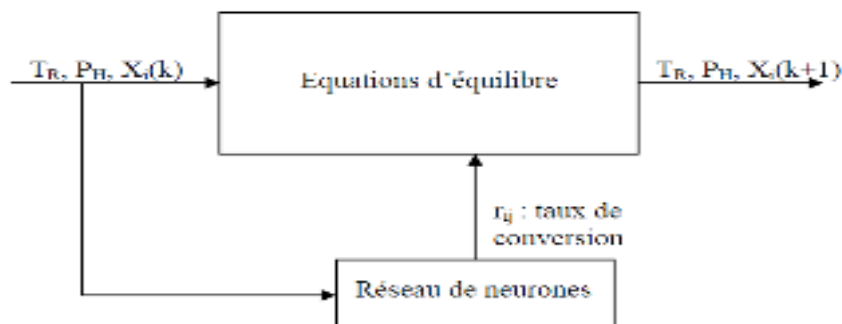


Figure II.28 : Diagramme schématisé de modèle neuronal hybride d'un réacteur chimique.

Un modèle hybride (figure II.28) peut être considéré comme un compromis entre un modèle de connaissance et un modèle « boîte noire ». Il peut prendre en considération toutes

les connaissances que l'ingénieur possède sur le processus, à condition que celles-ci puissent être exprimées par des équations algébriques ou différentielles, et, de surcroît, ce modèle peut utiliser des fonctions paramétrées, dont les paramètres sont déterminés par apprentissage.

Dans la mesure où l'on met en œuvre davantage de connaissances expertes, les données expérimentales nécessaires pour estimer les paramètres d'une manière significative peuvent être en quantité plus réduite.

La conception d'un modèle hybride exige que l'on dispose d'un modèle de connaissance, qui se présente habituellement sous la forme d'un ensemble d'équations algébriques, différentielles, et aux dérivées partielles, non linéaires couplées. Par la suite, on doit procéder à l'apprentissage de ce modèle (ou une partie de celui-ci) à partir de données obtenues par intégration numérique du modèle de connaissance, et de données expérimentales.

Psychogios et Ungar (1992) ont introduit l'idée du premier principe d'un modèle neuronal hybride (modèle boîte grise) [114]. Un tel modèle hybride utilise toute connaissance accessible et possible. Par exemple, habituellement, la modélisation de réacteurs chimiques repose sur les équations de transfert thermique et de masse. Dans le modèle hybride, le réseau de neurones est utilisé pour représenter les éléments inconnus de cette modélisation. Très souvent, la cinétique de la réaction est un élément non suffisamment reconnu par le modèle du réacteur.

II.12. Conclusion

Tout au long de ce chapitre, nous avons tenté de donner un maximum d'informations sur les réseaux de neurones d'une façon générale tout en mettant l'accent sur les réseaux de neurones non bouclés ou statiques, type multicouches (MLP). En effet, après la définition de l'architecture du réseau MLP, vraisemblablement les plus utilisés dans les modélisations statiques du fait qu'ils sont doués de la propriété d'approximation universelle, nous avons présenté un algorithme d'apprentissage très utilisé dans la communauté connexionniste: l'algorithme de rétropropagation du gradient. Plusieurs travaux lui ont été consacrés et un certain nombre de modifications ont été proposées pour accélérer le temps d'apprentissage et améliorer la capacité de généralisation ainsi que les performances du réseau obtenu. Nous avons également présenté dans ce chapitre une procédure plus au moins détaillée avec un organigramme d'optimisation d'une modélisation statique en utilisant les MLP. Cette démarche devrait, nous aider dans le calcul des isothermes d'adsorptions de césium dans le sol qui fera l'objet du chapitre IV.

III.1. Introduction

Le présent chapitre est subdivisé en deux parties : la première est dédiée à la présentation des procédures expérimentales de détermination des paramètres physicochimiques utilisées dans cette modélisation à savoir : La capacité d'échange cationique, la matière organique, la fraction fine, les carbonates et les énergies d'activation. En plus évidemment, de l'isotherme d'adsorption et du coefficient de distribution K_d . Quant à la deuxième partie, elle est consacrée exclusivement à la présentation des démarches méthodologiques de la conception du modèle neuronal.

III.2. Echantillonnage

Les sites d'échantillonnage ont été soigneusement sélectionnés de façon à bien représenter la région d'intérêt. Quatre endroits sont investis correspondant aux 4 points cardinaux du site (Nord, Sud, Est et Ouest), à chaque point et après avoir désherbé l'endroit à la main, le sol est prélevé sur une profondeur allant jusqu'à 40 cm. Les points sélectionnés sont:

- Le point (A), situé à environ 10 km du réacteur à la sortie Ouest de la ville de Birine ;
- Le point (B), situé au Nord-est du site à environ 20 km ;
- Le point (C), situé au Sud-ouest à environ 20 Km du réacteur ;
- Le point (D), situé à proximité du réacteur, à 800 m de l'installation.

III.3. Procédures expérimentales

Dans cette partie, la fraction fine a été déterminée par une analyse granulométrique des sols étudiés. La matière organique et les carbonates ont été obtenues par la méthode de la perte au feu (PAF). Tandis que, les capacités d'échanges cationiques des sols ont été mesurées par saturation à l'acétate d'ammonium au pH neutre. En revanche, les énergies d'activation ont été calculées à l'aide de l'équation (III.11). Nous précisons que, les constantes de vitesse utilisées pour le calcul des énergies d'activation sont celles utilisées dans les travaux de Thèse de Doctorat (en préparation) de Mr A. BOUZIDI (Encadreur de ce travail). Les isothermes d'adsorption et les coefficients de distribution du césium dans les sols étudiés ont été déterminés expérimentalement.

III.3.1. Isotherme d'adsorption

Dans le but de savoir le type d'adsorption du césium dans le sol de la région d'Ain Oussera, nous avons tracé l'isotherme d'adsorption qui met en évidence une relation entre la quantité du césium retenue par unité de masse de sol et sa concentration dans la solution à l'équilibre. L'isotherme est obtenue en faisant varier le poids du sol de 0.01 à 2g. Un ensemble de 31 points ont été utilisés pour le traçage de ces isothermes. Cette manipulation est effectuée à température ambiante (23 °C) ? et aux conditions expérimentales suivantes :

- Poids du sol variable : 0.01, 0.02....., 2 g.
- Concentration du césium qui est de 1 ppm.
- Volume de la phase liquide qui est de 50 ml.
- Temps d'agitation qui est de 30 min.
- pH du milieu : 5.7
- Force ionique : 8.46 $\mu\text{S/cm}$.
- Vitesse d'agitation qui est de 350 min^{-1} .
- Vitesse de centrifugation 3500 rpm.
- Temps de centrifugation 30 minutes.
- Diamètre des pores des membranes de filtration qui est de 0.2 μm .

Les analyses de césium sont effectuées, avant et après mélangeage des phases, à l'aide d'un Spectromètre d'Absorption Atomique (Perkin Elmer AAnalyst 400).

Cette détermination a été effectuée en employant la technique en «Batch» qui consiste à mélanger une quantité de sol avec un volume déterminé de liquide (50 mL), contenant le Cs, pendant un certain temps (temps d'équilibre, 30 minutes) et aux conditions expérimentales arrêtées. Ensuite, les deux phases sont séparées par centrifugation et filtration sous vide et l'élément (Cs) est dosé dans le filtrat.

III.3.2. Détermination de K_d

Le K_d est défini comme étant la concentration (en masse ou en activité) du radionucléide (Cs) dans la phase solide divisée par la concentration du même élément dans la phase liquide à l'équilibre.

$$Kd = \frac{C_s}{C_l} \quad (\text{III.1})$$

Avec :

K_d : Coefficient de distribution en l/kg

C_s : Concentration du césium dans la phase solide en $g\ kg^{-1}$.

C_l : Concentration du césium dans la phase liquide en $g\ L^{-1}$.

C'est à dire :

$$K_d = \frac{\text{g de radionucléide} / \text{kg de sol}}{\text{g de radionucléide} / \text{l de la solution}} \quad (\text{III.2})$$

Le taux (noté R) de Césium retenu par le sol est exprimé par la relation suivante:

$$R = \left(1 - \frac{C_f}{C_0}\right) 100 \quad (\text{III.3})$$

D'où :

$$R = \left(\frac{C_0 - C_f}{C_0}\right) 100 \quad (\text{III.4})$$

C_0 : Concentration initiale de C_s (avant mélange avec le sol) en ppm.

C_f : Concentration finale de C_s (après séparation des phases) en ppm.

R : Le Taux de rétention du césium dans le sol en %.

En concordance avec sa définition, le K_d devient :

$$K_d = \left(\frac{R}{100 - R}\right) \left(\frac{V}{M}\right) \quad (\text{III.5})$$

Où

V : Le volume de la phase liquide en mL.

M : La masse du sol en g.

Selon l'équation (III.4) exprimant R en fonction des concentrations C_0 et C_f , K_d peut s'écrire :

$$K_d = \left[\frac{\left(\frac{C_0 - C_f}{C_0} \right) 100}{100 - \left(\frac{C_0 - C_f}{C_0} \right) 100} \right] \left(\frac{V}{M} \right) \quad (\text{III.6})$$

Après simplifications :

$$K_d = \left(\frac{C_0 - C_f}{C_f} \right) \left(\frac{V}{M} \right) \quad (\text{III.7})$$

Nous avons utilisé les résultats expérimentaux des isothermes pour calculer les valeurs des coefficients de distribution K_d .

III.3.3. Fraction fine

La fraction fine est déterminée sur la base d'une analyse granulométrique effectuée selon le principe de la méthode internationale à la pipette de Robinson [115].

Les sols sont constitués de particules solides organiques et minérales qui s'organisent sous la forme d'agrégats. Du point de vue granulométrique, un sol est assimilable à un assemblage de particules dont la taille varie de plusieurs centimètres à moins d'un micron, les éléments « grossiers » étant la plupart du temps des fragments de roche dont l'altération conduit aux éléments les plus fins, les minéraux argileux [116]. Ces microstructures sont maintenues par des liants colloïdaux constitués de mélange de carbonates, d'oxydes-hydroxydes, d'argiles et de matière organique. Le fractionnement physique d'un sol nécessite donc de casser non seulement les macrostructures, souvent instables, mais aussi les microstructures résistantes, pour ensuite effectuer le classement granulométrique des particules [117]. Ce classement granulométrique appelé encore granulométrie, ou « texture », étudie la répartition des constituants minéraux d'un sol en fonction de leur taille,

indépendamment de leur nature et de leurs propriétés chimiques et minéralogiques. Une des classifications des fractions granulométriques d'un sol, reconnue par une convention internationale, et qui a permis de classer les particules en fonction de leur diamètre, est décrite dans le Tableau III.1:

Tableau III.1 : Echelle granulométrique des constituants du sol, [116]

<i>Cailloux et graviers</i>	<i>Sables</i>	<i>Limons</i>	<i>Argile (ou fraction fine)</i>
Diamètre $\varnothing > 2000 \mu\text{m}$	Sable grossier : $200 < \varnothing < 2000 \mu\text{m}$ Sable fin : $50 < \varnothing < 200 \mu\text{m}$	Limons grossier : $20 < \varnothing < 50 \mu\text{m}$ Limon fin : $2 < \varnothing < 20 \mu\text{m}$	$\varnothing < 2 \mu\text{m}$

Les « argiles », au sens textural du terme, ne contiennent pas seulement les argiles minérales. En effet, cette fraction dont le diamètre de particule est inférieure à $2 \mu\text{m}$ est communément mais improprement appelée « argile », car elle comporte souvent aussi d'autres minéraux tels que des quartzs, des micas, d'autres composés silicatés ainsi que des oxyhydroxydes [118, 116].

III.3.3.1. Préparation des échantillons

Trois prises de chaque sol de 30g chacune, séchées à l'air libre et tamisées à 2 mm, sont utilisées pour cette détermination.

III.3.3.2. Destruction de la matière organique

La matière organique (MO) présente dans les échantillons des sols est détruite par une attaque à l'eau oxygénée (H_2O_2 , 30%). L'opération consiste à mélanger environ 50ml d' H_2O_2 avec 30g de sol sous agitation intermittente et pendant des durées allant jusqu'à 24 heures. L'attaque est renouvelée à froid et à chaud jusqu'à l'obtention d'un surnageant clair et disparition des bulles de gaz. La durée totale de cette étape est de 72 heures.

III.3.3.3. Tamisage

Les sols débarrassés de la MO sont ensuite tamisés à l'eau sur tamis de 50 μ m, on obtient une suspension. La partie des sols, dont le diamètre des particules est supérieur à 50 μ m qui représente le sable (fin et grossier), a été à son tour fractionnée à l'aide des tamis selon les indications portées sur le tableau III.1.

III.3.3.4. Dispersion minérale

La suspension de sol, c'est-à-dire, celui dont les diamètres des particules sont inférieurs à 50 μ m, est ensuite, mélangé avec l'hexamétaphosphate de sodium. Le processus de la dispersion consiste à substituer les cations de valence supérieure avec des cations monovalents pour fractionner les agrégats extrêmement fins. Le schéma de la figure III.1 illustre le mécanisme de cette dispersion. Les sphères en couleur bleue schématisent les grains de la fraction fine.

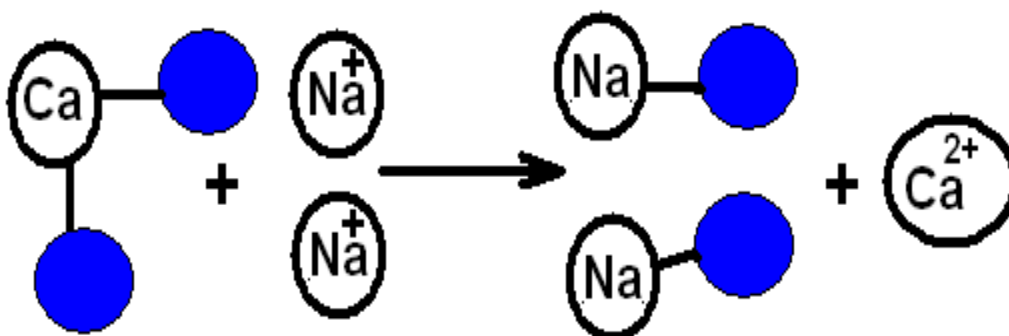


Figure III.1 : Mécanisme de la dispersion minérale de la fraction fine

Après dispersion de la matière minérale, le limon grossier (diamètre compris entre 50 et 20 μ m) est récupéré à son tour par tamisage à 20 μ m. La suspension restante est mise en allonge de sédimentation.

III.3.3.5. Sédimentation et prélèvement de la fraction fine et du limon fin

Selon le principe de la pipette de Robinson [115], la solution de sol est agitée fortement ensuite laisser sédimenter. Au temps t , prélever à hauteur de 10 cm une aliquote V de 25 ml, qui est ensuite évaporée dans une boîte à tare.

A la température ambiante, environ 20 °C

A $t = 4$ minutes 48 secondes, l'aliquote prélevée contient du limon fin et la fraction fine (LF + FF), autrement dit, les particules dont les diamètres sont compris entre 0 et 20 μm .

A $t = 8$ heures, elle contient uniquement de la fraction fine (FF), diamètre compris entre 0 et 2 μm .

Calcul :

Le pourcentage de la fraction fine :

Soit : P_2 le poids de LF + FF, P_1 le poids de FF et $P_0 = 30\text{g}$ (le poids de l'échantillon de sol au départ)

$$\%FF = P_1 \times \frac{1000}{V} \times \frac{100}{P_0} \quad (\text{III.8})$$

$$\%LF = \frac{P_2 - P_1}{VP_0} 10^5 \quad (\text{III.9})$$

Le tableau suivant récapitule les vitesses de sédimentation des différentes fractions en fonction de la température :

Tableau III.2 : Vitesse de sédimentation [119]

Température C°	Temps de chute pour 10 cm en heures minutes	Argiles ($d \leq 2 \mu\text{m}$)			Limons (2 à 20 μm) Temps de chute pour 10cm en min secondes
		Profondeur de décantation de prélèvement en cm après			
		8 heures	7 heures	6 heures	
10	10.23	7.7	6.7	5.8	6.14
11	10.06	7.9	6.9	5.9	6.03
12	9.49	8.1	7.1	6.1	5.54
13	9.34	8.4	7.3	6.3	5.44
14	9.19	8.6	7.5	6.4	5.35
15	9.05	8.8	7.7	6.6	5.27
16	8.51	9.0	7.9	6.8	5.19
17	8.37	9.3	8.1	7.0	5.10
18	8.24	9.5	8.3	7.2	5.03
19	8.12	9.8	8.6	7.3	4.55
20	8.00	10.0	8.8	7.7	4.48
21	7.48	10.3	9.0	7.5	4.41
22	7.37	10.5	9.2	7.9	4.34
23	7.26	10.8	9.4	8.1	4.28
24	7.16	11.0	9.7	8.3	4.22
25	7.06	11.3	9.9	8.5	4.15
26	6.56	11.5	10.1	8.7	4.10
27	6.47	11.8	10.3	8.9	4.04
28	6.38	12.1	10.6	9.1	3.59
29	6.29	12.3	10.8	9.3	3.54
30	6.21	12.6	11.0	9.5	3.48

III.3.4. Matière organique et carbonates

Pour la détermination de ces deux paramètres, nous avons utilisé la technique de la perte au feu (PAF). Elle consiste à porter un échantillon de sol successivement à des températures élevées et à chaque fois mesurer la différence de poids pour déterminer tel ou tel paramètre.

Quatre creusets en porcelaine sont nettoyés et mis dans un four à moufle porté à 1000°C pendant 5 heures, ensuite refroidis dans un dessiccateur. Le but de cette manipulation est de débarrasser ces creusets de toutes matières calcinables qui pourraient interférer lors de la détermination de la matière organique ou des carbonates. Une fois refroidis, ils sont pesés et réintroduits dans le dessiccateur à l'abri de l'humidité.

III.3.4.1. Matière organique

Trois échantillons de chaque sol de 10g chacun sont mis dans les creusets en porcelaine, un quatrième creuset est intentionnellement laissé vide, comme témoin. Nous précisons que, ces échantillons ont été préalablement débarrassés de l'humidité en les portant à 150°C dans une étuve pendant 16 heures. Les creusets contenant le sol sont pesés et leurs poids notés. L'ensemble est introduit dans un four à moufle porté à 375 °C pendant 16 heures. Après refroidissement dans le dessiccateur, les creusets sont à nouveau pesés, y compris le creuset vide pour estimer les fluctuations de son poids. La différence de poids avant et après étuvage nous fournit le taux de la matière organique.

$$\%MO = \frac{(p_0 - p_1) - (p_i - p_f)}{p_0} \times 100 \quad (\text{III.10})$$

MO : Taux de la matière organique, %

P₀ : Poids du sol initial, g

P₁ : Poids du sol après calcination, g

P_i : Poids du creuset (témoin) avant calcination, g

P_f : Poids du creuset (témoin) après calcination, g

III.3.4.2. Carbonates

Les carbonates sont, à leurs tours, déterminées selon la même procédure que les matières organiques avec une température du four à moufle portée à 980°C et ce pour une durée de 5 heures.

III.3.5. Capacité d'échange cationique

La capacité d'échange cationique (CEC) est liée à la teneur en argile et en matière organique dans le sol. Le principe de mesure de la CEC consiste à déterminer approximativement tous les sites d'adsorption qu'offre un solide. Dans un premier temps, nous avons mélangé 1g de sol avec 50 ml de $MgCl_2$ 0.5 M pendant 01 heure avec agitation douce. Ensuite, nous avons séparé les deux phases par centrifugation et filtration. Pour éliminer le sel de magnésium libre dans le sol, nous avons effectué plusieurs lavages à l'alcool éthylique. Pour déterminer la CEC, nous avons dû extraire le Mg lié aux sites d'adsorption sur le sol et le mesurer, pour ce faire, nous avons utilisé une solution concentrée d'acétate d'ammonium 1M. La CEC est exprimée en meq de mg/100g de sol [115].

La détermination de la capacité d'échange cationique du sol a été effectuée au laboratoire d'analyse chimique du Centre de Recherche Nucléaire de Birine (CRNB). Les analyses du magnésium ont été effectuées à l'aide d'un spectromètre d'absorption atomique.

III.3.6. L'énergie d'activation

Les énergies d'activation ont été calculées sur la base de l'expression suivante :

$$\ln \frac{K_2(T_2)}{K_2(T_1)} = \frac{E_a}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \quad (\text{III.11})$$

Où R est la constante des gaz parfaits ($8.314 \text{ J mol}^{-1} \text{ K}^{-1}$) et K_2 est la constante de vitesse à la température T_1 et T_2 . Les valeurs de ces constantes sont regroupées dans le

tableau III.3. Rappelons que, l'énergie d'activation peut être conçue comme étant l'énergie cinétique minimale requise pour qu'une réaction donnée se produise.

Tableau III.3 : Paramètres cinétiques pour les quatre sols

Sols	T °C	K_2 (g.mg ⁻¹ .min ⁻¹)
A	23	4.174
	60	6.759
B	23	6.420
	60	9.460
C	23	2.225
	60	3.580
D	23	5.570
	60	9.260

III.4. Méthodologie de la conception des réseaux de neurones

Cette partie de ce chapitre sera donc consacrée à la description exhaustive de la méthodologie de conception du réseau de neurones pour la modélisation de nos isothermes.

Les réseaux de neurones multicouches de type non bouclés (MLP), dont l'apprentissage est supervisé constituent une technique relativement récente de modélisation (boîte noire) qui a montré ses performances. Il suffit d'avoir un ensemble de variables mesurées et d'un ensemble de mesures d'une grandeur de nature quelconque et en supposant qu'il existe une relation suffisamment régulière entre les variables et la grandeur alors, il est possible d'approximer cette relation par un réseau de neurone non bouclé [94]. C'est parfaitement ce que nous estimons avoir.

La procédure de conception du RNA type (MLP) peut être schématisée conformément à l'organigramme suivant :

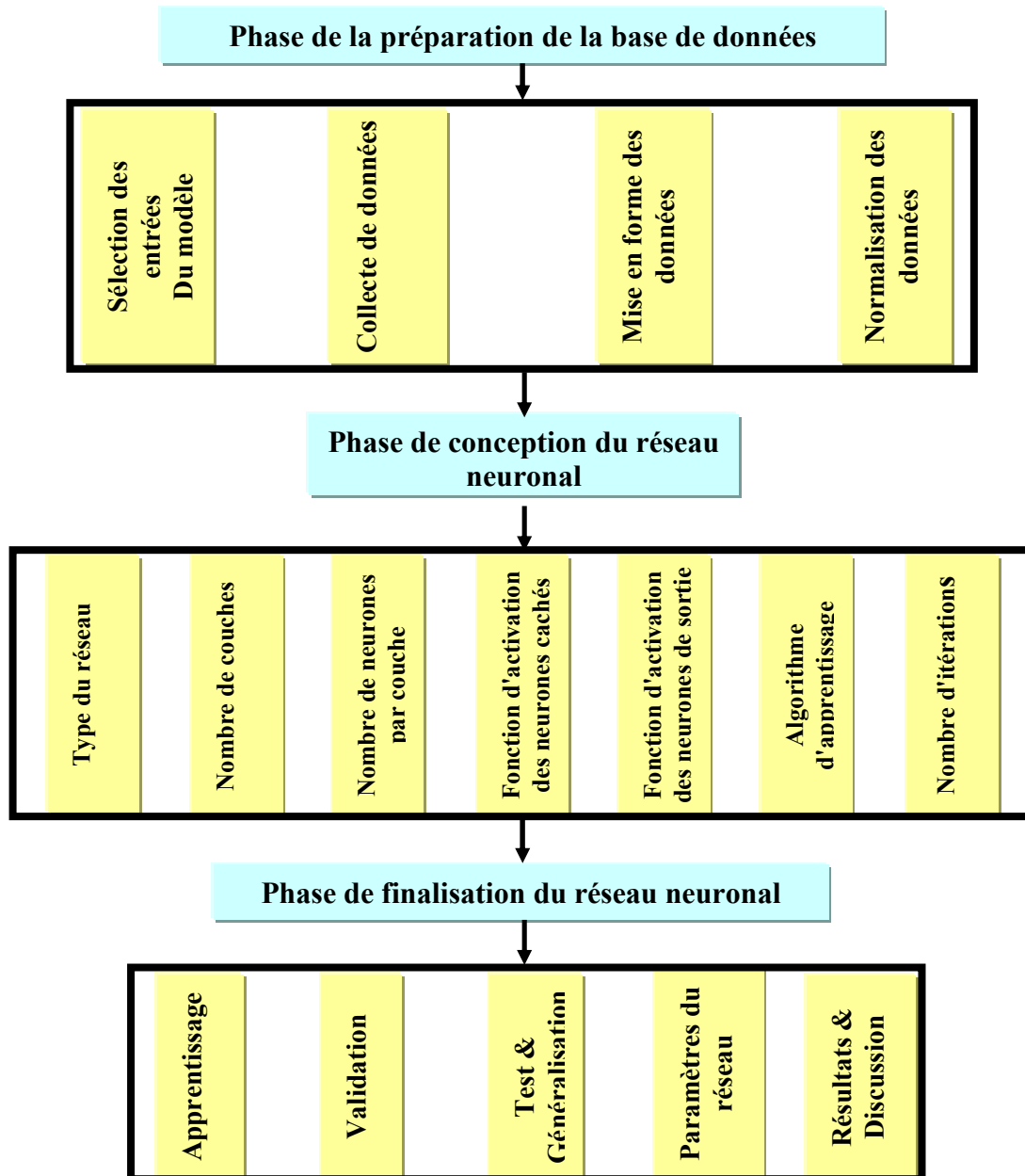


Figure III.2: Méthodologie de calcul des différents paramètres par les réseaux de neurones

III.4.1. Base de données

L'élaboration d'une base de données est une phase décisive dans la conception d'un modèle neuronal. Elle doit être exécutée avec la plus grande précaution car toute erreur à ce stade, va connaître des influences importantes tant sur la convergence du modèle que sur la généralisation.

Dans le but de constituer une base de données fiable et représentative, nous avons suivi une méthodologie, inspirée de l'étude bibliographique [94], qui revêt les étapes suivantes :

- sélection des entrées du modèle neuronal,
- collecte des données,
- mise en forme des données,
- normalisation des données.

III.4.1.1. Sélection des entrées et des sorties du modèle neuronal

La sélection des entrées consiste à déterminer les variables pertinentes par rapport à la grandeur à modéliser. Elle vise deux points essentiels [94] :

- réduction de la dimension de l'espace de représentation des variables du modèle,
- élimination des entrées qui n'ont pas ou peu d'influence sur la sortie.

Il est important de mentionner que chaque ajout d'une variable encombre le modèle et entraîne une multiplication des paramètres du réseau. Raison pour laquelle, la représentation des entrées doit contenir tous les facteurs pertinents par rapport à la grandeur à modéliser, mais seulement ceux-ci.

Dans notre cas, nous avons abouti aux grandeurs mesurables qui sont : la concentration du soluté dans la phase aqueuse (C_e), taux de la matière organique (MO), taux des carbonates (C_o) et taux de la fraction fine (FF), dont on a utilisé comme entrées de notre modèle. Ces entrées sont utilisées dans un réseau à quatre sorties qui sont : la quantité de soluté retenue par le solide (Q_e), la capacité d'échange cationique (CEC), le coefficient de distribution (K_d) et l'énergie d'activation (E_a)

Sachant que chaque paramètre d'entrée et de sortie de notre modèle est lié à l'autre, par exemple, la capacité d'échange cationique (CEC) est liée à la teneur en argile et en matière organique dans le sol tandis que le coefficient de distribution est définie comme étant le rapport entre la concentration du césium dans la phase solide et la phase liquide....etc. Alors nous pouvons dire que tous les paramètres que nous avons utilisés dans ce modèle ont des rapports de dépendance entre eux.

III.4.1.2. Collecte des données

Une fois les entrées pertinentes du modèle sont sélectionnées, l'objectif de cette étape sera de les recueillir. La base de données ainsi constituée doit répondre à deux exigences [94] à savoir :

- contenir le maximum d'échantillons ou d'exemples (génériques et/ou expérimentaux),
- être représentative des données susceptibles d'intervenir en phase d'utilisation du modèle neuronal.

La base de données que nous avons pu constituer comporte **124 exemples** et ce pour les **4 sols** (A, B, C et D) étudiés.

III.4.1.3. Mise en forme des données

Dans cette étape les données sont d'abord subdivisées en trois sous ensembles ; un ensemble pour l'apprentissage et deux autres sont réservés à la validation et la généralisation. L'échantillonnage des trois ensembles est fait de façon qu'ils reflètent parfaitement la base de données de départ. Ensuite, l'ensemble des données est écrit sous forme matricielle. Comme il s'agit là d'un apprentissage supervisé, on a donc des matrices de quatre lignes représentant les 4 entrées du modèle et des matrices également de 4 lignes représentant les 4 sorties désirées.

La matrice utilisée pour l'apprentissage comporte **86%** de la totalité de la base de données, soit environs **107** (exemples). Les **14%** restantes sont divisées équitablement entre la validation et la généralisation.

III.4.1.4. Normalisation des données

La normalisation est une étape importante dans le processus de l'élaboration des données. Elle devient dans la plupart des cas une exigence pour les matrices des inputs, du fait qu'on utilise souvent des fonctions de transfert bornées de type sigmoïdes dans les modélisations statiques.

Dans notre travail, on a utilisé l'instruction « normc » en langage Matlab qui effectue une normalisation des colonnes de la matrice des inputs selon l'expression suivante [94] :

$$x_{ij}^N (\text{normalisé}) = \frac{x_{ij}}{\sqrt{\sum_{i=1}^5 x_{ij}^2}} \quad (\text{III.12})$$

La matrice de sortie n'est pas normalisée de la même façon que celle des entrées, du fait que les fonctions de transfert attribuées aux neurones de la couche de sortie sont des fonctions identités. Mais vu l'écart plus au moins important observé entre les valeurs de cette matrice, nous l'avons transformé en logarithmique pour compresser ces écarts.

III.4.2. Phase de conception du réseau neuronal

La conception d'un modèle neuronal consiste à faire une étude d'évaluation des éléments constitutifs du réseau en fonction de la performance de modélisation souhaitée.

Elle vise à fixer les éléments suivants :

1. le type du réseau,
2. le nombre de couches cachées,
3. le nombre de neurones dans les couches cachées,
4. la fonction de transfert des neurones,
5. l'algorithme d'apprentissage et
6. le nombre d'itérations (cycles).

Dans notre cas, il s'agit d'approximer l'isotherme d'adsorption de césium par un réseau de neurones. Il est donc question, de faire une modélisation statique indépendamment du temps, le réseau adéquat au vu de la monographie consultée [86, 120, 121, 122, 123, 124, 125, 126, 127, 128, 94] ne peut être en général qu'un réseau non bouclé type MLP (Multi Layer Perceptron). La propriété fondamentale de ces réseaux est l'approximation universelle, qui s'énonce de la façon suivante (théorème d'existence):

*« 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** [94, 84] ».*

III.4.2.1. Choix du type de réseau de neurones

Comme il s'agit d'une modélisation statique type boîte noire pour approcher quatre paramètres représentant des fonctions bornées et continues, la meilleure approche neuronale envisageable selon l'énoncé précédent [94, 84] est celle d'un MLP.

II.4.2.2. Choix du nombre des couches cachées

Evidemment quant il s'agit d'approcher une seule sortie, on utilise un MLP à une seule couche cachée, mais dans notre cas, nous avons à modéliser simultanément quatre sorties. Pour cette raison et pour des contraintes liées à la convergence, notre réseau aura besoin de plus d'une couche cachée pour pouvoir satisfaire une approximation acceptable.

III.4.2.3. Choix du nombre de neurones dans la couche cachée

Il n'existe pas de règle permettant de prévoir à l'avance le nombre de neurones dans la couche cachée. On pourrait penser qu'en plaçant un grand nombre de neurones dans la couche cachée, on obtiendrait systématiquement de meilleurs résultats. Mais il convient de prendre garde d'une part au temps de calcul (plus le nombre de neurones placés dans la couche cachée est important, plus le temps de calcul est important), d'autre part au risque de sur-apprentissage (diminue la capacité de généralisation du réseau) ou de sous-apprentissage lorsque leur nombre est insuffisant.

D'autre part, nous savons que le nombre des paramètres croît linéairement avec le nombre des variables d'entrées du modèle (parcimonie). Il va falloir donc de trouver un compromis entre la parcimonie et le sur ou le sous-apprentissage.

III.4.2.4. Choix de la fonction de transfert

La fonction de transfert neuronale, que l'on appelle aussi fonction d'activation, est la fonction mathématique appliquée à la sortie de chaque neurone. Là encore, il n'existe pas de méthode permettant de prévoir a priori quelle sera la fonction qui donnera les meilleurs résultats. Mais pour les MLP, il faut des fonctions de type sigmoïde pour les neurones de la couche cachée et généralement une fonction linéaire pour le neurone de la couche de sortie.

III.4.2.5. Algorithme d'apprentissage

Un algorithme d'apprentissage permet de déterminer le meilleur ensemble de paramètres du modèle (poids et biais) pour le problème considéré. Trouver l'ensemble optimal est souvent un compromis entre temps de calcul, minimisation de l'erreur et

conservation de la capacité de généralisation pour le réseau. L'algorithme d'apprentissage utilisé pour les MLP est dit de rétropropagation du gradient ou de l'erreur qui fournit un moyen simple d'entraîner les neurones des couches cachées.

III.4.2.6. Nombre d'itérations (cycles)

Le nombre d'itération est également un paramètre clé pour la convergence du réseau. Un nombre trop élevé de cycle ou d'itération peut conduire à un sur-apprentissage, par contre, un moins nombre peut induire au sous-apprentissage. Pour cette raison ce paramètre devra également faire l'objet d'une recherche d'un optimum.

III.4.3. Phase de finalisation du réseau neuronal

La phase de finalisation du réseau neuronal est la plus importante étape car elle constitue la phase de conception du réseau. En effet, elle comprend l'apprentissage qui définit le réseau neuronal, la validation et le test de généralisation.

III.4.3.1. Apprentissage

On appelle apprentissage des réseaux de neurones la procédure qui consiste à estimer les paramètres (poids et biais) des neurones du réseau, afin que celui-ci remplisse au mieux la tâche qui lui est affectée [123]. L'apprentissage est une phase du développement du réseau de neurones durant laquelle on calcule les poids des neurones de telle manière que les sorties du réseau soient aussi proches que possible des sorties désirées.

III.4.3.2. Validation

La validation consiste à tester le réseau au cours de l'apprentissage avec des données proches des données utilisées en apprentissage. Les résultats de ce test permettront au modèle d'éviter les situations de sous ou de sur-apprentissage.

III.4.3.3. Généralisation

La généralisation s'effectue avec des données tout à fait nouvelles et différentes de la base de données de départ. Elle permet de savoir si le réseau a développé une capacité de généralisation assez performante.

III.4.3.4. Estimation des paramètres du réseau

Pour les modélisations statiques au moyen des MLP, il s'agit de faire un apprentissage supervisé.

Le critère d'arrêt de l'apprentissage est souvent calculé à partir d'une fonction de coût, caractérisant l'écart entre les valeurs de sortie obtenues et les valeurs désirées. Ce qui permet un arrêt adéquat de l'apprentissage pour obtenir de bonnes performances de généralisation.

L'obtention des paramètres (poids et biais) et l'architecture optimale se fera par comparaison des performances obtenues pour différentes architectures de réseaux de neurones [129].

III.4.3.5. Résultats et discussion

Cette étape couronne la procédure de conception et de finalisation du réseau neuronal. Les résultats obtenus à l'issue de cette modélisation feront l'objet de comparaisons et de discussions.

III.5. Conclusion

Tout au long de ce chapitre, nous avons tenté de donner un maximum d'informations d'une part, sur les procédures expérimentales pour la détermination des paramètres physicochimiques utilisés dans cette modélisation et d'autre part, sur la procédure, plus au moins détaillée, d'une modélisation statique en utilisant les MLP pour approcher les quatre sorties visées dans cette étude.

IV.1. Introduction

Ce chapitre est consacré à la présentation des résultats obtenus à l'issue de cette étude. Nous présentons d'abord les résultats liés au tracé des isothermes ainsi que les différents paramètres physicochimiques utilisés dans cette modélisation. Ensuite, nous exposons les résultats concernant la modélisation des isothermes d'adsorption des sols étudiés et les paramètres de sortie qui se calculent en parallèle (coefficient de distribution, l'énergie d'activation et la capacité d'échange cationique). Le modèle neuronal conçu fera l'objet d'une confrontation à d'autres modèles théoriques.

IV.2. Résultats liés aux travaux expérimentaux

IV.2.1. Isothermes

Dans le cadre de ce travail, nous avons tracé les isothermes du césium pour les 4 sols étudiés. Rappelons que les isothermes sont obtenues en variant la quantité de chaque sol de 0.01 à 2g. Les figures IV.1, IV.2, IV.3 et IV.4 présentent respectivement les isothermes du césium pour les sols (A), (B), (C) et (D) de la région de Birine.

D'après leurs allures, nous pouvons dire que l'isotherme du sol A est proche de la classe II selon la classification de BET. Rappelons que cette classe d'isothermes correspond en général à l'adsorption multicouche sur des surfaces ouvertes. Qui peut aussi résulter d'une somme d'isothermes I + III (remplissage de micropores suivi d'une adsorption multicouche sur une surface externe). L'isotherme du sol C correspond plutôt au type I, typique d'une adsorption en monocouche sur support microporeux, où la monocouche d'adsorbat suffit de remplir jusqu'à saturation tous les micropores. Cependant vers la fin de cette isotherme, on ne voit pas clairement l'établissement du plateau indiquant la saturation des tous les sites d'adsorption, comme stipule cette classe. Ce constat laisse supposer que les interactions adsorbat-adsorbant ne perdent pas d'affinité et conduisent probablement à une superposition de couches rapprochant ainsi l'isotherme plutôt à la classe V. En revanche, celles des sols B et D appartiennent à la classe III qui reflète une faible interaction adsorbat-adsorbant comparée aux interactions adsorbat-adsorbant significativement plus élevées.

Le but de cette partie n'est pas de s'étaler sur les explications de la nature des isothermes mais plutôt de fournir les données nécessaires qui serviront dans la modélisation. Rappelons que les valeurs de C_e et Q_e exprimées en mg/l et mg/g vont être

utilisées respectivement comme input et output de réseau neuronal. Pour cette fin, nous avons dressé des tableaux de ces deux paramètres en annexe A.

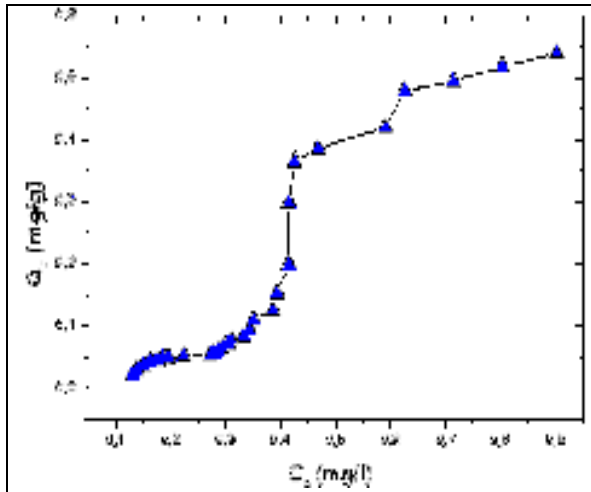


Figure IV.1 : Isotherme d'adsorption du Cs sur le sol (A)

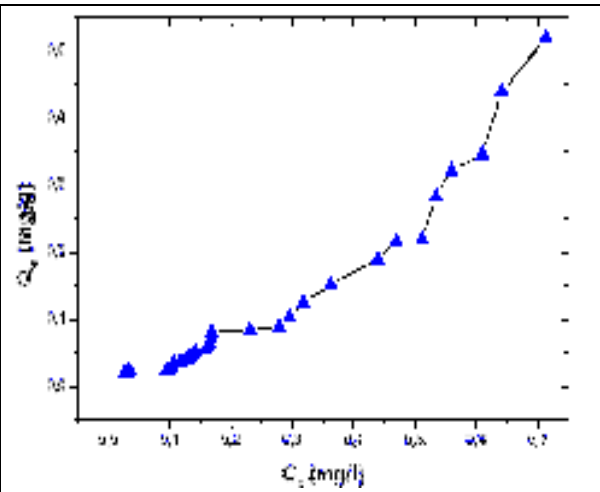


Figure IV.2 : Isotherme d'adsorption du Cs sur le sol (B)

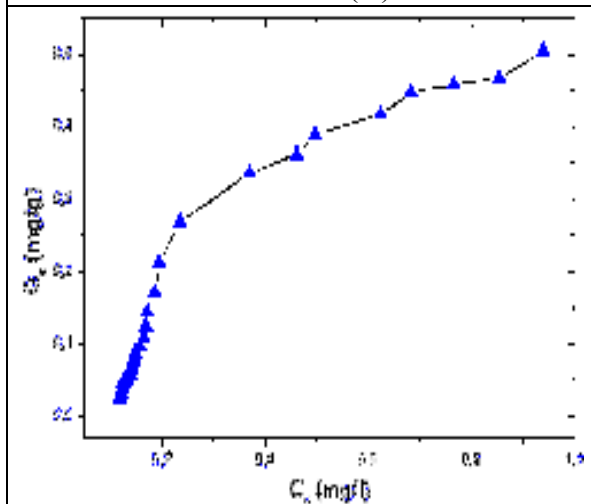


Figure IV.3 : Isotherme d'adsorption du Cs sur le sol (C)

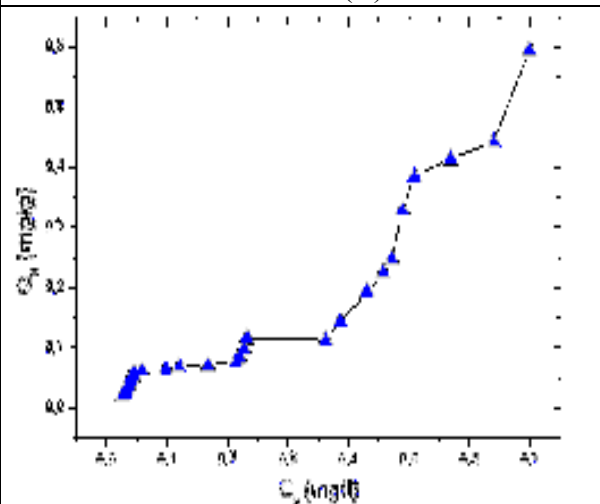


Figure IV.4 : Isotherme d'adsorption du Cs sur le sol (D)

IV.2.2. Détermination de Kd

Le coefficient de distribution ou Kd est défini, à l'équilibre, comme le rapport entre la concentration d'un polluant associé à la phase solide (sol) à sa concentration en phase liquide (solution du sol) équation (III.5) :

$$Kd = \frac{C_s}{C_l}$$

Ce coefficient est calculé en utilisant les données expérimentales des isothermes des quatre sols. Ces valeurs sont également présentées en annexe A. Elles vont nous servir

pour vérifier le calcul du réseau de neurones, puisque comme nous l'avons déjà indiqué plus haut dans le chapitre III, les réseaux de neurones type MLP sont à apprentissage supervisé.

IV.2.3. Fraction fine

La fraction fine est déterminée sur la base d'une analyse granulométrique effectuée selon le principe de la méthode internationale à la pipette de Robinson [115]. Les résultats sont regroupés dans le tableau IV.1. Nous précisons que ces résultats présentent la moyenne de trois répétitions effectuées pour chaque type de sol.

Tableau IV.1 : Pourcentage de la fraction fine pour chaque sol.

<i>Sols</i>	<i>Argile (%)</i>
<i>A</i>	14,73
<i>B</i>	16,90
<i>C</i>	24,75
<i>D</i>	12,57

Les fractions fines sont utilisées comme un paramètre d'entrée dans notre réseau neuronal.

IV.2.4. Matière organique et carbonates

Nous avons utilisé la technique de la perte au feu (PAF) pour déterminer successivement des paramètres suivants : L'humidité, la matière organique et les carbonates. La procédure de cette méthode est détaillée dans le chapitre précédent. Les résultats de la matière organique et des carbonates, exprimés en %, sont présentés dans le tableau suivant :

Tableau IV.2 : Détermination de la matière organique et des carbonates par la PAF.

<i>Sols</i>	<i>Matière organique MO (%)</i>	<i>Carbonates (%)</i>
<i>A</i>	1.56	27.09
<i>B</i>	2.14	18.53
<i>C</i>	1.61	32.63
<i>D</i>	0.62	13.99

Ces deux paramètres sont utilisés comme entrées du modèle neuronale.

IV.2.5. Capacité d'échange cationique

La détermination de la CEC des quatre types de sol a été effectuée comme précisé dans le chapitre précédent par saturation à l'ion de magnésium ($MgCl_2$). Les résultats, de cette analyse, sont présentés dans le tableau suivant :

Tableau IV.3 : Valeurs de la CEC pour chaque sol.

<i>Sols</i>	<i>CEC (meq/100g)</i>
<i>A</i>	14.06
<i>B</i>	14.22
<i>C</i>	11.62
<i>D</i>	8.31

Nous précisons que la CEC est un paramètre de sortie du modèle.

IV.2.6. L'énergie d'activation

Les valeurs des énergies d'activation ont été calculées sur la base de l'expression III.11. Ces valeurs sont regroupées dans le tableau IV.4.

Tableau IV.4 : Valeurs des énergies d'activation des quatre sols.

<i>Sols</i>	<i>E_a (kJ/mol)</i>
<i>A</i>	10.67
<i>B</i>	8.58
<i>C</i>	10.53
<i>D</i>	11.26

Ce paramètre est aussi une sortie de notre modèle.

IV.3. Résultats liés à la conception

IV.3.1. Conception du réseau

Pour la conception du réseau il existe deux types de paramètres : ceux qui sont ajustables moyennant une étude expérimentale et ceux qui peuvent directement découler des propriétés des réseaux eux-mêmes comme énoncé dans le théorème d'existante.

Les paramètres qui découlent des propriétés des réseaux sont :

- Le type de réseau

- L'algorithme d'apprentissage
- La fonction de transfert

Pour notre réseau il est de type MLP à apprentissage supervisé. L'algorithme est celui de rétropropagation du gradient de « Levenberg-Marquardt ». Les fonctions de transfert sont de deux types : tangente sigmoïdes pour les neurones cachés et fonctions d'identité pour les neurones de sortie.

Les paramètres ajustables sont :

- Nombre de couches cachées
- Nombre de neurones dans les couches cachées
- Nombre d'itérations ou de cycles d'apprentissage

Dans ce qui suit, nous allons présenter les démarches expérimentales suivies pour leurs déterminations.

IV.3.2. Nombre de couches cachées

Comme nous l'avons mentionné dans le chapitre précédent, notre réseau aura besoin de plus d'une couche cachée pour pouvoir satisfaire une approximation acceptable. Effectivement c'est le cas, la figure ci-dessous IV.5, qui exprime la variation de l'erreur relative en fonction du nombre de couches, montre que le point optimum satisfaisant à la fois l'apprentissage, la validation et la généralisation est situé à deux couches.

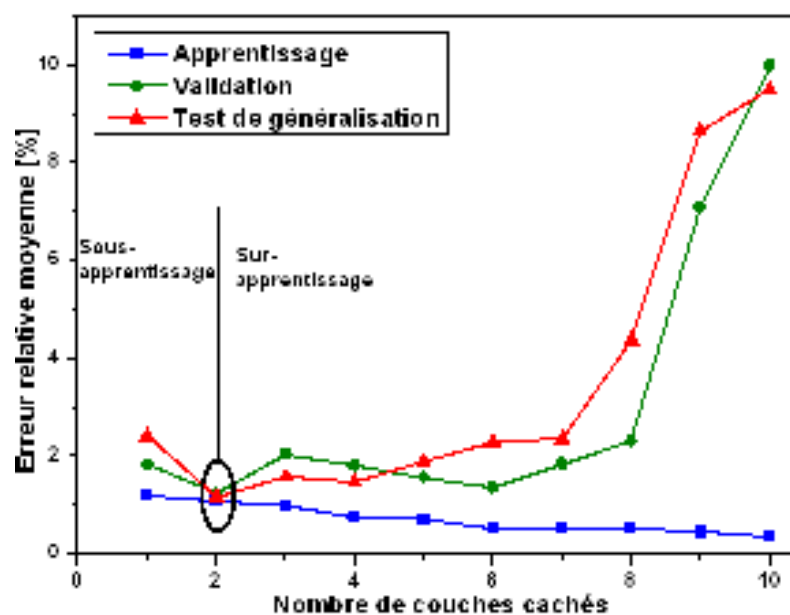


Figure IV.5 : Evaluation du nombre de couches cachées.

IV.3.3. Nombre de neurones cachés

Le nombre de neurones dans la couche cachée demeure l'un des paramètres clés de la convergence du réseau de neurones. Il ne dépend pas uniquement du nombre des entrées mais aussi de la complexité de la fonction à approximer [94, 130]. Pour fixer ce nombre, il faut chercher un compromis entre la convergence lors de l'apprentissage et une bonne généralisation. En effet, si le nombre de neurones cachés est trop faible, le réseau ne convergera pas, il aura trop de données à apprendre et pas assez de neurones pour pondérer, on dit qu'on est en phase de sous apprentissage. En revanche, si le réseau a trop de neurones, il convergera parfaitement, mais en dehors des points d'apprentissage, les réponses calculées seront médiocres, on parle alors de sur apprentissage. Pour résoudre ce dilemme, nous avons varié le nombre de neurones cachés et nous avons suivi la répercussion sur la performance de l'apprentissage et de la généralisation.

Suite à cette évaluation, nous avons parvenu au nombre de 30 neurones, 10 neurones pour la première couche cachée et 20 neurones pour la seconde. Ce nombre semble être le point qui satisfait à la fois la convergence et la généralisation comme le montre la figure IV.6.

Par contre, la couche de sortie comporte quatre neurones, correspondant aux quatre grandeurs à calculer.

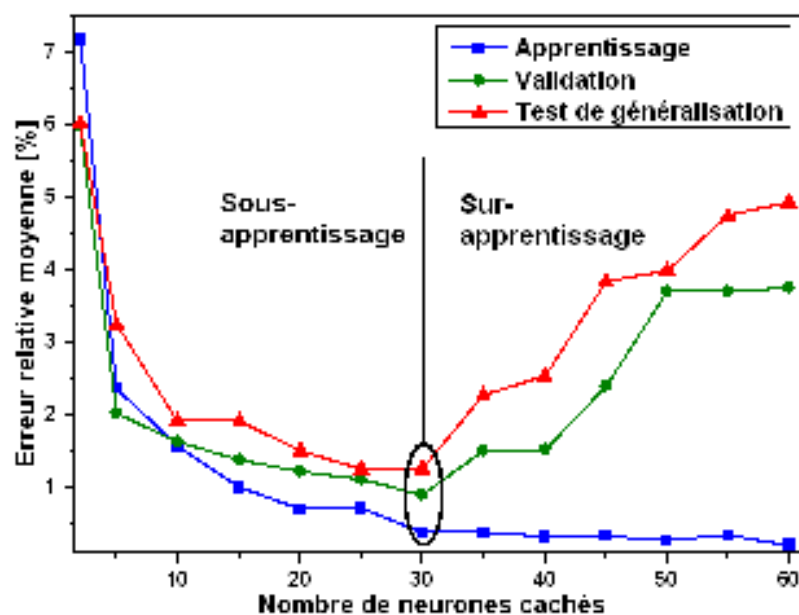


Figure IV.6 : Evaluation du nombre de neurones de la couche cachée.

IV.3.4. Effet du nombre d'itération

Le même problème de compromis est rencontré dans la détermination du nombre de cycles lors de l'apprentissage. Dans ce cas, il faut travailler avec un nombre de cycle suffisamment élevé pour pouvoir localiser le point qui donne une bonne performance d'apprentissage et une généralisation acceptable. La figure IV.7 montre cette évaluation du nombre d'itération dont l'optimum correspond à la valeur de 1300 cycles.

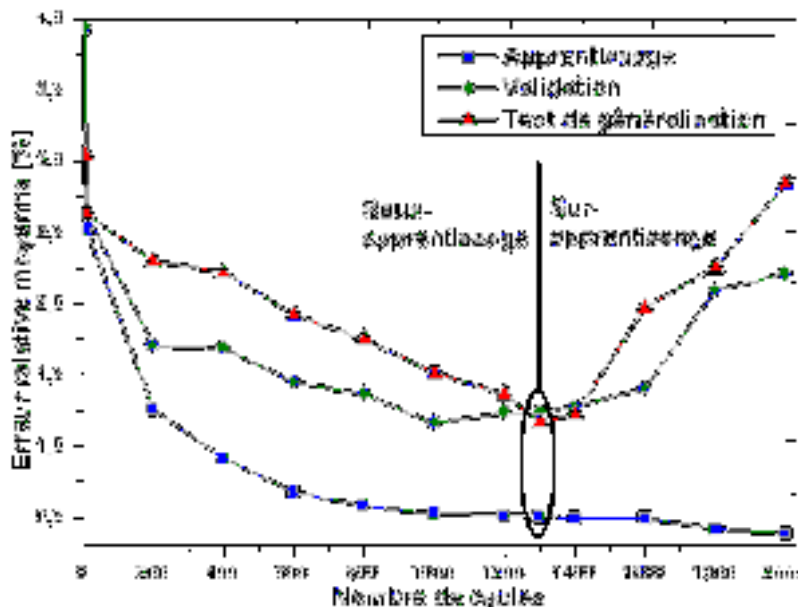


Figure IV.7: Evaluation du nombre de cycles d'apprentissage.

IV.3.5. Le réseau neuronal conçu

Après avoir arrêté les paramètres structuraux de notre réseau neuronal utilisé pour la modélisation des isothermes d'adsorption de césium, nous les avons regroupés dans le tableau IV.5. La figure IV.8 présente l'architecture du réseau conçu pour la modélisation de nos isothermes.

Tableau IV.5 : Structure du réseau neuronal utilisé dans la modélisation des isothermes d’adsorption.

Type du réseau	Nbr. de couches cachées	Nbr. de neurones dans les couches cachées	Fonction de transfert des neurones des couches cachées	Fonction de transfert des neurones de sortie	Algorithme d’apprentissage	Nombre d’itération ou de cycles
MLP à 3 couches	2 couches cachées	1 ^{ère} : 10 2 ^{ème} : 20	Sigmoïde Tangentielle « TANSIG »	Fonction identité « PURELIN »	Levenberg-Marquardt	1300

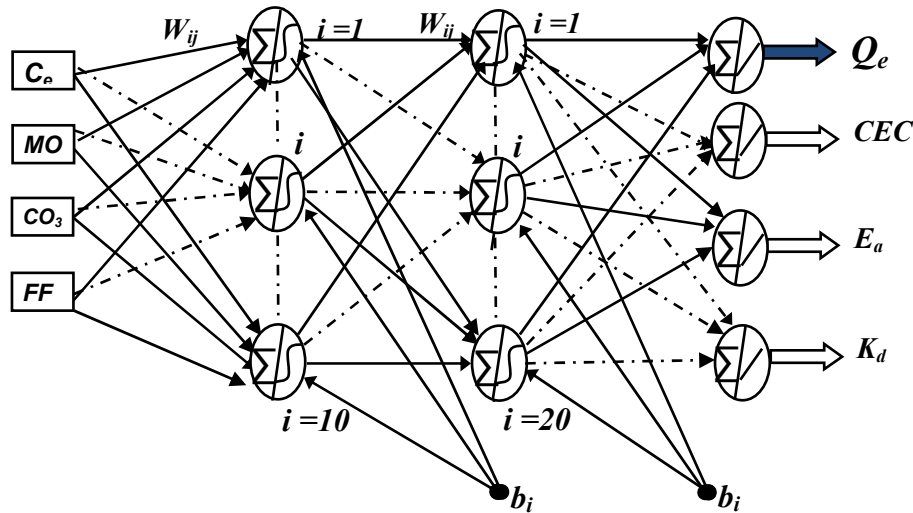


Figure IV.8 : Schéma simplifié de l’architecture du réseau neuronal avec la fonction de transfert globale.

IV.3.6. Paramètres du réseau

Une fois le nombre de neurones cachés et le nombre de cycles d’apprentissage sont fixés (après apprentissage et validation), nous pouvons donc extraire les paramètres du réseau neuronal, c’est-à-dire l’ensemble de ses poids synaptiques et ses biais. Le tableau B.1, présenté en annexe B, regroupe ces paramètres pour le réseau finalisé.

IV.4. Résultats et discussion

Le modèle développé dans cette étude est à quatre sorties (Q_e , K_d , CEC et E_a), de ce fait, nous présentons les résultats de chaque sortie séparément. Nous commençons par les

résultats obtenus lors des phases de conception du modèle neuronal, c'est-à-dire la phase d'apprentissage, de validation et de généralisation.

IV.4.1. La première sortie (Q_e)

Les figures (IV.9) et (IV.10) et (IV.11) présentent respectivement les résultats du calcul des valeurs de Q_e pour l'apprentissage, la validation et le test de généralisation. Nous remarquons que les valeurs moyennes des erreurs relatives par rapport aux données utilisées (apprentissage, validation et généralisation) ne dépassent pas les 1.24%. Le coefficient de corrélation R de la courbe de linéarisation pour les trois sous-ensembles est proche de l'unité (environ 0.999).

Le modèle neuronal reproduit également d'une façon acceptable les données qui ont servi à l'apprentissage, à la validation et au test de généralisation. En effet, l'erreur commise, sur l'ensemble de la validation et de la généralisation (figure (IV.10) et (IV.11)), montre bien que le réseau de neurones arrive à suivre l'évolution de la quantité adsorbée Q_e entre les points d'apprentissage ce qui le met en dehors des domaines de sous ou de sur apprentissage.

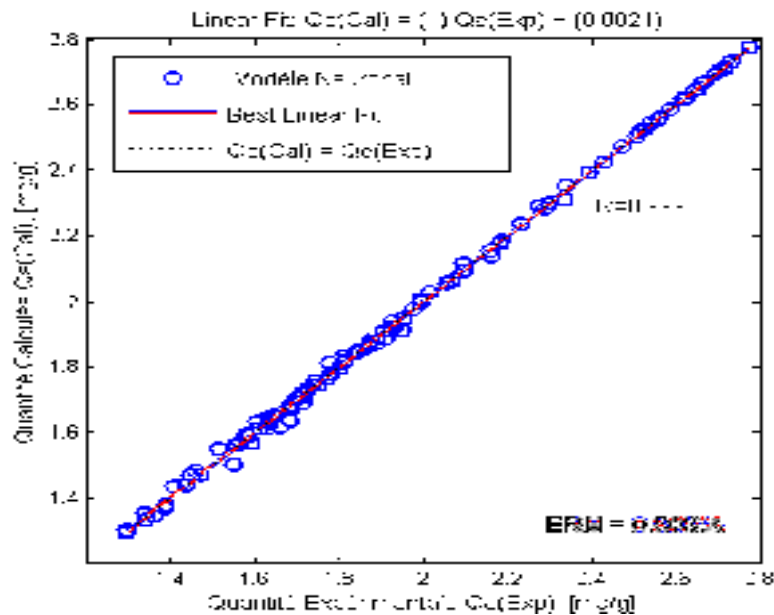


Figure IV.9 : Résultats d'apprentissage pour les valeurs de la quantité de soluté retenue par le solide (Q_e).

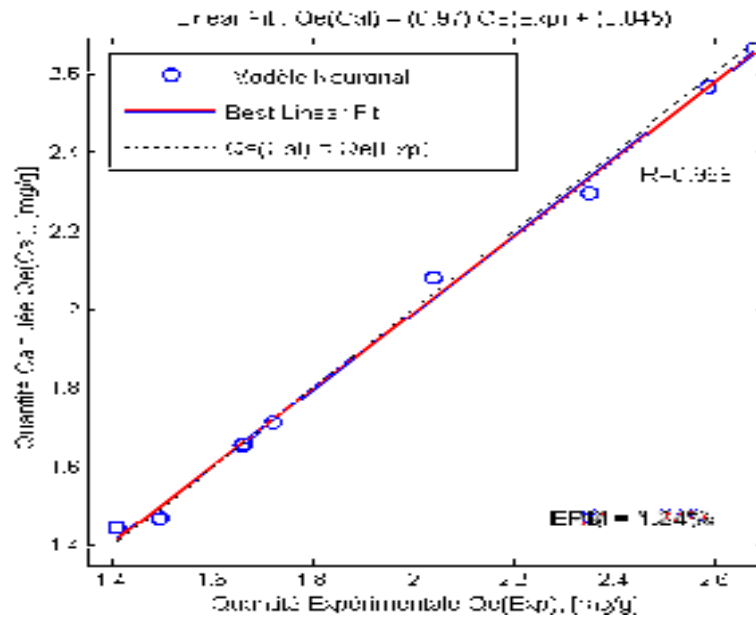


Figure IV.10 : Résultats de validation pour les valeurs de la quantité de soluté retenue par le solide (Q_e).

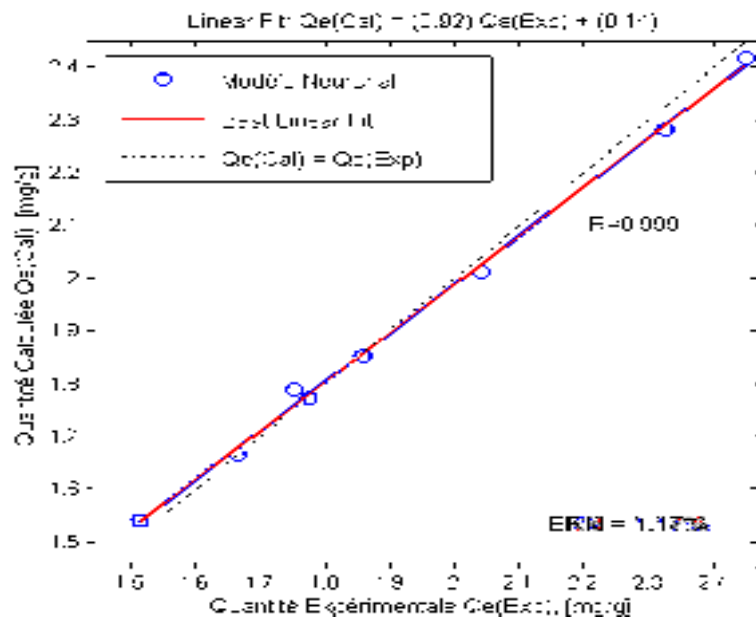


Figure IV.11 : Résultats du test de généralisation pour les valeurs de la quantité de soluté retenue par le solide (Q_e).

IV.4.2. La deuxième sortie (CEC)

Les figures (IV.12) et (IV.13) et (IV.14) présentent respectivement les résultats du calcul des valeurs des capacités d'échanges cationiques des quatre sols étudiés (CEC) et ce pour l'apprentissage, la validation et le test de généralisation. Nous remarquons dans ce cas que les erreurs relatives moyennes par rapport aux données utilisées (apprentissage,

validation et généralisation) ne dépassent pas les 0.017%. Le coefficient de corrélation R de la courbe de linéarisation pour les trois sous-ensembles est égal à l'unité.

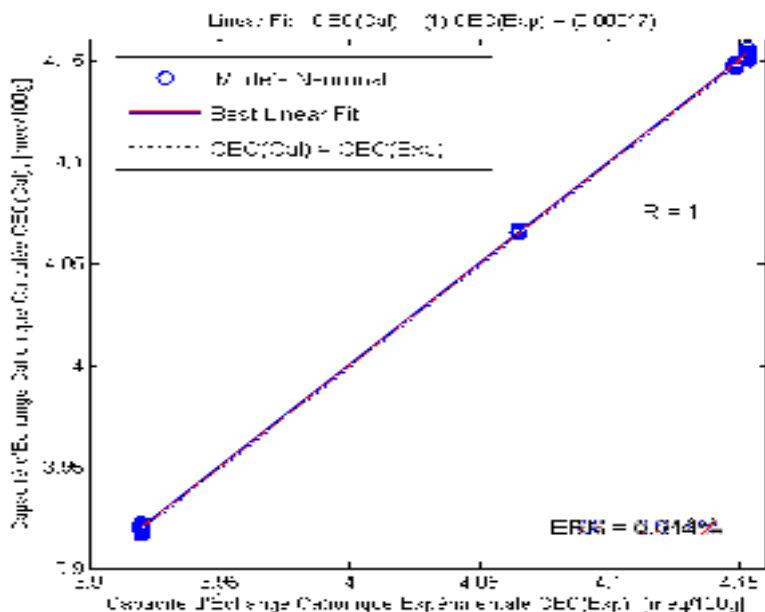


Figure IV.12 : Résultats d'apprentissage pour les valeurs de la capacité d'échange cationique (CEC).

Le modèle neuronal reproduit parfaitement les valeurs des CEC pour les trois ensembles (apprentissage, validation et également test de généralisation).

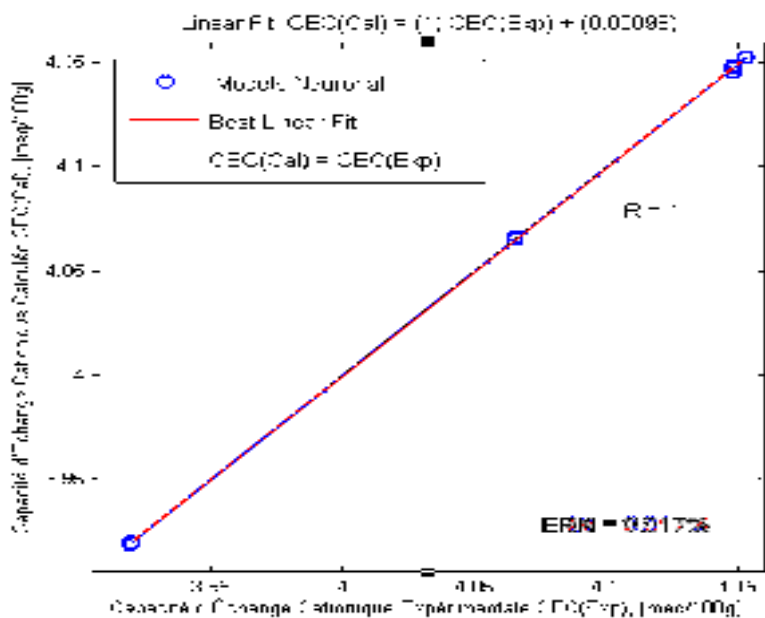


Figure IV.13 : Résultats de validation pour les valeurs de la capacité d'échange cationique (CEC).

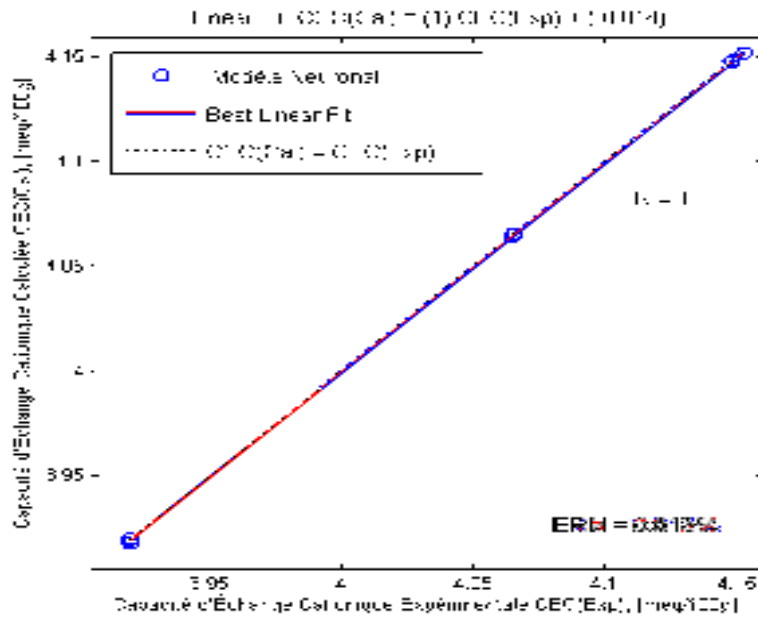


Figure IV.14 : Résultats du test de généralisation pour les valeurs de la capacité d'échange cationique (CEC).

IV.4.3. La troisième sortie (Ea)

Les figures (IV.15) et (IV.16) et (IV.17) présentent respectivement les résultats du calcul des valeurs des énergies d'activation (Ea) pour l'apprentissage, la validation et le test de généralisation. Pour ce paramètre, nous constatons aussi que, l'erreur relative moyenne par rapport aux données utilisées (apprentissage, validation et généralisation) ne dépasse pas les 0.011%. Le coefficient de corrélation R de la courbe de linéarisation pour les trois sous-ensembles est presque égal à l'unité (0.9999).

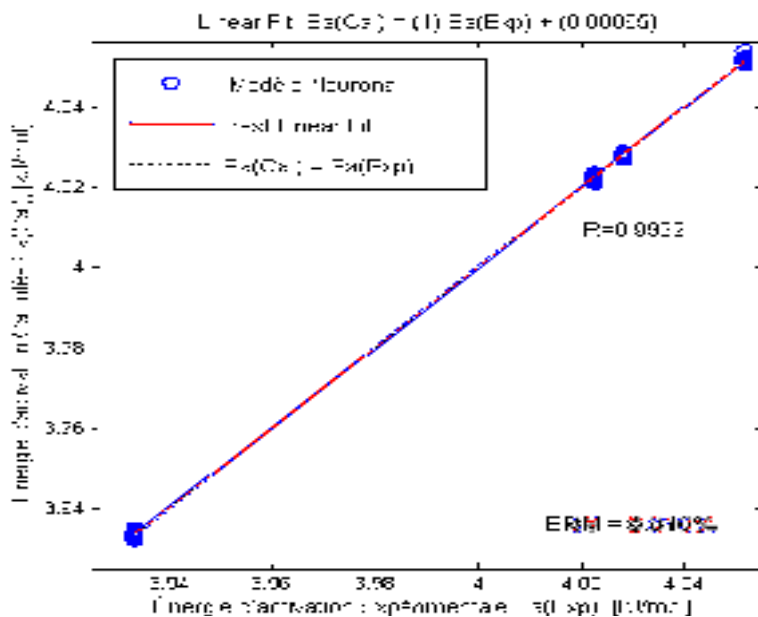


Figure IV.15 : Résultats d'apprentissage pour les valeurs de l'énergie d'activation (Ea).

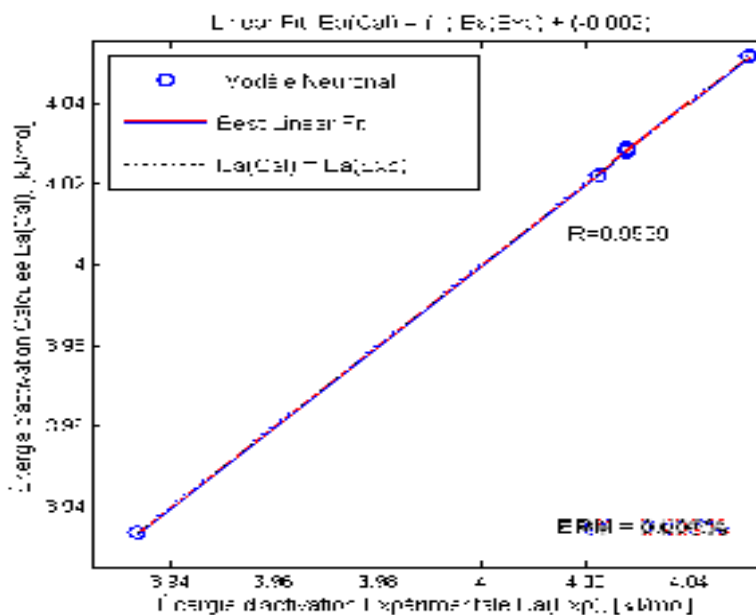


Figure IV.16 : Résultats de validation pour les valeurs de l'énergie d'activation (E_a).

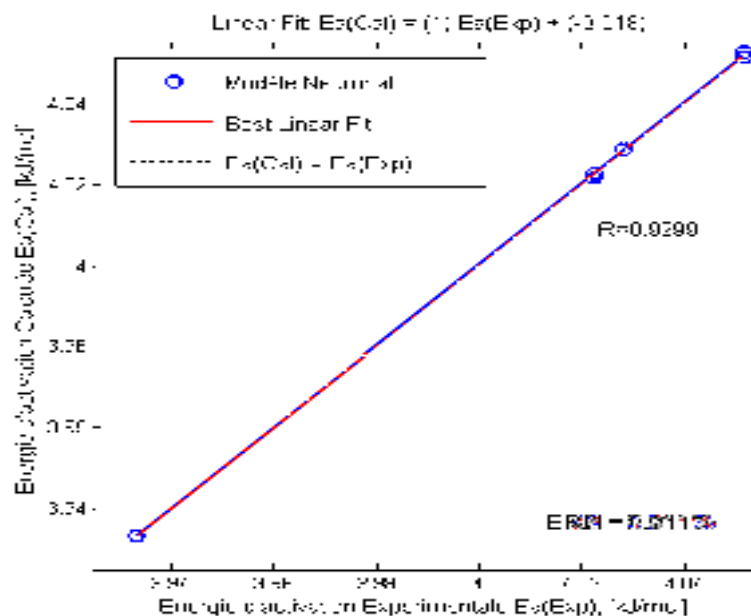


Figure IV.17 : Résultats du test de généralisation pour les valeurs de l'énergie d'activation (E_a).

Ici également le modèle neuronal reproduit parfaitement les valeurs des énergies d'activation pour les trois ensembles (apprentissage, validation et également test de généralisation). A vu des erreurs commise sur les ensembles de validation et de généralisation, il y'a pas lieu d'évoquer le problème de sous ou de sur apprentissage.

IV.4.4. La Quatrième sortie (K_d)

Pour les coefficients de distribution K_d , les figures (IV.18) et (IV.19) et (IV.20) présentent respectivement les résultats du calcul pour l'apprentissage, la validation et le

test de généralisation. Nous remarquons que les valeurs moyennes des erreurs relatives par rapport aux données utilisées (apprentissage, validation et généralisation) ne dépassent pas les 0.42%.

Le coefficient de corrélation R de la courbe de linéarisation pour les trois sous-ensembles tourne autour de la valeur 0.994.

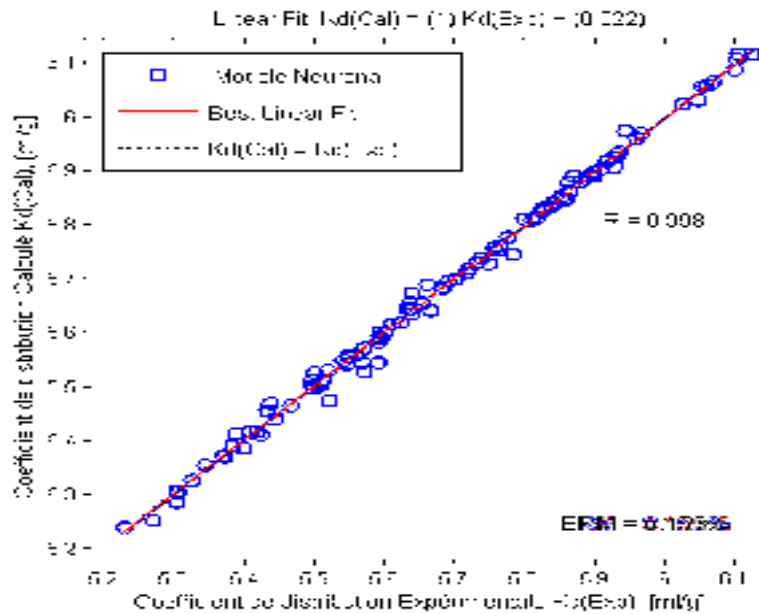


Figure IV.18 : Résultats d'apprentissage pour les valeurs du coefficient de distribution (K_d).

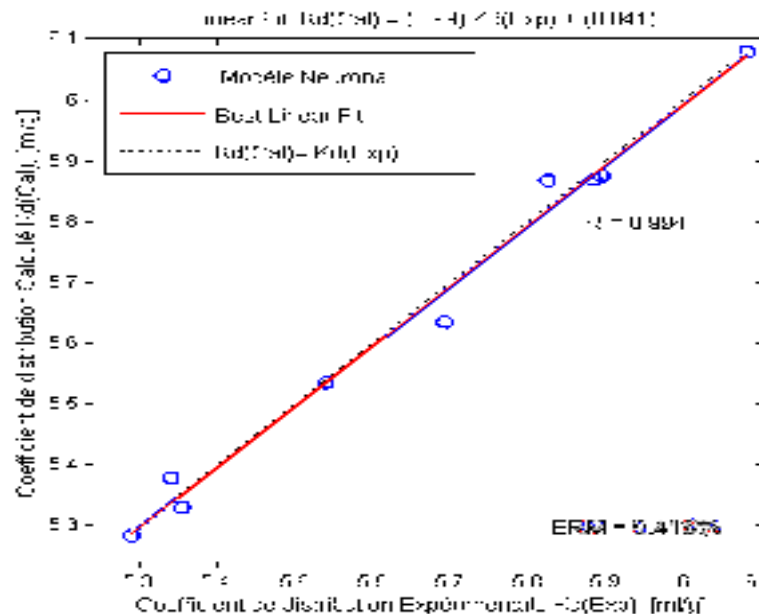


Figure IV.19 : Résultats de validation pour les valeurs du coefficient de distribution (K_d).

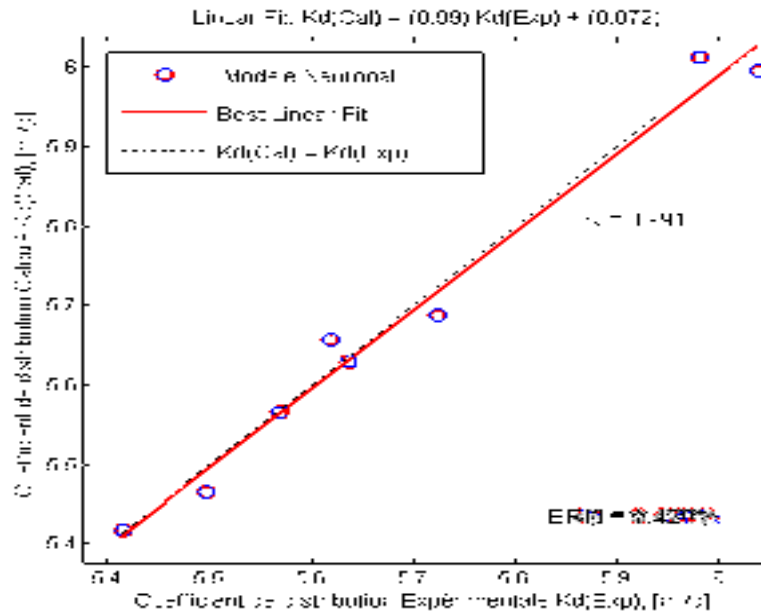


Figure IV.20 : Résultats du test de généralisation pour les valeurs du coefficient de distribution (K_d).

Là encore nous constatons que le modèle neuronal reproduit d'une manière convenable les valeurs de K_d des sols étudiés et ce pour les trois sous-ensembles (apprentissage, validation et également test de généralisation). A vu des erreurs commises sur les sous-ensembles de validation et de généralisation (0.41 et 0.42%), le réseau arrive à interpoler correctement entre les points d'apprentissage.

Les tableaux qui présentent une comparaison entre les valeurs expérimentales et les valeurs calculées à partir de réseau de neurones, de chaque sol (A, B, C et D) et ce pour les quatre paramètres (Q_e , CEC, E_a et K_d), sont présentés en annexe A. Au vu des erreurs commises, nous considérons que cette modélisation est tout à fait réussie puisqu'elle arrive à faire une bonne approximation des valeurs expérimentales.

IV.5. Comparaison des isothermes

Toujours dans le but de mesurer les performances du modèle neuronal conçu, particulièrement en ce qui concerne les isothermes qui sont, rappelons-le, le pivot de cette étude, nous l'avons confronté aux modèles proposés dans la littérature (voir chapitre II). Pour ce faire, nous avons sélectionné cinq modèles : Langmuir, Temkin, Langmuir-Freundlich, Freundlich et Radke-Prausnitz. Ces modèles ont été choisis, d'une part, sur la base de leurs fondements théoriques, et d'autre part, sur la base des paramètres qu'ils

calculent. En effet, nous avons besoins de comparer tous les paramètres calculés par le réseau de neurone à ceux des modèles théoriques.

Nous allons présenter les résultats de cette comparaison en termes de graphiques pour chaque sol séparément. On a intentionnellement séparé la présentation des modèles sur deux figures différentes, en rassemblant les modèles qui convergent mieux dans une figure et les modèles dont la convergence est moins importante dans une autre figure.

Mais avant de passer à l'interprétation des résultats de la comparaison des isothermes d'adsorption étudiées, on a jugé utile de rappeler les équations ainsi que les fondements théoriques de ces modèles (tableau IV.6).

Tableau IV.6 : Description des modèles étudiés avec leurs équations.

Nom du modèle	Equations	Description
Langmuir	$Q = Q_m \frac{KC_e}{1 + KC_e}$	Adsorption monocouche sur surface homogène (absence d'interactions adsorbat-adsorbat).
Freundlich	$Q = K_f C_e^{1/n}$	Adsorption sur surface hétérogène (sites d'adsorption énergétiquement différents). il ne repose sur aucune théorie particulière.
Freundlich- Langmuir	$Q = Q_m (K C_e)^\gamma / [1 + (K C_e)^\gamma]$	Adsorption monocouche sur surface hétérogène.
Radke-Prausnitz	$Q = Q_m \frac{KC_e}{(1 + (K C_e)^\gamma)}$	Modèle empirique d'adsorption sur surface hétérogène (développé notamment pour décrire l'adsorption sur charbon actif).
Temkin	$Q = Q_m \frac{RT}{\Delta Q} \cdot \ln (K_T C_e)$	Les adsorbants dont la surface est considérée comme énergétiquement homogène (TÓTH, 1971) [36].

IV.5.1. Sol A

Les figures IV.21 et IV.22 présentent respectivement les modélisations des valeurs expérimentales des isothermes d'adsorption par les modèles (Freundlich-Langmuir, Radke-Prausnitz et modèle neuronale), et (Langmuir, Freundlich, Temkin et modèle neuronale).

La figure IV.21 montre que le réseau de neurone suit parfaitement l'évolution des valeurs expérimentales de l'isotherme. Le coefficient de corrélation R^2 est pratiquement égal à l'unité (tableau IV.7). Tandis que, pour les deux autres modèles (Freundlich-Langmuir et Radke-Prausnitz), bien que leur convergence n'égal pas celle du réseau de neurone, elle demeure aussi importante avec des coefficients de corrélation dépassant les 0.90 (tableau IV.7). Cette convergence des modèles (Freundlich-Langmuir et Radke-Prausnitz) est justifiée par le fait que ces derniers, dans leur fondement théorique et plus particulièrement dans leurs hypothèses, se rapprochent du contexte des sols notamment à travers le paramètre surface qu'ils considèrent comme hétérogène.

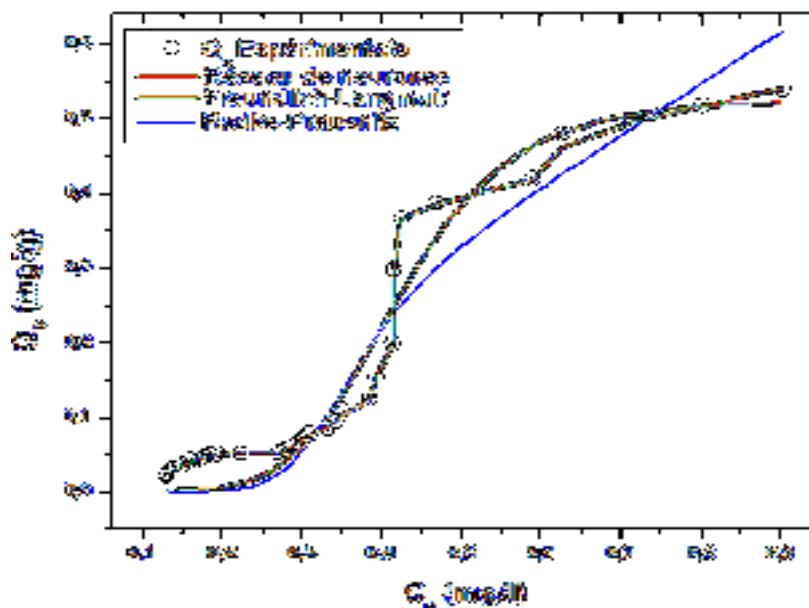


Figure VI.21 : Comparaison de l'isotherme expérimentale avec les modèles (neuronale, Freundlich-Langmuir et Radke-Prausnitz) pour le sol A.

Dans le but d'examiner les autres sorties du modèle neuronal à savoir : la CEC, l'énergie d'activation E_a et le coefficient de distribution K_d , nous avons dressé un tableau comparatif (tableau IV.7). Ce tableau comporte les valeurs expérimentales de ces paramètres et celles calculées par les modèles de Freundlich-Langmuir, de Radke-Prausnitz et du RNA.

Tableau IV.7 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol A.

Paramètre	Valeur Exp.	RNA			Freundlich-Langmuir			Radke-Prausnitz			
		Valeur	ERM (%)	R^2	Valeur	ERM (%)	R^2	Valeur	ERM (%)	R^2	
CEC (méq/100g)	14.06	14.058	0.014	0.99	0.39	97,18	0.95	0.17	98,76	0.93	
E_a (kJ/mol)	10.67	10.671	0,009		-	-		-	-		-
K_d (ml/g)	378.96	376.805	0.568		2.4^{E6}	6.2^{E5}		2.9^{E5}	7.8^{E4}		

Au vu de ces résultats, nous estimons que le RNA reproduit parfaitement les valeurs expérimentales des paramètres (CEC, E_a et K_d) avec des erreurs relatives très faibles (0.014 pour la CEC, 0,009 pour l' E_a et 0,568 pour le K_d). Alors que les modèles de Freundlich-Langmuir et Radke-Prausnitz ne reproduisent pas ces paramètres, en effet, les erreurs commises sont toutes bien au-delà de 100% atteignant même les 1000000% dans le cas du coefficient de distribution K_d . Ceci montre que le réseau de neurone développé, en plus qu'il suit parfaitement l'isotherme du sol A, arrive à reproduire avec exactitude les paramètres de ce sol (CEC, E_a et K_d) comparativement aux deux autres modèles.

Pour ce qui est des trois autres modèles (Langmuir, Temkin et Freundlich), nous remarquons (figure IV.22) effectivement que ces modèles reproduisent mal les valeurs expérimentales des isothermes particulièrement ceux de Langmuir et de Temkin avec des coefficients de corrélation avoisinant les 0.77. Ceci est en adéquation avec leurs fondements théoriques qui considèrent que les surfaces d'adsorption sont homogènes. Ce constat contredit la nature des sols qui sont des matrices géologiques à surfaces très hétérogènes. Tandis que le modèle de Freundlich présente une convergence nettement meilleure que les modèles précédents avec un R^2 de 0.87 du fait qu'il considère que la surface d'adsorption est hétérogène. Néanmoins, ces modèles demeurent largement inférieurs au modèle neuronale, en matière de convergence.

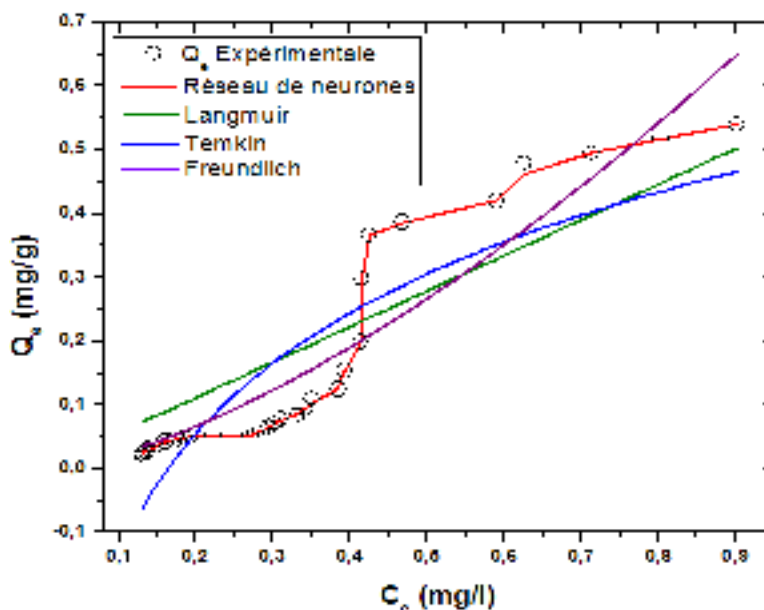


Figure VI.22 : Comparaison de l’isotherme expérimentale avec les modèles (neuronale, Langmuir, Temkin et Freundlich) pour le sol A.

Quant aux valeurs de la CEC, de l’ E_a et du K_d , nous remarquons que les modèles n’arrivent pas également à reproduire ces paramètres comparativement au modèle neuronale excepté, peut être, la valeur de la CEC calculée par le modèle de Langmuir qui se rapproche de la valeur expérimentale avec une erreur avoisinant les 23% (tableau IV.8).

Tableau IV.8 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol A.

Paramètre	Val. Exp.	RNA			Langmuir			Temkin			Freundlich		
		Val.	ERM	R ²	Val.	ERM	R ²	Val.	ERM	R ²	Val.	ERM	R ²
CEC (mék/100g)	14.06	14.058	0.014	0.99	17.3	23.11	0.76	1.02	92.72	0.77	-	-	0.87
E_a (kJ/mol)	10.67	10.671	0,009		-	-		12.07	13.14		-	-	
K_d (ml/g)	378.96	376.805	0.568		2.4 ^{E4}	0.6 ^{E4}		6,1 ^{E6}	1.6 ^{E6}		7.6 ^{E5}	2 ^{E5}	

IV.5.2. Sol B

Les mêmes étapes suivies pour l’interprétation des résultats dans le sol précédent (sol A) sont également adoptées pour ce sol, c’est-à-dire qu’on commence d’abord par la présentation des figures de la modélisation des isothermes d’adsorption pour les différents modèles ensuite, on verra les résultats des autres paramètres (CEC, E_a et K_d).

La Figure VI.23 présente la modélisation des valeurs expérimentales des isothermes d’adsorption de césium par les modèles de Freundlich-Langmuir, de Radke-Prausnitz et de réseau de neurone.

Nous remarquons, d’une part, que les deux isothermes (Freundlich-Langmuir et Radke-Prausnitz) sont identiques, et d’autre part, qu’elles arrivent à suivre les valeurs expérimentales justifiant de coefficients de corrélation de l’ordre de 0.97. Toute fois, le réseau de neurone, avec son coefficient de corrélation égale à 0.99, demeure le plus fidèle à reproduire les valeurs expérimentales des isothermes.

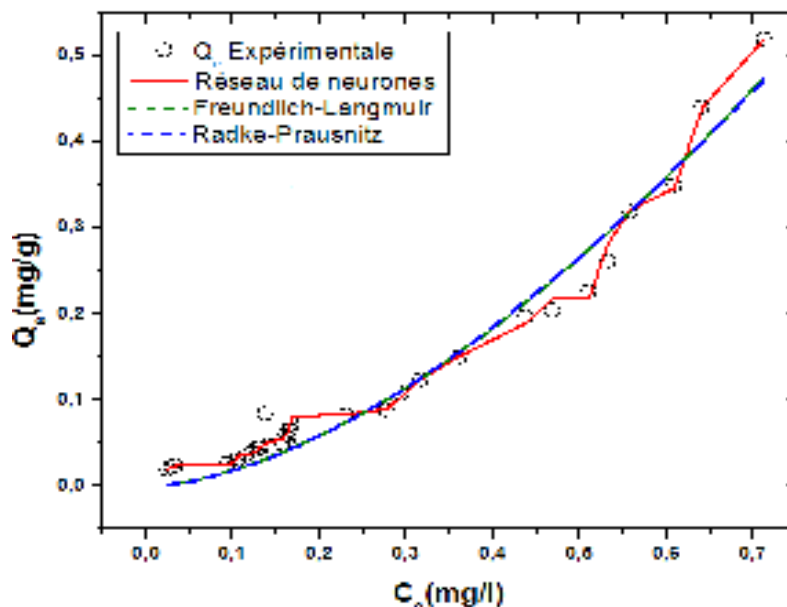


Figure VI.23 : Comparaison de l’isotherme expérimentale avec les modèles (neuronale, Freundlich-Langmuir et Radke-Prausnitz) pour le sol B.

Bien que les modèles de Freundlich-Langmuir et Radke-Prausnitz génèrent les valeurs expérimentales des isothermes d’une manière amplement acceptables, ils n’arrivent pas à reproduire parfaitement les autres paramètres (CEC et K_d) particulièrement le coefficient de distribution K_d dont les erreurs relatives atteignent des valeurs extrêmement élevées (tableau IV.9). Excepté, les valeurs de la CEC dont les erreurs sont raisonnables allant de 9.69 à 15.85%. Par contre, le RNA donne des résultats, tout à fait, proches des valeurs expérimentales avec des erreurs n’atteignant pas les 0.5%.

Tableau IV.9 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol B.

Paramètre	Valeur Exp.	RNA			Freundlich-Langmuir			Radke-Prausnitz			
		Valeur	ERM (%)	R^2	Valeur	ERM (%)	R^2	Valeur	ERM (%)	R^2	
CEC (méq/100g)	14.22	14.218	0.014	0.99	16.47	15.85	0.97	15.59	9.69	0.97	
E_a (kJ/mol)	8.58	8.5801	0.0012		-	-		-	-		-
K_d (ml/g)	445.07	442.975	0.471		1.4^{E5}	3.1^{E4}		1.7^{E5}	3.9^{E4}		

Pour la deuxième série de modèles (Langmuir, Temkin et Freundlich) et d'après la figure (Figure VI.24), nous remarquons comme précédemment que les modèles de Langmuir et de Temkin destinés pour les surfaces homogènes ne suivent pas l'évolution de l'isotherme ($R^2 = 64$ pour Temkin). En plus, nous rappelons que ces modèles stipulent dans leurs hypothèses que l'adsorption est en monocouche excluant toute affinité entre adsorbats. Alors que l'isotherme expérimentale est de type III indiquant une forte affinité adsorbat-adsorbat et conduisant ainsi à une adsorption en multicouches. Tandis que, le modèle de Freundlich, destiné pour les surfaces hétérogènes et incluant la possibilité de formation de multicouche, s'adapte bien aux valeurs expérimentales de l'isotherme ($R^2=0.97$).

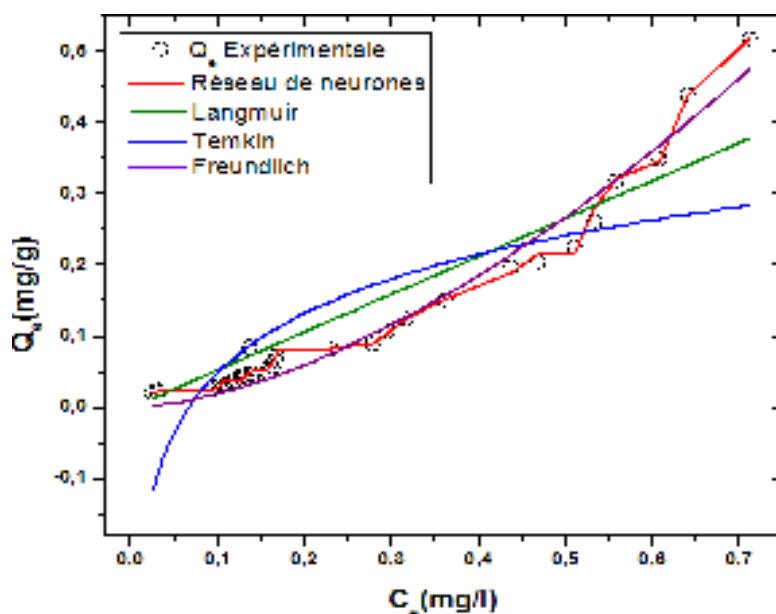


Figure VI.24 : Comparaison de l'isotherme expérimentale avec les modèles (neuronale, Langmuir, Temkin et Freundlich) pour le sol B.

Pour ce qui est des autres paramètres, présentés dans le tableau IV.10 (CEC, E_a et K_d), nous remarquons, comme précédemment, que ces modèles ne les reproduisent pas. Excepté la valeur de la CEC pour le modèle de Langmuir (27.14%) et l'énergie d'activation pour le modèle de Temkin (9,93%) qui peuvent être considérées comme plus au moins acceptables.

Tableau IV.10 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol B.

Paramètre	Val. Exp.	RNA			Langmuir			Temkin			Freundlich		
		Val.	ERM	R ²	Val.	ERM	R ²	Val.	ERM	R ²	Val.	ERM	R ²
CEC (mék/100g)	14.22	14.218	0.014	0.99	18.08	27.14	0.89	0.3	97,55	0.64	-	-	0.97
E_a (kJ/mol)	8.58	8.5801	0.001		-	-		9.4	9,93		-	-	
K_d (ml/g)	445.07	442.975	0.47		2.2 ^{E4}	0.5 ^{E4}		15.1 ^{E6}	3.4 ^{E6}		8.3 ^{E5}	1.9 ^{E5}	

IV.5.3. Sol C

La figure ci-dessous présente la modélisation de l’isotherme d’adsorption de césium par les modèles de Freundlich-Langmuir, de Radke-Prausnitz et de réseau neuronale.

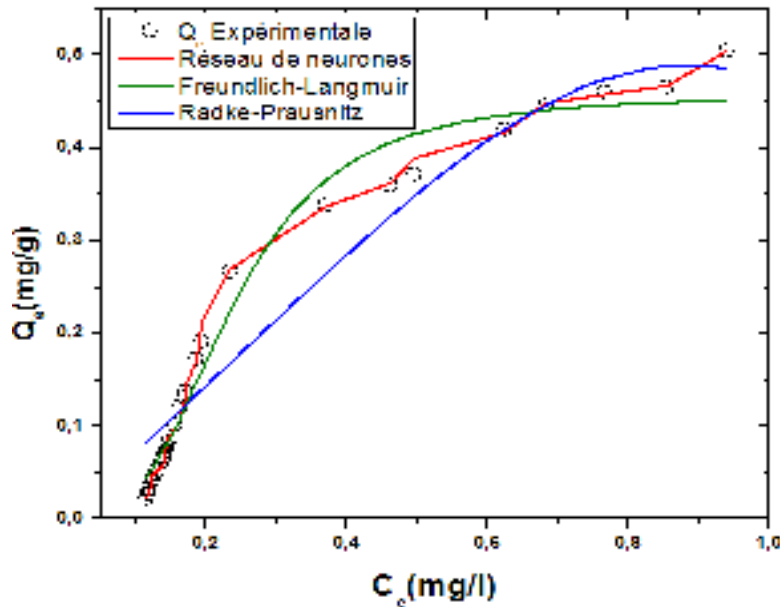


Figure VI.25 : Comparaison de l’isotherme expérimentale avec les modèles (neuronale, Freundliche-Langmuir et Radke-Prausnitz) pour le sol C.

D’après la figure VI.25, on remarque que le modèle neuronale converge parfaitement, son coefficient de corrélation avoisine l’unité. Le modèle de Freundliche-Langmuir présente également une convergence tout à fait comparable à celle du RNA avec un coefficient de corrélation de 0.98. Tandis que, le modèle de Radke-Prausnitz et au regard de son coefficient de corrélation (0.93) se classe en troisième position.

Tableau IV.11 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol C.

Paramètre	Valeur Exp.	RNA			Freundlich-Langmuir			Radke-Prausnitz			
		Val.	ERM (%)	R^2	Val.	ERM (%)	R^2	Val.	ERM (%)	R^2	
CEC (mécq/100g)	11.62	11.6204	0.003	0.99	0.34	97,07	0.98	0.63	94,54	0.93	
E_a (kJ/mol)	10.53	10.529	0.0095		-	-		-	-		-
K_d (ml/g)	575.741	574.895	0.147		4.4 ^{E6}	7.6 ^{E5}		8. 6E5	1.5 ^{E5}		

Pour ce qui des valeurs de la CEC, l’ E_a et le K_d , comme dans les autres sols, le modèle neuronal reproduit parfaitement ces paramètres. Alors que les deux autres modèles donnent des valeurs incomparables aux valeurs expérimentales (tableau IV.11).

En ce qui concerne le deuxième groupe de modèles, on remarque (figure VI.26) que le modèle de Temkin se distingue par rapport à ceux de Freundlich et Langmuir par un coefficient relativement élevé 0.96.

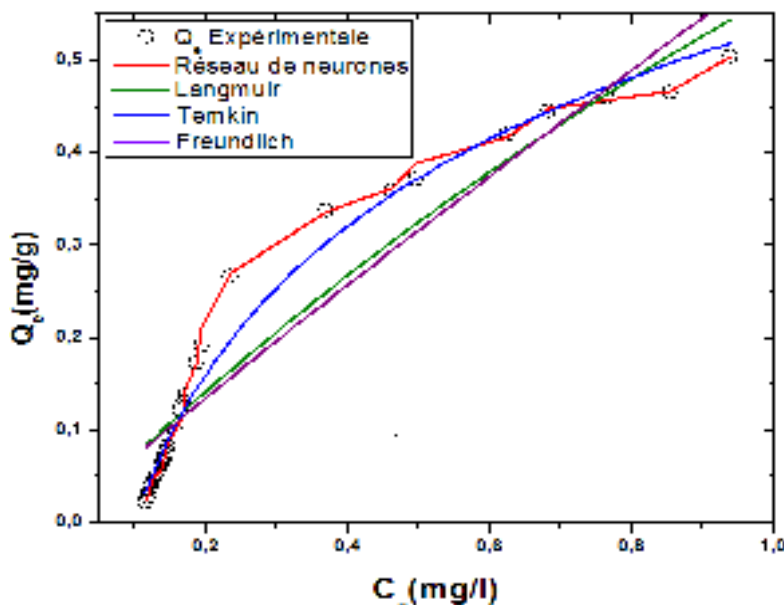


Figure VI.26 : Comparaison de l’isotherme expérimentale avec les modèles (neuronale, Langmuir, Temkin et Freundlich) pour le sol C.

Les résultats du Tableau IV.12 montrent que le modèle neuronale est toujours performant car il arrive à reproduire les paramètres (CEC, E_a et K_d) avec des erreurs ne dépassant pas les 0.15%. Tandis que le reste des modèles sont loin d’arriver à cette performance, à l’exception de la CEC pour le modèle de Langmuir (11.4%) et la valeur de l’énergie d’activation pour le modèle de Temkin (2,28%).

Tableau IV.12 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol C.

Paramètre	Val. Exp.	RNA			Langmuir			Temkin			Freundlich		
		Val.	ERM (%)	R ²	Val.	ERM (%)	R ²	Val.	ERM (%)	R ²	Val.	ERM (%)	R ²
CEC (mécq/100g)	11.62	11.6204	0.003	0.99	12.95	11,44	0.9	0.69	94,01	0.96	-	-	0.9
E_a (kJ/mol)	10.53	10.529	0.009		-	-		10.77	2,28		-	-	
K_d (ml/g)	575.741	574.895	0.147		3.7 ^E 4	6.4 ^E 3		1.1 ^E 7	2 ^E 6		6 ^E 5	1.05 ^E 5	

IV.5.4. Sol D

La figure VI.27 montre que le modèle neuronal suit pratiquement point par point l'évolution de l'isotherme. Quand aux modèles de Freundlich-Langmuir et Radke-Prausnitz, on remarque, d'une part, qu'ils sont presque superposés, et d'autre part, qu'ils reproduisent mal les valeurs expérimentales de l'isotherme ($R^2 = 0.92\%$).

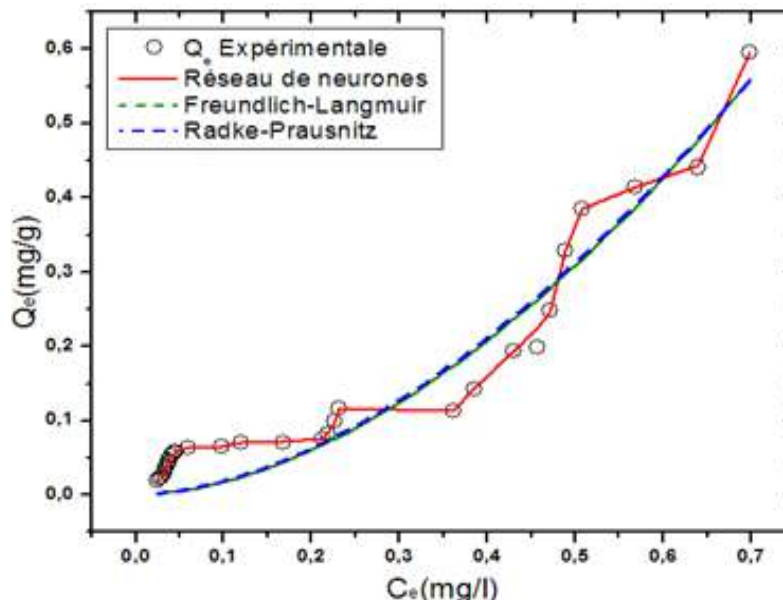


Figure VI.27 : Comparaison de l'isotherme expérimentale avec les modèles (neuronale, Freundlich-Langmuir et Radke-Prausnitz) pour le sol D.

Tableau IV.13 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol D.

Paramètre	Valeur Exp.	RNA			Freundlich-Langmuir			Radke-Prausnitz		
		valeur	ERM	R ²	valeur	ERM	R ²	valeur	ERM	R ²
CEC (mécq/100g)	8.31	8.308	0.024	0.99	10.26	23,5	0.92	10.67	28,41	0.92
E_a (kJ/mol)	11.26	11.261	0.009		-	-		-	-	
K_d (ml/g)	784.514	784.305	0.027		2.4 ^E 5	3.1 ^E 4		2.8 ^E 5	3.6E4	

Pour ce qui est des autres modèles et d’après la figure IV.28, nous remarquons que le modèle Freundlich présente un coefficient de corrélation de 0.92 indiquant une performance moyenne. Tandis que, les modèles de Langmuir et de Temkin leurs convergences est insuffisantes.

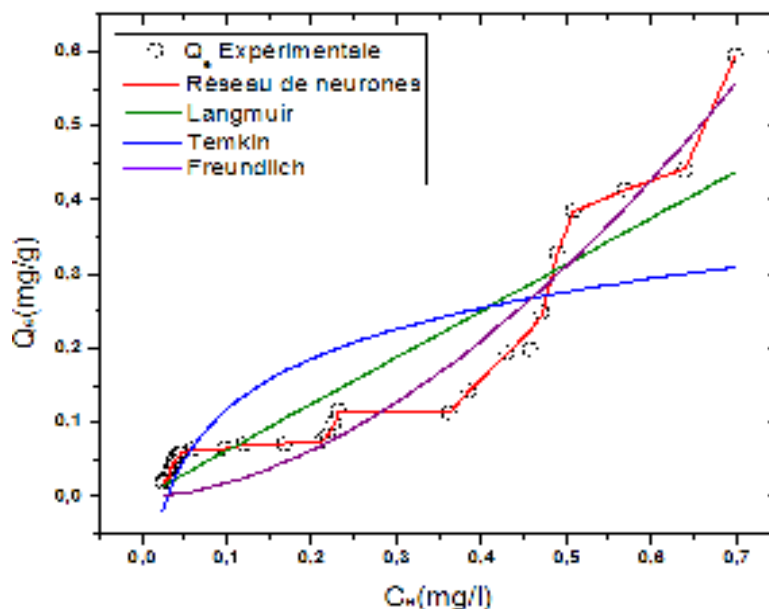


Figure VI.28 : Comparaison de l’isotherme expérimentale avec les modèles (neuronale, Langmuir, Temkin et Freundlich) pour le sol D.

En ce qui concerne les autres paramètres (CEC, E_a et K_d) et comme dans tous les sols, seul le modèle neuronal traduit parfaitement les valeurs expérimentales.

Tableau IV.14 : Etude comparative des paramètres (CEC, E_a et K_d) pour le sol D.

Paramètre	Val. Exp.	RNA			Langmuir			Temkin			Freundlich		
		Val.	ERM	R ²	Val.	ERM	R ²	Val.	ERM	R ²	Val.	ERM	R ²
CEC (méq/100g)	8.31	8.308	0.024	0.99	11.3	36,02	0.87	0.29	96,47	0.64	-	-	0.92
E_a (kJ/mol)	11.26	11.261	0.009		-	-		9.55	15,22		-	-	
K_d (ml/g)	784.51	784.305	0.027		4.3 ^{E4}	5.4 ^{E3}		3.3 ^{E7}	4.2 ^{E6}		9.5 ^{E5}	1.2 ^{E5}	

IV.6. Conclusion

A la lumière de ces résultats, nous pouvons dire que le modèle neuronal, conçu pour la modélisation des isothermes des sols étudiés, est performant. En effet, il arrive à suivre parfaitement l'évolution des isothermes. Les valeurs des coefficients de corrélation témoignent de cette performance. La confrontation du RNA aux modèles théoriques montre également que ce dernier les devance de loin en matière de convergence. Pour les sorties restantes du modèle (CEC , E_a et K_d), nous constatons qu'elles sont amplement approchées par le RNA comparativement aux autres modèles théoriques.

Sommaire

Partie Théorique

Partie Pratique

Introduction Générale

Chapitre I

Isothermes d'adsorption

Chapitre II

Réseaux de Neurones Artificiels

Chapitre III

Protocoles Expérimentaux

Chapitre IV

Résultats et Discussions

***Références
Bibliographiques***

Annexes

***Conclusion
&
Perspectives***

Nomenclature
&
Abréviation

Liste des Figures

Liste des Tableaux

Résumé

Dédicace

Remerciements

NOMENCLATURE

Caractères Latins

ΔG^0	Variation d'enthalpie libre standard	$J.mol^{-1}$
ΔH^0	Variation d'enthalpie standard	$J.mol^{-1}$
ΔS^0	Variation d'entropie standard	$J.mol^{-1}$
ΔQ	Variation d'énergie d'adsorption	$J.mol^{-1}$
a	Activité	$mol.L^{-1}$
C_0	Concentration initiale en soluté	$mg.L^{-1}$
C_e	Concentration en équilibre en soluté dans la phase liquide ou gazeuse	$mg.L^{-1}$
D	Coefficient de fréquence de collision	
D ou D_0	Diffusivité ou coefficient de diffusion effective	$m^2.s^{-1}$
E_1	Energie d'activation des réactions d'adsorption	$kJ.mol^{-1}$
E_{-1}	Energie d'activation des réactions de désorption	$kJ.mol^{-1}$
G	Energie libre d'interaction	$kJ.mol^{-1}$
H, F et p	Constantes de Jossens	
K_E	Constante de Elovich	
K_F	Constante de Freundlich, relative à la capacité d'adsorption	$L.mg^{-1}$
K_L	Constante d'équilibre de Langmuir	$mg^{(1-n)}.L^n.g^{-1}$
K_T	Constante de Temkin	$L.mg^{-1}$
K^0	Constante d'équilibre standard	$L.mg^{-1}$
k_{ads} ou k_1	Constantes de vitesse d'adsorption	
K_d	Coefficient de distribution	$L.g^{-1}$
n	Constante de Freundlich, relative à l'intensité de l'adsorption	
q_0	Concentration en soluté initialement présente sur le solide	$mg.g^{-1}$
q_e	Quantité de soluté retenue à l'équilibre par le solide	$mg.g^{-1}$
q_m	Quantité maximum de soluté retenue à l'équilibre par le solide (en monocouche pour Langmuir)	$mg.g^{-1}$
q_{mDR}	Capacité maximum d'adsorption dans les micropores	$mg.g^{-1}$
$q_{m,i}^0$	paramètre de l'isotherme mono-soluté de Langmuir de l'espèce i , à utiliser pour les modèles multi-solutés	
q_t	Concentration en soluté à chaque temps t sur le solide	$mg.g^{-1}$

T	Température absolue	<i>K</i>
V	Volume de solution	<i>L</i>
C_s	Concentration du césium dans la phase solide	<i>g kg⁻¹</i>
C_l	Concentration du césium dans la phase liquide	<i>g L⁻¹</i>
R	Taux de rétention du césium dans le sol	<i>%</i>
C_f	Concentration finale de Cs (après séparation des phases)	<i>Mg.L⁻¹</i>
V	volume de la phase liquide	<i>mL</i>
M	masse du sol	<i>g</i>
P_0	Poids du sol initial	<i>g</i>
P_1	Poids du sol après calcination	<i>g</i>
P_i	Poids du creuset (témoin) avant calcination	<i>g</i>
P_f	Poids du creuset (témoin) après calcination	<i>g</i>
CEC	Capacité d'Echange Cationique	<i>mg/100g de sol</i>
K_2	Constante de vitesse à la température T1 et T2	
C_o	Taux des carbonates	<i>%</i>
E_a	Energie d'activation	<i>kJ.mol⁻¹</i>
R^2	Coefficient de corrélation	
b_i	Biais du neurone	
d_i	Sortie désirée	
D	Vecteur de sortie désirée	
e	Erreur	
E	Fonction de coût (fonction d'erreur)	
F	Fonction d'activation du neurone	
$H(w')$	Matrice Hessienne	
I	Matrice identité	
n	Nombre d'entrées du neurone	
P	Potentiel d'action (somme pondérée)	
s_i	Sortie du neurone	
S	Vecteur de sorties du neurone	
w_i	Poids synaptiques du neurone	
w_{ij}^t	Ancien poids	
w_{ij}^{t+1}	Nouveau poids	
w^*	Poids minimum	
Δw^t	Adaptation des poids à l'itération t	

x_i	Entrées du neurone
X	Vecteur d'entrée

Caractères Grecs

δ_i	Gradient d'erreur
ε	Erreur admissible
ε_i	Variable aléatoire
η^t	Coefficient d'adaptation
θ	Seuil
μ	Pas de modification des poids (pas d'apprentissage)
μ^t	Pas de la descente du gradient
β	Constante reliée à l'énergie d'adsorption par $E = \frac{1}{(\beta)^{0.5}}$
ε_p	Potentiel de Polanyi = $RT \ln (C_s/C_e)$
θ_r	Taux de recouvrement

Indices et exposants

i	Neurone i
j	Neurone j
t	Itération ou transposé

ABREVIATION

Adaline	Ad aptative Linear Element
ANN	Artificiel Neural Network
BET	Brunauer, Emmet et Teller
BFGS	Méthode développée par « Broyden, Flecher, Goldfarb et Shanno »
CRNB	Centre de Recherche Nucléaire de Birine
ERM	Erreur Relative Moyenne
F	Freundlich
FF	Fraction Fine
F-L	Freundlich- Langmuir
L	Langmuir
LF	Limon Fin
MLP	Régression Linéaire Multi-paramètres
MLR	Multi Layer Perceptron (Perceptron multicouches)
MN	Modèle Neuronal
MO	Matière Organique
PAF	Perte Au Feu
PMC	Perceptron Multi Couches
RBF	Radial Basis Function
RN	Réseau de Neurones
RNA	Réseau de Neurones Artificiels
R-P	Radke-Prausnitz
T	Temkin

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

وقل ربّي

زدني

علما

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Tableau 5: Valeurs expérimentale de la concentration du soluté dans la phase aqueuse (Ce) en mg.L⁻¹.

<i>Concentration du soluté dans la phase aqueuse Ce (mg.L⁻¹)</i>			
<i>Sol A</i>	<i>Sol B</i>	<i>Sol C</i>	<i>Sol D</i>
0,902	0,713	0,939	0,698
0,803	0,641	0,853	0,640
0,713	0,609	0,765	0,568
0,626	0,559	0,682	0,508
0,590	0,534	0,623	0,488
0,468	0,511	0,496	0,471
0,425	0,469	0,461	0,457
0,415	0,438	0,369	0,430
0,414	0,362	0,234	0,386
0,393	0,317	0,193	0,362
0,385	0,295	0,185	0,232
0,350	0,278	0,171	0,227
0,343	0,231	0,168	0,219
0,332	0,168	0,164	0,213
0,310	0,166	0,156	0,167
0,302	0,164	0,147	0,120
0,294	0,161	0,145	0,098
0,285	0,158	0,143	0,059
0,278	0,141	0,141	0,045
0,271	0,138	0,140	0,044
0,223	0,132	0,138	0,041
0,194	0,130	0,136	0,039
0,185	0,128	0,131	0,038
0,178	0,121	0,129	0,037
0,172	0,117	0,124	0,036
0,160	0,107	0,123	0,035
0,149	0,102	0,122	0,034
0,138	0,095	0,120	0,032
0,136	0,034	0,118	0,029
0,131	0,030	0,117	0,027
0,130	0,025	0,115	0,025

ANNEXE B

Couches cachées																	
Première couche cachée					Deuxième couche cachée												
Matrice des poids				Vecteur de biais	Matrice des poids												Vecteur de biais
-64.923	-3.392	12.85	-15.130	1.7128	4.292	-0.895	-1.39	0.542	0.193	0.717	-0.1779	-2.086	1.843	-0.725	-2.6697		
					1.701	-2.407	-10.283	0.093	2.281	-1.25	8.84	1.598	-3.474	7.18	3.5501		
9.686	20.157	19.855	-13.138	-10.9779	-0.73	0.24	0.89	1.3	0.05	-0.13	-0.12	-1.36	0.88	-0.07	-2.8784		
					1.223	-0.508	-0.725	-2.155	-0.392	-0.712	0.264	1.895	0.35	-1.041	-0.1644		
100.777	8.079	16.836	-8.71	-12.3329	-1.783	-1.02	-7.191	-0.376	-2.036	2.888	4.08	-1.897	-0.814	0.203	-1.2596		
					0.365	-0.228	-0.486	-1.69	0.0064	-0.111	0.697	1.635	0.141	-3.379	0.9822		
72.252	12.524	16.304	12.618	-23.791	-2.944	5.678	0.506	-0.11	2.375	-1.64	-9.816	1.082	4.157	-2.275	1.3064		
					-0.725	1.997	4.21	-1.73	0.11	-0.06	2.836	-0.021	-0.218	2.209	2.1221		
90.229	-24.94	-21.151	-0.618	15.698	1.009	0.144	-0.774	-0.098	0.318	1.451	0.054	0.868	0.046	-0.327	-0.2043		
					7.701	2.473	3.808	2.231	3.75	-11.995	5.83	-2.942	-6.93	7.921	-1.8432		
-50.01	-39.74	1.008	-16.619	11.1759	0.526	0.502	1.74	2.343	-0.853	-2.971	1.77	-2.32	-0.525	11.402	-2.3042		
					0.917	-0.398	3.607	-0.252	-2.582	-8.632	5.84	0.695	5.344	-5.186	0.3342		
-121.174	-42.984	7.895	4.347	-5.1843	-1.864	2.903	1.4	-1.868	-0.379	1.543	2.988	0.522	1.502	1.556	1.4351		
					0.159	-0.585	-0.531	4.985	3.019	-0.886	-2.37	1.022	-2.281	2.077	0.0292		
37.516	36.28	-22.294	-4.705	20.292	-1.787	1.662	-3.636	0.166	4.874	1.658	0.408	-1.757	-1.535	3.96	0.4595		
					-9.414	2.021	3.198	5.228	16.972	1.409	12.93	-2.737	1.513	-1.389	-2.4453		
65.71	45.089	2.424	-8.354	-0.8899	-4.406	0.491	-5.159	2.006	1.036	14.332	7.02	-5.963	-0.526	28.487	-4.2242		
					2.022	-1.694	-3.008	-0.291	1.508	0.031	4.703	0.708	-0.45	-1.003	2.7762		
-98.248	32.749	-3.157	-8.081	6.3725	-0.668	0.499	1.023	-0.095	-1.82	2.038	2.079	-1.612	-0.151	1.1	-2.5938		
					-1.217	-1.728	-0.698	-1.472	-2.49	-3.359	4.315	-1.06	-1.51	-0.277	-3.1037		

119

Couche de sortie																				
Matrice des poids																				Vecteur de biais
0.731	7.363	0.812	-0.893	-4.666	-0.0048	-2.48	2.468	-2.183	-3.865	2.291	1.256	6.549	3.117	-2.487	0.472	-5.391	1.133	-0.178	2.736	-0.748
-0.022	-0.046	-0.865	1.577	0.02	1.681	0.011	-1.421	-0.041	0.058	0.082	-0.0034	1.418	-0.018	-0.095	-0.0062	0.037	-0.002	2.288	-0.009	2.2529
0.055	0.046	-2.359	0.032	-0.023	-0.919	-0.014	0.7	0.022	0.013	-0.031	0.0058	-0.651	0.02	0.039	0.004	-0.035	0.0052	1.137	0.012	1.4076
-2.543	7.276	-1.547	-0.137	-4.575	-0.698	-2.43	6.306	-0.667	-3.821	2.283	1.238	3.078	3.035	-2.498	0.426	-5.338	1.123	1.287	2.755	-0.4132

Tableau 5 : Poids et biais du modèle neuronal

ANNEXE A

Tableau 1 : Comparaison entre les valeurs expérimentale est les valeurs calculées par le réseau de neurones pour les quatre paramètres (Q_e , CEC, E_a et K_d), pour le sol A.

Quantité de soluté retenue par le solide (Q_e)			Capacité d'échange cationique (CEC)			Energie d'activation E_a			Coefficient de distribution K_d		
Valeur exp. (mg/g)	Valeur calc. (mg/g)	ER (%)	Valeur exp. (meq/100g)	Valeur calc. (meq/100g)	ER (%)	Valeur exp. (kJ/mol)	Valeur calc. (kJ/mol)	ER (%)	Valeur exp. (ml/g)	Valeur calc. (ml/g)	ER (%)
0,540	0,540	0,047	14,06	14,073	0,096	10,67	10,646	0,222	598,670	598,687	0,003
0,518	0,517	0,186	14,06	14,070	0,072	10,67	10,673	0,031	644,458	644,466	0,001
0,495	0,496	0,183	14,06	14,057	0,020	10,67	10,683	0,123	694,250	693,585	0,096
0,480	0,462	3,671	14,06	14,041	0,135	10,67	10,683	0,123	766,773	737,564	3,809
0,420	0,420	0,017	14,06	14,041	0,135	10,67	10,678	0,077	712,770	713,181	0,058
0,387	0,386	0,200	14,06	14,073	0,096	10,67	10,666	0,038	827,228	828,705	0,179
0,366	0,366	0,008	14,06	14,028	0,227	10,67	10,678	0,077	860,294	860,796	0,058
0,298	0,298	0,017	14,06	14,051	0,066	10,67	10,671	0,008	717,018	714,003	0,420
0,199	0,199	0,300	14,06	14,093	0,234	10,67	10,641	0,268	480,068	480,507	0,092
0,154	0,154	0,075	14,06	14,057	0,020	10,67	10,676	0,054	392,494	392,555	0,016
0,125	0,125	0,209	14,06	14,048	0,089	10,67	10,683	0,123	324,675	325,537	0,265
0,110	0,102	6,866	14,06	14,060	0,003	10,67	10,686	0,146	314,286	291,608	7,216
0,095	0,095	0,492	14,06	14,067	0,049	10,67	10,678	0,077	277,825	275,486	0,842
0,085	0,087	2,211	14,06	14,077	0,119	10,67	10,668	0,015	255,381	260,196	1,885
0,078	0,075	3,028	14,06	14,080	0,142	10,67	10,656	0,130	250,896	242,941	3,171
0,071	0,071	0,408	14,06	14,073	0,096	10,67	10,656	0,130	234,437	235,342	0,386
0,065	0,066	1,807	14,06	14,060	0,003	10,67	10,659	0,107	221,398	225,632	1,912
0,060	0,060	0,010	14,06	14,038	0,158	10,67	10,664	0,061	211,988	212,520	0,251

0,056	0,056	0,795	14,06	14,012	0,341	10,67	10,668	0,015	202,546	201,651	0,442	
0,053	0,052	2,055	14,06	13,986	0,525	10,67	10,676	0,054	194,781	191,558	1,655	
0,052	0,053	0,257	14,06	14,028	0,227	10,67	10,686	0,146	235,277	234,963	0,133	
0,051	0,051	0,085	14,06	14,064	0,026	10,67	10,683	0,123	262,887	262,060	0,315	
0,049	0,048	0,162	14,06	14,070	0,072	10,67	10,683	0,123	262,321	261,818	0,192	
0,046	0,046	0,566	14,06	14,073	0,096	10,67	10,681	0,100	259,675	261,577	0,732	
0,044	0,045	1,686	14,06	14,080	0,142	10,67	10,678	0,077	256,246	261,336	1,986	
0,043	0,041	3,214	14,06	14,090	0,211	10,67	10,671	0,008	265,625	258,226	2,785	
0,036	0,036	1,557	14,06	14,093	0,234	10,67	10,666	0,038	240,772	245,471	1,952	
0,031	0,029	5,672	14,06	14,070	0,072	10,67	10,664	0,061	225,673	213,059	5,589	
0,027	0,028	0,876	14,06	14,064	0,026	10,67	10,664	0,061	201,340	203,049	0,848	
0,024	0,024	3,570	14,06	14,038	0,158	10,67	10,666	0,038	186,387	179,226	3,842	
0,022	0,023	2,962	14,06	14,031	0,204	10,67	10,668	0,015	169,231	173,660	2,617	
Moy.	0,164	0,163	0,554	14,060	14,058	0,017	10,670	10,671	0,009	378,957	376,805	0,568

Tableau 2 : Comparaison entre les valeurs expérimentale est les valeurs calculées par le réseau de neurones pour les quatre paramètres (Q_e , CEC, E_a et K_d), pour le sol B.

Quantité de soluté retenue par le solide (Q_e)			Capacité d'échange cationique (CEC)			Energie d'activation E_a			Coefficient de distribution K_d		
Valeur exp. (mg/g)	Valeur calc. (mg/g)	ERM (%)	Valeur exp. (meq/100g)	Valeur calc. (meq/100g)	ERM (%)	Valeur exp. (kJ/mol)	Valeur calc. (kJ/mol)	ERM (%)	Valeur exp. (ml/g)	Valeur calc. (ml/g)	ERM (%)
0,520	0,519	0,116	14,22	14,243	0,161	8,58	8,590	0,118	729,313	730,298	0,135
0,440	0,440	0,011	14,22	14,210	0,069	8,58	8,594	0,164	686,427	683,597	0,412
0,347	0,349	0,551	14,22	14,197	0,161	8,58	8,586	0,072	569,239	572,928	0,648

0,322	0,319	1,090	14,22	14,181	0,276	8,58	8,590	0,118	576,596	566,370	1,774	
0,283	0,262	7,421	14,22	14,200	0,138	8,58	8,580	0,003	529,963	487,978	7,922	
0,219	0,226	3,468	14,22	14,223	0,023	8,58	8,570	0,112	427,733	442,079	3,354	
0,217	0,205	5,767	14,22	14,256	0,254	8,58	8,561	0,227	463,080	439,137	5,170	
0,190	0,197	3,775	14,22	14,269	0,346	8,58	8,559	0,250	432,648	452,585	4,608	
0,152	0,150	0,800	14,22	14,263	0,300	8,58	8,562	0,204	418,969	416,006	0,707	
0,125	0,124	1,147	14,22	14,210	0,069	8,58	8,582	0,026	394,322	388,329	1,520	
0,104	0,107	2,565	14,22	14,168	0,368	8,58	8,600	0,233	353,898	361,493	2,146	
0,090	0,089	0,880	14,22	14,125	0,665	8,58	8,612	0,372	323,141	319,595	1,097	
0,084	0,082	2,411	14,22	14,259	0,277	8,58	8,568	0,135	362,400	484,507	33,694	
0,081	0,072	11,601	14,22	14,230	0,069	8,58	8,574	0,066	482,887	431,023	10,740	
0,072	0,065	9,913	14,22	14,210	0,069	8,58	8,576	0,043	435,743	396,826	8,931	
0,065	0,056	14,279	14,22	14,187	0,230	8,58	8,580	0,003	398,171	354,324	11,012	
0,060	0,059	0,691	14,22	14,194	0,184	8,58	8,578	0,020	370,412	367,705	0,731	
0,055	0,049	10,382	14,22	14,204	0,115	8,58	8,580	0,003	347,574	350,348	0,798	
0,052	0,048	7,557	14,22	14,207	0,092	8,58	8,582	0,026	369,903	348,980	5,657	
0,084	0,084	0,207	14,22	14,312	0,647	8,58	8,580	0,003	362,399	363,413	0,280	
0,046	0,045	1,076	14,22	14,210	0,069	8,58	8,582	0,026	345,960	342,610	0,968	
0,043	0,044	2,722	14,22	14,210	0,069	8,58	8,582	0,026	330,288	339,703	2,851	
0,041	0,043	6,152	14,22	14,210	0,069	8,58	8,582	0,026	316,636	336,434	6,253	
0,039	0,039	0,557	14,22	14,213	0,046	8,58	8,582	0,026	320,803	322,924	0,661	
0,037	0,037	0,926	14,22	14,213	0,046	8,58	8,582	0,026	313,638	316,738	0,988	
0,036	0,032	9,973	14,22	14,220	0,000	8,58	8,580	0,003	331,776	298,676	9,977	
0,030	0,030	0,024	14,22	14,227	0,046	8,58	8,578	0,020	292,075	292,012	0,022	
0,026	0,027	5,709	14,22	14,240	0,138	8,58	8,574	0,066	270,944	286,154	5,614	
0,024	0,024	3,264	14,22	14,272	0,369	8,58	8,566	0,158	719,669	703,720	2,216	
0,022	0,023	3,545	14,22	14,230	0,069	8,58	8,576	0,043	728,704	757,879	4,004	
0,020	0,020	0,429	14,22	14,164	0,391	8,58	8,590	0,118	792,000	777,857	1,786	
Moy.	0.127	0.125	1.222	14.22	14.218	0.014	8.58	8.580	0,001	445.07	442.975	0.471

Tableau 3 : Comparaison entre les valeurs expérimentales et les valeurs calculées par le réseau de neurones pour les quatre paramètres (Q_e , CEC, E_a et K_d), pour le sol C.

Quantité de soluté retenue par le solide (Q_e)			Capacité d'échange cationique (CEC)			Energie d'activation E_a			Coefficient de distribution K_d		
Valeur exp. (mg/g)	Valeur calc. (mg/g)	ERM (%)	Valeur exp. meq/100g)	Valeur calc. (meq/100g)	ERM (%)	Valeur exp. (kJ/mol)	Valeur calc. (kJ/mol)	ERM (%)	Valeur exp. (ml/g)	Valeur calc. (ml/g)	ERM (%)
0,505	0,505	0,021	11,62	11,617	0,024	10,53	10,517	0,122	537,806	538,022	0,040
0,467	0,466	0,117	11,62	11,639	0,160	10,53	10,544	0,132	546,875	547,394	0,095
0,458	0,460	0,516	11,62	11,628	0,068	10,53	10,554	0,224	598,536	597,585	0,159
0,448	0,445	0,752	11,62	11,617	0,024	10,53	10,554	0,224	657,277	652,229	0,768
0,417	0,420	0,560	11,62	11,614	0,047	10,53	10,546	0,155	669,952	678,578	1,288
0,389	0,371	4,626	11,62	11,623	0,022	10,53	10,524	0,053	783,410	749,894	4,278
0,362	0,359	0,725	11,62	11,623	0,022	10,53	10,517	0,122	784,978	779,292	0,724
0,336	0,338	0,741	11,62	11,609	0,093	10,53	10,508	0,214	909,214	912,851	0,400
0,269	0,267	0,554	11,62	11,623	0,022	10,53	10,498	0,305	1148,148	1143,668	0,390
0,212	0,192	9,515	11,62	11,620	0,001	10,53	10,510	0,191	1097,150	991,745	9,607
0,171	0,173	1,229	11,62	11,617	0,024	10,53	10,515	0,145	924,324	934,329	1,082
0,145	0,138	4,472	11,62	11,614	0,047	10,53	10,520	0,099	846,979	807,979	4,605
0,125	0,131	4,926	11,62	11,614	0,047	10,53	10,522	0,076	741,497	777,141	4,807
0,110	0,121	10,150	11,62	11,614	0,047	10,53	10,524	0,053	667,683	734,683	10,035
0,098	0,101	2,894	11,62	11,612	0,070	10,53	10,527	0,030	629,630	647,739	2,876
0,089	0,082	8,252	11,62	11,612	0,070	10,53	10,529	0,007	606,042	556,288	8,210
0,081	0,077	4,946	11,62	11,612	0,070	10,53	10,532	0,016	561,129	533,826	4,866
0,075	0,074	1,579	11,62	11,612	0,070	10,53	10,532	0,016	522,727	514,873	1,503
0,069	0,070	1,201	11,62	11,612	0,070	10,53	10,534	0,040	490,453	496,707	1,275
0,064	0,068	6,166	11,62	11,612	0,070	10,53	10,534	0,040	459,184	487,978	6,271
0,060	0,065	7,965	11,62	11,612	0,070	10,53	10,534	0,040	435,749	470,869	8,060

	0,057	0,062	9,283	11,62	11,612	0,070	10,53	10,537	0,063	415,441	454,360	9,368
	0,053	0,054	1,481	11,62	11,614	0,047	10,53	10,539	0,086	408,172	414,477	1,544
	0,051	0,051	1,427	11,62	11,614	0,047	10,53	10,539	0,086	392,334	398,016	1,448
	0,048	0,043	9,935	11,62	11,620	0,001	10,53	10,539	0,086	388,795	350,026	9,972
	0,046	0,042	9,226	11,62	11,623	0,022	10,53	10,539	0,086	372,764	338,221	9,267
	0,038	0,040	3,793	11,62	11,623	0,022	10,53	10,539	0,086	313,525	325,312	3,760
	0,033	0,035	7,962	11,62	11,631	0,091	10,53	10,534	0,040	273,810	295,461	7,907
	0,029	0,031	6,297	11,62	11,639	0,160	10,53	10,529	0,007	244,174	259,478	6,268
	0,026	0,028	9,119	11,62	11,644	0,206	10,53	10,524	0,053	219,136	239,001	9,065
	0,023	0,022	3,680	11,62	11,660	0,345	10,53	10,515	0,145	201,087	193,731	3,658
Moy.	0,173	0,172	0,367	11,620	11,6204	0,003	10,530	10,529	0,0095	575,741	574,895	0,147

Tableau 4 : Comparaison entre les valeurs expérimentale est les valeurs calculées par le réseau de neurones pour les quatre paramètres (Q_e , CEC, E_a et K_d), pour le sol D.

Quantité de soluté retenue par le solide (Q_e)			Capacité d'échange cationique (CEC)			Energie d'activation E_a			Coefficient de distribution K_d		
Valeur exp. (mg/g)	Valeur calc. (mg/g)	ER (%)	Valeur exp. (meq/100g)	Valeur calc. (meq/100g)	ER (%)	Valeur exp. (kJ/mol)	Valeur calc. (kJ/mol)	ER (%)	Valeur exp. (ml/g)	Valeur calc. (ml/g)	ER (%)
0,595	0,596	0,111	8,31	8,291	0,230	11,26	11,262	0,014	852,436	852,315	0,014
0,443	0,441	0,256	8,31	8,327	0,207	11,26	11,246	0,124	691,406	692,468	0,154
0,415	0,415	0,035	8,31	8,337	0,323	11,26	11,251	0,078	730,634	730,971	0,046
0,386	0,385	0,269	8,31	8,314	0,046	11,26	11,267	0,060	760,335	762,606	0,299
0,329	0,329	0,170	8,31	8,291	0,230	11,26	11,275	0,129	673,837	671,583	0,334
0,247	0,248	0,520	8,31	8,281	0,345	11,26	11,272	0,106	524,719	522,757	0,374
0,225	0,198	12,074	8,31	8,289	0,253	11,26	11,262	0,014	492,341	431,320	12,394
0,194	0,193	0,314	8,31	8,304	0,069	11,26	11,259	0,009	450,168	451,336	0,259
0,144	0,143	0,128	8,31	8,320	0,115	11,26	11,249	0,101	372,193	372,735	0,145

	0,114	0,114	0,201	8,31	8,321	0,138	11,26	11,241	0,170	315,007	314,413	0,189
	0,117	0,117	0,410	8,31	8,314	0,046	11,26	11,249	0,101	504,310	500,495	0,756
	0,098	0,100	1,672	8,31	8,316	0,069	11,26	11,243	0,147	433,186	440,453	1,678
	0,085	0,083	2,299	8,31	8,318	0,092	11,26	11,238	0,193	390,085	382,737	1,884
	0,076	0,076	0,757	8,31	8,316	0,069	11,26	11,241	0,170	354,705	359,584	1,375
	0,072	0,071	1,524	8,31	8,270	0,483	11,26	11,295	0,314	432,468	426,481	1,384
	0,070	0,070	0,246	8,31	8,255	0,666	11,26	11,321	0,545	580,833	577,431	0,586
	0,065	0,066	0,453	8,31	8,308	0,023	11,26	11,282	0,199	666,976	667,882	0,136
	0,063	0,063	0,895	8,31	8,368	0,693	11,26	11,233	0,239	1061,086	1058,035	0,288
	0,059	0,059	1,395	8,31	8,320	0,115	11,26	11,262	0,014	1332,184	1308,579	1,772
	0,055	0,057	3,435	8,31	8,316	0,069	11,26	11,264	0,037	1270,115	1308,579	3,028
	0,052	0,053	1,983	8,31	8,310	0,000	11,26	11,267	0,060	1261,789	1283,217	1,698
	0,049	0,047	2,484	8,31	8,306	0,046	11,26	11,267	0,060	1256,665	1225,180	2,505
	0,046	0,046	0,503	8,31	8,306	0,046	11,26	11,267	0,060	1205,882	1200,881	0,415
	0,043	0,043	1,036	8,31	8,306	0,046	11,26	11,264	0,037	1171,171	1162,251	0,762
	0,041	0,041	0,352	8,31	8,306	0,046	11,26	11,264	0,037	1128,687	1137,104	0,746
	0,039	0,037	4,452	8,31	8,308	0,023	11,26	11,262	0,014	1117,143	1074,484	3,819
	0,033	0,035	6,279	8,31	8,308	0,023	11,26	11,262	0,014	959,559	1028,490	7,184
	0,028	0,030	6,216	8,31	8,310	0,000	11,26	11,256	0,032	876,116	940,589	7,359
	0,025	0,024	1,933	8,31	8,310	0,000	11,26	11,254	0,055	849,138	838,880	1,208
	0,022	0,022	1,287	8,31	8,306	0,046	11,26	11,256	0,032	812,757	800,756	1,477
	0,020	0,020	1,493	8,31	8,299	0,138	11,26	11,262	0,014	792,000	788,860	0,396
Moy.	0,137	0,136	0,629	8,310	8,308	0,024	11,260	11,261	0,009	784,514	784,305	0,027

Références bibliographiques

- [1] Desjardins R., Le traitement des eaux 2ème édition revue. Édition de l'école polytechnique de Montréal, (1990).
- [2] El Azzouzi M., Persistance, mobilité et photodégradation de l'imazapyr dans le sol et l'eau, Thèse d'état, (1999).
- [3] Arias M., T.Barral M., Mejuto J. C., Chemosphere J., 48 (2002) 1081–1088.
- [4] H.Rachidi M., Etude de l'adsorption de l'Octaéthylporphyrine de Nickel sur les apatites phosphocalciques. Mémoire de CEA. (1994).
- [5] Mechrafi E., Adsorption, désorption et mobilité des herbicides au contact des adsorbants organiques et minéraux. Thèse de doctorat, (2002).
- [6] El Madani M., Contribution à l'étude de l'adsorption de l'imazethapyr et de la photodégradation catalysée de l'imazethapyr et du diuron, Thèse d'état (2004).
- [7] Mekaoui M., Etude de l'adsorption, de la persistance et de la photodégradation de l'herbicide Tribenuron Méthyle (TBM) dans le sol et l'eau. Thèse d'état. (2001).
- [8] Montacer H., Contribution à l'étude du devenir de la Therbuthylazine dans le sol et l'eau. Doctorat d'état, 1999.
- [9] Senesi N., J. The science of the total environment, 123-124 (1992) 63-76.
- [10] Chiou C. T., Peters L. J., Freed V. H. J. Environmental science and technology, 11(1979) 475–478.
- [11] Karickhoff S. W., Brown D. S., Scott T. A., J. Water Research, 13 (1979) 241-248.
- [12] Perrat C., Devenir des pesticides dans les sols : Validation des méthodes d'analyse et isothermes d'adsorption. Mémoire de DEA (2001).
- [13] Bekouche S., Etude de l'adsorption du phénol sur le TiO₂, Université Mentouri Constantine, Thèse Magistère, (2003).
- [14] Gao J. P., Maguhn J., Spitzauer P., J. Water Research, 32 (1998) 1662-1672.
- [15] Pussemier L., J. Revue de l'agriculture, 31(1978) 405-411.
- [16] El M'rabet M., Contribution à l'étude de l'adsorption du carbofuran et du phénamiphos par les complexes argilo-humiques et par les sols et de la biodégradation du carbofuran. Thèse d'état, aguadir, (2002).
- [17] Seiber J. N., (Ed) .Ragsdale N.N. and Kuhr R. J., Acs. Symposium series, 336 (1987) 80-105.
- [18] Bois L., Ribes A., Petit-Ramel M., Grenier-Loustalot M.F., J. Chem. And Eco., 19 (2003) 263- 273.
- [19] Houas A., Bakir I., Ksibi M., Elaloui E., J.Chim.Phys., 96 (1996) 497-486.
- [20] Memento, technique de l'eau. 9^{ème} Edition (1989).
- [21] Bellir K., Caractérisation de la rétention du cuivre par des matériaux naturels utilisés dans l'imperméabilisation des décharges, Université Mentouri Constantine, Thèse Magistère, (2002).
- [22] Slejko E. L. and Dekker M., Adsorption Technology; A step by step approach to process evaluation and application (Chap 2, P 37) Ed M. Dekker, New York (1985).
- [23] Giles C.H., Smith D., Huitson A., A general treatment and classification of the solute adsorption isotherm I Theoretical, Colloid Interface Science, 47 (1974), 755–765.
- [24] Limousin G., Gaudet J. P., Charlet L., Szenknect S., Barthes V., Krimissa M. (2007), Sorption isotherms: A review on physical bases, modelling and measurement - Applied Geochemistry, 22, 249-275.
- [25] Hinz C., Description of sorption data with isotherm equations, Geoderma, 99 (2001) 225-243.
- [26] Brunauer S., Emmett P.H., Teller E., Adsorption of Gases in Multi-Molecular Layers, Journal of the American Chemical Society, 60 (1938) 309 – 319. Reference 26
- [27] Hamdaoui O., Naffrechoux E., Modeling of adsorption isotherms of phenol and chlorophenols onto granular activated carbon. Part II .Models with more than two parameters, Hazardous Materials, 147 (2007b) 401-411.
- [28] Brasquet C., LE CLOIREC P. (2000), Experimental data modelling using classical models and neural networks, Chemical Engineering Science , 55, 2767-2778.
- [29] Van Bemmelen J.M., Die Adsorption Verbindungen und das Adsorption vermo"gen der Ackererde. *Die Landwirtschaftlichen Versuchs-Stationen*, 35 (1888) 69–136.
- [30] Freundlich H., Kapillarchemie. Akademische Verlagsgesellschaft, Leipzig, Germany, (1909).

- [31] Halsey G.D., The role of surface heterogeneity, *Journal of Advanced Synthesis and Catalysis*, 4 (1952) 259–269.
- [32] Hamdaoui O., Naffrechoux E., Modeling of adsorption isotherms of phenol and chlorophenols onto granular activated carbon. Part I. Two-parameter models and equations allowing determination of thermodynamic parameters, *Hazardous Materials*, 147 (2007a) 381-394.
- [33] Langmuir I., The adsorption of gases on plane surfaces of glass, mica and platinum, *Journal of American Chemical Society*, 40(1918) 1361–1403.
- [34] Hall K.R., Eagleton L.C., Acrivos A., Vermeulen T., Pore and solid diffusion kinetics in fixed-bed adsorption under constant patterns conditions, *Industrial and Engineering Chemistry Fundamentals*, 5 (1966) 212-223.
- [35] Tempkin M.J., Pyzhev V., Kinetics of ammonia synthesis on promoted iron catalysts, *Acta Physicochim*, URSS, 12 (1940) 217-256.
- [36] Toth J., State equations of the solid gas interface layer, *Acta Chemistr Academia Science*, Hungaria, 69 (1971) 311-328.
- [37] Hamdaoui O., Batch study of liquid-phase adsorption of methylene blue using cedar sawdust and crushed brick, *Hazardous Materials*, B 135 (2006), 264-273.
- [38] Srivastava V.C., Swamy M.M., Malli D., Prasad B., Mishra I.M., Adsorptive removal of phenol by bagasse fly ash and activated carbon: Equilibrium, kinetics and thermodynamics, *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 272 (2006) 89-104.
- [39] Hameed B.H., Equilibrium and kinetic studies of methyl violet sorption by agricultural waste, *Journal of Hazardous Materials*, doi:10.1016/j.jhazmat.2007.10.010, (2007).
- [40] Gimbert F., Crini N.M., Renault F., Badot P.M., Crini G., Adsorption isotherm models for dye removal by cationized starch-based material in a single component system: Error analysis, *Journal of Hazardous Materials*, doi :10.1016/j.jhazmat.2007.12.072, (2008).
- [41] Elovich S.Y., Larinov O.G., Theory of adsorption from solutions of non electrolytes on solid (I) equation adsorption from solutions and the analysis of its simplest form (II), verification of the equation of adsorption isotherm from solutions, *Izv. Akad. Nauk. SSSR, Otd. Khim. Nauk*, 2 (1962) 209–216.
- [42] Kennedy L. J., Vijaya J.J., Kayalvizhi K., Sekaran G., Adsorption of phenol from aqueous solutions using mesoporous carbon prepared by two-stage process, *Chemical Engineering Journal*, 132 (2007) 279-287.
- [43] Choy K.H.H., Porter J.F., McKay G., Single and Multicomponent Equilibrium Studies for the Adsorption of Acidic Dyes on Carbon from Effluents, *Langmuir*, 20 (2004) 9646- 9656.
- [44] Özkaya B., Adsorption and desorption of phenol on activated carbon and a comparison of isotherm models, *Hazardous Materials*, 129 (2006) 158–163.
- [45] Karagag D., Modeling the mechanism, equilibrium and kinetics for the adsorption of Acid Orange 8 onto surfactant-modified clinoptilolite: The application of nonlinear regression analysis, *Dyes and Pigments*, 74 (2007) 659-664.
- [46] Ncibi M.C., Applicability of some statistical tools to predict optimum adsorption isotherm after linear and non-linear regression analysis, *Hazardous Materials*, doi:10.1016/j.jhazmat.2007.08.038, (2007).
- [47] Kumar A., Kumar S., Kumar S., Gupta D.V., Adsorption of phenol and 4-nitrophenol on granular activated carbon in basal salt medium: Equilibrium and kinetics, *Hazardous Materials*, 147 (2007) 155-166.
- [48] Sing r.K., Kumar S., Kumar A., Developemnt pf parthenium based activated carbon and its utilization for adsorptive removal of p-cresol from aqueous solution, *Journal of Hazradous Materials*, doi:10.1016/j.jhazmat.2007.11.117, (2008).
- [49] Redlich O., Peterson D.L., A useful adsorption isotherm, *Journal of Physical Chemistry*, 63, 1024-1026, (1959).
- [50] Pikaar I., Koelmans A.A., Van Noort P.C.M., Sorption of organic compounds to activated carbons. Evaluation of isotherm models, *Chemosphere*, 65 (2006) 2343- 2351.
- [51] Rudzinski W., Plazinski W., Theoretical description of the kinetics of solute adsorption at heterogeneous solid/solution interfaces on the possibility of distinguishing between the diffusional and the surface reaction kinetics models, *Journal of Applied Surface Science*, 253 (2007) 5827-5840.
- [52] Toth J., Gas- (Dampft) – Adsorption an festen oberflächen in homogener aktivität, III, *Acta Chim. Hung.*, 32 (1962) 39.
- [53] Sips R., On the structure of a catalyst surface, *Journal of Chemistry Physical*, 16 (1948) 490–495.

- [54] Jin Y., Row K.H., Adsorption isotherm of Ibuprofen on molecular imprinted polymer, *Korean J. Chem. Eng.*, 22 (2005), 264-267.
- [55] Radke C.J., Prausnitz J.M., Adsorption of organic solutions from dilute aqueous solution on activated carbon, *Industrial and Engineering Chemistry Fundamentals*, 11(1972) 445–451.
- [56] Kumar K.V., Porkodi K., Relation between some two- and three-parameter isotherms models for the sorption of methylene blue onto lemon peel, *Journal of Hazardous Materials*, 138 (2006) 633-635.
- [57] Fritz W., Schluender E.U., Simultaneous adsorption equilibrium of organic solutes in dilute aqueous solution on activated carbon, *Chemical Engineering Science*, 29 (1974) 1279–1282.
- [58] Jossens L., Prausnitz J.M., Fritz E.U., Myer A.L., Thermodynamics of multi, solute adsorption from dilute aqueous solutions, *Chemical Engineering Science*, 33 (1978) 1097–1106.
- [59] Kiselev A.V., Vapor adsorption in the formation of adsorbate molecule complexes on the surface, *Kolloid Zhur*, 20 (1958) 338–348.
- [60] Boyd G.E., Adamson A.W., Myres L.S., Kinetics of ionic exchange adsorption processes, *J. Am. Chemistry*, 69 (1947) 2836-2848.
- [61] Fowler R.H., Guggenheim E.A., *Statistical Thermodynamics*, Cambridge University Press, London, (1939) 431–450.
- [62] Coughtrey P.J., Thorne M.C., Radionuclide distribution and transport in terrestrial and aquatic ecosystems. A critical review of data (A.A. Balkema, Rotterdam), 1 (1983) 330-346.
- [63] Avery S.V., “Fte of cesium in the environment: distribution between the abiotic and biotic components of aquatic ecosystems”, *Journal of Environmental Radioactivity* 30 (1996) 139-171.
- [64] Sawhney B.L., Selective Sorption and Fixation of Cations by Clay Minerals: a Review. *Clays and Clay Minerals*, 20 (1972) 93-100.
- [65] Cornell R.M., Adsorption of Caesium on Minerals: a Review. *Journal of Radioanalytical and Nuclear Chemistry, Articles*, 171 (1993) 483-500.
- [66] Vincent I., *Eléments de mathématiques appliquées*, Sciences.ch, (2005).
- [67] Gerstner W. et Kistler W.M., *Spiking Neuron Models Single Neurons, Populations, Plasticity*, Cambridge University Press, (2002).
- [68] Hagan M.T. et Demuth H.B., *Neural network design*, PWS publishing company, USA, (1996).
- [69] Fine T.L., *Feedforward neural network methodology*, Statistics for engineering and information science, Springer-verlag New York, (1999).
- [70] Guanghui S. et al., Application of an artificial neural network in reactor thermohydraulic problem: Prediction of critical heat flux, *journal of nuclear science and technology*, 39 N°5 (2002) 564-571.
- [71] Virole B., Réseau de neurones et psychométrie, Etude prospective des applications possibles des réseaux de neurones formels dans le traitement des données psychométriques, Editions du Centre de Psychologie Appliquée – ECPA, (2001).
- [72] P.D. Mc Nelis, *Neural Networks in Finance: Gaining Predictive Edge in the Market*, Elsevier Academic Press, (2005).
- [73] Rivals I., Les réseaux de neurones formels pour le pilotage de robots mobiles FLUX, revue de l’Association amicale Les Ingénieurs SUPÉLEC, N° 178, (1996).
- [74] Dricot L. et Lison P., Conception d’un classifieur homme/femme pour la reconnaissance de visages, Université Catholique de Louvain, (2005).
- [75] Mc Culloch W.S. et Pitts W., A logical calculus of the ideas immanent in nervous activity, *Bulletin of Math. Biophysics*, 5 (1943) 115-133.
- [76] Hebb D.O., *The organisation of behavior*, Wiley, New-york, (1949).
- [77] Rosenblatt F., The perceptron : a probabilistic model for information storage and organization in the brain, *Psychological Review*, 65 (1958) 386-408.
- [78] Widrow B. et Hoff D.E., adaptative switching circuits, *IRE Western Electric Show and convention Record*, 4 (1960) 96-104.
- [79] Minsky M. et Papert S., *Perceptrons: an introduction to computational geometry*, MIT Press, expanded edition, (1988).
- [80] Hopfield J., Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences*, 79 (1982) 2554-2558.
- [81] Kohonen T., Self organized formation of topologically correct feature maps, *Biol Cybernetics*, 43 (1982)59-69.
- [82] Rumelhart D. E. et Mc Clelland. J. L., *Parallel Distributed Processing: Exploration in the MicroStructure of Cognition*, MIT Press, Cambridge, (1986).
- [83] Le Cun Y., Une procédure d’apprentissage pour réseau à seuil assymétrique, *Cognitiva* 85, Paris, (4-7 Juin 1985).

- [84] Hassani D., Méthodes des Réseaux de Neurones pour le Calcul de la Dispersion Axiale dans les Lits Fixes, Mémoire de Thèse de Magister, Centre Universitaire de Médéa, (2002-2003).
- [85] Davalo E. et P. Naïm, Des Réseaux de Neurones, Eyrolles, 2ème Edition, Mars, (1993).
- [86] Bouzidi A., Méthode des réseaux de neurones pour le calcul des coefficients pertinents des phénomènes de transport : Viscosité, Conductivité et Diffusion Moléculaire des Gaz, Mémoire de Thèse de Magister, Centre Universitaire de Médéa, (2006).
- [87] Hassani D., Méthodes des Réseaux de Neurones pour le Calcul de la Dispersion Axiale dans les Lits Fixes, Mémoire de Thèse de Magister, Centre Universitaire de Médéa, (2002-2003).
- [88] DasGupta B., Schnitger G., The Power of Approximating: a Comparison of Activation Functions, In Hanson S. J., Cowan J.D., and Giles C.L. editors, *Advances in Neural Networks Information Processing Systems*, 5 (1993) 615-622.
- [89] Mhaskar H.N., Micchelli C.A., How to Choose an Activation Function, In J.D. Cowan, G. Tesauro and J. Alspector, editors, *Advances in Neural Information Processing Systems*, 6 (1994) 319-326.
- [90] Fombellida M.R.J., Minsoul M.J.M. et Destine J.L.O., Perceptrons Multi-Couches et Fonctions d'Activation Non-Monotones, In Proc. Neuronimes'90, EC2 & Cie, Nîmes, (1990).
- [91] Parizeau M., Réseaux de neurones, Université LAVAL, (2004).
- [92] Demuth H., Beale M. et Hagan M., Matlab, Neural Network Toolbox 6, User's Guide, MathWorks, Inc., (2009).
- [93] Jain A.K. et Mao J., Artificial neural networks: A tutorial, *Cybernetics IEEE*, (1996).
- [94] Dreyfus G. et al., Réseaux de Neurones Méthodologie et Applications, Editions Eyrolles, Paris, (2002).
- [95] Rosenblatt F., Principles of Neurodynamics, Spartan Books, New York, (1962).
- [96] Hornik K., Approximation Capabilities of Multilayer Feedforward Networks, *Neural Network*, 4 (1991) 251-257.
- [97] Hinton G., Connectionist Learning Procedures, *Artificial Intelligence*, (1989).
- [98] Touzet C., Les Réseaux de Neurones Artificiels : Introduction au Connexionnisme, (1992).
- [99] Widrow B. et Hoff M.E., Adaptive Switching Circuits, IRE WESCON, New York, (1960) 96-104.
- [100] Freeman J.A. et Skapura D.M., Neural Networks Algorithms, Applications and Programming Techniques, Addison-Wesley Publishing Company, New York, (1992).
- [101] Elie F., Conception et Réalisation d'un Système Utilisant des Réseaux de Neurones pour l'Identification et la Caractérisation, à Bord de Satellites, de Signaux Transitoires de type Sifflement", Thèse de Doctorat, Université d'Orléans, (1997).
- [102] Lee Y., Hoon S. Oh et Myung Won Kim, The Effect of Initial Weights on Premature Saturation in Back-Propagation Learning, *International Joint Conference on Neural Networks*, 1 (1991) 765-770,
- [103] Burel G., Réseaux de Neurones en Traitement d'Image : des Modèles Théoriques aux Pratiques Industrielles, Thèse de doctorat de l'Université de Bretagne Occidentale, (1991).
- [104] Abdi H., Les réseaux de Neurones. Sciences et Technologies de la Connaissance, PUG, (1994).
- [105] Hassibi B. et Stork D.G., Second order Derivatives for Network Pruning : Optimal Brain Surgeon, In S. J. Hanson, J. D. Cawnan and C.L. Giles editors, *Advances in Neural Information Processing Systems* 5 (1993) 164-171.
- [106] Tsutsumi A., Wie Chen, Neural Network for Prediction of the Dynamic Heat-Transfer Rate in Bubble Columns, *Ind. Eng. Chem. Res*, 40 (2001) 5358-5361.
- [107] Watrous R.L., Learning Algorithms for Connectionist Networks: Applied Gradient Methods for Nonlinear Optimization, In *Proceedings of the 1st International Conference on Neural Networks*, San Diego, IEEE, Press, New York, 2 (1987) 619-627.
- [108] Duprat A., Huynh T., Dreyfus G., Towards a Principled Methodology for Neural Network Design and Performance Evaluation in QSAR; Application to the Prediction of LogP, *Journal of Chemical Information and Computer Sciences*, 38 (1998) 586-594.
- [109] Levenberg K., A Method for the Solution of Certain Non-linear Problems in Least Squares, *Quarterly Journal of Applied Mathematics* II, (1944) 164-168.
- [110] Marquardt D.W., an Algorithm for Least-Squares Estimation of Non-linear Parameters, *Journal of the Society of Industrial and Applied Mathematics*, 11 (1963) 431-441.
- [111] Thiria S., Lechevallier Y., Gascuel O. et Canu S., *Statistiques et Méthodes Neuronales*, Dunod, Paris, (1997).
- [112] Boucheron S., *Théorie de l'Apprentissage*, Hermès, Paris, (1992).
- [113] Tibshirani R., A Comparison of Some Error Estimates for Neural Network Models", *Neural Computation*, 8 (1996) 152-163.
- [114] Psychogios D.C. et Ungar L.H., A hybrid neural network-first principles approach to process modelling, *Aiche J.*, 38 (1992) 2269-2276.

- [115] Pansu M., Gautheyrou J., Handbook of Soil Analysis Mineralogical, Organic and Inorganic Methods, Springer, Netherlands, (2006), 15–48 and 593– 604.
- [116] Duchaufour P., Abrégé de Pédologie : sol, végétation, environnement, MASSON 5è f1C Edition (1997) 291.
- [117] Feller C., Burtin G., Gérard B. et Balesdent J., Utilisation des résines sodiques et des ultrasons dans le fractionnement granulométrique de la matière organique des sols .intérêt et limites, sciences du sol, 29 (1991) 77-93.
- [118] Gri N., étude du comportement du ^{137}Cs et du ^{90}Sr - dans les sols contaminés de la région de Tchernobyl, Thèse de Doctorat, (1999), CEA, France.
- [119] Maurice Bonneau et Bernard Souchier, constituants et propriétés du sol, France, (1994) 621.
- [120] Bouzidi A., Souahi F., Hanini S., Sorption behavior of cesium on Ain Oussera soil under different physicochemical conditions, Journal of Hazardous Materials 184 (2010) 640–646.
- [121] Lu M. L., Mcgreavy C. et Kam E. K. T., Prediction of Thermal Conductivity of Pure Liquids and Mixtures Using Neural Network, Journal of Chemical Engineering of Japan, 3 (1997) 30.
- [122] Lee M. J., Hwang S. M. et Chen J. T., Density and Viscosity Calculations for Polar Solutions via Neural Networks, Journal of Chemical Engineering of Japan, 6 (1994) 27.
- [123] Homer J. et al. , Artificial Neural Networks for Prediction of Liquids Viscosity, Density, Heat of Vaporisation, Boiling Point and Pitzer’s Acentric Factor, Phys. Chem. Chem. Phys., 1 (1999) 4075-4081.
- [124] Rai P. et al., Prediction of Viscosity of Clarified Fruit Juice Using Artificial Neural Network: A Combined Effect of Concentration and Temperature, Journal of Food Engineering, 68 (2005) 527-533.
- [125] Scalabrin G. et Cristofoli G., The Viscosity Surface of Propane in the Form of Multilayer Feed Forward Neural networks, International Journal of Thermophysics, 5 (2003) 24.
- [126] Scalabrin G., Corbetti C. et Cristofoli G., A Viscosity Equation of State for R123 in the Form of Multilayer Feed Forward Neural networks, International Journal of Thermophysics, 5 (2001) 22.
- [127] Scalabrin G. et Cristofoli G., The Viscosity Surfaces of R152a in the Form a Multilayer Feed Forward Neural Networks, International Journal of Refrigeration, 26 (2003) 302-314.
- [128] Cristofoli G., Piazza L., Scalabrin G., A Viscosity Equation of State For R134a Through a Multilayer Feed Forward Neural Networks, Elsevier Science, Fluid Phase Equilibria, 199 (2002) 223-236.
- [129] Gosse A., Déroulède A., et Dutheuil C., Propriétés de Transport des Gaz à Pression Modérée; Viscosité, Conductivité Thermique, Coefficient de Diffusion, Données Thermodynamiques des Fluides, Banques et Bases de Données en Thermodynamique, Technique de l'Ingénieur, Paris, (1991), K425 (1-17), K530 (1-17) et K535 (122).
- [130] Division Scientifique de L’AIR LIQUIDE, Encyclopédie des Gaz, Elsevier, Netherlands, (1976)

CONCLUSION ET PERSPECTIVES

Les réseaux de neurones formels ou artificiels envahissent tous les domaines de notre vie, non seulement en matière de modélisation statique mais aussi en modélisation dynamique des processus évolutifs dans le temps, la reconnaissance de forme, le contrôle non destructif, le filtrage d'information textuelle, la bio-ingénierie, la formulation de matériaux nouveaux, la modélisation des procédés industriels, contrôle de l'environnement, l'interprétation des spectres, la classification des images (satellites), etc. Tous les cursus relatifs au traitement de l'information dans les universités ou grandes écoles consacrent une place de plus en plus importante à ces techniques. La pénétration de celles-ci dans l'industrie des pays développés comme les Etats-Unis, la Grande Bretagne, La France et les pays du nord en général, ne cesse d'augmenter.

Ces techniques neuronales se sont donc imposées par leurs simplicités relatives et leurs efficacités irréfutables. Choses que, nous avons perçues et vérifiées de près à travers ce mémoire. En effet la modélisation des isothermes des sols étudiés ainsi que le calcul des propriétés en relation de la sorption du césium dans les sol (CEC, Ea et Kd), que nous avons effectué par une technique basée sur les réseaux de neurones multicouches, a amplement montré ses performances. Les différents tests effectués sur le modèle neuronale conçu ont permis de mettre en évidence une très bonne concordance des résultats obtenus à ceux fournis par les exemples expérimentaux. Les erreurs relatives commises sont très acceptables. Des études comparatives effectuées entre le RNA conçu les modèles proposés dans la littérature ont également apporté des preuves tangibles quant ses performances.

A la lumière de ces résultats tout à fait remarquables, nous pouvons dire que ce modèle neuronale est un outil parfaitement fiables et performants, pour la modélisation des isothermes des sols de la région du Birine aux alentours de l'installation nucléaire Es-Salam. Il faut rappeler que ce modèle neuronal est conçu spécifiquement à ces sols, son application pour d'autres type sols n'a pas été envisagé dans ce projet de mémoire. Néanmoins, il est toujours possible d'étendre son champs d'application, quand on dispose de suffisamment de données.

En perspective, tout en estimant avoir atteint notre objectif tracé pour cette étude, nous envisageons de développer des modèles neuronaux globaux qui peuvent répondre à une variété plus large des sols. Toujours dans cette perspective, il serait souhaitable d'essayer de concevoir pour chaque catégorie de sols un unique modèle capable de reproduire leurs isothermes. C'est un défi ambitieux mais tout à fait envisageable, néanmoins il demande beaucoup de travail et surtout énormément de données.